A novel approach to dynamic flux balance analysis that accounts for the dynamic transfer of information by internal metabolites

Kazuki Iizuka

PhD

University of York

Biology

December 2016

Abstract

Understanding the dynamics of information feedback amongst components of complex biological systems is crucial to the success of engineering desirable metabolic phenotypes. Flux Balance Analysis (FBA) is a structural metabolic modelling procedure that allows for local topological constraints to be related to steady-state global behaviors of metabolic systems. A vast majority of biological systems of interest, such as microbial communities, however do not exist under steady-state conditions. Therefore, extending FBA methods to the dynamical setting has been a major challenge to metabolic modelling. In dynamic FBA (dFBA), the representation of feedback dynamics is made possible by combining the methods of FBA with those of Ordinary Differential Equations (ODE). Although numerous dFBA models have been constructed to date, very little effort has gone into the theoretical analysis of how static FBA models and dynamic ODE models should be combined in dFBA. To develop a better understanding of the mathematical structure of dFBA, we investigate the properties of FBA. In order to predict time-derivatives of population growth, every dFBA model must make the assumption that the underlying metabolic network modeled via FBA optimizes a phenotypic function of growth rate. We show however, that under certain circumstances, this requirement introduces a rigid correspondence between growth rate, and a related quantity, the growth yield. The consequence of this is that the dFBA models become rigid in its predictions, effectively becoming a near-static representation of metabolism. In this thesis, we show that this tight correspondence between yield and rate may be broken by combining two inversely related approaches to formulating the FBA problem.

Table of Contents:

List of Tables:

List of Figures:

List of Accompanying Materials:

An electronic appendix of custom codes

Preface:

The aims of the work that is presented in this document are two fold, but are related by the common goal of making improvements to the way by which dynamic FBA (dFBA) methods are commonly implemented. Specifically, we will argue that dFBA models that represent biomass compositions statically are too rigid in their model predictions, and are thus of limited use for representing the dynamics of real biological systems. The properties of FBA models are first investigated. From this investigation, we identify the cause of this rigidity to be related to the use of biomass objective functions in predicting optimal FBA flux distributions. Biomass yields and growth rates will be shown to become tightly coupled variables in FBA models when a biomass objective is used. Hypothesizing from this that the construction of more dynamically significant dFBA models will become possible, provided that the FBA correspondence between yield and rate can be broken, we investigate whether the dFBA problem may be formulated in such a way that these two variables decouple. We will find that by combining two inverse approaches to formulating the FBA problem, that such a decoupling is possible, and demonstrate that dFBA models that are defined with this FBA formulation predict dynamical trajectories that are not possible with simpler dFBA methods.

This thesis is written in eight chapters starting with two introductory chapters (chapter 1 and chapter 2) and ending with a concluding chapter where we discuss our research process in a broader framework (chapter 8). The introductory chapters are divided into a chapter dealing with general biological topics (chapter 1), and a chapter dealing with the mathematics of structural metabolic modelling (chapter 2). In chapter 3, we investigate the biomass objective and identify two inverse approaches to formulating an FBA problem based on whether composition is treated as a model parameter or a model prediction. These approaches are referred to respectively as the "Palsson-style" and the "Fell-style" formulation of FBA. In chapter 4, we investigate the Palsson-style method and demonstrate mathematically that biomass yields and growth rates are identical cellular objectives under this FBA formulation. The consequences of this correspondence to FBA predictions are investigated in chapter 5. Chapter 6 is an experimental chapter where we characterize growth phenotypic differences amongst wild type and mutant *E.coli*. In chapter 7, we use our results from the previous chapters to develop a novel dFBA formulation that combines the Fell and Palsson style FBA approaches to decouple the yield-rate correspondence that introduces rigidity in traditional dFBA models.

Acknowledgements:

This work would not have been possible without the continued guidance of my supervisors, Dr A. Jamie Wood and Dr Gavin H. Thomas. Their willingness to share their expertise and to communicate using two fundamentally different languages – mathematics, and biology – has made me a better systems biologist. For this, and their limitless patients which I have tested over these past four years, I cannot pay enough gratitude.

I would also like to thank Dr Jon W. Pitchford and Professor Simon J. McQueen-Mason for serving on my Thesis Advisory Panel; I am grateful to Dr Pitchford for serving also as my thesis examiner and similarly to Professor David Fell. Our conversations not only gave me invaluable insights into microbiology, metabolic modelling, and research practice in general; they encouraged me to stay on track and to move this project towards a logical and satisfying conclusion through constructive criticism.

Finally, I would like to thank my family and friends in, Japan, America, and the United Kingdom, for the support they have continuously provided me with. They have been the source of the emotional and the intellectual strength and discipline with which I have been able to complete this research.

Declaration:

I declare that this thesis is a presentation of original work and the I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

CHAPTER 1: Biology and modelling of *Escherichia coli*

Abstract:

This chapter is an introduction to the biology and the modelling of the microbe, *Escherichia coli*. We focus specifically on those aspects of this organism that are relevant to bioengineering, and summarize the ways in which *E.coli* are characterized computationally in models that aid in theoretical studies of metabolism and growth.

1.1: Plant-based biofuels, xylose, and microbial engineering:

It is currently widely accepted, in light of diminishing fossil-fuel reserves and of an increasing social awareness from the public of the damaging environmental effects of a petroleum-based energy economy, that renewable energy sources must replace currently existing unsustainable practices in the near future (Schmidt and Dauenhauer 2007, Fritsche, Sims et al. 2010, Liu and Khosla 2010, Naik, Goud et al. 2010). Liquid transportation fuels derived from the conversion of plant biomass, are a highly desirable alternative to petroleum for a number of reasons. Carbon-neutrality for instance refers to the property of the former that there are only small net increases in greenhouse gasses resulting from its combustion due to recycling from subsequent phototrophic biosynthesis (Cheng and Timilsina 2011). Similarly biofuels exhibit advantages over other alternative forms of renewable energy sources. Bioethanol is a liquid fuel like petroleum. Therefore existing infrastructure for fuel distribution offers a potential for a smoother transition to a renewable biofuel energy economy than with alternatives such as wind and solar energies (Lee, Chou et al. 2008, Liu and Khosla 2010). In order to compete as a realistic alternative source of energy however, the production of plant-based biofuels must be viable economically; in particular the processes of degradation of biomass to monomers and the subsequent metabolism of the said monomers must be optimized (Hill, Nelson et al. 2006, Schmidt and Dauenhauer 2007, Liu and Khosla 2010). With advances in microbial molecular biology platforms in the recent decades, the engineering of biocatalysts capable of efficient conversion of depolymerized

plant matter to fuel compounds has received considerable attention (Hernández-Montalvo, Valle et al. 2001, Tao, Gonzalez et al. 2001, El-Mansi 2004, Lee, Chou et al. 2008, Singh, Soh et al. 2011, Liao, Mi et al. 2016).

The pentose sugar D-Xylose occurs naturally in lignocellulosic biomass. It is a monosaccharide having the molecular formula $C_5H_{10}O_5$ and a molar mass of 150.13 $g \cdot mol^{-1}$. Xylose, glucose and arabinose, are the three most terrestrially abundant plant breakdown products from which biofuels such as ethanol may be derived (Groff, Benke et al. 2012). Studies of microbial (bacterial and yeast) utilization of xylose as a carbon source have been carried out well before attentions shifted towards xylose (xylan) as a potential raw material for biofuel production (Davis and Henderson 1987, Sumiya and Jf 1989). Building on these works the development of microbial metabolic systems for the conversion of xylose to commercially useful compounds is recognized today as an important goal of synthetic biology and engineering.

1.2: *E.coli* as a model system:

Escherichia coli is a Gram negative facultatively anaerobic bacteria. With most strains naturally occurring in the lower intestine of mammals, *E.coli* as a taxon represents a highly diverse group of genetically distinguishable organisms whose members uniformly share merely 20% of their genes with one another (Lukjancenko, Wassenaar et al. 2010). Since around 1920, when the first *E.coli* culture was deposited in the National Collection of Type Cultures (NCTC) of the United Kingdom, and 1922 when the strain most commonly used for genetic analyses K-12 was isolated in the Unites States, extensive effort has gone into elucidating the molecular biology of *E.coli* (Daegelen, Studier et al. 2009). The complete genome of *E.coli* K-12 composed of roughly 4500 genes was sequenced in 1997 (Blattner, Plunkett et al. 1997, Alon 2006). *E.coli* strain K-12 is an established and widely

used model organism for molecular biological and genetic engineering studies owing greatly to its amenability to gene manipulation (Lee 1996). The extensive curation of the information available from over 30,224 publications (as of March, 2013) on the K-12 strain is an ongoing effort and has culminated in Ecocyc, a comprehensive bioinformatic database supporting *E.coli* research (Karp, Riley et al. 2002, Keseler, Collado-Vides et al. 2005). A library of non-essential single gene deletion strains of *E.coli* K-12 (BW25113 derivatives) is maintained at the Keio University (the Keio Collection) and is available for distribution (Baba, Ara et al. 2006). Likewise, the ASKA library (A Complete Set of *E.coli* K-12 ORF Archive) also maintained in Japan makes available a set of open reading frames of *E.coli* K-12 (W3110 derivatives) cloned into the high copy number plasmid pCA24N (Kitagawa, Ara et al. 2006).

E.coli is both a simple and challenging organism to study from the perspective of quantitative and systems biology. While the malleability of *E.coli* combined with common biochemical platforms afford an experimental system within which hypotheses conjectured from theory may be tested, *E.coli* cells are also capable of appropriately adapting their physiology to an exceptionally wide range of environmental conditions. Thus the information flows across the genetic regulatory circuitry in an *E.coli* cell may be substantially more involved than in any individual cell of a multicellular organism that has specialized in function within a distribution of labour.

1.3: Xylose transport and metabolism by *E.coli*:

The uptake of external D-xylose by *E.coli* is facilitated by one of at least two main transport systems. First the sugar transport may be energized by the proton motive force across the inner membrane in a process catalysed by the proton symporter XylE (Sumiya and Jf 1989). The existence of a proton-symport process responsible for xylose transport was first proposed by Lam et al. and has subsequently confirmed via genetic analysis by Davis

and Henderson (Lam, Daruwalla et al. 1980, Davis and Henderson 1987). The second possible avenue for xylose transport is through ATP-energized catalysis by a xylose- ATP-binding cassette transporter XylFGH. First studied in 1981, the functional xylose-ABC transporter consists of five protein domains – two each of a cytoplasmic ATP-binding and a membrane-bound channel domain, together with a periplasmic substrate binding domain; the protein subunits are respectively XylG, XylH, and XylF (Ahlem, Huisman et al. 1982, Sumiya, Davis et al. 1995, Song and Park 1997). Like the substrate binding domain of many other importers of the ABC family of transporters, XylF is expressed in the periplasm to search for substrate diffusing from the external media through the outer membrane. Periplasmic xylose bound by XylF is brought to XylH where the sugar is transported across the inner membrane via a conformational change in XylH affected through ATP hydrolysis/hydrolyses by the XylG protein (Sofia, Burland et al. 1994, Sumiya, Davis et al. 1995). Based on reported Km values for the two transport processes – 0.2-4.0μM for the ABC transporter and 63.0-169μM for the symport transporter – it can be concluded that whereas the xylose ABC transporter is a high energy high affinity transporter, the xylose symporter is a low energy low affinity transporter (Sumiya, Davis et al. 1995). Therefore, it would seem reasonable to suspect that at higher concentrations of external substrate (on the 1 mM order as commonly encountered in the laboratory) that the symport system would intuitively appear to be the more important. The relative importance to *E.coli* growth of the two transport processes for xylose however has not yet been established as evidenced by the lack of a consensus in the literature (Tao, Gonzalez et al. 2001, Hasona, Kim et al. 2004, Khankal, Chin et al. 2008).

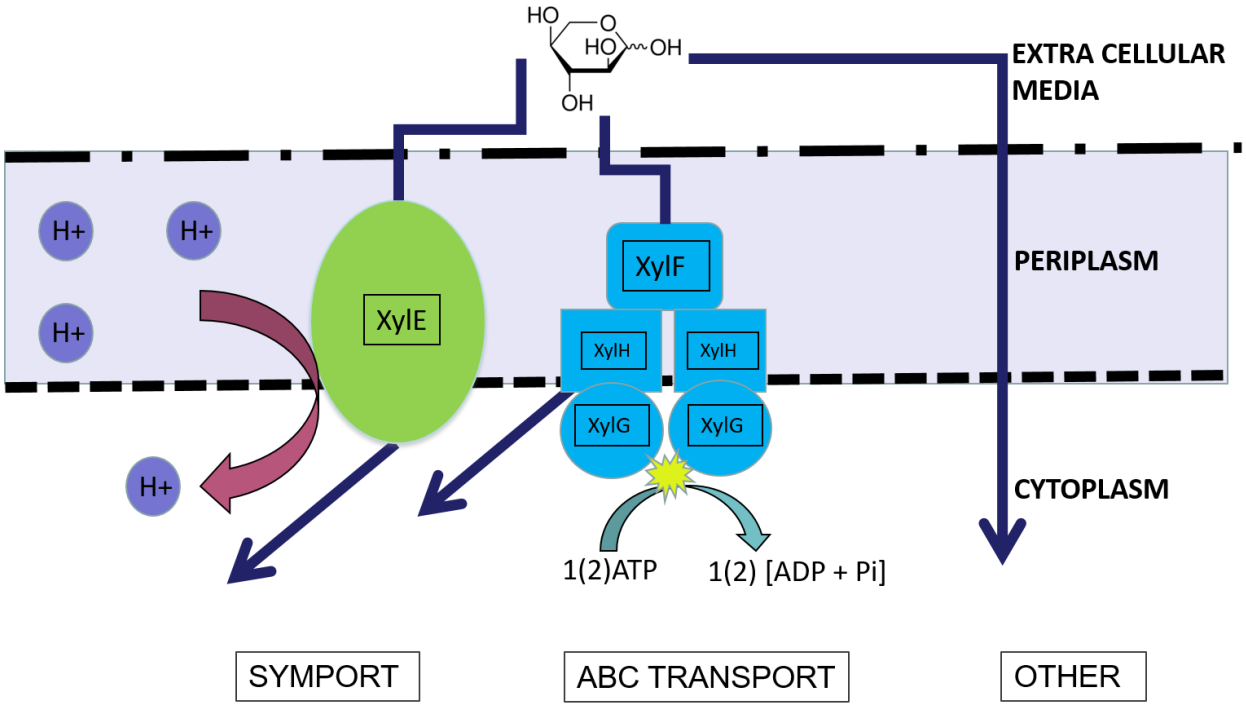


Figure 1.3.1: The transport of xylose by *E.coli* primarily uses two transporters specific for this sugar. Alternative uncharacterized transport processes are possible and may involve diffusion as well as contributions from promiscuous transporters specialized for sugars other than xylose. The symport protein XylE uses directly the proton electrochemical gradient across the inner membrane to energize the transport of xylose. The ABC complex in contrast, composed of the five subunits shown in blue made up from proteins XylF XylG and XylH, uses the energy of ATP hydrolysis. However, the stoichiometry of hydrolysis to transport has not yet been determined.

Xylose degradation by *E.coli* proceeds as follows. Upon entry into the cytoplasm, xylose is first isomerised to the ketose xylulose by xylose isomerase (XylA); in a second step, xylulose is phosphorylated to D-xylulose-5-phosphate by xylulokinase (XylB); both XylA and XylB are enzymes specific to xylose metabolism (Kim and Gadd 2008). Common central metabolic enzymes carry out further metabolism of Xylulose 5-phosphate towards energy harvesting, storage, and biosynthesis, the specifics of which depend upon gene regulation and growth-conditions (Berg, Tymoczko et al. 2006).

1.4: Regulation of xylose transport and metabolism by *E.coli* and carbon catabolite repression

The genes coding for the three subunits of the xylose ABC transporter, *xylF, xylG* and *xylH*, are transcribed as a single operon which also includes a fourth gene encoding the

xylose transcriptional regulator XylR located immediately downstream of *xylH* (Figure 1.4.1).

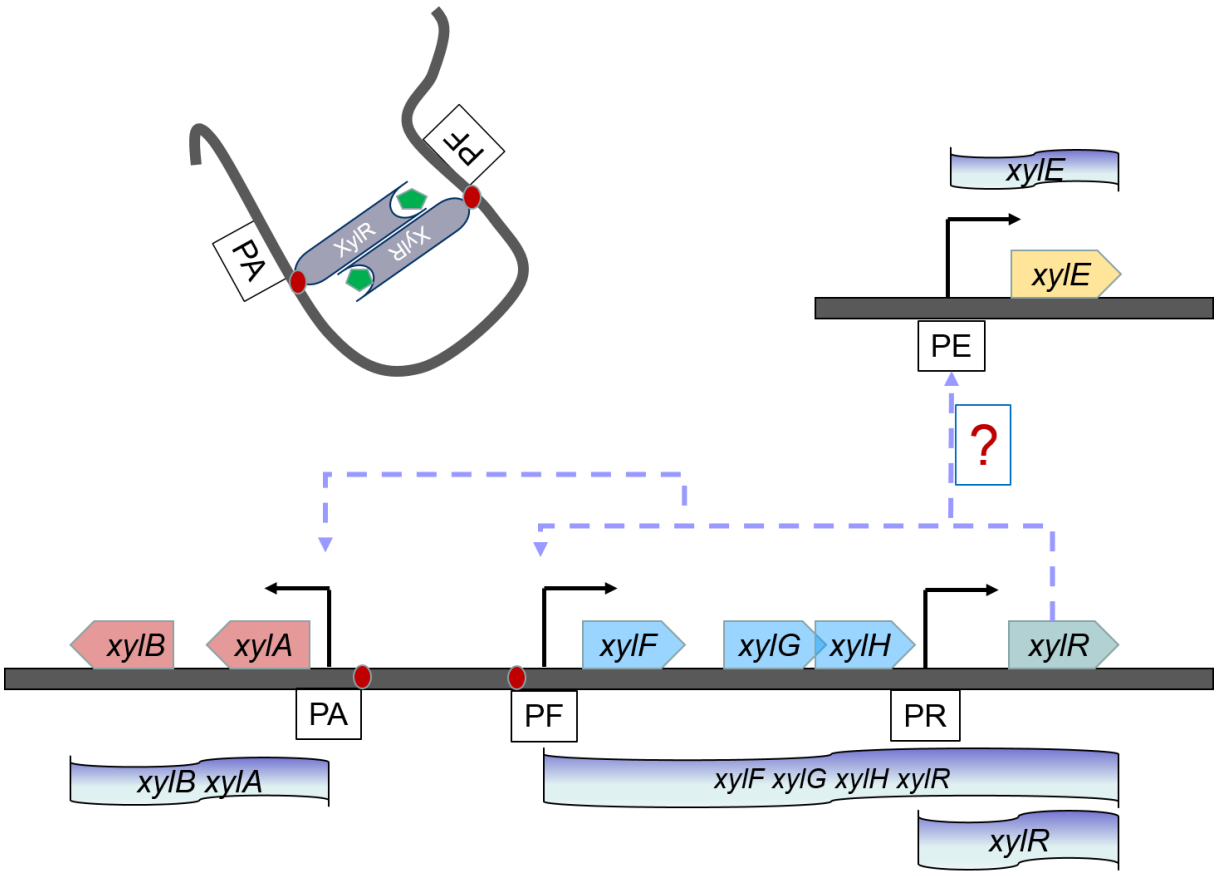


Figure 1.4.1: The genetic regulation of xylose transport and metabolism involves the upregulation of the xylose transport and metabolic genes by the transcription factor XylR. XylR is a dimeric protein that is activated when bound by xylose. Expression from three promotors, PA PF and PE, are thought to be responsive to transcriptional activation by XylR although direct evidence of activation at PE has not been demonstrated. XylR is also constitutively expressible from the XylR independent promotor PR. There is evidence that XylR binding at the PA/PF sites involves a DNA looping mechanism pictured in the upper left of the figure. However a clear looping was visualized between two copies of the PF promotor in an artificial strand of DNA rather than between PF and a PA promotors. The ribbons in the figure show the possible transcripts that may be synthesized from the xylose operon assuming the RNA polymerase completes all transcriptions initiated from the promotors. (Ni, Tonthat et al. 2013)

The *xylA* promoter (PA) is located 360 base pairs upstream of the *xylF* promoter (PF); the genes for the two xylose metabolic enzymes are transcribed from PA in the opposite direction with respect to *xylFGHR* transcription; the *xylE* gene is located elsewhere in the

genome and is expressed from its own separate promoter (PE) (Figure 1.4.1) (Davis and Henderson 1987, Song and Park 1997). The XylR protein is thought to act as a transcriptional activator of the xylose operons based on evidence provided by Song and Park (1997). DNA mobility shift assays showed that XylR binds the PA and PF promoters in the presence of xylose, and *lacZ* reporter assays showed the increased expression and no expression in the presence and absence of active xylR respectively (Song and Park 1997). While these authors provide an indirect evidence of XylR dependent upregulation of the XylE protein – a mutation in *xylR* was shown to retard substrate transport with a greater degree than by a *xylG* mutation – it is not known presently whether this is a consequence of direct XylR-to-PE interaction (Song and Park 1997). Similarly in support of a XylR activation of *xylE* expression are cDNA hybridization assays which show indirectly comparable expression patters between the xylFGH and xylE genes by *E.coli* during fermentative growth on xylose (Tao, Gonzalez et al. 2001). Structural analyses revealed XylR to be a dimeric protein in which individual monomers align in an antiparallel orientation with the corresponding DNA/substrate binding of each monomer facing the opposite direction with respect to the corresponding region in the other monomer; Atomic Force Microscopy suggests that upon activation by xylose binding, the XylR dimer binds the PA and PF promoters in a 1 promoter to 1 XylR monomer stoichiometry as the DNA in between the promoters is looped (Ni, Tonthat et al. 2013). While evidence was provided by Ni et al. for the looping, a clearly visualised image of the loop is captured only for an artificially lengthened piece of DNA encoding two copies of the PF rather than one each of PF and PA. If XylR is binding at the *xylE* promoter, the stoichiometry and mechanism of activation remains to be investigated. Likewise, it is thought that further investigations are warranted of functions of XylR protein other than the known activation of the xylose operons. Evidence has for instance recently been provided of a repressive role of XylR at the

arabinose metabolic operon (Koirala, Wang et al. 2016). Similarly for the xylose operons, while there is now evidence of a looping mechanism, the exact role played by the structure remains unknown, and its treatment may be beneficial towards a better understanding of how *E.coli* cells regulate their xylose metabolic system.

XylR itself has a promoter (PR) located 5 base pairs downstream from the *xylH* translational stop codon; results of chromosomal *lacZ* insertion downstream this promoter revealed a sugar-independent (glycerol xylose and glucose) expression level 10% that from PF and increasing after the mid-log phase of growth (Song and Park 1997). As mentioned earlier, *xylR* is also a part of a transcriptional read-through from the PF promoter: XylR expression is possible from either one of PF and PR in the absence of any repressive conditions.

By both mass and energy, protein expression is known to be a costly investment to cells and bacteria for this reason tightly control the composition of their proteome (Scott, Gunderson et al. 2010, Basan, Hui et al. 2015). Proteins required for the first steps of acquisition and conversion of substrate are required to be expressed at high concentration so as to not induce a bottle neck to catabolism and biosynthesis. A potential evolutionary consequence of this effect is the phenomena of carbon catabolite repression; in the presence of several substrates in comparable concentrations bacterial cells preferentially choose one over the other, and thus cultures in mixed media will sequentially rather than simultaneously consume the available substrates (Görke and Stülke 2008).

The end product of thermochemical plant biomass depolymerisation is a diverse mixture of simple oligomers. This means that carbon catabolite repression poses a hurdle to the development of microbial biocatalysts for biofuel production (Hernández-Montalvo, Valle et al. 2001, Görke and Stülke 2008). The three most prevalent sugars from degraded plant biomass in order of abundance are glucose, xylose, and arabinose; it is known that *E.coli*

exhibit the preference hierarchy of metabolizing glucose first, then arabinose, then xylose (Desai and Rao 2010). The glucose-xylose repression is thought to involve the same cAMP-CRP global regulation that has been reported for the repression of lactose metabolic genes in the presence of glucose, and a likely CRP-cAMP binding sequence has been located in the region of DNA between PA and PF (Song and Park 1997, Desai and Rao 2010). Repression from Arabinose has been shown to involve the arabinose-bound-AraC transcription factor interacting with the xylose promoters (Desai and Rao 2010). Subsequent work on catabolite repression of xylose metabolic genes have shown that the effects may be overcome to varying degrees with genetic interventions; strains constitutively expressing the active CRP have been shown to co-utilize glucose and xylose; similarly an overexpression by a factor of 10-30 of the *xylR* gene has been demonstrated to allow to outcompete the AraC for the xylose promotors (Khankal, Chin et al. 2009, Groff, Benke et al. 2012).

1.5: The numeric characterization of *E.coli*:

A quantitative characterization of a generic *E.coli* cell is required for the modelling of this bacteria; that is to say a computationally representable characterization of a typical *E.coli* cell. The *E.coli* cell is biochemically a very well characterized system, as is discernible from our brief description in section 1.2 of the current state of knowledge. However, the distillation of the known biology of this organism into a computationally tractable conceptualization requires simplification. This section describes how *E.coli* often appears as modelled in an equation or a piece of computer code.

The concept of a "general" *E.coli* is genetically unfounded as can be seen from the breadth of *E.coli* as a taxon (Lukjancenko, Wassenaar et al. 2010). This fact is largely ignored in the quantitative representation of *E.coli*. Thus numeric parameters calculated for one strain are often used in the description of another; the difference between *E.coli* strains B/r and K-12 are assumed negligible for the purposes of computational modelling though

clearly likely to be biologically relevant. Similarly, *E.coli* cells within a population of genetically related individuals are known to differ by such factors as age (Stewart, Madden et al. 2005). On the contrary, the computational population is typically taken to be unstructured. The computational *E.coli* is best conceptualized to be a cylinder with a base radius of 0.35-0.7 μm, a height of 2-4 μm, and thus a volume of 0.8-6.2 μm^3 and a population is a homogenous distribution (an ensemble) of such cylinders over a reaction volume (Bionumbers ID: 100002, 103714, 104113, 102065, 100001, 100004). With regard to population density, an OD_{600} of 1.0 is thought to correlate with a cell concentration of $10^8 \sim 10^9 \ \frac{\text{cells}}{\text{mL}}$ (Bionumbers ID: 100985, 103625, 10831) (Milo, Jorgensen et al. 2010).

The surface surrounding the cylindrical volume is by definition the thermodynamic boundary separating the cell from its environment. Therefore the concept of a periplasm and an outer membrane is generally absent. It is noted that biochemical models of membrane transport phenomena in Gram negative bacteria often assume the freely diffusive interaction of substrate concentrations across the outer membrane via porins. Therefore the simplification is to reduce the effect of any periplasmic processes to the permeability arising from the more selective of the two membranes, the inner membrane, as a function of the presence of transporters. As a reference, it is known that an *E.coli* cell growing on lactose is capable of expressing 10,000 copies of the lactose transporter whose turnover number is reported to be 10 reactions sec^{-1} at 25 degrees Celsius (Nelson, Lehninger et al. 2008, Stein and Litman 2014).

Inside the cylinder is the cytoplasm within which the majority of biochemical processes occur. The fluidic property of the cytoplasm is reported to have a glass-like consistency, quite different from the aqueous solutions for which such numeric values as Michaelis Menten enzymatic parameters are measured and reported *in vitro* (Parry, Surovtsev

et al. 2014). The prokaryotic cytoplasm is a highly ordered structure; the biochemical environment of the nucleoid region as an RNA polymerase enzyme might experience is quite different from what would be experienced by an average cytoplasmic ribosome or a plasmid (Nelson, Lehninger et al. 2008). This fact is ignored in *E.coli* dynamics and chemical processes are modelled through ordinary differential equations and kinetics based on the principle of mass action (Alon 2006, Bolouri 2008).

Taking the cell volume to be 1.0 femto-litre (fl, 10^{-15} l), a useful conversion factor between molecular concentration and copy number is as follows: a compound (we assumed is called "S") whose concentration, in the *E.coli* cell in mM is [S], is represented by approximately $[S] \times 10^6$ copies; a compound represented by n copies is in turn present at a concentration of $n \times 10^{-6}$mM. This is because

$$[S]\left(\frac{\text{mmol}}{\text{L}}\right)\left(6.23 \times 10^{20}\frac{\text{copies}}{\text{mmol}}\right)10^{-15}\text{L} = 6.23 \times 10^5 \text{copies}[S] \qquad \textbf{(1.5.1)}$$

The enzyme database BRENDA (BRaunschweig ENzyme DAtabase) reports that the *E.coli* xylose isomerase has a substrate affinity (Km) on the order of 10 mM. Approximating a typical concentration of the sugar during growth on xylose by the Km value of the enzyme metabolizing the sugar suggests that one would expect there to be about a hundred-thousand copies of the sugar present in the cell at steady state.

1.6: Models of microbial growth and specific growth rate

A fundamental assumption concerning the growth of populations is that overall growth rate is proportional to the size of the reproducing fraction of the total population; if each reproducing individual can be approximated by a single per-capita rate, then the rate of increase in population size should be approximated to be the size of the reproducing fraction multiplied by this rate. This latter per-capita rate is commonly referred to as the "specific

growth rate." In a homogeneous population model, each existing individual is assumed to contribute equally to the overall growth rate, so that there is no "population structure" in the model distinguishing members by reproductive fecundity. Thus, let μ (in units of reciprocal time) designate the specific growth rate of an individual, and x designate population size. With this notation, the overall reproductive rate of a homogeneous population may be expressed by the elementary differential equation, the Malthus equation, in the time-variable "t"

$$\frac{dx}{dt} = \mu x \qquad \textbf{(1.6.1)}$$

If in the above setting, the specific growth rate is a constant (μ does not change with time), then equation 1.6.1 is seen to have the solution

$$x(t) = x_0 e^{\mu t} \qquad \textbf{(1.6.2)}$$

Equation 1.6.2 describes the time-dependence of the size of a population starting with an initial value of x_0 and growing obeying the rate equation 1.6.1. Note that although we have assumed, by calling μ a "reproductive rate," that $\mu \geq 0.0$, the solution in 1.6.2 is equally valid for $\mu < 0.0$ where 1.6.1 would be a model of population decay. Consider for instance, a chemostat model where the population growing by 1.6.1 is also diluted at a rate D > 0.0, where we have used the capital "D" to avoid confusion with the differential notation. This modification may be expressed using the equation

$$\frac{dx}{dt} = (\mu - D) \times x \qquad \textbf{(1.6.3)}$$

In the experimental setting of a chemostat, the system whose dynamics is represented by equation 1.6.3 is progressed to a steady-state where $D = \mu$. The investigator can control the specific growth rate of a population at this steady-state, by experimentally controlling the

dilution rate. Thus, the dilution rate is fixed to a constant by the experimenter and determines the steady-state growth-rate of the population, which in turn is also constant.

In more realistic biological settings, it is rarely the case that the specific growth rate, of even a simple homogeneous population, is a constant over time. The growth rate is likely to be a very complex function of many growth parameters for which their cumulative effect on reproduction is unknown. To account for observations that the specific growth rate varies over population growth, modifications to 1.6.1 have been introduced to model the dependence of μ on the state of the growth system. We mention two such modifications: a dependence on population size, and a dependence on the growth substrate.

One straightforward way by which μ may be modelled to vary is by writing

$$\mu = \mu(x) \tag{1.6.4}$$

In equation 1.6.4, the specific growth rate of a population is dependent on the size of the population x. As an example, the simple logistic growth model is given by

$$\frac{dx}{dt} = \left[\mu^{Max}\left(1 - \frac{x}{N}\right)\right]x \tag{1.6.5}$$

Where

$$\mu(x) = \left[\mu^{Max}\left(1 - \frac{x}{N}\right)\right] \tag{1.6.6}$$

(Verhulst 1838). Equation 1.6.6 provide the following interpretation of μ. If the population is small ($x \approx 0.0$), the specific growth rate should be at a maximal value (μ^{Max}) which may be for instance because of minimal intrapopulation competition; equation 1.6.5 is approximately equation 1.6.1 with this maximal specific growth rate. Near the environmental carrying capacity (N) however, the specific growth rate approaches zero, and the population size converges to the carrying capacity. Thus, the model given by equation 1.6.5 results in a sigmoidal population growth from below and population decay from above a non-zero

attracting equilibrium located where the population size equals the carrying capacity N. Generalizations maintaining the same form as the simple logistic model are possible; these are the Richards model and its limiting form, the Gompertz model (Gompertz 1825, Richards 1959).

When bacteria are grown in the laboratory in batch, the growth trajectory of populations is known generally to conform well to a sigmoidal description, and the carrying capacity appears similarly to be a biologically relevant growth parameter. As an example, it has been demonstrated that *E.coli* grown in Luria Bertani Lysogeny Broth (LB) will cease to increase beyond an environmental carrying capacity (density) of an OD_{600} of 7.0 (Sezonov, Joseleau-Petit et al. 2007). Because sigmoidal growth is so wide spread across microbial growth experiments, models such as the Logistic model provide a convenient way by which to compare independent growth data through a common set of parameters (μ^{Max} and N in the case of equation 1.6.5).

As a second example of a variable specific growth rate, we consider that μ may depend on the concentration of a limiting growth substrate in the growth medium (which we again call "S" as in section 1.5).

$$\mu = \mu([S]) \qquad \textbf{(1.6.7)}$$

One instance of equation 1.6.7 is provided by the Monod microbial growth model, derived empirically in the 1940s by Jacques Monod. In Monod's model specific growth rate is defined as a saturating function of the external concentration of a limiting substrate (Monod 1949).

$$\frac{dx}{dt} = \left[\frac{\mu^{Max}[S]}{K_m + [S]}\right] x \qquad \textbf{(1.6.8)}$$

where

$$\mu([S]) = \left[\frac{\mu^{Max}[S]}{K_m + [S]}\right] \quad \textbf{(1.6.9)}$$

The form of equation 1.6.9 can be seen to be identical to that of the Michaelis-Menten formula for the catalytic rate of enzymes; the specific growth rate in equation 1.6.9 approaches a maximal rate (μ^{Max}) with increasing [S], and is $\frac{\mu^{Max}}{2}$ when $[S] = K_m$. The derivation of 1.6.9 is the same as the derivation of the Michaelis-Menten formula but with two notable differences. First, the enzyme concentration appearing in the Michaelis-Menten derivation is replaced with the number of transporters across the cell membrane. This is because mass action kinetics between the external substrate and the transporter is assumed to determine the consumption rate of the substrate by individual cells. Second, whereas the rate that is predicted by the Michaelis-Menten model is that of a single reaction, the rate that is predicted by equation 1.6.9 is interpreted to be a population specific growth rate. Thus, in writing equation 1.6.9, one also assumes that growth rates are approximately determined by the rate of the consumption of a limiting substrate. Equation 1.6.9 also predicts that the slightest increase in [S] from 0.0 results immediately in a positive value of μ. In reality, this is not seen to be the case, most likely because a portion of the available substrate must be invested towards the survival of the reproducing parent; thus, equation 1.6.9 ignores the fact that microbial growth systems often come with threshold substrate concentrations below which no growth is possible (because there is not enough substrate for novel biosynthesis) (Kovárová-Kovar and Egli 1998). This phenomenon is treated theoretically in models by including what are called "maintenance energies." The theory of maintenance energies is reviewed in van Bodegom and details are also provided in section 5.2 of this thesis (Van Bodegom 2007). Finally, we mention that a theoretical connection does exist between the Monod and the Logistic growth models. With a handful of assumptions, one equation can be derived from the other through a rather nasty sequence of steps; the derivation was felt to be

beyond the scope of this section owing to the technical details and is therefore excluded (A. Jamie Wood; personal communication).

CHAPTER 2: Stoichiometric models of metabolism

Abstract:

A mathematical introduction to the principles of flux balance analysis (FBA) and dynamic flux balance analysis (dFBA) is provided. FBA and dFBA are considered in the general setting of structural metabolic modelling. In the general setting, metabolic system dynamics are represented mathematically as a composition of two mapping operations - a nonlinear velocity map, and a linear map represented by the stoichiometry matrix. By starting with this description, we are able to make explicit, the assumptions inherent in flux-balance models, as well as provide a unified perspective within which these methods can be explicitly related to other structural modelling methods such as Metabolic Control Analysis.

2.1: Introduction

The metabolic engineering of microorganisms towards the synthesis of commercially relevant compounds is now, with the availability of precise genetic manipulation techniques and in the presence of industrial demand, a highly active area of research. In the previous chapter, we have considered the biological aspects of this engineering effort with an explicit focus towards the role played by the model organism, *Escherichia coli*. Single gene manipulation of this organism has been successfully employed in the past as a means of engineering desired phenotypes in the laboratory, however the limitations in the scope of the approach becomes apparent in light of the sheer size and intricacies of cellular metabolism. Therefore, the genetic engineering of cells towards desired metabolic and growth phenotypes calls for a more holistic perspective to complement experimental methods. Such a global perspective to phenotypes is afforded by systems biology (Kitano 2002, Kitano 2002).

In this chapter, we will describe the basic principles behind two related systems biological models of metabolism: flux balance analysis (FBA) and dynamic flux balance analysis (dFBA). The reader is first introduced to the stoichiometric matrix representation of metabolic systems, the biological rationale underlying this representation, as well as the motivations behind its use in molecular systems biology. We will then consider the

mathematical representation of biomolecular systems dynamics, describing how the stoichiometry matrix plays a fundamental role when describing fluxes across interconnected sets of reactions modelled as ordinary differential equation (ODE) systems. The differential equation representation will be seen to be only symbolic and generally insolvable; it will however also be readily observed that, by applying some basic results of mathematical analysis, that practical results of empirical relevance can be derived for linearized systems in the proximity of stable equilibria. By introducing the stoichiometry matrix in this way, two goals are achieved. Firstly, the separation between the structural and dynamical aspects of metabolism, as well as their connection, is made clear as is relevant to dFBA. Second, by starting with the full dynamical representation of metabolism, we are able to interpret FBA methods in a unified framework that also includes such methods as Metabolic Control Analysis (MCA). In particular, we are able to compare and contrast these two methods by considering how the steady-state assumption is used in slightly different ways in FBA and MCA when reducing a fully kinetic model. As part of this process we make explicit the fundamental assumptions made by both FBA theory and the new ideas introduced in this thesis. The most important of these new ideas will be seen to be that of the cellular objective. Together with the stoichiometric matrix, the objective will be seen to turn FBA into an instance of a constrained optimization problem. Finally, we give a description of the dynamic extension of FBA, the dFBA method.

FBA methods are often described in the literature by mentioning only the null-space mapping by the stoichiometry matrix explicitly. We emphasize however that this linear mapping, between the flux space and space of metabolite concentration derivatives, is only one step in a larger mapping procedure, whose true domain is the space of metabolite concentrations and system parameters. This view does not affect the way we interpret results of FBA. However, we find that the mathematical assumptions underlying dFBA may be best

understood by considering what restrictions are necessary to be placed on the first step of the overall mapping operation.

2.2: Conventions

When referring to a general reaction, we will use the substitute "rxn" in place of a specific reaction name. Where it makes sense to qualify a symbolic notation with a reaction name, we will qualify with rxn when the indicated quantity is being discussed in the general.

Analogously, we will use "comp" to refer to a general metabolic compound, and extend this notation to all notations for which a qualification by a specific compound name is valid. We will assume that a directionality has been assigned to all reactions and further, that for all irreversible reactions, the same direction as the one indicated by the reaction arrow in its chemical equation representation has been assigned to be the forward (*positive*) direction. A compound residing to the *negative* side of the chemical equation representation of a reaction will be called a *substrate*; likewise, a compound residing to the positive side will be called a *product*. The name of an enzyme catalysed reaction will be identical to the name of the catalyst. The names for reactions that are not catalysed by an enzyme will be treated individually. These exceptions arise for uncatalyzed processes and lumped processes. A special class of uncatalyzed processes is that of the exchange reactions. An *exchange reaction* is defined to be a reaction with a single substrate having a stoichiometric coefficient of 1.0 and no products. These reactions are not physical reactions but are computational features that allow the modelled metabolic system to interact with its surroundings. In naming exchange reactions, we will follow the convention of attaching the name of the compound associated with the reaction to "EX_." This is the same convention followed for *E.coli* metabolic models such as the iAF1260 model (Feist, Henry et al. 2007). Thus, the name given to the following exchange reaction

$$\text{comp} \quad \begin{matrix} \rightarrow \\ \leftarrow \end{matrix} \qquad \textbf{(2.2.1)}$$

will be "EX_comp." Note also that we will generally not write the stoichiometric coefficient explicitly if the value of the coefficient is 1.0. Every exchange reaction is a reversible reaction.

Given a particular reaction rxn, we define the *velocity*, or equivalently the *rate*, of the reaction as follows. Let comp be any substrate participating in the reaction with a non-zero stoichiometric coefficient which we denote here by s_{comp}. The velocity of the reaction, generally denoted by v_{rxn}, is defined as

$$v_{rxn} = \left(\frac{-1.0}{s_{comp}}\right)\frac{d[comp]}{dt} \qquad \textbf{(2.2.2)}$$

Note that, as a result of stoichiometric equivalence, the velocity of a reaction will be of the same value regardless of which substrate is chosen when evaluating equation 2.2.2. Likewise, comp in 2.2.2 can be replaced with a product without altering the value of v_{rxn} provided the numerator in the ratio scaling the derivative is given a positive sign. Thus, at any given time, a reaction will exhibit a single value for velocity, and it makes sense to speak of *the* velocity of a reaction at that time. The unit on velocity is unit amount of substance per unit of time, as in mmol h^{-1}. A reaction velocity has the interpretation of a rate – namely the number of conversion events per unit time. A slightly different interpretation occurs when modelling metabolic systems using networks, where metabolites will be represented by nodes and reactions by directed edges connecting the nodes. In this view, a reaction whose velocity is v_{rxn} may be thought of as carrying a unit of mass that passes through the edge from one node to another, and that such "fluxes" across individual reactions combine systemically to form mass currents across the overall network. Thus, we will refer to a *flux across* a reaction

to indicate specifically, the transport of mass (via chemical conversion) between two classes of metabolites. Importantly, unlike for velocity that was seen to be a property of a reaction, the value of flux will depend on the unit of matter that is being considered. If, in equation 2.2.2, comp is a five-carbon compound appearing alone on the substrate side with a stoichiometric coefficient of 1.0, it would be valid to say either that $|v_{rxn}|$ is the flux of comp across rxn, or that $5|v_{rxn}|$is the flux of carbon across rxn. Note that a flux *of* a unit of matter will always be a positive value. When no other qualifications are given, *the* flux of a reaction will be taken to mean the velocity of the reaction, except that whilst it makes sense to speak of the flux *across* the reaction or the flux *carried by* the reaction, the same would not be the case for either the velocity or the rate. For this reason, we will use the same notation v_{rxn} to indicate both the velocity of rxn, and the flux of rxn.

In moving forward with the network perspective of metabolic systems, we will need to represent, the net flux of mass either into or away from a node, representing the net synthesis and utilization rate respectively of the associated metabolite. That is, we will need to represent, not just the velocity of a particular reactive step, but also the cumulative flux to or away from a particular metabolite that results from summing over velocities of several reactions in which the metabolite is a participant. In a steady-state model, the production and utilization rates for a particular metabolite must balance. Therefore we will denote the absolute value of the production rate, and equivalently, the absolute value of the utilization rate, of a metabolite by subscripting the lower case v with the name of the metabolite. This is in analogy to how we have treated reaction velocities. As an example, v_{ATP} will be used to represent the net rate of ATP production, and equivalently, the net rate of ATP hydrolysis by a metabolic system. Because the system is assumed to be at steady-state, we find that subtracting the ATP hydrolysis rate from the ATP production rate results for this example in the steady-state equation

$$v_{atp} - v_{atp} = 0 \quad \textbf{(2.2.3)}$$

One noteworthy case occurs when the metabolite in question has an associated exchange reaction. If flux to or from the metabolite is balanced by a flux from or to the surrounding, then with respect to the part of the metabolic system that is being explicitly modelled, the metabolite in question is not conserved. Whilst the convention described above may still be applied to expressing the overall production and consumption rates of such metabolites, it should be noted that the biological interpretation of these rates are different. Whereas a production rate of an internally balanced metabolite such as ATP represents a dynamic equilibrium within the explicitly modelled part of the metabolic system, an analogous rate for a sink or source metabolite represents a consumption or expulsion of matter by the modelled system. As an example, consider the production rate of oxygen v_{o2}. As in equation 2.2.3, we may still write

$$v_{o2} - v_{o2} = 0 \quad \textbf{(2.2.4)}$$

To represent the fact that production and consumption rates balance in the overall flux-balanced model. However, oxygen is an external metabolite that is consumed by a respiring system, and absorbed from the environment by an exchange reaction EX_o2. Thus, we have

$$v_{o2} - v_{EX\_o2} = 0 \quad \textbf{(2.2.5)}$$

From which it is seen that whilst the overall model balances oxygen, oxygen is lost by the metabolic component that only explicitly accounts for its consumption.

Provided with a particular quantity, we will often need to indicate that the value of the quantity satisfies a chemical steady state or, less frequently, a chemical equilibrium condition. These facts will be indicated by subscripting the notation for the quantity by "ss" for steady state, and "eq" for equilibrium. Thus, $[comp]_{ss}$ is the steady-state concentration of the metabolite named comp. Similarly, we will also often need to indicate that a quantity meets an extreme value condition; either the quantity itself has been extremized or its value is

such that some other quantity has been extremized as a function of this quantity. These facts will be indicated with superscripts. To indicate that a quantity satisfies an extreme value condition, the notation for the quantity will be given the superscript "opt" (for optimal). If we also want to indicate the direction of extremization, then the superscripts "Max" and "Min" will be used to indicate maxima and minima. An exception to this rule is given for the chemical parameter V_{max} whose notation is already well established in the biochemical literature. At no point in this thesis will we make explicit references to (finite) mathematical suprema and infima that are not also respectively maximal or minimal values to which variables may evaluate.

Finally, it will also be noted here that the qualifications declared in this section will only be applied if it the information they provide is not already clear from context. Thus, we will refrain for example from repetitively writing reaction velocities as v_{rxnss} if it has been declared in the beginning that the metabolic system under discussion is being analysed under steady-state conditions.

2.3: Local versus global metabolic models

In this section, we highlight the key differences between traditional versus systems biological approaches to metabolic modelling. In so doing, we illustrate to the reader, the conceptual paradigms and the motivations that underlie the systems framework of modelling within which flux-balance methods fit. Thus, by way of introduction, suppose we are interested in describing the kinetics of a single reversible biochemical reaction converting a substrate S into product P. Letting the forward and the reverse rate constants for the process be denoted by $k_{+1}$and k_{-1}, we may write down the following chemical and mathematical equations to describe this process

$$S \underset{k_{-1}}{\overset{k_{+1}}{\rightleftarrows}} P \qquad \textbf{(2.3.1)}$$

$$\frac{d[S]}{dt} = k_{-1}[P] - k_{+1}[S] \qquad \textbf{(2.3.2)}$$

Upon reaching chemical equilibrium where [S] and [P] cease to change with time, but remain at fixed values $[S]_{eq}$ and $[P]_{eq}$ equation 2.3.2 can be used to solve for the equilibrium constant for the process k_{eq}defined to be

$$k_{eq} = \frac{k_{+1}}{k_{-1}} \qquad \textbf{(2.3.3)}$$

Or equivalently,

$$k_{eq} = \frac{[P]_{eq}}{[S]_{eq}} \qquad \textbf{(2.3.4)}$$

This empirical result may be taken as a thermodynamic characterization of the reaction given by 2.3.1. In the biochemical context of a cell, this single reaction will most certainly be only one step in a lengthy set of steps that together form a chemical pathway such as the glycolytic pathway or the tricarboxylic acid cycle. Thus it may seem possible and tractable to characterize the pathway by modelling thermodynamically or kinetically the individual steps in such a way as in equation 2.3.2, and to draw conclusions about the global behaviour of the pathway as a whole. This approach to pathway modelling has been shown to come with two potentially dangerous limitations. First, parameters of reactions such as the equilibrium constant have most likely been empirically derived from experiments that investigated a reaction *in vitro* and in isolation. Thus, the parametrization will likely only apply to the reaction when it has been taken out of the context of a chemical pathway. Second, when a

pathway is described as a sequence of reactive steps, but without a consideration towards the interdependencies that exist amongst the rate determinants for those steps, an apparently "dominating" local quantity may erroneously be attributed an undue global significance (Fell 1997). An essentially irreversible step (one with a high equilibrium constant) may for instance give an impression that this step is a thermodynamic driver of material flux through the pathway in which it is embedded; from this, an erroneous conclusion may be drawn that such a step is rate-limiting for the pathway as a whole. Though intuitively satisfying, it may be shown with precise mathematical analyses of empirical data that most real biochemical pathways do not possess such a rate-determining step (Fell 1997, Cornish-Bowden and Cárdenas 2001, Gunawardena 2002). Rather, the "control" of material flux through a pathway is more often distributed amongst several reactions instead of being localized to any single step. This well-known result, may be proved using the theory of metabolic control analysis; the resulting conclusion is formally asserted by the (flux control) Summation Theorem (Gunawardena 2002). The effort to model metabolism as an integrated system is motivated in part by these two shortcomings of local analysis just mentioned. There is a need to first conceptualize metabolism as a complex system of interdependent components and second, a need to represent the dynamical aspects of said interdependencies in a way amenable to computational interrogation. One practical consequence this has had to modelling is a push by modellers in the recent years towards the network representation and analyses of metabolic systems (and also of other unrelated systems arising elsewhere in biology) (Barabasi and Oltvai 2004). Network models are structural representations of large biosystems that use nodes and edges in place of system components and interactions. The next section will describe in detail how the network view of complex biosystems is used in metabolic representation.

Writing simple equations like 2.3.2 relies on mass-action kinetics. If more information is known about the process, such as enzyme concentration [E], and k_{Cat} and K_m values, a more detailed mass-action model of the process might utilize a rate equation such as (for the forward process)

$$\frac{d[S]}{dt} = \frac{-k_{Cat}[S][E]}{K_m + [S]} \qquad \textbf{(2.3.5)}$$

(Cornish-Bowden 2012, Klipp, Liebermeister et al. 2016). Note that writing the derivative of metabolite concentration as equation 2.3.5 or some generalization of that equation states that only three types of quantities in general will factor into the determination of the rate on the LHS. These are respectively metabolite concentrations, enzyme concentrations, and parameters. Whilst equation 2.3.5 describes the dependence for a single isolated process, it does not account for the local rate dependence on concentrations and parameters of other processes of a larger system. The network perspective of metabolism will be shown to bridge this gap.

2.4: The stoichiometric matrix

Consider a biochemical reaction system consisting of n reactions and m metabolites. We assume that each metabolite has been uniquely indexed by an integer i, with $1 \leq i \leq m$, and that each reaction has been uniquely indexed by an integer j with $1 \leq j \leq n$. The stoichiometric matrix (which we generally denote by **S**) representation for this system is a real matrix consisting of m rows and n columns. Let the rows and the columns of **S** each be indexed by positive integers in the usual manner. By matching the indices, to each metabolite is associated a row of **S** and similarly, to each reaction is associated a column of **S**. The absolute value of the entry in the ith row and jth column of **S** is the stoichiometric coefficient of the ith metabolite as it participates in the reaction j; this entry is given a positive sign if the ith metabolite resides to the positive side of the jth reaction, and is

otherwise given a negative sign. Denoting this general entry by s_{ij}, we see that $s_{ij} > 0$ if metabolite i lies to the product side of the reaction j, $s_{ij} < 0$ if metabolite i lies to the substrate side of reaction j, and $s_{ij} = 0$ if metabolite i does not participate in reaction j. Hence we see that in **S** is contained the structural information and the stoichiometric information of our metabolic system; that is to say, the connectivity amongst metabolites defined as reachability by a sequence of reaction edges and the stoichiometric ratios by which metabolites participate in these individual reactions.

2.5: Dynamical representation of metabolism

The stoichiometric matrix is a static representation of metabolism. However, as we now describe, **S** in combination with velocity results in a complete dynamical representation; specifically, we are interested in deriving a self-contained system of differential equations in metabolite (as opposed to catalyst) concentrations using **S**. The metabolic models described in this and the coming sections require that algebraic equations written using the stoichiometric matrix, provided they are consistent, are solvable. For this computational reason, we will assume that the inequality $n \geq m$ is satisfied by all metabolic models without loss of generality. Empirically, we find that this assumption is generally justifiable for metabolic systems; biological reaction networks are known to be such that the number of reactions far outnumber the number of metabolites (Orth, Thiele et al. 2010).

As described earlier, when considering the local aspects of metabolism, the velocity of any single reaction, defined as substrate utilization rate or product formation scaled by a stoichiometric coefficient, will depend generally on three factors: the concentration of metabolites, the concentration of catalysts, and the parameters of the process. In this section, we will consider only the dynamics of metabolite concentrations. Therefore, the concentration of catalysts will be assumed to be static, and thus be treated as process

parameters rather than as dynamic variables. Note that this means that a general expression such as $k_{cat} \times [E]_{ss}$, where $[E]_{ss}$ is an enzyme concentration, may be combined as a single parameter V_{max} representing the product of the turnover rate and a static concentration of the enzyme E. Note also that the assumption that [E] is static at a constant value of $[E]_{ss}$ may be justified by separation of timescales; enzyme concentrations (but not activity), being functions of gene expression, is expected to change negligibly on the time-scale of the reactions they catalyse.

Let c_i and v_j denote respectively the concentration of the metabolite indexed i, and the velocity of the reaction indexed j. Further, we assume that there is a total of k parameters describing the system, denoted p_l indexed by integers l with $1 \leq l \leq k$. Based on the previous paragraph, we see that this velocity will be a function of the m metabolite concentration variables, and the k parameters. In other words,

$$v_j = v_j(c_1, \dots, c_m, p_1, \dots, p_k) \qquad \textbf{(2.5.1)}$$

Let

$$\mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \in \mathbb{R}^m \qquad \textbf{(2.5.2)}$$

$$\mathbf{p} = \begin{bmatrix} p_k \\ \vdots \\ p_k \end{bmatrix} \in \mathbb{R}^k \qquad \textbf{(2.5.3)}$$

With n velocity functions, we may symbolically express the systems velocity by a mapping of the form

$$\mathbf{v} = \begin{bmatrix} v_1(\mathbf{c}, \mathbf{p}) \\ \vdots \\ v_n(\mathbf{c}, \mathbf{p}) \end{bmatrix} : \mathbb{R}^{m+k} \to \mathbb{R}^n \qquad \textbf{(2.5.4)}$$

or more concisely as,

$$\mathbf{v} = \mathbf{v}(\mathbf{c}, \mathbf{p}) \qquad \textbf{(2.5.5)}$$

This equation states that the velocities for the n reactions of the system are deterministically quantified when the concentrations of the m metabolites of the system and the k parameters of the system are given. Further, because equation 2.5.4 has been written generally, the domain of **v** has not been restricted to any biologically realistic subset of $\mathbb{R}^{m+k}$. Before imposing a few mathematical requirements we need of **v**, we first consider the product

$$\boldsymbol{\phi} = \mathbf{S} \cdot \mathbf{v}(\mathbf{c}, \mathbf{p}) \qquad \textbf{(2.5.6)}$$

for some fixed **c** and **p**. The object on the LHS ($\boldsymbol{\phi}$) is a well-defined point of $\mathbb{R}^m$ since **v** is now a fixed n-vector and **S** has dimensions m × n. We also note that the ith entry of $\boldsymbol{\phi}$ is the inner product of the ith row of **S** corresponding to the ith metabolite, and the rate vector **v**.

$$\varphi_i = \sum_{j=1}^{n} v_j \times s_{ij} \qquad \textbf{(2.5.7)}$$

Thus, the ith entry of $\boldsymbol{\phi}$ is seen to be the sum over all reactions of the stoichiometric coefficient of the ith metabolite for the jth reaction multiplied by a velocity of the jth reaction. Therefore, the quantity φ_i is the net difference between the rates of production and consumption of the ith metabolite by the system. Recognizing this quantity as a derivative, we arrive at the following general expression for the concentration dynamics of the metabolites

$$\frac{\mathbf{dc}}{\mathbf{dt}} = \mathbf{S} \cdot \mathbf{v}(\mathbf{c}, \mathbf{p}) \qquad \textbf{(2.5.8)}$$

In sum, the stoichiometric matrix acts on a vector of velocities to define a vector of concentration derivatives. If **v** were constant as in equation 2.5.6, equation 2.5.8 represents an algebraic relation between reaction velocities and concentration derivatives. However, in the general setting, it is seen that this is not the case as **v**, even with a static **p**, depends (nonlinearly) on the concentration vector that need not remain static. As a result, equation

2.5.8 does not generally represent an integrable system. In the following, we will consider the maps **S** and **v** to be representative of the two aspects of the dynamics of metabolic systems: **S** defines a metabolic structure whilst **v** defines (local) reaction dynamics. The systems-level behaviour is represented by the differential equation 2.5.8. For technical reasons to be explained in the foot note at the end of this section, but are not all crucial to FBA, we will make several assumptions about equation 2.5.8. First, we assume for simplicity that **S** is of full rank. Second, we assume for the purposes of MCA that **v** is continuously differentiable in both **c** and **p**. Given any **p** of interest, it is required that a stable equilibrium solution to equation 2.5.8 exists – that is, we may provide a particular **c** for which the LHS of the equation evaluates to the zero-vector and that equation 2.5.8 will return the system to that state after a small perturbation is applied to this **c**. With these assumptions in hand, we discuss the steady-state properties of the system

$$\begin{cases} \boldsymbol{\Phi}(\mathbf{c},\mathbf{p}) = \mathbf{S}\cdot\mathbf{v}(\mathbf{c},\mathbf{p})\colon \mathbb{R}^{m+k} \to \mathbb{R}^{m} \\ \mathbf{S}\colon \mathbb{R}^{n} \xrightarrow{\text{linear}} \mathbb{R}^{m} \\ \mathbf{v}\colon \mathbb{R}^{m+k} \xrightarrow{C'} \mathbb{R}^{n} \end{cases} \qquad \textbf{(2.5.9)}$$

*****We have assumed for simplicity that **S** is of full rank. Most biological stoichiometric matrices are rank deficient, due to the presence of conserved moieties that introduce row-dependencies (e.g. adenylate is conserved in ATP, ADP, and AMP). However, it is possible to convert **S** into a matrix of full rank by eliminating these dependencies. The metabolic model and its mathematical analyses will remain unchanged by this operation (Gunawardena 2002). **S** was assumed to be of full rank so that the Jacobian in the system 2.5.9 is invertible.

We assume for the purposes of MCA that **v** in 2.5.5 is a C' mapping. This assumption is required so that **Φ** in 2.5.9 is C' to meet the requirements of the Implicit Functions Theorem (Rudin 1964). For dFBA, the smooth dependence of the velocity map on concentration and parameters is a structural stability requirement of **Φ** and has been discussed throughout.

2.6: Steady-state metabolism and Metabolic Control Analysis

Let $\mathbf{p_p}$ denote a particular fixed parameter combination. Starting with the system in 2.5.9, our stability assumptions on **v** allow us to write the equation

$$\boldsymbol{\Phi}\left(\mathbf{c_{ssp}}, \mathbf{p_p}\right) = \mathbf{0} \qquad \textbf{(2.6.1)}$$

Where $\mathbf{c_{ssp}}$ is the metabolite concentration that satisfies 2.6.1 for the parameter combination $\mathbf{p_p}$. Recalling that $\mathbf{\Phi}$ represents metabolite concentration derivatives (equations 2.5.7 and 2.5.8), it can be seen that equation 2.6.1 is a biological statement about a global metabolic steady-state. Namely, it states that for a particular choice of parameters, the metabolic system has arrived at a (stable) dynamic equilibrium at which point the concentrations of metabolites remain unchanging at $\mathbf{c_{ssp}}$. For this pair of values for **c** and **p**, the velocity function evaluates to a constant $\mathbf{v_{ssp}}$ defined by

$$\mathbf{v_{ssp}} = \mathbf{v}(\mathbf{c_{ssp}}, \mathbf{p_p}) \qquad \textbf{(2.6.2)}$$

and equation 2.5.8 is seen to simplify to

$$\mathbf{S} \cdot \mathbf{v_{ssp}} = \mathbf{0} \qquad \textbf{(2.6.3)}$$

which is a computationally tractable algebraic system in m independent equations and $n \geq m$ variables. In order to ensure that outputs are computable, large-scale computational models that are concerned with rates of metabolic systems must make this simplification in one form or another. For this reason, such models can be characterized as having the two common features represented by equation 2.6.3: a stoichiometric matrix and a steady-state assumption. The system in 2.5.9 represents metabolic dynamics as a composition of two mapping procedures with three associated spaces. Namely, these are the $\mathbb{R}^{m+k}$ space of parameters and metabolite concentrations, the $\mathbb{R}^{n}$space of reaction velocities, and the $\mathbb{R}^{m}$ space of metabolite concentration derivatives. Although different steady-state models of metabolism may approach this equation differently, our discussion makes clear of the fact that there is only one dynamical metabolic equation: equation 2.5.9. Two prominent approaches to steady-state metabolic systems modelling are metabolic control analysis (MCA), and flux balance analysis (FBA). Although we are primarily concerned with the latter approach, a consideration of the connection between FBA and MCA will be useful

towards a better understanding of where FBA methods fit in the larger context of metabolic modelling, and with respect to the general mapping operations given in 2.5.9. In the next section, it will be made clear that FBA is concerned with only the mapping by **S** between $\mathbb{R}^n$ to $\mathbb{R}^m$. To set FBA up in a context in which **S** is explicitly understood to be only one component of a larger mapping operation, we give now a brief description of MCA.

The overall goal of MCA is to characterize the sensitivity of metabolic steady-state properties (e.g. metabolite concentrations, reaction rates, and pathway fluxes) to perturbations in parameters (e.g. enzyme concentration and activity) (Gunawardena 2002). The MCA modelling framework requires the steady-state assumption (equation 2.6.1), but for its consequences to the mapping **v** (equation 2.5.5) from $\mathbb{R}^{m+k}$ to $\mathbb{R}^n$, thus for a somewhat different mathematical reason than for FBA. The main result from analysis that is required by MCA models is the Implicit Functions Theorem. By asserting the C' requirement on the velocity map **v**, MCA methods use the steady-state equation 2.6.1 to arrive at the equation for parameter dependence

$$\mathbf{\Phi}\left(\mathbf{c_{ss}}(\mathbf{p}), \mathbf{p}\right) = \mathbf{0} \qquad \textbf{(2.6.4)}$$

This equation is similar to equation 2.6.1, but the parameters are now variables, and steady-state metabolite concentrations are explicitly shown to be (implicitly defined) functions of **p**. The Implicit Functions Theorem states that, provided **Φ** is continuously differentiable, that steady-state concentrations may be written implicitly as a continuously differentiable mapping of **p** in the vicinity of a point of $\mathbb{R}^{m+k}$ satisfying equation 2.6.1 at which the Jacobian $\frac{\mathbf{d\Phi}}{\mathbf{dc}}$ is invertible (Rudin 1964). The invertability requirement is satisfied in the biological setting because **S** is of full rank and because $\frac{\mathbf{dv}}{\mathbf{dc}}$ is empirically unlikely to be singular (Gunawardena 2002).

We will not go further in describing the theory behind MCA, as our main interest is FBA. The method is introduced to illustrate how MCA and FBA approach equation 2.5.9 in contrasting ways. We have stated that, in the overall dynamical description of metabolic systems, **v** represents dynamics whereas **S** represents structure. It will be seen however, that methods such as FBA are developed starting with equation 2.6.3 by largely ignoring the role that is played by the stoichiometric mapping in the overall dynamical equation 2.5.9. This point is highly relevant to dFBA methods that tread the line between purely structural (FBA) and fully dynamic descriptions (equation 2.5.9). In contrast to the formulation of FBA, the formulation of MCA requires the steady-state assumption first and foremost as a justification to writing

$$\mathbf{c_{ss}} = \mathbf{c_{ss}}(\mathbf{p}) \qquad \textbf{(2.6.5)}$$

The main advantage for MCA of being able to write equation 2.6.5, is that equation 2.6.4 may then be differentiated directly to relate metabolic structure to parameter dependences. The most important consequences of MCA, such as the Summation Theorem mentioned in section 2.3, are structural implications that arise from equation 2.6.3 (Iglesias and Ingalls 2010). Namely, these results all depend crucially on combining equation 2.6.3 with the fact that **S** emerges "intact" upon the application of the Chain Rule to 2.6.4 (because **S** is a linear transformation). Based on our brief description, it may be observed that MCA is an analytic theory; the phenotypes of interest to MCA are continuously defined metabolic system properties. This is a point that we find is worth emphasizing as the distinction between the continuous and the discrete aspects of metabolism will be seen to be less clearly delineated when FBA results are reported. We may now proceed to give a proper introduction to FBA methods, with an explicit understanding that FBA is one component of a larger family of metabolic models that are derived from placing equation 2.5.9 in steady-state.

2.7: Flux Balance Analysis and the objective function

It was seen in the previous section that MCA is concerned with the local sensitivity of metabolic steady states to perturbations made to the parameters of the system. In this section we proceed by describing the FBA framework of metabolic modelling. In FBA, the steady-state assumption is asserted for a purely computational reason, and we find that the properties of the velocity function **v** are quite immaterial; rather, FBA is more or less concerned with the geometry of the space into which **v** maps in relation to imposing direct algebraic constraints on the space itself. As a result, FBA is concerned primarily with the discrete and static aspects of metabolic systems.

Broadly stated, FBA is concerned with the following problem: starting with the steady-state equation (equation 2.6.1), how to derive a particular point in the velocity space in the absence of detailed kinetic descriptions of the rates (that is without equation 2.6.2). The proposed solution to this problem is that we introduce one further biological assumption to the modelling framework; we assume that the parameters of the system **p** are quantified in such a way that an optimality criterion is met by the system. This is an evolutionary assumption. From being exposed to selective pressures to meet an optimality criterion, the parameters **p** in a population are assumed to be tuned in such a way that the members in the evolved population expresses the optimal phenotype. Recalling that the rate equation for any single reaction will be some generalization of the form given in equation 2.3.5, the evolutionary "tuning" is assumed to be of a genetic origin. Enzymatic parameters such as k_{cat} and K_m for instance may be tuned to some degree via mutations that structurally alter the final protein product; similarly, a parameter such as V_{max} may be tuned to some degree by expression-regulatory mutations that alter the steady-state concentration of the catalyst. Before making precise what is meant by an "optimality criterion" we restate the steady-state equation 2.6.3 as it is understood in FBA. Let $\mathbf{p}^{\mathbf{opt}}$ denote a parameter combination for

which the optimality criterion is met by the system, let $\mathbf{c}_{\mathbf{ss}}^{\mathbf{opt}}$ be a steady-state concentration distribution of metabolites corresponding to $\mathbf{p}^{\mathbf{opt}}$, and let $\mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}}$ be the optimal velocity the system exhibits with parameters $\mathbf{p}^{\mathbf{opt}}$ and concentrations $\mathbf{c}_{\mathbf{ssp}^{\mathbf{opt}}}$ so that from equation 2.6.2, we have

$$\mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}} = \mathbf{v}\left(\mathbf{c}_{\mathbf{ssp}^{\mathbf{opt}}}, \mathbf{p}^{\mathbf{opt}}\right) \tag{2.7.1}$$

If **v** is constrained as in MCA, we could similarly apply equation 2.6.5 to arrive at

$$\mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}} = \mathbf{v}\left(\mathbf{c}_{\mathbf{ss}}(\mathbf{p}^{\mathbf{opt}}), \mathbf{p}^{\mathbf{opt}}\right) \tag{2.7.2}$$

but this is not an FBA requirement. Note that we have not written $\mathbf{v}_{\mathbf{ssp}^{\mathbf{opt}}}$. With velocity obeying the steady, state assumption as in MCA, but now also satisfying an optimality criterion, the steady-state equation 2.6.3, for FBA reads

$$\mathbf{S} \cdot \mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}} = \mathbf{0} \tag{2.7.3}$$

where either expressions 2.7.1 or 2.7.2 may be substituted for $\mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}}$. Thus the optimality assumption in FBA converts the general null-space equation 2.6.5, to the optimal equation 2.7.3.

We now discuss the mathematical conditions with which to decide when the equation

$$\mathbf{v} = \mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}} \tag{2.7.4}$$

holds true. Thus, starting with a velocity **v** satisfying the steady-state equation 2.6.5, we seek conditions to decide if **v** also satisfies the optimal steady-state equation 2.7.4 (or equivalently, equation 2.7.3). Towards this end, we define the "cellular objective" (or just "the objective") to mean any real-valued function of velocity. Denoting a general objective function, using standard function notation, we may express the cellular objective symbolically as

$$f(\mathbf{v}) = f(v_1, \ldots, v_n) \ : \mathbb{R}^n \to \mathbb{R} \qquad \textbf{(2.7.5)}$$

In writing f as in equation 2.7.5, we have already assumed that the objective depends implicitly on metabolite concentrations and on systems parameters (e.g. temperature and enzyme concentrations). This is because we assumed in equation 2.5.5 that **v** is a function of metabolite concentrations and of system parameters. It is of course possible to express explicitly, the fact that f might also depend explicitly on a set of parameters (for instance as f = f(**v**, **q**) if **q**, possibly identical to **p,** were to denote a vector of parameters). However, we find this unnecessary, and will therefore adhere to the notation in 2.7.5.

Having introduced the objective, we may now state the mathematical condition for equation 2.7.5. Let f_S^{opt} denote the supremum (or infimum) of f(**v**) over $\mathbf{v} \in S \subseteq \mathbb{R}^n$ with ∞ (or $-\infty$) allowed.

$$\mathbf{v} = \mathbf{v_{ss}^{opt}}$$

if **(2.7.6)**

$$\mathbf{S} \cdot \mathbf{v} \ = \mathbf{0} \text{ and } f(\mathbf{v}) = f_{\mathrm{Null}(\mathbf{S})}^{\mathrm{opt}}$$

Where Null(**S**) denotes the null-space of the matrix **S**. Note that writing $f(\mathbf{v}) = f_{\mathrm{Null}(\mathbf{S})}^{\mathrm{opt}}$ on its own in the above is enough to assert that $\mathbf{S} \cdot \mathbf{v} \ = \mathbf{0}$. We will also give an explanation later why we have not written "iff." Since the entries of **v** represent rates of biochemical reactions, it is seen to be a trivial physical requirement that $\|\mathbf{v}\| < \infty$ holds strictly. Thus **v** is physically restricted to reside inside of some bounded subset of $\mathbb{R}^n$, which we will denote by S_{bd}. If we further assume that $S_{bd}$is also a closed set, then provided that f is restricted to and continuously defined over S_{bd}, there will exist a finite number f^{max} such that $f(\mathbf{v}) \leq f^{max}$ and so $f_{S_{bd}}^{opt} = f^{max}$ exactly (an analogous result is clearly true for a minimal value of f). In this light, the (maximizing) FBA problem may be formulated as follows:

Given a stoichiometric matrix **S**, a closed bounded set $S_{bd} \subseteq \mathbb{R}^n$, and a cellular objective f(**v**) defined continuously over S_{bd}, solve for a velocity $\mathbf{v} \in S_{bd} \cap \text{Null}(\mathbf{S})$ such that $f(\mathbf{v}) = f^{max}$. As per equation 2.7.3, provided with one **v** meeting the above conditions, $\mathbf{v_{ss}^{opt}}$ is defined using equation 2.7.4 as $\mathbf{v} = \mathbf{v_{ss}^{opt}}$.

The issues of what specific objective to use and similarly of how to define S_{bd} are biological, in that they depend on there being a specific investigative setting, and do not readily conform to a general mathematical treatment beyond that already given in this section. Thus, the next two sections provide a biologically oriented discussion of objectives and of bounds. Based on what we have described already however, it may be apparent that efficient computational methods for finding the optimal value of f(**v**) over $S_{bd} \cap \text{Null}(\mathbf{S})$ are required for the above formulation of FBA to be of any practical use. In accordance with this requirement, we find that the objective is generally chosen to be a linear function of **v** (e.g. extremizing an inner product of **v** with a constant n-vector) or a quadratic function of **v** (e.g. extremizing $\|\mathbf{v}\|$); similarly, $S_{bd} \cap \text{Null}(\mathbf{S})$will be seen generally to be a convex subset of $\mathbb{R}^n$.

An important observation to be noted in the above formulation of FBA is the potential non-uniqueness of the quantity $\mathbf{v_{ss}^{opt}}$. There is no reason for a **v**, at which an objective is extremized, to be unique. In practice, FBA solutions are well-known to have multiple such **v**'s, an issue termed "degeneracy" which, although we recognize is important, find to be beyond the scope of our simple description of FBA. Thus, in the above formulation, we have assumed that one **v** has been found to optimize f, and that this **v** is arbitrarily used to define $\mathbf{v_{ss}^{opt}}$. It should be noted that the fact that a set of optimal solutions may exist in general, and that $\mathbf{v} = \mathbf{v_{ss}^{opt}}$is just one member of such a set is quite important to many FBA applications. A good example of this is when FBA is used for gene essentiality analyses, as will be seen in

the next chapter. However, we find that degeneracies are ignored in all dFBA applications that we are aware of.

Another issue worth considering is the question of whether a particular FBA solution such as $\mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}}$ contains information about steady-state metabolite concentrations. The traditional view appears to be that FBA methods cannot be used in predicting steady-state metabolite concentrations because FBA methods do not contain kinetic information (apart from the bounds) (Orth, Thiele et al. 2010). We find in light of equation 2.5.5 however that this is not entirely the case. Having quantified $\mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}}$ using FBA, one would have also implicitly quantified the LHS of one or both of equations 2.7.1 and 2.7.2 regardless of the formulation of FBA not depending explicitly on a velocity map. As a consequence of this fact, in combination with a particular $\mathbf{p}^{\mathbf{opt}}$, possibly found independently of an FBA formulation, a quantification of a $\mathbf{v}_{\mathbf{ss}}^{\mathbf{opt}}$ is also an implicit quantification of a steady-state vector of concentrations $\mathbf{c}_{\mathbf{ssp}^{\mathbf{opt}}}$. This point may be illustrated by observing that a kinetic equation such as 2.3.5 for a single reaction asserts a steady-state metabolite concentration when both the velocity and the parameters are given. Provided that the latter have been quantified within a theoretically consistent framework, a concentration is asserted within that same theory by knowledge of velocity and parameters even when these have been independently derived. Thus, FBA solutions do factor into the assertion of steady-state concentrations as they may be used to reverse-engineer concentrations when Michaelis-Menten parameters are known.

2.8: Biological constraints

Whilst we have given a simple common-sensical reason as to why **v** should be bounded, the theoretical formulation of FBA does not give any indications as to what those bounds should be numerically. Moreover, the reduction of Null(**S**) to $S_{bd} \cap \text{Null}(\mathbf{S})$ may

rely, not just on the assertion that the rates of individual reactions are bounded, but perhaps also on an assertion that the velocities of several reactions must satisfy some algebraic equation (we provide a specific example of such a case below). Defining the set S_{bd} requires biochemical knowledge or assumptions specific to the context of an FBA application. The same can be said of the objective function f(**v**). The FBA formulation requires that a cellular objective be defined, but does not indicate specifically what that function should be. We now discuss the FBA constraints and, in the next section, the cellular objective.

Consider the general Null-space equation for **S**,

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0} \qquad \textbf{(2.8.1)}$$

In this section, **v** is not a function, but is interpreted to represent a general point of $\mathbb{R}^n$. FBA seeks the set of points that satisfy 2.8.1 provided with **S**. The vector **v** of rates is taken from $\mathbb{R}^n$ which may be decomposed (as a direct sum)

$$\mathbb{R}^n = \text{Null}(\mathbf{S}) \oplus \text{Null}(\mathbf{S})^{\perp} \qquad \textbf{(2.8.2)}$$

where we have used the "inverted T" to indicate the orthogonal complement of Null(**S**). The biological steady-state assumption requires that $\mathbf{v} \in \text{Null}(\mathbf{S})$. Asserting that **v** should also be constrained to obey a set of equations (inequalities and equalities) serves to further bipartition Null(**S**) into those vectors that satisfy these equations, and those that do not. We will henceforth refer to these equations as "constraints," and refer to the subset of Null(**S**) consisting of those **v** satisfying the constraints as the "feasible region" of the FBA problem and denote it by S_f. Thus we write

$$\text{Null}(\mathbf{S}) = S_f \cup \text{Null}(\mathbf{S}) \backslash S_f \qquad \textbf{(2.8.3)}$$

And require that $v \in S_f$. Since S_f is the intersection of a bounded subset of $\mathbb{R}^n$ and Null(**S**), we may also write

$$S_f = S_{bd} \cap \text{Null}(\mathbf{S}) \quad \textbf{(2.8.4)}$$

to see again that the feasible region is the subset of $\mathbb{R}^n$which results from asserting a steady-state assumption together with further biological constraints.

Two constraints may be applied to the rate of any single reaction (v_j) on its own which are an upper bound and a lower bound. Denoting these by v_j^{Max} and v_j^{Min}, the rate of the jth reaction may be constrained as

$$v_j \in \left[v_j^{Min}, v_j^{Max}\right] \quad \textbf{(2.8.5)}$$

The value that is assigned to the bounds depends on the reaction and the context. Thus, we illustrate with some biological examples. If a reaction is known to be irreversible, we may write $v_j^{Min} = 0.0$ for this reaction. The consumption rate of oxygen is commonly assumed to be bounded above at a rate of about $20.0\ \text{mmol GDW}^{-1}\ \text{h}^{-1}$. Letting v_{O_2} denote the consumption rate of oxygen, if it is assumed for simplicity that a respiring cell does not produce oxygen, this fact may be represented by writing $v_{O_2} \in [0.0, 20.0]$. If the rate of a particular reaction is known exactly, we may set $v_j^{Min} = v_j^{Max}$ for this reaction. A common situation where such a bound may be used is when the rate of consumption of a substrate has been measured exactly in an experiment. Let v_{subst} and v_{subst}^{obs} denote the FBA rate of substrate consumption and the experimentally observed rate of substrate consumption respectively. Then the uptake rate of substrate may be constrained as $v_{subst} \in \left[v_{subst}^{obs}, v_{subst}^{obs}\right]$. A second example of where a reaction rate is known exactly is when the gene for the enzyme catalysing the process has been deleted; in this case, the reaction is not expected to occur at all, and we may write $v_j \in [0.0, 0.0]$. Finally, if no information is available about the bounds of a reaction v_j, this fact may be represented by simply asserting that $v_j$should at the very least be finite. Thus, taking a finite, but large value for velocity,

10^6mmol GDW^{-1} h^{-1}, for instance, we may write $v_j \in [0.0, 10^6]$ if the reaction is known to be irreversible, and $v_j \in [-10^6, 10^6]$ otherwise. In some less common instances, an FBA application will require that a set of velocities obeys a simple algebraic equation which is usually linear. As an example of this, Feist et al. have made the assumption that the rates across the two NADH dehydrogenases at the start of the *E.coli* electron transport will be in a 1:1 ratio (Feist, Henry et al. 2007). Letting v_{NADH1} and v_{NADH2} represent the FBA predicted rates for these dehydrogenases, this assumption is represented as $v_{NADH1} = v_{NADH2}$, a linear constraint on **v** that is not a simple min/max constraint.

Equation 2.8.4 illustrates that the one effect of applying biological constraints is that the resulting feasible region will have a smaller volume relative to the unbounded set Null(**S**) of all steady-state vectors. When all constraints are linear, as in the example $v_{NADH1} = v_{NADH2}$ from above, the feasible region will also be seen to be the polyhedral subset of $\mathbb{R}^n$ described by the system of the form

$$\begin{aligned} \mathbf{S} \cdot \mathbf{v} &= \mathbf{0} \\ \mathbf{L} \cdot \mathbf{v} &= \mathbf{w} \\ \mathbf{v^{Min}} \leq \mathbf{v} &\leq \mathbf{v^{Max}} \end{aligned} \qquad \textbf{(2.8.6)}$$

In 2.8.6, **L** and **w** are constant matrix and vector representing the equality constraints whose number of rows equal the number of such equations; $\mathbf{v^{Min}}$ and $\mathbf{v^{Max}}$ are n-vectors whose jth entry is respectively, the lower and upper bound corresponding to v_j; the inequality signs are understood to be applied component-wise. We do not fully consider in this section, the adjustments to the algebraic description of Null(**S**) that is required to describe S_f as they are technical points on which our results do not depend. However, it is worth mentioning that while any point of S_f will be a linear combination of the entries of a basis for Null(**S**) (since $S_f \subseteq$ Null(**S**)) the converse is not generally true. To illustrate this point using a simple example, we note that whilst any point in the first quadrant of $\mathbb{R}^2$ may be represented as a

linear combination of members in the standard basis for $\mathbb{R}^2$, that the span of this basis covers also the second, the third, and the fourth quadrants. A consideration of only those linear combinations whose coefficients are positive will restrict the results to the first quadrant. However, if we introduce the additional requirement that the result also lie in a polygonal subset in the first quadrant, then further restrictions on the coefficients are seen to be needed. The requirement will be in general that we start with a convex basis (of vectors corresponding to the corners of the polyhedron) and to take a convex combination of these elements (the coefficients must not just be positive, but also sum to 1). Interpreting this representation biologically, we are asserting that a biological steady-state velocity **v** lies in a convex subset S_f of $\mathbb{R}^n$and that a point of S_f may be described by a convex combination of velocities each corresponding to a extreme point (a corner) of S_f.

2.9: Biological objectives

We have defined the objective function to generally be any real valued function f(**v**) of velocity (equation 2.7.5), and have assumed that f(**v**) is continuous over the set of velocities for which it is defined. The biological rationale for the latter assumption is that phenotypes are expected to be continuously defined functions of metabolic velocities. In light of the previous section, we make the additional assumption in this section that the domain of f(**v**) has also been restricted to the feasible region S_f for an FBA application. In light of the description of FBA given in section 2.7, we see that FBA can be expressed by the following optimization problem.

Optimize:

$$f = f(\mathbf{v})$$

Subject to: **(2.9.1)**

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0}$$
$$\mathbf{L} \cdot \mathbf{v} = \mathbf{w}$$
$$\mathbf{v}^{\mathbf{Min}} \leq \mathbf{v} \leq \mathbf{v}^{\mathbf{Max}}$$

Or equivalently, to solve for a **v** such that

$$f(\mathbf{v}) = f_{S_f}^{opt} \qquad \textbf{(2.9.2)}$$

In order to ensure that the optimization problem is computationally tractable, we will also require that the objective is either a linear or quadratic function of velocity. Note that this is a computability requirement and not a biological one. Having explicitly described the FBA framework, we now discuss some examples of biological objectives that are commonly used.

The most common cellular objective used in FBA modelling is to maximize the flux of metabolites towards the production of a unit of biomass. We provide a comprehensive treatment of this objective function which we call "the biomass objective" in section 3.3. The biomass objective is an inner product of **v** with a constant n-vector representing the metabolite composition that is stoichiometric with respect to a unit of biomass. In choosing a **v** that maximizes this objective, one assumes that cellular reaction rates are such that the reproduction rate, as represented by a stoichiometrically balanced biosynthetic rate, is maximized. Another example of a related linear objective is the maximization of just one or few of the reaction rates $v_1 \ldots v_n$ (it will be seen in section 3.3 that the biomass objective is a special case). As an example, if v_j represents the rate across a reaction needed for the production of a commercially relevant compound, the FBA problem of maximizing the objective $f(\mathbf{v}) = v_j$ may be used to solve for a **v** which may through engineering help increase the production rate of the desired compound.

Quadratic objective functions often arise from FBA problems that consider Euclidian distances involving $\mathbf{v}$. One example occurs when the FBA is formulated to minimize the objective $f(\mathbf{v}) = \|\mathbf{v}\|$ which is an instance of "minimization of overall flux". A linear alternative to flux minimization would be to represent reversible reactions by two separate reactions, for the forward and the reverse processes, so that all velocities would be positive (Holzhütter 2004). The biological rationale for this objective is as follows. Since enzymes are required to catalyse reactions and a cell's proteomic composition is budgeted, the combined rates across reactions, each term being proportional to an enzyme concentration, would be minimized so as to make the most efficient use of the available proteome. Another example where a quadratic optimization is required involves "minimization of metabolic adjustment" (MOMA) (Segre, Vitkup et al. 2002). Here one considers two velocities, $\mathbf{v_{WT}}$ and $\mathbf{v_{MT}}$, representing the velocities associated with a wild-type and a mutant metabolic system. In MOMA, the wild-type velocity is calculated from an FBA application is fixed to be that solution. In calculating the mutant velocity, one introduces further constraints, to the wild-type FBA problem, representing the metabolic perturbations that differentiates the mutant system from the wild-type system. Biologically, because mutations such as those that have been engineered have not been present during the evolutionary adaptive phase of populations, the perturbed metabolic system has likely not yet had the time to be tuned to deal optimally with conditions for which the wild-type system does behave optimally. In modelling such sub-optimal mutant behaviour, the MOMA approach is to make the assumption that a metabolic system will respond to a perturbation by attempting to best approximate the velocity $\mathbf{v_{WT}}$ exhibited by the wild-type. In this way, MOMA models sub-optimal metabolism by solving for a $\mathbf{v_{MT}}$ in the perturbed feasible region for which the objective function $f(\mathbf{v_{MT}}) = \|\mathbf{v_{WT}} - \mathbf{v_{MT}}\|$ with a fixed $\mathbf{v_{WT}}$ is minimized.

With cellular objectives now having been defined, this section completes our description of general flux balance analysis. The method is summarized as follows. FBA is an approach to structural metabolic modelling comparable to methods such as MCA where metabolism is treated holistically as a system rather than by focusing on any local set of reactions in isolation. The dynamical representation of metabolism is given by equation 2.5.9 which characterizes how dynamics and structure combine to determine the time-derivatives of metabolite concentrations as functions of metabolite concentrations and system parameters. The general approach to dealing with equation 2.5.9 in practice is by bringing it to a steady-state and studying its linearization. The formulation of FBA however starts, not with equation 2.5.9, but with 2.8.1; that is to say that FBA starts with the stead-state equation without an explicit consideration of velocity as a mapping as in equation 2.5.5. Whereas MCA is a sensitivity analysis, FBA is an optimization problem. Starting with equation 2.8.1, the FBA approach uses biological constraints to arrive at a convex subset of Null(**S**), the feasible region, and a cellular objective is optimized over this subset of possible velocities. In this way, any solution to the FBA problem is a vector **v** of velocities that satisfies the steady-state equation and optimizes the objective over the feasible region.

2.10: Dynamic Flux Balance Analysis

The metabolic models described up to this section both require that metabolite concentrations, except for those with exchange reactions, are at steady-state values. This is because both approaches depend on equation 2.6.3 holding for some particular parameter set. Most metabolic phenomena occur however outside of steady-state conditions. Thus, it would be desirable to be able to extend steady-state metabolic models to more realistic dynamic settings. Dynamic Flux Balance Analysis (dFBA) is a non-steady state extension of FBA that combines the FBA approach with ODE modelling. Computationally, dFBA is implemented by iteratively performing optimization and integration over some time interval of interest

(Mahadevan, Edwards et al. 2002). We first describe the algorithm and then consider the biological details. Thus, assume that we are starting with some initial condition of metabolites and enzymes for the system of interest $\mathbf{c_0}$ and $\mathbf{e_0}$. Further, assume that there are two models, an FBA model and an ODE model denoted $\mathbf{M_{FBA}}$ and $\mathbf{M_{ODE}}$ respectively, with parameters $\mathbf{v^{Min}}, \mathbf{v^{Max}}, \mathbf{L}, f,$ and $\mathbf{p_{ODE}}$. With these inputs, a simple dFBA algorithm starting at time t_0 and moving with a step-size Δt in N steps to the terminal point $t_0 + N\Delta t$ may be expressed as

Input:

Variables: Metabolite_concentrations = $\mathbf{c_0}$

Enzyme_concentrations = $\mathbf{e_0}$

Models: FBA_model = $(\mathbf{M_{FBA}}, \mathbf{v^{Min}}, \mathbf{v^{Max}}, \mathbf{L}, f)$

ODE_model = $(\mathbf{M_{ODE}}, \mathbf{p_{ODE}})$

Time: $t \in [t_0, t_0 + N\Delta t]$

Algorithm: dFBA

START

Set FBA_model = $(\mathbf{M_{FBA}}, \mathbf{v^{Min}}, \mathbf{v^{Max}}, \mathbf{L}, f)$

Set ODE_model = $(\mathbf{M_{ODE}}, \mathbf{p_{ODE}})$

Set Metabolite_concentrations = $\mathbf{c_0}$

Set Enzyme_concentrations = $\mathbf{e_0}$

Set $t = t_0$

Set total_steps = N

Set setp_size = Δt

while $(t < t_0 + N\Delta t)\{$

1. Record Metabolite_concentrations
2. Record Enzyme_concentrations
3. Optimise the FBA_model

4. Retrieve $\mathbf{v_{ss}^{opt}}$ and $f_{S_f}^{opt}$ from the optimisation step
5. Update $\mathbf{p_{ODE}}$ using $\mathbf{v_{ss}^{opt}}$ and $f_{S_f}^{opt}$

6. Set Initial_condition = (Metabolite_concentrations, Enzyme_concentrations)
7. Integrate ODE_model over $[t,\ t + \Delta t]$ from Initial_condition
8. Update Metabolite_concentrations
9. Update Enzyme_concentrations
10. Update $\mathbf{v^{Min}}$, $\mathbf{v^{Max}}$, and $\mathbf{L}$ using Metabolite_concentrations and Enzyme_concentrations
11. Update f($\mathbf{v}$) using Metabolite_concentrations and Enzyme_concentrations <if applicable >

12. Set $t = t + \Delta t$

} // **end while**

STOP

Output:

A time-series of metabolite concentrations and enzyme concentrations:

$$\mathbf{t} = (t_0 \ldots t_{N-1})$$

$$\mathbf{C} = (\mathbf{c_0} \ldots \mathbf{c_{N-1}})$$

$$\mathbf{E} = (\mathbf{e_0} \ldots \mathbf{e_{N-1}})$$

Thus, dFBA is an algorithm that iteratively applies FBA optimization and ODE integration, thereby introducing a feedback of information between the parameter sets of the two models. The optimization step (step 3) involves the application of FBA at a time point to arrive at a flux distribution under the assumption of an instantaneously achieved metabolic steady state. The results of the FBA are used to update the parameters of the ODE (step 5), and the resulting equations are integrated in the ODE step to advance the algorithm over a small partition of the time interval (step 7). The results of the integration step are used to update the parameters of the FBA step (step 10). It is possible, though not often done, for the

objective to be redefined in response to the ODE outputs (step 11). Two examples are found in Pramanik and Keasling 1997, Meadows et.al. 2010 where the biomass objective with a growth-rate dependent composition is used (Pramanik and Keasling 1997, Meadows, Karnik et al. 2010). The two computations – optimization and integration – are repeated sequentially until the final time point is reached and the dFBA algorithm terminates. The output of the algorithm is a simulated time series in metabolite and enzyme concentrations (Mahadevan, Edwards et al. 2002, Antoniewicz 2013). Note that we have used the term "metabolite" in our description to include any node of an FBA model, including possibly a "biomass node." The significance of this node is made clear in the next chapter, where we discuss the biomass reaction and "Palsson-style" FBA models.

It should be noted that the concentrations of metabolites treated dynamically in dFBA models should be considered now to be parameters in equation 2.5.5 because they are not required to return to a steady-state value as in equation 2.6.2. With this in mind, the key biological assumption of dFBA models is represented by step 3 of the routine in the while-loop. dFBA models assume that both the cellular objective and the steady-state condition are restored at every FBA step. This is seen to be both an assumption on the stability of the system 2.5.9 as well one on the biological properties of metabolism itself. It is recalled that the mathematical formulation of FBA, unlike MCA, does not explicitly require equation 2.6.4, but only equation 2.6.1 (although stability may be biologically implied). For dFBA, the structural stability of equilibria is now an explicit mathematical requirement, for otherwise, there is no guarantee for steady-state conditions to be returned to after parameter perturbations occur. More subtly, dFBA, being a sequence of FBA operations, requires that the steady-state equation 2.6.4 holds for the first term of that sequence. However, as pointed out by Gunawaradena, there is no mathematical reason for the system 2.5.9 to have a stable fixed point in the first place; a well-known example is glycolysis, a metabolic system that has

been shown experimentally to exhibit oscillatory dynamics rather than arriving at a simple steady-state (Gunawardena 2002). Thus, by assuming the structural stability of the system 2.5.9 to achieve instantaneous steady-state conditions over the FBA iterations, dFBA methods may exclude several important biological phenomena from the scope of its modelling. The second assumption of dFBA is that not only is the steady-state condition restored for **S** (equation 2.6.3), but the optimality condition is also restored (equation 2.7.3). Thus, the parameter set **p** in equation 2.5.5 is assumed biologically to be flexible such that, after a perturbation to its entries, equation 2.7.2 is restorable by reconfiguring its other entries in response.

With some success, the dFBA approach has been used towards reproducing experimental growth response to dynamic respiratory constraints in yeast, therapeutic protein production by *E.coli*, diauxic growths, and co-culture fermentation of lignocellulosic feedstock (Mahadevan, Edwards et al. 2002, Meadows, Karnik et al. 2010, Hanly and Henson 2011, Hanly, Urello et al. 2012, Jouhten, Wiebe et al. 2012). However, unlike for FBA, it remains to be investigated for dFBA whether results of practical use may be gained from this modelling approach.

2.11: Discussion

In this chapter, we have introduced the reader to the fundamental ideas underlying structural metabolic modelling. These models are required as a part of the systems biological approach to the interrogation of metabolism as they allow local kinetic models of reactions (e.g. equation 2.3.5) to be related to global metabolic behaviour (equation 2.5.9). Fundamental to this approach is the stoichiometric matrix which represents the topology of a metabolic network. FBA methods use the stoichiometric matrix to write equation 2.8.1 to solve the problem formulated in 2.9.1. As a member of a broader family of metabolic models

however, we find it may be beneficial to consider FBA methods in the context of the dynamical equation for metabolism that is given by equation 2.5.9. In doing so, we have made explicit in this chapter, the mathematical and biological assumptions that are made by FBA and its dynamic extension, dFBA.

CHAPTER 3: Two approaches to Flux Balance Analysis

Abstract:

Most commonly, FBA models used in practice are genome-scale models that employ a biomass objective, where the biomass composition is treated statically using stoichiometric coefficients derived from the existing literature. Here we derive the mathematical closed form of this objective from the FBA steady-state equation, and deconstruct the biomass objective of the iAF1260 *E.coli* metabolic model to reveal the exact sources of its stoichiometric coefficients. FBA models are neither required to use a biomass objective nor a genome-scale metabolic representation. In consideration of this fact, we find that a large number of FBA methods in practice may be characterized into one of two inverse types based on scale, and on their treatment of growth rates.

3.1: Introduction

Formally, FBA and dFBA are members of a broad collection of methods known as COnstraint-Based Reconstruction and Analysis (COBRA). The "constraint" aspect of COBRA methods has been described in chapter 2 where it was seen that FBA is mathematically formulated as a constrained optimization problem (formulation 2.9.1). The "reconstruction" on the other hand refers to the integration of data in the construction of the stoichiometric matrix. A protocol for reconstruction has been developed by Palsson et al. (2010) and involves the synthesis of a detailed metabolic pathway map of all conversions between metabolites that are known to occur in an organism (Thiele and Palsson 2010). This process commonly termed "network reconstruction" involves the extensive manual data mining of the literature that is available for a particular organism. Largely bioinformatic by nature, network reconstruction depends critically on there being a strong knowledge base for the organism such as that represented by EcoCyc for *E.coli* (see section 1.2). The reconstructed network is a computational representation of an organism's metabolic system in the form of a stoichiometry matrix. Since an extensive knowledge base is available for *E.coli*, this organism is thought to have one of the most complete network representations. We will henceforth refer to the available genomic and metabolomic data as "genotype data"

for short. Thus, as a systems biological approach, FBA methods seek to relate genotype data to growth phenotype through the mathematical formulation 2.9.1.

In the effort to relate genotype data to growth phenotype, there is currently a noticeable trend in the COBRA literature which is towards describing metabolic systems in ever finer detail; in other words, by striving towards a complete a stoichiometric model as possible based on the availability of data. This trend is seen decidedly in the iJE660, iJR904, iAF1260, iJO1366 series, where the final model is of such a complete genomic coverage that it is thought to be expandable only by the addition of newly characterized reactions (Edwards and Palsson 2000, Reed, Vo et al. 2003, Feist, Henry et al. 2007, Orth, Conrad et al. 2011, McCloskey, Palsson et al. 2013). Work is currently under way in the construction of (steady-state) Metabolism and Expression models (ME-models) that combine genomic scale protein expression data with the metabolic data (O'Brien, Lerman et al. 2013). Similarly, we note that this trend towards increasing scale is represented, not just by the expansion of models, but also by the increasing scale of the bio-systems the models are constructed to represent; ecosystem models, such as of bio-films and of interspecies interactions, as well as whole-organism models (e.g. of *Arabidopsis thaliana*) have now been published (Poolman, Miguet et al. 2009, Hanly and Henson 2011, Hanly, Urello et al. 2012, Biggs and Papin 2013). Thus, the most common COBRA models encountered in practice are large and account for an extensive volume of genotype data, a fact often strongly implied as when FBA models are referred to as "genome scale" models.

There is an assumption represented by these trends that a more comprehensive coverage of genotype will lead to correspondingly more accurate predictions of phenotype. Model size is advocated to increase with data availability, for instance as when the construction of a biomass objective is described to progress through three levels of detail (from basic to intermediate to advanced) based on this criteria (Feist and Palsson 2010).

However, the relation between accuracy and detail must certainly be contingent upon exactly what phenotype is being investigated in a COBRA framework; that is, how one has defined a "growth phenotype" in the first place. Moreover, insofar that there is a diversity amongst biological phenotypes which may be investigated by FBA methods, it would be expected that there should also be a correspondingly wide continuum of metabolic resolutions with which to best investigate. As a demonstration of phenotype ambiguity, authors Feist and Palsson mention in one article several phenotypes referring to "growth rates," "cellular yields," "flux distributions," "optimal network states" and "phenotypic states," "essentialities," "topological properties," and "biophysical capabilities" (Feist and Palsson 2010). Neglecting the fact that some of these terms come with somewhat vague definitions (i.e. "network states" and "biophysical capabilities"), it is also important to point out that lists such as this one combine phenotypes that become relevant at differing biological scales. Flux distributions describe "possible metabolic routs through the metabolic network of an organism" and are thus phenotypes that require a high resolution description of the network to investigate (Schuster, Pfeiffer et al. 2008). Specific growth rates in contrast, describe the per-capita reproductive contribution of individuals to the total population growth rate (equation 1.7.1) and, may not be a phenotype requiring as highly a detailed metabolic model to investigate. Whilst it is the stated goal of systems biology to bridge the gaps between various biological scales, nowhere has it been demonstrated that such bridges are formed automatically with increased descriptive resolution; likewise, nowhere has it been demonstrated that high resolution models are prerequisite to the investigation of multi-scale biological phenomena (Kitano 2002, Kitano 2002). The Palsson protocol for network reconstruction invariably culminates with a genome-scale model however, in which an organism's metabolism is described at the finest possible resolution with respect to the available biological information. Thus, we may not readily identify from the protocol, what is the relevant scale at which reconstructed

networks should be studied when a particular set of growth phenotype predictions is of interest. It is therefore an open question if large-scale stoichiometric models are always appropriate.

One area in which large-scale FBA has been notably successful is in the identification of unintuitive pathways. We define an "unintuitive pathway" here to mean a functional path along a metabolic network, between two metabolites, that is made up of a previously uncharacterized combination of reactive steps. Relying on large FBA models (and growth-rate optimization), investigators have been able to identify engineering targets for the microbial production of commercial compounds, 1,3-propanediol and L-threonine; the experimental modifications of the predicted targets were later shown to result in increased product yields (Nakamura and Whited 2003, Lee, Park et al. 2007, McCloskey, Palsson et al. 2013). Another topological growth phenotype for which large-scale FBA models are known to make good predictions is gene essentiality. FBA models have been used to correctly identify "essential genes" defined to be those which, if knocked out, cause the growth of cells to cease completely. Note that if essentiality predictions do not match experimental observations, then a "knowledge gap" has been identified which would equally be of value to biology (McCloskey, Palsson et al. 2013). Thus, for topological growth phenotypes, it appears unsurprisingly that larger metabolic models do in fact correspond to better predictions.

In contrast to these examples, many applications of FBA are concerned first and foremost with metabolic rate phenotypes. The dFBA method (section 2.10) for instance has been developed specifically for the investigation of the cellular reprogramming of flux distributions in response to dynamically changing metabolic rate parameters (Mahadevan, Edwards et al. 2002). Topology and rates are two distinct aspects of metabolic systems however. As a result, it does not follow automatically that larger metabolic models will

correspond to more accurate predictions of rates as it seems to do for topology. Thus the necessity of genome scale metabolic representations in applications of FBA that are concerned primarily with rates warrant further investigation. We consider in this chapter how growth rates are represented by FBA models, and identify two reciprocal classes of FBA models based on the treatment of growth rates.

3.2: The biomass objective function

The application of FBA methods requires not just a network structure, but also a network function. This is what a metabolic network is assumed to have been programmed to do. In sections 2.7 and 2.9 of the previous chapter, it was seen that this functional aspect of metabolism is represented in FBA models by objective functions. In this section, we consider one specific objective function, called the "biomass objective," which is the one most commonly used in practice (Feist and Palsson 2010, Yuan, Cheung et al. 2016). The biomass objective is used whenever a modeler assumes that a metabolic network has been programmed to maximize the specific growth rate (μ) of a cell. As was described in section1.7, μ is the per-capita growth rate as understood in a standard Malthusian growth equation. Here, we describe how μ may be formulated mathematically in an FBA model by deriving a closed-form expression for μ starting with the steady-state equation 2.8.1.

In FBA models, specific growth rate is defined to be the rate of flow of a stoichiometric collection of metabolites to a unit of biomass. We will later make a distinction between two inverse approaches to FBA based on whether stoichiometry (of biomass metabolic composition) is used to predict μ, or instead, if μ is used to predict stoichiometry. In this section, the former situation is assumed. We will also continue to adhere to the notational conventions that were introduced in chapter 2. So as to be consistent with the

notation in section 1.7, where the Greek letter μ was used to signify specific growth rate, we will represent the biomass objective as

$$\mu(\mathbf{v}) = \mu(v_1, \ldots, v_n) \ : \mathbb{R}^n \to \mathbb{R} \qquad \textbf{(3.2.1)}$$

Thus, we replace the generic function notation introduced in equation 2.7.5 with μ to indicate that we are now dealing explicitly with a biomass objective. The replacement of f with μ will be carried through to all symbolic notations for which f may be used. Thus for instance, a maximal value for the biomass objective will be indicated by μ^{Max}. Note that this notation in this example coincides with the notation used to denote a generic maximal specific growth rate, appearing in equations such as 1.7.5 and 1.7.8, which may be defined without an FBA model. It now follows from the FBA formulation in 2.9.1 that an FBA problem with the objective of maximizing a biomass objective may be expressed as

Maximize:

$$\mu = \mu(\mathbf{v})$$

Subject to: **(3.2.2)**

$$\begin{aligned} \mathbf{S} \cdot \mathbf{v} &= \mathbf{0} \\ \mathbf{L} \cdot \mathbf{v} &= \mathbf{w} \\ \mathbf{v^{Min}} \leq \mathbf{v} &\leq \mathbf{v^{Max}} \end{aligned}$$

For succinctness, we will often only write the steady-state constraint explicitly when defining an FBA problem; similarly, we may write matrices and vectors in an expanded form if doing so leads to increased clarity. Thus, the above formulation may alternatively be written as

Maximize:

$$\mu = \mu(\mathbf{v})$$

Subject to: **(3.2.3)**

$$\begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Note that we are using n and m as in chapter 2 to represent the total number of reactions and the total number of metabolites occurring in the model respectively. In the following, we will derive an explicit formula for the biomass objective 3.2.1.

Starting with the steady-state equation for FBA 2.8.1,

$$\begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.2.4}$$

we augment the system with an extra reaction to arrive at the $m \times (n + 1)$ dimensional equation

$$\begin{pmatrix} s_{11} & \cdots & s_{1n} & s_{1b} \\ \vdots & \ddots & \vdots & \vdots \\ s_{m1} & \cdots & s_{mn} & s_{mb} \end{pmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \\ v_{biomass} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.2.5}$$

Isolating the newly added term of the equation gives

$$\begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = (-1) v_{biomass} \begin{bmatrix} s_{1b} \\ \vdots \\ s_{mb} \end{bmatrix} \tag{3.2.6}$$

It will be seen that the FBA rate predicted for $v_{biomass}$ will be taken to be the specific growth rate prediction. Thus, the RHS of 3.2.6 represents the rate at which metabolites must be stoichiometrically drawn towards biomass so as to balance the metabolic fluxes on the LHS. The vector that is scaled by $v_{biomass}$ contains the coefficients for each metabolite comprising a unit of biomass. We define this vector as $\mathbf{b_{rxn}}$

$$\mathbf{b_{rxn}} = \begin{bmatrix} s_{1b} \\ \vdots \\ s_{mb} \end{bmatrix} \tag{3.2.7}$$

Re-expressing 3.2.6 with this notation, we arrive at

$$\mathbf{S} \cdot \mathbf{v} = (-1) v_{biomass} \mathbf{b_{rxn}} \tag{3.2.8}$$

Taking an inner product with $\mathbf{b_{rxn}}$,

$$\langle \mathbf{S} \cdot \mathbf{v},\ \mathbf{b_{rxn}} \rangle = (-1) v_{biomass} \|\mathbf{b_{rxn}}\|^2 \tag{3.2.9}$$

and isolating $v_{biomass}$ gives

$$(-1)\frac{\langle \mathbf{S}\cdot\mathbf{v},\ \mathbf{b_{rxn}}\rangle}{\|\mathbf{b_{rxn}}\|^2} = v_{biomass} \quad \textbf{(3.2.10)}$$

Replacing the stoichiometric matrix with its transpose gives

$$(-1)\frac{\langle \mathbf{v}, \mathbf{S^T}\cdot\ \mathbf{b_{rxn}}\rangle}{\|\mathbf{b_{rxn}}\|^2} = v_{biomass} \quad \textbf{(3.2.11)}$$

Let the constant vector **b** be defined as

$$\mathbf{b} = \left(\frac{-1}{\|\mathbf{b_{rxn}}\|^2}\right)\mathbf{S^T}\cdot\ \mathbf{b_{rxn}} \quad \textbf{(3.2.12)}$$

Substituting expression 3.2.12 into 3.2.11, we arrive at the desired expression for $v_{biomass}$

$$v_{biomass} = \langle \mathbf{v}, \mathbf{b}\rangle \quad \textbf{(3.2.13)}$$

Defining the rate prediction for $v_{growth}$to be the specific growth rate prediction, we may write

$$\mu = v_{biomass} \quad \textbf{(3.2.14)}$$

Taking the specific growth rate as the objective, the FBA problem 3.2.3 may be expressed as

Maximize:

$$\mu(\mathbf{v}) = \langle \mathbf{v}, \mathbf{b}\rangle$$

Subject to: **(3.2.15)**

$$\mathbf{S}\cdot\mathbf{v} \quad = \mathbf{0}$$

The biomass objective is often expressed as a *biomass reaction* in the COBRA literature (Varma and Palsson 1994, Yuan, Cheung et al. 2016). The value of the biomass objective is interpreted, when this is done, to be a flux across the biomass reaction. We now briefly discuss what is meant by these statements. Let a_i for $1 \leq i \leq m$ denote the chemical name that is given to the metabolite that has been indexed by i (e.g. a_1 might stand for

"valine"). Likening a unit of biomass to a defined metabolic object, a conceptual chemical equation such as

$$|s_{1b}|a_1 + \cdots + |s_{mb}|a_m \rightarrow \text{biomass} \qquad \textbf{(3.2.16)}$$

may be written as a model of growth. Note that we use absolute values in this equation as the signs on the stoichiometric coefficients are given by their location relative to the reaction arrow. Note also that the notation v_{biomass} was used in accordance with section 2.2 as this represents the rate of production of biomass treated as an individual metabolite. Growth in equation 3.2.16 is the irreversible transfer of mass from a stoichiometric collection of metabolites on the negative side to the biomass on the positive side. The stoichiometric coefficient for the ith metabolite in 3.2.16 is seen to be

$$s_i = |s_{ib}| \qquad \textbf{(3.2.17)}$$

Because biomass reactions often have more terms than would be practical to write out at every occurrence, a notation analogous to the standard notation for sums is used to compact expressions such as 3.2.16.

$$\sum_{i=1}^{m} s_i a_i \rightarrow \text{biomass} \qquad \textbf{(3.2.18)}$$

The expression on the LHS of 3.2.18 is not a mathematical sum. The operations of addition appearing there are not defined in any mathematical sense, but rather, represent the combining of chemical terms (stoichiometric coefficient together with the name of a compound) in expressing a chemical equation. This being stated, we also note that the indices that appear in 3.2.18 are useful as they may be used as a means of isolating particular terms of the chemical equation when appropriate. As one example, it would be appropriate to use the expression

$$s_1 a_1 + \sum_{i=2}^{m} s_i a_i \rightarrow \text{biomass} \qquad \textbf{(3.2.19)}$$

to single out the first term of the overall expression.

3.3: Static biomass objective of the iAF1260 model

Most descriptions of FBA are given using equation 3.2.14 (Orth, Thiele et al. 2010) or the biomass reaction 3.2.18 (Varma and Palsson 1994). These simple equations contains an important assumption about the biomass composition. Namely by fixing **b** to be a constant, the equation asserts that biomass composition ($\mathbf{b_{rxn}}$) is static. This assumption becomes significant when, as is very commonly done, the composition vector is constructed by extrapolating data from measurements that have been obtained under several differing growth conditions. Similarly, the assumption will have implications to dFBA model predictions, where population growth is profiled with a constant biomass vector but in a variable environment.

The application of equation 3.2.13. assumes implicitly that the biomass composition, invariantly of such possible factors as strain, media, and growth conditions in general, remains approximately unchanged. Unfortunately, the FBA method is often presented in the language of linear programming that is foreign to biology and that obscure this assumption (Varma, Boesch et al. 1993, Varma and Palsson 1994, Varma and Palsson 1994, Edwards and Palsson 2000, Edwards, Ibarra et al. 2001, Edwards, Covert et al. 2002, Orth, Thiele et al. 2010). For a more lucid illustration of this assumption, we find it useful to complement the definition of μ in equation 3.2.14 with a biochemically explicit description. In the above setup, FBA identifies the empirical μ with a specific physiological quantity – the optimal velocity of steady-state biosynthesis as represented by a metabolic flux through a phenomenological biomass reaction (equation 3.2.18). In this section and the next, we describe the biomass objective that is used by the iAF1260 metabolic model and reverse-

engineer its biomass constituents to identify the experimental sources of its stoichiometric coefficients (Feist, Henry et al. 2007).

Amino Acids		**Stoich. Abs(mmol /gdw)**
Gly-L	Glycine	0.595
Ala-L	Alanine	0.499
Leu-L	Leucine	0.438
Val-L	Valine	0.412
Lys-L	Lysine	0.333
Arg-L	Arginine	0.287
Ile-L	Isoleucine	0.282
Gln-L	Glutamine	0.256
Glu-L	Glutamate	0.256
Thr-L	Threonine	0.247
Asn-L	Asparagine	0.234
Asp-L	Aspartate	0.234
Pro-L	Proline	0.215
Ser-L	Serine	0.210
Phe-L	Phenylalanine	0.180
Met-L	Methionine	0.149
Tyr-L	Tyrosine	0.134
His-L	Histidine	0.092
Cys-L	Cysteine	0.089
Trp-L	Tryptophan	0.055
Ribonucleotides		
Gtp	GTP	0.209
Utp	UTP	0.140
Ctp	CTP	0.130
Atp	ATP (biomass component)	0.170*****
Deoxy-ribonucleotides		
dCtp	Deoxy-CTP	0.026
dGtp	Deoxy-GTP	0.026
dATP	Deoxy-ATP	0.025
dTtp	Deoxy-TTP	0.025
Inorganic ions		
H2o	Water	54.61
K	Potassium(1+)	0.169
Nh4	Ammonium(1+)	0.011
Mg2	Magnesium(2+)	0.008
Fe2	Iron(2+)	0.007
Fe3	Iron(3+)	0.007
Ca2	Calcium(2+)	0.005
Cl	Chlorine(1-)	0.005
So4	Sulphate(2-)	0.004
Cobalt2	Cobalt(2+)	0.003
Cu2	Copper(2+)	0.003
Mn2	Manganese(2+)	0.003
Mobd	Molybdenum(1+)	0.003
Zn2	Zinc(2+)	0.003
Energy Currencies		
Atp	ATP (GAM)	59.81*
Nad	NAD+	0.002
Nadph	NADPH	0.0003
Nadp	NADP+	0.0001
Nadh	NADH	0.00005**
Other		
Glycogen	Glycogen	0.154
Ptrc	Putrescine	0.033
Pe160[p]	phosphatidylethanolamine	0.032
Pe161[p]	phosphatidylethanolamine	0.025
Pe181[p]	phosphatidylethanolamine	0.013
Pe160	phosphatidylethanolamine	0.012
Pe161	phosphatidylethanolamine	0.010
Colipa	core oligosaccharide lipid A	0.008
Spmd	Spermidine	0.007
Pg160	Phosphatidylglycerol	0.006
Murein4p4p[p]	Murein	0.005
Murein4px4p[p]	Murein	0.005
Pe181	phosphatidylethanolamine	0.005
Pg160[p]	Phosphatidylglycerol	0.005
Pg161	Phosphatidylglycerol	0.004
Pg161[p]	Phosphatidylglycerol	0.004
Clpn160	cardiolipin	0.003
Clpn161	cardiolipin	0.003
Pg181	Phosphatidylglycerol	0.002
Pg181[p]	Phosphatidylglycerol	0.002
Clpn181	cardiolipin	0.001
Murein3p3p[p]	Murein	0.001
Murein3px4p[p]	Murein	0.001
Murein4px4px4p[p]	Murein	0.001
Accoa	Acetyl-CoA	0.0003
10fthf	10-formyltetrahydrofolate	0.0002
2dmmql8	2-Demethylmenaquinol 8	0.0002
5mthf	5-Methyltetrahydrofolate	0.0002
adocbl	Adenosylcobalamin	0.0002
Amet	s-adenosyl-methionine	0.0002
Chor	chorismate	0.0002
Coa	Coenzyme-A	0.0002
Enter	Enterochelin	0.0002
Fad	FAD	0.0002
Gthrd	Reduced glutathione	0.0002
hemeO	Heme O	0.0002
Mlthf	5,10-methylenetetrahydrofolate	0.0002
Mql8	Menaquinol 8	0.0002
Pheme	Proto-heme	0.0002
Pydx5p	Pyridoxal-5'-phosphate	0.0002
Q8h2	Ubiquinol-8	0.0002
Ribflv	Riboflavin	0.0002
Sheme	Siroheme	0.0002
Thf	Tetrahydrofolate	0.0002
Thmpp	Thiamine diphosphate	0.0002
Succoa	Succinyl-CoA	0.0001
Udcpdp	Bactoprenol	0.0001
Malcoa	Malonyl-CoA	0.00003
Products		
Adp	ADP	59.81
H	H(1+)	59.81
Pi	Orthophosphate	58.81
Ppi	Pyrophosphate	0.750

TABLE 3.3.1 (page 72): The biomass composition of the iAF1260 model. With the exception of the GAM requirement described below, one gram dry weight (GDW) of biomass is assumed to be comprised of the compounds listed in the left-most column in the stoichiometric amount given by the coefficients in the corresponding rows of the right-most column. When the FBA model predicts that these compounds are synthesized in the stoichiometric ratio given in this table over a period of one hour, the specific growth rate is predicted to be exactly $\mu = 1/hr$. *ATP fraction of the biomass composition has been split to indicate respectively, the amount needed to form new biomass and the amount needed to run the biomass reaction. The latter demand is termed Growth Associated Maintenance (GAM) energy. **stoichiometric coefficients rounding to zero at the third decimal place have been reported to the decimal place where the first non-zero entry occurs.

The iAF1260 biomass reaction consists of ninety-nine metabolites (Table 3.3.1). In view of the biomass reaction, an entry from either of the first two columns of this table is seen to be a valid substitute for a_i given a particular metabolite indexed by i. For that same metabolite, the stoichiometric coefficient s_i is the number located in the third column of the corresponding row. Note that in writing the table above, we have ignored any metabolite for which $s_i = 0.0$; these are the metabolites that are assumed to not be a part of the biomass composition. Including all metabolites, even those with coefficients of zero, will result in a table with m rows. Thus, to each row of the complete table is associated a metabolite with a corresponding row in the stoichiometric matrix **S**. In the final column of the augmented matrix $(\mathbf{S}|\mathbf{b}_{\mathbf{rxn}})$, the number that appears in a given row is the number appearing in the third column of the full table for the corresponding row multiplied by -1.0. The only exception to this are those metabolites that appear as products; their stoichiometric coefficients will appear in the augmented matrix with positive signs, as a positive flux across the biomass reaction will lead to their formation. The table indicates that four (non-biomass) metabolites – ADP, protons, orthophosphate, and pyrophosphate are formed via the biomass reaction. Their function is to mass-balance a large

ATP hydrolysis accompanying the flux across the biomass reaction of the iAF1260 model. The significance of this hydrolysis is discussed in section 4.2 where "maintenance energies" are defined. The metabolites present on the reactant side (left-hand-side; LHS) of the biomass reaction, not involved in the accompanying ATP hydrolysis, represent one's assumptions of the composition of a unit (GDW) of biomass. Such an assumption on biomass composition is made by a large majority of the FBA models but rarely is it asserted in an explicitly biochemical language as it has been done here (Varma and Palsson 1994, Varma and Palsson 1994, Edwards, Ibarra et al. 2001, Mahadevan, Edwards et al. 2002, Meadows, Karnik et al. 2010, Orth, Thiele et al. 2010, Hanly and Henson 2011, Hanly, Urello et al. 2012).

In sum, the biomass composition in FBA models is represented as a stoichiometric collection of simple biomolecules. The next section discusses the experimental characterization of the stoichiometric coefficients. In the common setting, this collection is placed to the LHS of a biomass reaction and a unit flux across this reaction equates with a unit specific growth rate (equation 3.2.18). Maximizing the value of the biomass objective is equivalent to maximizing the value of this flux; the flux of the LHS metabolites across the biomass reaction and in those proportions given by their stoichiometric coefficients (Equation 3.2.18). In this way, the metabolism of a given substrate is modelled as the maximization of flow through pathways starting from its acquisition by transporter(s) and ending at terminal products of well-known anabolic pathways. By optimizing the biomass objective, an FBA model formulated as 3.2.15, predicts μ to be the maximal possible cellular rate of synthesis of a static set of basic biomolecules.

FBA models are structured to bridge the gap between genotype data and growth phenotype. The optimized value of the biomass objective is a rate phenotype, and there is a clear advantage to modelling growth rates via FBA. Unlike with growth rate models which have been introduced in section 1.6, there is a potential with FBA methods, to provide mechanism solutions to the question of what determines metabolic rates in a particular genetic and environmental setting. Developing this insight into mechanisms is a critical step towards metabolic engineering progress. This is why phenomenological models such as those of section 1.7 are seen to be inadequate for an engineering setting. Phenomenological models make no statement of causality, and therefore, provide little insight with which to predict system responses to experimentally induced perturbations. On the other hand, the extent to which flux across a biomass reaction can realistically represent specific growth rate remains to be explored. For suppose that even if a unit of biomass can be approximated to be a cumulative sum of a handful of elementary biomolecules. There are currently no reliable empirical methods by which one can easily quantify the potential variation in composition over differential physiological states. Results of recent studies on this matter do in fact support the hypothesis that biomass composition can vary considerably across conditions (Godin, Delgado et al. 2010, Yamamotoya, Dose et al. 2012, Schmidt, Kochanowski et al. 2016). Therefore, the applicability of FBA methods are seen to be limited by the appropriateness of a static biomass approximation over various timescales, conditions, cell types, and cell states. Some (d)FBA applications have addressed this problem, for instance by the use of a growth-rate-dependent biomass composition (Pramanik and Keasling 1997, Meadows, Karnik et al. 2010). Yuan et al. notes that for many FBA models, the biomass composition is not measured, but extrapolated from literature values.

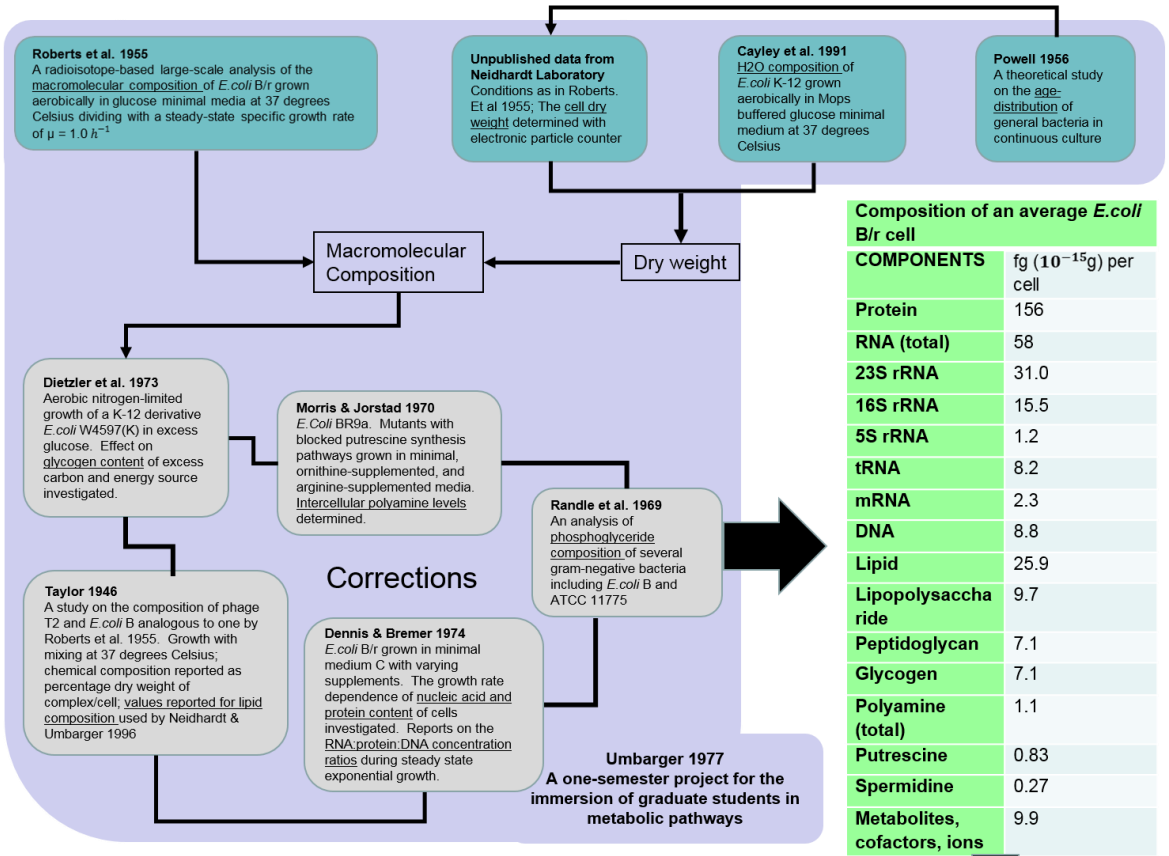
In their analysis of twenty-one publications using *Arabidopsis* metabolic models, they have found that only five employed biomass compositions that were directly measured by the modelers. Moreover, these authors have found that the measured biomass composition was significantly different across those five studies. To investigate the sensitivity of rate predictions to the assumed biomass composition, Yuan et al. studied the biomass composition of three published models by permuting the stoichiometric coefficients amongst the three models (Poolman, Miguet et al. 2009, de Oliveira Dal'Molin, Quek et al. 2010, Arnold and Nikoloski 2014, Yuan, Cheung et al. 2016). It was found, with few exceptions (e.g. AKGDH), that FBA predictions of central metabolic reactions as well as specific growth rate predictions were largely insensitive to the assumed composition. In a similar study, Dikicioglu et al investigated the sensitivity of Yeast FBA model predictions to the biomass composition. Whilst these authors report that specific growth rate predictions are not sensitive to composition under a particular growth condition (consistently with the findings of Yuan et al.), it was observed that pair-wise alterations of composition and growth condition do lead to discernible changes. In light of these results, we find that a closer investigation between the rate prediction and the biomass composition is warranted for *E.coli* metabolic models. In the remaining sections of this chapter, we consider two distinct ways by which FBA models may account for biomass composition. In the two chapters following, we investigate the mathematical relation between biomass carbon composition and rate predictions and its consequences under the assumption of a biomass objective.

3.4: The compositional characterization of *E.coli*:

In the previous two sections, it was seen that many FBA models require, in addition to a reconstructed metabolic network, an inventory of molecular components that are assumed to

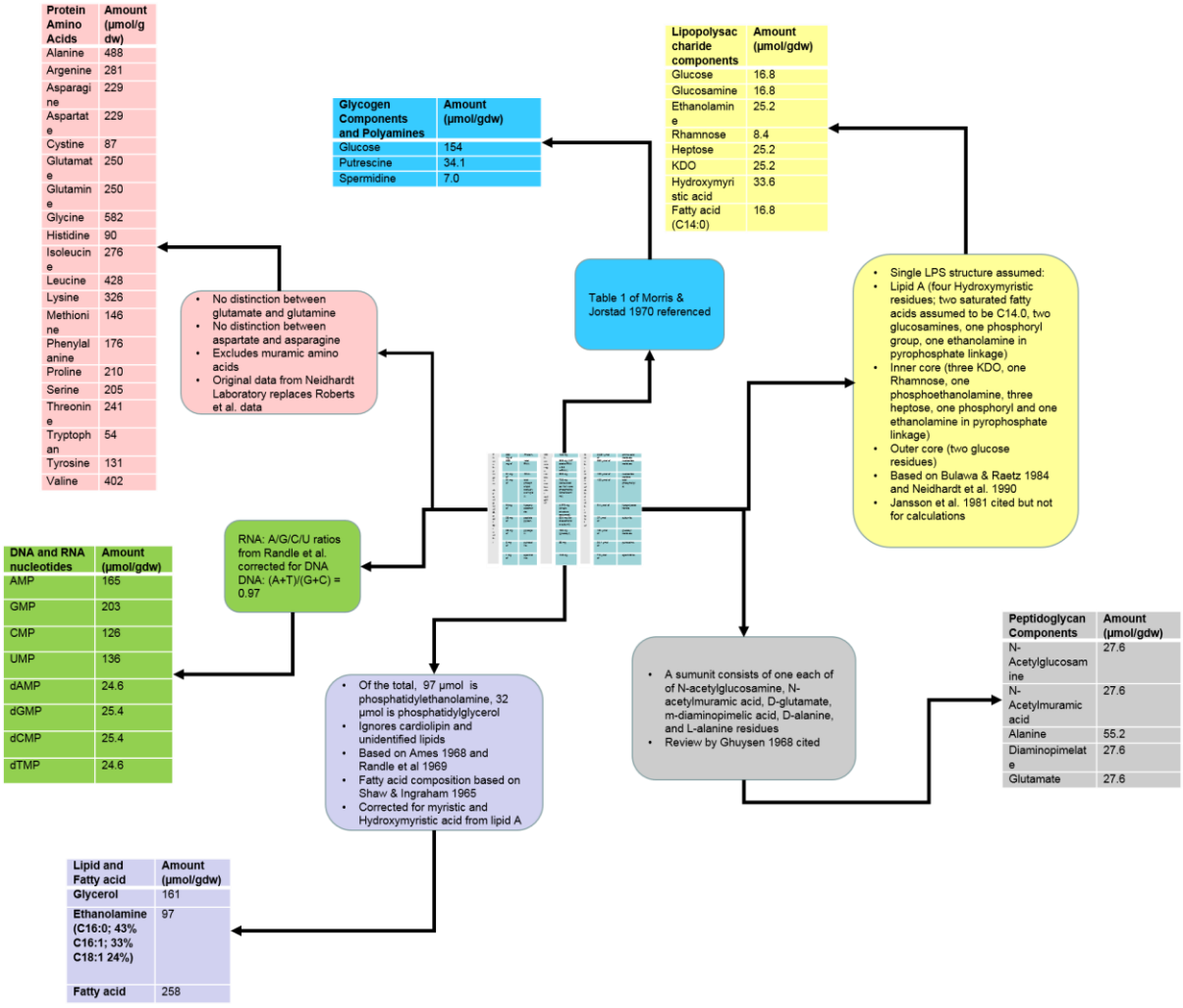
constitute a unit of biomass. If an FBA model is to optimize a biomass objective, then a stoichiometric coefficient will need to be specified for each member of the inventory so that a biomass reaction may be written. Note however, that the formulation of equation 3.2.13 does not require the flux across a biomass reaction to be the objective. Thus, to be more precise, we may say that FBA models that make a growth rate prediction require a stoichiometric molecular inventory and a biomass reaction. We have noted in the previous section that *E.coli* has one of the most comprehensively characterized metabolic reconstructions. It is also the case that *E.coli* has one of the most comprehensively characterized stoichiometric inventory of biomass metabolites. This situation is owed largely to the extensive characterization of the molecular composition of this bacteria that has been provided by Neidhardt (Neidhardt, Ingraham et al. 1990). In this section, we deconstruct the biomass objective of the iAF1260 model by reviewing the workflow that has gone into this derivation of *E.coli* composition.

The iAF1260 model assumes a molecular parts-list of a typical *E.coli* cell that can be traced to a table of data published by Neidhardt et al. in 1990 in their book *Physiology of the Bacterial Cell: a Molecular Approach* (Neidhardt, Ingraham et al. 1990). That same data also appears as two tables published in an article by Neidhardt and Umbarger in the first volume of *Escherichia coli and Salmonella* (Dempsey, Neidhardt et al. 1987). In the latter article, is a detailed description of how the numbers were obtained. These values are interpreted today to be masses of biomolecules that are stoichiometric with respect a gram dry weight (GDW) of an *E.coli* cell. The workflow which will now be summarized is shown schematically in figure 3.4.1.



Composition of an average *E.coli* B/r cell	
COMPONENTS	fg (10^{-15}g) per cell
Protein	156
RNA (total)	58
23S rRNA	31.0
16S rRNA	15.5
5S rRNA	1.2
tRNA	8.2
mRNA	2.3
DNA	8.8
Lipid	25.9
Lipopolysaccharide	9.7
Peptidoglycan	7.1
Glycogen	7.1
Polyamine (total)	1.1
Putrescine	0.83
Spermidine	0.27
Metabolites, cofactors, ions	9.9

There is per gram-dry-weight of *E.coli* B/r cells,	550 mg of	Protein.	**With an average residue molecular weight of**	108 mg,	**there is, per gram of dried cells, a total of**	5,081 μmol of	amino acid residues.
	205 mg of	total RNA.		325 mg (197 stable RNA + 8.3 mRNA),		630 μmol of	nucleotide residues.
	31 mg of	DNA.		309 mg,		100 μmol of	nucleotide resides.
	91 mg of	total phospholipid exclusive of lipid A.		705 mg (calculated as if all were phosphatidylethanolamine),		129 μmol of	total phospholipid.
	34mg of	lipopolysaccharide.		4,070 mg (single structure assumed),		8.4 μmol of	lipopolysaccharide.
	25 mg of	peptidoglycan.		904 mg (for disaccharide subunit),		27 μmol of	subunits.
	25 mg of	glycogen.		162 mg (glucosyl),		154 μmol of	glucosyl residues.
	3 mg of	putrescine.		88 mg,		34.1 μmol of	putrescine.
	1 mg of	spermidine.		145 mg,		7.0 μmol of	spermidine.



<u>Figure 3.4.1(continued on page 78)</u>: The flow of information described by Neidhardt & Umbarger from which the macromolecular composition of a gram dry weight of an *E.coli* B/r cell was determined. The area surrounded in purple in the top chart illustrates guidance from the 1977 article by Umbarger. The table to the right in green is a column in T1 taken directly from Neidhardt & Umbarger. The figure and text have been modified from Neidhardt & Umbarger. The full list of references from which the information was derived in Neidhardt & Umbarger is as follows: (Taylor 1946, Smith and Wyatt 1951, Roberts, Abelson et al. 1955, Powell 1956, Dunn and Smith 1958, Shaw and Ingraham 1965, Ames 1968, Ghuysen 1968, Randle, Albro et al. 1969, Morris and Jorstad 1970, Dietzler, Leckie et al. 1973, Dennis and Bremer 1974, Umbarger 1977, Jansson, Lindberg et al. 1981, Bulawa and Raetz 1984, Neidhardt, Ingraham et al. 1990, Cayley, Lewis et al. 1991) (Dempsey, Neidhardt et al. 1987).

The inventory of molecular components published by Neidhardt & Umbarger is of an average *E.coli* B/r strain grown aerobically in glucose minimal media at 37 degrees Celsius dividing with a steady-state specific growth rate of $\mu = 1.0\ h^{-1}$ (Dempsey, Neidhardt et al.

1987). The construction of the inventory has a strong bioinformatic component relying on published literature values of *E.coli* composition combined with basic dimensional analysis and unpublished measurements made in Neidhardt's own laboratory. Seventeen primary sources are cited in the reference to the Neidhardt & Umbarger article. Ten of this set of seventeen are explicitly stated and shown to have contributed to the construction of the first table (T1) reporting on the macromolecular composition of *E.coli* B/r. This number increases to eleven with the inclusion of the last column of T1 separating macromolecular composition by mass into composition by different kinds of molecules of a given type; this is information not typically used for FBA modelling (Dempsey, Neidhardt et al. 1987, Orth, Thiele et al. 2010). The second table (T2) reports on the stoichiometric composition of a GDW of *E.coli* B/r by residue. It has been constructed by combining the information of T1 (cellular mass composition by macromolecules) with known mean molecular weights and approximate per-capita copy numbers of residues. Some minimal assumptions have been included in the derivation of T2 from T1; the number of directly referenced sources is increased in the process by two for the calculations concerning respectively lipopolysaccharide and peptidoglycan content (Dempsey, Neidhardt et al. 1987).

Four references are particularly noteworthy. First, the information of T1 and T2 in the 1990 text itself by Neidhardt et al. is cited not just by Neidhardt & Umbarger, but also importantly by subsequent publications that concern the network reconstruction of *E.coli* (Dempsey, Neidhardt et al. 1987, Varma, Boesch et al. 1993, Varma and Palsson 1993, Varma and Palsson 1993, Varma and Palsson 1994, Pramanik and Keasling 1997, Edwards and Palsson 2000, Feist, Henry et al. 2007, Thiele and Palsson 2010). The second reference, which developed a framework for stoichiometric analysis of biochemical pathways and served as

guidance for the construction of T1 and T2, is a 1977 publication by Umbarger which is an article appearing in Biochemical Education (Umbarger 1977). Pedagogically oriented, this article outlines a syllabus for a fourteen week learning exercise, designed and implemented by Umbarger at Purdue University for first year graduate students in biochemistry, involving the whole-cell mass and energy stoichiometric characterization of *E.coli* metabolic pathways using literature information.

The currently accepted dry mass of an *E.coli* cell is on order of 10^{-13} g (Bionumbers ID: 103904) (Milo, Jorgensen et al. 2010). The exact number, 280 fg, reported by Neidhardt & Umbarger and widely used today, is an original measurement determined in the Neidhardt laboratory (Dempsey, Neidhardt et al. 1987). The amount per cell in grams of macromolecules, the key information in T1 needed for T2, is calculated from this measurement combined with percentage total dry weight values. Deriving this information required the assumption on the percent composition of water of cells. For this quantity, Neidhardt & Umbarger assumes a 'common textbook value' of 70%; however it is noted that this is a relatively accurate assumption as corroborated by the referenced work by Cayley et al. (Dempsey, Neidhardt et al. 1987, Cayley, Lewis et al. 1991).

The fourth and most important reference in the Neidhardt & Umbarger study is a work on *E.coli* whole-cell composition that was undertaken by Roberts et al. in the 1950s (Roberts, Abelson et al. 1955). Their measurements carried out using radioisotopes is cited to be the starting point for the derivation of all information published in T1 and T2 (Dempsey, Neidhardt et al. 1987). As mentioned already, the information on T1 of macromolecular composition is translatable to an inventory of residue composition provided the average residue molecular weight and per-capita residue copy number are known. The information provided by Roberts et

al. has been combined with T1 to generate T2. The primary contributions from Neidhardt & Umbarger are the replacement of amino acid values of Roberts with values measured in the Neidhardt Laboratory (under the same conditions) and few basic biochemical assumptions; one example to illustrate is the assumption that the (A+T)/(G+C) ratio is 0.97 for DNA (Dempsey, Neidhardt et al. 1987).

3.5: The biomass composition as a parameter and as a response

Combining a large-scale stoichiometric model with a biomass objective is a standard way of using FBA when implementing COBRA methods, so much so that the term "FBA" appears to have almost become synonymous with this style of application. This is however only one particular approach to FBA which is not universal in the literature. To make the distinction clear, we will find it useful to refer to genome-scale FBA models that employ a biomass objective as "Palsson-style" models (David Fell, Jon Pitchford; personal communication). More generally, we will refer to FBA models that predict growth rates using a biomass reaction as Palsson-style models. In making this distinction, we find that a strikingly different approach to FBA modelling is possible. This method, which we refer to as the "Fell-style" FBA deserves to be distinguished as an important alternative approach related inversely to Palsson-style metabolic modelling. To put the trend succinctly, a Fell-model starts with experimental flux data whereas a Palsson-model starts with genomic data; whereas a Fell-model starts with experiment, a Palsson-model starts with bioinformatics. Both the Fell and Palsson approaches eventually result in a network flow model with a structure (a stoichiometric matrix) and currents (flux distributions). However, in the Palsson approach, μ is a prediction, whereas in the Fell approach, μ is a fixed measured quantity. In the Palsson framework, μ is the predicted flux across a fixed biomass

reaction (equation 3.2.18). In the Fell framework, the biomass composition, rather than μ, is often the model prediction.

Section 3 of this chapter discussed how a biomass reaction may be used by FBA models to predict specific growth rates. The requirement for using a biomass reaction was seen to be the knowledge of the stoichiometric coefficients of the metabolites comprising a unit of biomass ($\mathbf{b_{rxn}}$). For a biomass composition that is highly detailed in description, as would naturally be seen to occur with a large-scale **S**, the coefficients often need to be extrapolated from bioinformatic data as was seen in the previous sections for the iAF1260 model. Thus, the FBA method that has been described in this chapter is characteristic of the Palsson-style approach. In contrast, the Fell-style alternative of biomass representation requires that the biomass metabolites be drawn out from the metabolic system individually. To see how this may be done, consider the following alternative way of writing the augmented equation 3.2.5

$$\begin{pmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{pmatrix} \begin{vmatrix} 1.0 & \cdots & 0.0 \\ \vdots & \ddots & \vdots \\ 0.0 & \cdots & 1.0 \end{vmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \\ v_{n+1} \\ \vdots \\ v_{n+m} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \textbf{(3.5.1)}$$

Here the stoichiometric matrix has been augmented with the $m \times m$ identity matrix, and the velocity vector has been augmented with m new variables, each representing the rate of exchange of a metabolite between the system and its surroundings. If the rate of exchange of a biomass metabolite is given by v_j, $n + 1 \leq j \leq n + m$, then with a measured growth rate μ^{obs}, the stoichiometric coefficient for this metabolite may now predicted to be

$$s_{jb} = \frac{v_j}{\mu^{obs}} \qquad \textbf{(3.5.2)}$$

Comparing with the corresponding equation for Palsson-style optimization

$$\mu = \langle \mathbf{v}, \mathbf{b} \rangle \qquad \textbf{(3.5.3)}$$

It can be seen that the parameter and the prediction have swapped between the two formulations (recall that **b** is fixed from the composition $\mathbf{b_{rxn}}$).

It is also possible in the Fell approach that the biomass composition has also been measured directly. Thus if, further to the above setup, the stoichiometric coefficients are also known, then it would be possible to individually constrain the fluxes corresponding to the biomass composition using strict bounds as

$$v_j \in \left[s_{jb}^{obs} \times \mu^{obs}, s_{jb}^{obs} \times \mu^{obs}\right] \qquad \textbf{(3.5.4)}$$

reducing the FBA solution space to a more experimentally consistent feasible region. For instance Cheung et al. describes a very coarse (*Arabidopsis*) biomass composition consisting of only thirty one metabolites (Cheung, Williams et al. 2013). The coarse resolution with which the biomass composition was described however, had made the direct experimental measurement of each stoichiometric coefficient possible. Similarly, was the case for the metabolic system description. In their study, all carbon exchanges including the substrate influx rate (glucose consumption) were also directly measured leaving CO_2 efflux rate as the only free boundary condition.

We find that this example is very typical of the Fell-style approach to FBA modelling. While Palsson-style FBA models increase in scale with genomic information, Fell-style models often do so in consideration of available flux measurements. As a result, although the dimensions of Fell-FBA models are sometimes modest relative to Palsson-style models, Fell models are superior in their level of experimental characterization. It is of course possible to construct Palsson-style models on the small scale and likewise, Fell-style models on the genome-scale. However, in all cases, it is seen that the Fell approach allows for a flexibility in which

biomass composition may be treated as response of a metabolic system to a hypothesized (or an actual) cellular objective. This is in contrast to the Palsson approach in which composition is a parameter rather than a response. This aspect of growth phenotypes may be safely ignored in some applications of FBA. As we have seen in section 3.1, FBA methods may be employed towards investigating strictly topological phenotypes in isolation from rates. One notable advantage of starting with the genome and not the experiment is revealed by the examples mentioned in that section. Genome-scale models are superior to small-scale models in topological discovery because they are able to explore a wider null-space.

The Fell approach does not generally allow for the prediction of μ. For this reason, a direct application of Fell optimization is not appropriate for extensions such as dFBA where μ, by definition is required to be a prediction. However, the properties of Palsson models that lead to predictions of μ (for instance through the LP defined in 3.2.1) can be better understood by leveraging on the Fell approach.

3.6: Discussion

Extensions of FBA such as dFBA require an FBA formulation where growth rates are predictions. We have therefore, in this chapter, investigated the biomass composition of the commonly used iAF1260 model of *E.coli* metabolism. We first provided a mathematical definition of the biomass objective function and discussed the biological rationale behind hypothesizing this objective. Next, we have deconstructing the coefficients that appear in the iAF1260 biomass objective by investigating the literature sources from which its assumed values have been derived. The FBA approach of using this objective function, described at the level of detail as is done by the iAF1260 model, with a genome-scale metabolic reconstruction is typical.

However, we recognize that alternatives to this approach are possible, and suggest that their consideration may be useful to dFBA modelling.

The number of ways to apply FBA methods is limitless. Each application is distinguishable by such factors as the size of the model and the objective function that is chosen. For this reason, there is no way to decisively decide if an FBA model is strictly of one type or of another; the notion that every FBA model can be put on a spectrum between two types would be an oversimplification. In extending FBA to methods such as dFBA however, a useful distinction is made based on how growth rates are treated. By treating μ as a model response, "Palsson-style" FBA methods allow for the prediction of growth rates. However, in doing so, this method fixes the carbon demands of the growth process to stoichiometric values that may not readily extrapolate beyond the settings in which they have been measured. Thus, the accuracy of rate predictions depends strongly on the extent to which a fixed composition extrapolates across various growth conditions. The opposite approach to the Palsson-style FBA method is to treat growth composition, rather than growth rate as a response. In this approach to FBA, which we have chosen to call the "Fell-style" approach, the constituents of the biomass are drawn out individually as exchange reactions. As a result, the constraints on the rate of biosynthesis may be applied for each metabolite individually; similarly, the biomass composition may be treated as a model prediction by investigating individually, the flux across exchange reactions. Often, the Fell method is employed by using experimentally measured constraints on the exchange reactions corresponding to biomass constituents, thus treating the product of stoichiometry and growth rate as an experimentally measured system response. Whilst both the Palsson and Fell approaches have been employed for both genome-scale and small-scale metabolic models, requiring that the individual components of biomass composition be measured leads to models

that are thought to be smaller in scale but experimentally better characterized for making rate predictions. On the other hand, genome scale models are better suited for topological phenotype analyses. These examples reinforce the idea that FBA methods and FBA models, though unified in their mathematical formulation as a constrained-based optimization problem, should be quite varied in the way they are executed towards the investigation of a particular biological problem. In this light, it is surprising to see that in practice, there is not a greater level of diversity amongst simple FBA models with respect to scale and to cellular objectives (although the number of extensions such as dFBA are many). At the very least, we expect that FBA methods should be tailored in their formulation to the growth phenotype under investigation, and that the latter must be properly defined. There is no indication at the present that genome scale metabolic models using the biomass objective is always an appropriate choice; the success of Fell-style applications suggest to us that they may not be. Is the Palsson-style formulation of FBA therefore the most prudent choice for dFBA and are there other possible formulations that allow for growth-rate predictions that have not yet been considered? We explore these questions by first investigating the consequences to growth rate predictions of using a biomass objective in the Palsson FBA framework.

CHAPTER 4: The relation between growth rate and biomass yield in flux-balanced systems

Abstract:

The prediction of flux distributions using FBA models requires that a cellular objective be first hypothesized. Yet it is rarely clear in applications as to what choice of objective to use and FBA models commonly adopt the most parsimonious choice of growth rate maximization. In this chapter, we investigate the stoichiometric consequences of using this objective function. Using conservation laws, we derive a mathematical relation between growth rate and a related quantity, growth yield. Though rate and yield have different biological meanings, we show that under certain growth conditions, the two quantities are identical as an optimality criterion. We conclude that the predictions of FBA methods that hypothesize growth rate maximization may also be rationalized by the hypothesis of growth yield maximization. We therefore advocate that FBA outputs be reported in the framework of optimizing yields if, compared to optimizing rates, this principle leads to simpler explanations of optimal phenotype predictions.

4.1 Introduction

It is a well-established fact that carbon is a dual currency in metabolic systems. Carbon is a synthetic currency to growing cells, since without an external carbon source, basic biomolecules such as DNA and protein cannot be synthesized, and the formation of new biomass will cease. Biosynthesis is a highly endergonic (energy-consuming) process however which requires not just a source of raw materials such as carbon but also a source of energy to drive its reactions uphill. This source of energy is typically stored in the form of the energetic cofactors ATP and NAD(P)H. To continuously replenish these internal energy pools, heterotrophic systems such as *E.coli* cells require an environmental source of highly reduced carbons. Thus, carbon plays a second role as an energetic currency in metabolism. In one scenario, a substrate molecule such as glucose may be oxidized to by-products of energy biogenesis (such as CO_2, acetate, and ethanol), which are exported to the environment while the energy liberated in the process is stored as a disequilibrium across reactions that interconvert between pairs of cofactors.

In a second scenario, that same substrate molecule may be converted to biosynthetic precursors and subsequently incorporated into biomolecules. For this single hypothetical molecule, these two scenarios are clearly mutually exclusive; the carbon may be used as a source or mass or as a source of energy but not both.

It was seen in section 3.2 how the Palsson approach to FBA uses a biomass reaction. Often, the flux across the biomass reaction, interpreted to be the specific growth rate (equation 3.2.14), is also used as the objective (formulation 3.2.15). All FBA models require the hypothesis of a cellular objective to predict flux distributions. Biologically speaking however, objectives are a highly abstracted representation of phenotype with a wide and open-ended range of possibilities for a choice. As a result of this, it is rarely clear to modelers whether the chosen objective function is biologically relevant. The practice of adopting the biomass objective as a default in the presence of these uncertainties make quite a bit of sense. For many systems, growth rate optimization can be seen empirically to be a parsimonious optimality criterion given our knowledge of evolution. For internally flux-balanced metabolic systems, optimizing growth rate require flux distributions that direct as much of the substrate available in the environment towards the biomolecular constituents of biomass. However, in order to pay for the thermodynamic costs incurred from operating the chemical pathways necessary for this direction, a fraction of the substrate must be diverted away from biosynthesis towards oxidized byproducts of energy biogenesis. Internal flux balance by definition implies the carbon flux across a metabolic boundary sums to zero. As a result, the rate of energy production and the rate of biomass production for steady-state systems form a pair of competing quantities that are constrained to obey a conservation relation; in order to balance a fixed influx of substrate carbon,

the carbon flow to biomass must decrease if carbon flow to byproducts of energy biogenesis are to increase and vice versa.

Up until this point, we have been concerned exclusively with FBA predictions of rates of reactions. However, our consideration of the competition between steady-state flux of carbon to biomass and to energetic by-products suggests that a measure of growth efficiency will also be needed. Consider the example of dFBA. In order for a dFBA model to correctly match experimental time-series of substrate consumption, biomass production, and by-product production, the underlying FBA model must be parametrized so that ratios of rates between consumption and production are consistent with observations. In the microbial literature, one such measure is the *biomass yield on substrate* which, for growing populations, is defined to be the ratio between the increase in population size and the amount of substrate consumed over a given time interval. Letting these values be denoted by Δx and ΔS respectively (as in equations 1.7.1 and 2.3.1), the biomass yield on substrate S, denoted here by Y_S, is the ratio

$$Y_S = \frac{\Delta x}{\Delta S} \qquad \textbf{(4.1.1)}$$

Although we will not use the biomass yields exactly as defined in 4.1.1 (see section 4.4), it appears that FBA predictions of growth rates and growth yields should be closely related. In this chapter, we demonstrate that this is in fact the case, by deriving an explicit equation between the two quantities involving FBA parameters. We will then proceed in the next chapter to investigate the consequences of this equation to the predictions of Palsson-style FBA models.

We begin in the following section, by defining cellular "maintenance energies" and by describing how FBA models may account for these quantities. From this discussion, we proceed to make explicit, a linear relation between specific growth rate and ATP hydrolysis rate that is implied by FBA models that account for maintenance rates of ATP dissipation. Throughout the

chapter, we assume an aerobically modeled system having a carbon source from the environment, a fixed biomass composition, and maximizing biosynthetic flux. Thus, our results apply to Palsson-type FBA models of substrate-limited growth.

4.2 The linear relation between growth rate and ATP hydrolysis rate

From the consideration of carbon-balancing across the boundary of a steady-state metabolic system, the maximal carbon flow to biomass for a given substrate consumption rate is seen to occur when all incoming carbon flow is directed towards biomass constituents. Therefore, for Palsson-FBA models of steady-state growth, the maximal value that a biomass objective can achieve, in the total absence of constraints, is the influx rate of substrate scaled by the carbon content ratio between the substrate and the biomass. To illustrate, consider *E.coli* cells modelled with the biomass objective to grow on the pentose (5-carbon) sugar xylose, and the uptake rate is fixed at the value of v_{xyl}. Denoting by C_{bio}, the carbon content of a unit of biomass, the maximal possible value for the objective is constrained as a result of carbon-conservation by the upper bound

$$\mu^{Max} \leq v_{xyl} \frac{5}{C_{bio}} \qquad \textbf{(4.2.1)}$$

Because of the need for metabolic systems to allocate a portion of the acquired substrate towards energy production however, it should be apparent that FBA predictions of steady-state μ^{Max} will typically fall well below the upper bound in response to stoichiometric constraints that divert carbon flow away from biomass precursors. We show in this section that the actual optimal value for μ can be expressed by a linear equation in the predicted ATP hydrolysis rate involving two parameters with respect to which FBA model predictions are highly sensitive.

Towards this end, we first consider why biologically it might make sense for ATP synthesis rate to exhibit a linear relation with μ. Suppose a quantity of energy is expended in the form of ATP hydrolyses over some time interval and results in population growth. This energetic demand may be partitioned categorically into energy that has been applied towards the production of *new* biomass and energy that has been applied towards maintenance of *existing* biomass. In the consideration of the latter energetic demand, the quantity Non-Growth Associated Maintenance (NGAM) rate of ATP hydrolysis (we use the unit $\mathrm{mmol\ GDW^{-1}h^{-1}}$) is defined to mean the energy that is required to maintain homeostasis for a unit biomass over a unit time interval. NGAM represents the collective energy needed by existing biomass for such housekeeping processes as maintaining a constant membrane potential and flagellar rotation for movement (Marr, Nilson et al. 1963, Pirt 1965, Pirt 1982, Van Bodegom 2007, Wang and Post 2012). To account for the energy expenditure not included in NGAM, we define the quantity Growth Associated energy (GA) (we use the unit $\mathrm{mmol\ GDW^{-1}}$) to mean the stoichiometric amount of ATP hydrolyses required per unit increase in specific growth rate. Note that unlike NGAM, GA is a gradient between the rates of growth and ATP hydrolysis as can also be inferred from its unit. GA itself may be partitioned categorically into two terms. On the one hand are biological processes directly leading to the synthesis of new biomass such as the conversion of substrate to metabolic precursors. We will refer to this component of GA as Growth Associated Anabolic energy (GAA). Processes associated with replication that do not involve the anabolic production of new biomass include as examples transcription, translation, and cell division; these processes collectively contribute to the second component of GA. This quantity, representing the ATP demand from *non-synthetic* processes, is termed Growth Associated Maintenance energy (GAM) in analogy to NGAM. This is the rationale behind embedding an ATP hydrolysis

in the biomass reaction, which was seen in table 1 of sections 3.3 (Thiele and Palsson 2010). Let the total rate of ATP hydrolysis by a steady-state flux distribution across a metabolic network be denoted by v_{ATP}. Our discussion suggests that v_{ATP} and μ should be linearly related as

$$v_{ATP} = NGAM + \mu GA \quad \textbf{(4.2.2)}$$

and equivalently as

$$v_{ATP} = NGAM + \mu(GAA + GAM) \quad \textbf{(4.2.3)}$$

Writing v_{ATP} in this way relies simply on categorically subdividing the net rate of ATP hydrolysis by cellular functions with respect to replication. Thus, there is nothing mathematically inconsistent about writing the above equations; we have defined the parameters in such a way that these equations are mathematically and biologically self-contained. Extrapolating these relations over a range of μ with a constant choice of the parameter values however requires further experimental justification. Unfortunately, the available evidence that a simple linear model may be used to describe the general relation between growth rate and ATP hydrolysis rate, are limited to chemostat growth. Performed most commonly with *E.coli*, a plot of μ against an indicator for v_{ATP} (see below and the next section for what is meant here by "indicator"), has been shown generally to conform well to a simple linear model (Andersen and von Meyenburg 1980, Varma and Palsson 1994, Edwards, Ibarra et al. 2001). The growth conditions involved in chemostat settings are, in many ways, highly unrepresentative of more realistic environmental conditions to which cells are exposed on a daily basis or during the course of batch growth. Thus, whether a simple linear model is sufficient for describing the general relation between ATP hydrolysis rate and specific growth rate for metabolic systems in a more biologically realistic setting is most certainly open to discussion.

Before considering the implications of assuming a (fixed) linear relation between growth rate and ATP hydrolysis rate, we describe how maintenance energies are represented in FBA models. In the Palsson setting, NGAM is modeled as the flux through one reaction, ATP hydrolysis of Maintenance (ATPM), which is constrained to hydrolyze ATP to ADP and orthophosphate at a constant rate by setting both the lower and upper bounds of this reaction to a fixed positive value NGAM. Thus, for an unconstrained growth-rate,

$$v_{ATPM} \in [NGAM, NGAM] \quad \textbf{(4.2.4)}$$

$$v_{growth} \in [0.0, 10^6] \quad \textbf{(4.2.5)}$$

Recall that we may choose arbitrarily, the value $10^6 \text{mmol GDW}^{-1}\text{h}^{-1}$ to represent the computational bound of a flux for which a biological bound has not been provided (section 2.8). In Fell-type models, a reaction such as the one termed 'generic ATP-ase' in Poolman et al. serves an analogous role (Poolman, Miguet et al. 2009). Importantly however, the flux through the generic ATP-ase reaction is often treated as a variable to be predicted rather than a fixed parameter. Thus, with an experimentally observed growth rate μ^{obs} and stoichiometric coefficients s_{jb}^{obs} (as in 3.5.4)

$$v_{ATPase} \in [0.0, 10^6] \quad \textbf{(4.2.6)}$$

$$v_j \in \left[s_{jb}^{obs} \times \mu^{obs}, s_{jb}^{obs} \times \mu^{obs}\right] \quad \textbf{(4.2.7)}$$

In the *E.coli* series of Palsson, ATPM is constrained strictly to 7.6, 8.39, and 3.15 $\text{mmol GDW}^{-1}\text{h}^{-1}$ in iJR904, iAF1260, and iJO1366 respectively. The quantities were chosen from linearly fitting a series of FBA solutions against oxygen or glucose consumption rates (Reed, Vo et al. 2003, Feist, Henry et al. 2007, Thiele and Palsson 2010, Orth, Conrad et al. 2011). GAM is modelled into the stoichiometric coefficients of ATP hydrolysis by the biomass reaction. The values for GAM are 45.56, 59.81, and 53.95 mmol GDW^{-1} in the iJR904,

iAF1260, and iJO1366 models respectively. The ATP hydrolysis stoichiometry of the biomass objective is lowest in an earlier FBA model by Edwards and Palsson at 23.2 mmol GDW^{-1} and greatest in iAF1260 (Edwards, Ibarra et al. 2001, Feist, Henry et al. 2007). It can be concluded from observing how the parameters have been varied in the past that there is a wide range of numeric possibilities for maintenance energy demands. This is an observation that may be corroborated experimentally. In one attempt to characterize NGAM, *E.coli* cells were grown aerobically in chemostat under various growth limiting conditions (glycerol, oxygen, ammonium, and sulphate limitation). It was found there that the experimentally calculated value for NGAM varied between 2.2 $\text{mmol GDW}^{-1}\text{h}^{-1}$ under oxygen-limitation, and 30.8 $\text{mmol GDW}^{-1}\text{h}^{-1}$ under sulphate limitation, a wide range within which the computationally required interval (3.15 to 8.39 $\text{mmol GDW}^{-1}\text{h}^{-1}$) fits comfortably. The corresponding values for glycerol and ammonium limitation were found respectively to be 2.3 and 16.8 $\text{mmol GDW}^{-1}\text{h}^{-1}$ (Farmer and Jones 1976).

Whilst there is nothing inherent to the stoichiometric formulation of FBA models that allow us to predict what value the maintenance parameter should be assigned *a priori*, FBA outputs are known to be quite sensitive to its maintenance parameters. In light of this fact, it is reasonable to surmise that flux-balance methods would be improved with a more careful and explicit treatment of maintenance energies, and more generally of energetic constraints to growth rate and growth yield predictions. To consider one example from the dFBA literature, Meadows et al. have found that writing a dynamic expression for NGAM as a function of externally present compounds and inorganic ions lead to better fits of their dFBA growth curves to their experimental data (Meadows, Karnik et al. 2010). Specifically, they have used the dynamic NGAM formula (in the unit $\text{mmol GDW}^{-1}\text{h}^{-1}$)

$$\mathrm{NGAM} = 0.1[\mathrm{acetate}] + 0.175[\mathrm{Na}] + 40.0\left(10.0^{|\mathrm{pH}-7.25|} - 1.0\right) \qquad \textbf{(4.2.8)}$$

with a GAM value of 40.0 mmol GDW^{-1}.

The results of Meadows et al. demonstrate the potential importance of treating maintenance parameters dynamically in dFBA. By defining NGAM to account for the changing effects on membrane potentials of an increasingly acidic environment, these authors have shown that not only can the quantification of maintenance parameters be made upon an explicit mechanistic foundation, but also that doing so may result in a flux-balance model that more accurately describes real biological data. Having stated this however, it should also be recognized that a hypothesis such as the one represented by equation 4.2.8 is rather complex and cannot be unequivocally justified. In consideration of the fact that an FBA model already makes a highly complex hypothesis when choosing a specific cellular objective (section 2.9), a challenge to FBA modelling is recognized. Whilst a set of hypotheses may be supported when a model is able to reproduce experimental observations, the level of abstraction that is inherent in the hypotheses themselves render their verification difficult.

In summary, published experimental observations suggest that the specific growth rates are directly related to ATP hydrolysis rate in steady state *E.coli* growth. The relation of v_{ATP} to μ is a linear relation so that it is determined by exactly two parameters: a slope and an intercept. The biological significance of these two parameters can be explained by the phenomenological theory of maintenance energies as given by equations 4.2.2 and 4.2.3. In FBA models, mechanism quantities are introduced into these equations in two ways. Firstly, by introducing a stoichiometric network, the slope term – GA – is split into two terms (GA = GAA + GAM), where one term – GAA – is quantified through fluxes through biological reactions with known ATP demands. Secondly, by defining μ as metabolite flux to biomass and defining the

composition of the said biomass, a correspondence is asserted between fluxes through a metabolic network, and the resulting growth rate.

4.3: The linear relation between growth rate and unbound oxygen consumption rate

We discussed in the previous section that FBA predictions of *E.coli* growth rate are closely related to predictions of ATP hydrolysis rate through maintenance energy parameters. Experimentally, the net rate of ATP hydrolysis is not a quantity which can be measured directly. Suppose however that the stoichiometric relation between ATP hydrolysis and the consumption rate (or production rate) of an externally observable metabolite is known approximately. Then it would follow that the rate of steady-state ATP hydrolysis can be deduced by the known stoichiometry and flux measurements at the cellular boundary. This observation provides a strategy for the experimental determination of ATP hydrolysis rates using external metabolites as proxies. Fluxes of CO_2 and O_2 for *E.coli* grown in chemostat have in fact both been used for this purpose and have been shown to be linear in μ under chemostat-controlled growth (Farmer and Jones 1976, Andersen and von Meyenburg 1980, Varma and Palsson 1994). We consider in this section, the rate of oxygen consumption (v_{O2}) by respiring *E.coli*.

The biological rationale behind using oxygen consumption rate is summarized as follows. The function of molecular oxygen in supporting aerobic respiratory growth is solely to serve as an electron acceptor; the incoming oxygen atoms of O_2 molecules will leave as oxygen atoms in outgoing H_2O molecules and as part of no other compound. As a result, there is a simple stoichiometric relation between the rate of conversion between O_2 and H_2O, and the rate of ATP synthesis. Provided with such a stoichiometry between O_2 and ATP, v_{O2} will be an appropriate substitute for v_{ATP}. For the time being, we will assume that this stoichiometry is known, and

shall denote it by the Greek letter δ. More precisely, we define δ by writing the stoichiometric equation

$$v_{ATP} \cong \delta v_{O2} \qquad \textbf{(4.3.1)}$$

Substituting the relation 4.2.2 into equation 4.3.1 gives the stoichiometrically equivalent flux balance equation

$$v_{O2} = \frac{NGAM}{\delta} + \frac{GA}{\delta}\mu \qquad \textbf{(4.3.2)}$$

which, unlike equation 4.2.2, relates growth rate linearly to the observable rate of metabolite flux across the cell boundary. In the interest of attempting to quantify maintenance parameters, this has been recognized to be an important property of steady-state growth. For example, the now commonly accepted metabolic reconstruction protocol by Thiel & Palsson suggest maintenance parameters to be quantified by linearly regressing growth rate in chemostat against a measure of ATP hydrolysis (protocol steps 32 -34) (Thiele and Palsson 2010).

As will be developed further in chapter 5, oxygen consumption rate is a particularly good proxy for ATP hydrolysis rate in *E.coli* models because experimental bounds on the former have been relatively well documented. However, we find it also important to consider the fact that since any pair of fluxes will exhibit a linear correspondence to each other at a metabolic steady-state, the fact that an equation of the form 4.3.2 should relate consumption rate to production rate is not a result that is unique to oxygen consumption and biomass production. Flux to the biomass in particular will be linear with respect to any given flux at the cellular boundary during steady-state, including substrate consumption flux. This fact together with equation 4.3.2 is used in the next section to derive the main result of this chapter, a relation between FBA predictions of biomass yields and of the rate of biomass production.

4.4: The relation between FBA predictions of growth rate and of growth yield

In the introductory section, we have provided explicit reasoning as to why the maximal rate of biomass production relative to a given rate of substrate consumption must diminish in response to increasing ATP demand for a flux-balanced metabolic system with no internal source of carbon. This relation arises from the mutually exclusive roles of biosynthesis and energy production that is served by any unit of substrate carbon. Biomass yield is the ratio by mass of biomass production to substrate consumption (equation 4.1.1). In this section, we complete the derivation of the equation relating steady-state predictions of biomass yield to predictions of growth rate. In order to generalize our result across substrates that differ in unit carbon content, we will first normalize the ratio in equation 4.1.1. We thus introduce a normalized measure of growth efficiency, Yield on Substrate Carbon (YSC), which we define as follows. Let C_{sub} be the carbon content of a substrate. If specific growth rate and substrate consumption rates assume values $\mu\ h^{-1}$ and v_{sub} mmol $GDW^{-1}h^{-1}$ respectively, YSC is defined to be the following ratio

$$YSC = \frac{\mu}{C_{sub}v_{sub}} \qquad \textbf{(4.4.1)}$$

This is simply the ratio of growth rate and the carbon influx from substrate. Energetic diversion of substrate carbon flux away from biomass results in the production of organic by-products and carbon dioxide. To describe such effects, several new quantities are needed for the purposes of carbon book-keeping (Table 4.4.1).

Notation	Quantity	Unit
v_{Csub}	Carbon uptake flux	$\text{mmol GDW}^{-1}\text{h}^{-1}$
v_{CO2}	CO_2secretion flux	$\text{mmol GDW}^{-1}\text{h}^{-1}$
v_{Cbio}	Carbon flux across biomass reaction	$\text{mmol GDW}^{-1}\text{h}^{-1}$
v_{Cbp}	Carbon secretion as by-product other than CO_2	$\text{mmol GDW}^{-1}\text{h}^{-1}$
C_{bio}	Total carbon content of (GDW) biomass	mmol GDW^{-1}
RQ	v_{CO2}/v_{O2}	Unit-less
BQ	v_{Cbp}/v_{Csub}	Unit-less

TABLE 4.4.1: New notations used to facilitate the quantitative description of carbon flows and ratios of carbon flows across general metabolic boundaries.

With these notations, we start by applying the steady-state assumption at the metabolic boundary which is usually taken to be the outer membrane for *E.coli* FBA models. Because the influx and efflux rates across the boundary must balance exactly for each element, the flux of carbon is constrained to obey the following flux-balance equation:

$$v_{Csub} = v_{CO2} + v_{Cbio} + v_{Cbp} \qquad \textbf{(4.4.2)}$$

We assume that the total carbon composition of a unit of biomass is fixed to be C_{bio}, and are interested in relating an FBA predicted rate (μ; the flux to biomass) to FBA predicted ratios of rates (YSC; a measure of biomass yield). Towards this end, we make use of the rate ratios RQ and BQ to derive a linear relation between the variables $\frac{(1-BQ)}{YSC}$ and $\frac{1}{\mu}$. From equation 4.4.2, we have

$$v_{Csub} = RQ \times v_{O2} + \mu \times C_{bio} + BQ \times v_{Csub} \qquad \textbf{(4.4.3)}$$

Rearranging this expression gives

$$(1 - BQ)v_{Csub} = \mu \times C_{bio} + RQ \times v_{O2} \qquad \textbf{(4.4.4)}$$

Or equivalently,

$$\frac{(1-\mathrm{BQ})}{\mathrm{YSC}} = \mathrm{C_{bio}} + \frac{\mathrm{RQ} \times \mathrm{v_{O2}}}{\mu} \qquad \textbf{(4.4.5)}$$

If RQ is assumed to be approximately 1.0, as is the case for carbohydrates, the previous equation simplifies to

$$\mathrm{YSC} = \left[\frac{\mu\left({}^{1}/_{\mathrm{C_{bio}}}\right)}{\mu + \left({}^{\mathrm{v_{O2}}}/_{\mathrm{C_{bio}}}\right)}\right](1-\mathrm{BQ}) \qquad \textbf{(4.4.6)}$$

(Andersen and von Meyenburg 1980).

Equation 4.4.6 is the desired result. In this equation, the efficiency of growth – YSC – is seen to diminish by a factor that decreases from 1.0 to 0.0 as more carbon influx is diverted to by-product; that is as BQ ranges from 0.0 to 1.0. Equation 4.4.6 can be combined with equation 4.2.2 to reveal a closed relation between YSC and μ through the FBA ATP hydrolysis parameters.

$$\mathrm{YSC} = \left[\frac{\mu\left(\frac{\delta}{\delta \mathrm{C_{bio}} + \mathrm{GA}}\right)}{\mu + \left(\frac{\mathrm{NGAM}}{\delta \mathrm{C_{bio}} + \mathrm{GA}}\right)}\right](1-\mathrm{BQ}) \qquad \textbf{(4.4.7)}$$

Neither an assumption of an optimality criterion nor of a stoichiometric metabolic model was needed in the derivation of the above equations. Rather, the key requirement for both equations is that the influx and the efflux of carbon balance at the cellular boundary; thus, equation 4.4.6 is a statement of how growth rate and growth yields must relate for a metabolic system with no internal sources of carbon. Equation 4.4.7 then follows if the rate of oxygen consumption is linear in the rate of ATP hydrolysis. The experimental evidence for this as has been previously stated in section 4.2 is that chemostat plots of oxygen consumption and growth rate can be shown to be linear. When equations 4.4.6 and 4.4.7 are applied with a stoichiometric metabolic model, exact values for the parameters that appear may be derived. Inspecting this

relation in turn reveals where FBA predictions of growth yields and growth rates come from. By separating the two components of GA as in equation 4.2.3, it can be seen that the parameter GAM makes an individual contribution to the slope of v_{ATP} with respect to μ. The GAA term in turn arises from ATP flux summation over metabolic pathways. Thus, this second term may be viewed as the mechanistic component of the gradient, determined stoichiometrically by a mode of substrate acquisition combined with biomass formation.

4.5: Discussion:

In this chapter, we have used stoichiometric balancing laws to derive an equation relating FBA predictions of yield and rates involving FBA parameters that quantify metabolic energy demands. Equations 4.4.6 and 4.4.7 have two important implications for Palsson-style FBA methods that optimize the flux across a biomass reaction (equation 3.2.18). We are first left to question whether Palsson-style methods optimize for rates as claimed, or instead for yields; we are second left to question whether the optimization of either is a valid cellular objective, since objective functions are only hypotheses one uses to derive plausible flux distributions (section 2.9). Schuster, Pfeiffer and Fell consider both these issues and point out that the terms "growth rate" and "growth yield" are often confused in papers on (Palsson-style) FBA; they argue that the latter is often optimized when the optimization of the former is claimed. Further, the authors go on, citing numerous biological examples, to show that yield maximization is not in general a valid cellular objective (Schuster, Pfeiffer et al. 2008). Equation 4.4.6 is entirely consistent with these conclusions. In fact the relation is a special case of representing yields as a ratio of linear combination of rates; this form, given in equation 2 of Schuster, Pfeiffer, and Fell, may be recognized by using equation 4.4.6 with equations 3.2.13 and 3.2.14 (Schuster, Pfeiffer et al.

2008). However, we find that equation 4.4.6 makes a slightly stronger assertion than claimed by these authors, which is that for internally closed metabolic systems in steady-state, rate maximization and yield maximization are the same objective to begin with when CO_2 is the only byproduct that is evolved (BQ = 0.0); a flux distribution that maximizes one will necessarily maximize the other because of 4.4.6. Founded only upon conservation laws, this is an assertion that is applicable to all fluxed-balanced biological systems and not just to FBA models. Therefore it does not make sense in any FBA setting to attempt to distinguish between the two optimizations when BQ is zero. The issue becomes important however when the maximization of rate and the maximization of yield become conflicting cellular objectives. Thus, we are left to investigate how the by-product quotient BQ is quantified under the Palsson FBA formulation. This is the subject of the next chapter.

Chapter 5: Consequences of rate maximization by steady-state yield maximization

Abstract:

Flux Balance Analysis (FBA) represents a family of metabolic modelling procedures by which systems biological principles are used in conjunction with bioinformatics to interrogate and analyze metabolic phenotypes. The most abstract aspect of FBA is the cellular objective, which is a mathematical representation of an evolutionarily optimal biological phenotype towards which a cell, when conceptualized as a control system, is hypothesized to strive. This notion of a cellular objective is therefore at once, both a mathematical abstraction of a complex biological phenomenon and a computational representation of an investigator's hypotheses. This fact alone suggests that the number of ways by which to formulate an FBA problem are infinite in principle. Yet we find that an overwhelming majority of published FBA models are formulated under a single parsimonious objective of growth rate maximization by way of flux maximization across a biomass reaction. We have demonstrated that, for a flux-balanced system, growth rate maximization and growth yield maximization are identical objectives. In this chapter, we follow up on these mathematical results by investigating the predictive capabilities of the iAF1260 metabolic model. We find that the growth rate maximizing objective converts FBA into a theory that is rigidly coupled to a profoundly simpler theory of energy stoichiometric balancing, much of which as we show can be understood from small scale representations of central metabolism. We conclude by suggesting that future work in the field of dynamic FBA (dFBA) will benefit from either decoupling this correspondence or reconsidering whether the inclusion of genome-scale metabolic representations makes the most efficient use of the genomic information that is available from reconstruction efforts.

5.1: Introduction:

The most commonly used Flux Balance Analysis (FBA) models may be classified as being a Palsson-type or a Fell-type based on how specific growth rates (μ) are treated, and to a lesser degree, on the scale on which metabolism is represented (section 3.5). Only with the Palsson-type models however, can dynamic flux balance analysis (dFBA) models of population growth be currently formulated. This is because dFBA models require an FBA formulation within which μ is a model prediction, which is made possible by the Palsson approach by defining a biomass reaction but not with the Fell approach which treats this quantity as a constraint. In a very large class of Palsson-type FBA models, μ is not only a prediction, but is also considered to be the cellular objective (formulation 3.2.15). The relation between μ and

yield was derived in the previous chapter for internally flux-balanced metabolic systems. Equation 4.4.6 applies with any organic byproduct and BQ may become non-zero due to a number of reasons. In the examples mentioned in section 3.1, it was seen that byproduct efflux may be induced in metabolic systems by upregulating pathways that have been engineered for the production of a commercially relevant compound. Under most natural settings however, non-zero BQ result from the activity of energy biogenesis pathways other than aerobic respiration (Varma and Palsson 1994, Wolfe 2005).

Equation 4.4.6 shows that provided there is no byproduct being produced (other than CO_2), growth rate and growth yields may be simultaneously optimized since one increases monotonically with the other. On the other hand, if further gains in growth rate could be induced from the production of metabolic byproducts, equation 4.4.6 equally shows that the optimization of rates will come at the cost of yields since YSC would switch to a monotonically decreasing function of rate (and BQ). Thus, the Palsson approach to FBA asserts that the evolutionarily derived cellular objective is to maximize the value of μ by correspondingly minimizing the value of BQ. By-product minimization in flux-balanced models is closely associated with the maximization of the efficiency of energy production (Varma and Palsson 1994). In view of equation 4.4.7 therefore, we are led to suspect that a large proportion of FBA predictions of rates and of yields that are made under the Palsson-style formulation may be explained by the model's representation of energy flux balance constraints. If this is the case, then it would follow that the FBA component of dFBA models may be more efficiently described by small scale metabolic models of energy biogenesis.

In the following, we use the results of the previous chapter to investigate the response of the iAF1260 model to various growth constraints, with an emphasis towards understanding flux-

redistributions as we find them to be the predictions of greatest importance to dFBA modelling. By considering the biological information that is necessary for bringing about the model responses, we investigate whether genome-scale metabolic representations are truly necessary for the prediction of growth rates and growth yields under the Palsson-type FBA formulation.

5.2: Distinct growth states of *E.coli* are predicted *in silico* by Palsson FBA during aerobic growth on xylose

The goal of this section is to detail the stoichiometric response to increasing substrate consumption of the iAF1260 model under the constraint of a maximal oxygen consumption rate. To generate our data, we used the iAF1260 *E.coli* model with the objective of maximizing flux through the biomass reaction to produce a series of FBA solutions. The FBA solutions profile aerobic growth predictions over increasing rates of xylose consumption from a minimum value of 0.0 mmol $GDW^{-1}h^{-1}$ to a maximum value of 50.0 mmol $GDW^{-1}h^{-1}$ in intervals of 0.1 mmol $GDW^{-1}h^{-1}$. The maintenance parameters of the FBA model were kept at their originally published values of 8.39 mmol $GDW^{-1}h^{-1}$ for NGAM and 59.8 mmol GDW^{-1} for GAM (Feist, Henry et al. 2007) . To investigate the response of the FBA model to energetic constraints, we have constrained the upper bound of the oxygen consumption rate to 20.0 mmol $GDW^{-1}h^{-1}$. A maximal oxygen uptake rate is an appropriate constraint to consider for both the Fell and Palsson FBA modelling of aerobic *E.coli* growth since several independent measurements are now available for its upper bound. These measurements, made available from chemostat experiments, show that the maximal uptake rate (v_{O2}^{Max}) consistently falls near the range of 15.0~20.0 mmol $GDW^{-1}h^{-1}$ under ideal conditions (Stouthamer and Bettenhaussen 1975, Tempest 1978, Andersen and von Meyenburg 1980, Calhoun, Oden et al. 1993, Varma and Palsson 1994, Feist, Henry et al. 2007).

Of the 2381 reactions in the stoichiometric network of iAF1260 it was observed that only 532 were predicted to be active at some point in the series data suggesting that biosynthesis can be theoretically supported by a relatively small proportion (~22%) of the available reactions. 79 of the active reactions (~15%) were seen to be exchange reactions that do not map to genes; this is only slightly higher than the proportion of all reactions of the overall metabolic model that are exchange reactions (297 or ~ 12%). By-product production was observed at higher substrate consumption rates (> 0.85 $\mathrm{mmol\ GDW^{-1}h^{-1}}$); based on their profiles, three general growth phases could be identified for the iAF1260 network. The first phase involves zero by-product production and occurs at low growth rates. This phase is followed by an acetogenic phase corresponding to intermediate growth rates. Finally, at higher growth rates, a combination of acetogenesis and ethanologenesis was observed. These observations are presented in figure 5.2.1 where biomass yield on substrate carbon (YSC) is plotted as a function of growth rate (μ).

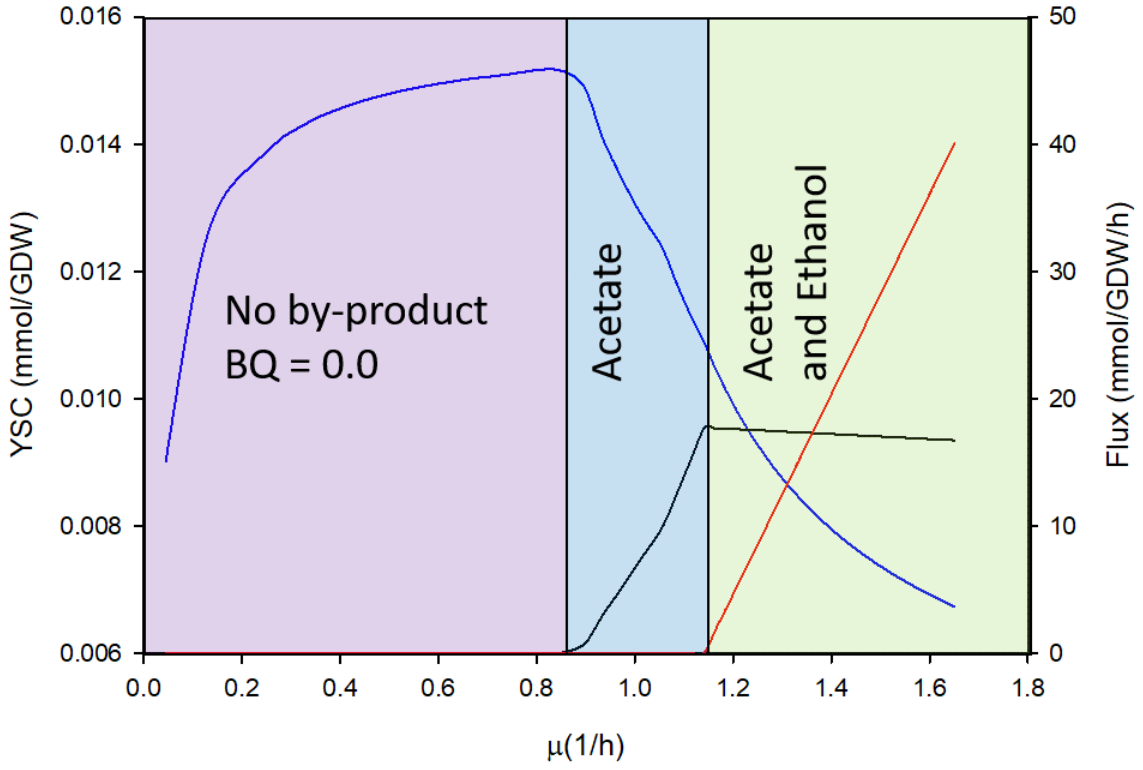


Figure 5.2.1: The iAF1260 model was grown aerobically on xylose over a consumption range from 0.0 to 50.0 mmol/GDW/h. The plot is of the predicted yield on substrate carbon (YSC) against predicted specific growth rate (left vertical axis). The YSC curve is seen to be hyperbolic when the by-product quotient (BQ) is zero (purple region). In the blue and green regions, BQ becomes positive first through acetogenesis and then through a joint acetogenesis and ethanologenesis (by-product fluxes correspond to the right vertical axis). As a result, YSC is seen to monotonically decrease with increasing specific growth rate in these two regions.

In accordance with our prediction (equation 4.4.6), it is seen that YSC exhibits a hyperbolic relation to μ under regimes of growth unaccompanied by energetic byproducts other than CO_2 (since BQ = 0.0). In contrast, YSC is seen to decrease monotonically with μ under growth regimes where BQ becomes increasingly positive.

An inspection of the flux distribution predictions show that these transitions occur as a result of energy flux-balance constraints. The first qualitative response from the flux distribution

(a *flux redistribution*) is seen to coincide with the maximization of oxygen consumption rate (Figure 5.2.2).

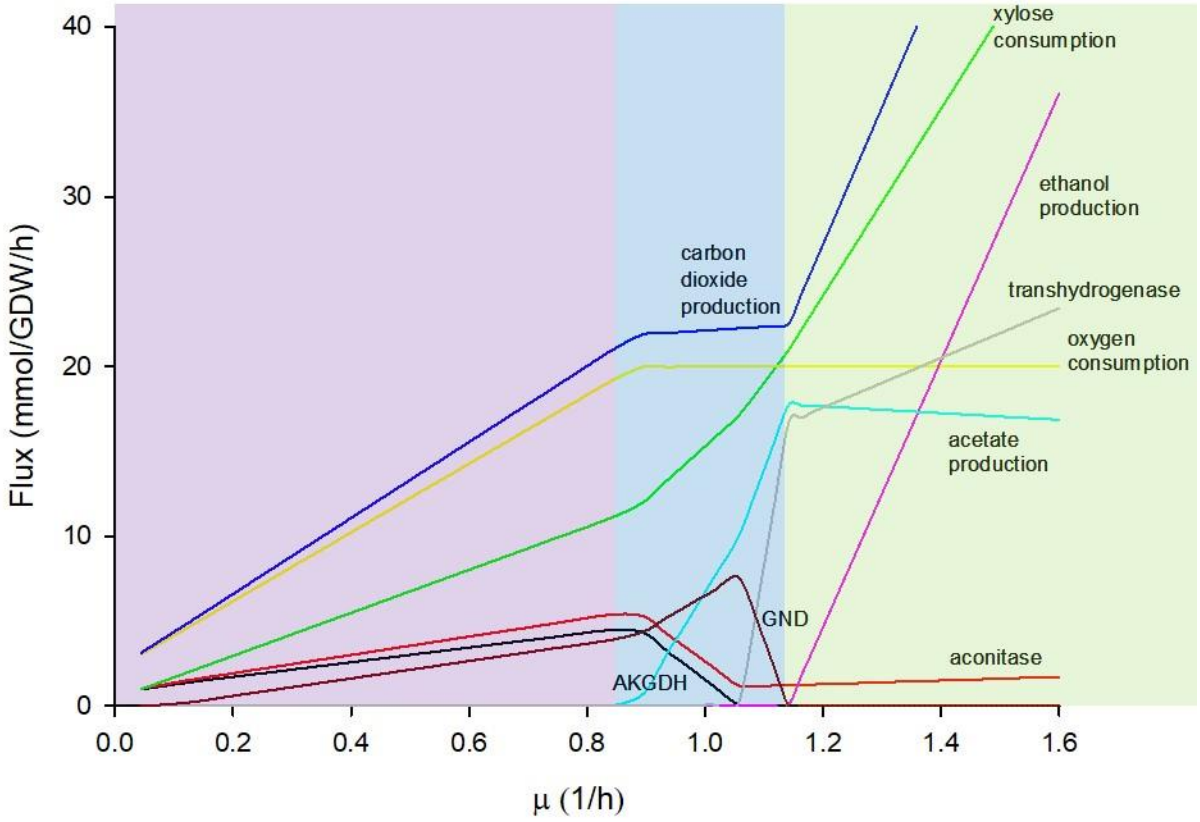


Figure 5.2.2: The iAF1260 model was grown aerobically on xylose. The flux response of several representative reactions have been plotted against specific growth rate to understand the flux redistributions that lead to changes in growth regimes. The AKGDH and aconitase represent TCA curves illustrate TCA branching, and this effect on PP is represented by the GND curve. Note that the acetogenic region can be further subdivided into one where GND is increasing with growth rate and GND is decreasing with growth rate. This behavior is explained by the transfer of NADP reductive function from PP to the transhydrogenase reaction (grey curve).

Upon the arrival of this constraint, the growth phase characterized by the absence of organic by-products transitions to one characterized by acetogenesis. As evidence by the continued production of CO_2 as well as the high flux through the tricarboxylic acid cycle (TCA),

this second growth phase is seen to be characterized energetically to be a respiro-fermentative phase where ATP is produced both oxidatively by the electron transport system (ETS) as well as by incomplete substrate-level oxidation by pathways leading to acetate (via glycolytic enzymes and acetate kinase).

The possibility of producing acetate as the only non-respiratory efflux metabolite in spite of the ETS carrying a maximized constant flux evidences that stoichiometric redox balancing can be achieved internally during this growth phase. The metabolic model was seen to employ two key strategies for internal redox balancing. In the early stages of the acetogenic phase, redox balancing was seen first to be supported by a linear response from fluxes through the tricarboxylic acid cycle (TCA) and the reductive pentose phosphate pathway (PP) (Figure 5.2.2). The iAF1260 model predicts that the TCA cycle would become increasingly branched through the suppression of its dehydrogenases upon maximization of oxygen flux. This is seen to be because both the fluxes through succinate dehydrogenase and the alpha ketoglutarate dehydrogenase contribute electrons to the ETS (Voet and Voet 2004, Keseler, Collado-Vides et al. 2005, Berg, Tymoczko et al. 2006, Nelson, Lehninger et al. 2008). In response to this linear decrease in TCA flux, figure 5.2.2 shows that a corresponding linear increase in the flux through the reductive pentose phosphate pathway is predicted to occur. This is indicative of an increasing contribution by the PP to NADP+ reduction in response to the decreasing contribution from the (prokaryotic) TCA. Notice that although the TCA branches, it is not suppressed completely as this pathway, in its non-cyclic form, is also needed for its biosynthetic functions.

The second strategy is applied when the TCA branches completely. When this occurs, a slight kink in the curve of YSC against μ may be observed in the exclusively acetogenic region. This transition was found to correspond to an onset of flux through the transhydrogenase

reaction. Recall that the periplasmic transhydrogenase reaction catalyzes an oxidation reduction between the cofactor pairs NADH/NAD+ and NADPH/NADP+ (Keseler, Collado-Vides et al. 2005, Berg, Tymoczko et al. 2006, Nelson, Lehninger et al. 2008) . This reaction was predicted by the iAF1260 model to proceed in the direction of NADH oxidation and NADP+ reduction. Thus, upon a complete TCA branching, the iAF1260 model predicts that the transhydrogenase functions to maintain a redox balance in the NADH/NAD+ pool by passing reductive potential to the NADPH/NADP+ pool. As the flux through the transhydrogenase reaction is increased, the redox balance in the NADP/NADPH pool is maintained by correspondingly decreasing the rate of NADP+ reduction by the PP pathway.

The same principle of redox balance which explains the flux redistribution that occurs in the acetogenic growth phase is seen to also explain the transition between the second and the final growth phase where a combination of acetogenesis and ethanologenesis is predicted. The onset of ethanologenesis is seen to coincide with the point at which the (reductive) PP becomes completely suppressed (Figure 5.2.2). Ethanologenesis was therefore predicted to occur when internal redox balancing is no longer possible and a flux (through the alcohol dehydrogenase reaction) diverting carbon away from biomass precursors becomes necessary.

In summary, we have been able to verify that equation 4.4.6 can be used to predict the behavior of Palsson-style FBA models. An explicit connection is shown to be made between energy flux-balance and byproduct formation when we introduce, as a constraint, the auxiliary information of a maximal oxygen consumption rate. In this common setting, we find that it is possible to reduce the predictive response of the Palsson FBA model to simpler energetic principles of by-product minimization. In all phases of growth, the Palsson-FBA solution is a flux distribution that most efficiently generates the energy required to support biosynthesis.

5.3: Are predictions unique to the iAF1260?

We have repeated the simulations of the iAF1260 metabolic model with the iJR904 and iJO1366. The results are presented in figure 5.3.1. It can be seen that despite the differences in dimension between the models, the qualitative behavior of flux distributions are identical.

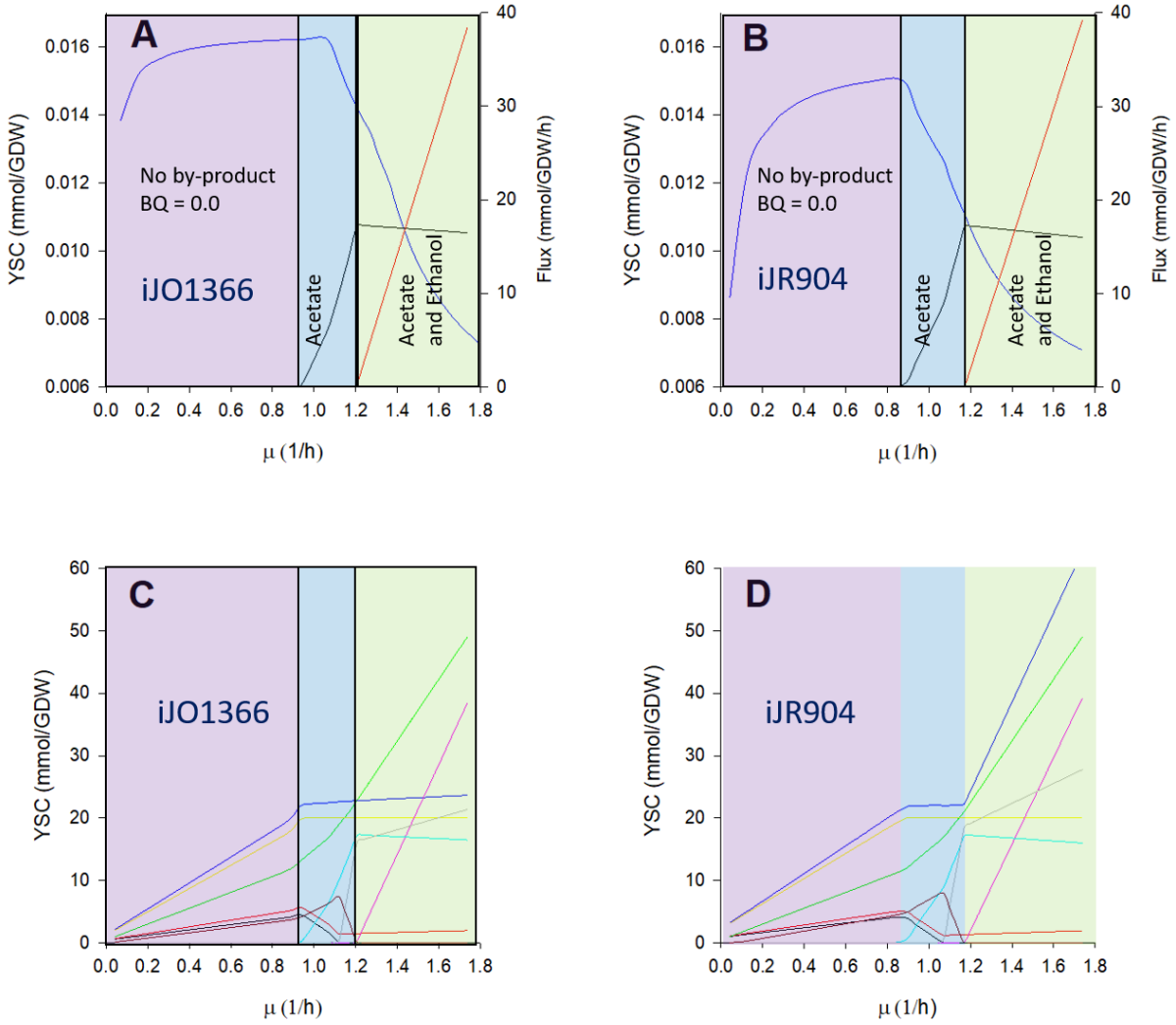


Figure 5.3.1: The aerobic xylose growth simulated for the iAF1260 model was repeated for the iJR904 and the iJO1366 *Ecoli* metabolic models. Plots A and B show the response of YSC to increasing growth rate as well as the byproduct fluxes for the two models. It is seen that like for the iAF1260 model, both the iJR904 and the iJO1366 model predict three growth regimes characterized by byproduct formation. Plots C and D show that the underlying flux redistributions that lead to regime changes are identical to the predictions of the iAF1260 model. For simplicity, the flux curves in plots C and D are not been given labels, and the reader is encouraged to compare these plots with figure 5.2.2 of this chapter.

5.4: Unequal distribution of energetic flux volume in the iAF1260 network

Our analysis of the response of the iAF1260 model detailed in section 5.2 lead us to appreciate the significance of a metabolic model's need to internally balance energetic cofactor concentrations to predictions of growth rate and growth yield. Under the simple constraint of a maximal oxygen consumption rate and the hypothesis of growth rate maximization, Palsson-style FBA models appear to predict that growth rates are optimized by minimizing the production of byproduct as much as possible for a particular value of μ^{Max}. Thus, yields are optimized secondarily in growth regimes involving by-products, and primarily (alongside rates) in those for which BQ = 0.0. In consideration of the energetic nature of the principles underlying this strategy of rate maximization, we turn to the fact that in the iAF1260 metabolic model, the number of reactions involving energetic cofactors is far less than the total number of reactions. Of the 2381 reactions available to the iAF1260 network, 340 reactions involve ATP, 111 reactions involve NADH, and 81 reactions involve NADPH (Feist, Henry et al. 2007). Thus, combined with the fact that conservation laws apply individually to each cofactor, we identify the potential to characterize the growth rate and growth yield predictions of Palsson models using a handful of energetic reactions involved with cofactor interconversions. In this section, we investigate which reactions of the iAF1260 model contribute most to the quantities $v_{ATP}, v_{NADH}, \text{and } v_{NADPH}$ (notation described in section 2.2). Towards this end, a single point FBA optimization was performed with iAF1260 under same setting as in section 5.2; maximal xylose and oxygen uptake rates were set at respectively 10 and 20 $\text{mmol GDW}^{-1}\text{h}^{-1}$. The maintenance parameters were set to NGAM = 8.39 $\text{mmol GDW}^{-1}\text{h}^{-1}$ and GAM = 59.81 mmol GDW^{-1}.

Because Palsson FBA models optimize for yield given μ^{Max} , an unconstrained optimization has been shown to predict that the most efficient energy conversion strategies will be used. For xylose transport, this would imply that flux through the symporter will be used in favor of flux through the ABC transporter (section 1.3). Likewise, the most energetically coupling NADH dehydrogenase and cytochrome oxidoreductase pair would be predicted to carry a maximally allowed flux at the electron transport system (section 1.6). In consideration of the transport step, it is seen from the setup that the energetic cost of symport may be stoichiometrically converted to an ATP cost much in the same way we have done for oxygen (equation 4.3.1). As we are interested only in investigating the contribution that the transport step makes to v_{ATP}, we have chosen to simplify matters by constraining the flux through the xylose symporter to zero so that, via flux through the ABC transporter, the transport step would make a direct contribution to ATP hydrolysis rate. We are similarly not interested in precisely determining what the stoichiometries are amongst the energetic cofactors. Therefore, no constraints were placed on the reactions of the electron transport system. This has left the iAF1260 model to predict that the most coupling NADH dehydrogenase and cytochrome oxidoreductase will exclusively carry fluxes, leaving their respective counterparts with a flux prediction of zero.

Under these conditions, xylose was predicted to be consumed at the maximal possible rate of 10 mmol $GDW^{-1}h^{-1}$; oxygen consumption was seen to occur at 18.7 mmol $GDW^{-1}h^{-1}$, with a carbon dioxide production rate of 20.2 mmol $GDW^{-1}h^{-1}$. No by-product production was predicted as can be explained, in light of section 2 of this chapter, that the predicted value of v_{O2} is below $v_{O2}^{Max} = 20.0$ mmol $GDW^{-1}h^{-1}$. The biomass objective was predicted to carry an optimal flux of $\mu^{Max} = 0.73\ h^{-1}$. From the resulting flux distribution, the locations in the

network where energy transactions were predicted to take place were identified. We discuss our findings in turn for ATP, NADH, and NADPH.

ATP synthesis was predicted to occur primarily via oxidative phosphorylation at the ETS (ATP synthase: 59.46 mmol $GDW^{-1}h^{-1}$), and to a slightly lesser extent at the substrate level (PYK = 9.819 mmol $GDW^{-1}h^{-1}$; PGK: 13.72 mmol $GDW^{-1}h^{-1}$ SUCCOAS: 4.458 mmol $GDW^{-1}h^{-1}$). The total rate of ATP synthesis was calculated to be 89.84 mmol $GDW^{-1}h^{-1}$. Compared to the five reactions of ATP synthesis, eighty-eight reactions were found to be actively consuming ATP (Figure 5.4.1).

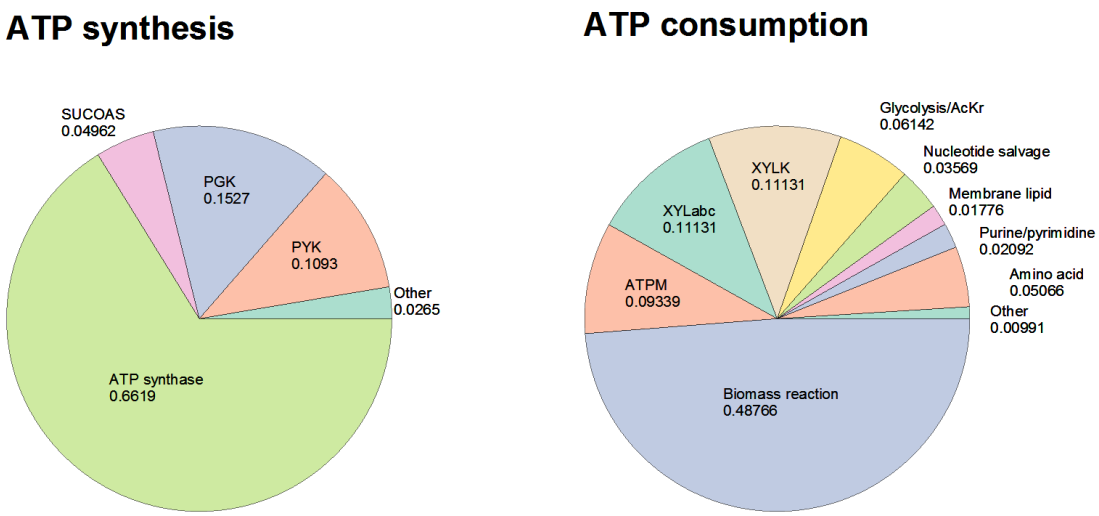


Figure 5.4.1: The fraction of total ATP synthesis and ATP consumption attributable to single reactions and groups of reactions.

It can be seen that the hydrolysis from the biomass reaction accounts for almost half of the energetic demand in the form of ATP consumption (~49% of the net rate of ATP synthesis). This fraction increases to ~58% with the inclusion of hydrolysis by an equally phenomenological 'ATP hydrolysis of maintenance reaction' (ATPM; discussed in section 4.2). Finally, with the inclusion of central metabolism and xylose acquisition, it is found that hydrolysis by just these six reactions account for ~87% of the total ATP demand in the iAF1260 model (Figure 5.4.1).

A similar observation was made for NADH production and consumption (Figure 5.4.2).

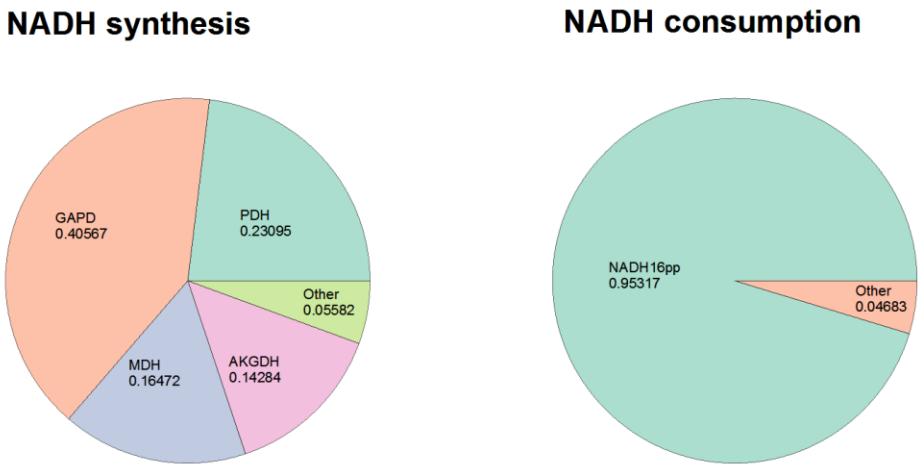


Figure 5.4.2: The fraction of total NADH synthesis and NADH consumption attributable to single reactions and groups of reactions.

Nineteen reactions were seen to reduce NAD+ to NADH at a combined rate of 33.82 mmol $GDW^{-1}h^{-1}$. The combined fluxes of four reactions of central metabolism (PDH,

GAPD, MDH, AKGDH) account for ~94% of the net synthesis; The coupling dehydrogenase of the ETS was seen to oxidize the resulting NADH at a flux of 32.3 mmol $GDW^{-1}h^{-1}$ or ~95% the total consumption rate. The remaining NADH demand was seen to be exerted by a folate metabolism reaction, cofactor biosynthesis, and the reactions of cell envelope biosynthesis. NADPH was found to be produced at a net rate of 11.8 mmol $GDW^{-1}h^{-1}$ by four reactions (Figure 5.4.2). Three central metabolic reactions of the reductive pentose phosphate pathway (PP), together with the tri-carboxylic acid cycle (TCA) are shown to account for ~94% of the synthesis.

The oxidation of NADPH was observed to be distributed over non-central metabolic reactions, with large contributions from amino acid and cell envelope biosynthesis (Figure 5.4.3).

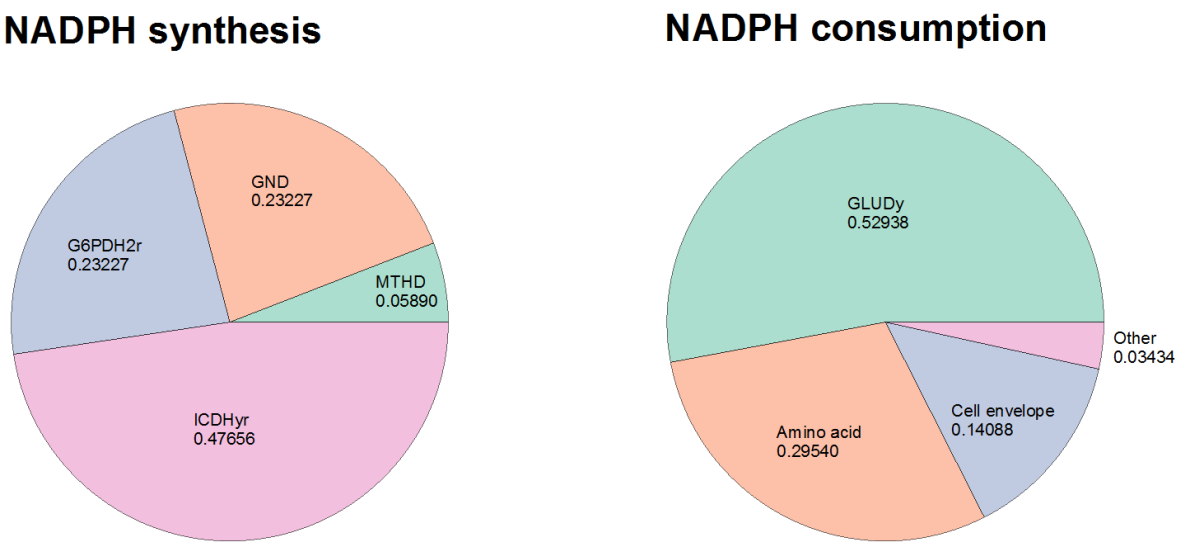


Figure 5.4.3: The fraction of total NADPH synthesis and NADPH consumption attributable to single reactions and groups of reactions.

The amination for alpha ketoglutarate by the NADP+ dependent glutamate dehydrogenase was seen to make a rather significant contribution by fraction (just over 50%) to the net NADPH consumption rate.

To summarize our observations of energetic flux profiles, we find that a small number of reactions may make a large contribution to the net production and consumption rates of energetic cofactors. This is most evidently true for the production rates of the three cofactors we have considered. For both ATP and NADH however, the same pattern is exhibited by their consumption rates. The majority of production fluxes for ATP, NADH, and NADPH may each be accounted for by just four reactions, all of which for ATP and NADH and three of which for NADH, are central metabolic reactions. The consumption flux of ATP has a relatively high fraction that may be attributed to biosynthetic reactions, however, it is evident that the large majority of the flux contribution comes directly from four reactions; these are namely the xylose ABC transporter, the xylulokinase reaction, the ATP maintenance reaction, and the biomass reaction. For NADH consumption rate, biosynthetic demands make up a very small fraction relative to the considerable contribution arising from the NADH dehydrogenase. In contrast, the net NADPH consumption rate appears to result from several biosynthetic reactions (note that we have presented the reactions as lumped reaction families in figure 5.4.3 for this reason). However, it is seen that the net NADPH consumption rate also depends on an unequally large contribution from just one reaction, the glutamate dehydrogenase reaction.

In view of the empirical context in which the Palsson FBA method is usually employed, that is, in the absence of large-scale kinetic data to constrain the model with, we find it highly relevant that a small number of reactions make an unequal contribution to the rate of energy production and energy use. This is because the efficiency of energy biogenesis is directly related

to the efficiency of biosynthesis, and so both are optimized under the assumption of a biomass objective (sections 4.4 and 5.2). We are encouraged by these observations to question again, the necessity of genome-scale metabolic reconstructions to FBA methods that are applied under the Palsson setting towards the prediction of metabolic rates. The reader is referred to sections 3.1 and 3.5 for discussions of why this statement does not apply generally to all FBA methods. In the case of rate predictions and yield predictions, both of which are relevant to dFBA modelling, the possibility that a genome-scale metabolic model may be greatly simplified and yet still quantitatively make the same desired predictions as the larger model is promising.

5.5: The iAF1260 model applies a common yield-maximizing strategy to different substrates

One of the most commonly used application of FBA methods is in the prediction of the response of growth phenotypes to different substrates. Substrate-specific phenotype phase planes (Phpp) have been studied in depth for *E.coli* and dFBA methods often require such predictions when a model must account for multiple growth phases in a single growth data (Edwards, Ibarra et al. 2001, Ibarra, Edwards et al. 2002, Mahadevan, Edwards et al. 2002, Meadows, Karnik et al. 2010, Hanly and Henson 2011). In this chapter, we have analyzed, the response of Palsson-style FBA models to varying xylose consumption rate with the additional assumption that oxygen consumption rate comes with a kinetic upper bound. We have discovered in doing so that Palsson-style models exhibit a growth rate optimizing strategy that optimizes for yield secondarily as an indirect cellular objective. This strategy was seen to lead to FBA solutions that favor pathways that are most efficient with respect to energy production and use. Strikingly however, we find that there is very little about this strategy as well as of how the models appear to implement it via central metabolic flux redistributions that is specific to the

substrate being xylose. Thus, we investigate the growth phenotype predictions of the iAF1260 model to substrates other than xylose. In this section, we define "growth phenotype" to mean the qualitative response of flux distributions to increased substrate uptake rate. The FBA series described in section 5.2 for xylose was repeated with glucose, malate, succinate, and acetate. The choice of compounds was based upon the set of substrates chosen by Edwards, Ibarra , & Palsson (2001), and Ibarra, Edwards, & Palsson (2002) (Edwards, Ibarra et al. 2001, Ibarra, Edwards et al. 2002). We now describe our observations for substrates in the order of similarity in model response to our observations for xylose (section 5.2).

For growth on glucose, the response of the iAF1260 model to an increasing rate of glucose exchange was identical to the observations already made for xylose (Figure 5.5.1).

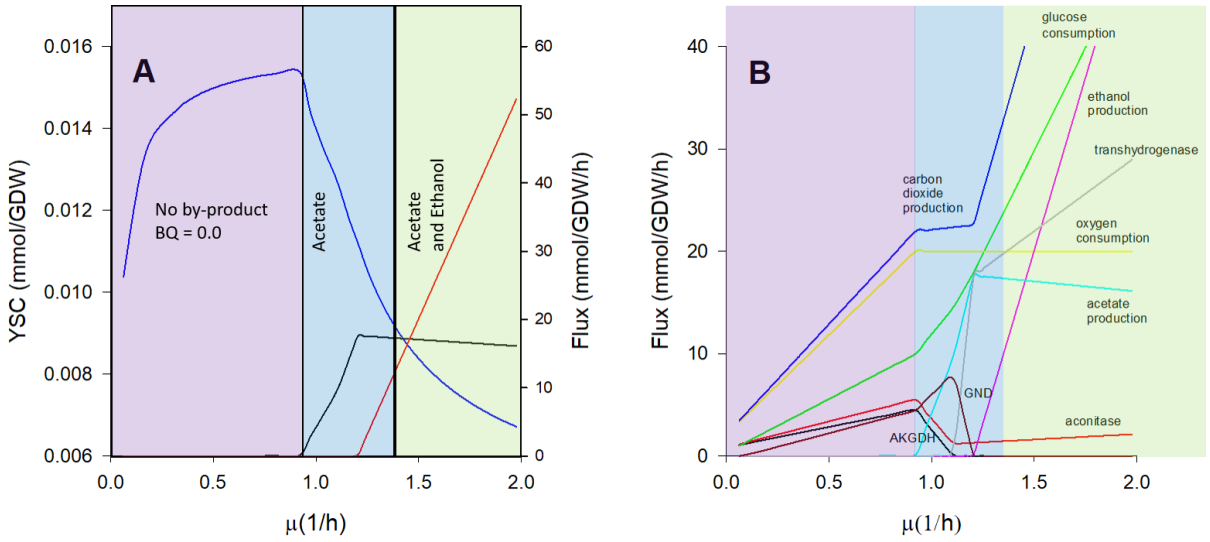


Figure 5.5.1: The iAF1260 model was grown aerobically on glucose. Plot A should be compared with figure 5.2.1 corresponding to xylose growth. Similarly, plot B should be compared with figure 5.2.2 also corresponding to xylose growth. We see from the glucose plots that the qualitative response of the iAF1260 model to increased glucose consumption under the constraint of a maximal oxygen uptake rate is identical to that for xylose.

The response during simulated aerobic growth on malate was similar to that observed for glucose. One notable difference was observed for the malate growth; namely the absence of a transhydrogenase phase partitioning the strictly acetogenic phase (Figure 5.5.2).

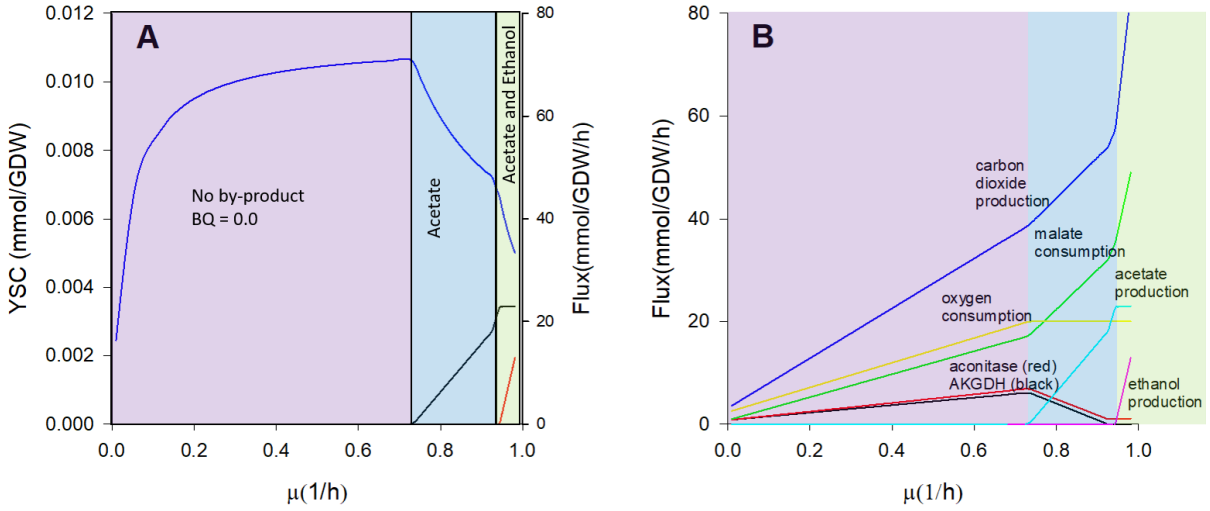


Figure 5.5.2: The iAF1260 model was grown aerobically on malate. We see that malate growth is qualitatively like both glucose and xylose growth except that for malate, there is no GND activity in any of the growth phases analyzed. A consequence of this is that, for malate, the transhydrogenase phase partitioning the acetogenic regime is lost, as this process involves the linear transfer of flux volume from PP to the transhydrogenase.

Transhydrogenase was never predicted to be utilized by the model during aerobic growth on malate. An inspection of the central metabolic flux distribution revealed this to be due to the fact that the iAF1260 network never employs a flux distribution using the reductive pentose phosphate pathway (PP) during non-acetogenic growth. Only the interconversion branch of PP was predicted to be active; in these flux distributions, PP function exclusively as a sugar-interconverting pathway rather than an energetic one. For growth on xylose, it is recalled that the transition from an exclusively acetogenic to a joint acetogenic and ethanologenic growth

involves the diminishing role of the reductive PP and correspondingly, an increase in flux through the transhydrogenase reaction with NADP+ reduction (section 5.2). The FBA prediction for malate arises, because the reductive PP never plays a role in NADP+ reduction; the branching of the tricarboxylic acid cycle (TCA) is immediately followed by ethanol production.

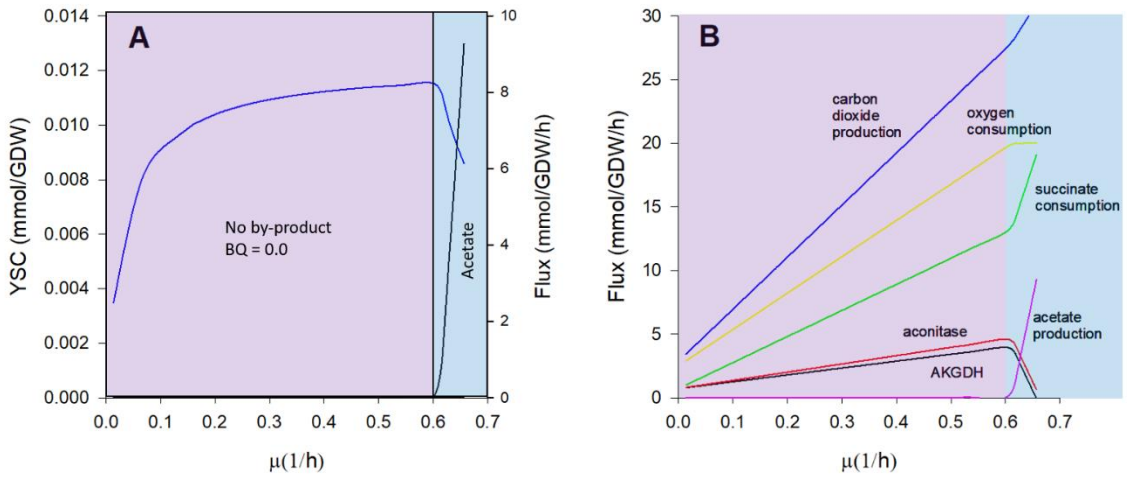


Figure 5.5.3: The iAF1260 model was grown aerobically on succinate. We see that succinate growth is qualitatively like malate growth except that there is no ethanologenic regime. Note that the succinate consumption curve (green) terminates below the maximal allowed uptake rate over the flux series (50 mmol/GDW/h), indicating that the iAF1260 model is unable to use ethanol formation as a strategy for further increasing growth rates. Succinate consumption is predicted below its FBA bound as further consumption is not possible without disrupting an internal redox balance.

Growth on succinate resembled growths on xylose, glucose, and malate but only up to TCA branching (Figure 5.5.3). Plot B of figure 5.5.3 reveals that succinate consumption never exceeds a value of 19.03 mmol $GDW^{-1}h^{-1}$. Thus, for exchange bounds placed beyond this limit, the maximal growth rate achievable with succinate metabolism is predicted to be constrained by factors other than substrate availability. The iAF1260 model predicts for

succinate, that once having entered the cytoplasm, this substrate is consumed primarily by the succinate dehydrogenase (SUCDi) of the TCA, and to a greatly lesser extent by Propanoyl-CoA-succinate CoA-transferase (for comparison, the latter reaction carried a final flux of 0.34 when, in the same flux distribution, the former carried a flux of 19.03 mmol $GDW^{-1}h^{-1}$). For xylose, glucose, and malate, redox- balancing was seen to require the suppression of SUCDi. As such, the stoichiometric constraint from the need to balance redox cofactors appears to prevent the consumption of this succinate beyond a limiting value. This value was seen to be set by maximal oxygen consumption (Figure 5.5.3 B). In sum, during aerobic growth on succinate, the iAF1260 model was shown to predict transitions analogous to that of malate but terminating prematurely at the point of ethanologenesis.

Finally, acetate metabolism was predicted to exhibit only a single state of growth in which all fluxes scale directly with μ (Figure 5.5.4).

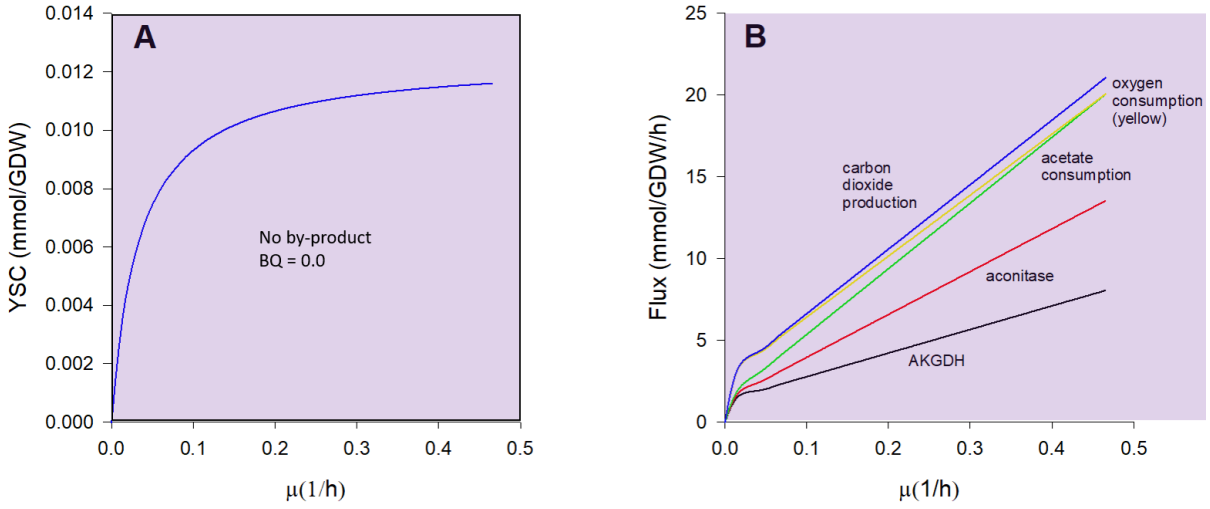


Figure 5.5.4: The iAF1260 model was grown aerobically on acetate. Acetate growth, like succinate growth terminates at a value below the maximal allowed consumption rate (green curve). Similarly for succinate, acetate growth does not have the three growth regimes predicted for xylose, glucose, and malate. Instead there is only a single growth phase for acetate, which is the one in which BQ is equal to zero. Note that we have not included a second axis in plot A, as there were no byproduct flux curves for acetate.

As it was the case for succinate, acetate consumption for large exchange bounds was observed to be limited below the constraint from substrate availability. It can be concluded from the energetic strategies used by the iAF1260 model in metabolizing the other substrates we have considered in this chapter, that this limit exists as a result of the acetate kinase catalyzing in the ATP-hydrolyzing direction during growth on acetate. Here, the metabolic strategy of increasing v_{ATP} by respiro-fermentative energy biogenesis is not feasible as acetate, the by-product of fermentative ATP synthesis, is also the substrate.

In this section, it was investigated whether qualitative predictions of the iAF1260 model during aerobic growth on xylose extrapolate to aerobic growth on substrates other than xylose. To a large extent the answer was found to be in the affirmative. A common strategy for increasing μ was seen to be shared by the growth simulations across various substrates; to grow without by-product production up to a limit set by oxygen consumption; to utilize acetogenesis for further ATP synthesis whilst branching the TCA; to pass the reductive function of the PP if appropriate to an alternative redox process. Differences were seen to arise when the processes of substrate acquisition conflict with the requirements of this strategy.

5.6: Substrates with similar acquisition pathways are not distinguishable by yield

The Palsson-style FBA optimizes for rates by optimizing for yields. As a consequence, differences in rate predictions in the Palsson framework may be attributed to differences in the ability of a metabolic network to satisfy energy flux balance constraints (sections 5.2 and 5.5). Yet we have seen that the aspects of a metabolic model that energetically determine growth rates may be relatively few with respect to the dimensions of the overall model, and even fewer when limited to those specific for a given substrate (sections 5.3, 5.4, and 5.5). The part of a metabolic

network whose activity is energetically relevant in a substrate-specific manner is often the small number of reactions required for the transport and early metabolism of the substrate (e.g. the ABC transporter and the kinase in figure 5.4.1). In light of this, we investigate the degree to which substrates with similar modes of acquisition may be distinguished by their predicted maximal yields under the Palsson framework.

The iAF1260 model was grown on six different sugar substrates – xylose, arabinose, ribose, glucose, mannose, and fructose. The choice of substrates was made based on the existence of previously demonstrated catabolite hierarchies and on the similarity of their mode of acquisition (Kang, Song et al. 1998, Desai and Rao 2010). Figures 5.6.1 and 5.6.2 summarize the steps by which these sugars are converted to common central metabolic intermediates.

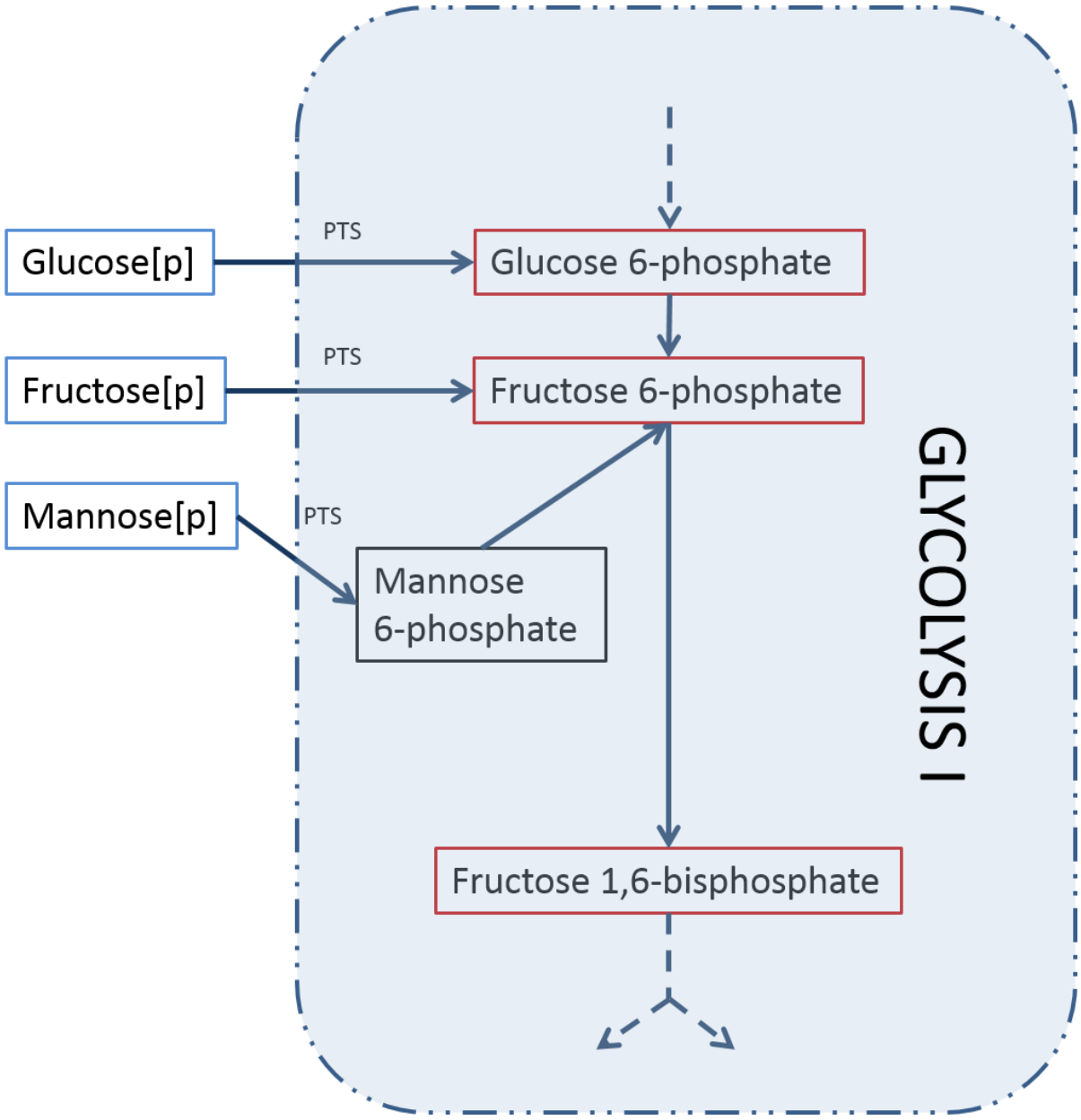


Figure 5.6.1: The hexose sugars -- glucose, mannose and fructose -- are transported via a phosphotransferase system (PTS) and subsequently enter metabolism through the first glycolytic branch. The '[p]' signifies that the sugars are being transported from the periplasm in this figure. This figure has been adapted from the textbook Bacterial physiology and metabolism (Kim and Gadd 2008).

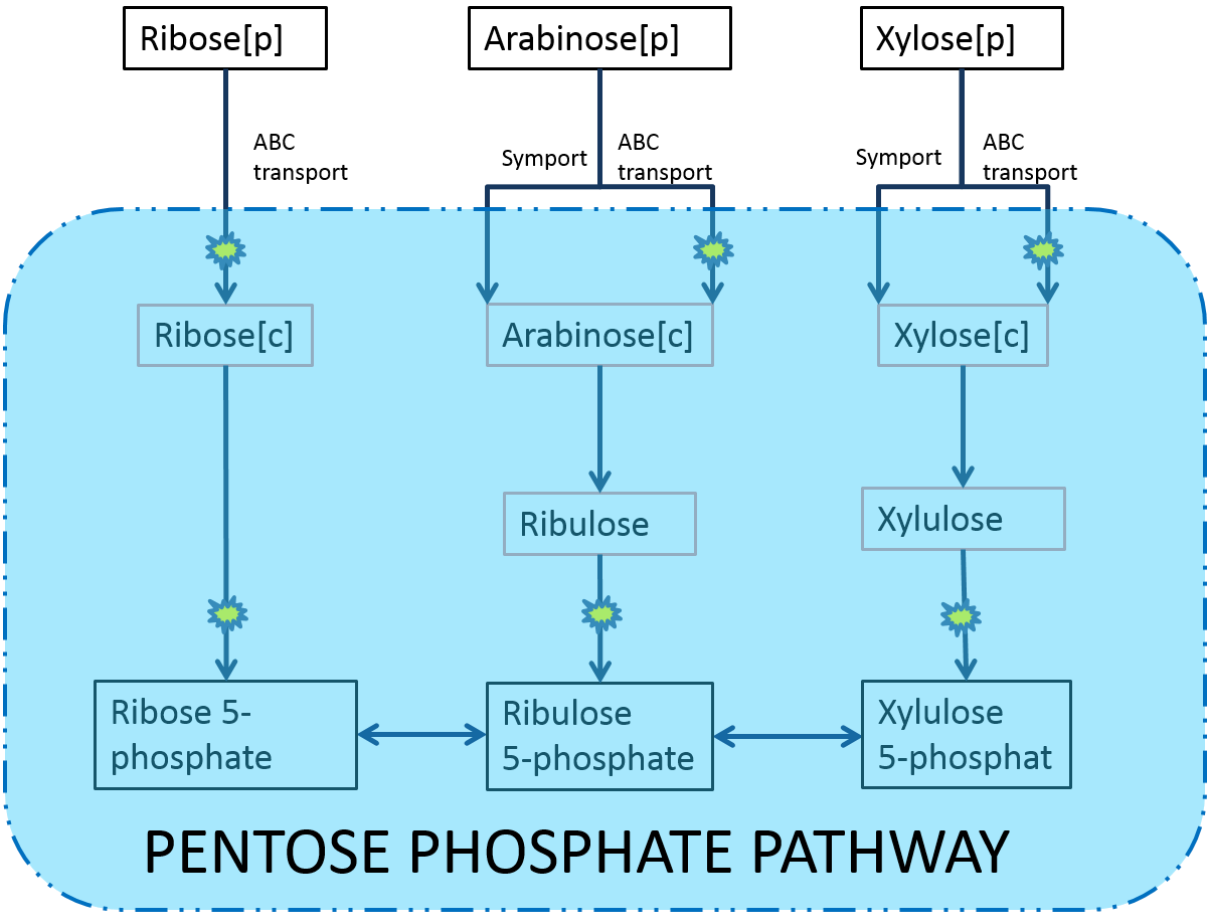


Figure 5.6.2: Xylose, arabinose, and ribose are pentose sugars that enter the central metabolism through the interconverting branch of the pentose phosphate pathway. *E.coli*, for the former two sugars has available an ABC transporter, and a proton symporter, whereas ribose can only be transported by an ABC transporter. The '[p]' signifies that the sugars are being transported from the periplasm in this figure. A bolt-mark superimposed upon a transport arrow indicates ATP hydrolysis. This figure has been adapted from the textbook Bacterial physiology and metabolism (Kim and Gadd 2008).

For each of the pentoses, three growth simulations were run with substrate uptake rates of respectively 10.0, 15.0, and 30.0 $\text{mmol} \cdot \text{GDW}^{-1} \cdot \text{h}^{-1}$. These consumption rates correspond to the three distinct growth phases observed previously in section 5.2 for the aerobic metabolism of xylose as characterized by by-product production. Growth was demonstrated to be accompanied by no by-product at 10.0, by acetogenesis alone at 15.0, and by joint acetogenesis and ethanologenesis at 30.0 $\text{mmol} \cdot \text{GDW}^{-1} \cdot \text{h}^{-1}$ uptake rates. The reader is referred to figure 5.2.2

to corroborate this information for xylose. Simulations were run in an identical manner for the hexoses as they were for the pentoses. In order to control for the total flux of carbon however, the hexoses for each growth phase were made available at five-sixth the rate for the pentoses; thus the influx of total carbon *atoms* was the same regardless of whether the iAF1260 model was grown on a hexose or a pentose. Higher yields correspond to the use of transporters with the lowest energy demands; thus, as in section 5.4, the FBA solution when unconstrained will favor a symport transporter over an ABC transporter under the Palsson formulation. To simulate the uptake of a given substrate through a symport process, the FBA model was optimized without constraints on transporters; to simulate the uptake through an ABC transporter, the model was optimized with the corresponding symport flux constrained to zero. As in the previous sections of this chapter, the maximal oxygen uptake rate was constrained to 20 $\mathrm{mmol} \cdot \mathrm{GDW}^{-1} \cdot \mathrm{h}^{-1}$similarly, the rate of ATP hydrolysis for non-growth-associated maintenance was constrained to a flux of 8.39 $\mathrm{mmol} \cdot \mathrm{GDW}^{-1} \cdot \mathrm{h}^{-1}$ (Feist, Henry et al. 2007).

Of the six sugars tested, two sugars – xylose and arabinose – may be transported by either a symport protein or an ABC transporter. For these sugars, the predicted specific growth rate was found to decrease very slightly (on the order of 0.01) when the transport was switched from a symporter to an ABC transporter (Table 5.6.1). This difference between predicted growth rates are seen to become more pronounced in regions 2 and 3 where growth is constrained by a combination of maximal oxygen consumption rate and redox balance. This trend, compared to the growth rate predicted for ribose, suggests that the observed differences arise as a result of the energetic efficiency of substrate acquisition rather than substrate identity. The growth rate prediction for ribose, whose transport and early metabolism require the expenditure of two units of ATP per unit of substrate, is seen to be identical to the growth rate predictions corresponding

to the ABC transport of xylose and arabinose. In a similar way, Table 5.6.1 shows that all sugars transported by a PTS exhibit the same growth rate predictions in all three regions of growth.

Sugar	Transport process used	Consumption rate (mmol/GDW/h)	μ (1/h)
Region 1: No by-product produced			
Xylose	Symport	10	0.76
Xylose	ABC transport	10	0.73
Arabinose	Symport	10	0.76
Arabinose	ABC transport	10	0.73
Ribose	ABC transport	10	0.73
Glucose	PTS	8.33	0.77
Mannose	PTS	8.33	0.77
Fructose	PTS	8.33	0.77
Region 2: Acetogenic but not ethanologenic			
Xylose	Symport	15	1.0
Xylose	ABC transport	15	0.90
Arabinose	Symport	15	1.0
Arabinose	ABC transport	15	0.90
Ribose	ABC transport	15	0.90
Glucose	PTS	12.5	1.0
Mannose	PTS	12.5	1.0
Fructose	PTS	12.5	1.0
Region 3: Acetogenic and ethanologenic			
Xylose	Symport	30	1.3
Xylose	ABC transport	30	1.1
Arabinose	Symport	30	1.3
Arabinose	ABC transport	30	1.1
Ribose	ABC transport	30	1.1
Glucose	PTS	25	1.4
Mannose	PTS	25	1.4
Fructose	PTS	25	1.4

TABLE 5.6.1: Comparison of specific growth rate predictions of the iAF1260 aerobic single-substrate growth for several sugars and different transport processes.

As was hypothesized, the output of the iAF1260 model demonstrates that Palsson-style FBA does not make significant distinctions in growth rate predictions amongst substrates when normalized for the total flux of carbon. The fact that we have normalized for the carbon flux is quite significant; the rate predictions given in this section are hypothetical growth rates

corresponding to hypothetical uptake rates of substrate. Note that one would expect the numbers under the final column of table 5.6.1 to show greater variation had we constrained the model with an experimentally observed uptake rate for each substrate. Thus, what is being compared is the rate of growth per unit of carbon delivered. The significance of the result therefore is that the Palsson-style method does not distinguish substrates by yield. However, it is recalled from equation 4.4.6 and from the discussions throughout this chapter that yields and rates exhibit a one-to-one relation within the Palsson framework. The results of this section therefore also suggest that the Palsson-style method does not distinguish substrates by growth rates unless an experimental uptake rate has been specified. If an experimental value is made available for the uptake rate of one substrate (e.g. glucose), then a carbon-normalized value of experimental growth rate may be calculated. Provided then with an experimental value for the uptake rate of a second substrate (e.g. mannose) then, the outputs of the iAF1260 model asserts that the resulting growth rate may be predicted a priori using the value of yield that is, as shown in table 5.6.1, identical for the two substrates.

5.7: Discussion

In this third and final chapter dealing with simple FBA models, we have presented results that demonstrate several consequences the relation between growth yield and growth rate (equation 4.4.6 and 4.4.7) have on the overall predictive capability of FBA models. It is important to point out that, because we have only considered FBA models maximizing for growth rate by maximizing the flux across a biomass objective, our main results apply to one very specific form of FBA which we refer to in this thesis as the "Palsson-style FBA." Given the wide-spread use of this method of FBA modelling however, we find that detailed analyses of its properties, both the strengths as well as the weaknesses, are useful. By applying equation 4.4.6 to the iAF1260 model, we have been able to show that Palsson models

optimize for growth rates first, and growth yields second. When no constraints are placed on the most efficient energy producing pathway (respiration), BQ remains zero; thus an FBA solution that maximizes for growth rate will also maximize for growth yield. The appearance of growth regimes within which yield and rate may no longer be maximized together requires an additional input of information. In this chapter, we have decided to introduce the hypothesis that the maximal rate of oxygen uptake is kinetically bounded, an assumption that is commonly used in the FBA community and is well founded experimentally (section 5.2).

In the general framework of FBA, a cellular objective is a hypothesis, and the biomass objective is only one of many possible hypotheses. However, towards applications such as dFBA that require models that predict growth rates, some mathematical connection must be asserted between the flux across a biomass reaction and a cellular objective. The most parsimonious solution is to assume that the two are one and the same; that is to assume the biomass objective. Because higher rates of biosynthesis means greater allocation of carbon to biomass, and equivalently lower carbon to by-products, Palsson FBA methods favour combinations of energy biogenic pathways that minimize the energetic efflux of carbon, whilst still obeying internal energy flux balance constraints. Thus, Palsson FBA solutions should all be considered to be yield-maximizing flux distributions. This is in the sense that although growth rate takes precedence over growth yield when the biomass objective is used, both quantities must be maximized with respect to energy biogenesis. The consequence of employing the Palsson-style FBA method is that a very strong coupling is assumed between growth rate, growth yield, and energetic efficiency which may not always be a biologically valid correspondence. Thus, we find that further FBA modelling, in settings such as dFBA where investigators are primarily concerned with rate predictions (as opposed to topological predictions), could be improved by devising modelling strategies that decouple the rigid correspondence between these three quantities in the Palsson framework.

Further to this goal of devising improved dFBA methods, the results of this chapter strongly indicate that dFBA models may be better served by small scale metabolic reconstructions. The results of

sections 5.3-5.5 indicate that the factors that determine the quantitative (rates and yields) as well as the qualitative (flux redistributions) growth phenotypes are surprisingly few in number relative to the total number of factors that are considered by genome-scale models. For example, of the 2381 reactions of the iAF1260 model, 300 reactions are those involving ATP; yet we have seen, in one simulation, that just 4 reactions (xylulokinase, xylose ABC transport, ATPM, and the biomass reaction), combined with the reactions of glycolysis, account for close to 90% of the total flux volume of ATP hydrolysis (Figure 5.4.1). It is likely the case that more complex behavioural predictions will become possible with the inclusion of more kinetic constraints and better objective functions. However, Palsson-style FBA models are useful in the absence of the experimental data that is required to verify that information. Where rate phenotypes are concerned, it is difficult to imagine realistic settings in which, given a set of reactions to model explicitly, defining the correct constraints and a biologically valid objective in a genome-scale model is not a task comparable in difficulty to measuring the required kinetic parameters. The quantitative predictions of the Palsson-style FBA modelling appear to depend strongly on the energetics of substrate acquisition and central metabolism. It has been shown that substrates whose transport and early metabolism share a common energetic stoichiometry cannot be distinguished by yield in the Palsson framework (section 5.6). These results appear, on a first glance, to indicate that there are flaws in the Palsson-FBA framework. However, we find that these results are actually quite encouraging from the standpoint of dynamic FBA modelling where the major source of model complexities is expected to be the differential equations. The metabolic reconstruction protocol outlined by Palsson describes how vast amounts of biological information may be integrated and summarised into a metabolic model and then curated. FBA methods provide ways of using this information to quantify globally relevant parameters such as biomass composition and maintenance energies. The fact that the prediction of FBA models depend on such quantities in a way that involves a small well studied component of a genome-scale network such as the central metabolism indicates to us that there may be more efficient ways of using the biological information contained in genome scale models in dynamic settings. In the remainder of this thesis, we explore the following two questions. First, we ask whether Palsson-style FBA predictions of

rates and of yields may be made by stoichiometric models that have been substantially reduced in dimension relative to a parental genome scale model. Second, we ask whether the Palsson-style FBA method may be reformulated in such a way that rates are predictable in a way decoupled from yields. By demonstrating that these two questions have affirmative answers, we will arrive at a novel formulation of dFBA the combines the growth rate predicting capabilities of the Palsson-style FBA method with the flexibility of Fell FBA.

CHAPTER 6: Experimental *E.coli* growth curves

Abstract:

Up to this point, we have relied on computer simulated outputs to describe E.coli growth yields and growth rates on the industrially relevant pentose sugar, xylose. In this chapter, we investigate the properties of *E.coli* growth on this sugar experimentally. Using transporter deleted strains required to consume xylose using different combinations of ABC transport and proton symport, we show that differing growth rates and growth yields (defined for this chapter alone as the maximal OD_{650}) result from transporter deletions. Comparatively analyzing these differences, we provide evidence that the wild type growth profile is supported by a combination of symport and ABC transport under the conditions investigated here. Finally, we show that, under shake-flask batch growth conditions, wild type xylose growth is a highly acetogenic process.

6.1: Introduction

Investigations into the transport by *Escherichia coli* of D-xylose (xylose), extend back from the present by roughly four decades and a dual-transporter model (xylose transport by an ABC-transporter and a proton-symporter) for the acquisition of this sugar had been established by 1989 (Sumiya and Jf 1989). The periplasmic substrate-binding domain of the xylose ABC transporter was first purified and biochemically characterized by Ahlem et al in 1981; (Ahlem, Huisman et al. 1982). A year prior, Lam et al. reported observations of increasing alkalinity of growth media of xylose-consuming *E.coli* cells together with the diminishing of the effect with the addition of uncoupling agents which led them to suspect the presence of a proton-linked xylose transport system; the existence of such a system was subsequently confirmed and the transporter responsible (XylE) was sequenced in the work of Davis and Henderson (Lam, Daruwalla et al. 1980, Davis and Henderson 1987). By the end of the decade following the publication of these works (and also building on prior works concerning xylose metabolic enzymes) a body of literature has been published about the components required for genetic induction of xylose transport and metabolism in the presence of environmental xylose,

specifically by the transcription factor XylR (Song and Park 1997). As with the closely related and by far better characterized system for *E.coli* arabinose transport and metabolism, it is fair to interpret the early goals underlying investigations into the bacterial regulation of xylose utilization as being focused on the characterization of a biochemical and genetic model system from which generalizable paradigms may be drawn and/or from which then-existing biotechnological methods be extended. While such goals appear to have remained largely unchanged to the present day a number of modern trends can be identified in the more recent research concerning *E.coli* xylose utilization; this is the inclusion of efforts towards investigating xylose as a potential source of biofuel in the framework of microbial synthetic biology.

As was discussed in section 1.1 of the introductory chapter, (D-) xylose and (L-) arabinose comprises 95% of hemicellulosic biomass and represent the largest sources of sugar second only to glucose (Kim, Block et al. 2010, Ni, Tonthat et al. 2013). *E.coli* bioconversion of hemicellulose to desired compounds such as ethanol, although promising owing to the possession by this microbe of the necessary chemical pathways, suffers the shortcoming of low product yields. An important hurdle to increased yields is carbon catabolite repression, the underlying causes for which are known to be largely genetic (Desai and Rao 2010). Much effort has gone into understanding carbon-catabolite repressions, with the aim of engineering strains capable of simultaneously utilizing multiple hemicellulose-derived sugars (Desai and Rao 2010, Kim, Block et al. 2010, Hanly and Henson 2011, Groff, Benke et al. 2012). Engineering studies of xylose metabolism have revealed several important properties of the underlying system; it has been shown that product (ethanol) yields can be increased, and similarly that repression

hierarchies (from glucose and arabinose) can be disrupted (Tao, Gonzalez et al. 2001, Kim, Block et al. 2010, Groff, Benke et al. 2012).

Clearly, a large volume of effort has gone into the experimental analysis of xylose metabolism in complex media. What we find however, is that effort is still needed in interrogating the basic properties of *E.coli* growth on xylose as a single substrate in a simple genetic background and in simple media. In this chapter, we describe the differences that are observed experimentally amongst growth phenotypes (growth rates and growth yields) of transporter-deleted *E.coli* mutants growing aerobically on xylose. Although this is mainly a theoretical thesis, I have had the opportunity to undertake some work in the laboratory from which the data described in the following experimental sections have been acquired with the help of an experimental colleague, Konstantinos Drousiotis. Also included in these sections are unpublished results acquired by another colleague, Henrique Neves; namely, these are plate reader data that will now be described.

6.2: Comparison of growth phenotypes of wild type and transport deletion mutants

In this section, we investigate the aerobic transporter utilization by *E.coli*. Comparative batch growth data were generated for WT and transporter-deleted mutant strains. The aerobic growth on xylose of WT *E.coli* K-12 was compared with growth of mutant strains lacking one or both of the genes coding for the respective xylose transporters. Growth data was generated for each strain in triplicate via plate-reader and was repeated with initial xylose concentrations of 2.5 $\text{mmol} \cdot \text{L}^{-1}$, 5.0 $\text{mmol} \cdot \text{L}^{-1}$, 10.0

mmol · L^{-1}, and 20.0 mmol · L^{-1}. Measurements of OD_{650} (OD) were taken in half-hourly intervals for 48 hours. In order to identify kinetic differences occurring within and across the growth profiles, the specific growth rates of the growing populations were approximated where the approximating formula is provided in the experimental methods section of this chapter. These results are presented in figures 6.2.1-6.2.4. In these figures, the final eight hours of growth data have been truncated as we have found that there were no significant events occurring in the extended death phases of the populations.

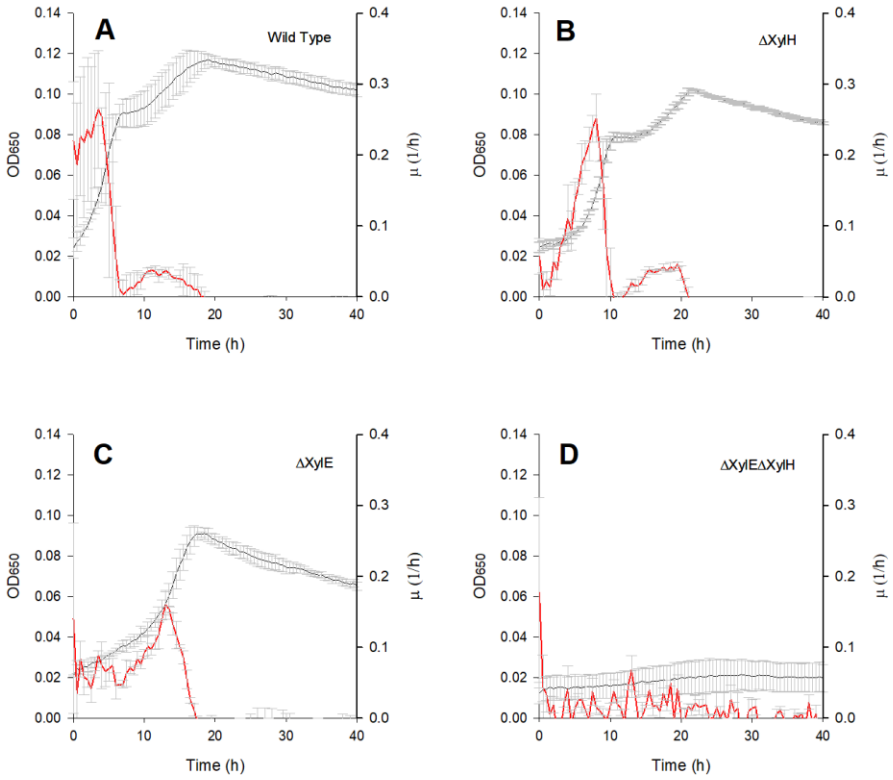


Figure 6.2.1: The aerobic plate reader growths for wild-type (WT) and transporter deleted strains of *E.coli* with an initial xylose concentration of 2.5 mmol · L^{-1} over 40 hours are plotted in black. The specific growth rates (μ) for each strain is plotted in red. Panels A, B, C, and D correspond respectively to WT, Δ*xylH*, Δ*xylE*, and Δ*xylE*Δ*xylH* curves. The growth data were generated in triplicate biological repetitions.

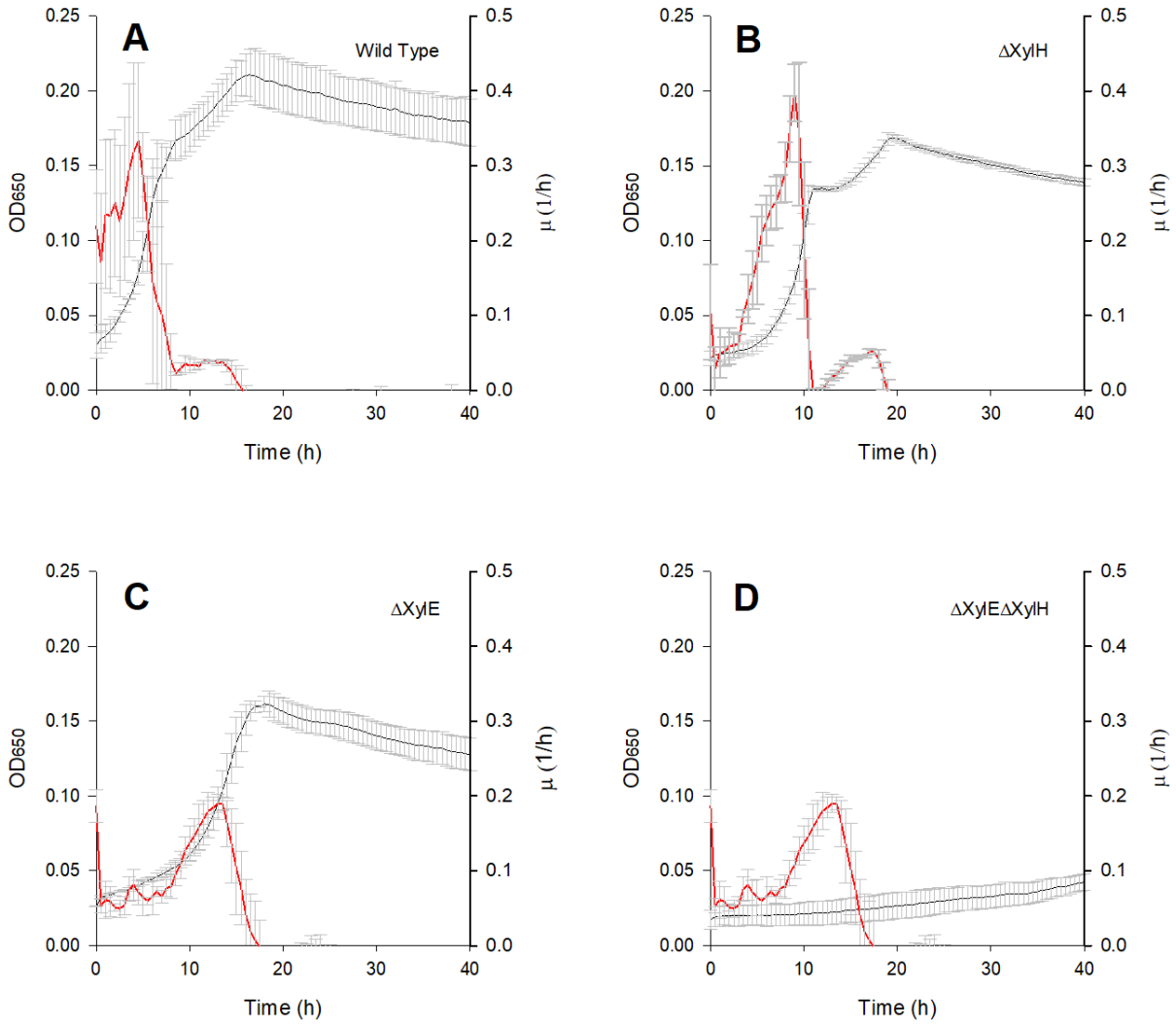


Figure 6.2.2: The aerobic plate reader growths for wild-type (WT) and transporter deleted strains of *E.coli* with an initial xylose concentration of 5.0 mmol · L^{-1} over 40 hours are plotted in black. The specific growth rates (μ) for each strain is plotted in red. Panels A, B, C, and D correspond respectively to WT, Δ*xylH*, Δ*xylE*, and Δ*xylE*Δ*xylH* curves. The growth data were generated in triplicate biological repetitions.

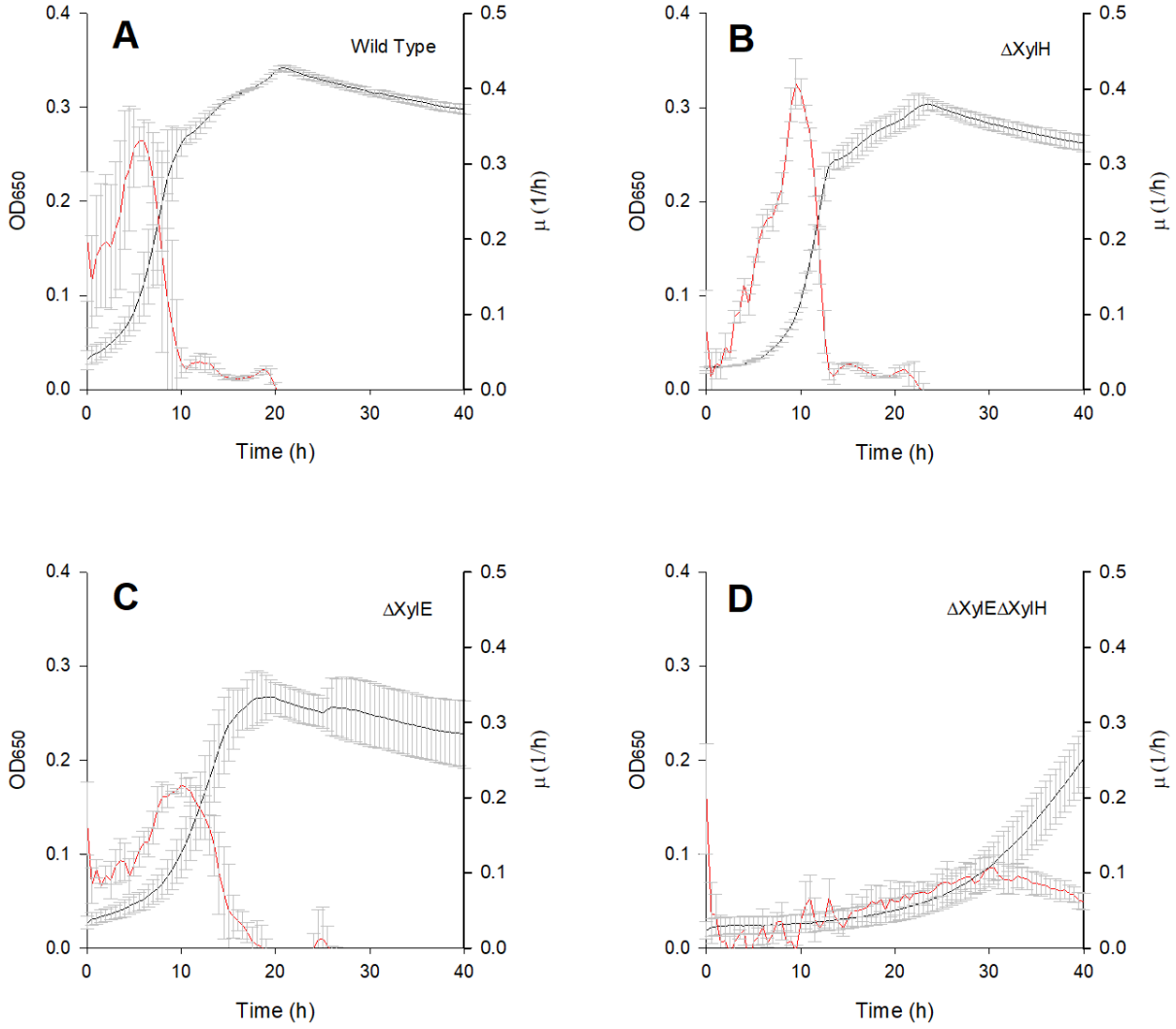


Figure 6.2.3: The aerobic plate reader growths for wild-type (WT) and transporter deleted strains of *E.coli* with an initial xylose concentration of 10.0 mmol · L^{-1} over 40 hours are plotted in black. The specific growth rates (μ) for each strain is plotted in red. Panels A, B, C, and D correspond respectively to WT, Δ*xylH*, Δ*xylE*, and Δ*xylE*Δ*xylH* curves. The growth data were generated in triplicate biological repetitions.

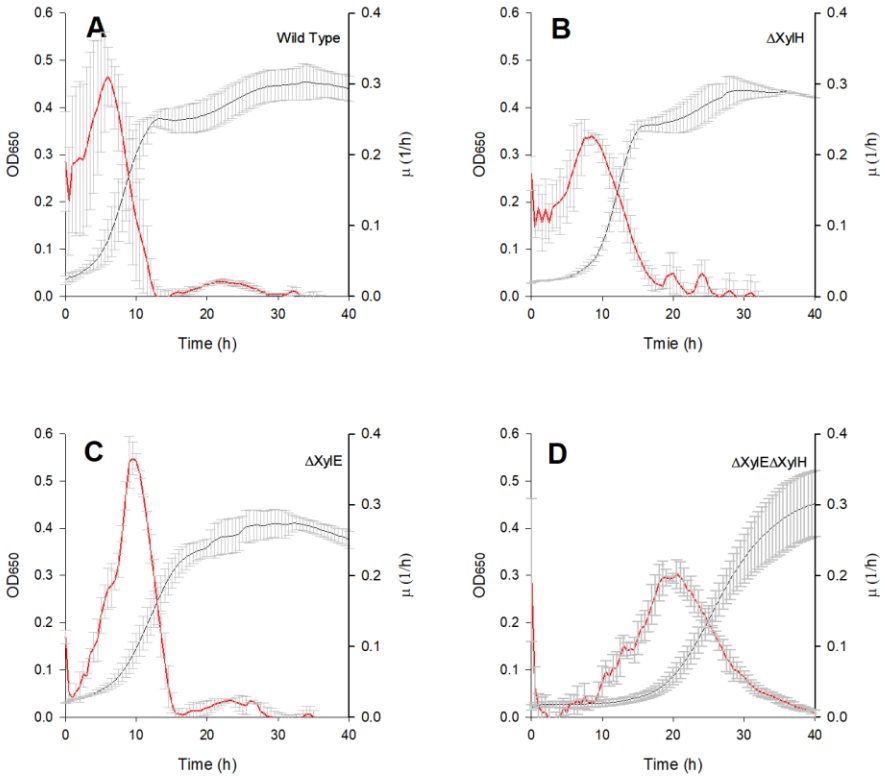


Figure 6.2.4: The aerobic plate reader growths for wild-type (WT) and transporter deleted strains of *E.coli* with an initial xylose concentration of 20.0 mmol · L^{-1} over 40 hours are plotted in black. The specific growth rates (μ) for each strain is plotted in red. Panels A, B, C, and D correspond respectively to WT, Δ*xylH*, Δ*xylE*, and Δ*xylE*Δ*xylH* curves. The growth data were generated in triplicate biological repetitions.

Growth was observed for the WT, Δ*xylE* mutant, and the Δ*xylH* mutant at all initial concentrations of xylose tested. The doubly transporter-deleted mutant only grew appreciably at 10 mmol · L^{-1} and 20 mmol · L^{-1} initial concentrations of xylose. The ability of the double mutant to grow at higher substrate concentrations demonstrates that means of xylose transport other than transport via the known xylose specific transporters exist – possibly this is attributable to promiscuous transporters or to membrane diffusion. The low-affinity kinetics associated with alternative processes is evidenced by the

inability of the double mutant to grow at 2.5 $\text{mmol} \cdot \text{L}^{-1}$ and 5.0 $\text{mmol} \cdot \text{L}^{-1}$ concentrations of xylose; at the same time it is noteworthy just how well growth can be supported by such non-xylose-specific processes at moderately higher xylose concentrations of 10 $\text{mmol} \cdot \text{L}^{-1}$ and 20 $\text{mmol} \cdot \text{L}^{-1}$. These concentrations are a full order of magnitude lower than a typical sugar concentration used for industrial-scale fermentations which may be as high as 400 $\text{mmol} \cdot \text{L}^{-1}$ (Gavin Thomas; personal communication).

The growth profiles of the singly-deleted transporter mutants are most prominently distinguishable by the presence and absence respectively in the Δ*xylH* and Δ*xyl*E growth of a diauxie. By exhibiting a diauxic shift the Δ*xylH* growth curve appears to more closely resemble the WT growth curve than does the Δ*xylE* curve. An inspection of the specific growth rate profiles at all concentrations reveals that the Δ*xylH* mutant is also kinetically similar to the WT. That the mutant whose growth is supported primarily by proton-symport exhibits a growth pattern more comparable to that of the WT relative to the ABC-deleted mutant is consistent with a hypothesis that a low-affinity transport process is used primarily for xylose acquisition by the WT. If however the XylE system is the most important, then why is a symport-deleted mutant capable of growing at all? At initial xylose concentrations of 10 $\text{mmol} \cdot \text{L}^{-1}$ and 20 $\text{mmol} \cdot \text{L}^{-1}$growth would be supportable in the absence of xylose transporters as was observed for the double mutant. However the symport deleted strain is distinguishable from the double mutant in its ability to grow well also at 2.5 mM and 5.0 mM initial xylose concentrations where the latter cannot. Therefore it must be concluded that in the absence of the xylose symporter, the xylose ABC complex may be expressed and that catalysis by this transporter is

sufficient to produce growth albeit with a lower rate and a lower biomass yield relative to the WT. Is the ABC transporter operative in the WT under the conditions here tested and the symporter-preference therefore non-exclusive? These observations raise the possibility that the purpose served by the ABC transporter may be more involved than simple scavenging at low (~ 10^{-3}mM) substrate concentrations as might be logically hypothesized from Km considerations.

6.3: Wild type growth on xylose in shake-flask is diauxic and acetogenic

In this section, we investigate the substrate consumption and by-product production rates of WT *E.coli*. As we required larger sample volumes than was possible with the plate reader for substrate quantifying assays, the experimental setting was switched from the plate-reader to shake-flask (SF) growth.

WT growth in SF was observed to mirror qualitatively the corresponding growths in the plate reader (Figure 6.3.1).

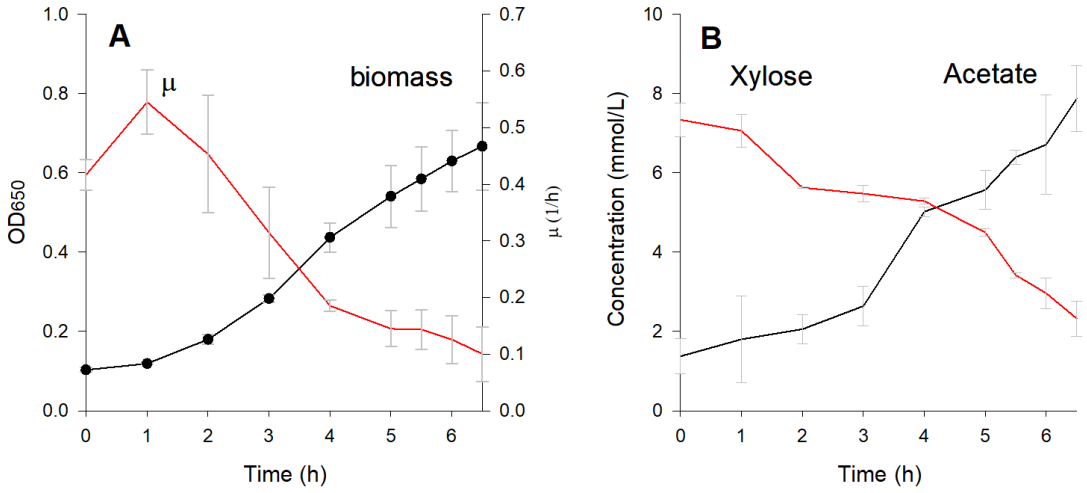


Figure 6.3.1: A) Experimentally measured time course of an aerobic growth of *E.coli* performed in triplicate (black) with approximations to μ for each partition considered (red). B) Depletion of substrate (xylose; black), and accumulation of by-product accompanying growth (acetate; red). Data have been generated in triplicate repeats.

The growth curve exhibited an approximately sigmoidal form saturating about a maximal average OD_{600} of 0.82 starting from an initial value of 0.10; a diauxic shift was observed two hours after the maximal specific growth rate was recorded. In panel A of figure 6.3.1, this is apparent from the sudden widening of the error bars past the hour-two time point. The specific growth rate varied during this growth over a range spanning 0.11 h^{-1} to 0.58 h^{-1}, with the maximal value being achieved three hours post-inoculation where the OD_{600} readings increase exponentially (Figure 6.3.1.A). Xylose is shown to be depleted immediately prior to the ninth hourly time point corresponding to the time of entry of the OD_{600} curve into the stationary phase; this is five hours past the onset of the diauxie. Despite the high aerobicity of the growth cycle, a large amount of acetate secretion into the reaction volume by the cells was recorded (Figure 6.3.1.B). Relative to

their plate reader counterparts, both the biomass yield and growth rate were measured to greater values during SF growth.

6.4: Discussion

The purpose of the growth experiments described in this chapter was to identify any discernible changes to *E.coli* growth phenotype occurring as a result of xylose transporter deletions. Wild-type (WT) *E.coli* growth was shown to exhibit a diauxic shift for which the underlying mechanism remains to be investigated. In particular, out SF data strongly suggests that what we observe is not an acetate diauxie. Exponential growth kinetics were observed prior to the diauxie but the maximal specific growth rates were seen to be short-lived. Singly and doubly deleted xylose transport mutants were seen to exhibit growth phenotypes non-identical relative to one another and to the WT. It is therefore likely that the completion of the WT growth is dependent upon the use of both the symport and the ABC transport of xylose. Based upon kinetic properties of the growth phenotypes, it is seen likely that the primary transport process supporting WT growth is symport-based.

The ability of *E.coli* to transport xylose by concentration-dependent non-xylose specific processes has been also demonstrated. We find that the contribution of alternative xylose transport processes to net xylose acquisition appears to be quite significant at concentrations of 10 $\mathrm{mmol} \cdot \mathrm{L}^{-1}$ and higher; this may explain the slightly diminished distinguishability amongst growth curves of genetically differing strains at the higher (10 $\mathrm{mmol} \cdot \mathrm{L}^{-1}$ and 20 $\mathrm{mmol} \cdot \mathrm{L}^{-1}$) xylose concentrations considered. These observations are quite relevant to industrial biotechnological applications in which *E.coli*

cells may be exposed to unnaturally high conditions towards which their genetic systems have not yet been adapted.

Interestingly, it has also be seen that higher final biomass concentration correspond to higher lower maximal specific growth rates under certain conditions. At $10.0\ \mathrm{mmol} \cdot \mathrm{L}^{-1}$ and $5.0\ \mathrm{mmol} \cdot \mathrm{L}^{-1}$ concentrations, it can be seen that WT cell achieve higher biomass concentrations with lower maximal growth rates compared to their closest counterpart, the Δ*xylH* mutant growing via xylose symport. We may not however draw a conclusion that this is a transport-energy dependent result. Under the same conditions, we find that the symport-deleted mutant growing via ABC transport achieves a final biomass concentration that is comparable to the Δ*xylH* mutant with a much reduced growth rate. Similarly, it is seen at $20\ \mathrm{mmol} \cdot \mathrm{L}^{-1}$concentrations that three strains achieve a comparable final biomass concentration yet with very different growth rates. It is concluded that the regulation of xylose transporter expression, as well as the functional relation amongst xylose transport, growth rates, and growth yields, is a complex property of the *E.coli* metabolic system that warrants deeper investigation.

6.5: Experimental methods

Strain	Description	Source
BW25113	Wild-type; parent to all Keio strains	(Baba, Ara et al. 2006)
BW*ΔxylE*	Symport deletion	D.B
BW*ΔxylH*	Deletion of membrane component of the xylose ABC transporter; XylR independent promotor missing	D.B
BW*ΔxylEΔxylH*	As in BW*ΔxylE* and BW*ΔxylH*	D.B
BW*ΔxylF*	Deletion of the periplasmic substrate binding domain of	D.B
	the xylose ABC transporter	
BW*ΔxylG*	Deletion of the cytoplasmic ATP binding domain of the xylose ABC transporter	D.B

Table 6.5.1: Table of strains used

Table 6.5.1 lists all of the *Escherichia coli* (*E.coli*) strains used for this study. In all experiments, the strains were grown in M9 minimal media with xylose and $MgSO_4$ added. The composition of M9 is as follows: 6 grams Na_2HPO_4, 3 grams KH_2PO_4, 0.5 grams NaCl, and 1.0 gram NH_4Cl dissolved in 1L of dH_2O at a PH of 7.0.

The growth experiments described below were carried out by Konstantinos Drousiotis (K.D) and Kazuki Iizuka (K.I), and by Henrique Iglesias Neves (H.N) as indicate. The xylose transporter deleted strains have been obtained from Dr. Daniel Bowden (D.B).

Growth of *Escherichia coli* strains in the plate reader (H.N)

Wild type and mutant *E.coli* were grown overnight in M9 minimal media (1mL M9, 1mM $MgSO_4$, and 20mM glucose) at 37 degrees Celsius. Cells from the culture were

then washed in 15mL of M9 by twice centrifuging at 4000 rpm for 10 minutes. Inocula were set up either in triplicate or quadruplicate on a 96 well plate at an initial OD_{650} reading 0.1. Initial xylose concentrations were set at 2.5mM, 5.0mM, 10.0mM, or 20.0mM. Growth was continued for a 48 hour period with optical density readings recorded every half hour.

Growth of *Escherichia coli* strains in shake flask and sampling (K.D; K.I)

Wild type (WT) *E.coli* were grown overnight in M9 minimal media (1mL M9, 1mM $MgSO_4$, and 20mM xylose) at 37 degrees Celsius. Cells from the culture were then washed in 15mL of M9 by twice centrifuging at 4000 rpm for 10 minutes. 50mL M9 solutions were prepared for triplicate growth assays and were inoculated to an OD_{600} reading of 0.1; xylose was provided as the sole carbon source starting with an initial concentration of 5mM. The shake-flask growth was conducted identically as for the overnight culture. Cells were grown to stationary phase with OD_{600} readings taken in hourly or half-hourly intervals. For subsequent substrate and by-product assays, samples were collected at every OD_{600} reading via centrifugation for 10 minutes and storing the resulting supernatant at -80 degrees Celsius.

Measurement of acetate concentration (KD; KI)

The acetate assay was carried out using the Megazyme acetic acid assay kit following manual procedure. This assay is based on the following. Acetate is converted to acetyl-CoA in a two-step process catalysed by the enzymes acetate kinase and phosphotransacetylase. The ADP that is generated in the first step (the kinase step) is re-phosphorylated to ATP in the presence of phosphoenolpyruvate and the gluconeogenic enzyme pyruvate kinase. As in gluconeogenesis the pyruvate kinase reaction produces

pyruvate which may in turn be reduced to lactic acid by the enzyme lactate dehydrogenase. The conversion of pyruvate to lactic acid involves the oxidation of NADH to NAD+ in a 1:1 stoichiometry with the conversion of acetate to acetyl-CoA. The Acetate assay therefore measures acetate concentration by measuring the NADH oxidation which in turn is indicated by a decrease in absorbance at $\lambda = 340$ nm. Assays were performed on samples taken from the shake-flask growth with an assay-volume of 2.66mL and at 25 degrees Celsius.

Measurement of xylose concentration (KD; KI)

The xylose assay was carried out using the Megazyme D-xylose assay kit following manual procedure. This assay is based on the following. Alpha-xylose furanose is converted to its beta-form isomer by the enzyme xylose mutarotase. The resulting beta-xylose furanose donates electrons to NAD+ in a reaction catalysed by beta-xylose dehydrogenase. Therefore, as was the case for the acetate assay, the change to NADH concentration measured as change (increase) in absorbance at λ=340 nm is used to measure xylose concentration. . Assays were performed on samples taken from the shake-flask growth with an assay-volume of 2.97mL and at 25 degrees Celsius.

6.6: Computational methods

Calculation of experimental specific growth rates and growth statistics

In calculating the specific growth rates of our growth curves, moving averages were used to more clearly observe the trends in the data. For each time interval of growth, the increase in OD_{650} readings observed over the interval was divided by the length of the time interval to approximate the total population growth rate. The resulting value was subsequently divided by

the OD_{650} reading for the left-end time point to derive a specific growth rate. This calculation was repeated for each interval, and a specific growth rate calculation was assigned to each time point with the exception of the final time point. A three-point smoothing procedure was applied to the resulting sequence of specific growth rates. For each time point, the corresponding rate was averaged with two succeeding terms in the sequence. Repeating this calculation over the sequence, we were able to calculate a specific growth rate value for each time point with the exception of the final two. Because a time point has already been lost in deriving the growth rate sequence, a total of three final time points were lost in our calculations.

All error-bars appearing in our data were generated using the sample standard deviation formula

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}, \qquad \textbf{(6.6.1)}$$

where N is the sample size, s is the standard deviation, x_i is a data value, and $\bar{x}$ is the sample mean. For our dataset, N = 3 because our experiments were performed in triplicate. Error bars are extended one standard deviation above and one standard deviation below an averaged data point. All statistical calculations were performed using the software R.

Flux Balance Analysis and Dynamic Flux Balance Analysis

The computation of FBA was performed using the GLPK (GNU Linear Programming Kit) software for linear programming. Structural metabolic models were represented using standard Excel files. Custom code was written in Java to translate this information, to carry out the optimization using GLPK, and to output this information as text files. All simulations described in this thesis were performed by codes running on a standard Windows hp home

computer. The visualization of all data included in this thesis, not just FBA outputs, were done using SigmaPlot.

In implementing the dFBA algorithm described in section 2.10, the Java routines written for FBA were iterated over a specified time interval. In all simulations, FBA solutions and ODE solutions were calculated alternatively every 0.3 seconds of model time. In calculating integrals numerically, ODE rates were solved, and the results were used to approximate net changes in concentrations using Newton's two-point approximation formula for derivatives.

Chapter 7: A novel approach to dFBA that accounts for dynamic biomass composition

Abstract:

A careful consideration of dynamic flux balance analysis (dFBA) formulations is a prerequisite for the construction of both better dFBA models and of better microbial community models. We suggest in this chapter that this improvement will require an investigation into how dFBA models should treat the boundaries between its dynamic and static components, as well as developing new methods where internal metabolic information feedback can be accommodated. In this chapter, we demonstrate how both these issues may be resolved by combining the Fell and Palsson approaches to flux balance analysis (FBA), and thus propose a new formulation of the dFBA method. We illustrate this "hybrid" dFBA approach, by implementing it using a small stoichiometric network that we parametrized against the boundary conditions of the iJR904 *E.coli* metabolic model. Through this implementation, we demonstrate how internal metabolites, potentially deeply located within a metabolic network, may be treated dynamically in a dFBA model without severely increasing the dimensionality of the differential equations describing the external environment. We extend this result by demonstrating how regulatory effects of deeply located metabolites, whether it be allosteric or genetic effects, can be represented by a dFBA model; thus, we demonstrate how dynamics of information transfer can be accommodated by a dFBA model.

7.1: Introduction

The ability to alter, and to be altered by the environment in turn, is a universal feature of biological systems that is shared across all branches of life. Understanding the nature of this feedback process is prerequisite, not just to understanding biology at the level of systems, but also to rationally engineering novel and useful cellular phenotypes through genetic and metabolomic means. For microbial communities, the interaction that takes place between one organism and its environment is often also an indirect interaction amongst several organisms sharing a common spatial domain. A very natural way of modelling such community-scale metabolic interactions is by coupling individual metabolic models through environmental variables, such as concentrations of substrates and metabolic by-products. Such models can, at least structurally, represent the complex interdependences that are known to exist between microorganisms that make up real communities (Stolyar, Van Dien et al. 2007, Chaganti, Kim et

al. 2011, Khandelwal, Olivier et al. 2013, Henson and Hanly 2014). Apart from a very small number of exceptions, such as communities that are grown in a chemostat, the chemical composition of the environment is a dynamically changing variable on the timescale of microbial growth and replication. The mammalian gut microbiota, for example, is exposed to a range of different growth media over the course of a single day, and similarly, cells cultured in batch will detect varying concentrations of substrates as population growth is progressed to a stationary phase. For this reason it is only for a limited number of cases, or over uninterestingly short timescales, that external metabolite concentrations can be modelled to be in a dynamic steady-state. In all remaining cases, external metabolites must be represented through differential equations that allow for their concentrations (and so the environment) to change in response to the metabolic behaviour of individual community members. The modelling framework within which steady-state metabolic models are coupled to differential equations is dynamic flux-balance analysis (dFBA). Therefore, the general problem of modelling the metabolism of microbial communities should be considered as a subset of dFBA modelling. By extension then, it becomes necessary to study the properties of dFBA models prior to exploring the full scope and potential of community flux-balance analysis.

The need to consider environmental dynamics in community FBA models brings to light a fundamental question applicable to all dFBA modelling efforts. What is the most appropriate method by which to couple individual metabolic systems? Each system will be modelled under the assumptions of steady state and an optimality criterion, but also incorporating appropriately chosen kinetic ODEs. What are the conditions required so that the final representation is a valid and useful systems biology model of the global complexity that emerges from local interactions? We argue that this is a problem of identifying the correct number, the types, and the placements

of feedback mechanisms between the individual dFBA components. The issue of which dFBA variables to treat dynamically via ODEs, and likewise, which variables to treat statically via FBA is context-dependent and decisions are usually left to the discretion of the investigator. Any choice of a partition of the variables will rest upon both convenience (what can feasibly be measured experimentally) and experimenter interest (what aspects of a system is being investigated). Nevertheless, there are several variables that are common across most metabolic systems under study that deserve special attention. In chapter 4, we have derived a mathematical relation (equations 4.4.6 and 4.4.7) between two key FBA predictions, growth yield and growth rate, both of which are fundamental to the dynamic modelling of microbial systems. The relation was shown to depend very strongly on the assumed biomass composition as well as the capability of central metabolism to meet steady-state energy demands; the consequences of this dependence to FBA predictions have been reported in chapter 5. The results we presented in these chapters lead us to conclude that, in addition to the chemical and biological composition of the environment, biomass composition and maintenance energies are variables that require further attention. Although we focus specifically with the biomass composition in this chapter, the carbon demand that is incurred dynamically on the FBA objective by the biomass reaction is readily recognized to be discussible analogously in terms of dynamic maintenance energies with the biomass composition held static; this is because of the stoichiometric interconvertibility of carbon and energy that has been emphasized throughout the previous chapters. An important point to recognize however, is that a very significant hurdle is introduced to the problem of integrating dFBA models when *both* biomass composition and maintenance energies are treated as dynamic variables. The simultaneous dynamic treatment of composition and maintenance is a significant problem on its own and the issue is therefore left to future studies.

7.2: Is the composition of biomass static?

If the biomass composition under investigation remains static over a timeframe and environmental range of interest, it may be either measured directly once (the Fell approach) or extrapolated from previously obtained data (the Palsson Approach). In both cases, the biomass composition may be treated as a part of an FBA model, for instance though not necessarily as a biomass objective. Thus, the only complication to dFBA that arises from a static biomass composition is the feasibility of deriving precise stoichiometric coefficients. Moreover, it is seen that any errors in the biomass composition that is used may be linearly compensated for in FBA by using phenomenological fitting parameters such as GAM (e.g. equations 4.2.3 and 4.4.2). However, for those cases where the biomass composition does change over a relevant range of growth conditions, then the question of how to represent these dynamics poses theoretical challenges to the modelling process.

There are several experimental observations that have been made to date that suggest the biomass composition, like environmental composition, changes over growth phases and over growth conditions, and so needs to be included in dFBA as a dynamic variable. Huang et al. showed that in the microalgae, *Chlorella sorokiniana*, the total lipid content of biomass differs depending upon whether the substrate of growth is acetate, glucose, or a combination of the two; similarly, when grown on a combination of glucose and acetate, the starch content of the biomass was shown to change over the course of batch growth (Huang, Huang et al. 2013). Likewise, Di Pasqua et al. provide evidence that *Escherichia coli* O157:H7, *Salmonella enterica* serovar typhimurium, and *Brochothrix thermosphacta*, all change their membrane fatty acid profiles in response to stress induced by sublethal concentrations of antimicrobial compounds in the growth

media; oppositely, it was found that the composition does not change in *Pseudomonas fluorescens* and *Staphylococcus aureus* (Di Pasqua, Hoskins et al. 2006). Godin et al. have shown that the buoyant mass of individual cells of *Bacillus subtilis*, *Escherichia coli*, *Saccharomyces cerevisiae*, and mouse lymphoblasts, increase with replication rate (Godin, Delgado et al. 2010). As a final example, Yamamotoya et al. have shown, for batch cultures of *E.coli* K-12 strains growing aerobically on glucose in MOPS medium, that the average measured intracellular concentrations of glycolytic metabolites (3-phosphoglycerate, glucose-6-phosphate, and phosphoenolpyruvate) as well as of glycogen can change over a range of several orders of magnitudes throughout the entire duration of growth from the lag phase to the stationary phase (Yamamotoya, Dose et al. 2012). Considering such data, we are led to conjecture, by way of exercising biological intuition, that perhaps a dynamically changing biomass composition is the norm rather than the exception. After all, is it truly reasonable to expect biological systems and phenomena of interest to exist or to occur predominantly under steady-state? By extension, how applicable to real systems are "simple dFBA" models that do not account for such changes in a realistic way?

Despite the many reasons to believe that the biomass needs to be treated as changing, we have found that very few published dFBA models consider this problem (Hanly and Henson 2011). In contrast, some dFBA authors have considered altering the biomass composition during growth, and reported improved dFBA fits to experimental growth data. Notably, Pramanik and Keasling used the observation, that fractional compositions of the various compounds that make up *E.coli* biomass change with growth rate, to derive statistical linear regressions between growth rate and concentrations of various molecules that comprise this organism; these regressions were then used to define biomass objectives with growth-rate-dependent coefficients

(Pramanik and Keasling 1997). In one application, the approach of Pramanik and Keasling was used to achieve qualitatively improved dFBA fits to aerobic growth data of *E.coli* growing on glucose and acetate (Meadows, Karnik et al. 2010).

Whilst using regressors against specific growth rate may lead to qualitatively acceptable fits of dFBA predictions to growth data, we would argue that this is still a static representation of biomass composition. In accounting for a changing composition by simply altering the coefficients of a biomass reaction, a dFBA model does not treat the metabolic cell constituents as dynamic concentrations; that is, as signals of feedback regulation. As a result, the feedback mechanisms from alterations to enzyme activity by substrates in pathways that transiently exhibit non-steady-state concentration changes are neglected. There is evidence that metabolites of even well studied pathways such as glycolysis exhibit non-steady-state concentration dynamics under certain growth conditions (Yamamotoya, Dose et al. 2012). The need for kinetic models to account for "local feedback" have already been demonstrated for fully kinetic models (Cornish-Bowden and Cárdenas 2001).

A second reason we find that the above treatment of biomass composition is a static representation is because metabolites such as those of central metabolism are not terminal products of pathways. Rather, they are precursors to anabolism or energy-sources. Therefore, the potential use of these metabolites as future growth assets is neglected when they are consumed in a model, for instance, by a biomass reaction. Thus, it is unclear how in dFBA to treat intracellular metabolites that, although contribute by mass to a dynamic biomass composition, do not contribute a service to an instantaneous and static cellular objective. Some dFBA authors have recently considered this problem. For example, Upton et al vary the

composition of phosphate products and glycogen storage in *aspergillus niger* to take into account the aggressive phosphate scavenging strategy of *A. niger* (Upton, McQueen-Mason et al. 2017).

We will propose solutions to overcome these hurdles to dFBA modelling. First however, we demonstrate using the ideas introduced in this section that the rigid correspondence between growth rate, growth yield, and energy flux in Palsson-style FBA models (chapters 4 and 5) leads to dFBA models that predict unrealistically rigid growth curves. We begin by introducing the notational conventions we will use for our dFBA models.

7.3: Notational convention for kinetic parameters

We will now describe the notational conventions that we will follow in describing our kinetic equations. The rate of most enzyme catalysed reactions will be described with at least three parameters. Representing the turnover rate, the half-maximal rate, and the product inhibition concentration, these parameters are respectively given the base notations k_{Cat}, K_m, and K_i as standard in biochemistry. For the most part, these values are parameters associated with a catalyst and not a substrate (each dynamically treated catalyst is modelled to have a single substrate and a single product). Therefore, in discussing these parameter values of a specific catalyst, the above three notations will be superscripted with the name of the catalyst. The notation K_i^{XylE} is therefore the product inhibition parameter for the XylE transporter; the fact that the inhibiting metabolite is xylose is understood as internal xylose is the only product of xylose transport (we have ignored any effects from protons). In later sections (7.11 and 7.12), we consider possibilities where a catalyst may be inhibited by metabolites other than its own product. Thus, inhibition constants other than those for product inhibition must be qualified with the inhibiting metabolite. To indicate this metabolite, we will follow the "i" in K_i with the name

of the metabolite. Thus, when a compound comp inhibits Rxn but is not a product of Rxn, the inhibition parameter for this interaction will be denoted by K_{icomp}^{Rxn}.

The next section deals with metabolites and enzymes that occur in real biological systems. However, we will later work with a hypothetical network model as a way of illustrating our ideas in the absence of the topological complexities of real metabolic networks. To facilitate the book keeping of metabolite and catalyst (and reaction) names there, we will assign to all hypothetical metabolites, a lower-case letter from the English alphabet. Catalysts of that network will then be named by sandwiching a "TO" between the sequence of its substrate and its product names written in alphabetical order. Thus, the catalyst for the reaction

$$a \rightarrow b + e, \qquad \textbf{(7.3.1)}$$

will be given the name "aTObe," and similarly for the reaction itself.

Introduced in section 7.6 are "indicator metabolites," which we use later on as a simple way of writing FBA constraints as functions of dynamic metabolite concentrations. Indicator metabolites are purely computational, but all indicator metabolites will have one real counterpart metabolite in the dFBA model. Thus, indicator metabolites will be named by prefixing the name of the associated metabolite with a capital "I." Constraints on exchange reactions involving indicator metabolites will be written as simple irreversible Michaelis-Menten equations with a maximal velocity parameter (V_{max}) and a half-maximal parameter (K_m). Because these parameters are associated with the metabolite, and not with any catalyst, we will qualify them by subscripting the base notation with the name of the indicator metabolite. The parameter V_{max} will be used in dFBA, as opposed to k_{Cat}, generally when the concentration of a catalyst cannot be represented as a dynamic variable. One such case occurs when an effector, but not the

catalyst, is a dynamic variable. In this case, we will qualify the V_{max} notation by indicating the name of the affected catalyst as a superscript. Note that there is a notational difference between doing this and what we have done for indicator metabolites; one denotes a parameter of a single reaction, whereas the other denotes a parameter of a single metabolite.

Less conventionally than for the kinetic parameters we have discussed so far, we will use the Greek symbols, β and κ, to describe the kinetics of gene expression. When an expression rate is described as an order-1 Hill function of a signalling metabolite or a transcription factor concentration, β will be used to represent the maximal rate of gene expression that occurs at saturation; in the same setting, κ will be used to represent the concentration of the signal at which expression rate is $\frac{1}{2}\beta$. These parameters will be specific to a protein whose expression is under consideration, and thus will be qualified using the protein name as a subscript. To describe the constant production rate of a protein due to basal gene expression, we will use the same conventions described here except that the notation β^0 will be used as the base notation for the constant rate. The parameter κ may be thought of as always having a value of zero for basal expression; therefore we will not indicate this explicitly in our equations.

Finally, we introduce two parameters that do not fall within the conventional scope of this section. First is the non-unit Hill coefficient that appears in one equation in section 7.11. We will use the notation n_{comp} to indicate the value of this coefficient where comp is the signalling metabolite. Second, we will use the notation k_{Diff}^{xyl} to indicate the diffusion coefficient of xylose across the inner cell membrane. With the notational convention having been described, we will now introduce a simple dFBA model of *E.coli* xylose metabolism.

7.4: A simple xylose dFBA model

The following system of differential equations was used to construct a simple dFBA model of aerobic *E.coli* growth on xylose.

$$[\mathrm{R2ss}] = \frac{[\mathrm{XylR}][\mathrm{xyl}]^2}{\left(\mathrm{k_d^{XylR}}\right)^2} \tag{7.4.1}$$

$$\frac{\mathrm{d[biomass]}}{\mathrm{dt}} = \frac{\mu[\mathrm{biomass}]}{3600} \tag{7.4.2}$$

$$\frac{\mathrm{d[acOut]}}{\mathrm{dt}} = \mathrm{v_{EX\_ac}}[\mathrm{biomass}]\left(\frac{2.8}{36.0}\right)\left(\frac{10^{-15}}{1.0}\right)\left(\frac{10^{-12}}{1.0}\right) \tag{7.4.3}$$

$$\frac{\mathrm{d[xylOut]}}{\mathrm{dt}} = (-1.0)\left[\mathrm{k_{Diff}^{xyl}}([\mathrm{xylOut}] - [\mathrm{xyl}]) + \frac{\mathrm{k_{Cat}^{XylE}}[\mathrm{XylE}][\mathrm{xylOut}]}{\mathrm{k_m^{XylE}}\left(1 + \frac{[\mathrm{xyl}]}{\mathrm{k_i^{XylE}}}\right) + [\mathrm{xylOut}]}\right]\left(\frac{10^{12}}{1.0}\right)\left(\frac{10^{-15}}{1.0}\right)[\mathrm{biomass}] \tag{7.4.4}$$

$$\frac{\mathrm{d[xyl]}}{\mathrm{dt}} = \left[\mathrm{k_{Diff}^{xyl}}([\mathrm{xylOut}] - [\mathrm{xyl}]) + \frac{\mathrm{k_{Cat}^{XylE}}[\mathrm{XylE}][\mathrm{xylOut}]}{\mathrm{k_m^{XylE}}\left(1 + \frac{[\mathrm{xyl}]}{\mathrm{k_i^{XylE}}}\right) + [\mathrm{xylOut}]}\right] - \left[\frac{\mathrm{k_{Cat}^{XylA}}[\mathrm{XylA}][\mathrm{xyl}]}{\mathrm{k_m^{XylA}} + [\mathrm{xyl}]}\right] - \frac{\mu[\mathrm{xyl}]}{3600} \tag{7.4.5}$$

$$\frac{d[XylE]}{dt} = \left[\frac{\beta_E[R2ss]}{[R2ss] + \kappa_E}\right] - \frac{\mu[XylE]}{3600} \quad \textbf{(7.4.6)}$$

$$\frac{d[XylA]}{dt} = \left[\frac{\beta_A[R2ss]}{[R2ss] + \kappa_A}\right] - \frac{\mu[XylA]}{3600} \quad \textbf{(7.4.7)}$$

$$\frac{d[XYLR]}{dt} = \beta_R^0 + \left[\frac{\beta_R[R2ss]}{[R2ss] + \kappa_R}\right] - \frac{\mu[XylR]}{3600} \quad \textbf{(7.4.8)}$$

The equation accounts for the upregulation of the proteins involved in the xylose system by the inclusion of the transcription factor (TF) XylR, which in its activated form, is written R2ss. The concentration of R2ss is assumed to achieve a rapid steady-state (ss) value, and is included as an algebraic equation expressing the quadratic dependence of the ss concentration on intracellular xylose. The presence of R2ss is modelled to activate the expression of the proteins XylR, XylE, and XylA via first-order Hill functions of R2ss; respectively, these proteins are the TF, the transporter, and the metabolic enzyme (section 1.4).

It should be noted that this ODE model accounts for only one transporter for xylose. In preliminary analyses of the equations, we have found that it makes very little difference to the overall behaviour of this system whether a second transporter is included. This is a consequence of our not including the mechanism details that distinguish the two transporters biologically in the interest of keeping the ODE simple for its integration into dFBA. Recall from section 1.4 that xylose is transported via an ABC mechanism by the XylFGH complex and via a proton-symport mechanism via the XylE protein. In particular, by assuming that the only biologically

distinguishing feature of the transporters is their K_m values, we have found that, because [xylOut] >> K_m for both the symporter and the ABC transporter, for the majority of the growth, the ss xylose concentration is determined summatively as a function of the ss concentrations of XylE, and XylFGH (correspondingly, by their combined transport rates). Specifically, with comparable values for the turnover parameter, the consumption rate of xylose (neglecting a small contribution from diffusion) remains essentially

$$v_{transport}^{xyl} \approx k_{Cat}^{XylE}([XylE] + [XylFGH])\,[xylOut] \qquad \textbf{(7.4.9)}$$

for most of the growth simulation, with the transporter concentrations driven to ss by a pair of decoupled Hill equations (note that k_{Cat}^{XylFGH} could have been used to substitute k_{Cat}^{XylE} in the above expression with the same effect). We have chosen to represent this dynamics as a single transport process as we find the inclusion of a second transporter to be unnecessary, unless a more complex regulatory model is assumed for their expression. Since we are not interested in this chapter in describing specific growth rates and yields at the resolution of the difference in Palsson-style FBA predictions that arises from using alternative transporters (table 5.6.1), we will assume that only the symport process is used.

The rate of xylose transport is assumed to be quantified as a combination of two processes. The first, is a simple diffusion, the rate of which is determined by the concentration difference between the external and the internal xylose. The flux across the diffusion step will be seen to be trivial in magnitude relative to the flux across catalysed transport, and so will have very little consequences to dFBA predictions. This is because k_{Diff}^{xyl} is assumed to be a small parameter. A diffusion step is required however to activate the xylose system when starting with an initial xylose concentration near 0.0 $mmol \cdot L^{-1}$; that is, the

model must accommodate, in some way, a process by which external xylose may be transported into the cell to activate the TF. An alternative way by which this could have been accomplished is by introducing a basal expression level of XylE, as we have done in the above equations for XylR. The second avenue of xylose transport is through the process catalysed by the transporter XylE. This step has been modelled as an irreversible Michaelis-Menten step with product inhibition. The simplest dFBA model of xylose metabolism we present has no feedback processes other than this inhibition that operate on the metabolic time-scale; the xylose consumption step is not regulated by "end metabolites" of pathways because those metabolites reside in the FBA and are at ss.

The intracellular xylose is processed by the xylose isomerase reaction catalysed by XylA, and, in the dFBA, the ODE-predicted flux across the XylA reaction is used to define constraints in the corresponding reaction in the linear program. Note that in the ODE, the XylA reaction is not product-inhibited as in the XylE reaction. This is because the product of this step, xylulose, is an FBA metabolite.

The spatial conversion factors have been included in the above to remind ourselves that the biomass is measured in OD-units, and that a volume-difference exists between the two compartments across which transport processes operate. Here, we have assumed that an OD_{600} of 1.0 corresponds to a cell concentration of $10^{12} \cdot \text{cells} \cdot \text{L}^{-1}$, that the external volume is 1.0 L, that the internal cell volume is 10^{-15}L , and that the GDW of a cell is 2.8 (10^{-13}) grams (Milo, Jorgensen et al. 2010). A conversion factor is also included for time to explicitly reflect the fact that the specific growth rates for *E.coli* are conventionally understood to carry the unit of h^{-1}, whereas the relevant timescale in the equations is sec^{-1}.

The *E. coli* metabolic model iJR904 was used to introduce the stoichiometric constraints of the FBA step (Reed, Vo et al. 2003). Xylose diffusion was added as an extra reaction as this step is not included in the published model. We have also used the NGAM and GAM values that have been published for the iJR904 model: respectively, 7.6 $\mathrm{mmol} \cdot \mathrm{GDW}^{-1} \cdot \mathrm{h}^{-1}$ and 45.56 $\mathrm{mmol} \cdot \mathrm{GDW}^{-1}$. The maximal oxygen consumption rate was constrained to a value of 15.0 $\mathrm{mmol} \cdot \mathrm{GDW}^{-1} \cdot \mathrm{h}^{-1}$ (Varma and Palsson 1994). The value was lowered from the maximal value of 20.0 $\mathrm{mmol} \cdot \mathrm{GDW}^{-1} \cdot \mathrm{h}^{-1}$which has been used throughout this thesis as we found from initial dFBA runs that our model does not predict an acetogenic growth regime for this v_{o2}^{Max}. Finally, we have used the Palsson-style method of assuming a biomass objective with the biomass composition of the published iJR904 model (Reed, Vo et al. 2003).

7.5: Parameters and the calculation of β

The summary of all numerical values we have assigned to parameters in this chapter are presented in table 7.5.1 with their units and sources.

Parameter	unit	Value used	Value cited	Source
NGAM	$mmol \cdot GDW^{-1} \cdot h^{-1}$	7.6	7.6, 8.39	(Reed, Vo et al. 2003, Feist, Henry et al. 2007)
GAM	$mmol \cdot GDW^{-1}$	45.56	45.56	(Reed, Vo et al. 2003)
v_{o2}^{Max}	$mmol \cdot GDW^{-1} \cdot h^{-1}$	15.0	15.0	(Varma and Palsson 1994)
C_{bio}	$mmol \cdot GDW^{-1}$	41.46	41.46	(Reed, Vo et al. 2003)
k_{Cat}^{XylE}	sec^{-1}	60.0	1.5~2320.0	(Stein and Litman 2014)
K_m^{XylE}	$mmol \cdot L^{-1}$	0.1	0.2~4.0	(Sumiya, Davis et al. 1995)
K_i^{XylE}	$mmol \cdot L^{-1}$	100.0	4.6-1174.0	BRENDA (www.brenda-enzymes.org) *XylA inhibition by xylitol
k_{Cat}^{XylA}	sec^{-1}	90.0	0.007-258.0, 47.0	BRENDA (www.brenda-enzymes.org) (Umemoto, Shibata et al. 2012)
K_m^{XylA}	$mmol \cdot L^{-1}$	10.0	0.076-605.0	BRENDA (www.brenda-enzymes.org)
K_{if}^{XylA}	$mmol \cdot L^{-1}$	20.0	4.6-1174.0	BRENDA (www.brenda-enzymes.org) *XylA inhibition by xylitol
k_{Diff}^{xyl}	sec^{-1}	$1.0 \cdot 10^{-5.0}$	NA	*required for initiation; arbitrarily chosen to be small
k_{cat}^{eTOg}	sec^{-1}	90.0	90.0	Value for k_{Cat}^{XylA}
K_m^{eTOg}	$mmol \cdot L^{-1}$	10.0	10.0	Value for k_m^{XylA}
K_i^{eTOg}	$mmol \cdot L^{-1}$	100.0	100.0	Value for k_i^{XylE}
k_{cat}^{gTOh}	sec^{-1}	90.0	90.0	Value for k_{Cat}^{XylA}
K_m^{gTOh}	$mmol \cdot L^{-1}$	10.0	10.0	Value for k_m^{XylA}
K_i^{gTOh}	$mmol \cdot L^{-1}$	100.0	100.0	Value for k_i^{XylE}
k_{cat}^{eTOh}	sec^{-1}	90.0	90.0	Value for k_{Cat}^{XylA}
K_m^{eTOh}	$mmol \cdot L^{-1}$	10.0	10.0	Value for k_m^{XylA}
K_i^{eTOh}	$mmol \cdot L^{-1}$	100.0	100.0	Value for k_i^{XylE}
k_{cat}^{hTOf}	sec^{-1}	90.0	90.0	Value for k_{Cat}^{XylA}
K_m^{hTOf}	$mmol \cdot L^{-1}$	10.0	10.0	Value for k_m^{XylA}
K_i^{hTOf}	$mmol \cdot L^{-1}$	100.0	100.0	Value for k_i^{XylE}
V_{max}^{Ie}	$mmol \cdot sec^{-1}$	0.02	0.75	(Lee 2009)*value for citrate synthase
K_m^{Ie}	$mmol \cdot L^{-1}$	100.0	0.076-605.0	BRENDA (www.brenda-enzymes.org)*value for XylA
V_{max}^{If}	$mmol \cdot sec^{-1}$	0.02	0.75	(Lee 2009)*value for citrate synthase
K_m^{If}	$mmol \cdot L^{-1}$	100.0	0.076-605.0	BRENDA (www.brenda-enzymes.org)*value for XylA
V_{max}^{Ig}	$mmol \cdot sec^{-1}$	0.02	0.75	(Lee 2009)*value for citrate synthase

K_m^{Ig}	$mmol \cdot L^{-1}$	100.0	0.076-605.0	BRENDA (www.brenda-enzymes.org)*value for XylA
V_{max}^{Ih}	$mmol \cdot sec^{-1}$	0.02	0.75	(Lee 2009)*value for citrate synthase
K_m^{Ih}	$mmol \cdot L^{-1}$	100.0	0.076-605.0	BRENDA (www.brenda-enzymes.org)*value for XylA
V_{max}^{Ixyl}	$mmol \cdot sec^{-1}$	0.02	0.75	(Lee 2009)*value for citrate synthase
K_m^{Ixyl}	$mmol \cdot L^{-1}$	100.0	0.076-605.0	BRENDA (www.brenda-enzymes.org)*value for XylA
k_d^{XylR}	$mmol \cdot L^{-1}$	$3.0 \cdot 10^{-3.0}$	$3.3 \cdot 10^{-3.0}$	(Ni, Tonthat et al. 2013)
β_R^0	$mmol \cdot L^{-1} \cdot sec^{-1}$	$1.0 \cdot 10^{-10.0}$	NA	*required for initiation; arbitrarily chosen to be small
β_R	$mmol \cdot L^{-1} \cdot sec^{-1}$	$5.0 \cdot 10^{-8.0}$	$5.0 \cdot 10^{-8.0}$	(Robison, McGuire et al. 1998, Milo, Jorgensen et al. 2010) Bionumber 102632; *TF concentration required to be two-three orders of magnitude below the value for XylE
κ_R	$mmol \cdot L^{-1}$	$1.0 \cdot 10^{-5.0}$	$1.0 \cdot 10^{-4.0}$	(Robison, McGuire et al. 1998, Milo, Jorgensen et al. 2010) Bionumber 102632;
β_E	$mmol \cdot L^{-1} \cdot sec^{-1}$	$5.0 \cdot 10^{-6.0}$	$2.7 \cdot 10^{-6.0}$	(Nelson, Lehninger et al. 2008) *calculated as described in text
κ_E	$mmol \cdot L^{-1}$	$1.0 \cdot 10^{-5.0}$	$1.0 \cdot 10^{-5.0}$	Value for κ_R
β_A	$mmol \cdot L^{-1} \cdot sec^{-1}$	$7.0 \cdot 10^{-6.0}$	$2.7 \cdot 10^{-6.0}$	Value for β_E
κ_A	$mmol \cdot L^{-1}$	$1.0 \cdot 10^{-5.0}$	$1.0 \cdot 10^{-5.0}$	Value for κ_R
β_{eTOg}	$mmol \cdot L^{-1} \cdot sec^{-1}$	$7.0 \cdot 10^{-6.0}$	$2.7 \cdot 10^{-6.0}$	Value for β_A
κ_{eTOg}	$mmol \cdot L^{-1}$	$1.0 \cdot 10^{-5.0}$	$1.0 \cdot 10^{-5.0}$	Value for κ_A
β_{gTOh}	$mmol \cdot L^{-1} \cdot sec^{-1}$	$7.0 \cdot 10^{-6.0}$	$2.7 \cdot 10^{-6.0}$	Value for β_A
κ_{gTOh}	$mmol \cdot L^{-1}$	$1.0 \cdot 10^{-5.0}$	$1.0 \cdot 10^{-5.0}$	Value for κ_A
β_{eTOh}	$mmol \cdot L^{-1} \cdot sec^{-1}$	$7.0 \cdot 10^{-6.0}$	$2.7 \cdot 10^{-6.0}$	Value for β_A
κ_{eTOh}	$mmol \cdot L^{-1}$	$1.0 \cdot 10^{-5.0}$	$1.0 \cdot 10^{-5.0}$	Value for κ_A
β_{hTOf}	$mmol \cdot L^{-1} \cdot sec^{-1}$	$3.0 \cdot 10^{-6.0}$	$2.7 \cdot 10^{-6.0}$	Value for β_A
κ_{hTOf}	$mmol \cdot L^{-1}$	$1.0 \cdot 10^{-5.0}$	$1.0 \cdot 10^{-5.0}$	Value for κ_A
V_{max}^{aTObe}	$mmol \cdot sec^{-1}$	7.0		
n_f	NA	1.3	1.0~4.0	(Alon 2006)
K_{if}^{aTObe}	$mmol \cdot L^{-1}$	$6.0 \cdot 10^{3.0}$	NA	*Discussed in text
n_h	NA	1.3	1.0~4.0	(Alon 2006)
K_{ih}^{aTObe}	$mmol \cdot L^{-1}$	50.0	NA	*Discussed in text

Table 7.5.1: Table of all parameter values used for our dFBA simulations in this chapter. Note that many of these parameters will not appear until the hybrid dFBA model is discussed. The simple xylose dFBA model presented in this section is modified to this second dFBA model, and its parameter values remain unchanged in the process.

We have found that one particularly difficult parameter to come across in the literature is the maximal gene expression rate β. In arriving at a value for β_E, we made the following calculation. We started with the assumption that the number of copies of the XylE transporter at steady-state is 10,000 using a published value for the *E.coli* lactose transporter (Nelson, Lehninger et al. 2008). According to equation 1.5.1, this value translates to 0.016 $\text{mmol} \cdot \text{L}^{-1}$ for the intracellular XylE concentration. Bringing the equation for XylE expression (equation 7.4.6) to steady-state, and assuming promotor saturation, we are able to write

$$\beta_E = \mu_{ss}[E]_{ss}. \qquad \textbf{(7.5.1)}$$

Using the maximal specific growth rate observed experimentally (0.6 h^{-1}) as an approximation to the steady-state specific growth rate, the RHS of equation 7.5.1 was evaluated to our approximation to β_E ($2.7 \cdot 10^{-6.0}$). In all of our dFBA models, we have assumed, when quantifying β_A, that this value would be of the same order of magnitude as β_E. This may not be strictly a valid assumption as the promotors of the symport protein and the promotor for the isomerase are located in different regions of the *E.coli* genome (section 1.4). However, the expression of the genes for the xylose ABC transporter has been shown to be coupled to the expression of the xylose metabolic genes via a DNA looping mechanism (Tao, Gonzalez et al. 2001, Ni, Tonthat et al. 2012). Thus, we make the assumption that the upregulation of the transport and metabolic processes for xylose should take place at comparable rates under optimal conditions. When quantifying β_R, we have used the fact that the intracellular concentration of transcription factors are often several orders of magnitude below that of metabolic enzymes (Milo, Jorgensen et al. 2010). We have assumed for XylR that its steady-state expression rate (as in equation 7.5.1) would be two orders of magnitude below that of XylE.

7.6: Predictions of the simple xylose dFBA model

The dFBA algorithm (section 2.10) was run with the ODE system of section 7.4 and the parameter values from section 7.6. The predicted growth dynamics are presented in figure 7.6.1.

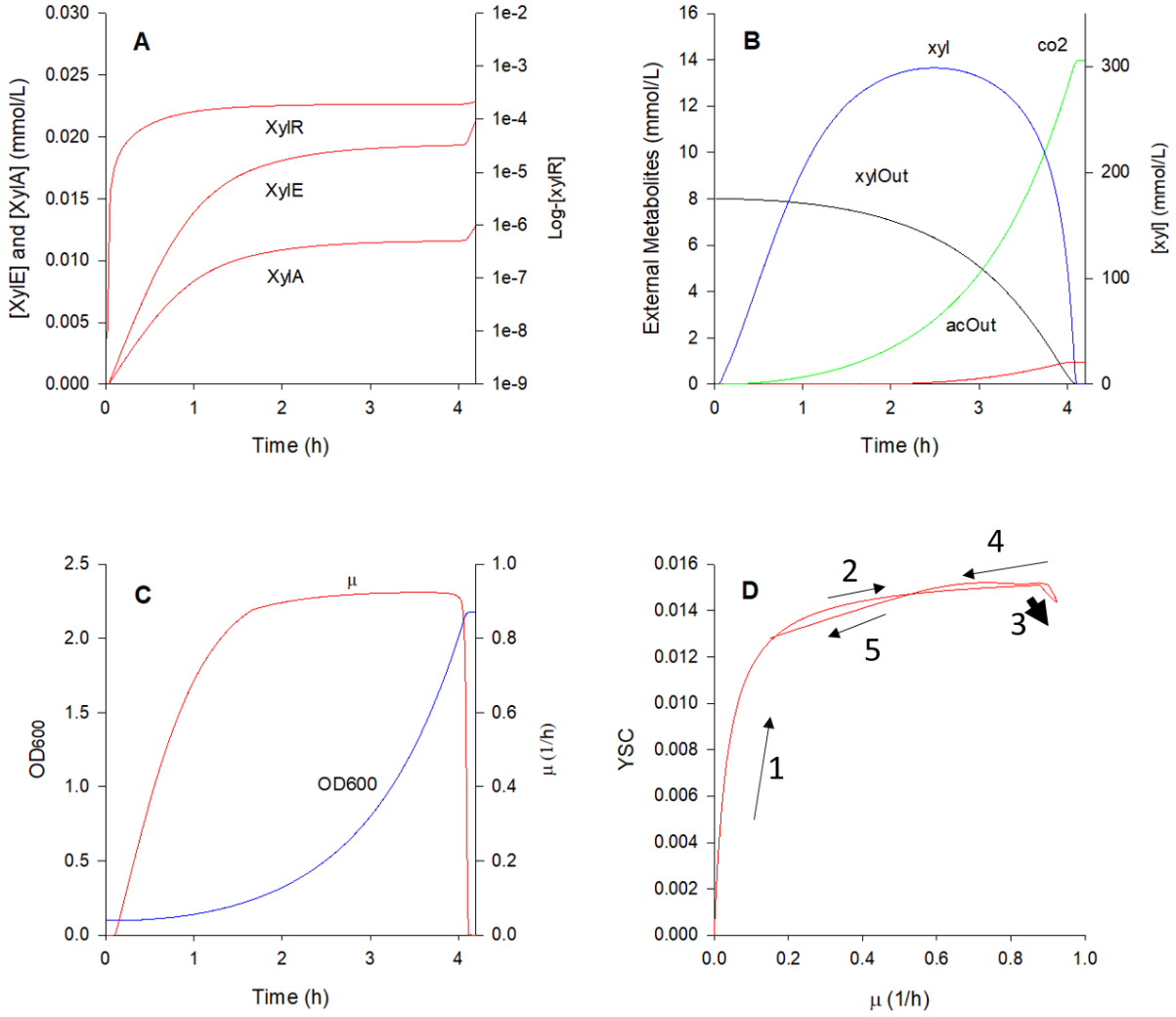


Figure 7.6.1: These plots show the output of the simple xylose dFBA model. A) The concentrations of the protein products of the xylose system upregulated by the presence of environmental xylose. Note that the concentration for the XylR protein is plotted using the right axis of the graph in panel A using a logarithmic scale. B) The concentration time series for the external metabolites (xylOut, acOut, and CO_2) are plotted together with the internal concentration of the substrate (xyl). C) The time series of population growth is plotted together with FBA predicted specific growth rates. D) The relation between yield and growth rate is tracked dynamically over the YSC-μ plane. The numbers and arrows indicate the direction of

movement along the depicted trajectory. A thick arrow is used at the third position to reflect that fact that the dFBA algorithm spends the majority of its time there.

It can be seen that, upon induction, the xylose transport and metabolic proteins tend to steady state concentrations, with the XylE protein achieving the highest steady-state value of $0.02\ \mathrm{mmol} \cdot \mathrm{L}^{-1}$, followed by XylA which achieves a concentration of $0.01\ \mathrm{mmol} \cdot \mathrm{L}^{-1}$. At all times, the concentration of the XylE protein exceeds that of the XylA protein. This is because $\beta_E$was assigned a slightly higher value relative to β_A. The consequence of this effect may be seen in the plot of internal xylose concentration in panel B of figure 7.6.1. The xylose curve is seen to increase toward a high steady-state concentration because the rate of transport is greater than the rate of metabolism. The xylose flux into the FBA model increases over the first hours of growth as the product of [xyl] and [XylA] increases. As a result, it is seen that the predicted specific growth rate also increases during this period (panel C). The specific growth rate prediction peaks, when the internal xylose concentration is maximal, around a value of $0.9\ \mathrm{h}^{-1}$ just over 1.5 hours into growth. Because the value of μ does not change appreciably once having approached this maximum, the biomass growth curve exhibits what is close to an exponential growth with a near-constant specific growth rate (panel C). The trajectory of the relation between growth yield and growth rate is plotted in Paned D of figure 7.6.1. It can be seen there that, at all time-points, the relation predicted by equation 4.4.6, and analysed as an FBA series in chapter 5, is obeyed. It is not until very close to internal xylose depletion that the trajectory is seen to make a return to region of the YSC-μ plane corresponding to lower growth rates. The region of the plane in which the dFBA model was seen to spend the most amount of time was determined, from inspecting the density of simulated points, to be the acetogenic region. This is indicated in panel D by a thick arrow. This fact may also be observed from, once again, noting

that μ is seen in panel C to be close to its maximal value for the majority of the growth. That this value of μ corresponds to acetogenic growth (at a v_{o2}^{Max} of 15.0 $mmol \cdot GDW^{-1} \cdot h^{-1}$) may be observed from the acetate accumulation that is seen in panel B. No ethanol production was seen to occur; note that, from the corresponding plots in chapter 5, this may be inferred from Panel D where the movement of the trajectory away from the jointly optimal hyperbola is too small to enter the ethanologenic regime.

The conclusion we draw from the simple model is that the dFBA approach of constraining metabolite uptake rates, as functions of external concentration, and using a biomass objective does introduce a systems level feedback of information, as claimed by dFBA authors (Mahadevan, Edwards et al. 2002). In the case of our model, the feedback involves two concentrations. The xylose concentration "informs" the biomass to grow, and the biomass concentration "informs" the xylose concentration to deplete. In this way, a feedback is established from the biomass onto itself, asserting that growth should cease as a function of population size.

On the other hand, a problem we identify with the simple xylose dFBA model is that its dynamic predictions are not qualitatively informative. The fact that the key variables that drive the dynamics of the system reach an early steady-state, combined with the insensitivity of μ to the applied constraints (xylose consumption rate) leads to a dFBA model which behaves essentially as an FBA model for the majority of the simulation. This is to say that rate-predictions of the simple dFBA model are very nearly static, or that the dynamics do not appreciably accelerate. A likely reason for this is because the just-mentioned feedback of information occurs on the timescale of population growth, and not metabolism. Thus we hypothesize that more dynamically significant dFBA models will need to account for feedback

transfer of information on the metabolic timescale, for example, by the inclusion of allosteric regulation. Fast feedback of information however requires that dFBA models consider not just the transfer of information that occurs through changes in environmental metabolite compositions, but also the information that is present internally to the cell; that is, the information that is available to a cell regarding its own metabolic system state as a function of internal metabolite concentrations (e.g. the internal xylose in the simple dFBA model). Thus, before discussing the possible ways by which to include faster information feedback, we will first deal with the problem of defining non-physical boundaries between the static and dynamic components of dFBA models. We start in the next section by considering the number of topologically distinct ways by which to set up dFBA problems.

7.7: How many distinct ways are there to set up a dFBA model?

As we have done with our simple xylose dFBA model, most microbial dFBA models that have been constructed to date and that we are aware of treat the environmental composition dynamically, but not the biomass composition (Mahadevan, Edwards et al. 2002, Meadows, Karnik et al. 2010, Hanly and Henson 2011). A notable exception is seen in the work of Upton et al (Upton, McQueen-Mason et al. 2017). This is because time-courses in biomass and metabolite concentrations in the external media are more easily measurable compared to their intracellular counterparts. As a simple thought experiment however, consider what might happen if the reverse situation were true; that is, if a particular cellular component was much easier to observe than the combination of cells and their external environment. In this case, it is entirely conceivable that modellers would opt to treat the observable cellular component as the dynamic component and to treat the rest of the cell together with the external environment as the

static component. In the real-world situation, the dFBA model is structured in such a way that a static system (the cell) is embedded in an ambient dynamic system (the environment). In the hypothetical reverse-world situation, the model is structured with a dynamic system embedded in an ambient static system. Biologically speaking, these are two very different models corresponding to two distinct physical systems. However, we will now argue that the two models mathematically exhibit the same structure. The main conjecture we present in this section is that the total number of ways to set up a dFBA model with a fixed number of FBA components is one. We build our argument in ten steps.

Step1: *Use formulation 2.9.1 to define an FBA model*

An FBA problem is a metabolic network and a global sink and an objective function and constraints.

Step2: *Enumerate FBA models by cellular objectives*

In any large model involving several FBA components, the number of FBA models is equal to the number of objective functions.

Step3: *Combine the sinks in separate FBA models into one*

All FBA models share a common global sink. This is because individual FBA sinks may be safely combined without altering the remaining components defining an FBA model in step 1 (the network, the objective, and the constraints).

Step 4: *Define how two FBA models are combined into one*

Two linear objective functions combine to form one when two networks are coupled by a common set of metabolites. Observe that for two metabolic networks represented by stoichiometric matrices $\mathbf{S_1}$ and $\mathbf{S_2}$, if the matrices are decoupled, then an FBA problem of the form

Optimize:

$$f = f(\mathbf{v_1}, \mathbf{v_2})$$

Subject to: **(7.7.1)**

$$\begin{bmatrix} \mathbf{S_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S_2} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{v_1} \\ \mathbf{v_2} \end{bmatrix} = \mathbf{0},$$

may be written as two separate problems

Optimize:

$$f_k = f_k(\mathbf{v_k})$$

Subject to: **(7.7.2)**

$$\mathbf{S_k} \cdot \mathbf{v_k} = \mathbf{0},$$

for k = 1, 2.

Step 5: *Define how an FBA model is transformed into a dFBA model*

An FBA model becomes a dFBA model when a network model is coupled to an ODE system. Formally, one partitions the set of all exchange reactions in the model into those that transfer flux between the network and the sink, and those that transfer flux between the network and the ODE system. The ODE system may be thought of in this way as the part of the FBA sink for which concentration information is explicitly accounted for in dFBA by integrating flux across the latter set of exchange reactions. The definite integrals evaluated in this way represent concentration information that provide the opportunity for feedback processes to be defined.

Step 6: *Restrict the transfer of information between two FBA components in dFBA*

Any transfer of information between two FBA models must be via an ODE component of a dFBA model. This is because information is lost to the sink as a consequence of not having integrated exchange fluxes.

Step 7: *Conclude that there is only one way to set up a dFBA problem given a fixed number of FBA problems*

We may conclude from steps 1-6 that, for a dFBA model with p FBA components and q = 1 ODE component, there is only one topologically distinct way by which to construct a dFBA model. Information may not be transferred via the sink (step 6), and fluxes may not be directly exchanged between FBA models without reducing the number p (step 4).

Step 8: *Observe that a dFBA model with one FBA model does not depend on the order of model compartmentalisation*

There is absolutely no concept of spatial orientation in FBA or dFBA. The dFBA algorithm (section 2.10) therefore does not distinguish between the physical spatial orders of embedding between its statically and dynamically treated components that occurs biologically. A sink in a steady-state network in a dynamic environment, has the same dFBA setup as a dynamic cellular compartment in a steady-state cell and environment surrounded by a sink.

Step 9: *Observe that two decoupled ODE models can be combined to form one ODE system*

This observation states simply that two differential equations may either form a single coupled system, or combine to form a single decoupled system. In either case, there is only one ODE system in the dFBA model.

Step 10: *Conclude that there is only one way of setting up a dFBA model with a fixed number of FBA models*

With p distinct FBA components and q decoupled ODE components, there is only one topologically distinct way by which to set up a dFBA model.

The conclusion in step 10 is the desired result. Whilst we have outlined a qualitative argument in this section that lead us to conjecture that the number of ways of setting up a dFBA problem for a fixed number of FBA and ODE models is one, the result remains to be proven mathematically or demonstrated computationally. Doing this is beyond the scope of our goals of this chapter. The significance of the result then, is the insight that this line of reasoning brings to the development of alternative dFBA formulations that account for dynamic biomass compositions. Specifically, we find that the partitioning the exchange reactions in step 5 should not in general be restricted by biological compartmentalisations that physically separate the various components of the systems because of our observation in step 8. In other words, topologically distant metabolic processes should be accountable under a common dFBA formulation. This idea is further explored in the next section.

Before proceeding, we find that it is very important to emphasize that we are interested in the number of ways of *setting up* dFBA models; what our conclusion does not state is that all dFBA models are mathematically equivalent. Such a statement would require a metric to be defined for a space of dFBA formulations and the proof of a uniqueness theorem for dFBA solutions. We find it unlikely that such a result is possible for general dFBA formulations. As one counterexample, if any one of the decoupled ODEs mentioned in step 9 where to exhibit a bifurcation, then we have no way of establishing a one-to-one mapping between initial conditions and dFBA trajectories. This is because the parameter set in $(\mathbf{M_{FBA}}, \mathbf{v^{Min}}, \mathbf{v^{Max}}, \mathbf{L}, \mathrm{f})$ of the dFBA algorithm becomes dependent on which branch of bifurcation the dFBA trajectory lands. One possible situation in which uniqueness of dFBA trajectories may be proven might be if the dFBA model is defined using only single-dimensional ODE models. In this case, each ODE is decoupled from the other; by sandwiching each equation between FBA models, one will have effectively assumed steady-state boundary conditions between the individual dynamical components of the larger metabolic system.

7.8: An alternative formulation of the xylose concentration dynamics

The result of section 7.7 states that the dFBA algorithm does not see the physical spatial order in which the static and dynamic components of the overall model have been embedded. This is because in every configuration, the number of differential equation systems is just one. The corollary to this result is that topologically separated biological compartments and subsystems may all be treated dynamically in dFBA without altering the overall architecture of the dFBA model. In this way, it is observed that the issue of how dynamic biomass compositions should be treated in dFBA is a special case of deciding on where to draw the biological boundaries. It is not a case of deciding amongst several possibilities for drawing

mathematical boundaries between static and dynamic model components. In this setting, the treatment of biomass composition dynamics, environmental composition dynamics, and community dynamics, are seen to be instances of the same problem from the perspective of dFBA topology. In any of these biological formulations, dFBA is equivalent to eliminating a set of FBA exchange reactions and replacing them with an ODE system.

In the Palsson FBA framework, a biomass reaction is employed and metabolites between the final step of substrate acquisition and biomass formation are assumed to be at steady-state concentrations. As a consequence, the physical boundary of metabolism is defined to coincide with the dynamic to static boundary of the FBA model, and there is no internal accumulation of mass. If therefore, in the Palsson-style dFBA setting, a compound that occurs at the periphery of the metabolic model is required a dynamic representation, this problem may be very easily resolved by extending the environmental boundary inward by several reactions. The apparent difficulty in the Palsson framework arises under settings where metabolites located far from the physical boundary need to be treated dynamically. As a result of asserting that internal metabolites may not accumulate, an ODE system in the Palsson framework is required to increase in dimension, in a coupled manner, as the distance of a dynamic metabolite from the physical boundary is increased. However, our discussion of dFBA topology suggests that ODE dimensions need not increase with topological distance between dynamically modelled dFBA components.

Two issues become immediately apparent from the discussion in the previous paragraph. First, we require a scheme for representing the concentration dynamics of internal metabolites that are not at steady-state values. Second, we require this representation scheme to be extendible towards deeper metabolic loci in a way that does not depend on metabolic distance

from the physical cell boundary. With the goal of arriving at a solution, we reconsider the Palsson formulation of the xylose dFBA model by asking how a Fell approach may have been different. Specifically, we consider how a Fell-FBA would have treated the concentration dynamics of intracellular xylose. Suppose that the xylose fraction was the only component of the biomass composition that is varying (the xylose model already makes this assumption implicitly). Our xylose model accounts for this dynamic strictly via its ODE component; that is to say that the FBA side of the overall model starts with xylulose, and not xylose. An alternative representation however would have been to draw xylose out in a separate exchange reaction, thus representing flux to the xylose composition of the biomass as a combination of ODE and FBA predictions. Having a xylose exchange while simultaneously using a biomass objective would clearly leave this exchange with zero-flux since Palsson-style FBA optimizes for (terminal) biomass yields. Thus, to ensure that flux imbalances can be predicted by FBA at the xylose node, the constraints on the reactions surrounding xylose must be strict LP constraints. The flux through the xylose exchange reaction would then be taken to be the difference in flux through the isomerase step and the transport step, which in turn, the ODE would determine. It is obvious from this description, of an alternative dFBA formulation, that its output would be identical to that of the original. The noteworthy difference is a technical one however. From the Fell-representation of xylose, but not from the Palsson-representation, it can be seen that by surrounding an FBA node for which there is an exchange reaction with strict LP bounds, integration of the flux through the exchange (set to infinity bounds) can be used to derive the concentration of a biomass metabolite that is not in steady state. The upshot to this representation is that, unlike with the strictly Palsson approach, the dynamical treatment of the non-steady state metabolite does not depend on the metabolite being coupled to the part of the

ODE system describing the dynamics of the environmental composition. Therefore, with the Fell-representation of variable biomass compositions, the modelling of non-steady state metabolite concentrations can be extended to metabolites located deeper within the metabolic system without increasing the dimensionality of an external ODE as a function of metabolic distance. We will now formally describe this method of doing dFBA.

7.9: Modifications for a hybrid dFBA algorithm for dynamic biomass compositions

In this section, we describe the modifications to the dFBA algorithm (section 2.10) that are needed for the dynamic representation of concentrations of biomass constituents. The reader is referred to section 2.10 as the same notations used in that section are used here as well.

1) A metabolite is treated dynamically or statically, but not both; thus, decide upon which subset of metabolites to treat dynamically. The set of dynamically treated metabolites will then be partitioned into those occurring internally to the cell and those occurring externally to the cell. Let $\mathbf{c_0^{in}}$ and $\mathbf{c_0^{ex}}$, be their initial concentrations so that

$$\mathbf{c_0} = \begin{bmatrix} \mathbf{c_0^{in}} \\ \mathbf{c_0^{ex}} \end{bmatrix}. \qquad \textbf{(7.9.1)}$$

2) For each dynamic metabolite, a_i, introduce a corresponding FBA metabolite, I_{a_i}. We will refer to this new metabolite as an indicator metabolite for a_i for reasons that will become immediately clear. Both metabolites should now be given separate exchange reactions. Thus, if we assume for concreteness that i = m, and assign to the indicator metabolite I_{a_m}, the index of m+1, the matrix equation 3.2.5 may be augmented as

$$\left(\begin{array}{ccc|cc|c} s_{11} & \cdots & s_{1n} & 0.0 & 0.0 & s_{1b} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ s_{m1} & \cdots & s_{mn} & -1.0 & 0.0 & s_{mb} \\ s_{(m+1)1} & \cdots & s_{(m+1)n} & 0.0 & -1.0 & 0.0 \end{array}\right)\begin{bmatrix} v_1 \\ \vdots \\ v_n \\ v_{biomass} \\ v_{EX\_a_m} \\ v_{EX\_I_{a_m}} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \textbf{(7.9.2)}$$

Note that a separate exchange reaction will not be needed if metabolite a_i has already an associated exchange reaction in the FBA model. This will be the case for external metabolites.

3) The exchange reaction for a_m in step 2 should now be given infinity bounds in both directions. Thus, let

$$v_{EX\_a_m} \in [-10^6, 10^6]. \qquad \textbf{(7.9.3)}$$

4) For each dynamic metabolite, a_i, consider the non-exchange reaction edges surrounding the metabolite. If an edge for a reaction (rxn) connects this metabolite to another dynamic metabolite $a_{j\neq i}$, then the rate of conversion across this edge is a part of the ODE system within which the concentrations of the two metabolites are variables. Thus, $\mathbf{M_{ODE}}$ has $[a_i]$ and $[a_j]$ as variables, and accounts for their interconversion by rxn in the definitions for $\frac{d[a_i]}{dt}$ and $\frac{d[a_j]}{dt}$. To ensure that the FBA predicted flux v_{rxn} is consistent with the ODE predicted flux v_{rxn}^{ODE}, the ODE predicted value should be assigned as a strict bound in the LP formulation.

$$v_{rxn} \in [v_{rxn}^{ODE}, v_{rxn}^{ODE}]. \qquad \textbf{(7.9.4)}$$

If, on the other hand, metabolite a_j is a steady-state metabolite, then modify the definition of rxn so that the indicator metabolite for a_i occurs on the same side and with the same stoichiometric coefficient as a_i. Thus, a reaction such as

$$s_i a_i \leftrightarrow s_j a_j \qquad \textbf{(7.9.5)}$$

would now be written as

$$s_i I_{a_i} + s_i a_i \leftrightarrow s_j a_j. \qquad \textbf{(7.9.6)}$$

With respect to the augmented matrix in 7.9.2, we define the entries of the row indexed by m+1 using the formula

$$s_{(m+1)j} = s_{mj} \qquad \textbf{(7.9.7)}$$

Computationally, the difference between a metabolite and its indicator is that, whilst the metabolite may participate as a variable in both the FBA and the ODE component of the dFBA model, the indicator metabolite is strictly an FBA variable.

5) Finally, for each indicator metabolite, assign a lower and/or upper constraint on its exchange reaction as functions, $v_{EX\_I_{a_i}}^{Min}([a_i])$ and $v_{EX\_I_{a_i}}^{Max}([a_i])$, of the concentration of the dynamic metabolite for which it has been defined. This step is required to constrain the flux to and/or from a non-steady-state node through reactions not explicitly modelled via ODEs.

$$v_{EX\_I_{a_i}} \in \left[v_{EX\_I_{a_i}}^{Min}([a_i]) , v_{EX\_I_{a_i}}^{Max}([a_i])\right]. \qquad \textbf{(7.9.8)}$$

As an illustration, if $[a_i] = 0.0$ were to hold, then we should not expect any outward flux from the node corresponding to this metabolite. Thus, we would constrain the exchange flux for the indicator metabolite as

$$v_{EX\_I_{a_i}} \in \left[0.0, v^{Max}_{EX\_I_{a_i}}([a_i])\right]. \quad \textbf{(7.9.9)}$$

With the modifications to the dFBA algorithm introduced in this section, we are able to now use the Palsson-style FBA formulation for the prediction of growth rates whilst accommodating for the biomass composition of intermediate metabolites in a dynamic fashion. The rationale behind the hybrid dFBA algorithm draws upon two key observations previously made about FBA and dFBA models. First is the observation that in the Fell-style formulation of the FBA problem, biomass constituents are drawn out individually. The major advantage of the Fell FBA is that stoichiometric coefficients with respect to a unit of biomass may be treated as model predictions (section 3.5). This is in contrast to the Palsson-style formulation of FBA in which the biomass composition is a static model parameter defined by the biomass reaction. Whilst our hybrid approach to dFBA still requires a biomass reaction (equation 7.9.2), the flexibility of the Fell-approach is employed towards the representation of intermediate metabolites. The second key observation is that the topological structure of a dFBA model is not altered by the addition of spatially decoupled ODE systems, as they will all become consolidated into a single differential equation (section 7.7). As a result of this, it becomes feasible to dynamically characterize several metabolite pools even when the metabolic distance between the pools may be arbitrarily large. Using a Palsson-style FBA formulation, one is able to use the information $\mathbf{c^{ex}}$ to model the dynamical feedback between environmental compositions and metabolic flux distributions, where the latter is constrained at each iteration of the dFBA algorithm as

$$\mathbf{v} \in [\mathbf{v^{Min}}(\mathbf{c^{ex}}), \mathbf{v^{Max}}(\mathbf{c^{ex}})]. \quad \textbf{(7.9.10)}$$

With our modification, the flux distribution is constrained instead as

$$\mathbf{v} \in [\mathbf{v}^{\mathbf{Min}}(\mathbf{c}^{\mathbf{in}}, \mathbf{c}^{\mathbf{ex}}), \mathbf{v}^{\mathbf{Max}}(\mathbf{c}^{\mathbf{in}}, \mathbf{c}^{\mathbf{ex}})], \qquad \textbf{(7.9.11)}$$

where the entries of $\mathbf{c}^{\mathbf{in}}$are derived exactly as they are for $\mathbf{c}^{\mathbf{ex}}$ by integrating FBA flux predictions across appropriate exchange reactions.

To summarize the approach we propose, we use the Fell-FBA formulation to leave the biomass composition of intermediates as variable predictions of the model, and the Palsson-FBA formulation to predict specific growth rates. The advantage of this hybrid method of implementing dFBA is that dynamic biomass compositions may now be treated as ODE variables rather than as simple FBA parameters. As a result, the regulatory effects of internal metabolites whose concentrations are not in steady-state concentrations may now be investigated in the framework of dFBA.

7.10: Network reduction of the iJR904 FBA model

Our results from section 7.7 suggest that fundamental to dFBA is the understanding of how best to draw biologically relevant boundaries when partitioning a complex metabolic system into dynamic and static components. As was seen in the previous section, this partition is not necessarily required to coincide with actual physical boundaries, such as membranes, when viewed purely from the standpoint of setting up a biological dFBA model. In this view, we conjecture that the more relevant aspect of a partitioned component of a metabolic model in dFBA than its structural resolution, is the ability of the subsystem to transfer relevant biological information across its boundary to other dFBA components. In the concluding section of chapter 5, we posed the following question about Palsson-style FBA: Can Palsson-style FBA predictions of rates and of yields be made by stoichiometric models that have been substantially reduced in dimension relative to a parental genome scale model? In this section, we demonstrate, by reducing the iJR904 metabolic model that the answer is in the affirmative. In showing that this

can be done, we will also argue that growth yields and growth rates in the Palsson-FBA framework represent information that may be transferred across a metabolic boundary without relying on a high resolution metabolic reconstruction. We argue that whilst the parametrization of the reduced network requires the biological information that is contained in a genome scale model, the genome scale model itself is unnecessary for the purposes of dFBA. Our reasoning begins as follows.

In sections 5.2 and 5.3, it was seen that despite the differences in dimension, the three *E.coli* metabolic models studied there (iJR904, iAF1260, and iJO1366) make essentially identical predictions of biomass yields and growth rates in response to increasing the consumption rate of xylose (figures 5.2.1 and 5.3.1). Whilst slight quantitative differences were observed (e.g. the iJO1366 model predicts higher yields), equation 4.4.7 shows that these deviations from equality may be minimized by parametrizing appropriately for maintenance energies and the coefficients of the biomass reaction (equations 3.2.7 and 3.2.18). These results indicate that the iJR904 model, which is the smallest of the three may be expanded by following a network reconstruction protocol to the iJO1366, the largest of the three without significantly altering yield and rate predictions. In this light, we ask if the reverse case is also true; that is if the iJR904 model may be put through a network reduction procedure to arrive at a minimal network model containing the sufficient amount of biological information (derived from iJR904) to make yield and rate predictions that quantitatively agree with those of its parental model.

In defining the target metabolic network model (M-model), we begin by defining its boundary as in definition 1 of section 7.7: An FBA model is a metabolic network combined with a global sink combined with an objective function and constraints (formulation 2.9.1). Thus, the M-model is one with a single boundary partitioning a yet unknown stoichiometric network from

a global sink, and flux are transferred across this boundary via exchange reactions. To specify that this model is a Palsson-style formulated FBA model, we assume that the M-model will try to optimize the efflux rate of a stoichiometrically fixed amount of carbon into the sink. For notational consistency (equations 4.4.2-4.4.7), we will denote this stoichiometric amount of carbon by C_{bio} and the normalized efflux rate by μ. Likewise, this process of transferring carbon from the network to the sink will be denoted "EX_bio." Thus, μ is identical to v_{EX_bio}, and μ=1.0 exactly when a C_{bio} (biomass) amount of carbon is transferred from the M-model to its sink via EX_bio. The value of carbon flux through the biomass reaction of the iJR904 model that corresponds to a specific growth rate prediction of 1.0 is 41.46 $\text{mmol} \cdot \text{GDW}^{-1}$. Thus, we will parametrize our model by setting

$$C_{bio} = 41.46. \qquad \textbf{(7.10.1)}$$

Similarly, we will need to assume that there are processes with which the sink is able to deliver a flux of carbon into the network. Imitating the processes of xylose transport that is available to the iJR904 network, we will introduce two processes by which a five-carbon compound which we also call "xylose," (or "xyl") is transferred into the M-model. These two processes will be called XylH and XylE. Recall from section 2.2 that enzyme-catalysed reaction names are given the same name as the catalyst. In the case of XylH, the ABC transporter, we have left out the "FG" part from the notation for succinctness. Finally, we will need to introduce exchange reactions that may potentially compete with EX_bio. For simplicity, we will consider two competitive carbon efflux processes. The first will be termed "EX_co2" and will be defined to be a process exporting single-carbon units from the network into the sink. The second will be termed "EX_ac" and will be defined to be a process exporting two-carbon units. Note that for a same value of v_{EX_co2} and v_{EX_ac}, that the latter flux will transfer a greater amount of carbon

from the network into the sink. Finally, we will assume that there is a compound denoted "o2Out" and that the one carbon unit that is exported by EX_co2 is one part carbon and one part o2Out. As a result, in order to export a unit of carbon as co2, EX_co2 will require an equal amount of flux of o2Out to be transferred from the sink into the network. The process by which o2Out may be transferred to the network will be termed EX_o2Out. With these exchange reactions, the boundary separating the M-model from its sink has been fully defined. Note that there are only two types of elementary objects (carbon and o2Out) that we have defined to be transferrable across this boundary. The system, as it currently stands, has a corresponding stoichiometric matrix of dimension m=5 by n=6 and rank 5. The only way for this network to satisfy a steady-state condition (equation 2.6.3) is by exchanging zero flux with the sink. To allow for non-trivial solutions to the null-space equation, internal metabolites, of one, two, and three carbon units were added to the matrix and reactions were added together with a stoichiometrically balanced set of reactions that form pathways between the incoming xylose and respectively, co2, ac, and bio. Since we are interested in whether networks with arbitrary metabolite interconversion topologies could reproduce the iJR904 simulated YSC curves, we have tried to introduce the internal reactions connecting the exchangeable metabolites without setting *a priori* topological requirements. There were however four important exceptions. First, in order to separate the transport of substrate and early metabolic steps unique to the substrate, it was required that xylose, upon entry into the network is first converted to a different five-carbon unit. Second, the pathway to co2 production was deliberately decoupled from the synthetic steps leading to bio as a measure of simplification. Third, a two-carbon unit was placed in the network as a node from which flux may be directed either towards the synthesis of bio, or the production of ac. Fourth we have ensured that the network is able to support the futile cycling of

metabolites. This final requirement, which we have come to find in retrospect to be unnecessary, was placed to detect computational difficulties that may arise from the presence of futile cycling when the model is used in a dFBA algorithm. As a further trivial requirement, we have also required that the underlying graph of the network model is connected. Note that step 4 in our argument presented in section 7.7 suggests that this requirement has already been asserted upon declaring that we arrive at a single (close to) minimal FBA model representative of *E.coli* metabolism.

Finally, to introduce an energy currency, we have added two metabolites, "ADP" and "ATP," whose summed concentration is constrained to be constant (moiety conservation). By coupling the conversion between these two metabolites to several reactions in the network, we have introduced an energy economy in our hypothetical network. The complete model is depicted in figure 7.10.1.

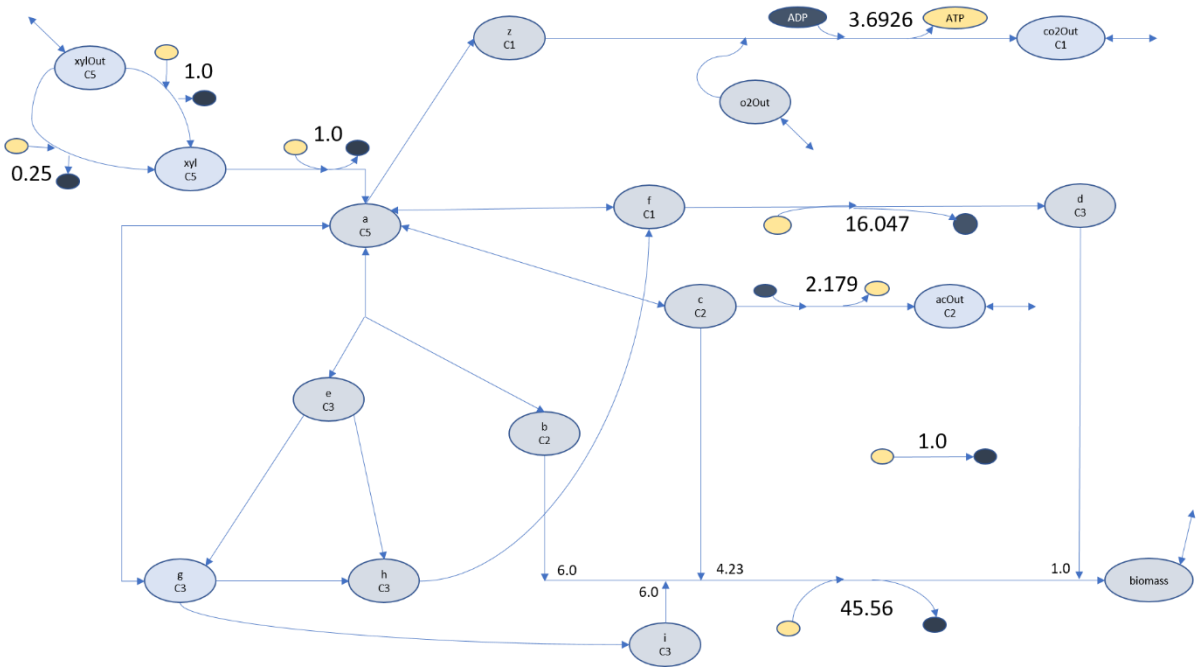


Figure 7.10.1: A pictorial representation of the network that has been reduced from the iJR904 metabolic model. Double arrows indicate reversible reactions and single arrows indicate irreversible reactions. Metabolites without biological counterparts have been assigned lower-case letter as names. These metabolites are represented as the nodes of the network. The carbon content of each metabolite is indicated below its given name. The conserved metabolite pair, ATP and ADP nodes, are depicted in blue (ADP) and yellow (ATP) respectively. Their interconversions are written alongside the respective arrows of the reactions they accompany. The stoichiometric coefficients for ATP and ADP are written within their transformation arrow. A biomass reaction has also been included. This transformation is represented by the bottom reaction draining internal metabolites towards the biomass node. The coefficient of each metabolite in the biomass reaction is indicated along the horizontal biomass reaction arrow. Finally, double arrows pointing to only a single node represent the exchange reaction for that node.

This network consists of 18 metabolites and 23 reactions, of which 5 are exchange reactions. In keeping track of the names of the hypothetical metabolites, we have assigned to each metabolite a lower-case alphabet letter indicating their name. For increased readability, these names will be italicised when appearing in text. The reader is also referred back to the notations section of this chapter.

Carbon flux from xylOut to the biomass was penalized energetically by assuming that some amount of current must flow from the ATP to the ADP node at certain steps along the way. The transport steps converting xylOut to xyl has been assigned a cost respectively of 1.0 and 0.25 units of ATP; similarly, the conversion of xyl to *a* has been also assigned a cost of one ATP unit. These values have been chosen to reflect the stoichiometric ATP costs for the biological xylose transport and isomerization steps. The conversion of a stoichiometric amount of internal metabolites to biomass has also been given an ATP cost. The value we assigned to the process is the numerical value of GAM (45.5) that is assigned to the biomass reaction of the iJR904 model. A cost to the synthesis of the metabolite *d*, required for biomass production, has been assigned a value such that GAA is consistent between the reduced network and the iJR904. Finally, a reaction was introduced to convert ATP to ADP without being coupled to any other model processes.

To meet the ATP demands, we have required that the reduced network sacrifice a portion of its incoming substrate carbon, thereby reducing the amount of carbon that may be transferred to biomass. The two reactions producing respectively, the metabolites co2Out and acOut, were coupled to ATP producing currents. The ATP and ADP stoichiometric coefficients for these two reactions have been assigned values that match the ATP production to by-product production gradients of the iJR904 FBA model. A series of FBA solutions was generated with the reduced network and the iJR904 network and their yield predictions were compared (figure 7.10.2).

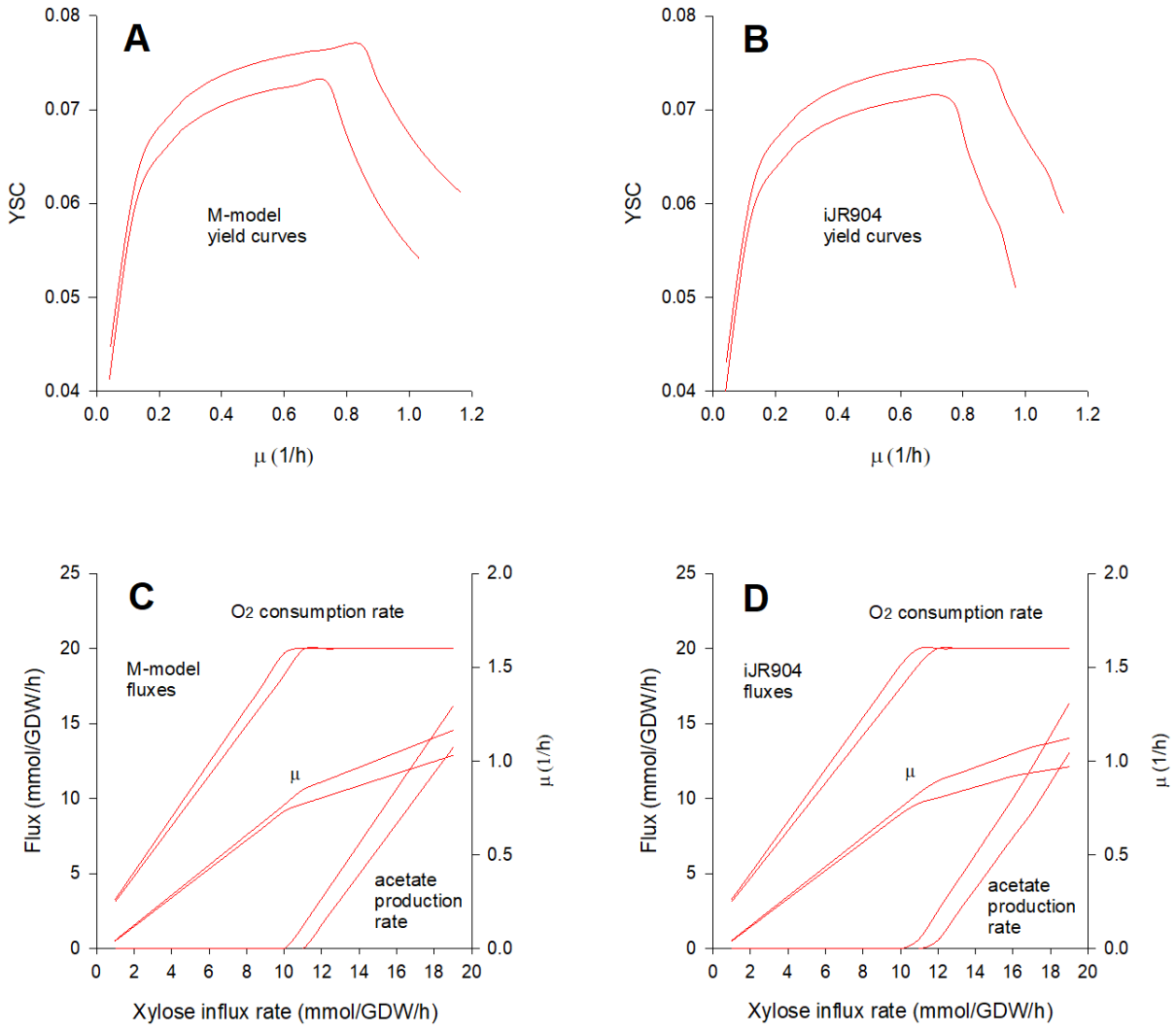


Figure 7.10.2: A comparison of the yield predictions between the complete and the reduced iJR904 model for aerobic xylose growth using high-energy and low-energy transport. Panels A and B compare the low-energy (top curves) and high-energy (bottom curves) yield predictions for the M-model (A) and the iJR904 model (B). Panels C and D compare the flux predictions for oxygen consumption, biomass production, and acetate production between the M-model (C) and the iJR904 model (D).

In figure 7.10.2, our results of comparing FBA series solutions between the iJR904 model and its reduced counterpart, the M-model, are presented. Two separate series of FBA solutions have been generated for each network model – one corresponding to high-energy cost substrate transport and another corresponding to low-energy cost substrate transport. It is apparent from

the plots of figure 7.10.2 that the predictions made by the two models for flux exchange between the steady-state network and the sink through exchange reactions are virtually identical. The optimal yield predictions made by the two FBA models agree, as do the predicted reduction in yield that occurs when substrate transport is switched from a high-energy cost to a low-energy cost transport process.

The results shown in figure 7.10.2 are possible most certainly because the reduced network has been constructed from, and deliberately parametrized to agree with, the genome-scale metabolic network that is the iJR904 model. However, what we can take away from the fact that procedure may be followed is the possibility that optimal yield predictions under the Palsson-style formulation of FBA may at times be blind to the internal topology of a metabolic system. Thus, when investigating non-topological growth phenotypes, such as optimal yields, it may be unnecessary under many circumstances to represent metabolism at the resolution of genome-scale or perhaps even small-scale network models. Whilst we have repeatedly emphasized in this section that the M-model is not a biological representation of metabolism from a topological perspective, we do not assert that the network does not contain biological information. To the contrary, we would argue that, by deriving the reduced network from a genome-scale model that has gone through a rigorous reconstruction procedure, the parameter values that are required by the M-model to match the predictions of the iJR904 contain integrative biological information. In the next two sections we use the reduced network model to demonstrate the application of the hybrid dFBA procedure that was introduced in section 7.9.

7.11: Building a hybrid dFBA model from the reduced iJR904 FBA model

The goal of this section is to demonstrate that a dFBA model may be formulated in such a way that an external dynamical system is topologically separated from an internal dynamical system. Towards this end, we describe how our modifications outlined in section 7.9 has been applied to the M-model to arrive at a hybrid dFBA model meeting this requirement.

In the first step, we partitioned the metabolites occurring in the network into those to be treated statically, and those to be treated dynamically. We have decided to treat all externally occurring metabolites with the exception of oxygen to be dynamic variables. As a way of making comparison with the simple dFBA model of xylose metabolism, we have included the same steps involved in the transport and metabolism of xylose as described by equations 7.4.1-7.4.8. Four internal nodes (e, f, g, and h) were chosen to be modelled dynamically. Nodes e, g, and h were chosen to investigate the mass-allocating properties of an internally mass-constrained dFBA model. Notice that there is no reason that incoming substrate mass should find a "way out" of a given metabolic network, and it is at this early stage, very likely that masses will accumulate internally. Whether such an effect can be circumvented by a system via simple feedback inhibition remains to be investigated. Node f was chosen as an end-metabolite with respect to the xylose flux arriving through the e-f-g system of nodes. For this pathway, node f can be seen topologically to be a potentially absorbing dead-end sink for incoming mass; for that same reason, the concentration accumulated at this node [f] may serve as a valuable signal for inhibiting the catalysis of the early steps.

Indicator metabolite nodes were introduced for the dynamic variables, and exchange processes were added for the xyl, e, f, g, and h nodes. The constraints on the non-indicator

exchange reactions were set to infinity in both directions so that fluxes may freely leave or enter the system provided the constraints from stoichiometry and the differential equations are obeyed. The indicator exchange fluxes were assumed to be bounded below by a simple Michaelis-Menten constraint as a function of the corresponding metabolite concentration. An upper bound was not specified for indicator exchange reactions. Thus, we have

$$v_{EX\_I_{a_i}} \in \left[(-1.0)\frac{V_{max}^{Ia_i}[a_i]}{K_m^{Ia_i} + [a_i]}, 10^{6.0}\right], \qquad \textbf{(7.11.1)}$$

where a_i is one of e, f, g, h, or xyl. This is as was done for the example given in equation 7.9.9. A generic value for a V_{max} corresponding to a central metabolic enzyme, the *E.coli* citrate synthase, was chosen arbitrarily for each indicator exchange; similarly, a value falling in the range published for the XylA enzyme was assigned as a generic K_m (table 7.5.1) (Lee 2009) .

The e, f, g, h group of nodes is interconnected by four reaction edges. These edges represent the reactions catalysed by the enzymes eTOg, gTOh, eTOh, and hTOf. In modelling the concentration dynamics of these enzymes, we used first order Hill functions as before (section 7.4); however, as we have not included a transcription factor for these catalysts, we have written the equations so that intracellular xylose will directly trigger the upregulation of their expression.

$$\frac{d[eTOg]}{dt} = \left[\frac{\beta_{eTOg}[xyl]}{[xyl] + \kappa_{eTOg}}\right] - \frac{\mu[eTOg]}{3600} \quad \textbf{(7.11.2)}$$

$$\frac{d[gTOh]}{dt} = \left[\frac{\beta_{gTOh}[xyl]}{[xyl] + \kappa_{gTOh}}\right] - \frac{\mu[gTOh]}{3600} \quad \textbf{(7.11.3)}$$

$$\frac{d[eTOh]}{dt} = \left[\frac{\beta_{eTOh}[xyl]}{[xyl] + \kappa_{eTOh}}\right] - \frac{\mu[eTOh]}{3600} \quad \textbf{(7.11.4)}$$

$$\frac{d[hTOf]}{dt} = \left[\frac{\beta_{hTOf}[xyl]}{[xyl] + \kappa_{hTOf}}\right] - \frac{\mu[hTOf]}{3600} \quad \textbf{(7.11.5)}$$

In writing equations 7.4.1 with 7.4.6-7.4.8, we have assumed that the presence of internal xylose will activate the genes for xylose transport and metabolism. In writing 7.11.2-7.11.5, we are also now assuming that the presence of internal substrate will also upregulate some general genes required for metabolism. For the numeric values of the β and κ parameters, we have assumed that they are equal to the values that have been assigned to the corresponding parameters for XylA regulation (equation 7.4.7).

The kinetics of the reaction steps catalysed by the four enzymes in equations 7.11.2-7.11.5 were written as irreversible Michaelis-Menten models with product inhibition. This is represented by the following equations.

$$v_{eTOg}^{ODE} = \frac{k_{Cat}^{eTOg}[eTOg][e]}{K_m^{eTOg}\left(1 + \frac{[g]}{K_i^{eTOg}}\right) + [e]} \quad \textbf{(7.11.6)}$$

$$v_{gTOh}^{ODE} = \frac{k_{Cat}^{gTOh}[gTOh][g]}{K_m^{gTOh}\left(1 + \frac{[h]}{K_i^{gTOh}}\right) + [g]} \quad \textbf{(7.11.7)}$$

$$v_{eTOh}^{ODE} = \frac{k_{Cat}^{eTOh}[eTOh][e]}{K_m^{eTOh}\left(1 + \frac{[h]}{K_i^{eTOh}}\right) + [e]} \quad \textbf{(7.11.8)}$$

$$v_{hTOf}^{ODE} = \frac{k_{Cat}^{hTOf}[hTOf][h]}{K_m^{hTOf}\left(1 + \frac{[f]}{K_i^{hTOf}}\right) + [h]} \quad \textbf{(7.11.9)}$$

The LHS of equations 7.11.6-7.11.9 may be interpreted as the rate of the corresponding reaction that would be predicted by a simple ODE model, of the internal dynamics for a given set of concentration values, had the model not been embedded into a larger network framework. Thus, to achieve constancy between the ODE rate predictions for the internal dynamics and the flux predictions of the surrounding FBA model, we have set dynamic strict bounds on the reactions whose catalyst concentrations are modelled dynamically by setting

$$v_{rxn} \in [\, v_{rxn}^{ODE}, v_{rxn}^{ODE}] \quad \textbf{(7.11.1)}$$

in the FBA, where rxn is one of eTOg, gTOh, eTOh, or hTOf. We have done similarly for xylose symport and xylose diffusion.

Finally, in order to demonstrate that the representation of internal dynamic concentrations may be used to model metabolic regulation, we have introduced two mechanisms by which early metabolic steps may be affected by metabolites occurring later in pathways. Firstly, we

introduce a rule that xylose metabolism by the isomerase protein may be inhibited by f accumulation. The Michaelis-Menten bound for this reaction has been set using

$$v_{XylA}^{ODE} = \frac{k_{Cat}^{XylA}[XylA][xyl]}{K_m^{XylA}\left(1 + \frac{[f]}{K_{if}^{XylA}}\right) + [xyl]} \quad \textbf{(7.11.11)}$$

Second, we also introduce the assumption that the reaction catalysed by aTObe may be dually inhibited by h and f as

$$v_{XylA}^{ODE} = V_{max}^{aTObe}\left(\frac{K_{if}^{aTObe}}{f^{n_f} + K_{if}^{aTObe}}\right)\left(\frac{K_{ih}^{aTObe}}{h^{n_h} + K_{ih}^{aTObe}}\right) \quad \textbf{(7.11.11)}$$

With these equations in hand, we will now describe the behaviour of the hybrid dFBA model.

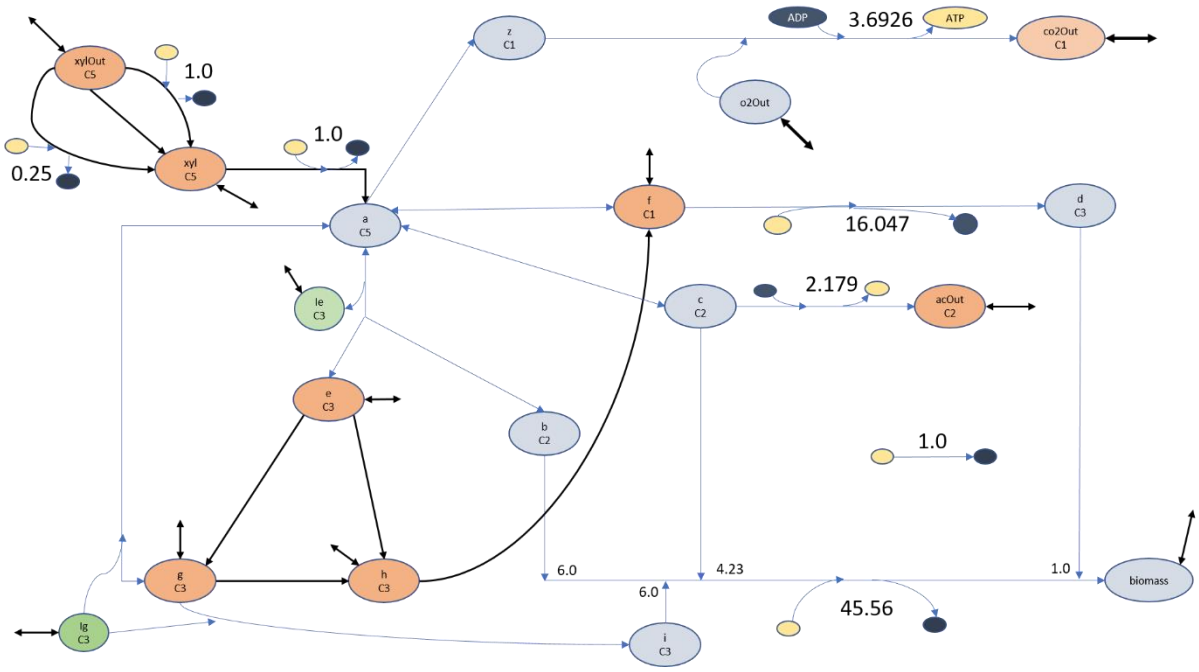


Figure 7.11.1: A pictorial representation of how the reduced network model in figure 7.10.1 has been transformed into a network appropriate for hybrid dFBA using the modifications suggested in section 7.9. The network is identical to the one in figure 7.10.1. However, dynamic metabolite nodes are now depicted in orange and dynamic edges in black. For clarity, we have shown only two indicator metabolite nodes (for g and e).

7.12: The predictions of the hybrid dFBA

We have applied the hybrid dFBA approach to the M-model as described in the previous section. We here present the predictions of the model. In figure 7.12.1, the fluxes across various reactions included in our model are presented. Similarly, in figure 7.12.2, the concentration dynamics of the non-steady state metabolites are presented.

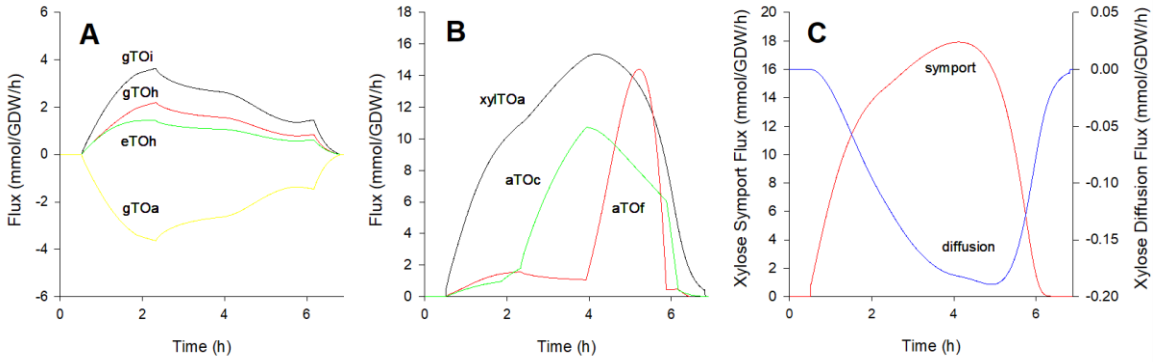


Figure 7.12.1: Fluxes across reactions of the M-model during the course of hybrid dFBA

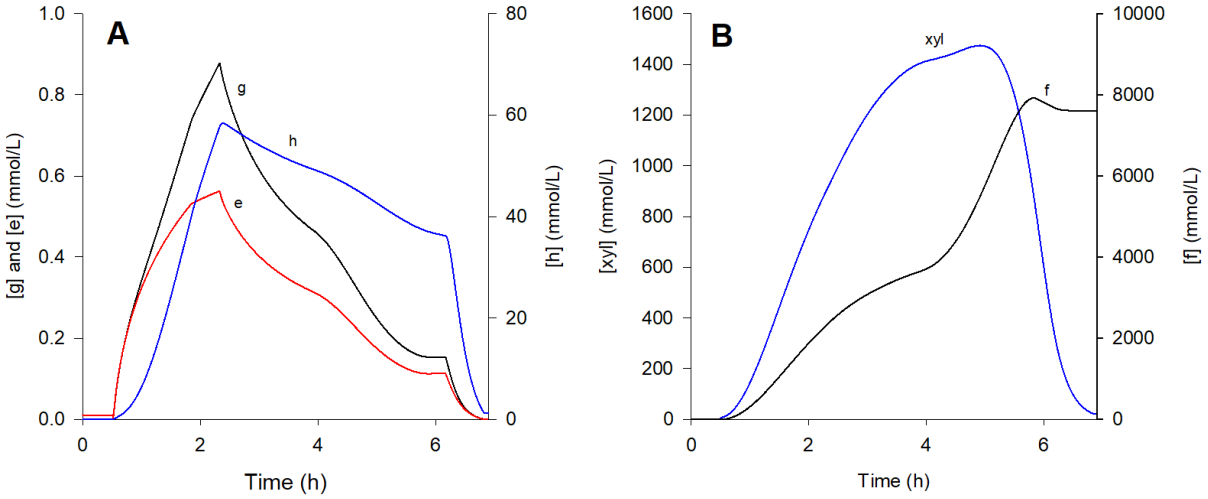


Figure 7.12.2: Concentration dynamics of non-steady state metabolites of the M-model hybrid dFBA

The initial trajectories of the fluxes in figure 7.12.1 are seen to be similar in behaviour to what was observed with the simple dFBA model of xylose growth. It is seen from panel B of figure 7.12.2 that during the early stages of growth, the internal xylose concentration increases as the xylose metabolic system is activated in a positive feedback between internal xylose and its transporter, XylE. Because we have assumed that the presence of intercellular xylose will activate the transcription of general metabolic enzymes, it is seen in panel A of 7.12.1 that the

fluxes through their corresponding reactions increase with xylose concentration. The metabolic system is not in an internal steady-state however; the concentrations of the e, f, g, h group of metabolites, separated from the xylose system via a steady-state reaction (aTObe), are dynamically changing in a way decoupled in the xylose ODE system. As a result, it is seen that masses are not allocated optimally by the FBA model, but rather increase and decrease depending upon what constraints are place in the model by the ODE system. The effect of these dynamically changing concentrations on the external dynamics of the model are shown in figure 7.12.3.

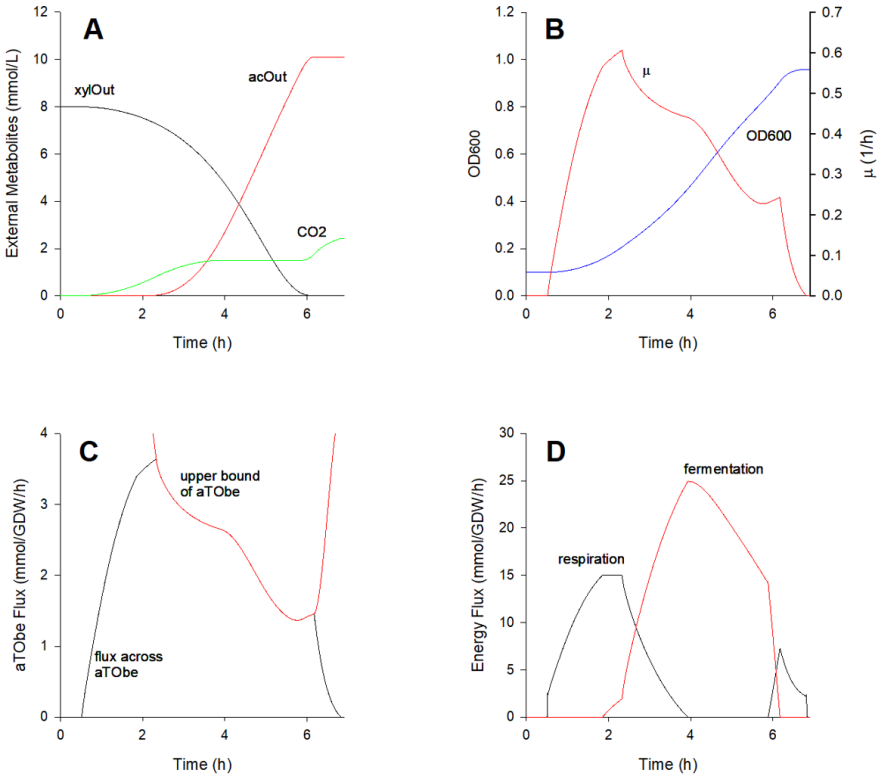


Figure 7.12.3: Dynamical consequences predicted by the M-model. A) External concentration dynamics of substrate and energy by-products. B) specific growth rate and biomass concentration predictions. C) the interaction between the optimal flux across the aTObe reaction and its dynamically changing constraint. D) Fluxes across fermentation and respiration over time.

External xylose is seen to be consumed by the system accompanied by CO_2 and acetate production. Whilst it may not be obvious from just observing the dynamics of substrates and by-product concentration, it can be readily concluded from the specific growth rate dynamics that the mode is predicting a more complicated growth behaviour then the simple xylose dFBA model. What is seen in panel B of figure 7.12.3 is that growth rate increases at first, like for the simple dFBA model, but becomes severely repressed as the growth trajectory is progressed. The reason for this growth rate depression can be understood from observing the curves of panel C of 7.12.3. What is shown there is the negative feedback transfer of information to an early metabolic step as a function of accumulating late metabolites. The red curve in the figure is the upper bound to the aTObe reaction in the FBA model which is a function of the dynamical concentrations of metabolites f and h. It is seen that as these regulatory metabolites increase in concentration (figure 7.12.2), their repressive effect on the early step of the pathway catalysed by aTObe becomes stronger. As a result of this, the dFBA trajectory transitions from a simple xylose-model like growth to a phase where the constraint to growth rate is no longer determined by the xylose system; rather, this control of growth is transferred to the e, f, g, h system. The exact point of this switch can be seen occur when the FBA constraint from f and h concentrations arrives to meet the optimal value of flux through aTObe under xylose-controlled growth. It is interesting to notice the effect of this arrival on the ability of the network to optimize for energy efficiency. It is seen in panel D of figure 7.12.3 that, whilst respiratory growth is possible and is thus used under xylose-controlled growth, this becomes no longer possible under e, f, g, h controlled growth. Fermentation is seen to increase in flux with the arrival of the aTObe constraint, and respiration flux is seen to decline. Why should it be the case that a dFBA model using a Palsson biomass objective predicts fermentation over respiration? This is because

acetate is effluxed into the media, not as an energetic by-product, but as a stoichiometric by-product. Thus what we can see is that this simple hybrid dFBA model predicts acetate production via a genuine overflow mechanism and not as an energetic strategy. As the aTObe constraint departs, the system is seen to be restored to respiratory growth.

One of the goals of this chapter has been to break the rigid correspondence between growth rate and growth yield predictions of the Palsson-style FBA formulation in dFBA. That a hybrid dFBA model is able to do this is represented in panel A of figure 7.12.4.

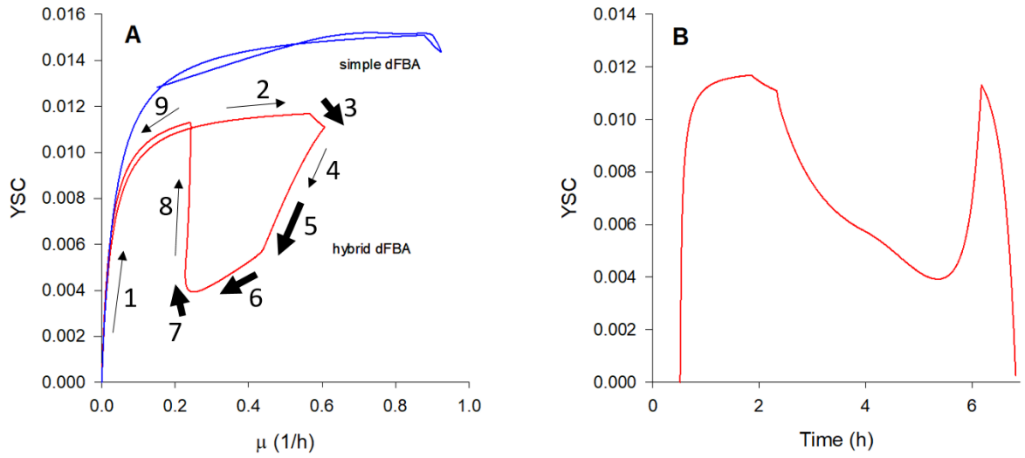


Figure 7.12.4: The time dynamics of YSC. A) The trajectory of the hybrid dFBA prediction along the YSC-μ plane (red) is compared with the corresponding trajectory of the simple xylose dFBA model (blue). Thick arrows are used to represent regions in which the hybrid model was seen to spend more time relative to the overall trajectory. B) The dynamics of YSC is plotted against time.

In figure 7.12.4, the dynamics of YSC is presented and compared with the YSC dynamics of the simple xylose dFBA model. What can be seen is that with the hybrid dFBA formulation, the metabolic system is able to explore a greater region of the YSC-μ plane. That is to say, that the dynamics of the model are no longer confined to the optimal curve described by equation

4.4.6. We have seen in sections 6.2 and 6.3 that yields and rates for real biological systems may relate in a way that is not as trivial as might be expected from a Palsson-style (d)FBA model. Thus, the hybrid dFBA approach which combines the key elements of the Fell FBA formulation with a Palsson-style biomass objective in dFBA brings the potential of modelling more complex dynamical systems behaviour than would be possible from a strict Palsson-style dFBA approach. Thus, we suggest that hybrid dFBA approach is more appropriate than the standard dFBA in modelling a wide range of real metabolic systems phenomena including, but not limited to, dynamic biomass compositions and microbial community interactions. The complete dynamic network model is presented in figure 7.12.5.

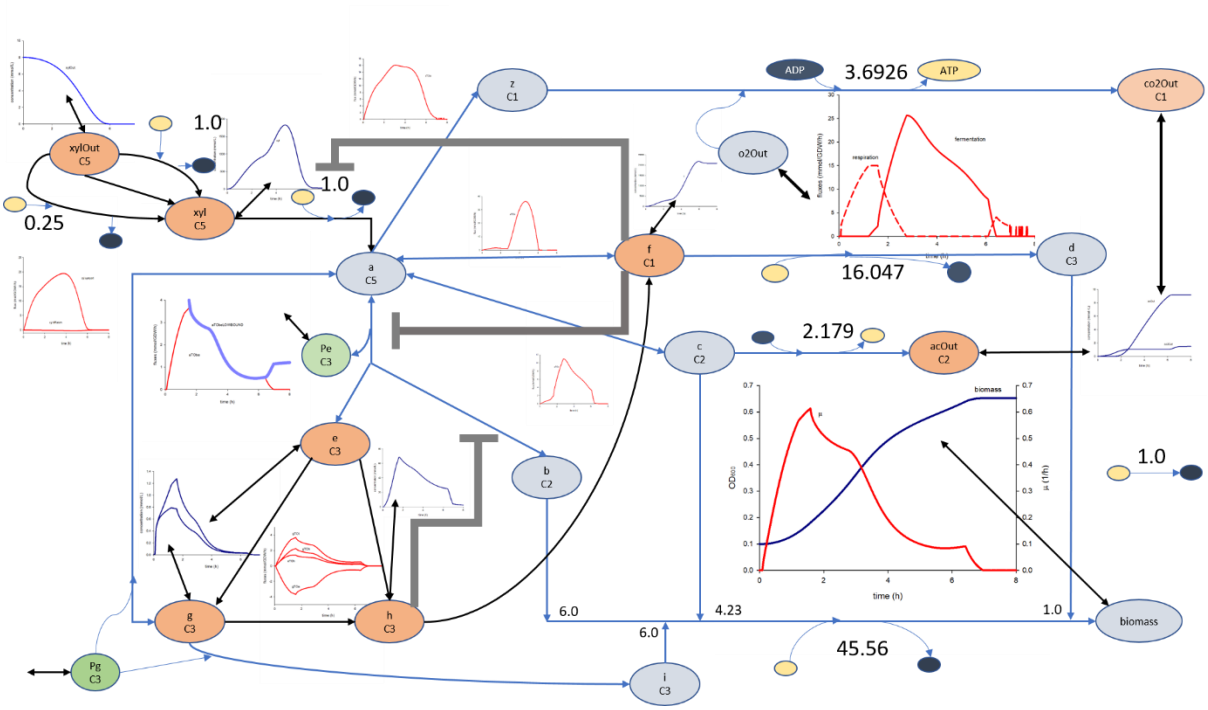


Figure 7.12.5: A pictorial representation of the dynamics of the M-model hybrid dFBA

We have presented, in this section, a novel approach to dFBA and used the M-model to illustrate the kinds of predictions our approach may make which are not possible with the traditional dFBA. It is important to emphasize however, that the M-model hybrid dFBA is only a qualitative representation of growth dynamics. Notice that the concentrations of internal metabolites that are predicted are far beyond biologically realistic values (we would suggest that $300\text{mmol} \cdot \text{L}^{-1}$ would be a good value for the total metabolite pool of a cell). Thus, the result that we have presented is a proof of principle; that yields and rates may be decoupled in dFBA.

7.13: Discussion

The goal of this chapter has been two-fold, and stem from our dissatisfaction with the apparent inability of dFBA methods that rely on a strict Palsson-style FBA formulation to produce biologically realistic growth curves without altering a set of phenomenological FBA parameters *posteriori* to meet investigator expectations. In chapters 4 and 5, it was seen that the main weakness of the Palsson approach to FBA modelling is the rigid correspondence that arises, when the biomass objective is assumed under steady-state conditions, between predictions of biomass yield and the specific growth rate. Thus, by extending our observations there, it was hypothesized that dFBA methods that are able to break this correspondence will lead to more qualitatively satisfying predictions of growth dynamics. Our first goal was therefore to generalize the notion of static-to-dynamic model boundaries in dFBA in such a way that these boundaries may be conceptualized independently of physical biological boundaries such as cell membranes. Our second goal was to use this more generalised notion of boundaries to introduce internal non-steady-state pockets of dynamic metabolite concentrations in a metabolic network. By introducing dynamic internal metabolite pools, the requirement of absolute flux balance

across physical cellular boundaries will be partially relaxed, and the potential of modelling internal metabolic feedback dynamics arises.

In meeting our goals, we have relied heavily on an alternative formulation of FBA that is in many respects, the inverse to the Palsson-style formulation. This "Fell-style" FBA brings a level of flexibility to predictions of biomass yields and growth rates by treating the composition of biomass as a variable to be predicted rather than as a model parameter. Unfortunately, the shortcoming of using a strict Fell-FBA formulation in dFBA is that the Fell-FBA requires specific growth rates to be parameters. However, our analysis of Fell style FBA has led us to conjecture that if model boundaries may be drawn independently of physical boundaries in FBA, then the same may hold true for dFBA. Pursuing this line of reasoning, we have been able to characterize the extension of FBA to dFBA as a process of choosing a set of exchange reactions over which to integrate fluxes so that information is retained over time intervals. Information is understood here to mean metabolite concentrations whose values depend on network inputs and outputs; thus information is "lost" to sinks through fluxes across exchange reactions that are not integrated in time. This formulation of dFBA does not depend on physical boundaries, but only on the boundaries across which an underlying FBA model interacts with its sink. Therefore we are able to combine the Fell-style approach to FBA with the Palsson style approach of using a biomass objective to formulate a dFBA procedure that predicts specific growth rates with the flexibility of Fell-FBA. We have referred to this approach of doing dFBA as a "hybrid" dFBA method.

The issue of drawing boundaries across FBA models leads to the following question; at what resolution should the internal metabolic network be described so that the appropriate information transfers are able to occur across steady-state boundaries? Our results of this

chapter, unsurprisingly perhaps, suggest that this depends on what information is being considered. Palsson-style FBA models and dFBA models that rely on the Palsson-approach are often corroborated by the ability of models to predict biomass yields; this may be for example by using a phenotype phase plane or matching biomass and by-product production to substrate depletion over time (Varma and Palsson 1994, Edwards, Ibarra et al. 2001, Ibarra, Edwards et al. 2002, Hanly and Henson 2011). By demonstrating that a substantially reduced metabolic network is still able to make optimal yield predictions of the iJR904 model however, we have demonstrated that yield is a phenotype which may be better characterized with consolidated parameters such as NGAM, GAA, and C_{bio}. This is jointly because first, flux-balance equations applied over the FBA model boundary do not "see" the underlying metabolic topology; second, because steady-state biomass yields are directly related to the flux of energetic by-products.

Having demonstrated that a reduced network is able to make yield predictions of a genome-scale model, we applied our hybrid dFBA approach to this hypothetical system. It was shown that by using the hybrid dFBA approach, we are able to describe feedback loops that occur internally within a metabolic system; thus, we have shown that the boundary of information transfer between the static and dynamic components of dFBA model need not coincide with the boundary between the environment and the cell. We find that characterizing dFBA models at this level of abstraction brings a unifying view with which to approach several seemingly separate modelling issues; these include the need to model feedback loops, the need to account for dynamic biomass compositions, and the need to extend FBA models to dynamic community models.

The M-model was used in this chapter to demonstrate how the hybrid dFBA may be computationally implemented as a method to introduce a flexibility between (d)FBA predictions

of yields and rates. We find that the reduced model is sufficient for meeting this goal. Despite this, the hybrid dFBA using the M-model is not representative of a realistic biological system. We point out however, that the model does make experimentally testable biological predictions. Biomass composition is predicted to be changing and doing so in a way that introduces internal feedback that renders substrate acquisition and early metabolism not rate limiting for growth. Similarly, energy maximization is not predicted to be an objective that has been yet achieved by our cells for xylose growth; although some observations suggest that perhaps with Adaptive Laboratory Evolution, this may be possible (Ibarra, Edwards et al. 2002). The model also makes the computational hypothesis that the appropriate level of resolution with which to describe metabolic topology to account for growth data may lie somewhere in between it, and the iJR904 model. What we take away from this is that the process of formulating and implementing novel dFBA methods is in it of itself a valid method for biological hypothesis generation.

CHAPTER 8: Discussion

Flux Balance Analysis (FBA) is a systems biological method that applies the known stoichiometric structure of metabolic networks towards microbial engineering. Genome-scale network reconstructions of metabolism are comprehensive and quality controlled through manual inspection. The flux-balance methods that rely on such reconstructions, not requiring detailed kinetic information, are easy to apply (Orth, Thiele et al. 2010, Thiele and Palsson 2010, McCloskey, Palsson et al. 2013). Therefore, in conjunction with the depth of existing knowledge available for a handful of model organisms and allied with the wide repertoire of techniques from modern experimental molecular biology, FBA methods offer an important interface between experimental and computational molecular biological efforts. We find however, that the question of how most efficiently to use the genomic information contained in metabolic reconstructions has not yet been answered. Strictly speaking, FBA models may be applied only to metabolic systems in steady-state, such as microbial populations replicating under a chemostat-controlled growth. Dynamic extensions of FBA are therefore required for modelling more general biological systems. The method of dynamic FBA (dFBA) combines FBA with ODEs to bridge this connection. This thesis concerns the investigation of the properties of FBA and dFBA models to arrive at improved dFBA implementations.

The need for improved dFBA methods represents two observations that we have made in our initial encounter with dFBA methods. First, we have been able to conclude that a number of separate but important biological systems to which FBA is applied today, particularly community FBA models of microbial interaction, may actually be

formulated under the dFBA framework. As we have argued in the introductory section chapter 7, community FBA models are in fact dFBA models. Thus an understanding of the properties of dFBA is a prerequisite to understanding community FBA modelling. The second observation, that prompted a need for better dFBA formulations, is the observation that when we ran simulations with dFBA models, the outputted growth trajectories were found to be unrealistically rigid. Variables such as specific growth rates were found to be highly insensitive to the outputs of the ODE component of our dFBA models. Thus, we have concluded, from experience, that dFBA methods have the potential to positively impact the analyses of a diverse group of important biological systems; however, the current dFBA method is not without flaws. How therefore to improve the dFBA method? More precisely, how to make dFBA predictions more "flexible?" Towards answering this question, we decided that the best place to start may be to try and better understand FBA methods. The results of this research effort has been presented in chapters 3, 4, and 5.

"What is FBA?" was the first question that was posed. The surprising answer to which we arrived, is that FBA is a structural metabolic model that predicts flux distributions that are consistent with the hypothesis of a cellular objective. Why this was found to be so surprising, was because we have also found that the vast majority of FBA model assume the common objective of maximizing a flux across a biomass reaction. Of course, the logic of the biomass objective is justified; it is a parsimonious representation of a metabolic goal in the absence of data. However, it had not occurred to us early on that using the biomass objective is in it of itself, a very strong hypothesis. Our conjecture at that point became that perhaps the rigidity of dFBA arises as a consequence of using a

biomass objective. We refer to this method of doing FBA as the "Palsson-style" FBA method. Two lines of investigation were initiated from this. First, we asked whether there are alternative means by which to do FBA, not just by choosing a separate objective, but alternative ways by which the FBA problem itself may be formulated. In the second line of investigation, we asked what the potential flaws are of the Palsson FBA method of using a biomass objective?

In engaging with the first problem, we have learned that an alternative FBA formulation exists within which the biomass composition, and not the growth rate, is predicted by the model. We refer to this style of FBA formulation as the "Fell-style" formulation. This method has been discussed in section 3.5 where it was argued that Fell-style FBA should be considered as an inverse to the Palsson-style FBA in that the predictions and parameters reverse in switching between the two.

The results of the second line of reasoning is presented in chapters 4 and 5. Our main theoretical finding is represented by equation 4.4.6, and its consequences to Palsson-style FBA predictions are explored in chapter 5. What we found is that under the Palsson-style FBA formulation, biomass yields and growth rates are two very rigidly coupled quantities. More specifically, optimizing for one was found to optimize for the other. Equation 4.4.6, in the form of 4.4.7, suggests that this is because Palsson FBA models optimize for the efficiency of energy biogenesis; equivalently put, Palsson FBA models find FBA solutions that minimize the efflux rate of energetic by-products. This is a consequence of assuming, simultaneously, a biomass objective, and a global steady state. In chapter 5, we have shown that energy biogenic constraints of metabolic networks used in Palsson-style FBA has a very strong central metabolic component as

well as a very strong phenomenological component (GAM). We have shown that the main contributors to the net exchange rates between energetic cofactors by the iAF1260 metabolic model are in fact central metabolic reactions and maintenance constraints. Therefore, to the extent that yields are coupled to rates in such a way as asserted by equation 4.4.6, it was reasoned that the genomic information contained in large-scale metabolic models may potentially be condensed as consolidated global parameters such as GAA. This was evidently suggested in section 5.3 that three dimensionally distinct metabolic reconstructions behave identically under the conditions there simulated.

With our findings, we returned to the question of improving dFBA in chapter 7. The statement that we made at the introduction of this chapter – that community FBA is dFBA – represents a conjecture that was made early on: that dFBA formulation is really about defining boundaries between components of metabolic systems, and modelling correctly, the transfer of information, in the form of dynamic metabolite concentrations, that occurs as topologically separated subsystems communicate. In chapter 7, we provided computational evidence that this is likely to be the case.

The original problem is recalled to be the following: how to make FBA modes more responsive to ODE models in dFBA? That is to say, how to make dFBA models more flexible? Through the findings of research process that has been described in the FBA chapters of this thesis, we are able to now state this question more precisely as well as provide a potential solution. Dynamic FBA models require a Palsson-style FBA because this is the formulation that allows for the prediction of growth rates. The rigidity of dFBA is a consequence of the rigidity in the coupling between yields and rates in Palsson-style FBA formulations. Thus, introducing a level of flexibility requires that

yields and rates be decoupled in the Palsson formulation. The solution to this problem comes when the Fell-style formulation of FBA is combined with our conjecture that dFBA modelling requires the modelling of boundaries between static and dynamic subsystems. Pursuing this line of reasoning, we arrive at three results which are presented in chapter 7. Our first result is that when the model boundary between an ODE system and a structural metabolic system is coincided with the physical boundary between a cell and its environment in the Palsson framework, that the overall FBA model may be substantially reduced in dimension and still make quantitative yield predictions that are consistent with a parental genome-scale model. The second result is that the Fell-style formulation of FBA may be applied to dFBA once it is realized that the model boundaries of dFBA need not coincide with physical boundaries. Combining these results together, we investigated the possibility of using the Fell-FBA approach of treating biomass composition as model predictions, whilst also using a Plasson-style approach of using a biomass objective to predict rates. Carrying out our work under the controlled setting of using a very small network which we refer to as the M-model, we arrive at our third and final result, which is a novel approach to dFBA modelling which we refer to as the "hybrid" dFBA.

In chapter 7, the dynamic predictions of yields and of rates of the hybrid dFBA were compared with those of the simple dFBA. What we have discovered was that in the hybrid approach, yield and rate predictions may be significantly decoupled, even when the biomass objective is used to predict growth rates. Most importantly, it was found that that the dFBA trajectory of the hybrid model is able to explore a wider region of the YSC-μ plane as a positive consequence of not being rigidly bound to the dual-optimality

curve represented by equation 4.4.6. Therefore, we have presented a novel approach to dFBA that is able to predict growth rates using a Palsson-style biomass objective whilst leveraging on the advantage of the Fell-style FBA of treating representing biomass composition as variables. The resulting hybrid dFBA exhibits the desired flexibility that we sought to find.

As future applications are concerned, it must first be noted that the result we presented in section 7.12 is a proof of principle that yields and rates may be decoupled in dFBA models. One of the main concerns we would have is that unless the underlying network is topologically capable of distributing masses amongst its nodes, that mass accumulations will inevitably occur. This was seen in the M-model where a metabolite (f) was unable to pass its concentrations off to another node, for example environmentally to a by-product. The consequence of this was seen to be that this metabolite acted effectively as a sink in the network in a similar manner that acetate behaves as an environmental sink in the simple dFBA formulation. Thus, what is suggested is that the M-model is far too reduced topologically to be of practical use. The "correct" topology is hypothesized currently to be represented by some network model that falls dimensionally in between the M-model and the iJR904 model; for example, perhaps a network representing just the central metabolism. Thus, the first step to follow is likely to be the demonstration that yields and rates may be decoupled by dFBA models in which internal concentration predictions are biologically realistic. In the hopeful assumption that this may be done, we find that the hybrid approach to dFBA which has been proposed in this thesis may have profoundly important implications to the dFBA modelling of complex systems such as microbial communities that are only just starting to be better understood.

REFERENCES:

Ahlem, C., et al. (1982). "Purification and properties of a periplasmic D-xylose-binding protein from Escherichia coli K-12." Journal of Biological Chemistry **257**(6): 2926-2931.

Alon, U. (2006). An introduction to systems biology: design principles of biological circuits, CRC press.

Ames, G. F. (1968). "Lipids of Salmonella typhimurium and Escherichia coli: structure and metabolism." Journal of Bacteriology **95**(3): 833-843.

Andersen, K. B. and K. von Meyenburg (1980). "Are growth rates of Escherichia coli in batch cultures limited by respiration?" Journal of Bacteriology **144**(1): 114-123.

Antoniewicz, M. R. (2013). "Dynamic metabolic flux analysis—tools for probing transient states of metabolic networks." Current opinion in biotechnology **24**(6): 973-978.

Arnold, A. and Z. Nikoloski (2014). "Bottom-up metabolic reconstruction of Arabidopsis and its application to determining the metabolic costs of enzyme production." Plant physiology **165**(3): 1380-1391.

Baba, T., et al. (2006). "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection." Molecular systems biology **2**(1).

Barabasi, A.-L. and Z. N. Oltvai (2004). "Network biology: understanding the cell's functional organization." Nature reviews genetics **5**(2): 101.

Basan, M., et al. (2015). "Overflow metabolism in Escherichia coli results from efficient proteome allocation." Nature **528**(7580): 99-104.

Berg, J. M., et al. (2006). Biochemistry: international edition, WH Freeman & Company Limited.

Biggs, M. B. and J. A. Papin (2013). "Novel multiscale modeling tool applied to Pseudomonas aeruginosa biofilm formation." PloS one **8**(10): e78011.

Blattner, F. R., et al. (1997). "The complete genome sequence of Escherichia coli K-12." science **277**(5331): 1453-1462.

Bolouri, H. (2008). Computational modeling of gene regulatory networks: a primer, World Scientific.

Briggs, K. A., et al. (1984). "Molecular cloning, DNA structure and expression of the Escherichia coli D-xylose isomerase." The EMBO journal **3**(3): 611.

Bulawa, C. E. and C. R. Raetz (1984). "The biosynthesis of gram-negative endotoxin. Identification and function of UDP-2, 3-diacylglucosamine in Escherichia coli." Journal of Biological Chemistry **259**(8): 4846-4851.

Calhoun, M. W., et al. (1993). "Energetic efficiency of Escherichia coli: effects of mutations in components of the aerobic respiratory chain." Journal of bacteriology **175**(10): 3020-3025.

Cayley, S., et al. (1991). "Characterization of the cytoplasm of Escherichia coli K-12 as a function of external osmolarity: implications for protein-DNA interactions in vivo." Journal of molecular biology **222**(2): 281-300.

Chaganti, S. R., et al. (2011). "Flux balance analysis of mixed anaerobic microbial communities: effects of linoleic acid (LA) and pH on biohydrogen production." international journal of hydrogen energy **36**(21): 14141-14152.

Cheng, J. J. and G. R. Timilsina (2011). "Status and barriers of advanced biofuel technologies: A review." Renewable Energy **36**(12): 3541-3549.

Cheung, C. Y., et al. (2013). "A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions." The Plant Journal **75**(6): 1050-1061.

Cornish-Bowden, A. (2012). Fundamentals of enzyme kinetics, Wiley-Blackwell Weinheim, Germany.

Cornish-Bowden, A. and M. L. Cárdenas (2001). "Information transfer in metabolic pathways." The FEBS Journal **268**(24): 6616-6624.

Daegelen, P., et al. (2009). "Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21 (DE3)." Journal of molecular biology **394**(4): 634-643.

Davis, E. O. and P. J. Henderson (1987). "The cloning and DNA sequence of the gene xylE for xylose-proton symport in Escherichia coli K12." Journal of Biological Chemistry **262**(29): 13928-13932.

de Oliveira Dal'Molin, C. G., et al. (2010). "AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis." Plant physiology **152**(2): 579-589.

Dempsey, W. B., et al. (1987). "Escherichia coli and Salmonella typhimurium: cellular and molecular biology."

Dennis, P. P. and H. Bremer (1974). "Macromolecular composition during steady-state growth of Escherichia coli B/r." Journal of bacteriology **119**(1): 270-281.

Desai, T. A. and C. V. Rao (2010). "Regulation of arabinose and xylose metabolism in Escherichia coli." Applied and environmental microbiology **76**(5): 1524-1532.

Di Luccio, E., et al. (2007). "Structural and kinetic studies of induced fit in xylulose kinase from Escherichia coli." Journal of molecular biology **365**(3): 783-798.

Di Pasqua, R., et al. (2006). "Changes in membrane fatty acids composition of microbial cells induced by addiction of thymol, carvacrol, limonene, cinnamaldehyde, and eugenol in the growing media." Journal of agricultural and food chemistry **54**(7): 2745-2749.

Dietzler, D. N., et al. (1973). "Rates of glycogen synthesis and the cellular levels of ATP and FDP during exponential growth and the nitrogen-limited stationary phase of Escherichia coli W4597 (K)." Archives of biochemistry and biophysics **156**(2): 684-693.

Dunn, D. B. and J. D. Smith (1958). "The occurrence of 6-methylaminopurine in deoxyribonucleic acids." Biochemical Journal **68**(4): 627.

Edwards, J. S., et al. (2002). "Metabolic modelling of microbes: the flux-balance approach." Environmental microbiology **4**(3): 133-140.

Edwards, J. S., et al. (2001). "In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data." Nature biotechnology **19**(2): 125-130.

Edwards, J. S. and B. O. Palsson (2000). "The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities." Proceedings of the National Academy of Sciences **97**(10): 5528-5533.

El-Mansi, M. (2004). "Flux to acetate and lactate excretions in industrial fermentations: physiological and biochemical implications." Journal of Industrial Microbiology and Biotechnology **31**(7): 295-300.

Farmer, I. S. and C. W. Jones (1976). "The energetics of Escherichia coli during aerobic growth in continuous culture." The FEBS Journal **67**(1): 115-122.

Feist, A. M., et al. (2007). "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information." Molecular systems biology **3**(1).

Feist, A. M. and B. O. Palsson (2010). "The biomass objective function." Current opinion in microbiology **13**(3): 344-349.

Fell, D. (1997). Understanding the control of metabolism, Portland press London.

Fritsche, U. R., et al. (2010). "Direct and indirect land-use competition issues for energy crops and their sustainable production–an overview." Biofuels, Bioproducts and Biorefining **4**(6): 692-704.

Ghuysen, J.-M. (1968). "Use of bacteriolytic enzymes in determination of wall structure and their role in cell metabolism." Bacteriological reviews **32**(4 Pt 2): 425.

Godin, M., et al. (2010). "Using buoyant mass to measure the growth of single cells." Nature methods **7**(5): 387-390.

Gompertz, B. (1825). "XXIV. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. FRS &c." Philosophical transactions of the Royal Society of London **115**: 513-583.

Groff, D., et al. (2012). "Supplementation of intracellular XylR leads to coutilization of hemicellulose sugars." Applied and environmental microbiology **78**(7): 2221-2229.

Gunawardena, J. (2002). "Notes on metabolic control analysis." Bauer Center for Genomics Research, Harvard University, Cambridge, MA **2138**.

Görke, B. and J. Stülke (2008). "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients." Nature Reviews Microbiology **6**(8): 613-624.

Hanly, T. J. and M. A. Henson (2011). "Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures." Biotechnology and bioengineering **108**(2): 376-385.

Hanly, T. J., et al. (2012). "Dynamic flux balance modeling of S. cerevisiae and E. coli co-cultures for efficient consumption of glucose/xylose mixtures." Applied microbiology and biotechnology **93**(6): 2529-2541.

Hasona, A., et al. (2004). "Pyruvate formate lyase and acetate kinase are essential for anaerobic growth of Escherichia coli on xylose." Journal of bacteriology **186**(22): 7593-7600.

Henson, M. A. and T. J. Hanly (2014). "Dynamic flux balance analysis for synthetic microbial communities." IET systems biology **8**(5): 214-229.

Hernández-Montalvo, V., et al. (2001). "Characterization of sugar mixtures utilization by an Escherichia coli mutant devoid of the phosphotransferase system." Applied microbiology and biotechnology **57**(1-2): 186-191.

Hill, J., et al. (2006). "Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels." Proceedings of the National Academy of sciences **103**(30): 11206-11210.

Holzhütter, H. G. (2004). "The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks." The FEBS Journal **271**(14): 2905-2922.

Huang, X., et al. (2013). "Effects of nitrogen supplementation of the culture medium on the growth, total lipid content and fatty acid profiles of three microalgae (Tetraselmis subcordiformis, Nannochloropsis oculata and Pavlova viridis)." Journal of applied phycology **25**(1): 129-137.

Ibarra, R. U., et al. (2002). "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth." Nature **420**(6912): 186-189.

Iglesias, P. A. and B. P. Ingalls (2010). Control theory and systems biology, MIT Press.

Jansson, P. E., et al. (1981). "Structural studies on the hexose region of the core in lipopolysaccharides from Enterobacteriaceae." European journal of biochemistry **115**(3): 571-577.

Jouhten, P., et al. (2012). "Dynamic flux balance analysis of the metabolism of Saccharomyces cerevisiae during the shift from fully respirative or respirofermentative metabolic states to anaerobiosis." FEBS Journal **279**(18): 3338-3354.

Kang, H. Y., et al. (1998). "Priority of pentose utilization at the level of transcription: arabinose, xylose, and ribose operons." Molecules and cells **8**(3): 318-323.

Karp, P. D., et al. (2002). "The ecocyc database." Nucleic acids research **30**(1): 56-58.

Keseler, I. M., et al. (2005). "EcoCyc: a comprehensive database resource for Escherichia coli." Nucleic acids research **33**(suppl 1): D334-D337.

Khandelwal, R. A., et al. (2013). "Community flux balance analysis for microbial consortia at balanced growth." PloS one **8**(5): e64567.

Khankal, R., et al. (2008). "Role of xylose transporters in xylitol production from engineered Escherichia coli." Journal of biotechnology **134**(3): 246-252.

Khankal, R., et al. (2009). "Transcriptional effects of CRP* expression in Escherichia coli." Journal of biological engineering **3**(1): 1.

Kim, B. H. and G. M. Gadd (2008). Bacterial physiology and metabolism, Cambridge university press.

Kim, J.-H., et al. (2010). "Simultaneous consumption of pentose and hexose sugars: an optimal microbial phenotype for efficient fermentation of lignocellulosic biomass." Applied microbiology and biotechnology **88**(5): 1077-1085.

Kitagawa, M., et al. (2006). "Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research." DNA research **12**(5): 291-299.

Kitano, H. (2002). "Computational systems biology." Nature **420**(6912): 206-210.

Kitano, H. (2002). "Systems biology: a brief overview." Science **295**(5560): 1662-1664.

Klipp, E., et al. (2016). Systems biology: a textbook, John Wiley & Sons.

Koirala, S., et al. (2016). "Reciprocal Regulation of l-Arabinose and d-Xylose Metabolism in Escherichia coli." Journal of bacteriology **198**(3): 386-393.

Kovárová-Kovar, K. and T. Egli (1998). "Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics." Microbiology and molecular biology reviews **62**(3): 646-666.

Lam, V. M., et al. (1980). "Proton-linked D-xylose transport in Escherichia coli." Journal of bacteriology **143**(1): 396-402.

Lee, K. H., et al. (2007). "Systems metabolic engineering of Escherichia coli for L-threonine production." Molecular systems biology **3**(1): 149.

Lee, S. K., et al. (2008). "Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels." Current opinion in biotechnology **19**(6): 556-563.

Lee, S. Y. (1996). "High cell-density culture of Escherichia coli." Trends in biotechnology **14**(3): 98-105.

Lee, S. Y. (2009). Systems biology and biotechnology of Escherichia coli, Springer.

Liao, J. C., et al. (2016). "Fuelling the future: microbial engineering for the production of sustainable biofuels." Nature Reviews Microbiology **14**(5): 288-304.

Liu, T. and C. Khosla (2010). "Genetic engineering of Escherichia coli for biofuel production." Annual review of genetics **44**: 53-69.

Lukjancenko, O., et al. (2010). "Comparison of 61 sequenced Escherichia coli genomes." Microbial ecology **60**(4): 708-720.

Mahadevan, R., et al. (2002). "Dynamic flux balance analysis of diauxic growth in Escherichia coli." Biophysical journal **83**(3): 1331-1340.

Marr, A. G., et al. (1963). "The maintenance requirement of Escherichia coli." Annals of the New York Academy of Sciences **102**(3): 536-548.

McCloskey, D., et al. (2013). "Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli." Molecular systems biology **9**(1): 661.

Meadows, A. L., et al. (2010). "Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation." Metabolic engineering **12**(2): 150-160.

Milo, R., et al. (2010). "BioNumbers—the database of key numbers in molecular and cell biology." Nucleic acids research **38**(suppl 1): D750-D753.

Monod, J. (1949). "The growth of bacterial cultures." Annual Reviews in Microbiology **3**(1): 371-394.

Morris, D. R. and C. M. Jorstad (1970). "Isolation of conditionally putrescine-deficient mutants of Escherichia coli." Journal of bacteriology **101**(3): 731-737.

Naik, S. N., et al. (2010). "Production of first and second generation biofuels: a comprehensive review." Renewable and Sustainable Energy Reviews **14**(2): 578-597.

Nakamura, C. E. and G. M. Whited (2003). "Metabolic engineering for the microbial production of 1, 3-propanediol." Current opinion in biotechnology **14**(5): 454-459.

Neidhardt, F. C., et al. (1990). "Physiology of the bacterial cell: a molecular approach."

Nelson, D. L., et al. (2008). Lehninger principles of biochemistry, Macmillan.

Ni, L., et al. (2012). "Structures of the Escherichia coli transcription activator and regulator of diauxie, XylR: an AraC DNA-binding family member with a LacI/GalR ligand-binding domain." Nucleic acids research **41**(3): 1998-2008.

Ni, L., et al. (2013). "Structures of the Escherichia coli transcription activator and regulator of diauxie, XylR: an AraC DNA-binding family member with a LacI/GalR ligand-binding domain." Nucleic acids research **41**(3): 1998-2008.

O'Brien, E. J., et al. (2013). "Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction." Molecular systems biology **9**(1): 693.

Orth, J. D., et al. (2011). "A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011." Molecular systems biology **7**(1): 535.

Orth, J. D., et al. (2010). "What is flux balance analysis?" Nat Biotechnol **28**(3): 245-248.

Parry, B. R., et al. (2014). "The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity." Cell **156**(1): 183-194.

Pirt, S. J. (1965). "The maintenance energy of bacteria in growing cultures." Proceedings of the Royal Society of London B: Biological Sciences **163**(991): 224-231.

Pirt, S. J. (1982). "Maintenance energy: a general model for energy-limited and energy-sufficient growth." Archives of Microbiology **133**(4): 300-302.

Poolman, M. G., et al. (2009). "A genome-scale metabolic model of Arabidopsis and some of its properties." Plant physiology **151**(3): 1570-1581.

Powell, E. O. (1956). "Growth rate and generation time of bacteria, with special reference to continuous culture." Microbiology **15**(3): 492-511.

Pramanik, J. and J. D. Keasling (1997). "Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements." Biotechnology and bioengineering **56**(4): 398-421.

Randle, C. L., et al. (1969). "The phosphoglyceride composition of Gram-negative bacteria and the changes in composition during growth." Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism **187**(2): 214-220.

Reed, J. L., et al. (2003). "An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR)." Genome Biol **4**(9): R54.

Richards, F. J. (1959). "A flexible growth function for empirical use." Journal of experimental Botany **10**(2): 290-301.

Roberts, R. B., et al. (1955). "Studies of Biosynthesis in." Escherichia coli: 607.

Robison, K., et al. (1998). "A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K-12 genome1." Journal of molecular biology **284**(2): 241-254.

Rudin, W. (1964). Principles of mathematical analysis, McGraw-hill New York.

Schellenberg, G. D., et al. (1984). "Xylose isomerase from Escherichia coli. Characterization of the protein and the structural gene." Journal of Biological Chemistry **259**(11): 6826-6832.

Schmidt, A., et al. (2016). "The quantitative and condition-dependent Escherichia coli proteome." Nature biotechnology **34**(1): 104-110.

Schmidt, L. D. and P. J. Dauenhauer (2007). "Chemical engineering: hybrid routes to biofuels." Nature **447**(7147): 914-915.

Schuster, S., et al. (2008). "Is maximization of molar yield in metabolic networks favoured by evolution?" Journal of theoretical biology **252**(3): 497-504.

Scott, M., et al. (2010). "Interdependence of cell growth and gene expression: origins and consequences." Science **330**(6007): 1099-1102.

Segre, D., et al. (2002). "Analysis of optimality in natural and perturbed metabolic networks." Proceedings of the National Academy of Sciences **99**(23): 15112-15117.

Sezonov, G., et al. (2007). "Escherichia coli physiology in Luria-Bertani broth." Journal of bacteriology **189**(23): 8746-8749.

Shaw, M. K. and J. L. Ingraham (1965). "Fatty acid composition of Escherichia coli as a possible controlling factor of the minimal growth temperature." Journal of bacteriology **90**(1): 141-146.

Singh, A., et al. (2011). "Manipulating redox and ATP balancing for improved production of succinate in E. coli." Metabolic engineering **13**(1): 76-81.

Smith, J. D. and G. R. Wyatt (1951). "The composition of some microbial deoxypentose nucleic acids." Biochemical Journal **49**(2): 144.

Sofia, H. J., et al. (1994). "Analysis of the Escherichia coli genome. V. DNA sequence of the region from 76.0 to 81.5 minutes." Nucleic acids research **22**(13): 2576-2586.

Song, S. and C. Park (1997). "Organization and regulation of the D-xylose operons in Escherichia coli K-12: XylR acts as a transcriptional activator." Journal of bacteriology **179**(22): 7025-7032.

Stein, W. D. and T. Litman (2014). Channels, carriers, and pumps: an introduction to membrane transport, Elsevier.

Stewart, E. J., et al. (2005). "Aging and death in an organism that reproduces by morphologically symmetric division." PLoS Biol **3**(2): e45.

Stolyar, S., et al. (2007). "Metabolic modeling of a mutualistic microbial community." Molecular systems biology **3**(1): 92.

Stouthamer, A. H. and C. W. Bettenhaussen (1975). "Determination of the efficiency of oxidative phosphorylation in continuous cultures of Aerobacter aerogenes." Archives of microbiology **102**(1): 187-192.

Sumiya, M., et al. (1995). "Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in Escherichia coli." Receptors & channels **3**(2): 117-128.

Sumiya, M. and H. P. Jf (1989). "The D-xylose binding protein of Escherichia coli." Biochemical Society Transactions **17**(3): 553-554.

Tao, H., et al. (2001). "Engineering a homo-ethanol pathway inEscherichia coli: Increased glycolytic flux and levels of expression of glycolytic genes during xylose fermentation." Journal of bacteriology **183**(10): 2979-2988.

Taylor, A. R. (1946). "Chemical analysis of the T2 bacteriophage and its host, Escherichia coli (strain B)." Journal of Biological Chemistry **165**(1): 271-284.

Tempest, D. W. (1978). "The biochemical significance of microbial growth yields: a reassessment." Trends in Biochemical Sciences **3**(3): 180-184.

Thiele, I. and B. Ø. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols **5**(1): 93-121.

Umbarger, H. E. (1977). "A one-semester project for the immersion of graduate students in metabolic pathways." Biochemical Education **5**(4): 67-71.

Umemoto, Y., et al. (2012). "D-Xylose isomerase from a marine bacterium, Vibrio sp. strain XY-214, and D-xylulose production from β-1, 3-xylan." Marine biotechnology **14**(1): 10-20.

Upton, D. J., et al. (2017). "An accurate description of Aspergillus niger organic acid batch fermentation through dynamic metabolic modelling." Biotechnology for biofuels **10**(1): 258.

Van Bodegom, P. (2007). "Microbial maintenance: a critical review on its quantification." Microbial Ecology **53**(4): 513-523.

Varma, A., et al. (1993). "Biochemical production capabilities of Escherichia coli." Biotechnology and bioengineering **42**(1): 59-73.

Varma, A., et al. (1993). "Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates." Applied and environmental microbiology **59**(8): 2465-2473.

Varma, A. and B. O. Palsson (1993). "Metabolic capabilities of Escherichia coli II. Optimal growth patterns." Journal of Theoretical Biology **165**(4): 503-522.

Varma, A. and B. O. Palsson (1993). "Metabolic capabilities of Escherichia coli: I. Synthesis of biosynthetic precursors and cofactors." Journal of theoretical biology **165**(4): 477-502.

Varma, A. and B. O. Palsson (1994). "Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use." Bio/technology **12**.

Varma, A. and B. O. Palsson (1994). "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110." Applied and environmental microbiology **60**(10): 3724-3731.

Verhulst, P.-F. (1838). "Notice sur la loi que la population suit dans son accroissement. correspondance mathématique et physique publiée par a." Quetelet **10**: 113-121.

Voet, D. and J. G. Voet (2004). "Biochemistry. Hoboken." John Wiley & Sons **1**: 591.

Wang, G. and W. M. Post (2012). "A theoretical reassessment of microbial maintenance and implications for microbial ecology modeling." FEMS microbiology ecology **81**(3): 610-617.

Wolfe, A. J. (2005). "The acetate switch." Microbiology and molecular biology reviews **69**(1): 12-50.

Yamamotoya, T., et al. (2012). "Glycogen is the primary source of glucose during the lag phase of E. coli proliferation." Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics **1824**(12): 1442-1448.

Yuan, H., et al. (2016). "Flux balance analysis of plant metabolism: The effect of biomass composition and model structure on model predictions." Frontiers in plant science **7**: 537.