Chapter 5

Optimal metabolic fluxes in constraint-based models

Daan de Groot, David Tourigny, Felipe Scott, and Jan-Pilipp Dapprich

Chapter overview

- An optimization objective can be added to constraint-based models to make more specific predictions.
- Different purposes can be served by choosing different optimization objectives and constraints
- The optimal solutions can be understood in terms of elementary flux modes

5.1. Can we use an optimality assumption to predict metabolic behavior?

In the previous chapter, we characterized an organism's metabolism by listing all the biochemical reactions that can be catalyzed by the enzymes encoded within the organism's genome. To understand how the genome constrains patterns of metabolic flux we needed to make several simplifying assumptions. The first important assumption was that intracellular metabolism is at steady-state, i.e., that the production and consumption of all metabolites is balanced such that their concentrations are constant in time. These resulted in the mass-balance constraints on the flux vector $\mathbf{v}$. The flux cone of all flux vectors satisfying the mass-balance constraints could be further reduced by additional constraints on $\mathbf{v}$, based on extra physical and biological assumptions about the magnitude and directionality of certain reactions within the network. We introduced several ways in which the entire flux space could be explored.

When applied to very large metabolic networks, the flux space will often contain an infinite number of flux vectors $\mathbf{v}$ that simultaneously satisfy all constraints. From a mathematical perspective, this implies that the constraints do not include enough information to uniquely specify a flux vector $\mathbf{v}$. This makes sense biologically, since if we imagine constraints are related to experimental observations it is very unlikely that we will ever be able to make enough to fully account for every reaction encoded within the entire genome of an organism (no matter how simple it might be). Often, however, researchers do want to further narrow down the set of flux vectors that they think biologically relevant to the organism and conditions they are studying, perhaps even to a unique $\mathbf{v}$ imagined to describe the metabolic state of an organism at a given moment in time. One popular approach for doing so is to provide an additional assumption (or set thereof) in the form of an objective function: it is assumed that the metabolic state of an organism is such that some function of $\mathbf{v}$ (e.g. growth rate) is maximized to satisfy some criteria (e.g. evolutionary selective advantage). The

computational problem then becomes one of constrained-optimization: find a flux vector $\mathbf{v}$ that is optimal in terms of the objective function(s) that simultaneously satisfies all constraints. The resulting space of optimal flux vectors (sometimes containing just one unique vector) is often considerably smaller than the space of those that satisfy only the constraints.

In this chapter, we will study metabolic models based on constrained-optimization. We will introduce a selection of commonly used objective functions and the computational methods used to solve the associated constrained-optimization problem. We will also characterize optimal solutions that we get in terms of the minimal metabolic strategies that we identified in the previous chapter: elementary flux modes. Finally, we will explain how we can handle the cases where the solutions are, even after optimization, not unique.

5.2. Metabolic models based on linear optimization problems

In the previous chapter, we described how *linear* homogeneous and inhomogeneous constraints arising from biological and physical knowledge can be combined into matrix and vector notation and written in the general form presented in Equations (4.11) and (4.12). The resulting space of all flux vectors $\mathbf{v}$ satisfying these constraints is called the flux polyhedron. The flux polyhedron can remain high-dimensional and, as explained above, an objective function f can be used to narrow down the set of flux vectors to only those that are optimal (i.e., maximize the objective function). The general form in which we can write the resulting constraint-based optimization problem is therefore:

$$\max_{\mathbf{v}} f(\mathbf{v}), \text{ such that } \mathbf{A}\mathbf{v} \geq \mathbf{b}, \tag{5.1}$$

with

$$\mathbf{A} = \begin{pmatrix} \mathbf{N} \\ -\mathbf{N} \\ \mathbf{I} \\ \mathbf{G} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{h} \end{pmatrix}. \tag{5.2}$$

Recall that $\mathbf{N}\mathbf{v} = \mathbf{0}$ models the steady-state assumption, while the multiplication with the identity matrix ($\mathbf{I}_{n\times n}\mathbf{v} \geq \mathbf{0}$) captures the fact that we forced all reactions to be irreversible by splitting reversible reactions into a forward and a backward reaction. Finally, $\mathbf{G}\mathbf{v} \geq \mathbf{h}$ can be used to impose additional 'inhomogeneous' constraints that can be used to input additional biological knowledge such as an experimentally measured upper bound on the uptake rate of a certain nutrient.

In many cases, the objective function is chosen to be a linear function of the fluxes, i.e.,

$$f(\mathbf{v}) = \sum_i c_i v_i, \tag{5.3}$$

where coefficients c_i weigh the relevance of the different reaction rates in the objective function. Problems of the form (5.1), (5.2), and (5.3) in general are called *linear programming problems* and as the name suggests can be solved using *linear programming*. Applied to metabolic models, linear programming is called *Flux Balance Analysis* (FBA). Linear programming problems are well studied, such that FBA is perhaps the most popular approach to genome-scale metabolic models [1, 2]. FBA problems are relatively easy to solve using specialized optimization software, which have been highly developed due to the general applicability of linear programming in economics, logistics, and many other fields also. In the following subsections we will briefly describe various choices that can be made for the linear objective function $f(\mathbf{v})$ in FBA.

As an example FBA problem, in Figure 5.1 we have extended the minimal example from the previous chapter

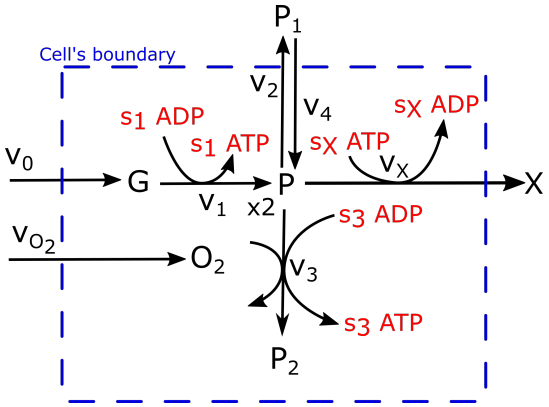

Figure 5.1: A simple representation of the metabolic reaction network for central carbon metabolism. Extracellular glucose is imported into the cell via a reaction with flux v_0 and converted via intracellular glucose, G, to pyruvate, P, via the reaction with flux v_1 that has a stoichiometric coefficient of two pyruvate molecules to each glucose molecule. Pyruvate can then either be converted to a fermentation product, P_1, via the reaction with flux v_2 or, in the presence of oxygen, O_2 imported via v_{O_2}, converted to an oxidative phosphorylation (OXPHOS) terminal product P_2 via the reaction with flux v_3. It can also be converted to biomass X with rate v_X. The reactions with flux values v_1 and v_3 produce ATP from ADP (in red) with stoichiometry s_1 and s_3, respectively, which can vary between species. The production of 1.0 grams of new cells, in a dry weight basis, requires one molecule of pyruvate and s_X molecules of ATP.

to include ATP and biomass (X) production, assuming the latter is produced from pyruvate using a single reaction that consumes s_X molecules of ATP with flux value v_X. We also introduce as a linear objective function the total rate of ATP production, v_{ATP}. Since in this example, reactions v_1 and v_3 produce ATP with stoichiometric coefficients s_1 and s_3, respectively, the total rate of ATP production is given by $v_{ATP} = s_1 v_1 + s_3 v_3 - s_X v_X$. The FBA problem is then given by simply maximizing v_{ATP} subject to $v_0, v_{O_2}, v_1, v_2, v_3, v_4, v_X$ satisfying the mass-balance constraints but, as we will see in the next subsection, this would result in a problem that is unbounded: the flux vectors and resulting optimal value of v_{ATP} could be indefinitely large. Biologically, this is because there are no bounds on the uptake rates of glucose v_0^{ub} and the fermentation product v_4^{ub}. Thus, if we re-impose these bounds as in the last chapter, the result is an FBA problem that is bounded and therefore has a finite objective value:

$$
\begin{aligned}
&\max_{\mathbf{v}} v_{ATP} = s_1 v_1 + s_3 v_3 - s_X v_X, \quad \text{such that :} \\
&0 = v_0 - v_1, \\
&0 = v_{O_2} - v_3, \\
&0 = 2v_1 - v_2 - v_3 + v_4 - v_X, \\
&v_0^{ub} \geq v_0, \\
&v_4^{ub} \geq v_4, \\
&v_0, v_1, v_2, v_3, v_4 \geq 0.
\end{aligned}
\tag{5.4}
$$

To illustrate a particular instance of this FBA problem, we consider the very simple case where $v_4^{ub} = 0$, $v_0^{ub} > 0$ and $s_3 = s_1 = 1$. It can be checked by hand that an optimal solution is given by $v_0 = v_1 = v_2/2 = v_0^{ub}$, with $v_2 = v_4 = v_X = 0$. The optimal objective value is given by $v_{ATP} = 3v_0^{ub}$.

Economic analogy 5.A : Linear programming and economic planning in the Soviet Union

Linear programming as an algorithmic approach to solving constrained linear optimization problems was first developed by soviet mathematician and economist Leonid Kantorovich in the 1930s [3, 4]. Kantorovich was tasked with helping to optimize production in the soviet plywood industry, but soon discovered that the underlying problems could not be solved using analytical methods. He instead developed a method for solving linear optimization problems using an iterative process through which a solution is continuously improved until an optimum is reached. Kantorovich argued that this could be used to make soviet economic planning more efficient.

Soviet planning was primarily based on material balancing, which aimed to create a consistent plan with regards to the inputs and outputs of various industries. For example, the input requirement of steel consuming industries ought not to exceed steel production targets. In a balanced plan the input requirements for steel would match the production of steel. But a balanced plan is not necessarily an optimal one. There can well be several consistent plans of which some lead to higher overall production output than others. Kantorovich observed that productive resources were often not used where they could yield the greatest benefit. By using linear programming, planners could in principle calculate a plan that made the best use of economic resources and maximized production output.

One of the problems that needed to be overcome by Kantorovich was that optimization always aims to optimize a singular objective function. However, there was no obvious way of measuring the output of qualitatively distinct products on a single scale. Without prior valuation of the products (for example through market prices) it is not clear whether 3 tanks and 10 trucks should be counted as more than 4 tanks and 8 trucks. Kantorovich circumvented this problem by assuming that outputs ought to be produced at given proportions. For example, it might be specified that 2 trucks ought to be produced for every tank. Linear programming can then be used to calculate the plan that maximizes output at these proportions. Unlike most contemporary economic applications of linear programming, this does not depend on a monetary objective function. So, whats being maximized is not monetary value. Instead, the objective function measures purely physical quantities (such as number of trucks or tons of steel).

In the context of economic planning, constraints are used to represent limits to available economic resources (such as fertile land). A plan that uses more resources than are available will not be feasible and must thus be excluded. Constraints can also be used to fix the proportions at which distinct outputs ought to be produced [5]. While it was first developed for economic planning, the fundamental principles of linear programming can also be applied to other problems (for example in biology).

5.2.1. Types of linear objective functions used in FBA

Solving the constraint-based optimization problem of (5.1) will reduce the set of flux vectors to those that are optimal (maximize the objective function), but the biological validity of this prediction is critically dependent on the particular choice of f. Consequently, there has been a lot of consideration and debate among researchers working on FBA about the appropriate objective functions to use in different contexts and how best interpret the results. Below, we will provide some popular examples, but for a more systematic comparison of different objective functions we refer the reader to [6, 7, 8].

Evolutionary justifications for objective functions: the rate of biomass production Objective functions are often based on evolutionary arguments: the objective is chosen to capture some proxy for the evolutionary fitness of an organism. The motivation behind this is that cells with a metabolic state that scores well on this fitness-proxy would come to dominate the cell-population because they outgrow their competitors. Proxies for fitness are in principle very hard to choose since evolutionary fitness is mostly related to the average net reproduction rate of a cell over a very long time[9]. Therefore, to know the metabolic objective that

aligns with the maximization of fitness would require us to know what the cell has been selected for in its evolutionary history. This is a non-trivial question, for example, is an *E. coli* cell growing in the human gut selected for the same metabolic objective as a muscle cell in your body?

An objective that is used very often is the maximization of a biomass production rate, because this is used as a proxy for maximizing growth rate. It is indeed arguable that unicellular organisms with high growth rates are selected, since in stationary conditions these cells will come to dominate the population. Indeed, FBA models in which the biomass production rate is optimized seem to predict metabolic states reasonably well [10, 11, 12].

But what exactly do we mean by biomass? This is extensively discussed in the Chapter 2, but for our purposes it is sufficient to say that it is the entirety of all components that constitute a new cell. In metabolic models, however, biomass refers to all precursors that are outputs of the model and that are needed to produce a new cell. This has two consequences. First, biomass in our model does not only consist of the components of which the cell is built, but also of components needed to do the building itself, such as a certain amount of ATP. Second, what is contained in biomass will depend on where we draw a line around the metabolic network - all necessary cell components that are not inside are regarded as biomass. In practice, biomass appears in metabolic networks in the form of a virtual *biomass reaction* that consumes all necessary precursor molecules in the right proportions and produces one unit of standard biomass. Maximizing the biomass production rate thus takes the very simple form of just maximizing the rate through the biomass reaction.

The use of such a fixed biomass reaction represents an important assumption, because in reality the biomass composition will be condition dependent. For example, if a cell grows faster and contains more ribosomes, this increases the cellular fraction of proteins and polynucleotides, and hence the need for the respective precursors (amino acids and nucleotides). Moreover, biomass composition can even depend on the choice of metabolic strategy. If a pathway includes enzymes that contain a lot of iron, then depending on the flux solution (which uses this pathway or not), more or less iron will be contained in the biomass. So, the flux solution must be known to know the biomass composition, but the biomass composition must also be known to get to a flux solution. To resolve this, we would need a model of the entire cell, including the synthesis reactions of all enzymes. Such models will be discussed later, in the Chapter 9 on large cell models.

Evolutionary justifications for objective functions: alternative fitness-proxies In some cases, modeling the maximization of the instantaneous growth rate through the biomass reaction is an unrealistic proxy of the evolutionary fitness. For example, in multicellular organisms each cell performs a task that contributes to the fitness of the whole organism, but this is not related to the reproduction rate of the individual cells. In those cases, we may still try to capture an evolutionary objective when we know the main task of the cell-type. For example, beta-cells in the pancreas have as their main task to produce insulin, and we may thus model their metabolism by maximizing the production of insulin.

In other cases, our metabolic model is focused only on a very small part of the true metabolic network, and therefore does not model the production of all biomass precursors. In such cases, energy production rate in the form of ATP production rate is often maximised. Yet other objective functions that are sometimes used and have a (somewhat vague) evolutionary motivation are the minimization of overall ATP usage and the minimization of overall fluxes.

Synthetic design-oriented objective functions Metabolic modeling can also be used to identify metabolic states that lead to a certain desired behavior of a microorganism. For example, we may seek to genetically perturb a microbe such that it produces a certain compound of industrial or medicinal interest, while it also retains a certain minimal growth rate [13]. Indeed, it is often desired to retain a certain minimal

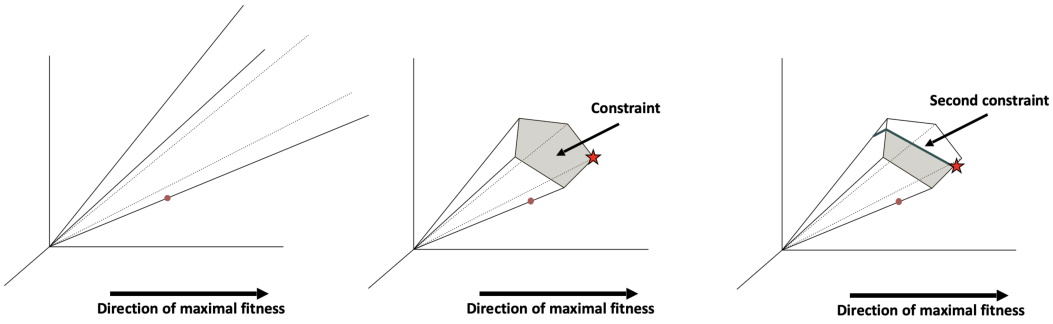

Figure 5.2: We show a cartoon of the solution space of a metabolic network, the so-called flux cone, with respectively zero, one and two constraints. With one constraint, the optimal solution for any linear objective can be attained in a vertex of the space, which means that it can be attained in a single EFM. With two constraints, we need to combine at most two EFMs to describe the optimal solution.

ability to grow such that the genetically engineered organisms can be lab-grown after which the produced compound of interest can be harvested. In that case, we can combine maximizing the production rate of the compound while imposing an inhomogeneous constraint that sets a lower bound on the biomass production rate. This can even be combined with a calculation in which we solely maximize the biomass production rate: maximizing the biomass production rate is a model for the wild-type cell, whereas maximizing the generation of the compound models the desired phenotype. By comparing the flux distributions between these 'strains', we can search for target genes that should be up- or downregulated.

5.3. Optimal metabolism in terms of elementary flux modes

In the previous chapter we introduced elementary flux modes (EFMs) and identified them as the fundamental metabolic pathways that carry flux through the metabolic reaction network. Here, we will show how elementary flux modes also can be very useful for describing optimal metabolic states. We briefly recapitulate the notion of elementary flux modes. All metabolic flux vectors $\mathbf{v}$ that satisfy both the mass-balance and irreversibility constraints form a pointed polyhedral cone, called the flux cone. The EFMs are the extreme rays of this cone, so that they can be used to decompose all steady-state flux vectors:

$$\mathbf{v} = \sum_i \lambda_i \mathbf{e}_i,$$

where $\lambda_i \geq 0$ and $\mathbf{e}_i$ is the i-th EFM. Moreover, the EFMs turn out to be the minimal metabolic subnetworks that a cell can use in steady-state without needing any other reaction, so that we can view EFMs as minimal metabolic strategies.

In Figure 5.2 we depict the EFMs as black lines, and the region in-between these lines is the steady-state solution space that is spanned by the EFMs. Note that this illustration is great simplification, usually the flux cone is a high-dimensional object that can only be visualized in trivial toy examples. In fact, the flux cone is a subspace of $\mathbb{R}^n$ where n (the number of reactions) can be in the thousands for a typical genome-scale metabolic network. Moreover, the number of extreme rays of the cone would be overwhelming, due to the complexity issues associated with EFM enumeration as described in the previous chapter.

Figure 5.2**a** also shows that there is a direction in which the objective increases fastest. This direction is determined by the choice of objective function, to be specific: the direction of maximal increase of the objective is given by the vector of coefficients, $[c_1 \cdots c_n]^T$, appearing in the linear objective function (5.3).

However, as long as we do not impose an inhomogeneous constraint, the flux cone is unbounded, so that we can usually reach infinite values. This makes sense when we think of the metabolic states in terms of elementary flux modes: when we have an EFM that reaches some nonzero objective value, we can always multiply it by any positive scalar. This multiplication will increase the objective value, while the steady-state and irreversibility constraints will not be affected.

Metabolism, however, is never unconstrained, so we will always have at least one inhomogeneous constraint. In the previous chapter, inhomogeneous constraint were written in the general form

$$\sum_i w_i^p v_i \leq h_p, \quad p = 1, \dots P \tag{5.5}$$

where each h_p corresponds to a component of the P-dimensional vector $\mathbf{h}$ and n weights w_i^p $(i = 1, \dots, n)$ are supplied for each of the P constraints. The second panel of Figure 5.2 shows how a single inhomogeneous constraint (i.e. the case $P = 1$) can constrain the flux cone and theory dictates an optimal flux vector is found at a vertex of the resulting flux polyhedron, which geometrically corresponds to the intersection of the flux cone and the hyperplane of the inhomogeneous constraint. One particular biological argument for such a constraint is related to resource allocation[14, 15]: only a limited number of macromolecules (proteins, ribosomes, etcetera) fit inside a cell. Since these molecules catalyze reactions, reaction rates are proportional to their concentrations:

$$v_i = e_i g_i(\mathbf{x}_{\text{metab}}), \tag{5.6}$$

where e_i is the concentration of the enzyme that catalyzes reaction i, and $g_i(\mathbf{x}_{\text{metab}})$ is a function that describes enzyme kinetics in a non-linear way that is for most reactions unknown. The resource-allocation constraint then takes the form

$$\sum_i w_i e_i \leq 1, \tag{5.7}$$

where w_i are weights that determine how much of the resources are taken up by one unit of the ith enzyme; these weights can for example be proportional to the volume, the mass, or the number of amino acids of the enzyme. Making a change of variables to express the constraint in terms of fluxes gives:

$$\sum_i \frac{w_i}{g_i} v_i \leq 1, \tag{5.8}$$

such that these resource-allocation constraints again fit the form presented in Equation (5.5). A well-known example of a modeling framework that uses such a constraint is *FBA with macromolecular crowding* (FBAwMC, [16]) where such a constraint arises due to a physical limitation on the number of enzymes contained within the cell.

It is not necessarily always the case that an inhomogenous constraint applies to all EFMs. For example, in a metabolic model of an organism able to grow on multiple carbon sources, many EFMs may remain unbounded. For treatment of these cases, the reader is referred to [17]. Moreover, we may have multiple inhomogenous constraints on flux values as Equation (5.5) suggests. The third panel of Figure 5.2 illustrates how a second inhomogenous constraint can further constrain the solution space where theory implies an optimal flux vector is found on a vertex lying on the edge between two EFMs (as shown in the example in the figure). Imposing additional inhomogenous constraints can therefore lead to the superposition of additional EFMs in the solution. In general, if we consider a constraint-based model with K inhomogeneous constraints it can be proved that an optimal flux vector will be built out of at most K EFMs [17]. We therefore see another important property of EFMs: not only do they form the minimal building blocks that span all

metabolic capabilities of the cell, they are also optimal building blocks. When metabolism is optimized, only few of these EFMs are used. As a result, solutions to linear constraint-based optimizations can usually be rationalized in terms of the properties of the available EFMs [18], for example, a flux balance analysis with only one constraint on a nutrient uptake will just return the EFM with the highest 'yield', i.e. the highest efficiency of making biomass per nutrient.

5.4. Phenotypic phase plane analysis

The analysis of the metabolic response to environmental changes is often sought assuming that there is only one substrate limiting growth (or other metabolic reaction). For example, we could be interested in the growth and ethanol production by *S. cerevisiae* under oxygen limitation in a chemostat. In this experimental setup, every other substrate should be provided in excess, including the carbon and energy source. If no oxygen is supplied, ATP must be produced only using oxidative phosphorylation reactions and a fermentation product, such as ethanol, will be produced. On the other extreme, if enough oxygen is available, a fraction of the carbon source will be completely oxidized, producing ATP via respiration. In both cases, the fraction of the carbon and energy source not used for energy generation will be used for the production of biomass at an specific growth rate equal to the dilution rate of the chemostat.

This behavior can be analyzed using the phenotypic phase plane analysis. To calculate a phenotypic phase plane (PPhP), the uptake fluxes values under analysis, typically the uptakes of oxygen and the carbon source are discretized between their upper and lower values and used to construct a meshgrid containing the 2-D grid coordinates based on the coordinates contained in the discretized vectors of oxygen and carbon uptake fluxes. At each tuple in the 2-D grid a FBA problem is solved after fixing the lower and upper bounds of the corresponding fluxes to the values in the tuple. Figure 5.3.A shows the PPhP of the metabolic network presented in Figure 5.1 with $s_X = 10$, $s_1 = 1$, $s_3 = 4$, $v_4 = 0$, $v_0^{UB} = 10$, and $v_{O_2}^{UB} = 15\,\mathrm{mmol g_{CDW}^{-1} h^{-1}}$.

At zero oxygen uptake, Figure 5.3.A shows that growth is possible reaching a specific growth rate of $1\,h^{-1}$ at the maximum glucose uptake. Notice that the slope of the line connecting the origin of coordinates and the point of the highest growth rate at a glucose uptake of $10\,\mathrm{mmol g_{CDW}^{-1} h^{-1}}$ is 0.1 $\mathrm{g_{CDW}(mmol)^{-1}}$. Biologically, this slope corresponds to the biomass yield on glucose under anaerobic conditions, and in terms of linear programming to the negative of the shadow price defined as:

$$\gamma_i = \frac{-\mathrm{d}z}{\mathrm{d}b_i^v}, \tag{5.9}$$

where z is the objective function optimal value (specific growth rate in this case) and b_i^v corresponds to the violation of a mass balance constraint and is equivalent to the uptake reaction of the i-th metabolite (glucose in this example)[19]. Figure 5.3.C shows that the glucose shadow price is equal to -0.1 at every point in the feasible region of the problem. Figure 5.3.D shows the shadow price values for oxygen uptake. For every unit increase in the oxygen uptake flux, the biomass specific growth rate increases by $0.4\,\mathrm{h}^{-1}$. Thus, the plane of increasing growth rate values in Figure 5.3.A can be described by the equation $0.1v_G + 0.4v_{O_2}$. Concomitantly, as the oxygen uptake increases, the flux of product P1 decreases as more ATP is generated in reaction v_3. For every constant glucose uptake flux, the specific growth rate increases and the production of P1 decreases until the optimally line (red line) is reached in Figure 5.3.A. This line represents the optimal relation between the two metabolic fluxes in the PhPP [19]. In this example, the optimally line represents the combinations of glucose and oxygen uptake fluxes leading to a complete oxidation of the substrate, and thus supporting the maximal biomass yield. Finally, increasing the oxygen consumption beyond the optimally line, at a constant glucose uptake, leads to an infeasible problem since there is no further glucose to be oxidized.

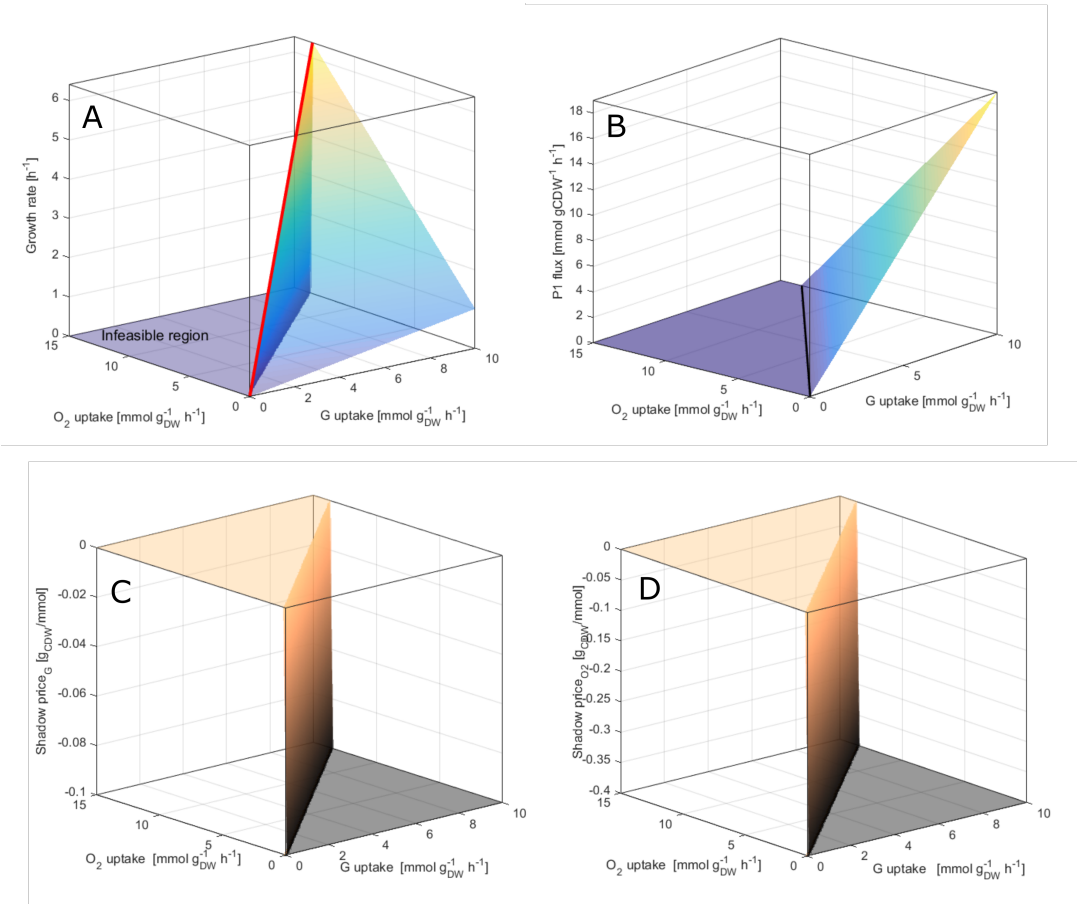

Figure 5.3: Phenotypic phase plane of the metabolic model shown in Figure 5.1, calculated as a function of the uptakes of oxygen and glucose

.

5.5. Non-uniqueness of the optimal metabolic state

Although the optimization of some objective function strongly reduces the number of solutions, it is still possible that many different metabolic states satisfy the constraints *and* reach the same maximal value for the objective. In that case, we are again undecided on which of the solutions gives the most useful information about the biological problem. This non-uniqueness of the optimum can be explained in terms of the elementary flux modes. In the second panel of Figure 5.2 we saw that the optimal solution was located on the vertex that was as far as possible in the optimization direction. One can imagine, however, that the flux cone can be located in the space such that there are two vertices that both reach out equally far into that direction. In that case, the two corresponding elementary flux modes perform equally well, and consequently, all convex combinations of these elementary flux modes also reach the same objective value. In metabolic modeling we often work in a high-dimensional space with constraints that concern only few of those dimensions (for example a bound only on a nutrient uptake rate). In such cases it is very likely that many elementary flux modes perform equally well, so that there is a whole subspace of equivalent solutions.

5.5.1. Flux Variability Analysis

The equality and inequality constraints of the FBA problem form a polytope where the problem is feasible, a cone if the problem is written in canonical form. The optimal solutions of the LP problem can lay on a vertex

of the polytope, and be unique, or be non-unique solutions if the objective function hyperplane is parallel to a facet of the constraint polytope at the solution. This means that one or several variables can change their values without affecting the value of the objective function. These variables can be identified using flux variability analysis (FVA), where each flux of the reactions in the metabolic network (the set of J reactions with N elements and I metabolites) maximized and minimized, one at a time, while fixing the value of the objective function to a fraction of the optimal value obtained in the original FBA problem.

$$\begin{aligned} &\max_{\mathbf{v}} v_j (\text{or} \min_{\mathbf{v}} v_j), \quad \text{such that :} \\ &\sum_{j \in J} S_{i,j} v_j = 0, \ \forall i \in I, \\ &LB_j \leq v_j \leq UB_j, \\ &v_{biomass} = f \cdot v^*_{biomass}, \\ &v_j \in \mathbb{R}, \ \forall j \in J. \end{aligned} \tag{5.10}$$

Hence, $2N$ optimization problems need to be solved if there are N unconstrained fluxes. The results of the VFA analysis should be carefully interpreted. Since the maximum and minimum fluxes are calculated one at a time, and although changes in this flux might not affect the objective function, this typically requires changes in the remaining fluxes. Therefore, the polytope that describes all alternate optimal solutions is not captured by VFA. Instead, FVA inscribes this polytope in the smallest possible "box" [20]. Besides being useful for the identification of alternative solutions, FVA can be utilized to identify blocked reactions under a given growth condition. These reactions are characterized by minimum and maximum flux values (as calculated by VFA) equal to zero and arise due to regulatory constraints imposed to the FBA or due to network gaps, for example, metabolites lacking a consumption or production pathways for whom a steady-state mass balance is impossible. Thus VFA, could help in the identification of dead-end metabolites, and in the long run, in model improvement.

5.6. Limitations of constraint-based metabolic models

In this and the previous chapter, we have introduced constraint-based analysis of genome-scale metabolic models. We started by pointing out many of the simplifying assumptions that are associated with the study of large metabolic reaction networks. For example, we only considered systems in chemical steady state with their environment, we ignored the effects of metabolite dilution, and we made semi-informed choices for which intracellular molecules are contained in our model or summarized in a biomass reaction. All these assumptions can be relaxed, at the cost of making models more complex. Although it is tempting to think that the more complex a model the more realistic it will be, there is not much use to adding additional complexity if we don't have the data to support it. Constraint-based analysis therefore provides one way to study metabolism at genome-scale when data are limited. In the following chapters we will study the consequences of lifting one or more of these simplifying assumptions.

In constraint-based analysis, one considers reaction rates (fluxes) as the variables in the model, giving the illusion that these are directly set by the cell to regulate its metabolic state. In reality, however, the reaction fluxes are the combined consequences of enzyme expression, regulation and metabolite concentrations. If we wish to model metabolism in more detail, we we should build models that incorporate gene expression and metabolite concentrations systematically. Some of the next chapters attempt this, but we have described that FBA is useful when experimental data are limited. Certain extensions of FBA discussed in later chapters

also move beyond the steady state assumption, allowing the environment to change with time. One example is the method dynamic FBA, which will also be discussed in a later chapter.

Philosophical remarks 5.B : Qualities of a model

When have we made a good model? Is the quality of a model determined by whether it fits all experimental observations? What is the ideal size of a model? Is the purpose of a model that it predicts, or rather that it provides insight into the biological processes?
The answers to these questions are as common as it is unsatisfying: 'it depends'. Sometimes a model can be very useful if it just predicts, and does not explain, as witnessed by the undebatable success that machine learning models have across the sciences. However, only true understanding of the studied process can lead to hypotheses and predictions on phenomena that are far away from the currently available data. The more a model is fitted to a specific dataset, the less we are able to extrapolate it beyond this dataset.
These questions are very relevant in the context of metabolic modeling. Metabolic models have many unknown parameters, stemming from our ignorance of the biological process: What is the true objective? What constraints are relevant for determining metabolism? It is a deceptive trap to view the success of the model in reproducing the observed data as a validation that the right parameters, objective and constraints were chosen. A successful model only indicates that the modeled mechanism can be similar to the true biological mechanism, but it does not show it actually is. The problem is that, since we have many different parameters to choose from, many different models can explain the same metabolic observations [21].
An especially important question is whether metabolism is truly optimized for some evolutionary function. It is now an attractive option to view the success of optimization-based models as proof that the cells are indeed optimized, but this would be wrong because we can also explain the data with models that do not require optimization. To really quantify whether metabolism is optimized we should therefore devise quantitative tests that distinguish between randomly chosen and optimized metabolic states. An interesting approach for describing the metabolic outcome of cells, relying on statistical mechanics rather than on a selected objective function, has already been introduced [22].

5.7. Concluding remarks

In this chapter we built upon the exploration of flux spaces derived from constraints by imposing optimality criteria in terms of an objective function. The choice of the objective function(s) and the constraints depend on the modeling purpose. We will summarize some of the possible choices by listing three purposes that this type of models can have.

First, constraint-based optimization can be used to collect, integrate and extrapolate data on the metabolism of a specific organism. In this case, as much experimental information as possible can be used to refine the model. For example, measured fluxes can be fixed with constraints, measured metabolite concentrations can be used to determine the thermodynamically feasible direction of reactions, and transcriptome information can be used to exclude some reactions because the corresponding genes are not expressed. One of the applications is then that unknown variables can be inferred such that they are in accordance with the metabolic network and all the measured variables.

Second, hypotheses can be tested on why the studied organism attains its metabolic state. By choosing an objective function we can propose what *drives* the metabolic behavior and by choosing the constraints we propose what *limits* the metabolic behavior. If the model is then in accordance with the experimental observations, we know that at least the hypotheses were not proven wrong. On the other hand, we must be careful not to conclude from this that the hypotheses must be right, as we discussed in the box with

philosophical remarks.

Third, we may use these models to search for a metabolic state that results in a certain desired behavior, for example in the secretion of a product that is useful for industrial or medicinal reasons. In this case, the objective function is picked such that exactly the desired behavior is maximized, often while requiring that some biomass production is still possible because the cells need to be able to grow before the harvesting of the product can start.

Despite these useful purposes, we have also identified several limitations of the FBA-type models that we described here, such as ignoring metabolite concentrations, enzyme kinetics, and the assumption of a stationary metabolic state. The reason that these models are still very popular is their computational simplicity: as long as the objective function and constraints are linear in the reaction rates, the optimal solution is relatively easy to find using linear programming. This makes it feasible to make and run these models on genome-scale metabolic networks, which are networks that comprise all the metabolic enzymes for which the genome encodes, and can include thousands of reactions.

Understanding the solutions of such large models can also be very difficult due to their dimensionality. This is made easier when one uses elementary flux modes: we have seen that a solution is always a combination of a relatively small number of EFMs. More precisely, the number of EFMs that are active in the optimal solution cannot exceed the number of imposed constraints. This means that to understand the solution, we only need to understand which EFMs are selected and why. As such, we can interpret optimal solutions in terms of the EFMs, i.e. the minimal metabolic strategies, that are used.

Recommended readings and tools

Escher FBA Escher FBA (`https://sbrg.github.io/escher-fba/`) is nicely illustrative. It does FBA on an *E. coli* core model. Bounds on all reactions can be changed and different objectives can be explored. The resulting flux distribution is shown graphically.

Problems

Problem 5.1 Augment the metabolic network of *Spirallus insilicus* (Problem 4.1) by adding the in-homogeneous constraint $v_{upt} \leq 10 \frac{\text{mmol}}{\text{gDWh}}$ and calculate the flux distribution if biomass is the objective function (maximize v_5).

1. Using a spreadsheet and its associated linear programming optimizer.
2. Using an LP solver in Python such as `linprog` available in `scipy.optimize`.
3. Is the flux distribution unique? Calculate the maximum and minimum values of each flux (except for the uptake of substrate and biomass production) if v_5 should be equal to its optimal value (v_5^*) and if this constraint is relaxed to $v_5 \geq 0.9\ v_5^*$.

Problem 5.2 The metabolic network illustrated in Figure 5.1, adapted from [19], was designed to include four phenotypes that can be reached depending on the ratios of the oxygen and carbon source (A) uptake, defining zones of single nutrient and dual nutrient limitation.

1. If the uptake of the carbon source A is bounded between 0 and $10\ \frac{\text{mmol}}{\text{gDWh}}$ and no restrictions on the oxygen uptake are imposed, prepare a plot showing the biomass, C, D and E fluxes attained at different uptakes of A.
2. Repeat the preceding analysis, but limit the maximum uptake rate of oxygen to $10\ \frac{\text{mmol}}{\text{gDWh}}$.

3. If substrates uptakes are bounded between 0 and $10\ \frac{\text{mmol}}{\text{gDWh}}$ for A and 0 and $20\ \frac{\text{mmol}}{\text{gDWh}}$ for oxygen, calculate the phenotype phase plane. In each region of the phase plane (defined by a different slope), pick a combination of A and oxygen uptakes and analyze the fluxes of C, D and E.

Table 5.1: Stoichiometry of the metabolic network for problem 5.2. Adapted from [23] after [19].

$\xrightarrow{\mathrm{q_A}} \mathrm{A}$	$\mathrm{ATP} \xrightarrow{\mathrm{R_{ft}}}$
$\mathrm{A} + \mathrm{ATP} \xrightarrow{\mathrm{R_1}} \mathrm{B}$	$\mathrm{C} + 10\,\mathrm{ATP} \xrightarrow{\mathrm{R_z}} \mathrm{Biomass}$
$\mathrm{B} \xrightarrow{\mathrm{R_2}} 2\,\mathrm{ATP} + 3\,\mathrm{NADH} + \mathrm{C}$	$\xrightarrow{\mathrm{q_{O_2}}} \mathrm{O_2}$
$0.2\,\mathrm{C} \xrightarrow{\mathrm{R_3}} 2\,\mathrm{NADH}$	$\mathrm{C} \xrightarrow{\mathrm{C_{out}}}$
$\mathrm{C} \xrightarrow{\mathrm{R_4}} \mathrm{ATP} + 3\,\mathrm{D}$	$\mathrm{D} \xrightarrow{\mathrm{D_{out}}}$
$\mathrm{C} + 2\,\mathrm{NADH} \xrightarrow{\mathrm{R_5}} 3\,\mathrm{E}$	$\mathrm{E} \xrightarrow{\mathrm{E_{out}}}$
$\mathrm{NADH} + \mathrm{O_2} \xrightarrow{\mathrm{R_{Res}}} 2\,\mathrm{ATP}$	

Bibliography

[1] Xin Fang, Colton J Lloyd, and Bernhard O Palsson. Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology*, 18(12):731–743, 2020. doi: 10.1038/s41579-020-00440-4.

[2] Daniel A Cuevas, Janaka Edirisinghe, Chris S Henry, Ross Overbeek, Taylor G OConnell, and Robert A Edwards. From dna to fba: how to build your own genome-scale metabolic model. *Frontiers in microbiology*, 7:907, 2016. doi: 10.3389/fmicb.2016.00907.

[3] Leonid V. Kantorovich. Mathematical methods of organizing and planning production. *Management Science*, 6(4):366–422, 1960. doi: 10.1287/mnsc.6.4.366.

[4] Leonid V. Kantorovich. *The Best Use of Economic Resources*. Pergamon Press, Oxford, 1965. URL https://archive.org/details/bestuseofeconomi0000kant/page/n391/mode/2up.

[5] Jan Philipp Dapprich. Optimal planning with consumer feedback: a simulation of a socialist economy. *Review of Political Economy*, 2022. doi: 10.1080/09538259.2021.2005367.

[6] Robert Schuetz, Lars Kuepfer, and Uwe Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli. *Molecular systems biology*, 3(1):119, 2007.

[7] Andrea L Knorr, Rishi Jain, and Ranjan Srivastava. Bayesian-based selection of metabolic objective functions. *Bioinformatics*, 23(3):351–357, 2007.

[8] Carlos Eduardo García Sánchez and Rodrigo Gonzalo Torres Sáez. Comparison and analysis of objective functions in flux balance analysis. *Biotechnology progress*, 30(5):985–991, 2014.

[9] Frank J Bruggeman, Robert Planqué, Douwe Molenaar, and Bas Teusink. Searching for principles of microbial physiology. *FEMS Microbiology Reviews*, 44(6):821–844, 09 2020. doi: 10.1093/femsre/fuaa034.

[10] Rafael U Ibarra, Jeremy S Edwards, and Bernhard O Palsson. Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–189, 2002. doi: 10.1038/nature01149.

[11] Nathan E Lewis, Kim K Hixson, Tom M Conrad, Joshua A Lerman, Pep Charusanti, Ashoka D Polpitiya, Joshua N Adkins, Gunnar Schramm, Samuel O Purvine, Daniel Lopez-Ferrer, et al. Omic data from evolved e. coli are consistent with computed optimal growth from genome-scale models. *Molecular systems biology*, 6(1):390, 2010. doi: 10.1038/msb.2010.47.

[12] William R. Harcombe, Nigel F. Delaney, Nicholas Leiby, Niels Klitgord, and Christopher J. Marx. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLOS Computational Biology*, 9(6):1–11, 06 2013. doi: 10.1371/journal.pcbi.1003091.

[13] Axel von Kamp and Steffen Klamt. Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nature communications*, 8(1):15956, 2017. doi: 10.1038/ncomms15956.

[14] Jan Berkhout, Evert Bosdriesz, Emrah Nikerel, Douwe Molenaar, Dick de Ridder, Bas Teusink, and Frank J Bruggeman. How Biochemical Constraints of Cellular Growth Shape Evolutionary Adaptations in Metabolism. *Genetics*, 194(2):505–512, 06 2013. doi: 10.1534/genetics.113.150631.

[15] Yu Chen, Eunice van Pelt-KleinJan, Berdien van Olst, Sieze Douwenga, Sjef Boeren, Herwig Bachmann, Douwe Molenaar, Jens Nielsen, and Bas Teusink. Proteome constraints reveal targets for improving microbial fitness in nutrient-rich environments. *Molecular Systems Biology*, 17(4):e10093, 2021. doi: 10.15252/msb.202010093.

[16] Qasim K Beg, Alexei Vazquez, Jason Ernst, Marcio A de Menezes, Ziv Bar-Joseph, A-L Barabási, and Zoltán N Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, 104(31): 12663–12668, 2007. doi: 10.1073/pnas.0609845104.

[17] Daan H De Groot, Coco Van Boxtel, Robert Planqué, Frank J Bruggeman, and Bas Teusink. The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. *PLOS Computational Biology*, 15(3):e1006858, 2019. doi: 10.1371/journal.pcbi.1006858.

[18] Eunice van Pelt-KleinJan, Daan H. de Groot, and Bas Teusink. Understanding fba solutions under multiple nutrient limitations. *Metabolites*, 11(5), 2021. doi: 10.3390/metabo11050257.

[19] Jeremy S. Edwards, Ramprasad Ramakrishna, and Bernhard O. Palsson. Characterizing the metabolic phenotype: A phenotype phase plane analysis. *Biotechnology and Bioengineering*, 77(1):27–36, 2002. doi: 10.1002/bit.10047.

[20] R. Mahadevan and C.H. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, 5(4):264–276, 2003. doi: https://doi.org/10.1016/j.ymben.2003.09.002.

[21] Daan H De Groot, Julia Lischke, Riccardo Muolo, Robert Planqué, Frank J Bruggeman, and Bas Teusink. The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. *Cellular and Molecular Life Sciences*, 77:441–453, 2020. doi: 10.1007/s00018-019-03380-2.

[22] Daniele De Martino, Anna Mc Andersson, Tobias Bergmiller, Călin C Guet, and Gašper Tkačik. Statistical mechanics for metabolic networks during steady state growth. *Nature communications*, 9(1):2988, 2018. doi: 10.1038/s41467-018-05417-9.

[23] Felipe Scott, Pamela Wilson, Raúl Conejeros, and Vassilios S. Vassiliadis. Simulation and optimization of dynamic flux balance analysis models using an interior point method reformulation. *Computers & Chemical Engineering*, 119:152–170, 2018. doi: 10.1016/j.compchemeng.2018.08.041.