Review

Basic concepts and principles of stoichiometric modeling of metabolic networks

Timo R. Maarleveld[1,2,3], Ruchir A. Khandelwal[2], Brett G. Olivier[2], Bas Teusink[2,4] and Frank J. Bruggeman[2,4]

[1] Life Sciences, Center for Mathematics and Computer Science, Amsterdam, The Netherlands
[2] Systems Bioinformatics, Amsterdam Institute for Molecules Medicines and Systems, VU University Amsterdam, Amsterdam, The Netherlands
[3] BioSolar Cells, Wageningen, The Netherlands
[4] Kluyver Centre for Genomics of Industrial Fermentation / NCSB, Delft, The Netherlands

Metabolic networks supply the energy and building blocks for cell growth and maintenance. Cells continuously rewire their metabolic networks in response to changes in environmental conditions to sustain fitness. Studies of the systemic properties of metabolic networks give insight into metabolic plasticity and robustness, and the ability of organisms to cope with different environments. Constraint-based stoichiometric modeling of metabolic networks has become an indispensable tool for such studies. Herein, we review the basic theoretical underpinnings of constraint-based stoichiometric modeling of metabolic networks. Basic concepts, such as stoichiometry, chemical moiety conservation, flux modes, flux balance analysis, and flux solution spaces, are explained with simple, illustrative examples. We emphasize the mathematical definitions and their network topological interpretations.

Supporting information available online

1 Mass balances for metabolites lie at the basis of constraint-based modeling

Metabolism is essentially a large network of coupled chemical conversions (reactions) catalyzed mostly by enzymes. In this process, nutrients are converted into building blocks, such as nucleotides, fatty acids, lipids, amino acids, and free-energy carriers, for the synthesis of macromolecules, such as DNA, RNA, and proteins. These macromolecules are required for the maintenance of cellular integrity and formation of new cells. Fundamental processes in metabolism are enzyme-catalyzed reactions. In a single reaction, substrates are converted into products and the number of atoms of a given type, such as C, H, O, N, P,

or S, and the net charge should balance on each side of the equation [1]. These balancing principles are followed in genome-scale metabolic reconstructions [2]. Some aspects of balancing remain ambiguous, such as the protonation states of some of the metabolites, because this may be dependent on intracellular properties, such as pH and ionic strength. Every reaction occurs at a rate that depends on the concentrations of the enzyme reactants, possibly a few effectors, and the enzyme kinetic properties described by enzyme kinetics. Any reaction j can be written as Eq. (1):

$$\sum_{i=1}^{m} n_{ij}^{-} x_i \rightarrow \sum_{i=1}^{m} n_{ij}^{+} x_i \qquad (1)$$

in which we consider a network with a total of m metabolites (reactants), denoted by x_i. The n_{ij}^{+} and n_{ij}^{-} coefficients denote product and substrate stoichiometric coefficients, respectively, and equal the number of molecules produced and consumed per unit reaction rate. The reaction rate is denoted by v_j and typical units are mM min^{-1} or mmol h^{-1} (g biomass)$^{-1}$. The net stoichiometric coefficient of metabolite i in reaction j is defined as $n_{ij} = n_{ij}^{+} - n_{ij}^{-}$.

Correspondence: Prof. Frank J. Bruggeman, Systems Bioinformatics, VU University, De Boelelaan 1087, 1081 HV, Amsterdam, The Netherlands
E-mail: f.j.bruggeman@vu.nl

Abbreviations: FBA, flux balance analysis; **FVA**, flux variability analysis; **EFMs**, elementary flux modes; **ExPas**, extreme pathways; **CoPE-FBA**, comprehensive polyhedra enumeration flux balance analysis

The rates of change of the concentration of every metabolite can be equated in terms of reaction rates and net stoichiometric coefficients, which gives rise to the set of ordinary differential equations given by Eq. (2):

$$\frac{d}{dt}\boldsymbol{x}(t,\boldsymbol{p}) = \boldsymbol{N} \cdot v(\boldsymbol{x}(t,\boldsymbol{p}),\boldsymbol{p}) \tag{2}$$

The metabolite or state vector $\boldsymbol{x}$ is $m \times 1$ in dimension. The $r \times 1$ rate vector, $\boldsymbol{v}$, contains the rate equations of the r reactions in the network, which are typically expressed in terms of enzyme kinetics. The stoichiometric matrix $\boldsymbol{N}$ is $m \times r$ in dimensions and contains as its i,jth entry the net stoichiometric coefficient, n_{ij}, of metabolite i in reaction j. The coefficient $n_{ij} < 0$ if metabolite i is a substrate in the net stoichiometry of reaction j and $n_{ij} > 0$ if metabolite i is a product in the net stoichiometry of reaction j. The kinetic and environmental parameters are elements of the vector $\boldsymbol{p}$ and t denotes time. Note that metabolites that are held at a fixed concentration (boundary metabolites) do not enter the stoichiometry matrix because they do not have a rate of change. They enter as parameters in the parameter vector $\boldsymbol{p}$.

Because our interest is in stoichiometric models, we will not discuss further enzyme kinetics that enter the rate vector $\boldsymbol{v}$, see [3]. The stoichiometric matrix is the principle object of study in stoichiometric modeling. Herein, we discuss basic analyses of the stoichiometric matrix.

2 Chemical moiety conservation

In metabolism, metabolites tend to occur that are solely recycled. Examples of such metabolites include ATP, NAD(P)H, and coenzyme A. As a consequence of recycling, the maximum concentration of those metabolites is constrained by a total concentration of a chemical moiety. For instance, in the case of phosphate and adenosine moiety conservation, the relationships given by Eq. (3) hold true at any moment in time:

$$P_T = 3ATP + 2ADP + AMP + P$$
$$A_T = ATP + ADP + AMP \tag{3}$$

with total phosphate and adenosine levels given as P_T and A_T. Taking the derivative of Eq. (3) with respect to time gives Eq. (4):

$$0 = 3\frac{dATP}{dt} + 2\frac{dADP}{dt} + \frac{dAMP}{dt} + \frac{dP}{dt}$$

$$0 = \frac{dATP}{dt} + \frac{dADP}{dt} + \frac{dAMP}{dt} \tag{4}$$

Equation (4) indicates linear relationships between the rows of the stoichiometry matrix and allows for the expression of the rate of change of one metabolite in terms of other rates of change [4]. In matrix form, we can write

this as Eq. (5) for the general case of metabolites of a metabolic network:

$$\frac{d\boldsymbol{x}^D}{dt} = \boldsymbol{L_0}\frac{d\boldsymbol{x}^I}{dt} \tag{5}$$

The vectors $\boldsymbol{x}^D$ and $\boldsymbol{x}^I$ denote the vector of dependent and independent metabolite concentrations. The matrix $\boldsymbol{L_0}$ expresses the linear combinations of the rates of changes of the independent metabolites. In integrated form, Eq. (5) becomes Eq. (6):

$$\boldsymbol{x}^D - \boldsymbol{L_0}\boldsymbol{x}^I = \boldsymbol{t} \tag{6}$$

with $\boldsymbol{t}$ as the vector of total concentrations of chemical moieties. By way of illustration, the vectors $\boldsymbol{x}^D$ and $\boldsymbol{x}^I$ and the matrix $\boldsymbol{L_0}$ are determined for the moiety-conservation relationships given in Eq. (3) by Eq. (7):

$$\frac{d}{dt}\begin{pmatrix} AMP \\ P \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ -2 & -1 \end{pmatrix}\frac{d}{dt}\begin{pmatrix} ATP \\ ADP \end{pmatrix} \tag{7}$$

Using this $\boldsymbol{L_0}$ matrix, the dynamics of all metabolites (ATP, ADP, AMP, and P) can be obtained from the dynamics of the independent metabolites (ATP, ADP). In other words, the dependent species are redundant for determining the species dynamics. Note that different combinations of independent metabolites can be chosen. This can intuitively be seen from the relationships given in Eq. (3). For example, by choosing ATP and AMP as independent metabolites, the concentration of ADP can be determined from $A_T = ATP + ADP + AMP$. Subsequently, the concentration of P can be determined from $P_T = 3ATP + 2\,ADP + AMP + P$.

The relationship given in Eq. (5) dictates the decomposition of the stoichiometric matrix into two blocks, given by Eq. (8):

$$\frac{d}{dt}\begin{pmatrix} \boldsymbol{x}^I \\ \boldsymbol{x}^D \end{pmatrix} = \boldsymbol{N}\boldsymbol{v} = \begin{pmatrix} \boldsymbol{N_R} \\ \boldsymbol{N_0} \end{pmatrix}\boldsymbol{v} \tag{8}$$

in which $\boldsymbol{N}$ is decomposed into blocks of $\boldsymbol{N_R}$, which is the reduced stoichiometry matrix, and $\boldsymbol{N_0}$ [5, 6]. Together, the relationships shown in Eqs. (5) and (8) give rise to Eq. (9):

$$\boldsymbol{L_0}\boldsymbol{N_R} = \boldsymbol{N_0} \Rightarrow (-\boldsymbol{L_0}\ \boldsymbol{I})\begin{pmatrix} \boldsymbol{N_R} \\ \boldsymbol{N_0} \end{pmatrix} = \boldsymbol{0} \Rightarrow \begin{pmatrix} \boldsymbol{N_R} \\ \boldsymbol{N_0} \end{pmatrix}^T(-\boldsymbol{L_0}\ \boldsymbol{I})^T = \boldsymbol{0} \tag{9}$$

and indicate that the moiety conservation matrix in Eq. (10):

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{L_0} \end{pmatrix} \tag{10}$$

can be derived from the left null-space of the stoichiometry matrix [7, 8]. Typically, $\boldsymbol{N_R}$ is identified in $\boldsymbol{N}$ after the null-space has been calculated. The number of independ-

ent metabolites, m_0, is denoted by the rank of $\boldsymbol{N}$. Thus, the reduced stoichiometry matrix $\boldsymbol{N_R}$ will be $m_0 \times r$ in size. This indicates that the stoichiometry matrix $\boldsymbol{N}$ has m_0 independent rows and $m - m_0$ moiety-conservation relationships.

In genome-scale models, a biomass reaction is typically used to describe cell growth. This biomass reaction is used as a sink for biomass precursors (e.g. DNA, RNA, proteins, lipids) that together define the biomass composition of the cell. These biomass precursors contain moieties, such as adenosine, that require the continuous synthesis of adenosine. Therefore, a non-zero flux through this biomass reaction results in a drain of the moieties. Hence, in such a genome-scale metabolic model, a strict

application of moiety conservation detection along the lines detailed above, will result in fewer moieties. Yet, to understand the dynamics of metabolic pathways, they are relevant because the turnover of ATP is much larger than the rate at which the adenosine moiety will be synthesized.

3   Steady-state flux modes

By definition, at a steady state of the metabolic network, Eq. (11) holds:

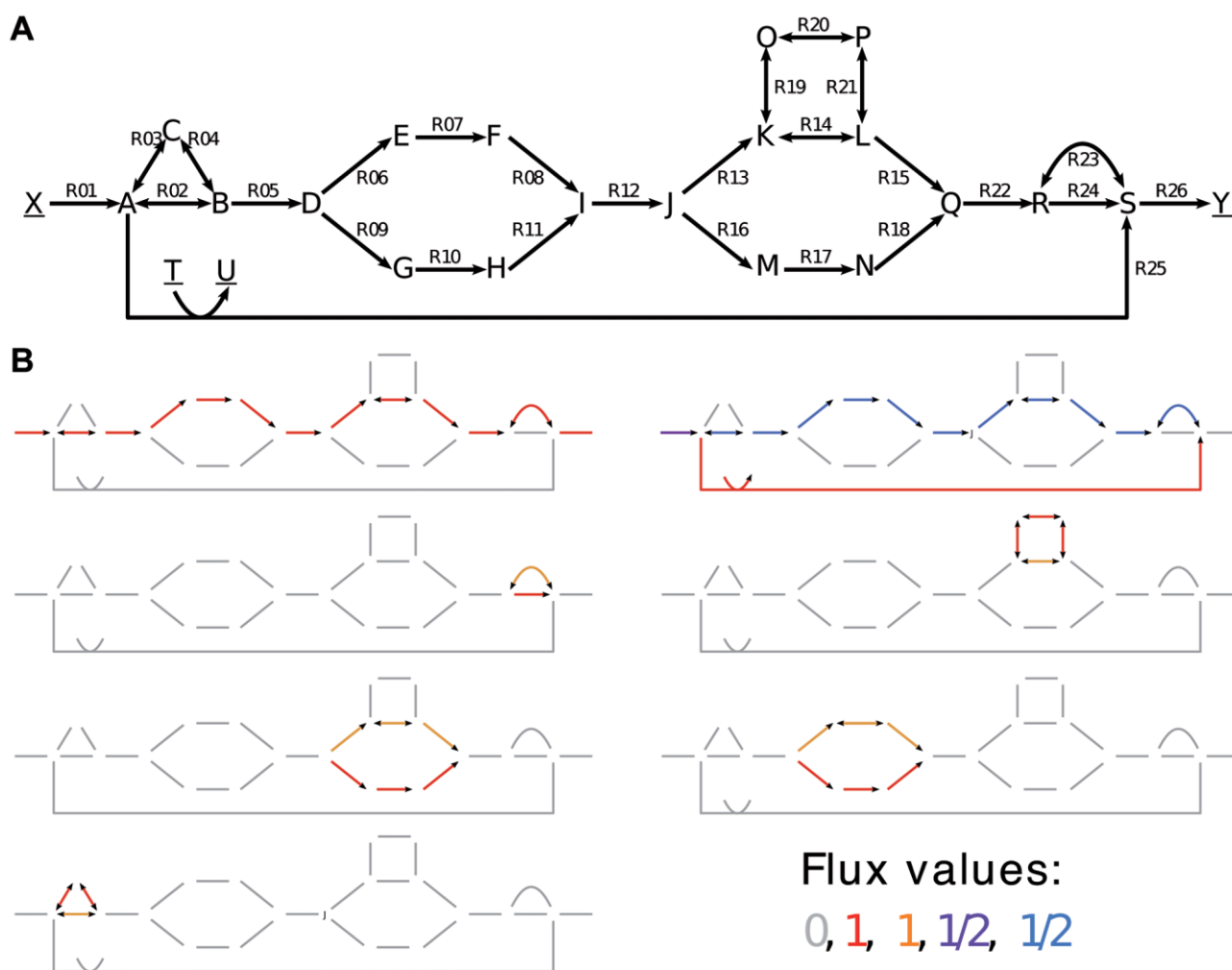$$\boldsymbol{N_R J} = \boldsymbol{0} \tag{11}$$

Figure 1. A simplified metabolic pathway to illustrate the concept of flux modes. (**A**) A network diagram of a simplified metabolic network. Arrows indicate reactions and are labeled as R*n*. Double-headed arrows indicate reversible reactions. Irreversible reactions are indicated by single-headed arrows, which point in the thermodynamically preferred direction. Underlined metabolites are considered to be fixed in concentration to allow for a steady state. Note that all reactions are uni–uni reactions, except R25, which has stoichiometry of $A + T \rightarrow U + \frac{1}{2}S$. We can rewrite this stoichiometry as $A_2 + P \rightarrow AP + A$ to illustrate that there is no stoichiometric inconsistency with the isomerization reactions. To deal with thermodynamic inconsistencies, imagine adding fixed metabolites V and W to R24 to drive this reaction forward. A description of this model in the SBML level 3 package can be found in the Supporting information. (**B**) An overview of the seven flux modes. Colors correspond to flux values.

Here we used the convention that the reaction rate vector, $\mathbf{v}$, at steady state is denoted by $\mathbf{J}$, which is the flux vector. Equation (11) gives rise to m_0 flux relationships, each of which represent a linear combination between fluxes. As a consequence, $r - m_0$ fluxes are minimally required to determine all fluxes. Hence, Eq. (12) must exist because it describes all linear combinations of independent fluxes ($\mathbf{J^I}$) that give rise to the dependent fluxes in $\mathbf{J^D}$:

$$\mathbf{J^D} = \mathbf{K_0}\mathbf{J^I} \Rightarrow \mathbf{J} = \begin{pmatrix} \mathbf{J^I} \\ \mathbf{J^D} \end{pmatrix} = \begin{pmatrix} \mathbf{I} \\ \mathbf{K_0} \end{pmatrix}\mathbf{J^I} \Rightarrow \mathbf{N}_R\begin{pmatrix} \mathbf{I} \\ \mathbf{K_0} \end{pmatrix} = \mathbf{0}$$

(12)

Thus, the right null-space of the (reduced) stoichiometry matrix equals the kernel matrix, given by Eq. (13):

$$\mathbf{K} = \begin{pmatrix} \mathbf{I} \\ \mathbf{K_0} \end{pmatrix}$$

(13)

Note that the columns of $\mathbf{N_R}$ may have to be reordered to write the null-space in this form. In addition, each column of $\mathbf{K}$ can be divided by any number; all resulting vectors continue to lie in the null-space of $\mathbf{N_R}$. The rows of $\mathbf{K}$ represent the flux values of a specific reaction.

If we denote the ith column of $\mathbf{K}$ by $\mathbf{k}_i$, any flux vector $\mathbf{J}$ can be written as a linear combination of the columns of $\mathbf{K}$, as given by Eq. (14):

$$\mathbf{J} = \sum_{i=1}^{r-m_0} \alpha_i \mathbf{k}_i$$

(14)

in which the weighting coefficients, α_i, can take any value. The set of all flux vectors of the metabolic network is contained within the null-space of $\mathbf{N_R}$. However, this is a huge space and below we discuss definitions that reduce this space by incorporating additional thermodynamic information or postulating optimal metabolic functioning. Equation (14) illustrates that the definition of $\mathbf{K}$ cannot be unique because multiplication of the multipliers α_i by any factor λ can be compensated for by the division of every element in $\mathbf{k}_i$ by λ. Because no restrictions apply to the values of the multipliers, $\mathbf{K}$ cannot be uniquely chosen.

The vectors $\mathbf{k}_i$ have a network topological interpretation. They represent routes through the network along which every metabolite is at steady state, if the fluxes carry the values dictated by $\mathbf{k}_i$. This is why the $\mathbf{k}_i$ terms are often called flux modes. In Fig. 1A, a toy metabolic network is shown. It contains 26 reactions and 23 metabolites. The external metabolites T, U, X, and Y are considered to be fixed; the other 19 metabolites are considered to be variable. The stoichiometry matrix has full rank and, therefore, no conserved chemical moieties occur. The number of independent fluxes equals seven (=26–19). Thus, seven flux modes exist and they are displayed in Fig. 1B. The color codes of the reactions indicate reaction rate values and it can be easily verified that all variable metabolites are at steady state for all flux modes. Because all of these metabolites along a flux mode are required to operate a steady state, the flux modes have to be either cycles or routes from source to sink metabolites.

Note that the flux modes do not necessarily agree with thermodynamics, since several irreversible reactions are forced to have a negative flux. In addition, the flux mode in the upper-right part of Fig. 1B is complex and could be swapped for a simpler flux mode if desired. This indicates the problems associated with analysis of the null-space and explains why alternative definitions for steady-state flux routes in metabolic networks have been developed; these are introduced and discussed in sections 4 to 9. These alternative definitions are unique representations of the null-space and agree with the thermodynamic preference of the reactions.

4 Flux balance analysis (FBA)

In Section 3, we discussed the steady-state relationship $\mathbf{N}_R\mathbf{J} = \mathbf{0}$. We illustrated that this system of equations was underdetermined; more unknown fluxes occurred than the number of linear relationships ($r > m_0$). Therefore, the null-space of the stoichiometric matrix $\mathbf{N}$ did not lead to a unique flux vector, but a whole solution space. To realistically narrow down the solution space, FBA selects only those flux values that together can optimize some biologically relevant objective, such as maximum biomass rate or maximum ATP production rate [9]. This optimization is achieved by a linear programming approach [10, 11] and FBA can be mathematically represented by Eq. (15):

Maximize or Minimize: $Z = \mathbf{c^T}\mathbf{J}$

Subject to,

$$\mathbf{N}\,\mathbf{J} = \mathbf{0}$$

$$\mathbf{J^{min}} \leq \mathbf{J} \leq \mathbf{J^{max}}$$

(15)

in which vector $\mathbf{c}$ dictates a linear relationship between fluxes of $\mathbf{J}$ that forms the objective function Z. $\mathbf{J^{min}}$ and $\mathbf{J^{max}}$ are vectors of minimum and maximum values, respectively, that any flux of vector $\mathbf{J}$ can attain during this optimization. These flux bounds can represent experimental measurements by bounding all known flux values within experimental errors or they can derive from thermodynamic considerations that force fluxes to be either strictly negative or positive. This linear program narrows down the feasible steady-state flux space of the stoichiometric matrix by applying stoichiometric, thermodynamic, and environmental constraints and by optimizing an objective function. Hence, this linear program results in an optimal solution space, which only contains those solutions for fluxes that, in combination, give a unique and maximum value for the objective function Z. This space is considerably smaller than the space dictated by $\mathbf{K}$.
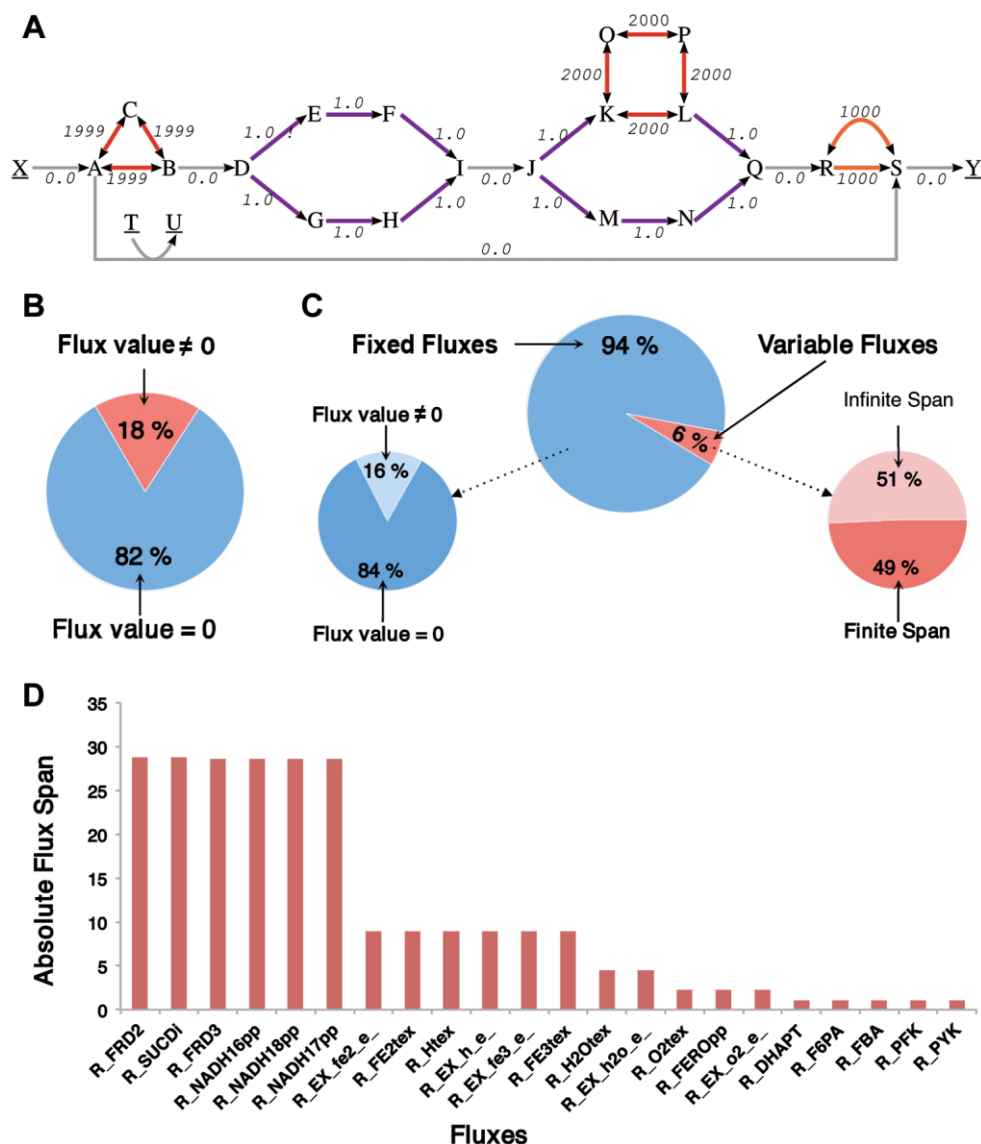
Figure 2. FBA and flux variability analysis (FVA) of *E. coli* model iAF1260 in a defined mineral medium. **(A)** FVA performed on the toy metabolic model. Resulting spans are shown for all reactions in which zero span is for fixed (in-)active reactions (gray, R25 is inactive), a span of one is for active but variable reactions (purple), and large spans are for reactions (red) in cycles. **(B)** Flux distribution resulting from FBA on the genome-scale model predicted that 82% of the reactions were inactive ($J_i = 0$) and only 18% of the fluxes carried a non-zero flux ($J_i \neq 0$). **(C)** Analysis of the results of FVA revealed that 94% of the metabolic network was fixed (fixed fluxes) and only 6% of all fluxes (variable fluxes) could vary without changing the growth rate. 49% of these variable fluxes have a finite span, while 51% have an infinite span, suggesting their involvement in infeasible cycles. Out of those fixed fluxes, 84% never carry any mass ($J_i = 0$) and 16% are active ($J_i \neq 0$). **(D)** Absolute spans of some reactions, resulting from FVA, are presented. All reaction names are taken from the model itself.

Performing FBA on the toy metabolic model shown in Fig. 1A with the input flux R01 ≤ 1, and maximization of the flux through R26 as the objective function yields a flux distribution, one of which is presented in the upper-left part of Fig. 1B. This example indicates that the flux distribution that results from FBA can be (a linear combination of the) flux modes of the stoichiometric matrix.

We also performed FBA on the iAF1260 model of *Escherichia coli* with input fluxes representing a mineral medium [12] supplemented with glucose (uptake flux of 8 mmol g^{-1}h^{-1}), in the presence of oxygen (uptake flux of 18.5 mmol g^{-1}h^{-1}) and free exchange of ammonia, water, carbon dioxide, protons, phosphate, sulfate, and other metal ions. As the metabolic objective, we considered maximization of the rate of biomass synthesis. This optimization predicts a growth rate of 0.73 h^{-1} and a flux distribution in which 82% reactions are inactive; only 18% reactions of the whole metabolic network are used (Fig. 2B).

5 Flux Variability Analysis (FVA)

FVA maximizes and minimizes each flux of the metabolic network, while satisfying all given constraints at the optimal objective function value [13]. It is a useful tool to gain an insight into network flexibility. It gives the span of the fluxes that exist within the optimal solution space defined by the linear program given in Eq. (15). The linear program for FVA can be written by Eq. (16):

Maximize or Minimize: $\quad\quad\quad\quad J_i$

$\quad\quad$ Subject to,

$$NJ = 0 \quad\quad\quad (16)$$
$$c^T J = Z_{obj}$$
$$J^{min} \le J \le J^{max} \text{ for } i = 1 \dots r$$

in which Z_{obj} is the objective function value of the previous FBA program. By fixing the objective function value at the value obtained from the FBA optimization, FVA determines the range for each flux within which all numerical values are valid FBA solutions. Using the results of FVA optimization, we can determine the spans ($= |J_i^{FVA\,max} - J_i^{FVA\,min}|$) as an absolute difference between the FVA maximum ($J_i^{FVA\,max}$) and FVA minimum ($J_i^{FVA\,min}$) values of each flux J_i within the optima. On the basis of these spans, we can determine the fixed and flexible parts of the metabolic network while it achieves a particular metabolic objective. These spans can hit infinity because, in an optimal flux distribution of the metabolic network, some reaction rates may not be constrained at all. The FVA span gives an indication of the range of values that a reaction may attain. However, the actual value it can take in a particular flux distribution depends on the entire reaction network: fluxes cannot be changed independently because this would violate the steady-state constraint [14].

In Fig. 2A, the absolute spans resulting from FVA for all fluxes in the toy metabolic model are shown. The constraints for this FVA are identical to the FBA calculations for this model discussed above and infinity flux bounds are represented by a large value of 1000. This analysis depicts the reactions that have a fixed flux and a span of zero (gray arrows) and reactions that have variable fluxes with a span of one (purple arrows). Here, reactions with a span of zero are either active (essential) reactions because alternative optimal paths are not present or inactive (non-essential) reactions because they yield a sub-optimal FBA solution (R25). Some reactions (red arrows) have large spans because they are part of metabolic cycles (R02–R04, R14 and R19–R21, and R23–R24). In any optimal FBA solution with a maximal flux of one through R01, a net flux of one is required from R01 to R05. Therefore, the allowed flux values of R02–R04 are between -999 and 1000 (span = 1999); a flux of -999 through R02 results in a flux of 1000 through R03 and R04 and vice versa. In contrast, no net flux is required through the second metabolic cycle (R14

and R19–R21) to obtain an optimal FBA solution. For example, a flux of one through R16–R18 results in no flux through R13–R15 in any optimal FBA solution. As a result, the reactions of this metabolic cycle can operate at their maximum bounds in both directions without violating optimal metabolic functioning (span = 2000). Also, a net flux of one is required from R22 to R26. Because R24 is irreversible, the maximal flux R23 can obtain in any optimal solution is one (with $J_{R24} = 0$). Since R23 is reversible, an optimal FBA solution can be obtained if the fluxes through R23 and R24 are -999 and 1000, respectively. As a consequence, the spans of R23 and R24 are 1000 (R23: –999 to 1 and R24: 0–1000).

We also performed FVA on the *E. coli* model iAF1260 with the same constraints and objective function value as the FBA optimization described above. Analyzing the spans of all fluxes revealed that only 6% reactions could vary in the optimal solution space. Of the 94% fixed reactions, 16% carry a non-zero flux and the remaining 84% are inactive (Fig. 2C). This means that 21.04% (6% + 0.94 × 16%) reactions can have a non-zero flux. The percentage of non-zero fluxes in a FBA outcome (Fig. 2B) will be equal or lower, because some variable fluxes can also be zero in a FBA solution.

Further analysis of the variable fluxes (6% of total fluxes) revealed that 49% of them had a finite span. The remainder (51% of fluxes with infinite spans) get their variability due to metabolic cycles in the metabolic network. In section 9 we explain how these cycles can be interpreted. The spans of some of the reactions in the model, as obtained by FVA, are shown in Fig. 2D.

6 Interpretation of the sensitivity parameters associated with FBA solution

Two sensitivity parameters – reduced costs and shadow prices – are associated with a FBA solution. A reduced cost (r_i) can be interpreted as the sensitivity of the objective function with respect to the change in the ith flux value. In biological terms, this can be interpreted in the following manner: If a flux J_i has a reduced cost of r_i in a particular FBA solution and this flux value is increased by ΔJ_i, then the objective function value will be changed to $Z + r_i \Delta J_i$. Reduced costs assigned to nutrient uptake fluxes give us an indication of the growth-limiting compounds in the medium. The reduced costs assigned to the uptake fluxes of substrates that are not allowed to be consumed identify which nutrients could be added to the medium to achieve a higher growth rate [15]. Sometimes, inactive substrate fluxes are of no interest and scaled reduced costs (sr_i) [16] are used to identify the limiting substrates. Scaled reduced costs can be represented by Eq. (17):

$$sr_i = \frac{r_i \cdot J_i}{Z} \quad\quad\quad (17)$$

A shadow price (γ_i) is the sensitivity of the objective function with respect to the change in a constraint [17]. Consequently, if metabolite i is added, then objective function value Z will change according to the shadow price. Shadow prices have been used to analyze the effects of substrate availability on the growth in phenotypic phase plane analysis [18].

7 Elementary flux modes

Elementary flux modes (EFMs) and extreme pathways (ExPas) were developed to uniquely characterize the right null-space **K** of a stoichiometry matrix. In contrast to FBA-related techniques, EFM and ExPa analyses are only based on network stoichiometry and therefore allow an unbiased analysis without imposing an optimization principle. These definitions rely on a convex set of flux vectors [19, 20]. By taking a convex combination of these flux vectors ($\boldsymbol{e}_n$), any possible steady-state flux distribution ($\boldsymbol{J}$) can be generated. Assuming that we have N EFMs, we can write this as Eq. (18):

$$\boldsymbol{J} = \sum_{n=1}^{N} \alpha_n \boldsymbol{e}_n \tag{18}$$

Here α_i are non-negative weighting coefficients that total one (Eq. 19)

$$\sum_{i=n}^{N} \alpha_n = 1, \alpha_n \geq 0 \tag{19}$$

Both EMFs and ExPas can be exploited to evaluate, for instance, pathway redundancy, to find (sub-)optimal pathways for the investigation of pathway properties, such as cost and length, and to study the effect of gene deletions [21–23]. Unfortunately, both approaches suffer from excessive running times, that is, characterizing the right null-space of the stoichiometric matrix is a non-deterministic polynomial-time (NP)-hard computational problem.

EFMs [19] fulfill three conditions: (i) (pseudo-)steady state, (ii) thermodynamic feasibility, and (iii) non-decomposability. These conditions have several consequences. First, internal metabolites of an EFM are neither net consumed or produced due to the steady-state condition. Second, all flux rates of an EFM are thermodynamically feasible in contrast to the flux modes, $\boldsymbol{k}_i$. Third, no subset of an EFM exists that fulfills the first two conditions without violating the third. The complete set of EFMs can be partitioned into three types: (I) all optimal yield pathways converting one or more substrates into a product (e.g. biomass), (II) all sub-optimal yield pathways converting one or more substrates to a product, and (III) internal loops in the metabolic network.

The toy metabolic network shown in Fig. 1A contains 28 EFMs, as shown in Fig. 3. In this toy model, we can identify all three types of EFMs earlier defined. To begin with, there are 24 type I EFMs that characterize all optimal pathways, which are shown in Fig. 3A–X. Any route from metabolite X to Y, ignoring reaction R25, is a type I EFM. In this network, there is only one type II EFM that gives a sub-optimal yield (Fig. 3Y). Thus, any route from metabolite X to Y, involving reaction R25, results in a sub-optimal yield because R25 has stoichiometry $A + T \rightarrow U + \frac{1}{2}S$. Finally, there are three type III EFMs that characterize the internal loops of this metabolic network, as shown in Fig. 3Z–AB. Generally, these cycles are responsible for the large number of EFMs. To illustrate this point, without these three cycles the toy metabolic network has only five EFMs. Then, there would be four EFMs to characterize all optimal pathways, one EFM to characterize the sub-optimal pathway, and zero EFMs to characterize internal loops.

8 Extreme pathways

The alternative approach, ExPas, determines the edges of the cone that describe the steady-state solution space and the thermodynamic preference of reactions [20]. The set of ExPas does not have to contain all pathways with an optimal and sub-optimal yield, in contrast to the EFMs. Convex combinations of ExPas that satisfy the three EFM conditions, however, can be used to obtain all optimal and sub-optimal pathways. In addition to the three conditions of EFMs, ExPas require two additional conditions: (iv) network reconfiguration and (v) systematic independence.

Network reconfiguration results in a classification of each reaction as an internal or exchange reaction. Moreover, each internal reversible reaction is split into two irreversible reactions: a reaction describing the forward reaction and a reaction describing the backward reaction. Systematic independence guaranties that an ExPa cannot be represented by a non-negative linear combination of other ExPas. Because of the systemic independence condition, ExPas are always a subset of the EFMs. In other words, each ExPa is also an EFM, but not necessarily vice versa. This can result in fewer ExPas than EFMs for the same metabolic network. For a metabolic model of the human red blood cell, the average number of EFMs used for a given ExPa was about four [24]. However, if all exchange reactions in a metabolic network are irreversible, the sets of relevant EFMs and ExPas are identical. This is a general property of EFMs and ExPas [24, 25]. Note that each originally reversible internal reaction, split into two irreversible reactions, fulfills all ExPa and EFM conditions, resulting in additional ExPas and EFMs. These ExPas and EFMs can be considered irrelevant [24, 25] because they only redefine reversibility. In more math-
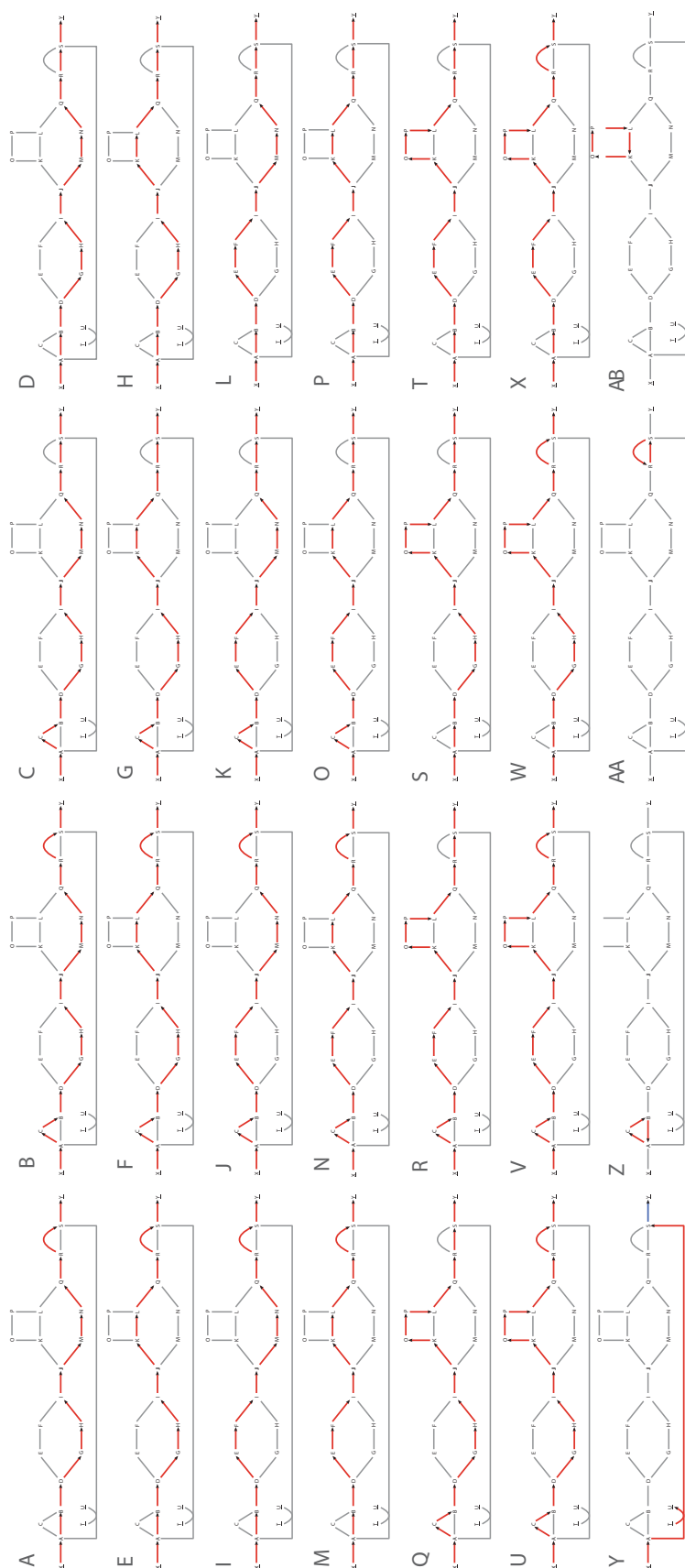
Figure 3. Topological characterization of all EFMs. **(A)**–**(X)** Type I EFMs. **(Y)** Type II EFM. **(Z)**-**(AB)** Type IV EFM. Visualizing ExPas requires decoupling of all reversible reactions into two irreversible reactions. Because all exchange reactions are irreversible, the set of relevant ExPas match this set of EFMs. Colors correspond to reaction values (red = 1, blue = 1/2).

ematical terms, the EFMs and ExPas are the extreme rays that span the flux cone, C, defined by $C = \{\boldsymbol{J} | \boldsymbol{NJ} = \boldsymbol{0}, \boldsymbol{J} \geq 0\}$.

Because the exchange reactions (R01 and R26) in this toy metabolic network are irreversible, the sets of relevant EFMs and ExPas are identical. Nevertheless, the network contains 38 ExPas and 28 EFMs. The additional 10 ExPas arise because of the network reconfiguration condition. Each of these 10 ExPas (not shown) is also an EFM that will be detected if they are determined after reconfiguring the network.

9 Unique representations of the optimal flux space

FBA can be exploited to calculate the maximum yield of a product on a certain substrate. FBA simulation provides a steady-state flux distribution, which corresponds to a point in the (optimal) solution space. Typically, a unique optimal steady-state flux distribution through the metabolic network cannot be guaranteed because the constraints defined by the stoichiometric network are insufficient. Accordingly, a solution space of optimal steady-state flux distributions exists that each give rise to the maximal yield. This solution space represents a polyhedron [26] and this space is considerably smaller than the entire steady-state solution space characterized by flux modes, EFMs, or ExPas. This reduction in solution space is achieved in FBA by the consideration of additional constraints, a particular nutrient environment, and the demand for flux distributions that optimize a metabolic objective. Characterizing the optimal solution space of FBA remains a NP-hard computational problem. Above, we characterized the variability of the flux values within the optimal solution space. Next, we characterize this solution space in network topological terms.

In contrast to both the EFMs and ExPas approaches, comprehensive polyhedra enumeration flux balance analysis (CoPE-FBA) characterizes only the optimal solution space, which is done in terms of a compact set of sub-networks [27]. These sub-networks account for all alternative flux distributions in the optimal steady state predicted by FBA. CoPE-FBA therefore provides the topological structure underneath flux variability, at least in the optimal solution. The solution space of optimal flux distributions contains three topological features: (i) vertices, (ii) rays, and (iii) linealities.

Vertices are optimal paths of the metabolic network, including reactions with fixed and variable fluxes. Rays are irreversible, thermodynamically infeasible cycles and linealities are reversible cycles in the metabolic network. No net conversion occurs in either the rays or linealities of the FBA polyhedron. The toy metabolic network has four vertices (Fig. 4B–E), one ray, and two linealities (Fig. 4A). Note that the EFMs and ExPas also consist of these three topological features. Therefore, rays and linealities are

responsible for the increase in the number of type I and II EFMs and ExPas. Alternatively, rays and linealities do not influence the number of CoPE-FBA sub-networks.

Any optimal flux distribution that satisfies the metabolic optimum obtained in the FBA calculation can be written in terms of the vertices, rays, and linealities using the Minkowski sum given in Eq. (20) [26, 27],

$$\boldsymbol{J}^{\text{opt}} = \sum_{i=1}^{s} \alpha_i \boldsymbol{\varphi}_i + \sum_{i=1}^{t} \beta_i \boldsymbol{\phi}_i + \sum_{i=1}^{u} \gamma_i \boldsymbol{\psi}_i \qquad (20)$$

in which the vectors $\boldsymbol{\varphi}_i$, $\boldsymbol{\phi}_i$, and $\boldsymbol{\psi}_i$ represent the vertices, rays, and linealities, respectively. The weighting coefficients obey the following restrictions: $\sum_{i=1}^{s} \alpha_i = 1$, $\alpha_i > 0$, $\beta_i > 0$, and γ_i can take any value. These definitions indicate that the vertices can be summed in a convex manner, the rays as a conical sum, and a linear combination can be taken over the linealities.

The sub-networks that can be identified with CoPE-FBA and explain the numbers of vertices for a given FBA problem satisfy three conditions: (i) only reactions belonging to a specific sub-network display correlation in flux values across the optimal solution space, (ii) fixed net input–output stoichiometry of reactants, and (iii) thermodynamic feasibility.

As a result, these sub-networks contain reactions that vary independently across all vertices of the optimal solution space. Therefore, without violating the optimality condition, sub-networks with alternative internal flux distributions can be independently chosen. For this reason, the number of vertices can be determined by multiplying the number of alternative internal flux distributions for each sub-network [27]. This illustrates the likely combinatorial explosion for the number of vertices of the optimal solution space for larger metabolic networks.

The toy metabolic network contains two CoPE-FBA sub-networks given in Fig. 4F. Each sub-network has two alternative internal fluxes distributions: the top and bottom branch. Multiplying the number of alternative flux distributions for each sub-network, $2 \times 2 = 4$, gives the number of vertices. Larger metabolic models tend to contain more vertices, while the number of sub-networks stays small. For instance, the genome-scale metabolic model iAF1260, consisting of 2374 reactions and 1668 metabolites, has about 1.7×10^6 vertices when studied under glucose growth conditions [27]. Still, only four sub-networks, which contain about 5% of the total number of reactions in this model, are enough to characterize the optimal solution space [27]. Comparing the number of EFMs (i), ExPas (j), vertices (k), and CoPE-FBA sub-networks (m) gives an indication of the level of compactness of these approaches. Typically, for larger models the number of EFMs, ExPas, and vertices will explode, which gives $i \geq j > k \gg m$.
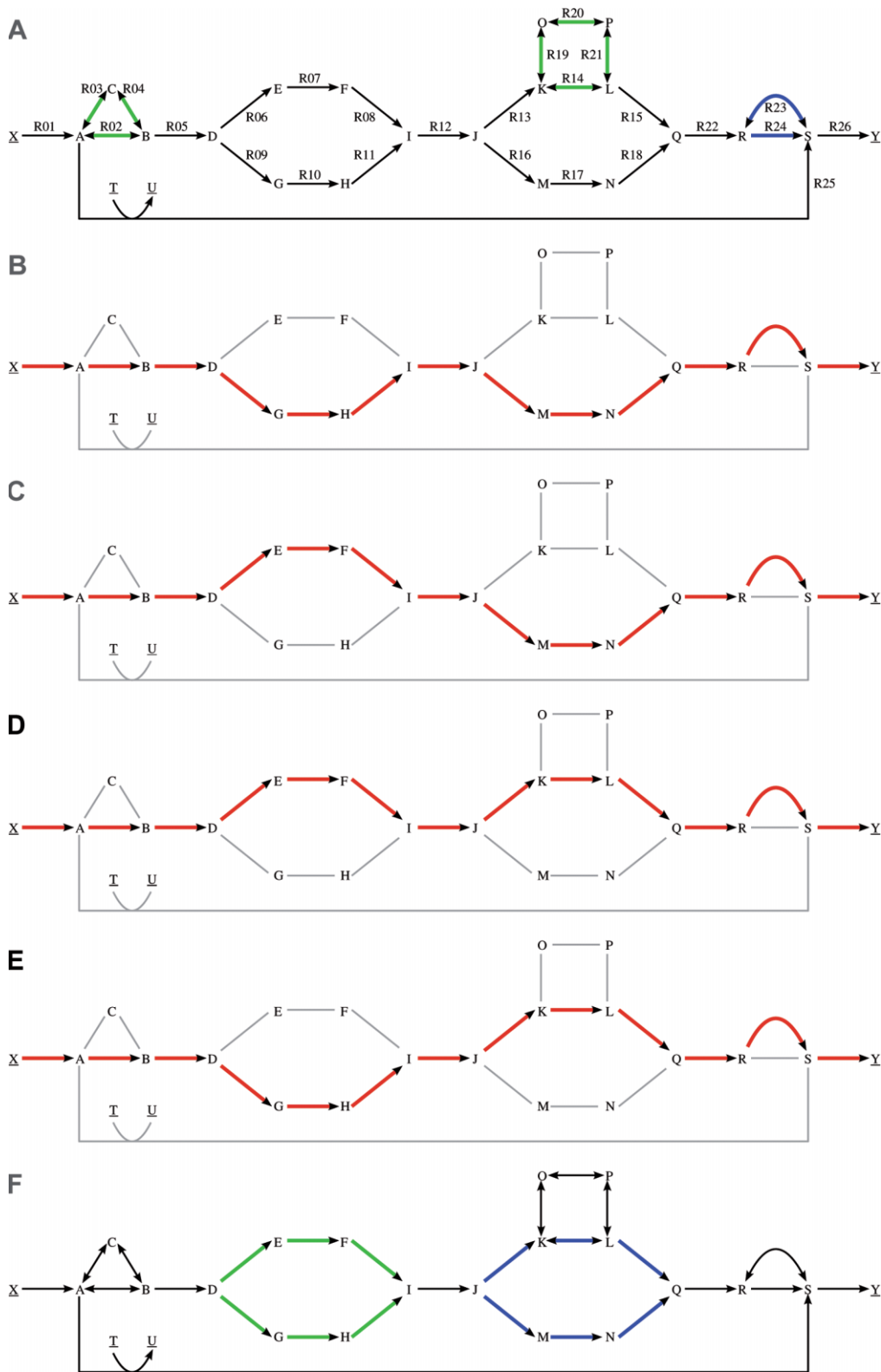
Figure 4. Topological characterization of the optimal FBA solution space. (**A**) This FBA program contains one ray (blue; R23 and R24) and two linealities (green; R02–R04 and R14, R19–R21). (**B**)–(**E**) Visualization of the four vertices this FBA program contains. Each vertex represents a route from substrate to product with a maximum yield. The values indicate the predicted flux values. Reaction R01 was bounded between zero and one. (**F**) The two sub-networks detected with CoPE-FBA. Both sub-networks contain two alternative flux distributions, resulting in 2 x 2 possible vertices shown in (**B**)–(**E**).

10 Biological implications of stoichiometric network analysis

The predictions made by any mathematical model depend heavily on the underlying assumptions. The definitions of ExPas, EMFs, and those related to FBA have the steady-state assumption in common. In general, the steady-state assumption is assumed to be valid because of the timescale separation between (fast) intracellular metabolic conversions and (slow) genetic regulation [28, 29].

In addition to the steady-state assumption, FBA assumed optimization of an objective function, which could, in some cases, be debatable from a biological perspective. Typical objective functions are the yield of the biomass reaction or ATP production. Optimization of these objectives is always bounded by capacity constraints of other reactions that ultimately bound the steady-state solution space. In other words, FBA optimizes an objective function relative to a limiting input flux. Thus, optimization of any reaction rate in FBA is always the optimization of a yield defined as the objective reaction rate divided by the limiting input. Optimization of growth rate rather than growth yield is a completely different strategy; this can be easily understood because the yield does not fix the rates of the metabolic processes. Selection for yield only occurs in the absence of competition for nutrients, which is an unlikely scenario in biology. The assumption of one objective may actually not always reflect reality: the occurrence of trade-offs between two metabolic objectives may cause cells to optimize both of them simultaneously (possibly, with different weights), leading to Pareto optimization problems [30].

11 Concluding remarks

We provided an overview of the most common mathematical techniques used in the stoichiometric analysis of metabolic networks. We have not described in any detail the application of these techniques to biological problems, which is found elsewhere [31–33]. These applications to biology are the reasons for the existence of pathway analysis, and there are a number of success stories [34, 35]. Yet, the simplifications and subsequent limitations of the described techniques are also clear and extensions to pathway analysis methods include the incorporation of dynamics (such as in dynamic FBA [36]), additional constraints (such as space or resource limitations [37–39], multidimensional optimality [30], and extensions to multi-species FBA [40–42]. It is therefore to be expected that such analysis will penetrate biology in increasingly many ways to provide rigorous and quantitative hypotheses and fundamental understanding.

The authors declare no conflict of interest.

12 References

[1] Roels, J. A., Application of macroscopic principles to microbial metabolism. *Biotechnol. Bioeng.* 1980, *103*, 2–59.

[2] Thiele, I., Palsson, B. O., A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 2010, *5*, 93–121.

[3] Cornish-Bowden A., *Fundamentals of Enzyme Kinetics*. Portland Press Limited, 1995.

[4] Hofmeyr, J., Metabolic control analysis in a nutshell. In *Proceedings of the 2nd International Conference on Systems Biology*, California Institute of Technology, 2001, pp. 291–300.

[5] Kholodenko, B. N., Sauro, H. M., Westerhoff, H. V., Control by enzymes, coenzymes and conserved moieties: generalization of the connectivity theorem of metabolic control analysis. *Eur. J. Biochem.* 1994, *225*, 179–186.

[6] Reder, C., Metabolic control theory: A structural approach. *J. Theor. Biol.* 1988, *135*, 175–201.

[7] Famili, I., Palsson, B. O., The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys. J.* 2003, *85*, 16–26.

[8] Sauro, H. M., Ingalls, B., Conservation analysis in biochemical networks: Computational issues for software writers. *Biophys. Chem.* 2004, *109*, 1–15.

[9] Schuetz, R., Kuepfer, L., Sauer, U., Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli. Mol. Syst. Biol.* 2007, *3*, 119.

[10] Fell, D. A., Small, J. R., Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem. J.* 1986, *238*, 781–786.

[11] Savinell, J. M., Palsson, B. O., Optimal selection of metabolic fluxes for in vivo measurement. I. Development of mathematical methods. *J. Theor. Biol.* 1992, *155*, 201–214.

[12] Feist, A. M., Herrgard, M. J., Thiele, I., Reed, J. L. et al., Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol. 2009, 7*, 129–143.

[13] Mahadevan, R., Schilling, C. H., The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 2003, *5*, 264–276.

[14] Burgard, A. P., Nikolaev, E. V., Schilling, C. H, Maranas, C. D., Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res. 2004, 14*, 301–312.

[15] Goffin, P., van de Bunt, B., Giovane, M., Leveau J. H. et al., Understanding the physiology of *Lactobacillus plantarum* at zero growth. *Mol. Syst. Biol. 2010, 6*, 413.

[16] Teusink, B., Wiersma, A., Molenaar, D., Francke, C. et al., Analysis of growth of *Lactobacillus plantarum* WCFS1 on a complex medium using a genome-scale metabolic model. *J. Biol. Chem*. 2006, *281*, 40041–40048.

[17] Palsson, B., *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, Cambridge 2006.

[18] Price, N. D., Reed, J. L., Palsson, B. O., Genome-scale models of microbial cells: Evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2004, *2*, 886–897.

[19] Schuster, S., Dandekar, T., Fell, D. A., Detection of elementary flux modes in biochemical networks: A promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* 1999, *17*, 53–60.

[20] Schilling, C. H., Letscher, D., Palsson, B. O., Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 2000, *203*, 229–248.

[21] Price, N. D., Papin, J. A., Palsson, B. O., Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. *Genome Res.* 2002, *12*, 760–769.

[22] Wiback, S. J., Palsson, B. O., Extreme pathway analysis of human red blood cell metabolism. *Biophys. J.* 2002, *83*, 808–818.

[23] Papin, J. A., Price, N. D., Palsson, B. O., Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Res*. 2002, *12*, 1889–1900.

[24] Papin, J. A., Stelling, J., Price, N. D., Klamt, S. et al., Comparison of network-based pathway analysis methods. *Trends Biotechnol.* 2004, *22*, 400–405.

[25] Klamt, S. Stelling, J., Two approaches for metabolic pathway analysis? *Trends Biotechnol.* 2003, *21*, 64–69.

[26] Grötschel, M., Lovasz, L., Schrijver, A., Geometric algorithms and combinatorial optimization. Springer-Verlag 1988.

[27] Kelk, S. M., Olivier, B. G., Stougie, L., Bruggeman, F. J., Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Sci Rep.* 2012, *2, 580*.

[28] Segre, D., Vitkup, D., Church, G. M., Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* 2002, *99*, 15112–15117.

[29] Lee, J. M., Gianchandani, E. P., Papin, J. A., Flux balance analysis in the era of metabolomics. *Briefings Bioinf.* 2006, *7*, 140–150.

[30] Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. et al., Multidimensional optimality of microbial metabolism. *Science* 2012, *336*, 601–604.

[31] Raman, K., Chandra, N., Flux balance analysis of biological systems: Applications and challenges. *Briefings Bioinf.* 2009, *10*, 435–449.

[32] Oberhardt, M. A., Palsson, B. O., Papin, J. A., Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* 2009, *5,* 320.

[33] Feist, A. M., Palsson, B. O., The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* 2008, *26*, 659–667.

[34] Ibarra, R. U., Edwards, J. S., Palsson, B. O., *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature,* 2002, *420*, 186–189.

[35] Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A. et al., Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* 2010, *6*, 390.

[36] Mahadevan, R., Edwards, J. S., Doyle, F. J., Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* 2002, *83*, 1331–1340.

[37] Beg, Q. K., Vazquez, A., Ernst, J., de Menezes, M. A. et al., Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc. Natl. Acad. Sci. USA* 2007, *104*, 12663–12668.

[38] van Hoek, M. J., Merks, R. M., Redox balance is key to explaining full vs. partial switching to low-yield metabolism. *BMC Syst. Biol*. 2012, *6*, 22.

[39] Branco Dos Santos, F., de Vos, W. M., Teusink, B., Towards metagenome-scale models for industrial applications – the case of lactic acid bacteria. *Curr. Opin. Biotechnol. 2013, 24*, 200–206.

[40] Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N. et al., Metabolic modeling of a mutualistic microbial community. *Mol. Syst. Biol.* 2007, *3*, 92.

[41] Zomorrodi, A. R., Maranas, C. D., OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput. Biol.* 2012, *8,* e1002363.

[42] Roling, W. F. M., Ferrer, M., Golyshin, P. N., Systems approaches to microbial communities and their functioning. *Curr. Opin. Biotechnol. 2010, 21*, 532–538.