Flux Balance Analysis

Images and text from:
E. Klipp, Systems Biology in Practice, Wiley-VCH, 2005 – Ch 5
Edwards JS, Palsson BO*, How will bioinformatics influence metabolic engineering?*
Biotechnol Bioeng. 1998 Apr 20-May 5;58(2-3):162-9.

Summary

"I thought there couldn't be anything as complicated as the universe, until I started reading about the cell."

(Systems Biologist Eric de Silva -astrophysicist by training-, Imperial College London.)

- Today:
  - A quick revision of the law of mass action
  - How to represent metabolic networks:
    - Stoichiometric coefficients
    - The stoichiometric matrix
    - System equations
  - A way of simplifying the cell metabolic complexity:
    - Flux Balance Analysis

A quick revision: The Law of Mass Action (1)

- The Law of Mass Action (Waage and Guldberg, 1864) states that the reaction rate is proportional to the probability of a collision of the reactants.
- This probability is in turn proportional to the concentration of reactants, to the power of the molecularity: e.g. the number in which they enter the specific reaction.

$$S_1 + S_2 \rightleftharpoons 2P$$

- For a reaction like:
- the reaction rate is:

$$v = v_+ - v_- = k_+ S_1 \cdot S_2 - k_- P^2$$

- *(where v is the net rate, v+ the rate of the forward reaction, v- the rate of the backward reaction, and k+ and k- are the respective proportionality factors, the so-called kinetic or rate constants)*

- A more general formula for substrate concentrations S_i, and product concentrations P_j is:

$$v = v_+ - v_- = k_+ \prod_i S_i^{m_i} - k_- \prod_j P_j^{m_j}$$

- where m_i, and m_j denote the respective molecularities of S_i and P_j

The Law of Mass Action (2)

- The equilibrium constant Keq characterizes the ratio of substrate and product concentrations in equilibrium (Seq and Peq), that is, the state with equal forward and backward rates.
- The rate constants are related to Keq:
-

$$Keq = \frac{k_+}{k_-} = \frac{\prod P_{eq}}{\prod S_{eq}}$$

- The dynamics of the concentrations can be described by Ordinary Differential Equations (ODE), e.g. for the S1+S2→ 2P reaction:
- (The time course of S1,S2 and P is obtained by integration of these ODEs)

$$1.\ \frac{d}{dt}S_1 = \frac{d}{dt}S_2 = -v$$

$$2.\ \frac{d}{dt}P = 2v$$

Text from: E. Klipp, Systems Biology in Practice, 2005

The Law of Mass Action (3)

- *An example*
- The kinetics of a simple decay (molecular destruction) such as:

$$S \longrightarrow$$

- is described by:

$$1.\ v = kS$$

$$2.\ \frac{d}{dt}S = -kS$$

- Integration of this ODE from time t = 0 with the initial concentration S_0 to an arbitrary time t with concentration S(t) yields the temporal expression:

$$\int_{S_0}^{S} \frac{dS}{S} = \int_{t=0}^{t} k\, dt \quad \text{or} \quad S(t) = S_0 e^{-kt}$$

Stoichiometric * coefficients

* *From the Greek word stoikheion "element" and metriā "measure," from metron*

- Stoichiometric coefficients denote the proportion or substrates and products involved in a reaction.
- For the example:

$$S_1 + S_2 \rightleftharpoons 2P$$

- The stoichiometric coefficients of S1 S2 and P are -1, -1, and 2.
- The assignment of stoichiometric coefficients **is not unique:**
  - If we consider that for producing a mole of P, half a mole of each S1 and S2 have to be used, we could choose -1/2, -1/2, and 1.
  - Or, if we change the direction of the reaction, then we may choose 1,1, and -2.

ODEs for one and many reactions…

- For our example reaction (with the first choice of stoichiometric coefficients), we have the (already seen) ODEs:

$$\frac{d}{dt}S_1 = \frac{d}{dt}S_2 = -v \text{ and } \frac{d}{dt}P = 2v$$

- *(This means that the degradation of S1 with rate v is accompanied by the degradation of S2 with the same rate and by the production of P with the double rate.)*

- For a metabolic network consisting of m substances and r reactions, the systems dynamics is described by systems equations *(or balance equations, since the balance of substrate production and degradation is considered)*:

$$\frac{dS_i}{dt} = \sum_{j=1}^{r} n_{ij} v_j \quad \text{for} i = 1,\ldots,m$$

- The quantities n_{ij} are the stoichiometric coefficients of metabolite i in reaction j. *(We assume that the reactions are the only reason for concentration changes and that no mass flow occurs due to convection or to diffusion.)*

Stoichiometric Matrix (1)

- The stoichiometric coefficients n_{ij} assigned to the substances S_i and the reactions V_j can be combined into the so-called stoichiometric matrix:

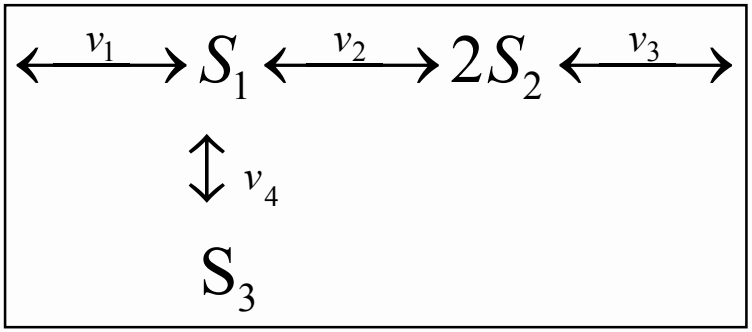
$$N = \{n_{ij}\} \text{ for } i = 1,\ldots,m \text{ and } j = 1,\ldots,r$$

- where each column belongs to a reaction and each row to a substance.

Text from: E. Klipp, Systems Biology in Practice, 2005

Stoichiometric Matrix (2)

- For the simple network:



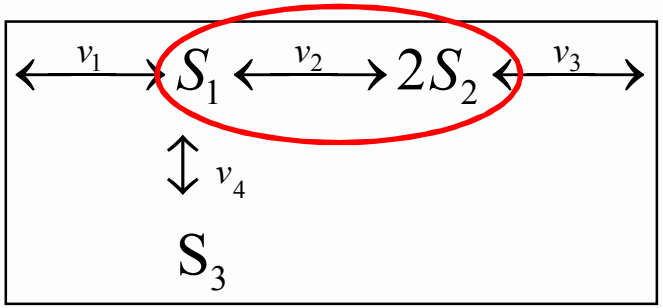
- the stoichiometric matrix is:

$$N = \begin{pmatrix} 1 & -1 & 0 & 1 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- *Note that in the network all reactions may be reversible. In order to determine the signs of N, the direction of the arrows is artificially assigned as positive "from left to right" and "from the top down."*

Stoichiometric Matrix (3)

- For the simple network:



When the reaction with rate v_2 occurs:

Row 1: 1 molecule of S_1 disappears (-1)

Row 2: 2 molecules of S_2 are produced (2)

Row 3: while S_3 stays the same (0).

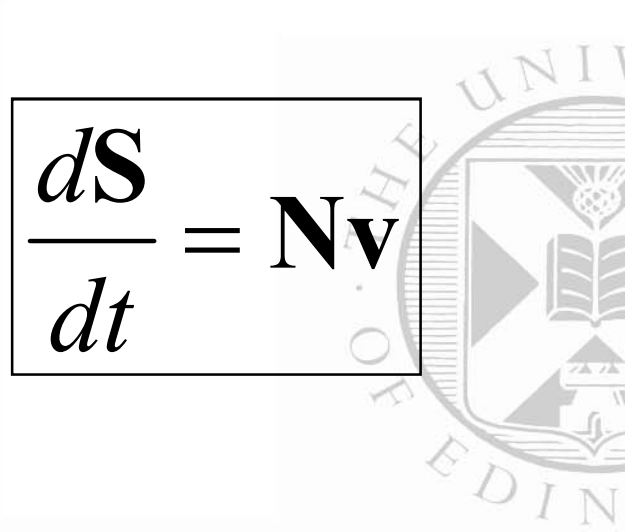
$$\text{reaction}: \quad \begin{matrix} v_1 & v_2 & v_3 & v_4 \end{matrix}$$

$$N = \begin{pmatrix} 1 & -1 & 0 & 1 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix}$$

Mathematical description of a metabolic system (1)

- Consists of:
  - a vector of concentration values, $\mathbf{S} = (S_1, S_2, \ldots, S_n)^T$
  - a vector of reaction rates, $\mathbf{v} = (v_1, v_2, \ldots, v_n)^T$
  - a parameter vector, $\mathbf{p} = (p_1, p_2, \ldots, p_n)^T$
  - and the stoichiometric matrix ***N***.
- With the above, the balance equation becomes:

$$\frac{d\mathbf{S}}{dt} = \mathbf{N}\mathbf{v}$$



Example 5-5 *(see also the appendix)*

For our running example (Example 5-1) of the upper glycolysis model, the concentration vector is

$$\boldsymbol{S} = \begin{pmatrix} Gluc6P \\ Fruc6P \\ Fruc1,6P_2 \\ ATP \\ ADP \\ AMP \end{pmatrix}. \tag{5-69}$$

the vector of reaction rates is $\boldsymbol{v} = (v_1, v_2, \ldots, v_8)^{\mathrm{T}}$, the parameter vector is given by

$$\mathrm{p} = \left(Glucose,\ V_{\max,1},\ K_{ATP,1},\ K_{Glucose,1},\ k_2,\ V^f_{\max,3},\ V^r_{\max,3},\ K_{Gluc6P,3},\ K_{Fruc6P,3},\ V_{\max,4},\ K_{F6P,4},\ \kappa_4,\ k_5,\ k_6,\ k_7,\ k_{8f},\ k_{8r}\right)^{\mathrm{T}}, \tag{5-70}$$

and the stoichiometric matrix reads

$$\boldsymbol{N} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ -1 & -1 & 0 & -1 & 0 & 1 & -1 & -1 \\ 1 & 1 & 0 & 1 & 0 & -1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}. \tag{5-71}$$

Text from: E. Klipp, Systems Biology in Practice, 2005

Flux Balance Analysis – FBA

- The stoichiometric analysis can be constrained in various ways to simplify the resolution of the system, and to limit the solution space.
- One of the techniques used to analyse the complete metabolic genotype of a microbial strain is FBA:
  - relies on balancing metabolic fluxes
  - is based on the fundamental law of mass conservation
  - is performed under steady-state conditions *(an example of constraint…)*
  - requires information only about:
    1. the stoichiometric of metabolic pathways,
    2. metabolic demands,
    3. and a few strain specific parameters
  - It does NOT require enzymatic kinetic data

Images and text from: Edwards JS, Palsson BO, *How will bioinformatics influence metabolic engineering? Biotechnol Bioeng. 1998 Apr 20-May 5;58(2-3):162-9.*

Flux Balance (1)

- The fundamental principle in FBA is the conservation of mass.
- A **flux balance** can be written for each metabolite (Xi) within a metabolic system to yield the dynamic mass balance equations that interconnect the various metabolites.

- Equating the rate of accumulation of Xi to its net rate of production, the dynamic mass balance for Xi is:

$$1.\ \frac{dX_i}{dt} = V_{syn} - V_{deg} - (V_{use} - V_{trans})$$

- where the subscripts, syn and deg refer to the metabolic synthesis and degradation of metabolite Xi.
- The uptake or secretion flux, Vtrans, can be determined experimentally.
- The growth and maintenance requirements, Vuse, can be accurately estimated from cellular composition

- Equation (1) therefore, can be written as: (b_i is the net transport out of our defined metabolic system):

$$2.\ \frac{dX_i}{dt} = V_{syn} - V_{deg} - b_i$$

Flux Balance (2)

- For a metabolic network that contains m metabolites and n metabolic fluxes, all the transient material balances can be represented by a single matrix equation:

$$3.\ \frac{d\mathbf{X}}{dt} = \mathbf{S}\cdot\mathbf{v} - \mathbf{b}$$

- where X is an m dimensional vector of metabolite amounts per cell, v is the vector of n metabolic fluxes, S is the stoichiometric m × n matrix, and b is the vector of known metabolic demands.

- The element S_{ij} is the stoichiometric coefficient that indicates the amount of the i^{th} compound produced per unit of flux of the j^{th} reaction.

Flux Balance (3)

- The time constants characterizing metabolic transients are typically very rapid compared to the time constants of cell growth, and the transient mass balances can be simplified to only consider the steady-state behaviour (dX/dt=0).
- Eliminating the time derivative in Equation (3) yields:

$$4.\ \mathbf{S}\cdot\mathbf{v}=\mathbf{b}$$

- (This equation simply states that over long periods of time, the formation of fluxes of a metabolite must be balanced by the degradation fluxes. Otherwise, significant amounts of the metabolite will accumulate inside the metabolic network.)

- *Note that this balance equation is formally analogous to Kirchhoff's current law used in electrical circuit analysis…*

Under-determination

- The flux-balance equation is typically **under-determined** ($m < n$), and cannot be solved using Gaussian elimination.
- Thus, additional information is needed to solve for all the metabolic fluxes: various techniques have been used to solve this equation.

- Several researchers have made sufficient measurements of **external fluxes** to either completely determine or over-determine the system
- In order for measurements of only the external fluxes to completely determine the system, additional assumptions are required, such as neglecting certain reactions occurring within the cell.

- Also **internal metabolic fluxes** have been measured and used to determine metabolic flux distributions.
- However, the measurement of internal fluxes is not always practical, and these measurements can only allow for the determination of the metabolic fluxes in *subsystems* of the metabolic network.

The null space of **S**

- For completely sequenced organisms, the cellular metabolism is defined, and the cellular inventory of metabolic gene products is expressed in the stoichiometric matrix (S).
- The metabolic genotype of an organism then is defined by all the allowable reactions that can occur with a given gene set.
- Mathematically, the metabolic capabilities of a metabolic genotype is defined as the **null space of S** (see next slide).
- The null space of S is typically large, and it represents the flexibility that a cell has in determining the use of its metabolic capabilities.

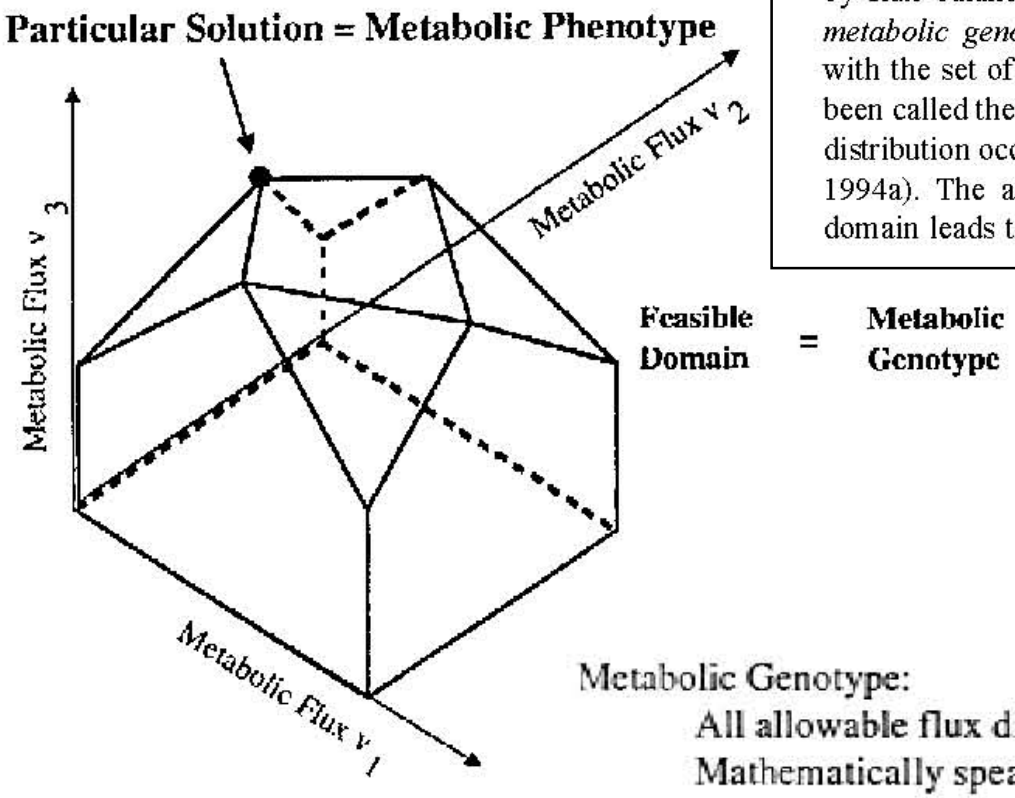


Figure 1. A schematic illustration of the solution domain that is defined by flux balance constraints. The domain illustrated has been called the *metabolic genotype,* because it represents all possible flux distributions with the set of metabolic enzymes given, whereas a specific solution has been called the *metabolic phenotype,* because it represents a particular flux distribution occurring under a defined set of conditions (Varma & Palsson, 1994a). The assessment of the metabolic flexibility in the use of this domain leads to the indicated definitions of redundancy and robustness.

Metabolic Genotype:

All allowable flux distributions by a set of metabolic genes. Mathematically speaking, the null space of S. (Varma and Palsson, 1994a)

Metabolic Phenotype:

A particular flux distribution used under given conditions. Mathematically speaking, an optimal solution obtained through linear programming. (Varma and Palsson, 1994a)

Metabolic Redundancy:

The ability of the metabolic circuit to adjust to in the absence of a gene without changes to the phenotype.

Metabolic Robustness:

The ability of the metabolic circuit to adjust to decreased fluxes through essential enzymes without changes to the phenotype.

Images from: Edwards JS, *Biotechnol Bioeng. 1998 Apr 20-May 5;58(2-3):162-9*

FBA: how to constrain the system (1)

- FBA (Varma and Pals son 1994a, 1994b; Edwards and Palsson 2000a, 2000b; Ramakrishna et al. 2001) investigates metabolism by involving **constraints** in the stoichiometric analysis:
  - The first constraint is set by the assumption of a steady state.
  - The second constraint is of a thermodynamic nature, respecting the irreversibility of reactions.
  - The third constraint may result from the limited capacity of enzymes for metabolite conversion.
  - Further constraints may be imposed by biomass composition or other external conditions.

- The constraints confine the steady-state fluxes to a feasible set but usually do not yield a unique solution.

- The determination of a particular metabolic flux distribution has been formulated as a **linear programming problem**. The idea is to maximize an objective function Z that is subject to the stoichiometric and capacity constraints *(where Cj represents weights for the individual rates)*:

$$Z = \sum_{i=1}^{r} c_i v_i \rightarrow max$$

- Examples of such objective functions are maximization of ATP production, minimization of nutrient uptake, maximal yield of a desired product or maximal growth rate.

FBA: how to constrain the system (2)

Table II. Questions that can be addressed using flux-balance analysis.

Question	Objective	Reference
What are the biochemical production capabilities?	Maximize metabolite product	Varma, Boesch, & Palsson, 1993
What is the maximal growth rate and biomass yield?	Maximize growth rate	Varma & Palsson, 1993; Varma & Palsson, 1994b
How efficiently can metabolism channel metabolites through the network?	Minimize the Euclidean norm	Bonarius et al., 1996
How energetically efficient can metabolism operate?	Minimize ATP production or minimize nutrient uptake	Majewski & Domach, 1990; Savinell & Palsson, 1992; Fell & Small, 1986
What is the tradeoff between biomass production and metabolite overproduction?	Maximize biomass production for a given metabolite production	Varma et al., 1993

Commercial applications of metabolic analysis

- Why do we need to improve bacterial growth? Shouldn't we try and kill all those nasty bacteria? Think again…
- Optimising E. coli growth is a commercial challenge.
- Recombinant E. coli bacteria *(through the addition of plasmids containing an inserted gene)* are grown in tanks and used for industrial production of:
  - insulin for diabetics
  - hepatitis B surface antigen (HBsAg) to vaccinate against the hepatitis B virus
  - factor VIII for males suffering from haemophilia A
  - erythropoietin (EPO) for treating anaemia *(and doping athletes…)*
  - human growth hormone (HGH)
  - granulocyte-macrophage colony-stimulating factor (GM-CSF) for stimulating the bone marrow after a bone marrow transplant
  - granulocyte colony-stimulating factor (G-CSF) for stimulating neutrophil production e.g., after chemotherapy
  - tissue plasminogen activator (TPA) for dissolving blood clots
  - and many others…

The E. coli example:

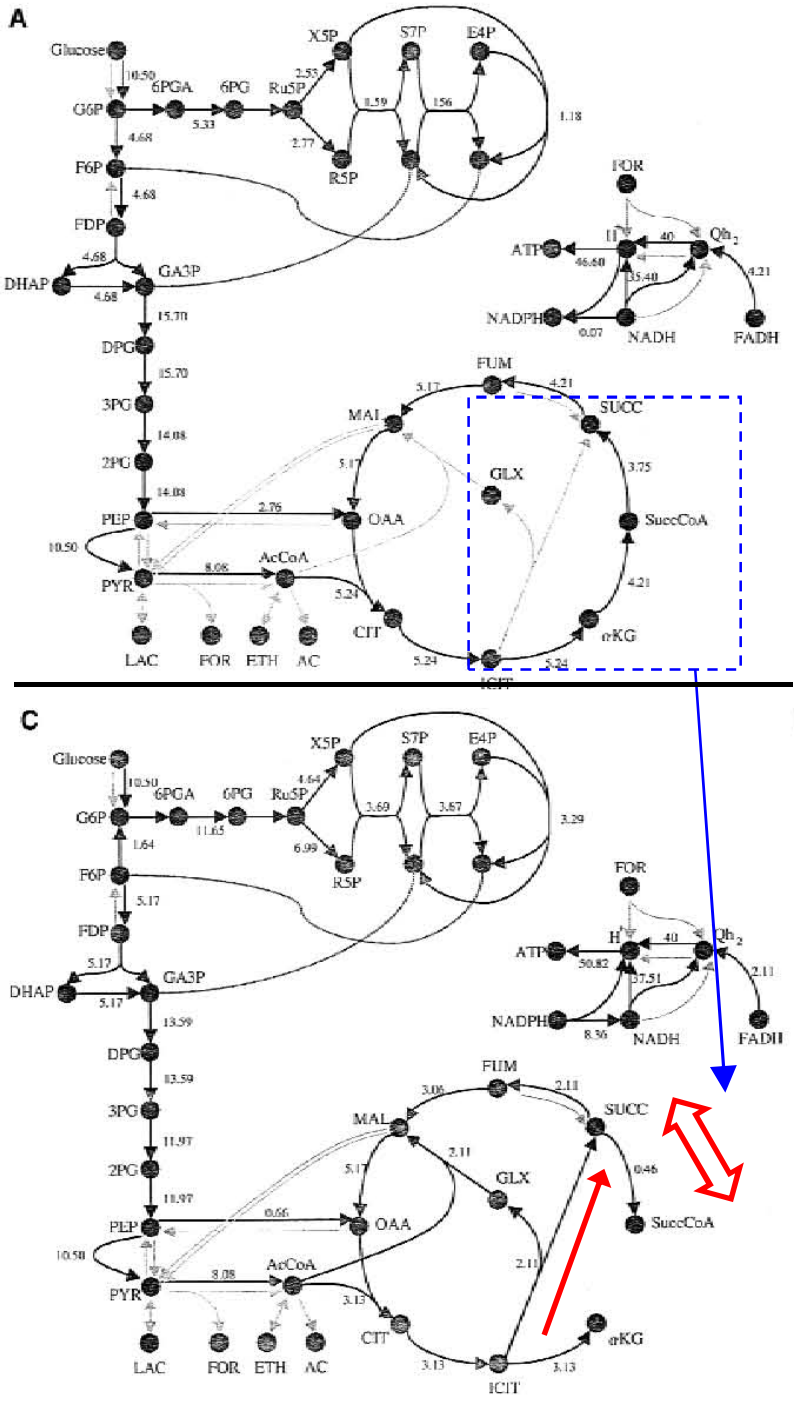
FBA to study metabolic flexibility (1)

- The complete genetic sequence for the Escherichia coli K-12 strain has been established (TIGR-Web Site, 1997).
- A genomically complete stoichiometric model for E. coli has been constructed in 1997, including 594 reactions and transport process that involve 334 metabolites.
- Using this model, the **flexibility** of E. coli metabolic flux distributions was examined during growth on glucose.
- Metabolic flexibility is the manifestation of two principal properties:
  1. Stochiometric **redundancy** means that the network can redistribute flexibly its metabolic fluxes
  2. **Robustness** is the ability of the network to adjust to decreased fluxes through a particular enzyme without significant changes in overall metabolic function

The E. coli example:

FBA to study metabolic flexibility (2)

- Figure A shows the metabolic flux distribution for the complete gene set present in E. coli for maximal biomass yield on glucose.
- The ability of E. coli to respond to the loss of an enzymatic function can be assessed by removing a gene from the basic gene set.
- Figure C shows the metabolic flux distribution when the sucA gene is removed. *The sucA gene codes for an essential component of the 2oxoglutarate dehydrogenase complex.*
- Note how the flux "by-passing" the missing enzyme increases, and how the reaction direction from SUCC to SuccCoA has been inverted



FBA: advantages and disadvantages

- Advantages:
  - It relies solely on stochiometric characteristics (can be used on any fully sequenced/characterised organism)
  - Does not need kinetic parameters (that are difficult to obtain)

- Disadvantages:
  - It does not uniquely specify the fluxes (the particular flux distribution chosen by the cell is a function of regulatory mechanisms that determine the kinetic characteristics of enzymes/enzyme expression)
  - Sometimes disagrees with experimental data (discrepancies can often be accounted for when regulatory loops are considered)
  - Cannot be used for modelling dynamic behaviour (but might be integrated with modal analysis etc.)

Reading

1. E. Klipp,

Systems Biology in Practice, Wiley-VCH, 2005

(Sections from chapter 5)

2. Edwards JS, Palsson BO,

How will bioinformatics influence metabolic engineering?

Biotechnol Bioeng. 1998 Apr 20-May 5;58(2-3):162-9. PMID: 10191386

Appendix: the glycolysis example (1)

Example 1

We will consider the first four reactions from the upper part of glycolysis as well as reactions balancing the energy currency ATP and ADP as represented in Fig. 5.1.

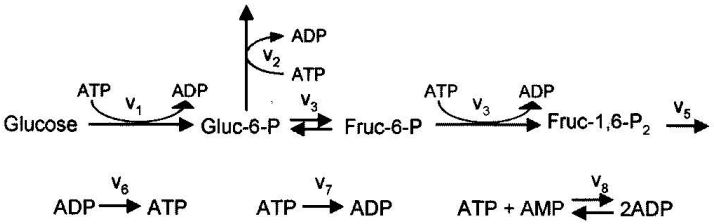


Fig. 5.1 Schematic representation of the upper part of glycolysis, i. e., the degradation of glucose in order to yield energy and building blocks for cellular processes. Abbreviations: Gluc6P: glucose-6-phosphate; Fruc6P: fructose-6-phosphate; Fruc1,6P_2: fructose-1,6-bisphosphate; ATP: adenosine-triphosphate; ADP: adenosine-diphosphate: AMP: adenosine-monophosphate. Reactions: v_1: hexokinase; v_2: consumption of glucose-6-phosphate in other pathways; v_3: phosphoglucoisomerase; v_4: phosphofructokinase; v_5: aldolase; v_6: ATP production in lower glycolysis; v7: ATP consumption in other pathways; v_8: adenylate kinase.

The ODE system for this reaction system is given by

$$\begin{aligned}
\frac{d}{dt}Gluc6P &= v_1 - v_2 - v_3 \\
\frac{d}{dt}Fruc6P &= v_3 - v_4 \\
\frac{d}{dt}Fruc1{,}6P_2 &= v_4 - v_5 \\
\frac{d}{dt}ATP &= -v_1 - v_2 - v_4 + v_6 - v_7 - v_8 \\
\frac{d}{dt}ADP &= v_1 + v_2 + v_4 - v_6 + v_7 + 2v_8 \\
\frac{d}{dt}AMP &= -v_8 \,.
\end{aligned} \tag{5-1}$$

Abbreviations are explained in the legend of Fig. 5.1. The individual rate expressions read

Appendix: the glycolysis example (2)

$$\nu_1 = \frac{V_{max,1}\, ATP(t) \cdot Glucose}{1 + \frac{ATP(t)}{K_{ATP,1}} + \frac{Glucose}{K_{Glucose,1}} + \frac{ATP(t)}{K_{ATP,1}} \cdot \frac{Glucose}{K_{Glucose,1}}} \quad \text{or} \quad \nu_1 = \frac{V_{max,1}\, ATP(t)}{K_{ATP,1} + ATP(t)} \tag{5-2}$$

$$\nu_2 = k_2\, ATP(t) \cdot Gluc6P(t) \tag{5-3}$$

$$\nu_3 = \frac{\frac{V^f_{max,3}}{K_{Gluc6P,3}} Gluc6P(t) - \frac{V^r_{max,3}}{K_{Fruc6P,3}} Fruc6P(t)}{1 + \frac{Gluc6P(t)}{K_{Gluc6P,3}} + \frac{Fruc6P(t)}{K_{Fruc6P,3}}} \tag{5-4}$$

$$\nu_4 = \frac{V_{max,4}\, (Fruc6P(t))^2}{K_{Fruc6P,4}\left(1 + \kappa\left(\frac{ATP(t)}{AMP(t)}\right)^2\right) + (Fruc6P(t))^2} \tag{5-5}$$

$$\nu_5 = k_5 Fruc1,6P_2(t) \tag{5-6}$$

$$\nu_6 = k_6 ADP(t) \tag{5-7}$$

$$\nu_7 = k_7 ATP(t) \tag{5-8}$$

$$\nu_8 = k_{8f}\, ATP(t) \cdot AMP(t) - k_{8r}\, (ADP(t))^2\,, \tag{5-9}$$

with the following parameters:

$Glucose = 12.8174\,\text{mM}$, $V_{max,1} = 1398.00\,\text{mM} \cdot \text{min}^{-1}$, $K_{ATP,1} = 0.10\,\text{mM}$, $K_{Glucose,1} = 0.37\,\text{mM}$, $V_{max,1} = 50.2747\,\text{mM} \cdot \text{min}^{-1}$

$k_2 = 2.26\,\text{mM}^{-1} \cdot \text{min}^{-1}$

$V^f_{max,3} = 140.282\,\text{mM} \cdot \text{min}^{-1}$, $V^r_{max,3} = 140.282\,\text{mM} \cdot \text{min}^{-1}$, $K_{Gluc6P,3} = 0.80\,\text{mM}$, $K_{Fruc6P,3} = 0.15\,\text{mM}$

$V_{max,4} = 44.7287\,\text{mM} \cdot \text{min}^{-1}$, $K_{Fruc6P,4} = 0.021\,\text{mM}^2$, $\kappa = 0.15$

$k_5 = 6.04662\ \text{min}^{-1}$

$k_6 = 68.48\ \text{min}^{-1}$

$k_7 = 3.21\ \text{min}^{-1}$

$k_{8f} = 432.9\,\text{mM}^{-1} \cdot \text{min}^{-1}$, $k_{8r} = 133.33\,\text{mM}^{-1} \cdot \text{min}^{-1}$

The temporal evolution of the concentrations starting from arbitrarily given values is shown in Fig. 5.2.

Appendix: the glycolysis example (3)

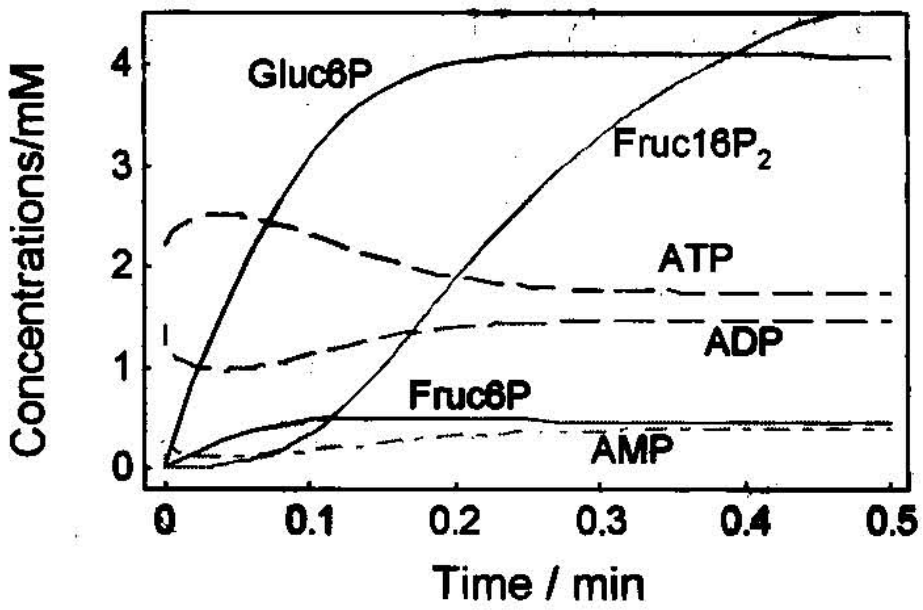


Fig. 5.2 Time courses of concentrations for the running model presented (Example 5-1). Parameters are stated there. Initial values: *Gluc*6*P*(0) = 1 mM, *Fruc*6*P*(0) = 0 mM, *Fruc*1,6P_2(0) = 0 mM, *ATP*(0) = 2.1 mM, *ADP*(0) = 1.4 mM, *AMP*(0) = 0.1 mM.