## Stoichiometric network analysis

In stoichiometric analysis of metabolic networks, one concerns the effect of the network structure on the behaviour and capabilities of metabolism.

Questions that can be tackled include:

- Discovery of pathways that carry a distinct biological function (e.g. glycolysis) from the network, discovery of dead ends and futile cycles, dependent subsets of enzymes
- Identification of optimal and suboptimal operating conditions for an organism
- Analysis of network flexibility and robustness, e.g. under gene knockouts

## Stoichiometric coefficients

Soitchiometric coefficients denote the proportion of substrate and product molecules involved in a reaction. For example, for a reaction

$$r : A + B \mapsto 2C,$$

the stoichiometric coefficients for $A, B$ and $C$ are $-1, -1$ and 2, respectively.

- Assignment of the coeefficients is not unique: we could as well choose $-1/2, -1/2, 1$ as the coefficients
- However, the relative sizes of the coeefficients remain in any valid choice.

  Note! We will denote both the name of a metabolite and its concentration by the same symbol.

## Stoichiometry and reaction rates

- The rate of change of concentration of metabolites is the most fundamental quantity in stoichiometric models
- Assume a reaction

$$r : A + B \mapsto 2C,$$

  operates at some rate or velocity $v$ (arbitrary units e.g. mol/hour)

- Then, the change of concentration of the reactants and the product are given by the reaction rate multiplied by the shoichiometric coefficients

$$\frac{dA}{dt} = -1 \cdot v, \frac{dA}{dt} = -1 \cdot v, \frac{dC}{dt} = 2 \cdot v$$

- Thus, $A$ and $B$ are consumed at the rate of the reaction, $C$ is produced at the double rate.

## Reversible reactions

- Many of metabolic reactions are reversible,

$$r : A + B \leftrightharpoons 2C,$$

  so they can work in either direction, depending on the conditions within the cell

- In stoichiometric models a reversible reaction can be modelled in two ways:
  - As a single reaction that can operate from left to right, indicated by positive reaction rate $v > 0$ or right to left, indicated by negative reaction rate $v < 0$.
  - As two separate reactions $r' : A + B \mapsto 2C$ and $r'' : 2C \mapsto A + B$, both with non-negative reaction rates $v', v'' \geq 0$.

## Concentration and rate vectors

- Let the reaction $R_i$ operate with rate $v_i$
- We collect the individual reaction rates to a *rate vector* $\mathbf{v} = (v_1, \ldots, v_r)^T$
- Similarly, the *concentration vector* $X(t) = (X_1(t), \ldots, X_r(t))^T$ contains the concentration of each metabolite in the system at time $t$

## Stoichiometric vector and matrix

- The stoichiometric coefficients of a reaction are collected to a vector $s_r$
- In $s_r$ there is a one position for each metabolite in the metabolic system, and the stoichiometric co-efficient of the reaction are inserted to appropriate positions, e.g. for the reaction

$$r : A + B \mapsto 2C,$$

$$s_r = \begin{array}{c} \cdot \\ \cdot \\ A \\ \cdot \\ \cdot \\ B \\ \cdot \\ \cdot \\ C \end{array} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

## Stoichiometric matrix

- The stoichiometric vectors can be combined into the stoichiometric matrix $S$.
- In the matrix $S$, the is one row for each metabolite and one column for each reaction.
- The coefficients $s_{*j}$ along the $j$'th column are the stoichiometric coeefficients of of the reaction $j$.
- The coefficients along the $i$'th row denote the relationship between the metabolite $M_i$'s concentration and the reactions consuming or producing it.

$$\mathbf{S} = \begin{bmatrix} s_{11} & \cdots & s_{1j} & \cdots & s_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{i1} & \cdots & s_{ij} & \cdots & s_{ik} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{l1} & \cdots & s_{lj} & \cdots & s_{lk} \end{bmatrix}$$

## Example: stoichiometric matrix

- Consider the set of reactions from the penthose-phospate pathway:
- The stoichiometric matrix is a 10-by-7 matrix:

$R_1$: $\beta$G6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH

$R_2$: 6PGL + $H_2O$ $\overset{pgl}{\Rightarrow}$ 6PG

$R_3$: 6PG + NADP$^+$ $\overset{gnd}{\Rightarrow}$ R5P + NADPH

$R_4$: R5P $\overset{rpe}{\Rightarrow}$ X5P

$R_5$: $\alpha$G6P $\overset{gpi}{\Leftrightarrow}$ $\beta$G6P

$R_6$: $\alpha$G6P $\overset{gpi}{\Leftrightarrow}$ $\beta$F6P

$R_7$: $\beta$G6P $\overset{gpi}{\Leftrightarrow}$ $\beta$F6P

$$S = \begin{array}{c} \\ \beta G6P \\ \alpha G6P \\ \beta F6P \\ 6PGL \\ 6PG \\ R5P \\ X5P \\ NADP^+ \\ NADPH \\ H_2O \end{array} \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

## Systems equations (1/2)

- Suppose that reactions $R_1, R_5$ and $R_7$ operate at rates 2, 1 (left to right) and $-2$ (right to left), respectively
- Multiply the reaction rates with stoichiometric coefficients to obtain the rates of change of concentration of $\beta$G6P caused by each reaction:
  $R_1 : (-1) \cdot 2 = -2$, $R_5 : 1 \cdot 1 = 1$, $R_7 : (-1) \cdot (-2) = 2$
- The *net rate* of change $\beta$G6P is therefore

$$\frac{d[\beta G6P]}{dt} = -2 + 1 + 2 = 1,$$

  thus the system is accumulating $\beta$G6P

$R_1$: $\beta$G6P + NADP$^+$ $\overset{zwf}{\Rightarrow}$ 6PGL + NADPH
$R_5$: $\alpha$G6P $\overset{gpi}{\Leftrightarrow}$ $\beta$G6P
$R_7$: $\beta$G6P $\overset{gpi}{\Leftrightarrow}$ $\beta$F6P

Stoichiometric coefficients from matrix $S$

$$S_{\beta G6P} = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$

## Systems equations (2/2)

In a network of $n$ metabolites and $r$ reactions, the dynamics of the system are characterized by the systems equations

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j, \text{ for } i = 1, \ldots, n$$

- $X_i$ is the concentration of the $i$th metabolite
- $v_j$ is the rate of the $j$th reaction and
- $s_{ij}$ is the stoichiometric coefficient of $i$th metabolite in the $j$th reaction.

Intuitively, each system equation states that the rate of change of concentration of a is the sum of metabolite flows to and from the metabolite.

## Systems equation example

- Assume our example metabolic network has the following rate vector $\mathbf{v} = (1, 1, 0, 0, 1, 0, 0)$
- Let us compute the rate of change for metabolites

$R_1$: $\beta$G6P + NADP$^+$ $\stackrel{zwf}{\Rightarrow}$ 6PGL + NADPH
$R_2$: 6PGL + $H_2O$ $\stackrel{pgl}{\Rightarrow}$ 6PG
$R_3$: 6PG + NADP$^+$ $\stackrel{gnd}{\Rightarrow}$ R5P + NADPH
$R_4$: R5P $\stackrel{rpe}{\Rightarrow}$ X5P
$R_5$: $\alpha$G6P $\stackrel{gpi}{\Leftrightarrow}$ $\beta$G6P
$R_6$: $\alpha$G6P $\stackrel{gpi}{\Leftrightarrow}$ $\beta$F6P
$R_7$: $\beta$G6P $\stackrel{gpi}{\Leftrightarrow}$ $\beta$F6P

$$\frac{d\beta G6P}{dt} = -1v_{R_1} + 1v_{R_5} - 1v_{R_7} = 0$$
$$\frac{d\alpha G6P}{dt} = -1v_{R_5} - 1v_{R_6} = -1 \Rightarrow \text{net consumption!}$$
$$\frac{d\beta F6P}{dt} = 1v_{R_6} + 1v_{R_7} = 0$$
$$\frac{d6GPL}{dt} = 1v_{R_1} - 1v_{R_2} = 0$$
$$\frac{d6PG}{dt} = 1v_{R_2} - 1v_{R_3} = 1 \Rightarrow \text{net production!}$$
$$\frac{dR5P}{dt} = 1v_{R_3} - 1v_{R_4} = 0$$
$$\frac{dX5P}{dt} = 1v_{R_4} = 0$$
$$\frac{dNADPH}{dt} = 1v_{R_1} + 1v_{R_3} = 1 \Rightarrow \text{net production!}$$
$$\frac{dNADP^+}{dt} = -1v_{R_1} - 1v_{R_3} = -1 \Rightarrow \text{net consumption!}$$
$$\frac{dH_20}{dt} = -1v_{R_2} = -1 \Rightarrow \text{net consumption!}$$

## Systems equations in matrix form

- The systems equation can be expressed in vector form as

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = S_i^T \mathbf{v},$$

  where $S_i$ contains the stoichiometric coefficients of a single metabolite, that is a row of the stoichiometric matrix

- All the systems equations of different equations together can then be expressed by a matrix equation

$$\frac{d\mathbf{X}}{dt} = S\mathbf{v},$$

- Above, the vector

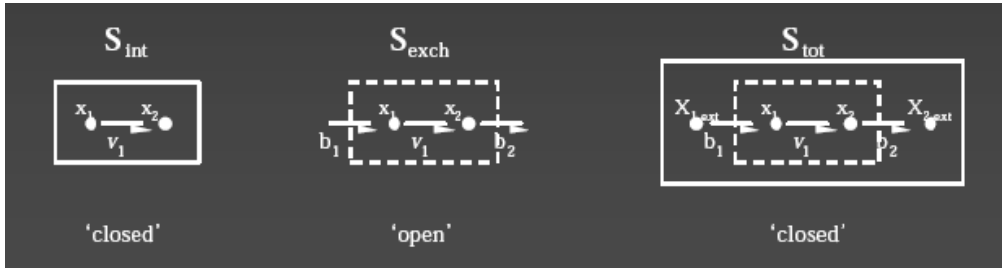
$$\frac{d\mathbf{X}}{dt} = \left( \frac{d\mathbf{X_1}}{dt}, \ldots, \frac{d\mathbf{X_n}}{dt} \right)^T$$

  collects the rates of concentration changes of all metabolites

## Defining the system boundary

When analysing a metabolic system we need to consider what to include in our system

1. Metabolites and reactions internal to the cell: this is a closed system with no matter flow to and from outside the system (cell)
2. (1) + exchange reactions transporting matter accross the cell membrane: this is an open system with the possibility of matter flow to and from the system
3. (1) + (2) + Metabolites outside the cell: This is again closed system with no matter flow to and from the system (cell + external metabolites)



## Defining the system boundary

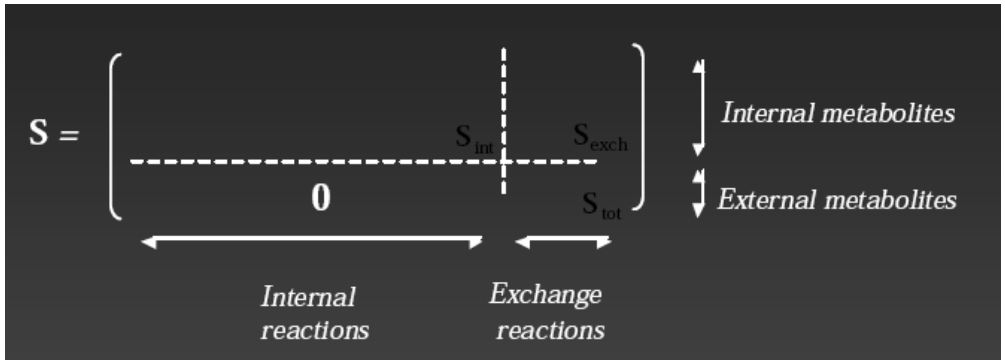
- Our example system is a closed one: we do not have exchange reactions carrying to or from the system.
- We can change our system to an open one, e..g by introducing a exchange reaction $R_8 :\Rightarrow \alpha G6P$ feeding $\alpha$G6P into the system and another reaction $R_9 : X5P \Rightarrow$ to push $X5P$ out of the system

$R_1$: $\beta$G6P + $NADP^+ \overset{zwf}{\Rightarrow}$ 6PGL + NADPH
$R_2$: 6PGL + $H_2O \overset{pgl}{\Rightarrow}$ 6PG
$R_3$: 6PG + $NADP^+ \overset{gnd}{\Rightarrow}$ R5P + NADPH
$R_4$: R5P $\overset{rpe}{\Rightarrow}$ X5P
$R_5$: $\alpha$G6P $\overset{gpi}{\Leftrightarrow}$ $\beta$G6P
$R_6$: $\alpha$G6P $\overset{gpi}{\Leftrightarrow}$ $\beta$F6P
$R_7$: $\beta$G6P $\overset{gpi}{\Leftrightarrow}$ $\beta$F6P

## System boundary and the stoichiometric matrix

The stoichiometric matrix $S = S_{tot}$ can be partitioned into according the system boundary:

- $S_{int}$ contains the stoichiometric coefficients of internal metabolites with respect to internal reactions
- $S_{exch}$ contains the stoichiometric coefficients of internal metabolites w.r.t. exchange reactions



## Example

The stoichiometric matrix of our extended example contains two extra columns, corresponding to the exchange reactions $R_8 :\Rightarrow \alpha G6P$ and $R_9 : X5P \Rightarrow$

$$
\begin{array}{c}
\beta G6P \\
\alpha G6P \\
\beta F6P \\
6PGL \\
6PG \\
R5P \\
X5P \\
NADP^{+} \\
NADPH \\
H_2O
\end{array}
\begin{bmatrix}
-1 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & -1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\
-1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

## Steady state analysis (1/2)

- Most applications of stoichiometric matrix assume that the system is in so called steady state
- In a steady state, the concentrations of metabolites remain constant over time, thus the derivative of the concentration is zero:

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = 0, \text{ for } i = 1, \ldots, n$$

- The requires the production equal consumption of each metabolite, which forces the reaction rates to be invariant over time.
- The steady-state reaction rates are also called the *fluxes*
- Note: Biologically, live cells do not exhibit true steady states, but in suitable conditions (e.g. continuous bioreactor cultivations) steady-state can be satisfied approximately. Pseudo-steady states or quasi-steady states are formally correct terms, but rarely used

## Steady state analysis (2/2)

- The requirements of non-changing concentrations

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} s_{ij} v_j = 0, \text{ for } i = 1, \ldots, n$$

  constitute a set of linear equations constraining to the reaction rates $v_j$.

- We can write this set of linear constraints in matrix form with the help of the stoichiometric matrix $S$ and the reaction rate vector $\mathbf{v}$

$$\frac{d\mathbf{X}}{dt} = S\mathbf{v} = \mathbf{0},$$

- A reaction rate vector $\mathbf{v}$ satisfying the above is called a *flux* vector.

## Null space of the stoichiometrix matrix (1/2)

- Any flux vector $\mathbf{v}$ that the cell can maintain in a steady-state is a solution to the system of equations

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

- The null space of the stoichiometric matrix

$$\mathcal{N}(S) = \{\mathbf{u}|S\mathbf{u} = 0\}$$

  contains all valid flux vectors

- Therefore, studying the null space of the stoichiometric matrix can give us important information about the cell's capabilities

## Null space of the stoichiometric matrix (2/2)

The null space $\mathcal{N}(S)$ is a linear vector space, so all properties of linear vector spcaes follow, e.g:

- $\mathcal{N}(S)$ contains the zero vector, and closed under linear combination: $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{N}(S) \implies \alpha_1 \mathbf{v}_1 + \alpha \mathbf{v}_2 \in \mathcal{N}(S)$
- The null space has a basis $\{\mathbf{k_1}, \ldots, \mathbf{k}_q\}$, a set of $q \leq \min(n, r)$ linearly independent vectors, where $r$ is the number of reactions and $n$ is the number of metabolites.
- The choice of basis is not unique, but the number $q$ of vector it contains is determined by the rank of $S$.

## Null space and feasible steady state rate vectors

- The kernel $K = (\mathbf{k}_1, \ldots, \mathbf{k}_q)$ of the stoichiometric matrix formed by the above basis vectors has a row corresponding to each reaction. (Note: the term 'kernel' here has no relation to kernel methods and SVMs)
- $K$ characterizes the feasible steady state reaction rate vectors: for each feasible flux vector $\mathbf{v}$, there is a vector $\mathbf{b} \in \mathbb{R}^q$ such that $K\mathbf{b} = \mathbf{v}$
- In other words, any steady state flux vector is a linear combination

$$b_1\mathbf{k}_1 + \cdots + b_q\mathbf{k}_q$$

of the basis vectors of $\mathcal{N}(N)$.

## Identifying dead ends in metabolism

- From the matrix $K$, one can identify reactions that can only have zero rate in a steady state.
- Such reactions may indicate a dead end: if the reaction is not properly connected the rest of the network, the reaction cannot operate in a steady state
- Such reactions necessarily have the corresponding row $K_j$ identically equal to zero, $K_j = 0$

## Proof outline

- This can be easily proven by contradiction using the the equation $K\mathbf{b} = \mathbf{v}$:
- Assume reaction $R_j$ is constrained to have zero rate in steady state, but assume for some $i$, $k_{ji} \neq 0$.
- Then we can pick the $i$'th basis vector of $K$ as the feasible solution $\mathbf{v} = \mathbf{k}_i$.
- Then $v_j = k_{ji} \neq 0$ and the $j$th reaction has non-zero rate in a steady state.

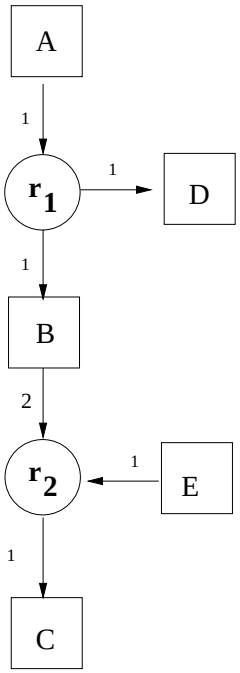
## Enzyme subsets (1/2)

- An enzyme subset is a group of enzymes which, in a steady state, must always operate together so that their reaction rates have a fixed ratio.
- Consider a pair of reactions $R_1$ and $R_2$ in the metabolic network that form a linear sequence.
- Let $B$ be a metabolite that is an intermediate within the pathway produced by $R_1$ and consumed by $R_2$ for which the steady-state assumption holds. Due to the steady state assumption, it must hold true that

$$v_1 s_{i1} + v_2 s_{i2} = 0$$

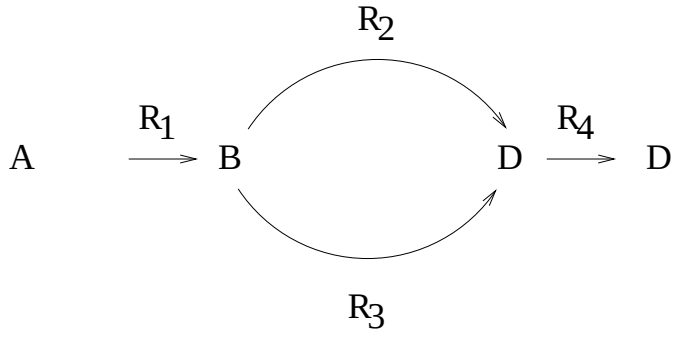
giving $v_2 = -v_1 s_{i1}/s_{i2}$.

- That is, the rates of the two reactions are linearly dependent.



## Enzyme subsets (2/2)

Also other than linear pathways may be force to operate in 'lock-step'. In the figure below t, $R1$ and $R4$ form an enzyme subset, but $R2$ and $R3$ are not in that subset.



## Identifying enzyme subsets

- Enzyme subsets are easy to recognize from the matrix $K$: the rows corresponding to an enzyme subset are scalar multiples of each other.
- That is, there is a constant $\alpha$ that satisfies $K_j = \alpha K_{j'}$ where $K_j$ denotes the $j$'th row of the kernel matrix $K$
- This is again easy to see from the equation

$$K\mathbf{b} = \mathbf{v}.$$

## Proof outline

- Assume that reactions along rows $j, j'$ in $K$ correspond to an enzyme subset.
- Now assume contrary to the claim that the rows are not scalar multiples of each other. Then we can find a pair of columns $i, i'$, where $K_{ji} = \alpha K_{j'i}$ and $K_{ji'} = \beta K_{j'i'}$ and $\alpha \neq \beta$.
- Both columns $i, i'$ are feasible flux vectors. By the above, the rates of $j$ and $j'$ differ by factor $\alpha$ in the flux vector given by the column $i$ and by factor $\beta$ in the flux vector given by the column $i'$.
- Thus the ratio of reaction rates of $j, j'$ can vary and the reactions are not force to operate with a fixed ratio.

## Independent components

Finally, the matrix $K$ can be used to discover subnetworks that can work independently from the rest of the metabolism, in a steady state.

Such components are characterized by a block-diagonal $K$: $K_{ji} \neq 0$ for a subset of rows $(j_1, \ldots, j_s)$ and a subset of columns $(i_1, \ldots, i_t)$. Given such a block we can change $b_{i_1}, \ldots, b_{i_t}$ freely, and that will only affect $v_{j_1}, \ldots, v_{j_s}$