Flux Balance Analysis (FBA)

- Flux balance analysis is a mathematical method for studying the flow of metabolites in a network. It is a method based on Constraints.
- In the past two decades, 'genome scale metabolic networks' have been constructed for specific organisms.
  These network reconstructions capture all the known metabolic reactions in an organism, and the information on the genes that encode each enzyme catalyzing these reactions.
- FBA calculates the flow of metabolites through this metabolic network. #
  These calculations can
  (i) predict the growth ~~of~~ rate of an organism (and hence)
  (ii) rate of production of ~~a metabolite important~~ metabolite important for biotechnology.
- Now, metabolic models for ~ 35 organisms available.
  (eg) Escherichia coli → 296 ~~1775~~ genes / ~~cap~~ in model
  488 reactions
  343 metabolites

  Saccharomyces cerevisiae → ~~618 gene~~ 708 gene in model
  1175 reactions.

Revision of few basic Concepts

① **gram molecular weight** (mole) is the mass of a substance in gram numerically equal to its molecular weight.

(eg) 1 mole of Nacl is 58.44 grams.

(1 mole of substance has 6.023×10^{23} particles (Avgadro number)

② **Active mass** : is the number of moles/litre. Also known as molar Concentration or molarity.

③ **The law of mass action and equilibrium Constant**

⊙ The law of mass action states that,

"the rate of a reaction at a given temperature is proportional to the product of the active masses of the reacting substances"

Consider the reaction,

$$A + B \rightleftharpoons C + D$$

According to law of mass action,

rate of forward reaction $v_1 = k_1 [A][B]$

rate of reverse reaction $v_2 = k_2 [C][D]$

where, k_1, k_2 are proportionality Constants

.... Contd..

In dynamic equilibrium, $V_1 = V_2$

$\therefore \quad K_1 [A][B] = K_2 [C][D]$

and thus,

$$\frac{K_1}{K_2} = \frac{[C][D]}{[A][B]} = K$$

where, K is the equilibrium Constant of the reaction at Constant temperature.

⊙ In the reaction where more than 1 molecule for each species participates, eg,

$$aA + bB \rightleftharpoons cC + dD,$$

$$K_c = \frac{[C]^c [D]^d}{[A]^a [B]^b}$$

(eg) take the reaction,

$$\underbrace{Na_2CO_3}_{aA} + \underbrace{CaCl_2}_{bB} \rightleftharpoons \underbrace{CaCO_3}_{cC} + \underbrace{2NaCl}_{dD}$$

$$\text{Eq. Const} = K_c = \frac{[CaCO_3]^1 [NaCl]^2}{[Na_2CO_3]^1 [CaCl_2]^1}$$

Eq. Const tells how far reaction has proceeded

* if reaction is almost complete, K is large
  if reaction is early stage, K is small.

page-4

⊙ There are 3 types of equilibrium Constants:

(i) K_c → where [A], [B], .. are in molara Concentrations. Usually for acquous solutions.

~~Conc equilib. Const~~

(ii) K_p → Pressure equilibrium Const, when [A], [B], ... are the partial pressures of species. (used for gases)

(iii) K_a → equilibrium Constant used for activities of species, and can be used for gases, liquids or solids.

~~When all species of reaction~~

When all reactants ~~are~~ and products are in same state (eg acquous solutions), it is a homogeneous reaction

when they are in a mixed state, it is heterogeneous reactions.

4. Stoichiometric Coefficients and Stoichiometric matrix:

Stoichiometric Coefficients are the numbers or ions or molecules written in front of atoms in balancing Chemical reactions. They establish the mole ratio between reactants and products.

$$2Na(s) + 2HCl(aq) \rightarrow 2NaCl(aq) + 2H_2(g)$$

* (Stochiometric Coef. Can be fractions also).

Stoichiometric Matrix

In a reaction network with many metabolites and reactions, the Stochiometric Coefficients of metabolites across reactions can be written as a ~~2×2~~ ~~2D~~ matrix:

Consider the reaction network:

$2A + B \xrightarrow{K_1} C + 3D \rightarrow R_1$ [$V_1 = K_1[A]^2[B]$]

$2A + 3B \xrightarrow{K_2} C + E \rightarrow R_2$ [$V_2 = K_1[A][B]$]

$A \xrightarrow{K_3} 2B + D + E \rightarrow R_3$ [$V_3 = K_3[A]$]

$4B \xrightarrow{K_4} D + A \rightarrow R_4$ [$V_4 = K_4[B]^4$]

$D + 2B \xrightarrow{K_5} C + 2E \rightarrow R_5$ [$K_5 = K_5[D][B]^2$]

$C + 4E \xrightarrow{K_6} 3B + D \rightarrow R_6$ [$V_6 = K_6[C][E]^4$]

We can write the stoichiometric Coeff. across reactions as a matrix:

(Metabolites)	R_1 (v_1)	R_2 (v_2)	R_3 (v_3)	R_4 (v_4)	R_5 (v_5)	R_6 (v_6)
A	−2	−1	−1	1	0	0
B	−1	−3	2	−4	−2	3
C	1	1	0	0	1	−1
D	3	0	1	1	−1	1
E	0	1	1	0	2	4

(Reactions)

(− Sign when metabolite is reactant
+ Sign when it is a product.)

Systems equations

In a network of n metabolites and r reactions, the dynamics of the system is characterized by the systems equation,

$$\frac{dX_i}{dt} = \sum_{j=1}^{r} S_{ij} V_j, \quad \text{for } i = 1, 2, \ldots, n$$

where X_i is the Concentration of i^{th} metabolite
V_j is the rate of j^{th} reaction
S_{ij} is the Stoichastic Coefficient of i^{th} metabolite in j^{th} reaction.

* Intuitively, each systems equation states that the rate of change of Concentration of a metabolite is the sum of metabolite flow to a metabolite and the flow away from it

if $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$, then, In general,

$$\frac{d\bar{X}}{dt} = S\bar{V}$$

where $\bar{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_r \end{bmatrix}$

page 8

Example - 1

$2A \xrightarrow{k_1} B$: Reaction R_1 with rate V_1

$A + B \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} C$: Reaction R_2 with rate V_2

$$\left.\begin{aligned} \therefore V_1 &= k_1 [A]^2 \\ V_2 &= k_2^+ [A][B] - k_2^- [C] \end{aligned}\right\} \Rightarrow \bar{V} = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

$$S = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{cc} R_1 & R_2 \end{array} \\ \begin{bmatrix} -2 & -1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \end{array}$$ is the stochiometric matrix.

$$\therefore \frac{d\bar{X}}{dt} = \begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \bar{S}\bar{V} = \begin{bmatrix} -2 & -1 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}$$

or,
$$\frac{dA}{dt} = -2V_1 - V_2 = -2k_1[A]^2 - k_2^+[A][B] + k_2^-[C]$$

$$\frac{dB}{dt} = V_1 - V_2 = k_1[A]^2 - k_2^+[A][B] + k_2^-[C]$$

$$\frac{dC}{dt} = 0 \cdot V_1 + V_2 = k_2^+[A][B] - k_2^-[C]$$

are the equations.

Example-2

$A + X \xrightarrow{k_1} 2X$: Reaction R_1 with rate V_1

$X + Y \xrightarrow{k_2} 2Y$: Reaction R_2 with rate V_2

$Y \xrightarrow{k_3} B$: Reaction R_3 with rate V_3

Stochiometric matrix

$$\bar{S} = \begin{array}{c} \\ A \\ X \\ Y \\ B \end{array} \begin{array}{c} \begin{array}{ccc} R_1 & R_2 & R_3 \end{array} \\ \begin{bmatrix} -1 & 0 & 0 \\ +1 & +1 & 0 \\ 0 & +1 & -1 \\ 0 & 0 & +1 \end{bmatrix} \end{array}$$

$$V_1 = k_1[A][X]$$
$$V_2 = k_2[X][Y]$$
$$V_3 = k_3[Y]$$

$$\therefore \bar{V} = \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}$$

$$\frac{d\bar{x}}{dt} = \begin{pmatrix} \frac{dA}{dt} \\ \frac{dX}{dt} \\ \frac{dY}{dt} \\ \frac{dB}{dt} \end{pmatrix} = \bar{S}\bar{V} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}$$

$$\therefore \frac{dA}{dt} = -V_1 = -k_1[A][X]$$

$$\frac{dX}{dt} = V_1 + V_2 = k_2[X][Y]$$

next page

.... Contd...

∴ The equations are,

$$\frac{dA}{dt} = -V_1 = -k_1[A][X]$$

$$\frac{dX}{dt} = V_1 + V_2 = k_1[A][X] + k_2[X][Y]$$

$$\frac{dY}{dt} = V_2 - V_3 = k_2[X][Y] - k_3[Y]$$

$$\frac{dB}{dt} = V_3 = k_3[Y]$$

---

* Steady state Analysis

The Under steady states, Concentrations don't change.

$$\therefore \boxed{\frac{dX_i}{dt} = \sum_{j=1}^{r} S_{ij} V_j = 0} \quad \text{for } i = 1, 2, \cdots, n$$

These are a set of linear equations Constraining to n reaction rates V_j

Above eqns for steady state can be written in matrix form as,

$$\boxed{\frac{d\bar{X}}{dt} = \bar{S}\,\bar{V} = \bar{0}}$$ under steady state

The reaction vector $\bar{V}$ satisfying the * above Condition is Called a "flux Vector"

page 11

Exercise :

Construct the Stoichiometric matrix and differential eqns for the following set of reactions from pentose-phosphate pathway :

R_1 : βG6P (Beta) + $NADP^+ \xrightarrow{k_1}$ 6PGL + NADPH

R_2 : 6PGL + $H_2O \xrightarrow{k_2}$ 6PG

R_3 : 6PG + $NADP^+ \xrightarrow{k_3}$ R5P + NADPH

R_4 : R5P $\xrightarrow{k_4}$ X5P

R_5 : αG6P $\underset{k_5^-}{\overset{k_5^+}{\rightleftharpoons}}$ βG6P

R_6 : αG6P $\underset{k_6^-}{\overset{k_6^+}{\rightleftharpoons}}$ βF6P

R_7 : βGCP $\underset{k_7^-}{\overset{k_7^+}{\rightleftharpoons}}$ βF6P

The methodology of Flux Balance Analysis

① The first step is the mathematical representation of the metabolism.

Metabolic reaction are represented as a Stoichiometric matrix of $\bar{S}$ of size $m \times r$.

The m rows are the Compounds and r Colums are the reactions.

The entries in each of $m \times r$ cells are the Stochiometric Coefficients of metabolites participating in the reaction

When a metabolite is Consumed, its Coefficient is negative.

when a metabolite is produced, its Coefficient is positive.

The stochiometric Coefficient is zero when the metabolite is not participating in the reaction.

$\bar{S}$ is generally a sparse matrix, Since each reaction involves only few metabolites.

The flux through the network is represented by a vector $\bar{V}$ (of length n)

The Concentration of metabolites is represented by a vector $\bar{X}$ (of length m)

⊙ At steady state, the system of mass balance equation is given by,

$$\frac{d\bar{x}}{dt} = \bar{S}\bar{v} = 0$$

where $\bar{x}_{m\times 1}$, $\bar{S}_{m\times r}$, $\bar{v}_{r\times 1}$

Any $\bar{v}$ that satisfies the above equation is said to be in the null space of $\bar{S}$

⊙ In any realistic large scale metabolic model, there are more reactions than the compounds (ie, $r > m$). In other words, there are more unknown variables than equations. So, there is no unique solution to this system.

Then how do we get specific solutions for a problem?

⊙ When many solutions exist, we generally impose constraints on the variables to reduce the solution space.

After reducing the ~~constraints~~, range of solutions with constraints, we define an 'optimal point' that solution should ~~satisfy~~. Correspond to.

(eg) ⊙ Maximum growth rate

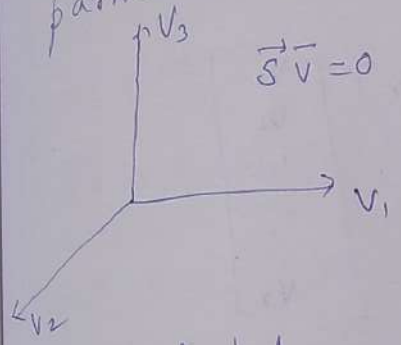
or

Maximum ATP production of organism

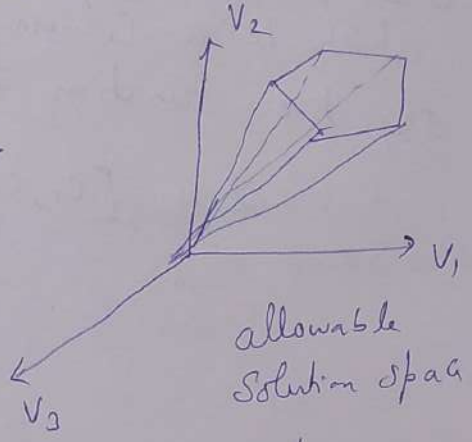
~~The optimal~~

page 14 ~~The optimal point in~~

Given a set of particular Constraints the optimal point helps us to arrive at a particular solution.



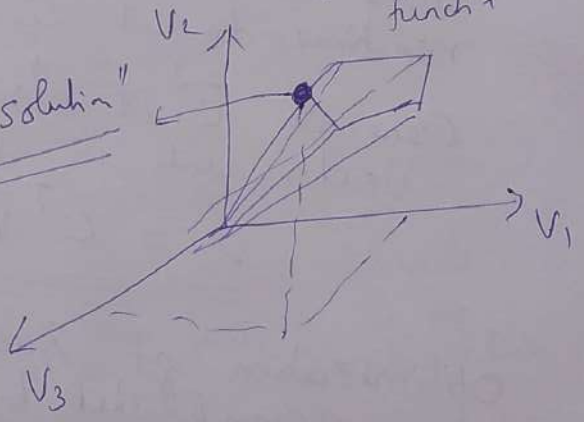
$\vec{S}\,\vec{V} = 0$ Constrain →



unrestricted Solution space

allowable Solution space

optimization (maximize something) → called objective function



"optimal solution"

The quantity that is maximized to get a particular solution is called ~~"Objective function"~~ the "Objective function"

The objective function Z is can be any linear combination of fluxes. once we are given a rate vector $\bar{V}$, which is a column vector of rates of reaction, objective function Z is,

$$Z = [C_1, C_2, C_3, \cdots C_r] \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \vdots \\ V_r \end{bmatrix}$$

where $\bar{C}$ is a vector of weights, indicating how much each reaction contributes to the objective function. $\bar{C}$ can be a vector with zeroes for some weights when the some reactions do not contribute to objective function.

Generally, $\bar{C}$ is also written as column vector and its Transpose is taken.

$$\therefore \quad Z = \bar{C}^T \bar{V}$$

Optimization of such a system is accomplished by linear programming

⊙ Therefore, FBA can thus be defined as the use of linear programming to solve the equation $\bar{S}\,\bar{V} = 0$, given a set of upper and lower bounds on reaction rates (or fluxes) $\bar{V}$, and a linear combination of fluxes as objective function.

The output of FBA is a particular flux (reaction rate) distribution, $\bar{V}$, which maximizes or minimizes the objective function.

(In eg, in a fermentation process, we get optimal reaction rates $\bar{V}$ that maximize some production!)

⊙ Constraints are imposed in 2 ways.

(i) as equations that balance reaction inputs and outputs.

ie, Some in/
(eg $[A] + [E] = K_0$ (a constant))

(ii) as inequalities that impose bounds on the system. These bounds define the allowable space of allowable flux distributions of the system, ie, the rate at which each metabolite is produced or consumed by each reaction.

Other constraints can be added.

Page 17

⊙ We should now understand that there are 2 approaches to solving the set of systems equations

Approach 1 : Solving set of differential equations with kinetic parameters to get the metabolic concentrations with time as solutions and to simulate conditions further.
[Kinetic models]

Approach 2 : The Flux Balance Analysis (FBA) to predict metabolic reaction fluxes (rates) that can simulate growth on different substrates or with genetic manipulation.

⊙ FBA does not require kinetic parameters and can be computed very quickly for even large networks.

(eg) Explore the effect on growth of deleting every pairwise combination of 136 of E. Coli genes to find double gene knockouts that are essential for survival of bacteria

Limitations of FBA

⊙ Because FBA does not use kinetic parameters, it cannot predict metabolite concentrations. No dynamic evolution of metabolic conc. under different conditions cannot be studied

⊙ It is only suitable for determining flux at steady state

⊙ FBA does not account for regulatory effects such as activation of enzymes by protein kinases or regulation of gene expression. Because of this, its prediction may not be accurate.

FBA is very useful in studies like,
- studying growth on different media
- genome scale synthetic biology
- simulation of multiple gene knockout
- to predict yields of important cofactors like ATP, NADH, NADPH etc

* Note : In FBA, more than one solution can yield desired optimization. Other methods of analysis are used to choose between them.