

Predicting rice (*Oryza sativa*) metabolism

Sudip Kundu

Department of Biophysics, Molecular

Biology & Bioinformatics,

University of Calcutta, WB, India .

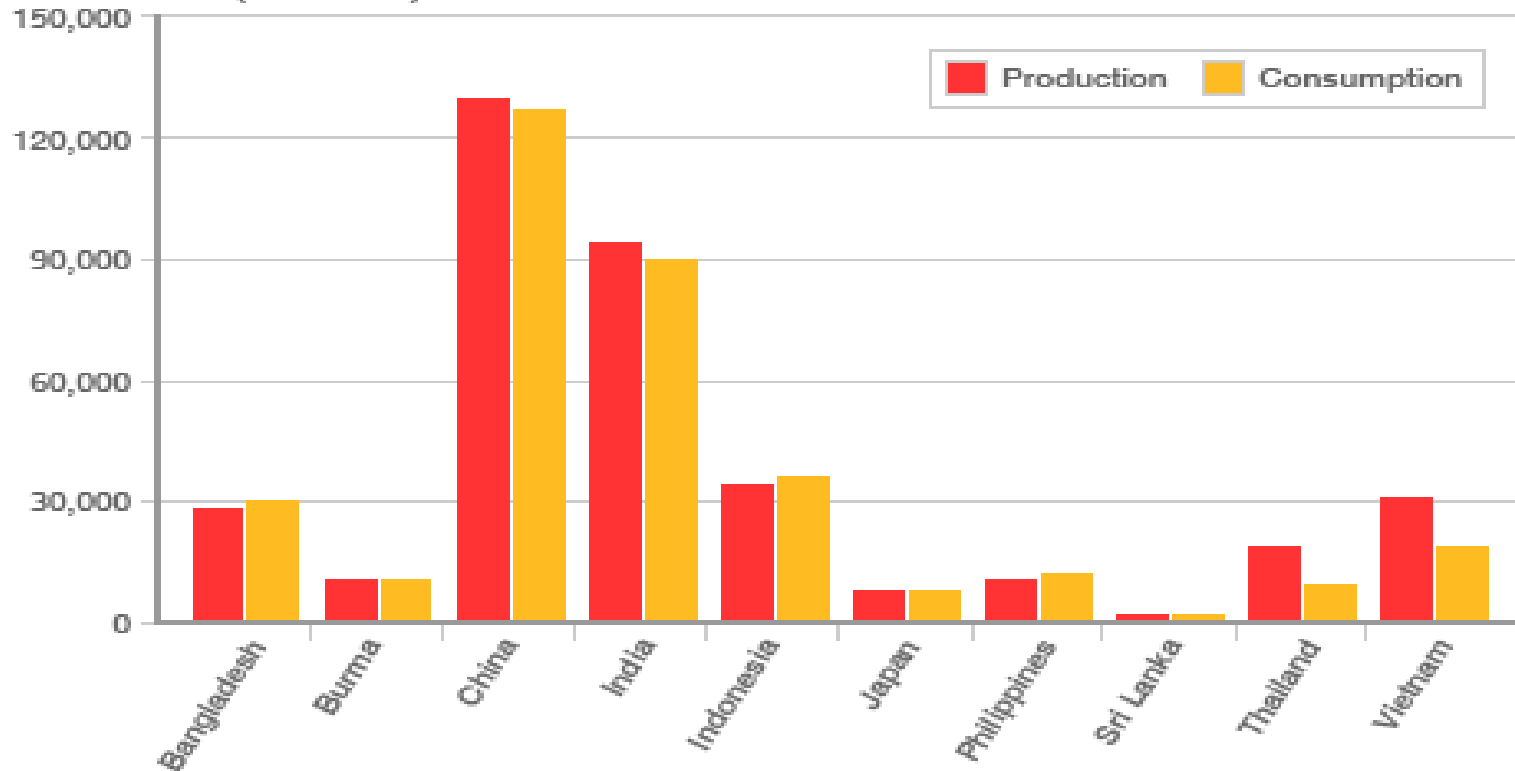
skbmbg@caluniv.ac.in

Collaborators:

Mark G Poolman and David A. Fell
Cell Systems Modelling Group, Oxford
Brookes University, Oxford, UK

RICE PRODUCTION AND CONSUMPTION

Metric tonnes (thousands)



Note: figures for milled rice for year ending March 2008

SOURCE: USDA

Rice production and consumption in Asian Countries

Rice makes up 20% of the total caloric intake for the human population as a whole and thus becomes one of the most important cereal crops in world economy, history and culture



Rice plant

It is also a model organism for the grass family of crops and other plants.

- **Understanding the cellular mechanism of rice would help one to develop more efficient rice cultivars as well as to understand the biology of crop grasses as well.**
- **Metabolic pathways are central to understanding cellular behaviour and to improving a particular properties of an organism.**

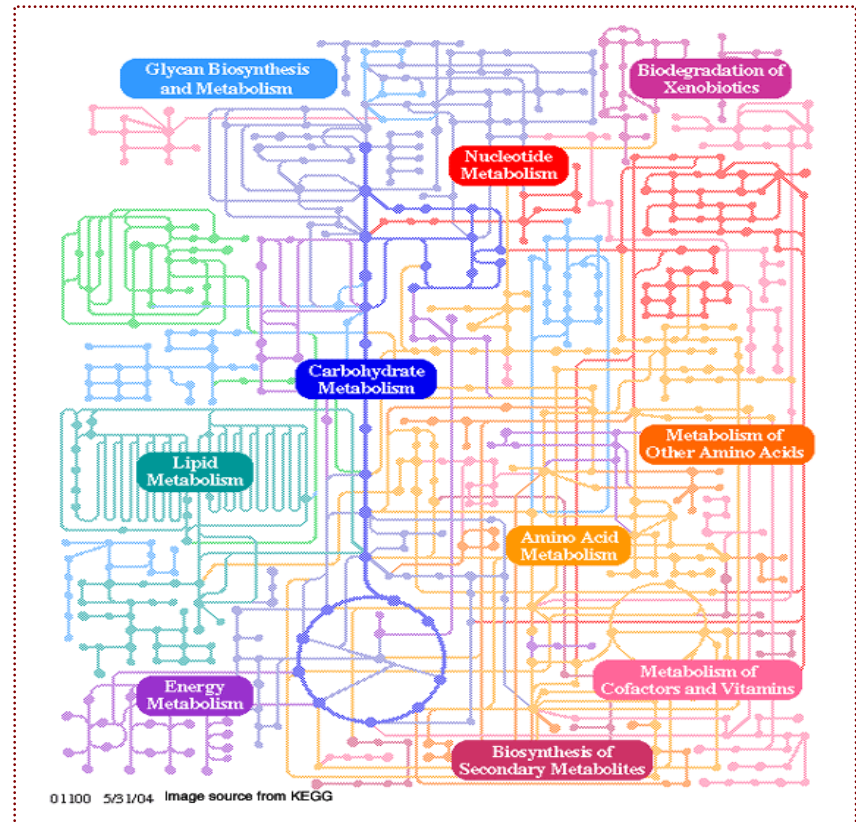
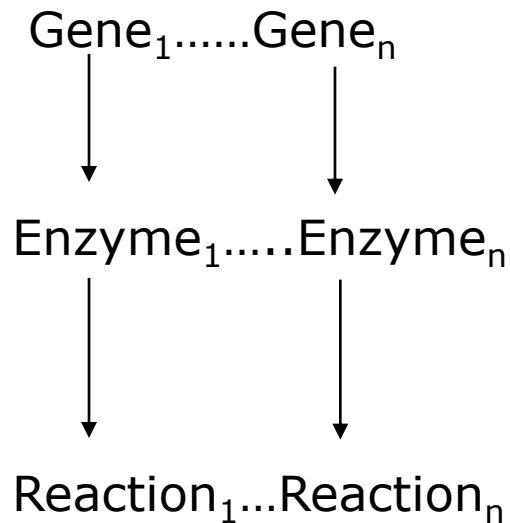


Rice plant

Motivation

- We expect the genome encodes the metabolic phenotype. Can we exploit that understanding to make correct predictions about metabolic responses?
- Application area metabolic engineering
- In order to cope with increasing world wide demand for food, as a first step with the aim of assisting the development of more efficient rice cultivars, we are constructing a genome scale metabolic model of *Oryza sativa*

Building Genome Scale Metabolic Models



Modelling approaches

Dynamics

(kinetic)

Kinetic rate equation for
each reaction + parameters



Simulation of system
behaviour

Structural

(stoichiometric)

Consider all possible
behaviours of the system
(large solution space)



Imposing constraints
(physicochemical laws,
biological constraints)



Smaller allowable solution
space

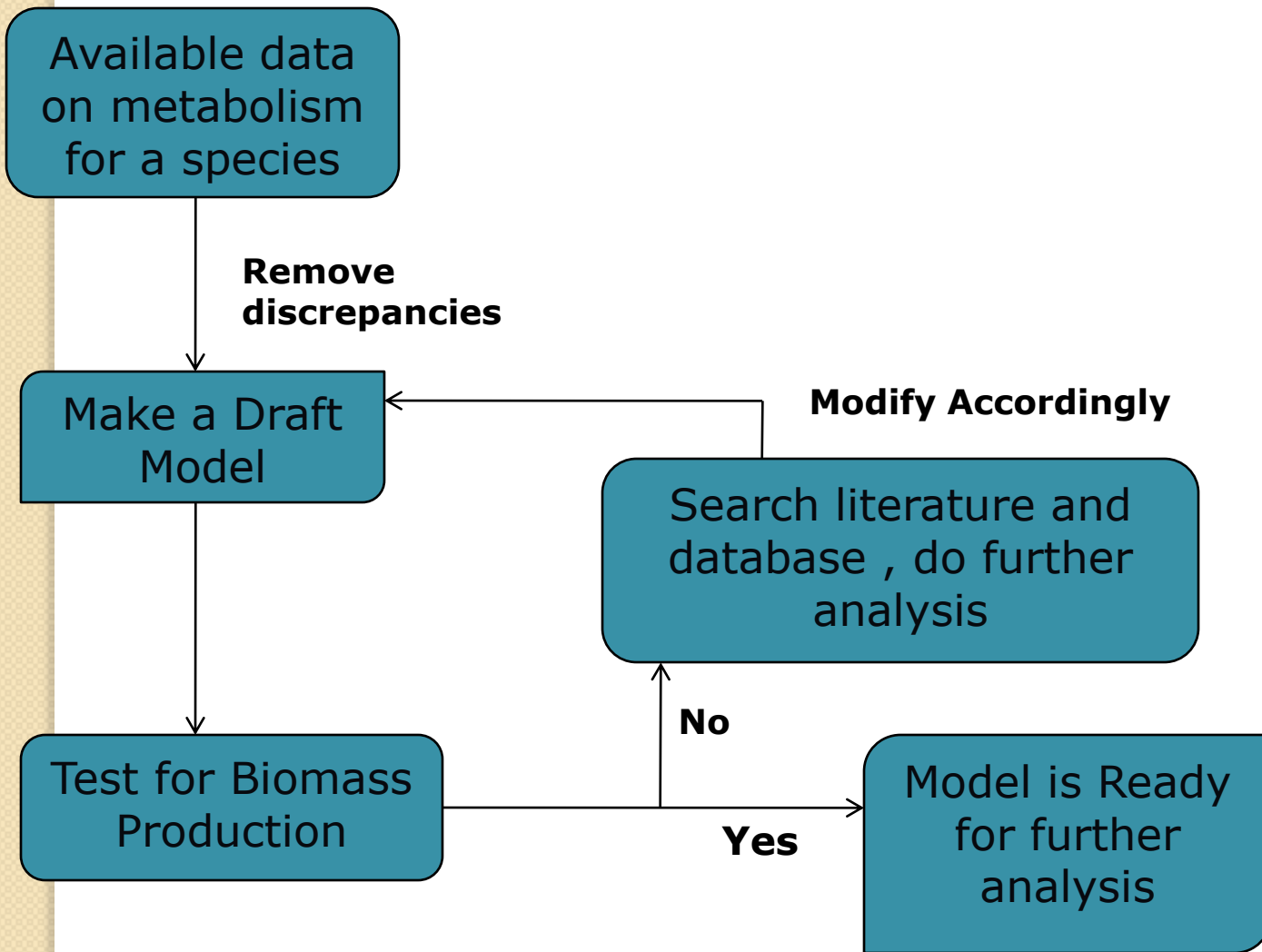
For Genome Scale Model - **STUDY Structure AND NOT Dynamics**

- Knowledge is more complete for network structure than for enzyme kinetics.
- Structural analysis involves simple linear equations; dynamic analysis involves non-linear enzyme kinetic functions.
- The network structure places limitations that constrain the network dynamics, irrespective of the kinetics, e.g.:
 - ◆ Whether viable routes exist from nutrients to stated metabolic products;
 - ◆ Whether some routes remain after deletion of the steps catalyzed by a particular enzyme;
 - ◆ What the maximum obtainable conversion yield is for formation of any metabolite from a given set of sources, and

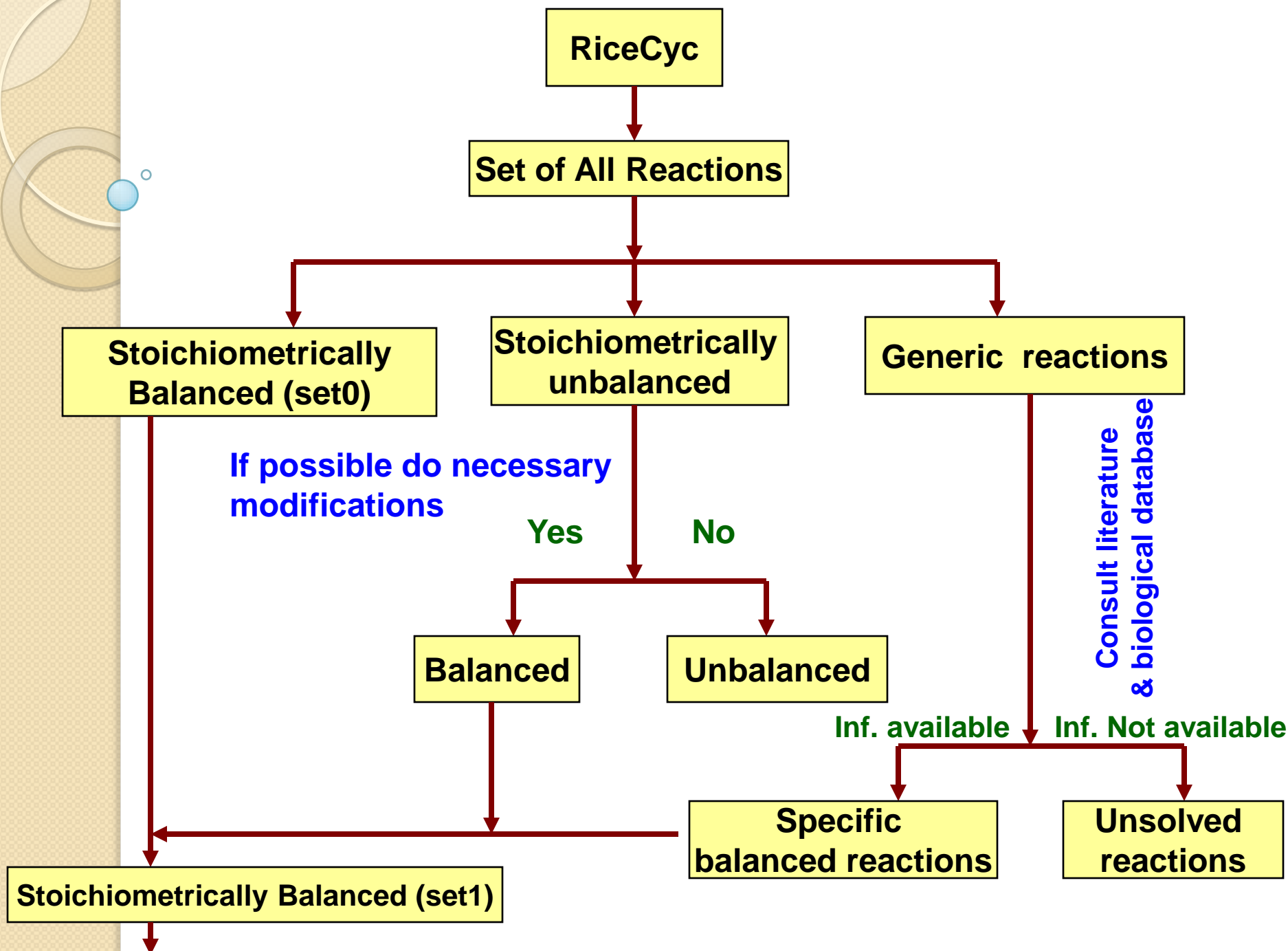
The Reaction list :

Where does the data for the structure of the metabolic mode come from?

- Biochemical literature : books , reviews , journal articles.
- Genome databases + annotation + enzyme database
- BioCyc : <http://biocyc.org/>
- KEGG : <http://www.genome.jp/kegg/>
- IntEnz at <http://www.ebi.ac.uk/intenz/>
- EXPASY Enzyme : [http:// www.expasy.org/enzyme/](http://www.expasy.org/enzyme/)
- Brenda :[http:// www.brenda.uni-koeln.de/](http://www.brenda.uni-koeln.de/)
- Others.....



Modeling of Metabolism



Stoichiometrically balanced (set1)

Remove other inconsistencies (like duplicate entries of reactions; presence of both – a set of individuals and combined reactions, etc.)

Stoichiometrically balanced (set2)

Remove dead reactions

Dead reactions are those reactions which do not carry any net flux through them at steady state.

Stoichiometrically balanced (set3)

Add transporters and take the giant cluster

The clusters in the metabolic network that contain most of the connected nodes

First (n) draft of working model

Test whether transporters are alive, if not do necessary modification

Next draft (n+1) of working model

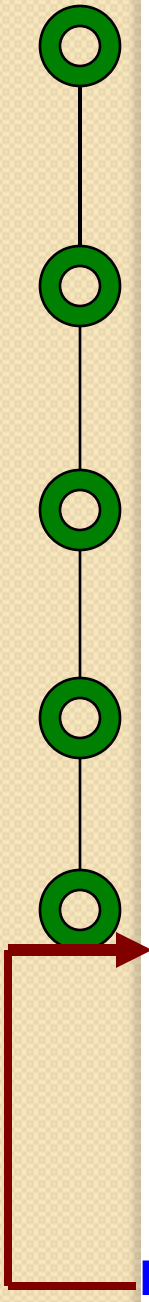
Check biomass productions.

If inconsistent

consistent

Do necessary modification

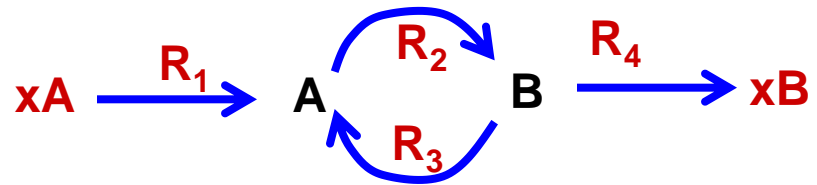
Latest working model



WHETHER THE MODEL IS CAPABLE TO PRODUCE ALL THE IMPORTANT BIOMASS FROM ITS NUTRIENTS?

- Source (Carbon-dioxide, Ammonia, Nitrate, Phosphate, Sulphate, Oxygen, water)
- Biomass - Starch, Sucrose, Cellulose, all amino acids and nucleotides, linoleate, chlorophyll

Structural Model: A set of reactions with defined stoichiometries.



	R1	R2	R3	R4
A	1	-1	1	0
B	0	1	-1	-1

Rows represent internal metabolites

Columns represent reactions

Elements represent stoichiometric coefficients

Mass conservation implies that the temporal change in the concentration of each metabolite equals the sum of all reaction rates producing that metabolite minus all reaction rates consuming it.

Using the stoichiometric coefficients, this can be written as

$$\frac{ds_i}{dt} = \sum_{j=1}^r n_{ij} v_j$$

when n_{ij} are the entries in the i^{th} row and; j^{th} column of the matrix.

$S_i \Rightarrow$ Concentration of the i^{th} metabolite

$v_j \Rightarrow$ Rate of the j^{th} reactions.

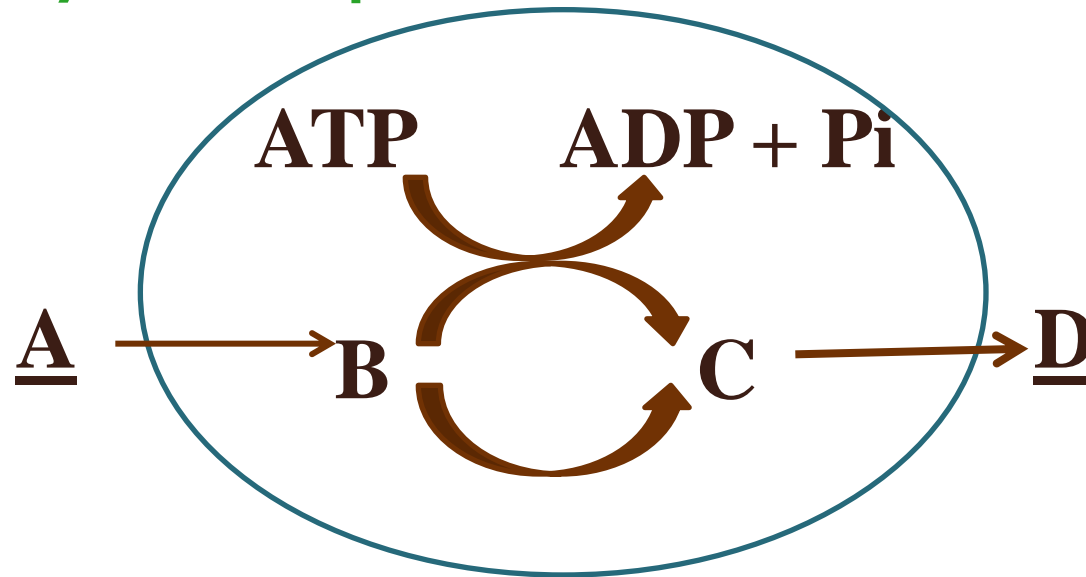
By setting S_i and v_j into vectors S and V

$$\frac{ds}{dt} = NV$$

when a system is at steady state, the reaction rate obey the following equation:

$$NV = 0$$

Flux Analysis and Optimisation



- **Knowing enough about the system is often difficult; assume for example that in the above “cell” we can only measure consumption of A and production of D.**
- **Is there a way to evaluate the internal fluxes? Are there any relationships among the fluxes?**
- **An assumption is needed; often this assumption is that the organism is operating OPTIMALLY**
- **e.g. assume the above organism optimal with respect to energetic efficiency**

Types of constraints

- **Physico-chemical constraints**

 - mass, charge and energy conservation, laws of thermodynamics

- **Biological constraints:**

 - external environment, regulatory constraints

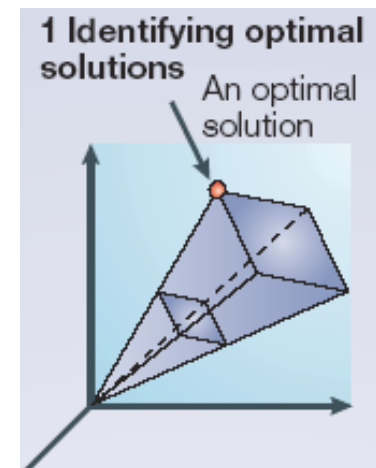
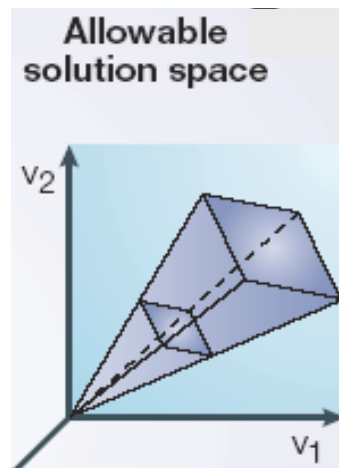
Problem:

Addition of constraints reduces the allowable solution space, but usually not to a single point (underdetermined system).

How to find a particular solution?



We can look for a solution which optimises a particular network function (e.g. production of ATP or biomass) → **FBA**



Linear Programming (LP)

- **LP model → objective function + linear constraints**
- **extensively used optimisation technique**
- **allocation of limited resources to competing activities in the optimal way**
- **examples of application: graphs, network flows, plant management, economics, business management**

minimise $c_1x_1 + c_2x_2 + \dots + c_nx_n$

subject to:

equality constraints

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = a_{i0}$$

inequality constraints

$$b_{i1}x_1 + b_{i2}x_2 + \dots + b_{in}x_n \leq b_{i0}$$

or in matrix form:

$$\min c^T x$$

subject to:

$$A \cdot x = a$$

$$B \cdot X \leq b$$



Flux Balance Analysis

- Relies on balancing metabolic fluxes
- Based on the fundamental law of mass conservation
- Performed under the steady state conditions
- Requires information only about:
 1. The stoichiometric of metabolic pathways
 2. Metabolic demands
 3. And a few strain specific parameters
- It does not require enzyme-kinetic data

USE LP FOR THIS...

Modifications in the model..

- **All reactions were initially assumed to be irreversible.**
- **A subset of these were subsequently made reversible based on other evidences**
- **All isomerases are made reversible.**
- **During the analysis, further subset of reactions seem to be reversible has been curated manually (if needed and if thermodynamically possible)**
- **Reversible reactions were split into irreversible forward and reverse components**

A VERY BRIEF OF LP

- The constraints can be formulated as:

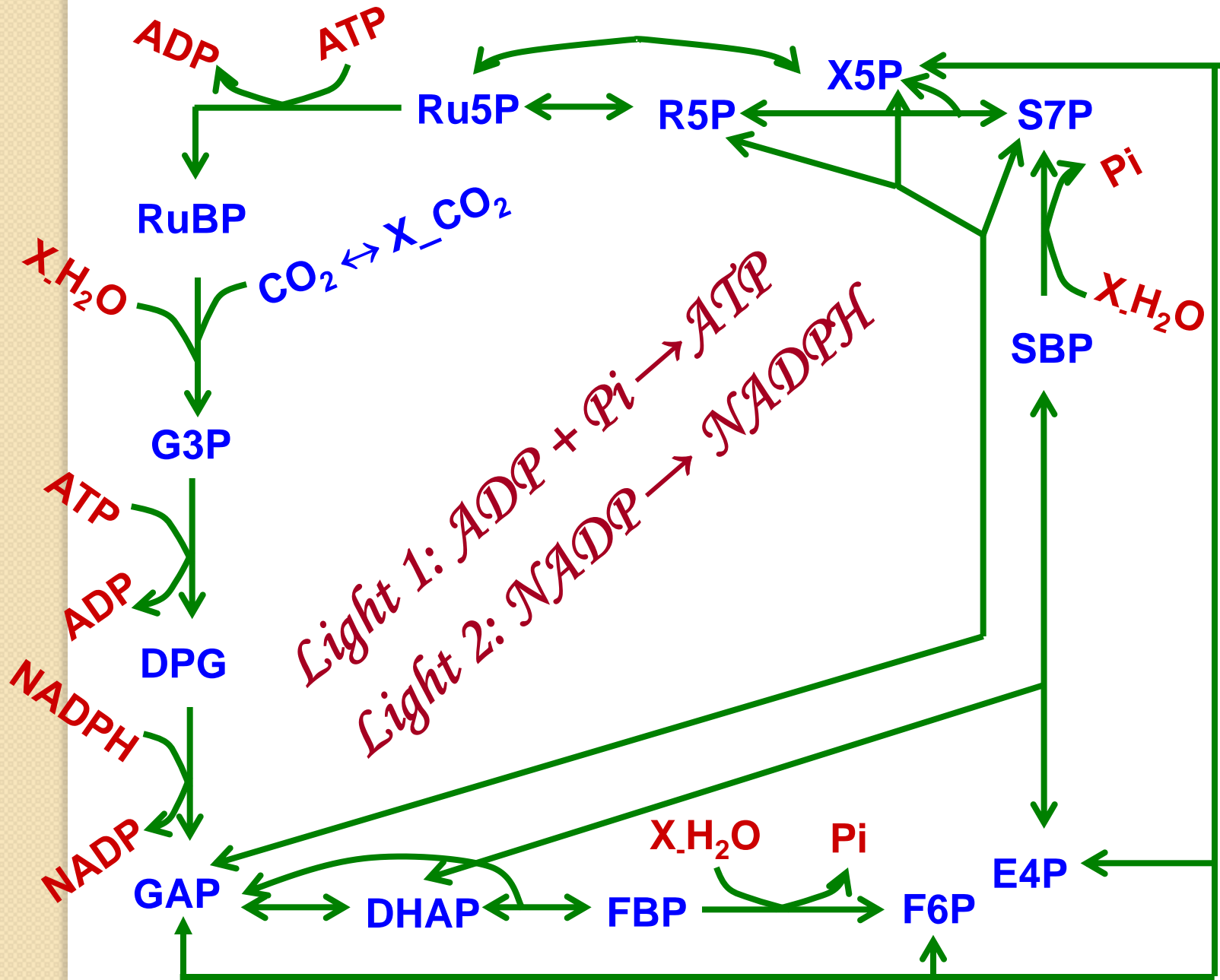
$$N_{\text{full}} \bar{v} = \bar{b}$$

where N_{full} is a full stoichiometry matrix, which includes both internal as well as external metabolites of the system.

Vector b represents the rate of the amount of uptake or excretion of external metabolites, and

the rate of change of internal metabolites is zero.

- **The objective function can be given as:**
- $\alpha_i \leq v_i \leq \beta_i$
- $\mathbf{c}^T \mathbf{r} = c_1 r_1 + c_2 r_2 + \dots + c_p r_q = \max$
- **Where α_i and β_i denote the lower and upper boundaries for the reaction rates and**
- **vector \mathbf{C} represents the optimization criteria such as weights of the reaction rates.**



Core Pathways for Starch, Cellulose and Sucrose from CO₂

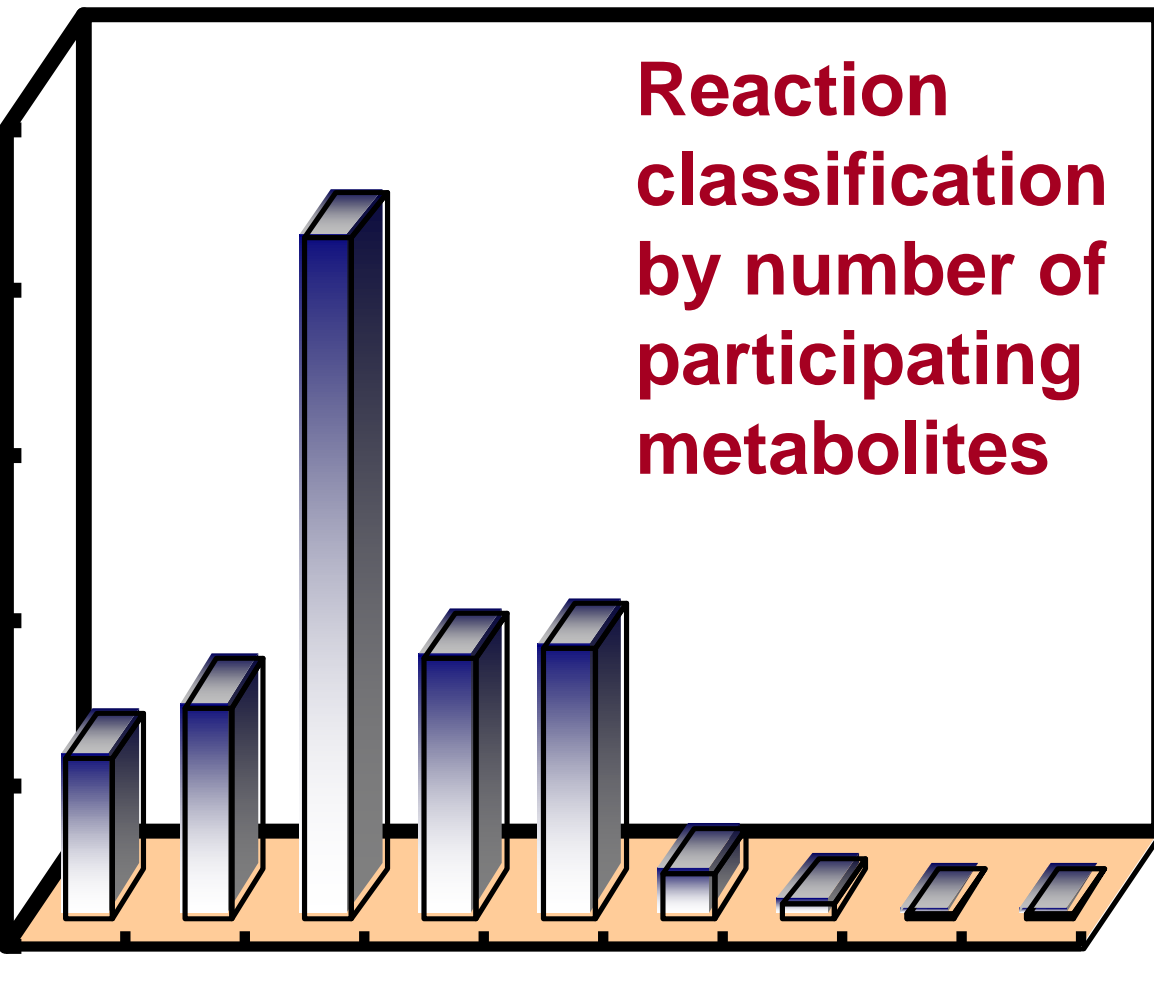
No. of Reactions

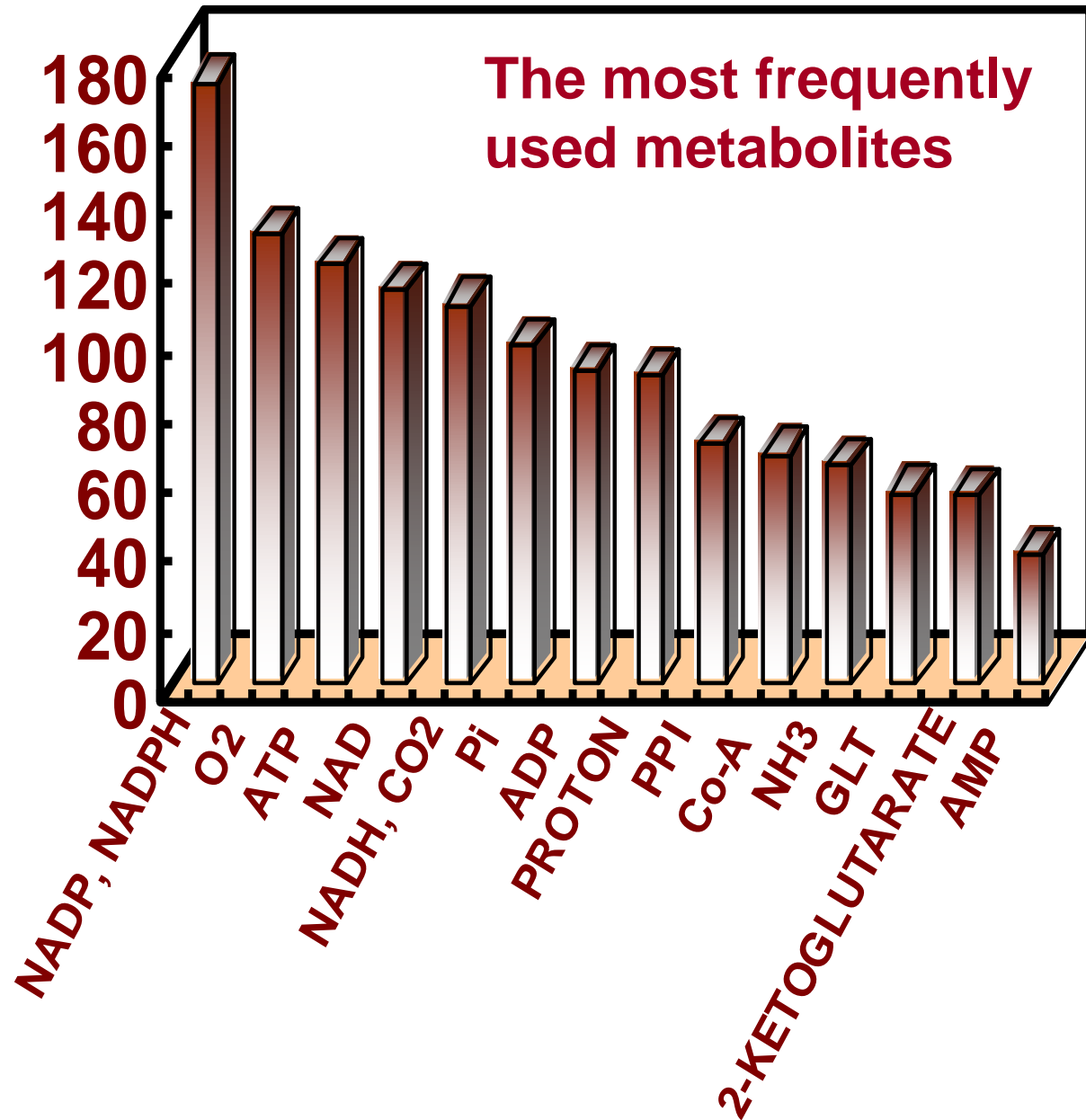
500
400
300
200
100
0

**Reaction
classification
by number of
participating
metabolites**

2 4 6 8 10

No. of Metabolites





Enzyme subset analysis

An enzyme subset is defined as a group of enzymes that carry flux in a fixed proportion in any steady state where they are active.

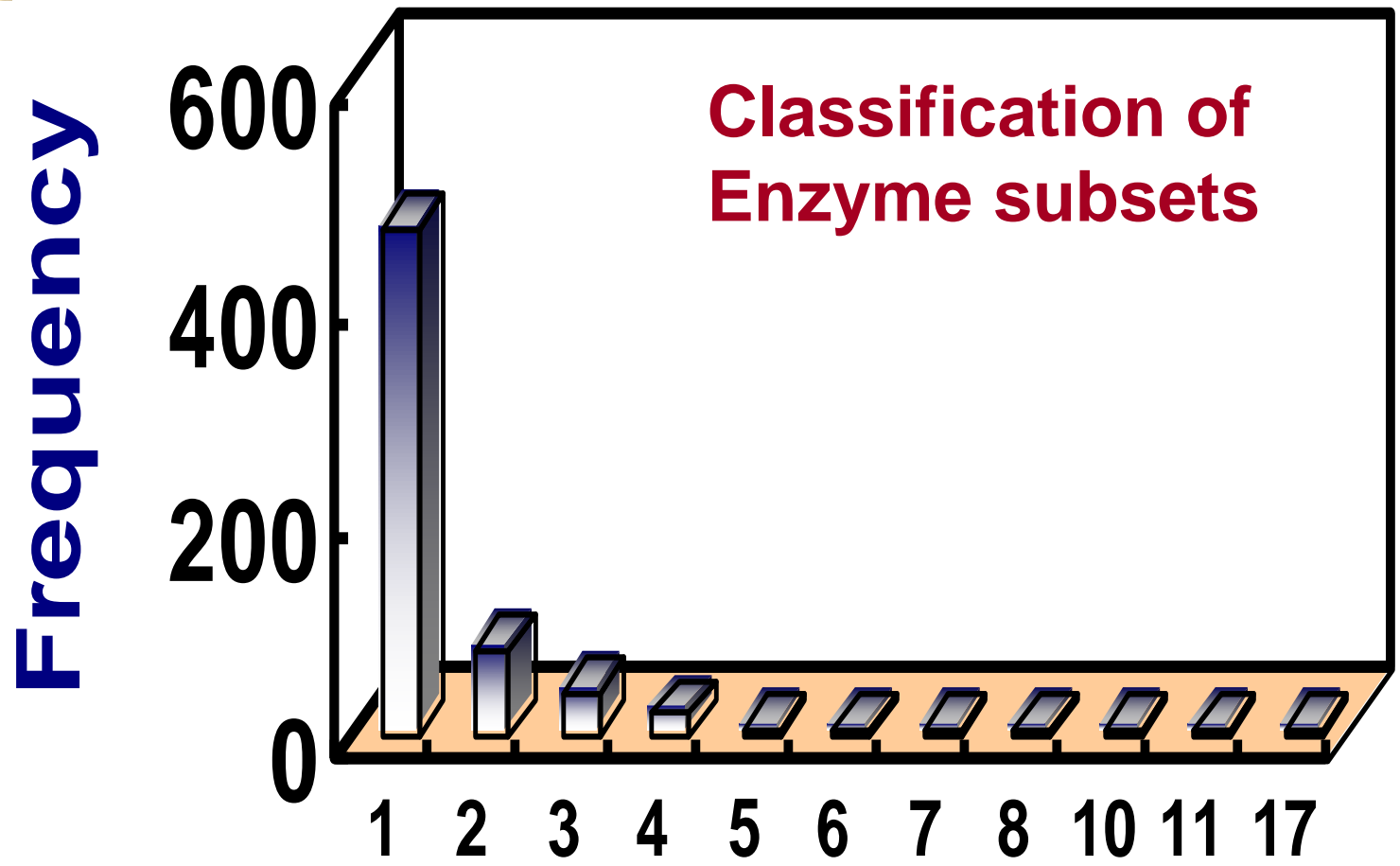
For a simple linear system, all enzymes are in a single enzyme subset.

Strictly ,they are reaction subsets, because they are identified from the stoichiometry matrix.

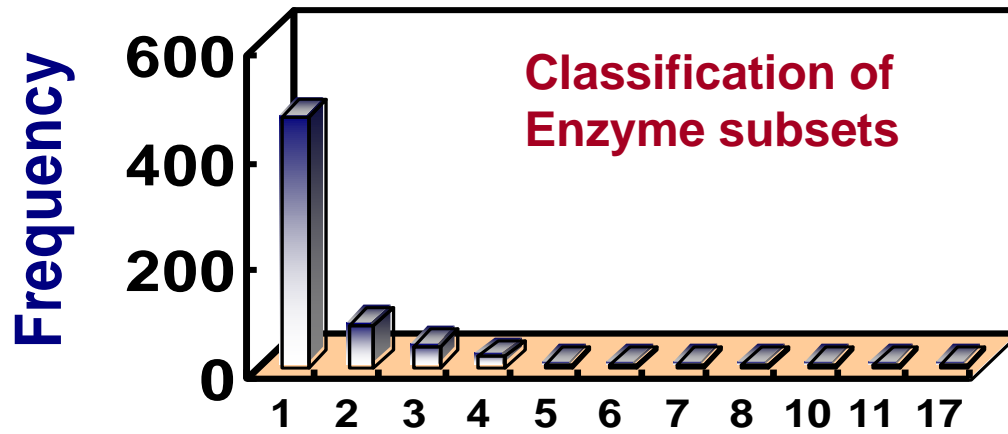
CALCULATING ENZYME SUBSETS....

A general outline of the algorithm for enzyme subset identification is as follows:

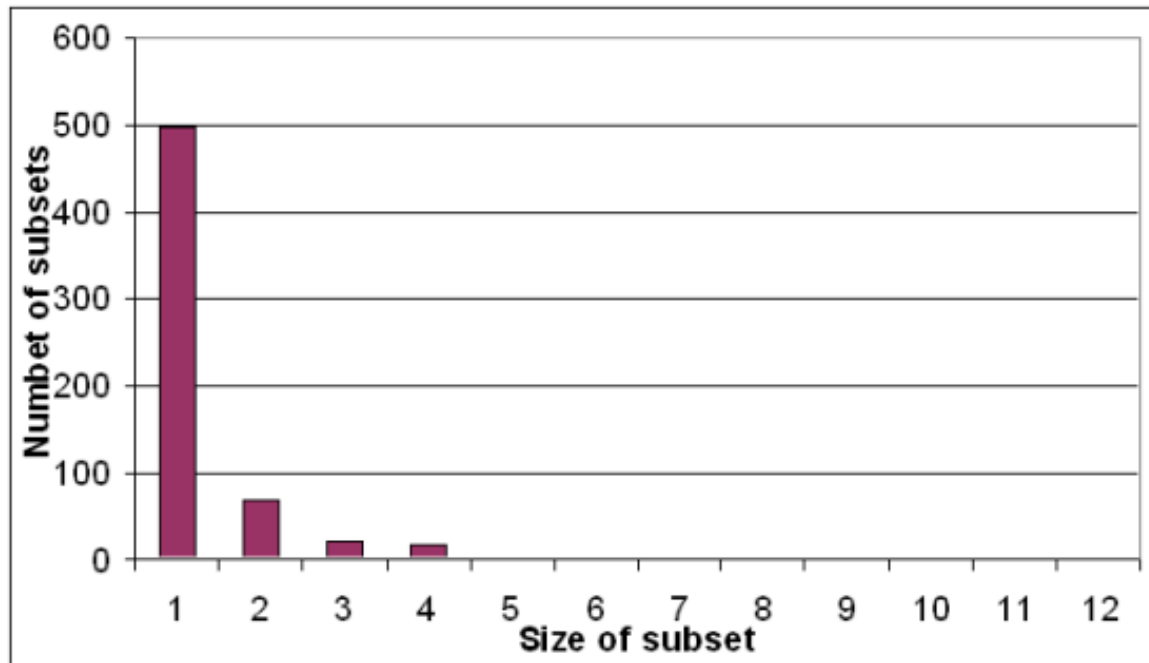
- **Detect all row vectors of K (null space of stoichiometry matrix) that are null vectors (i.e. Dead reactions)**
- **Normalise each of the remaining row vectors of K by dividing by its greatest common divisor.**
- **Compare any normalised row vector with any other. If they are the same and there are no contradictions in the directionalities of irreversible reactions, the corresponding reactions belong to the same subset.**



Reactions in a Subset



Reactions in a Subset



FREQUENCY OF ENZYME SUBSETS IN E.Coli

ACKNOWLEDGEMENT...

- **To all my teachers from childhood to today...**
- **BBSRC UK-India Partnering award (2009-2013)**

Thank you

Some inconsistencies---

- **"3.2.1.32-RXN":
NOTHING -> NOTHING**
- **"RXN-1802":
"1-KESTOTRIOSE" + "SUCROSE" -> "6G-KESTOTETRAOSE" +
"SUCROSE"**
- **"CHOLESTENOL-DELTA-ISOMERASE-RXN":
"CPD-4186" -> "CPD-4186"**

Some more inconsistencies --

- **presence of individual and summary reaction**

"2.5.1.32-RXN":

2 "GERANYLGERANYL-PP" -> "CPD-464" + "PPI"

"RXN-8002":

"CPD-464" -> "PHYTOENE" + "PPI"

"RXN1F-144":

2 "GERANYLGERANYL-PP" -> "PHYTOENE" + 2 "PPI"

← BACK