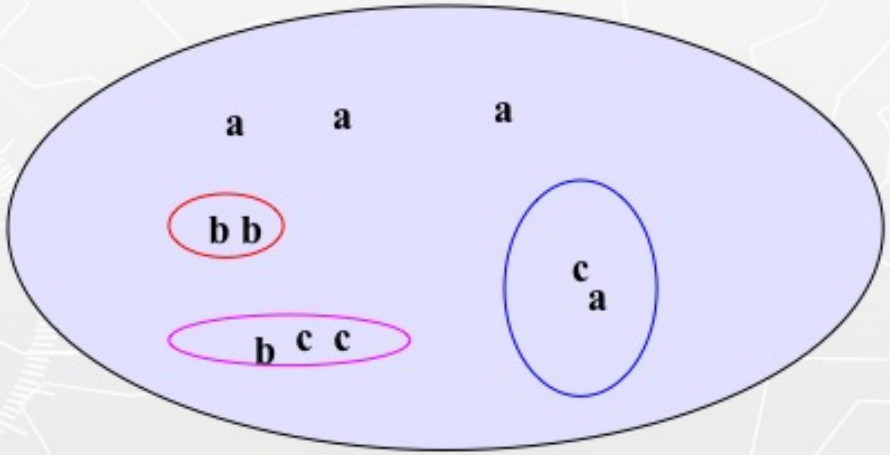
# Metabolic Networks analysis

Dr Giuditta Franco

Department of Computer Science, University of Verona, Italy

# Metabolic System

The **state** of a metabolic system is given by the **types**, the **localization**, and **quantity** of its metabolites.




**P systems the natural theoretical framework**

# Metabolic Systems

A *metabolic system* is a reactor where *n* different types of biomolecules are subjected to transformations (mainly according to stoichiometric rules, but also introduction and/or expulsion of matter are possible) which consume *reactants* and generate *products* in a quantity depending on the system state.

Such a structure may be naturally described by a dynamical system, whose states are *n*-dimensional numerical vectors (representing the quantities of the biomolecules) and whose dynamics keeps track of the evolution of the reactor content.

# What kind of dynamics?

Let P system states be infinite, $X_i \in \mathbb{N}_n$ (vectors or matrices).
Let us focus on the transitions

$$X_i \rightarrow X_j$$

that we have if "the rules of the P system make it to pass from the configuration $X_i$ to $X_j$".

Any function to represent the application of rewriting rules?
How to represent movements of objects between membranes?

It depends on the strategy. Two "deterministic" possibilities:

- Stochastic matrices (a probability of application is associated to each rule)
- Metabolic algorithm (a polynomial function) which in each step performs a distribution of objects among all the rules.

# What kind of dynamics?

Let P system states be infinite, $X_i \in \mathbb{N}_n$ (vectors or matrices).
Let us focus on the transitions

$$X_i \rightarrow X_j$$

that we have if "the rules of the P system make it to pass from the configuration $X_i$ to $X_j$".
Any function to represent the application of rewriting rules?
How to represent movements of objects between membranes?

It depends on the strategy. Two "deterministic" possibilities:

- Stochastic matrices (a probability of application is associated to each rule)
- Metabolic algorithm (a polynomial function) which in each step performs a distribution of objects among all the rules.

# What kind of dynamics?

Let P system states be infinite, $X_i \in \mathbb{N}_n$ (vectors or matrices).
Let us focus on the transitions

$$X_i \rightarrow X_j$$

that we have if "the rules of the P system make it to pass from the configuration $X_i$ to $X_j$".
Any function to represent the application of rewriting rules?
How to represent movements of objects between membranes?

It depends on the strategy. Two "deterministic" possibilities:

- Stochastic matrices (a probability of application is associated to each rule)
- Metabolic algorithm (a polynomial function) which in each step performs a distribution of objects among all the rules.

# What kind of dynamics?

Let P system states be infinite, $X_i \in \mathbb{N}_n$ (vectors or matrices).
Let us focus on the transitions

$$X_i \rightarrow X_j$$

that we have if "the rules of the P system make it to pass from the configuration $X_i$ to $X_j$".
Any function to represent the application of rewriting rules?
How to represent movements of objects between membranes?

It depends on the strategy. Two "deterministic" possibilities:

- Stochastic matrices (a probability of application is associated to each rule)
- Metabolic algorithm (a polynomial function) which in each step performs a distribution of objects among all the rules.

# What kind of dynamics?

Let P system states be infinite, $X_i \in \mathbb{N}_n$ (vectors or matrices).
Let us focus on the transitions

$$X_i \rightarrow X_j$$

that we have if "the rules of the P system make it to pass from the configuration $X_i$ to $X_j$".
Any function to represent the application of rewriting rules?
How to represent movements of objects between membranes?

It depends on the strategy. Two "deterministic" possibilities:

- Stochastic matrices (a probability of application is associated to each rule)
- Metabolic algorithm (a polynomial function) which in each step performs a distribution of objects among all the rules.

# Modeling and computation

Computational model: dynamical system, discrete in space and time. A computation as a sequence of states is a dynamics.

Metabolic P systems [Manca et al. '05] with algebraic and algorithmic procedures compute the dynamics of a reactor by a recurrence system.

New methodology. From time series[1] to an MP grammar which

1. includes the biological knowledge of the phenomenon,
2. has a 'own' dynamics fitting the time series.

[1] quantities sperimentally measured at macroscopic temporal scales

# Traditional modeling

## ODE for Metabolism

Autonomous Differential Equations
for computing the dynamics
(a state is an n-dimension vector)

$$dx_1/dt = f_1(x_1, x_2, \ldots, x_n)$$
$$dx_2/dt = f_2(x_1, x_2, \ldots, x_n)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$dx_n/dt = f_n(x_1, x_2, \ldots, x_n)$$

Eberhard O. Voit Computational Analysis of Biochemical Systems, C.U.P, 2000 (+ PLAS software)

6

# Change of perspective

## Mass Partition Perspective (*vs* Time Partition)

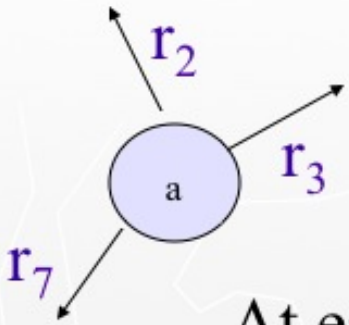
Time is an **observation/macroscopic time** *vs* the (infinitesimal) microscopic time of chemo-physical interactions :

steps 0, 1, 2, 3, …

Reaction reading is not individual, but rather **population/mole** driven: aab --> cd does not mean 2a + 1b molecules produces 1c + 1d, rather 2(population unit)a + …..

9

# Mass Partition Dynamics




$X = \{a, b, c, \ldots\}$ **R**

Observe a system at some steps

At each step "**determine**" the amount of reactants taken by each reaction

For every substance x, **remove** the amount of x which is consumed (by the reactions consuming x) and **add** the amount of x which is produced (by the reactions producing x).

10

Courtesy of prof. Manca, Univ. VR, IT

# Difference *w.r.t.* Gillespie

Gillespie essentially chooses the most probable reaction and applies it to a single reaction complex (2a+1b --> 1c+1d), therefore, in a macroscopic step, billions of individual reactions are applied (following, non-deterministically, a possible behavior)

Mass Partition moves non-determinism from reactions to population. At population level, the system evolves deterministically, but individual behavior remains unknown.

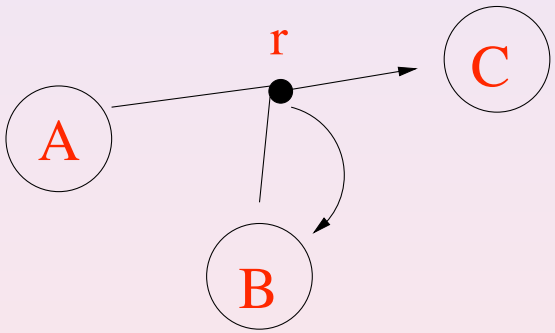
11

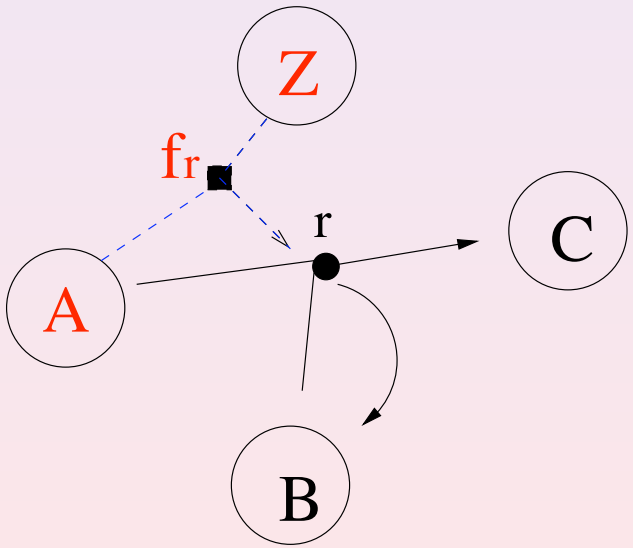
Courtesy of prof. Manca, Univ. VR, IT

# Metabolic P systems: an intuition

Multiset rewriting system: a set of reactions r (transforming metabolites), corresponding regulators $f_r$, whose value (*flux* $u_r$) is the molar quantity of matter transformed by the reaction.

$A + B \xrightarrow{r} B + C$ and $f_r = 3az^2$, with $y = |Y|$, $Y = A, B, C, Z$



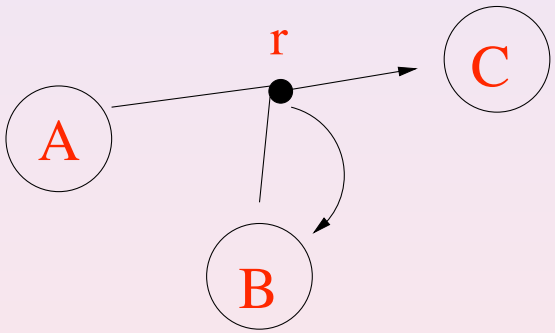

*Z parameter*
(massless entity)

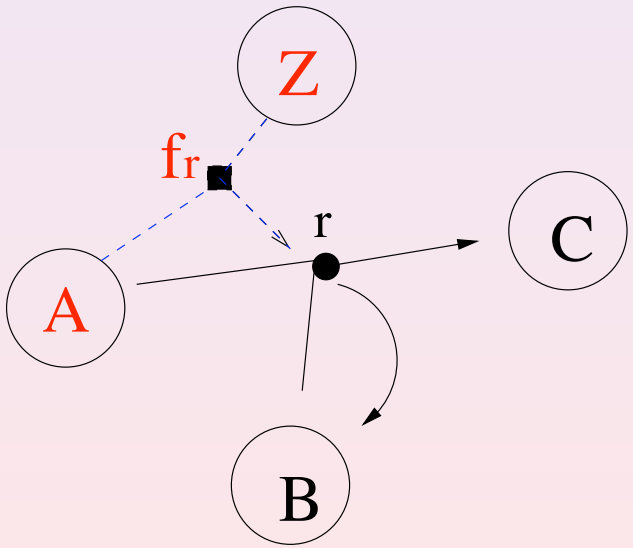
- $f_r = 1.3z^{-2}$ (inhibition)
- $f_r = 0.8a^3$ (promotion)
- $\lambda \xrightarrow{r} B,\ f_r = k$ (production)
- $C \xrightarrow{r} \lambda,\ f_r = k$ (degradation)

# Metabolic P systems: an intuition

Multiset rewriting system: a set of reactions r (transforming metabolites), corresponding regulators $f_r$, whose value (*flux* $u_r$) is the molar quantity of matter transformed by the reaction.

$A + B \xrightarrow{r} B + C$ and $f_r = 3az^2$, with $y = |Y|$, $Y = A, B, C, Z$




*Z parameter*
(massless entity)

- $f_r = 1.3z^{-2}$ (inhibition)
- $f_r = 0.8a^3$ (promotion)
- $\lambda \xrightarrow{r} B,\ f_r = k$ (production)
- $C \xrightarrow{r} \lambda,\ f_r = k$ (degradation)

# MP Systems - A Formal Definition

MP systems [2] have a structure

$$(S, R, H, \Phi, \tau, \nu, \mu)$$

where there are defined a set $S$ of metabolites, a set $R$ of rewriting rules, a (vector) function $H$ providing the parameter evolution, a set $\Phi$ of *regulators* associated to the rules, a time interval $\tau$, and a couples of units of measure $\nu \in \mathbb{R}$ and $\mu \in \mathbb{R}^n$, which respectively represent the number of molecules of a *conventional mole* and the molar masses of metabolites.

**Remark**: Regulators $f_r$ are real maps defined over the *system states* Q (i.e., vectors of metabolite and parameter values at a certain moment $i$): $f_r(Q[i]) = u_r[i], i \in \mathbb{N}$.

[2] Sections 3.1 and 3.2 + pag 134 (LGSS)

# Deterministic Evolution of a Metabolic System

Let us namely consider the evolution of the metabolite $c$:

$r_1 = c \to o + 12h + p$
$r_2 = c \to c + q$
$r_3 = c \to c + f$
$r_4 = o \to c$

The stoichiometric balance of $c$ is
$\Delta c[i] = c[i+1] - c[i] = -u_1[i] + u_4[i]$

By associating vectors to metabolites, rules, and fluxes:

$$\begin{array}{c} \Delta X[i] \\ \| \\ \begin{pmatrix} \Delta o[i] \\ \Delta c[i] \\ \Delta h[i] \\ \Delta p[i] \\ \Delta q[i] \\ \Delta f[i] \end{pmatrix} \end{array} = \left(\begin{array}{c|c|c|c} r_1 & r_2 & r_3 & r_4 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 \\ 12 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \times \begin{array}{c} U[i] \\ \| \\ \begin{pmatrix} u_1[i] \\ u_2[i] \\ u[3] \\ u[4] \end{pmatrix} \end{array}$$

# Deterministic Evolution of a Metabolic System

Let us namely consider the evolution of the metabolite $c$:

$r_1 = c \to o + 12h + p$
$r_2 = c \to c + q$
$r_3 = c \to c + f$
$r_4 = o \to c$

The stoichiometric balance of $c$ is
$\Delta c[i] = c[i+1] - c[i] = -u_1[i] + u_4[i]$

By associating vectors to metabolites, rules, and fluxes:

$$\begin{array}{c} \Delta X[i] \\ \| \\ \begin{pmatrix} \Delta o[i] \\ \Delta c[i] \\ \Delta h[i] \\ \Delta p[i] \\ \Delta q[i] \\ \Delta f[i] \end{pmatrix} \end{array} = \left(\begin{array}{c|c|c|c} \begin{array}{c} r_1 \\ \downarrow \end{array} & \begin{array}{c} r_2 \\ \downarrow \end{array} & \begin{array}{c} r_3 \\ \downarrow \end{array} & \begin{array}{c} r_4 \\ \downarrow \end{array} \\ \hline 1 & 0 & 0 & -1 \\ \hline -1 & 0 & 0 & 1 \\ \hline 12 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \end{array}\right) \times \begin{array}{c} U[i] \\ \| \\ \begin{pmatrix} u_1[i] \\ u_2[i] \\ u[3] \\ u[4] \end{pmatrix} \end{array}$$

# Dynamics of an MP System

The dynamics of an MP system is given by the recurrence equations EMA[i] (Equational Metabolic Algorithm):

$$X[i+1] = A \times U[i] + X[i]$$

$$U[i] = \Phi(Q[i])$$

where X is the metabolite vector, A the *stoichiometric matrix*, U the flux vector, $\Phi$ the regulator vector, and Q the system state.

# Metabolic computing - an example

Rules above for a metabolic system, where M represents the initial state $X(0)$. Compute $X(1)$, according to the flux maps: $u_1 = f_1(a, b, c) = ab$, $u_2 = f_2(a, b, c) = c^2$, $u_3 = f_3(a, b, c) = 2a$, $u_4 = f_4(a, b, c) = a$, $u_5 = f_5(a, b, c) = c$.

By EMA: $X[1] = A \times U[0] + X[0]$, where:

$$A = \begin{pmatrix} 1 & -1 & -1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 \end{pmatrix}, U[0] = \begin{pmatrix} 4 \\ 4 \\ 4 \\ 2 \\ 2 \end{pmatrix}, \text{ and}$$

$$X(0) = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \text{ therefore } X(1) = \begin{pmatrix} 0 \\ 10 \\ 4 \end{pmatrix}.$$

# MP System

$$M = (X, V, R, Q, \nu, \mu, \tau, \sigma_0, \Phi, \delta)$$

X = Substances

V = Parameters $\{h_v \mid v \in V\}$ $h_v : N \dashrightarrow R$

R = Reactions

Q = States

A state q is a function

$q : X \cup V \dashrightarrow R$

$\nu$ = Mole size

$q = \{x_1, x_2, \ldots, v_1, v_2, \ldots\}$

$\mu$ = Mole mass

$q[i] = (x_1[i], x_2[i], \ldots, v_1[i], v_2[i], \ldots)$

$\tau$ = Time unit

$\sigma_0$ = Initial state

H = Parameter Functions

$\Phi$= Flux Regulation Functions $\varphi_r : Q \dashrightarrow R$

$\delta$ = Dynamics

14

Courtesy of prof. Manca, Univ. VR, IT

# Metabolic Algorithm

**Dynamics**

$\delta$: N --> Q

$\delta(i) = (X[i], V[i])$

**Regulation**

$\phi = \{\varphi_r \mid r \in R\}$

$(X[0],V[0]) = \sigma_0$ $U[i] = \phi(X[i],V[i])$

$$X[i+1] = A \times U[i] + X[i]$$

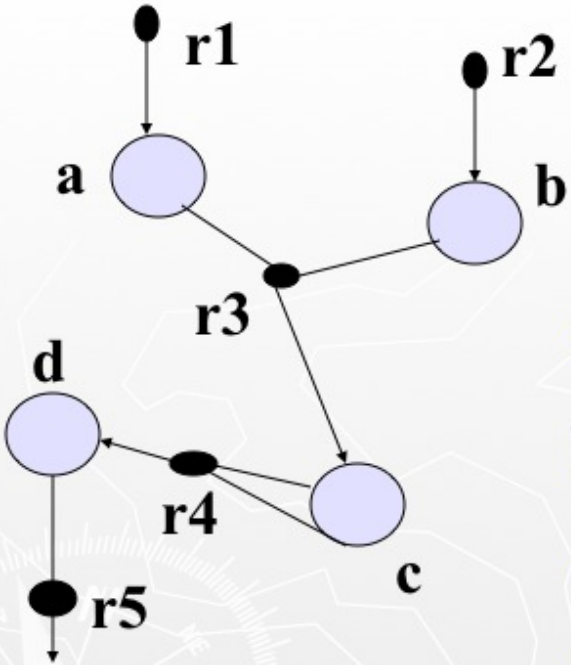
A =Stoichiometric Matrix, × = matrix product

Manca V., The Metabolic algorithm for P systems: Principles and Applications, TCS, 2008

15

Courtesy of prof. Manca, Univ. VR, IT

# From time series to a model




a[0], b[0], c[0], d[0]

Observation Times: 0, 1, 2, 3, …

Reaction Units: 0, 1, 2, 3, …

$a[i+1] - a[i] = u1[i] - u3[i]$

$b[i+1] - b[i] = u2[i] - u3[i]$

$c[i+1] - c[i] = u3[i] - 2u4[i]$

$d[i+1] - d[i] = u4[i] - u5[i]$

a[i], b[i], c[i], d[i]

u1[i], u2[i], … , u5[i]

i = 0, 1, 2, 3, …

⇒ a[i+1], b[i+1], c[i+1], d[i+1]

i = 0, 1, 2, 3, …

Courtesy of prof. Manca, Univ. VR, IT

# Estimation of flux parameters

$$a[i+1] - a[i] = u1[i] - u3[i]$$

$$b[i+1] - b[i] = u2[i] - u3[i]$$

$$c[i+1] - c[i] = u3[i] - 2u4[i]$$

$$d[i+1] - d[i] = u4[i] - u5[i]$$

Key values

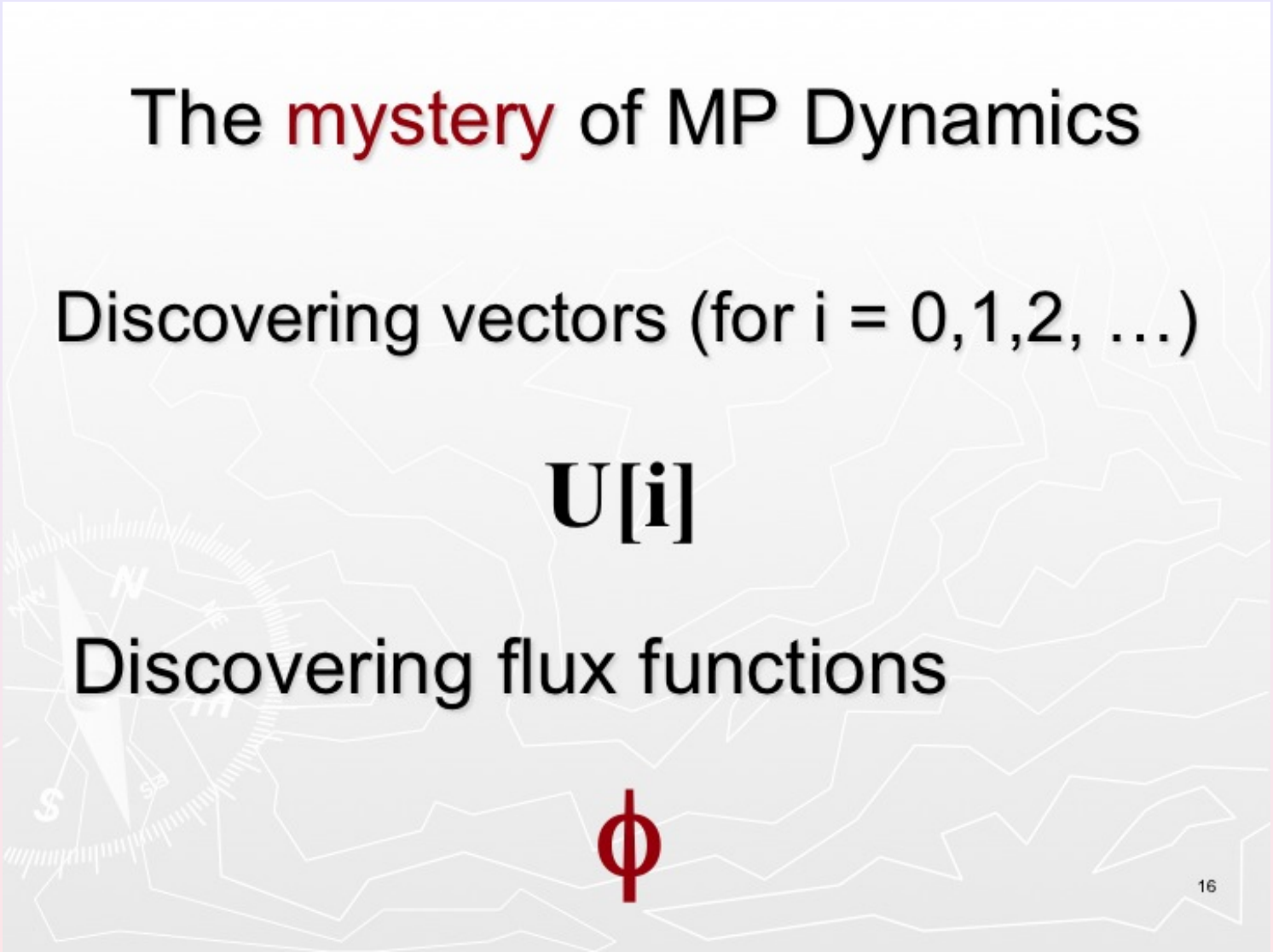
$$\begin{pmatrix} a[i+1] \\ b[i+1] \\ c[i+1] \\ d[i+1] \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} u1[i] \\ u2[i] \\ u3[i] \\ u4[i] \\ u5[i] \end{pmatrix} \quad \begin{pmatrix} a[i] \\ b[i] \\ c[i] \\ d[i] \end{pmatrix}$$

$$X[i+1] = A \quad U[i] \quad X[i]$$

8

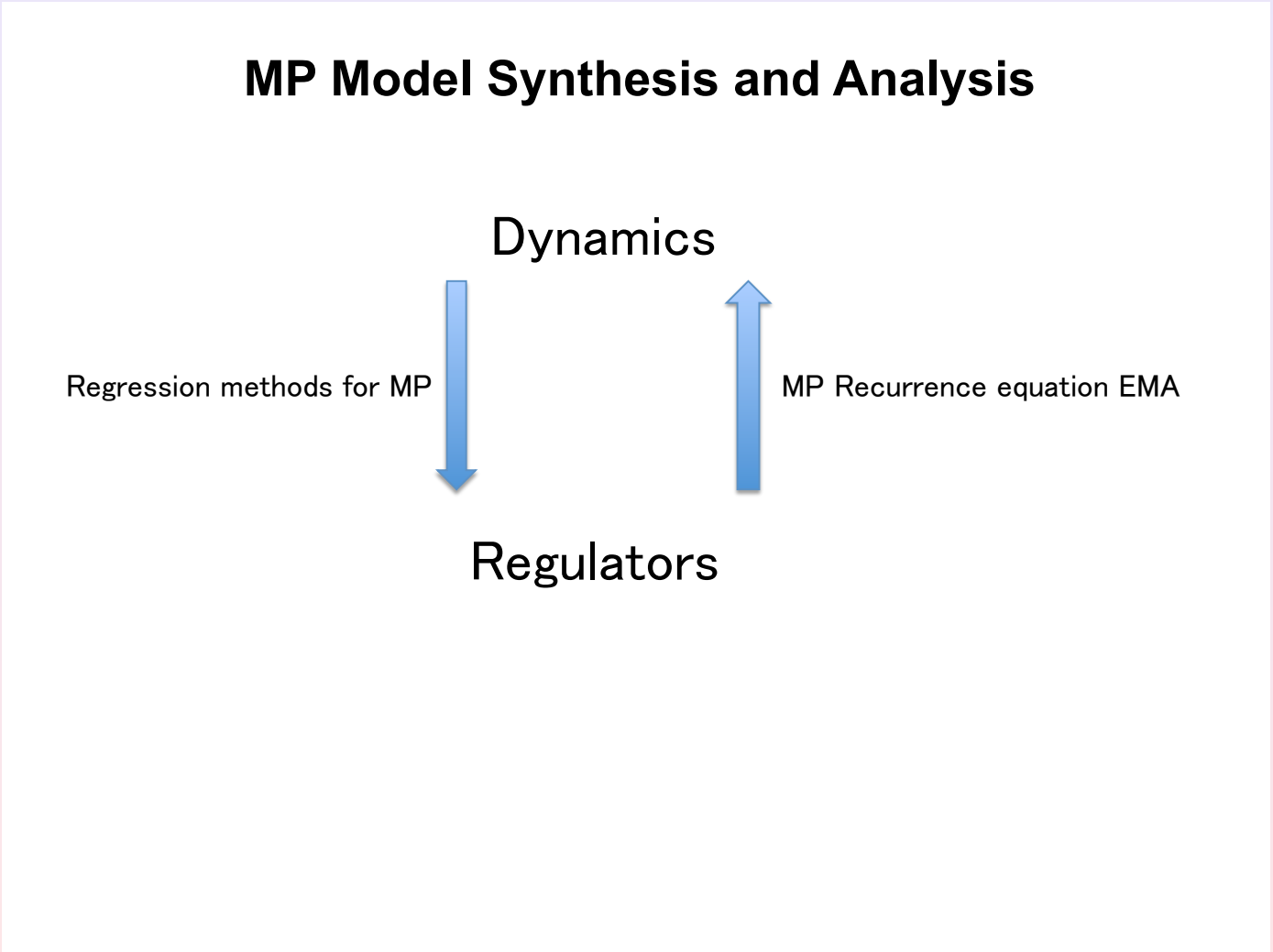
**A = Stoichiometric matrix deduced by R**

Courtesy of prof. Manca, Univ. VR, IT




Courtesy of prof. Manca, Univ. VR, IT

# Regulation Discovery Problem



# Log-gain Stoichiometric Stepwise Regression

Regulators are assumed to be polynomial functions:

$$f_r(Q) = c_1 p_1(Q) + \cdots + c_n p_n(Q).$$

Given a set of monomials $p_1, \ldots, p_n$, a first problem to synthesize an MP model is to find the coefficients $c_1, \ldots, c_n$ such that EMA approximates the observed temporal series.

Technically, monomials are admitted in a range of degrees, they are ranked by means of the log-gain principle [Manca '09], and orderly passed to a stepwise regression method, which is based on the Least Square Estimation and a Fisher Test for the analysis of variance.

# Log-gain Stoichiometric Stepwise Regression

$$A \times U[i] = \Delta X[i]$$

$U$ represents a vector of $m$ variables, over $n$ equations (and $m \geq n$, we have more constraints than variables)

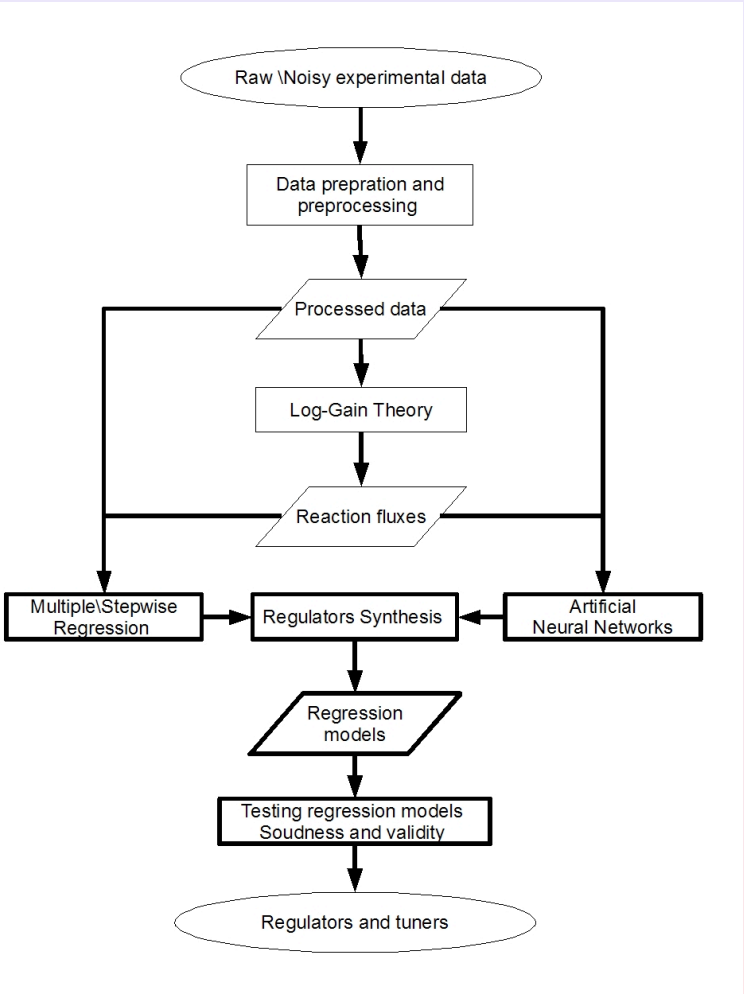
we assume $\Phi = C \times (g_1, \ldots, g_d)$ with $G = (g_1, \ldots, g_d)$ basic regressors (for example monomials)

$$A \times C \times G[i] = \Delta X[i]$$

now $n$ equations with $m \times d$ variables (coefficients in matrix C)

We take this system in $t$ time instants, so that to obtain $n \times t \geq md$ with a maximal rank (md), in order to apply the least square method to find values of C.

# Regression methods for MP systems

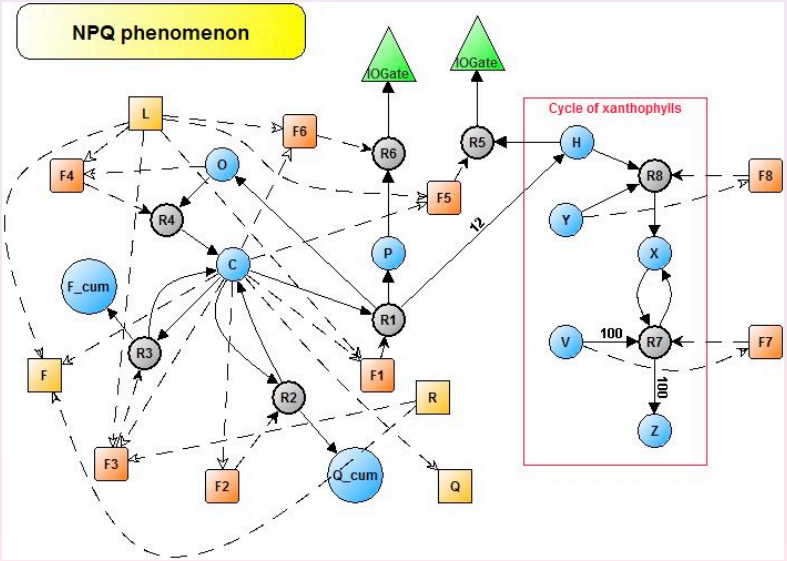



- Two regression techniques to synthesize MP regulators;
- An MP modeling procedure for data analysis from experimental data;
- Case studies: mitotic cycle, and photosynthetic phenomenon (NPQ).

• A. Castellini, **G. F.**, R. Pagliarini, *Data analysis pipeline from laboratory to MP models*, Natural Computing, 2010.

# An example

An MP model (10 metabolites, 4 parameters, 8 rules) for the Non-Photochemical Quenching phenomenon (NPQ), a relevant photosynthetic process - no differential model in literature.




| Grammar | Regulators |
|---|---|
| $r_1 = c \to o + 12h + p$ | $\varphi_1 = \alpha_1 - \beta_1 c + \gamma_1$ |
| $r_2 = c \to c + q^+$ | $\varphi_2 = -\alpha_2 + \beta_2 c - \gamma_2 c^2$ |
| $r_3 = c \to c + f^+$ | $\varphi_3 = \alpha_3 - \beta_3 c + \gamma_3 c^2 + \eta_3 \cdot r^{-1}$ |
| $r_4 = o \to c$ | $\varphi_4 = -\alpha_4 + \beta_4 o + \gamma_4$ |
| $r_5 = h \to \lambda$ | $\varphi_5 = \alpha_5 - \beta_5 c + \gamma_5$ |
| $r_6 = p \to \lambda$ | $\varphi_7 = -\alpha_7 + \beta_7 v$ |
| $r_7 = x + 100v \to x + 100z$ | $\varphi_6 = \alpha_6 - \beta_6 c + \gamma_5$ |
| $r_8 = y + h \to x$ | $\varphi_8 = \alpha_8 + \beta_8 y$ |

A. Castellini, **G. Franco**, R. Pagliarini, *Data analysis pipeline from laboratory to MP models*, Natural Computing, Springer. Article published online: April 30, 2010.

# Biological Networks

In biological literature, metabolic systems are usually represented by networks, where nodes are substances or enzymes and edges represent some “interaction”, such as:

- activation, inactivation
- catalysis (speeding up of the reaction, e.g., by means of binding)
- degradation, consumption
- engulfing, expulsion
- inhibition, promotion
- synthesis, production, “transformation”..

  ⋮

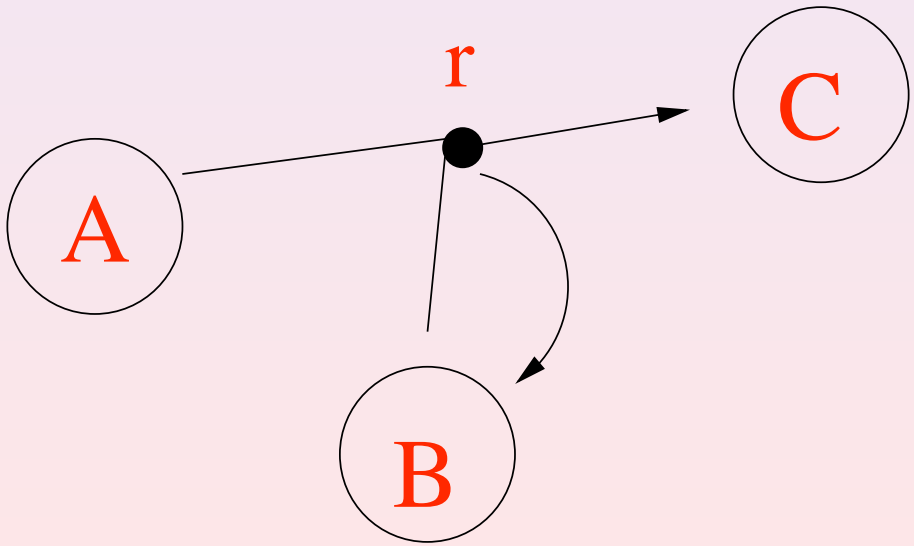
- “influence”

All concepts usually vague, ambiguous, depending on the specific contexts, formally not well-defined.

# Two Levels Graphs

The formalism we consider here suggests a natural representation of rules as graphs with two levels.

- The first level describes the reaction in itself (the *stoichiometry*, that is the network of physical connections between substances).

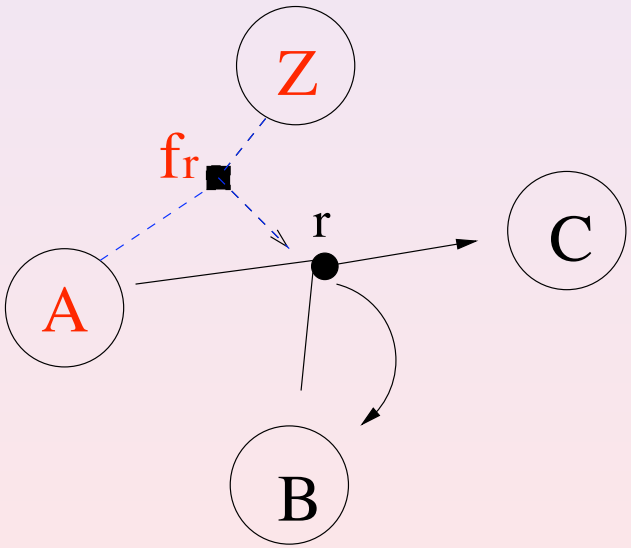



Example: $AB \xrightarrow{r} BC$

# Two Levels Graphs

- the second level expresses the layout of *regulation* which tunes the relative strengths of rules.

  Example: $f_r = 3az^2$, where $x = |X|$ for each object $X$




- $f_r = 1.3z^{-2}$ (inhibition)
- $f_r = 0.8a^3$ (promotion)
- $\lambda \xrightarrow{r} X,\ f_r = k$ (production)
- $X \xrightarrow{r} \lambda,\ f_r = k$ (consumption)

# MP Graphs
[Manca & Bianco, 2006]

Formally, a *MP graph* is a structure $G = (T, R, F, E, C)$ where T,R,F,E are sets of vertices and C is a set of edges

- $T$ is the set of circles nodes representing types and labeled by objects. Each $t \in T$ may be seen as a container holding a certain amount of a kind of substance.
- $R$ is the set of full bullet nodes representing biochemical reactions between types and labeled by reactions.
- $F$ is the set of full rectangles nodes labelled by reaction maps. These nodes are connected with a possibly empty set of circles (types) and with exactly one bullet node.
- $E$ is a set of nodes representing input or output gates.

# MP Graphs
[Manca & Bianco, 2006]

- *C* is a set of (plain, dashed) edges between nodes.

  i) *Plain edges* connect types from T (or E) to biochemical reactions from R. They specify *reactants* and *products* of the reaction, as arcs (lines) connect reactants to reactions while oriented arcs (arrows) connect reactions to products.

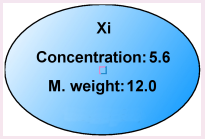
  ii) *Dashed edges* can connect types from T with reactivity nodes from F and *dashed oriented arcs* connect rectangles nodes from F with bullet nodes from R.

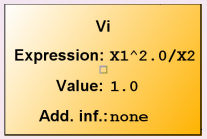
# MP graphs: basic concepts

**MP graphs** are a natural representation of MP systems as *bipartite graphs* with two levels [V. Manca, L. Bianco. BioSystems. To appear]:

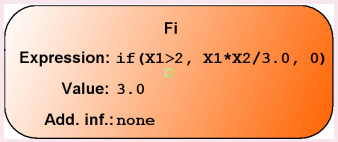
- **stoichiometric level** describes the stoichiometry of the reactions;
- **regulation level** tunes the flux of every reaction, i.e., the quantity of chemicals transformed at each step.

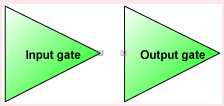
**MP graphs elements:**

Substances 


Parameters 

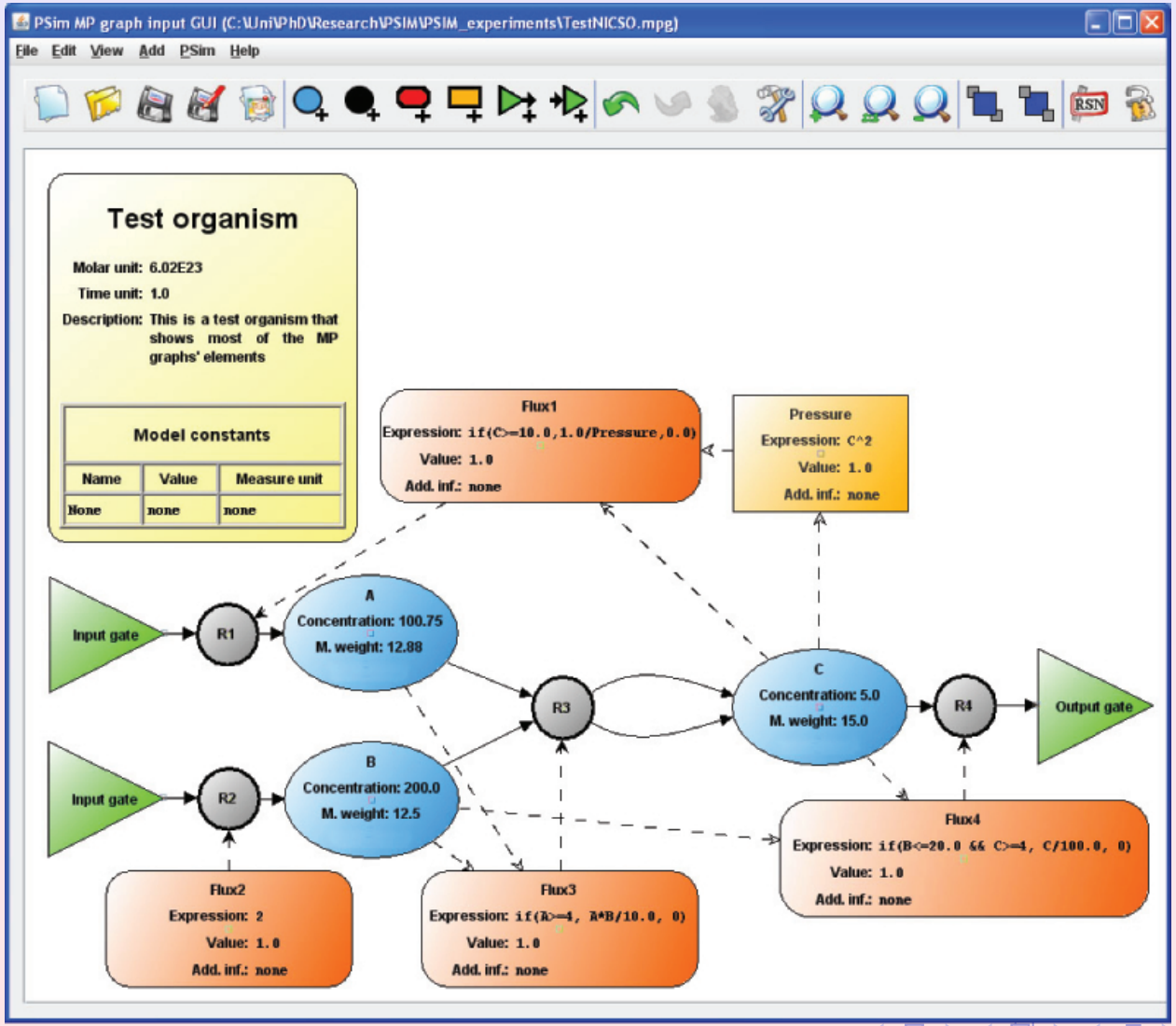

Reactions 


Fluxes 


Gates 

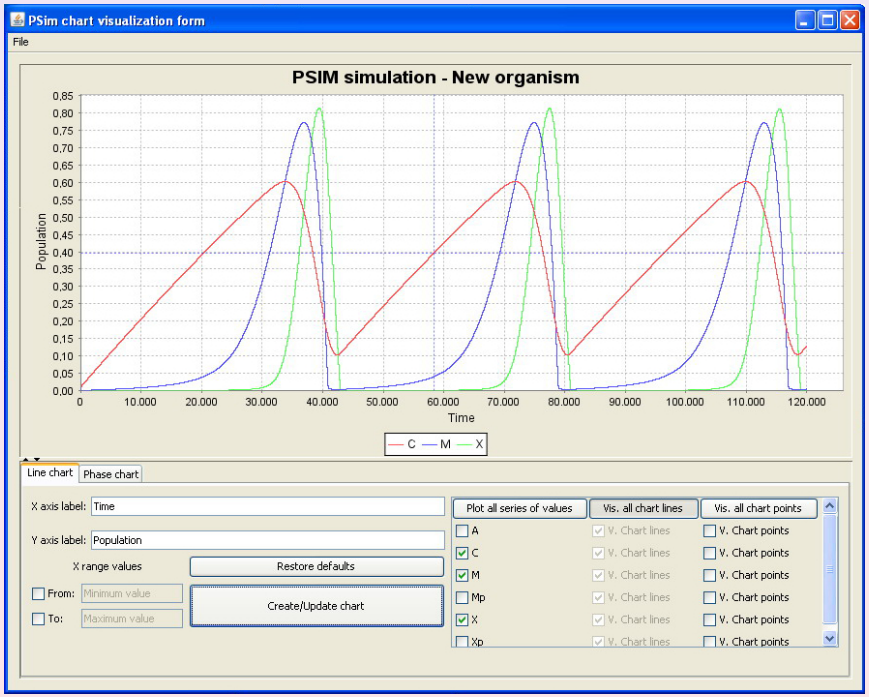

Arcs 

# A toy example



# MetaPlab and MpTheory

Software[3] was developed to provide modelers, as well as biologists, with a reliable and easy-to-use simulation environment to compute metabolic dynamics.




## Main features

- User-friendly interface
- Plugin architecture
- Flexibility and portability
- Cross application applicability

[3] `mplab.sci.univr.it` and `mptheory.scienze.univr.it`.

# Lotka-Volterra Equations

$$x' = ax - bxy$$
$$y' = -dx + exy$$

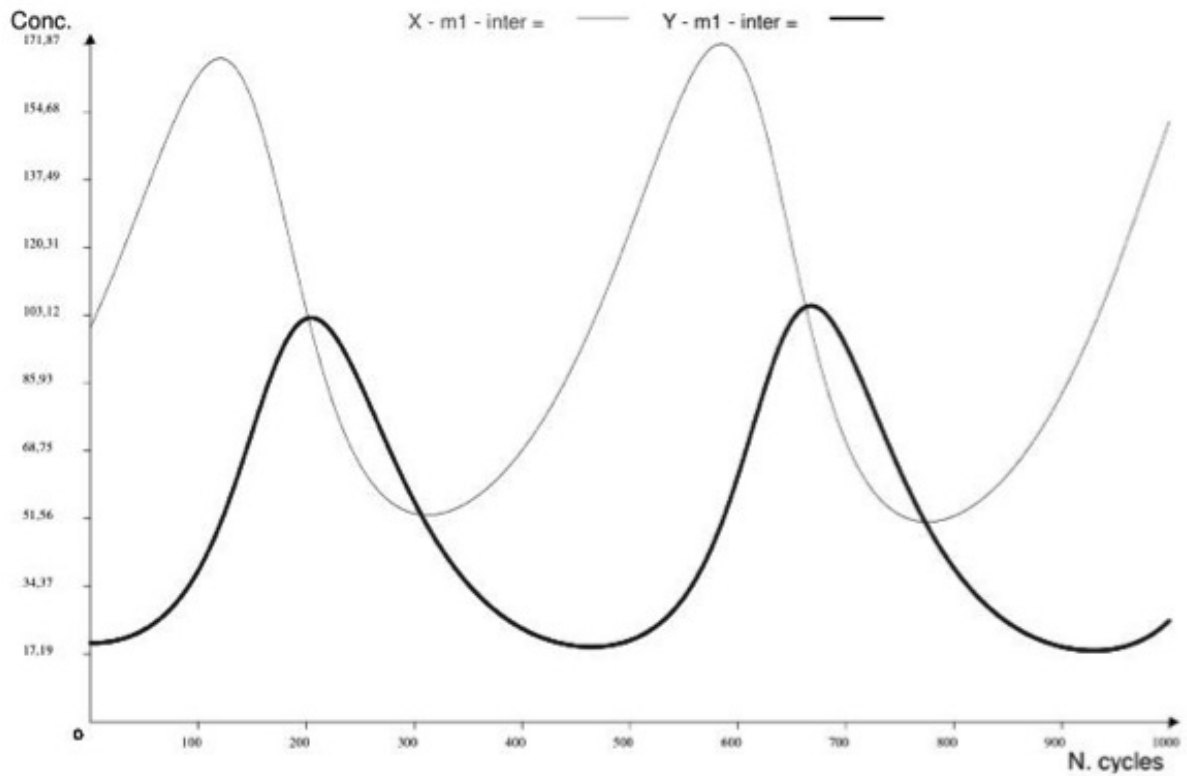
Prey-Predator Model

- R1 : X ---> XX
- R2 : XY ---> YY
- R3 : Y ---> $\lambda$

39

Courtesy of prof. Manca, Univ. VR, IT

Oscillations of the predator-prey model simulated by Psim with k1=0.01, k2=0.02, k3=0.02 and mu=100 (|M|=10000)




Courtesy of prof. Manca, Univ. VR, IT

# Belousov-Zabotinskii (The Troy Horse of Oscillations)

- Sulphuric Acid
- Malonic Acid
- Ferroin
- Sodium Bromate
- Cerium (catalyst)
- Organic
- After some initial time, sudden oscillations in color start which range from red to blue

33

Courtesy of prof. Manca, Univ. VR, IT

# Brussellator

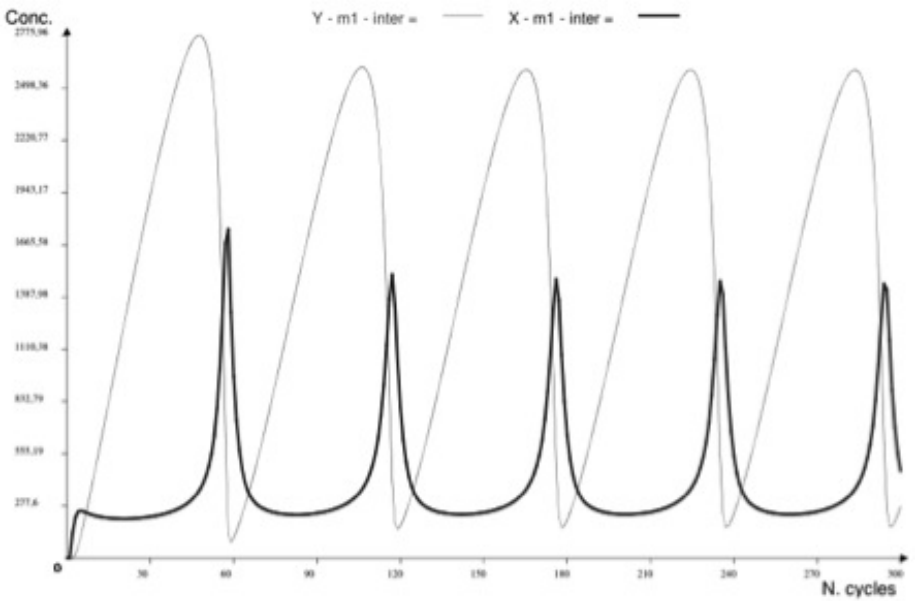
- R1 : A --> X
- R2 : BX ---> YD
- R3 : XXY ---> XXX
- R4 : X ---> C
- R5 : λ ---> A
- R6 : λ ---> B

Suzuki and Tanaka's Formulation

34

Courtesy of prof. Manca, Univ. VR, IT

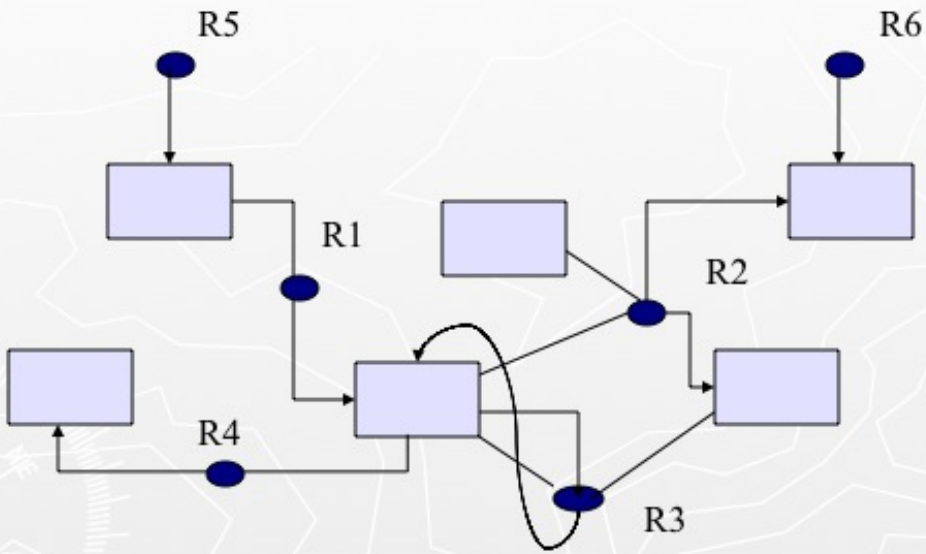
Oscillations of Belousov-Zhabotinskii reaction model simulated by Psim with parameters k1=0.9, k2=0.7, k3=0.36, k4=0.36, k5=0.1, k6=0.15 and mu=1000 (|M|=100000). Parameters could be rewritten in terms of k1 : k2=0.78 k1, k3=0.4 k1 and k4=0.4 k1, ...




35

Courtesy of prof. Manca, Univ. VR, IT

# Brusselator Metabolic Graph




X ---> Y ---> X

A neuron like membrane structure

45

Courtesy of prof. Manca, Univ. VR, IT

# Static/structural analysis of networks

Networks are high-dimentional graphs, often having a real world origin: social, economical, internet, biological.

Fundamental parameters are:

- (In, out – min, max, average) degree
- diameter (maximal distance between nodes)
- average path length is an average of averages [4]
- shortest path is the average of shortest paths between any couple of nodes.

[4] If this value is low for PPI, those proteins have tendency to form a functional complex.

# Scale free topology

Real (biological, social) [5] networks are scale-free: node degree distribution (portion of nodes having a given degree) follows a scale-invariant law

$$P(k) = ck^{-\gamma}$$

where $k > 0$, $c$ is a constant, $\gamma > 0$ is the characteristic exponent ($2 < \gamma < 3$ for biological networks).

Majority of nodes has small degree, living with few hubs [6].

Hypothesis of *preferential attachment*[Barabasi]

[5] citations, web access, books sold, phone calls, city populations, airports

[6] Nodes with degree much higher than the average degree

# Small world property

Each node of real networks may reach another one in a limited number of steps. This number is 5/6 in social networks, 8 in biological networks.

For a mathematical explanation of this phenomenon, see [Duncan Watts e Steven Strogatz, Nature 1998]

For fun, try to compute the (Kevin) Bacon number: `https://oracleofbacon.org`

# Node centrality

Network $N = (V, E)$.

- **Eccentricity**:

$$Ecc(v) = \frac{1}{max\{dist(v, w) : w \in V\}}$$

(high value if all nodes are close to $v$)

- **Closeness**:

$$Clo(v) = \frac{1}{\sum_{w \in V} dist(v, w)}$$

dist measures the shortest path (relaxed eccentricity)

- **Stress**:

$$Str(v) = \sum_{s \in V, s \neq v} \sum_{t \in V, t \neq v} \sigma_s p(v)$$

where $\sigma_s t(v)$ is the number of shortest paths between $s$ and $t$ involving also $v$

- Radiality (average centrality), betweeness (refined stress)

# Cytoscape

It is a publically available platform to analyze and visualize biological networks. It was published by a collaboration of Pasteur Institute, UCSD, UCSF, MSKCC, Unilever, ISB..

It is a plug-in architecture: open to be integrated by programmers from any part of the world.

www.cytoscape.org