Biologia de Sistemas para o desenvolvimento de fábricas celulares microbianas

Eugénio Campos Ferreira

Centre of Biological Engineering
University of Minho
Braga
Portugal

Universidade do Minho

OUTLINE

- Introduction to Systems Biology
- Metabolic Models
- Simulation
- Systems Metabolic Engineering (in silico)
- Bioprocess Engineering
- Conclusions

SYSTEMS BIOLOGY

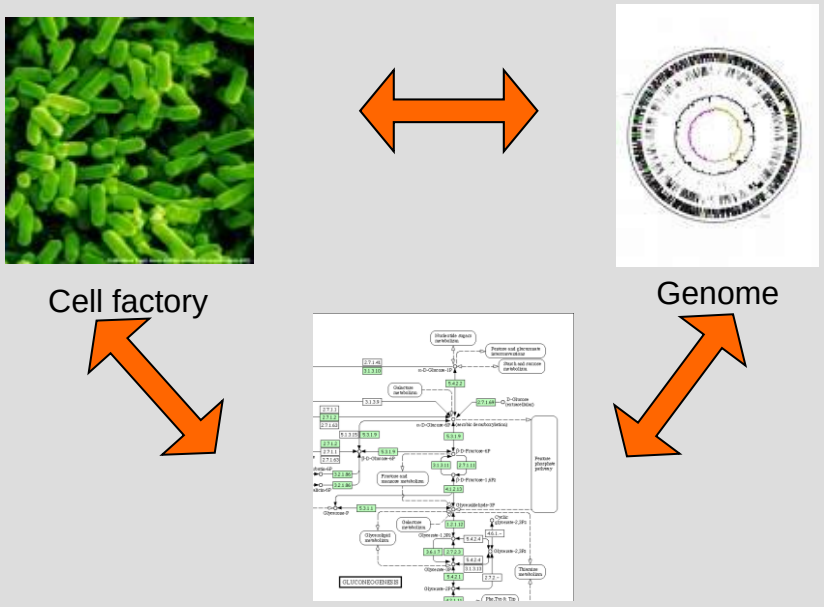
- **Systems Biology** does not investigate individual cellular components at a time, but the behaviour and relationships of all of the elements in a particular biological system while it is functioning
- **Metabolic Engineering** can gain major benefits from the systems biology approach

Systems biology involves the use of computer simulations of cellular subsystems (such as the networks of metabolites and enzymes which comprise metabolism, signal transduction pathways and gene regulatory networks) to both analyze and visualize the complex connections of these cellular processes.

Systems Biology approaches for modelling, optimization, and control of microbial cell factories

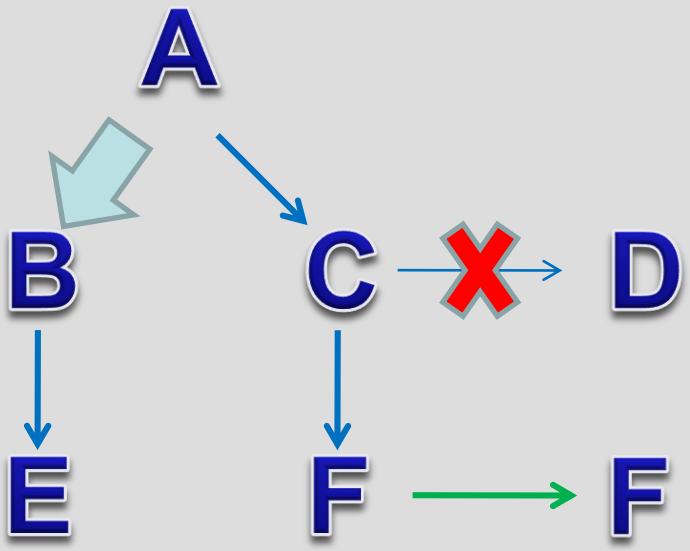
- Cellular Models for Metabolic Engineering: gene networks
- Inference of Biological Networks
  - From Genome-scale metabolic models
  - From experimental data
  - From literature data mining
- *In Silico* Metabolic Engineering Platforms: Optimization of Microbial strains – **OptFlux** tool

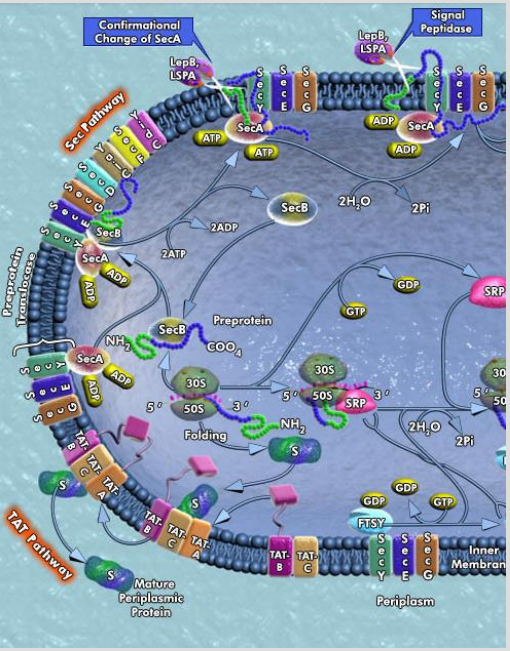
- Metabolic engineering - introduction of directed genetic modifications leading to desirable metabolic phenotypes



Metabolic Engineering Strategies:

- Gene Deletion
- Gene Addition
- Gene Overexpression
- Manipulation of environmental conditions

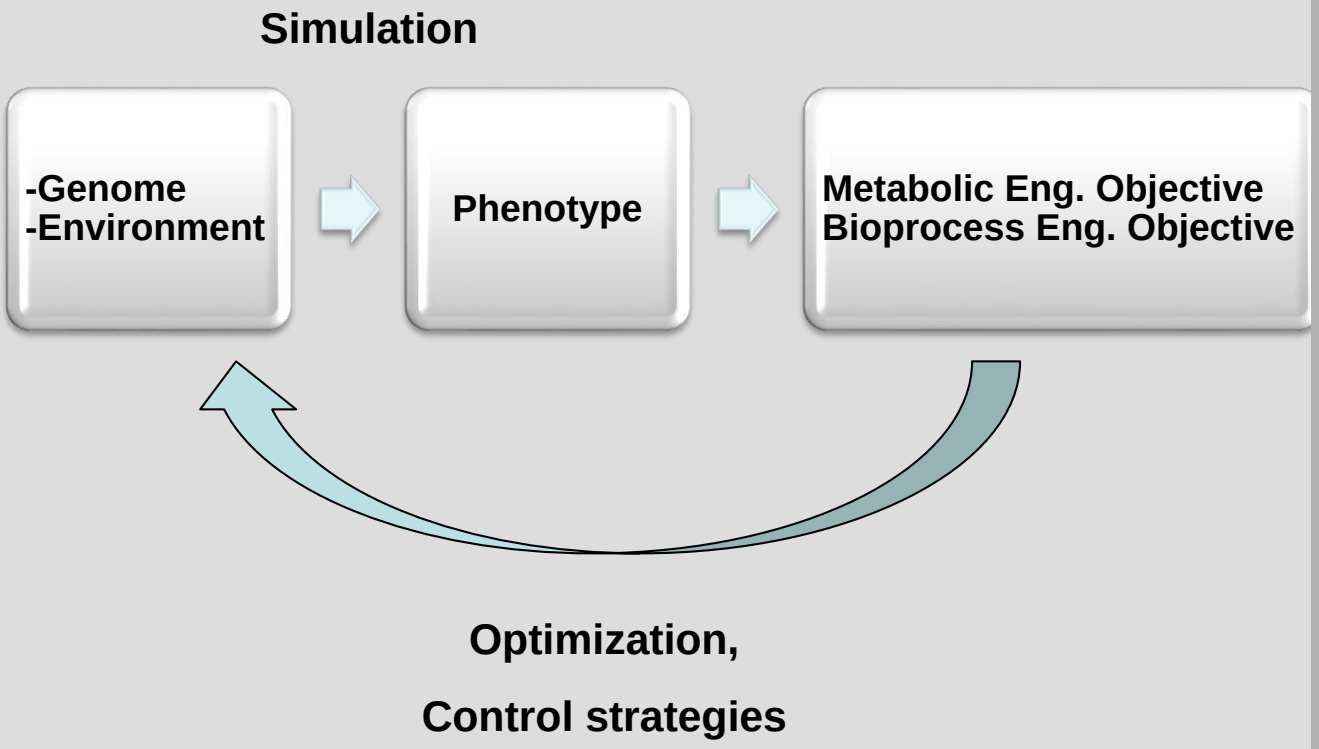




Bioprocess Engineering

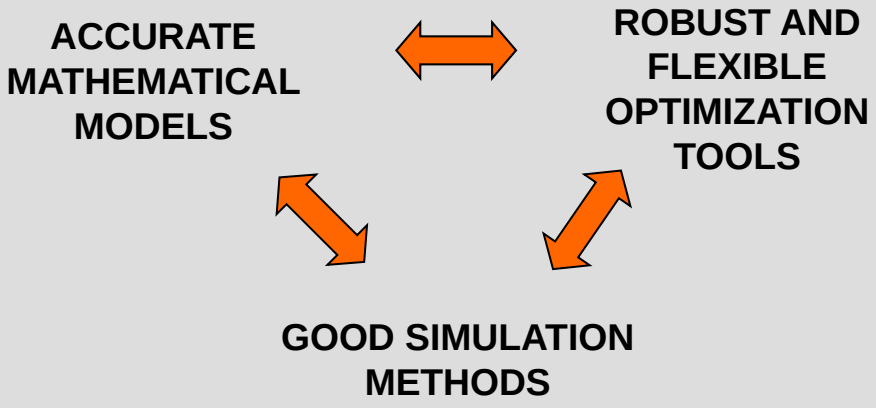
Metabolic Engineering

- A view of the Metabolic Engineering / Bioprocess Engineering Problem



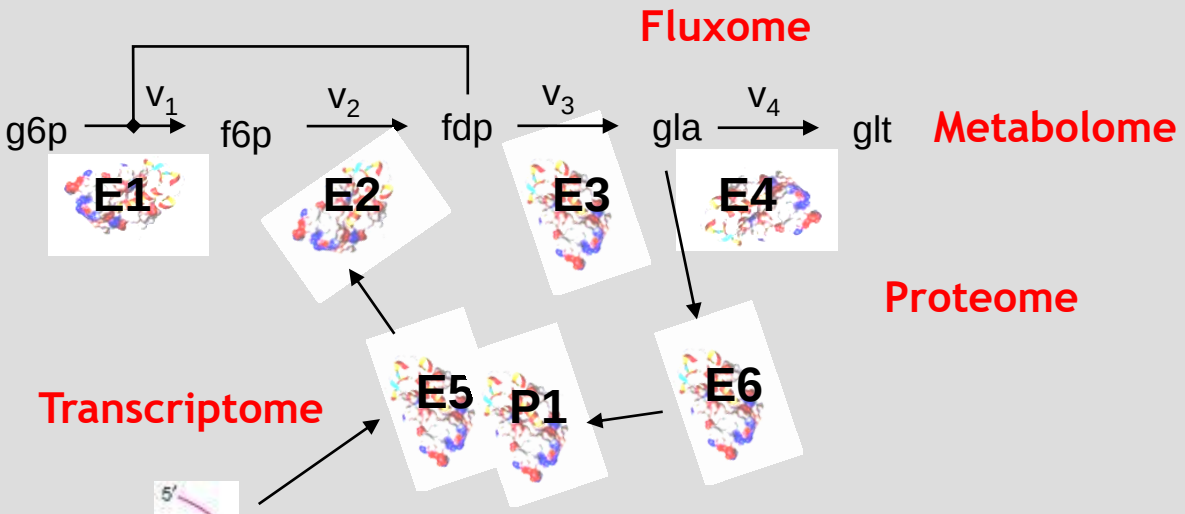
- In metabolic engineering problems, it is often difficult to identify *a priori* which genetic manipulations will originate a given desired phenotype

- In order to rationally design production strains with enhanced capabilities, it is essential to have:



Models should comprise different levels of information:

- Reactions stoichiometry
- Reactions kinetics
- Regulatory information



$$\frac{dC_{g6p}}{dt} = v_{PTA} - v_{PGI} - v_{G6PDH} - \mu C_{G6P}$$

$$v_{PFK} = \frac{v_{PFK}^{max} C_{atp} C_{f6p}}{\left(C_{atp} + K_{PFK,atp,s}\left(1 + \frac{C_{adp}}{K_{PFK,adp,s}}\right)\right) + \left(C_{f6p} + K_{PFK,f6p,s}\frac{A}{B}\right)\left(1 + \frac{L_{PFK}}{\left(1 + C_{f6p}\frac{B}{K_{PFK,f6p,s}}^{A}\right)^{nPFK}}\right)}$$

- Different model types are at different development stages...

	Stoichiometric	Regulatory	Kinetic
Genome-scale	**	***	*
Simul. Accuracy	*	**	***
Gene Deletions	***	***	***
Gene Over/ under express.	-	-	***

Framework for calculation of intracellular metabolic (net) fluxes is based on:

- **Mass balance** over intra-cellular metabolites

$$\frac{d[Tyr]}{dt} = v_1 - v_2 - v_3$$

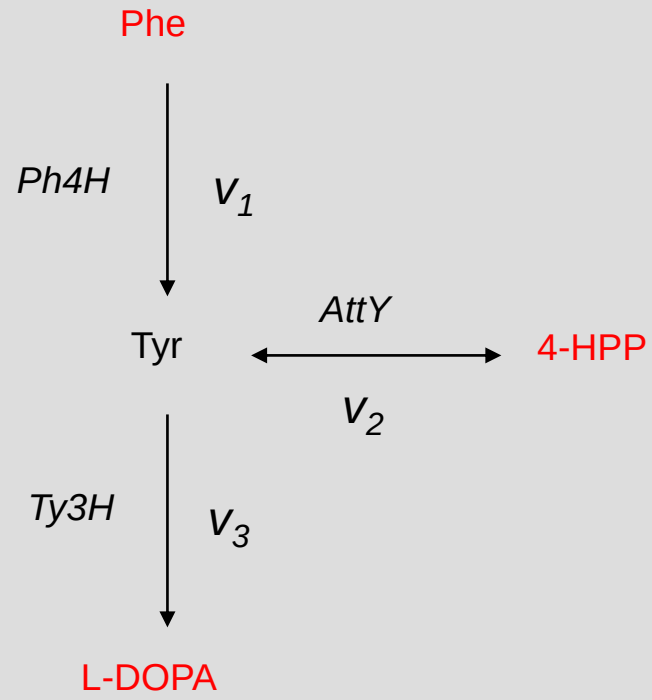
- Assumption of (pseudo) **steady state**

$$v_1 - v_2 - v_3 = 0$$

- For all metabolites:

$$S \cdot v = 0$$

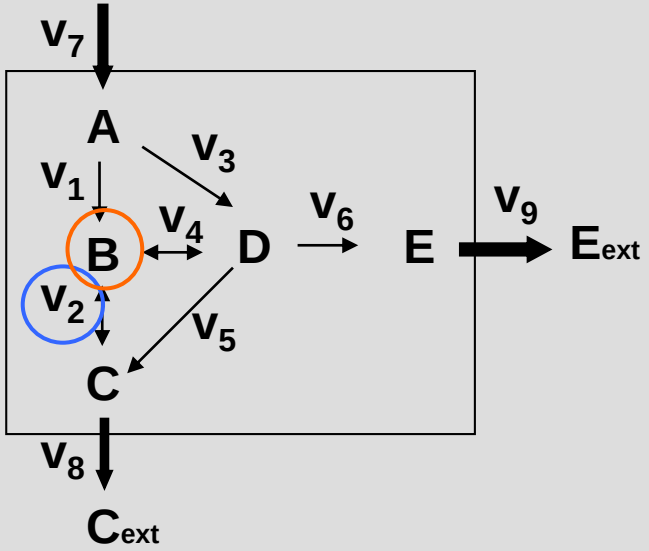
$$\beta_j \leq v_j \leq \alpha_j$$



This procedure is repeated for *all* considered metabolites and will originate the so-called stoichiometric model

The result is a Linear Equations system described by stoichiometric matrix *S.*

For an identified reaction set:



$A \xrightarrow{v_1} B$

$B \xleftrightarrow{v_2} 2C$

$A \xrightarrow{v_3} D$

$B \xleftrightarrow{v_4} D$

$D \xrightarrow{v_5} C$

$D \xrightarrow{v_6} E$

$A_{ext} \xrightarrow{v_7} A$

$C \xrightarrow{v_8} C_{ext}$

$E \xrightarrow{v_9} E_{ext}$

$A : -v_1 - v_3 + v_7 = 0$

$B : v_1 - v_2 - v_4 = 0$

$C : 2v_2 + v_5 - v_8 = 0$

$D : v_3 + v_4 - v_5 - v_6 = 0$

$E : v_6 - v_9 = 0$

$0 \le v_1 \le +\infty$

$-\infty \le v_2 \le +\infty$

$0 \le v_3 \le +\infty$

$-\infty \le v_4 \le +\infty$

$0 \le v_5 \le +\infty$

$0 \le v_6 \le +\infty$

$0 \le v_7 \le a$

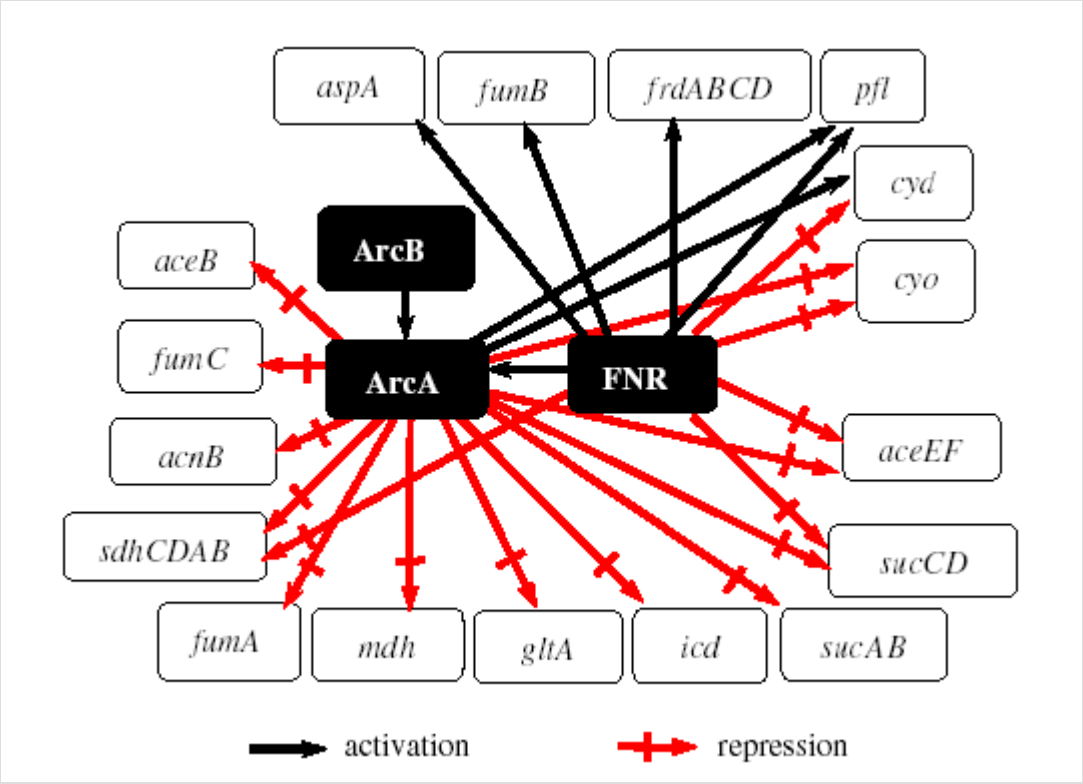
$0 \le v_8 \le +\infty$

$0 \le v_9 \le +\infty$

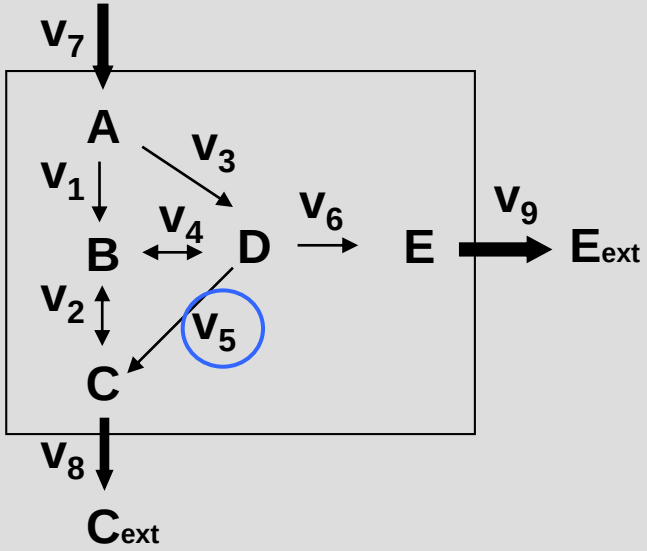
$$\begin{bmatrix} -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$S \cdot v = 0$$

- Gene Regulatory Networks represent regulatory elements and their interactions
- A regulatory network will direct the activation or repression of a set of genes in response to a specific environmental stimulus, like O_2 or pH
- In the figure, ArcA and FNR are transcription factors



For an identified reaction set:



$A \xrightarrow{v_1} B$

$B \xleftrightarrow{v_2} 2C$

$A \xrightarrow{v_3} D$

$B \xleftrightarrow{v_4} D$

$D \xrightarrow{v_5} C$

$D \xrightarrow{v_6} E$

$A_{ext} \xrightarrow{v_7} A$

$C \xrightarrow{v_8} C_{ext}$

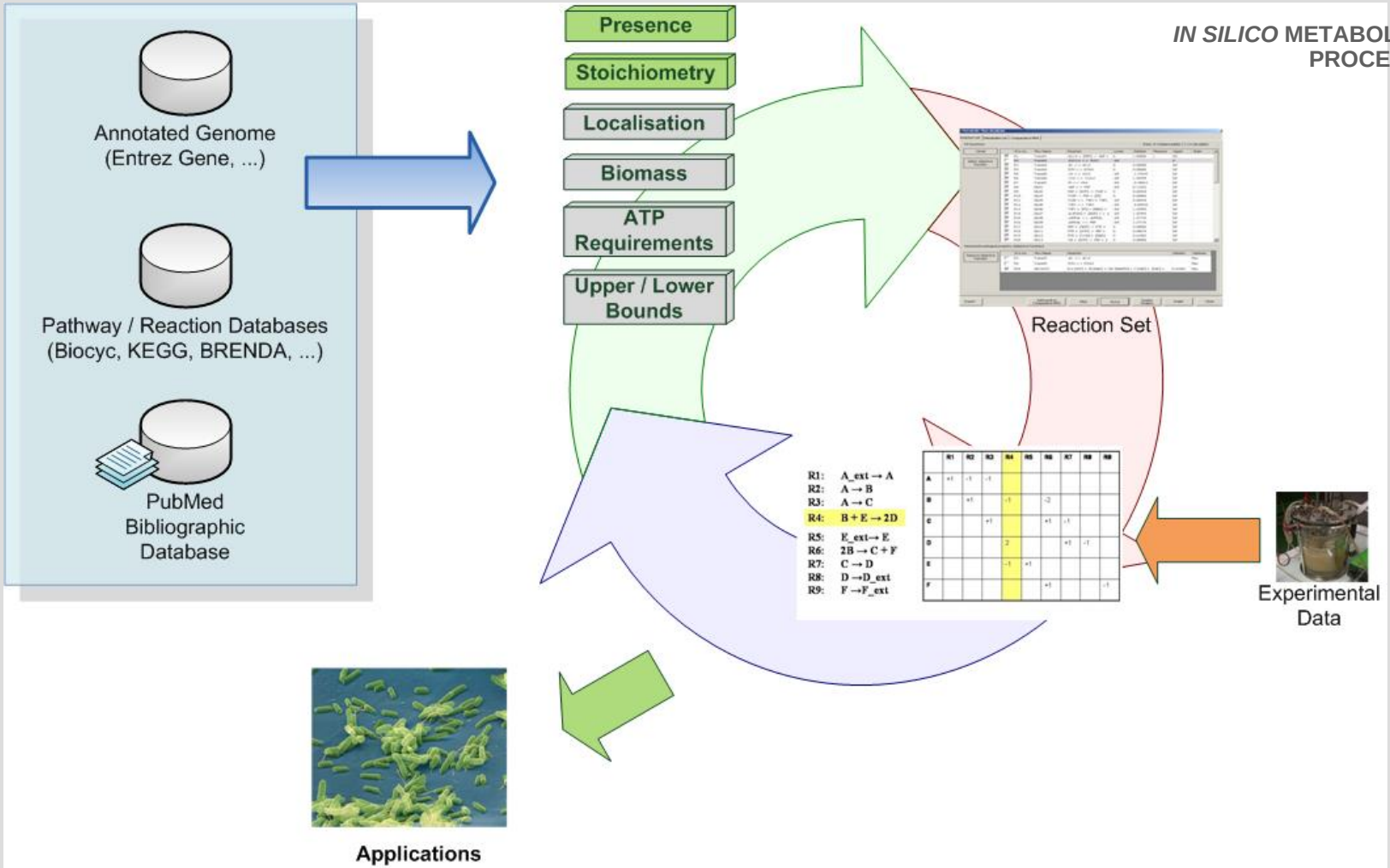
$E \xrightarrow{v_9} E_{ext}$

If for condition Z, reaction 5 does not occur

$$A: -v_1 - v_3 + v_7 = 0$$
$$B: v_1 - v_2 - v_4 = 0$$
$$C: 2v_2 + v_5 - v_8 = 0$$
$$D: v_3 + v_4 - v_5 - v_6 = 0$$
$$E: v_6 - v_9 = 0$$

$$0 \le v_1 \le +\infty$$
$$-\infty \le v_2 \le +\infty$$
$$0 \le v_3 \le +\infty$$
$$-\infty \le v_4 \le +\infty$$
$$\boxed{0 \le v_5 \le 0}$$
$$0 \le v_6 \le +\infty$$
$$0 \le v_7 \le a$$
$$0 \le v_8 \le +\infty$$
$$0 \le v_9 \le +\infty$$

$$\begin{bmatrix} -1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$



Rocha et al (2008) Methods in Molecular Biology, Vol. 416, Ch. 29, 409

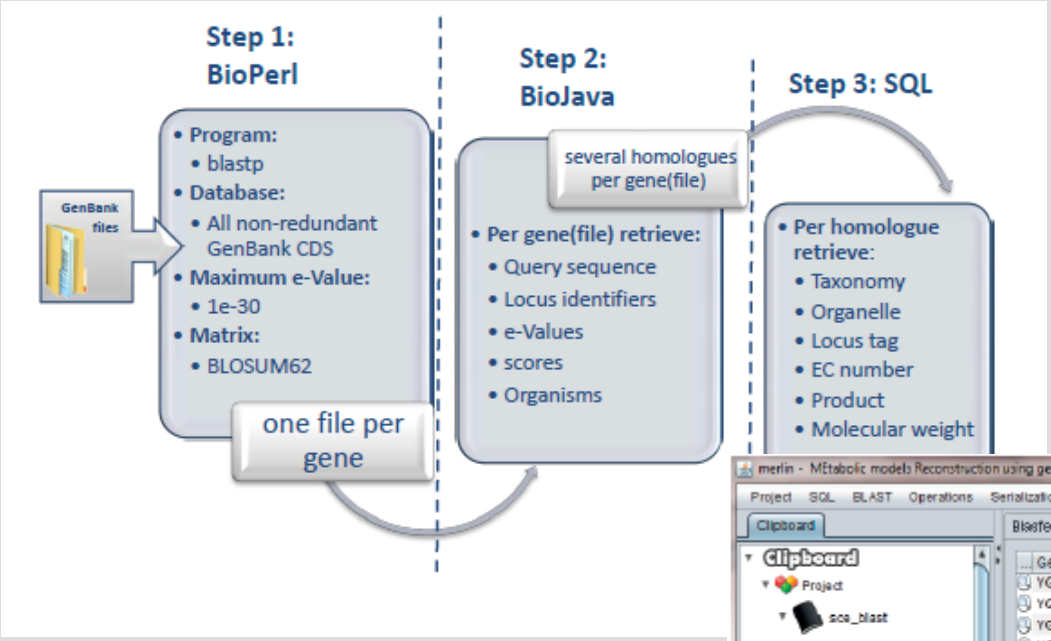
Microorganism	On-line availability
Haemophilus influenzae	http://gcrg.ucsd.edu/organisms/hinfluenzae.html
Escherichia coli	http://gcrg.ucsd.edu/organisms/ecoli_reactions.html
Helicobacter pylori	http://gcrg.ucsd.edu/organisms/hpylori.html
Saccharomyces cerevisiae	http://www.cpb.dtu.dk/models/yeastmodel.html http://systemsbiology.ucsd.edu/organisms/yeast.html
Aspergillus niger	http://blackwellpublishing.com/products/journals/suppmat/EJB/EJB3798/EJB3798sm.htm
Plasmodium falciparum	http://plasmocyc.stanford.edu
...	

Applications of Genome-scale metabolic models:

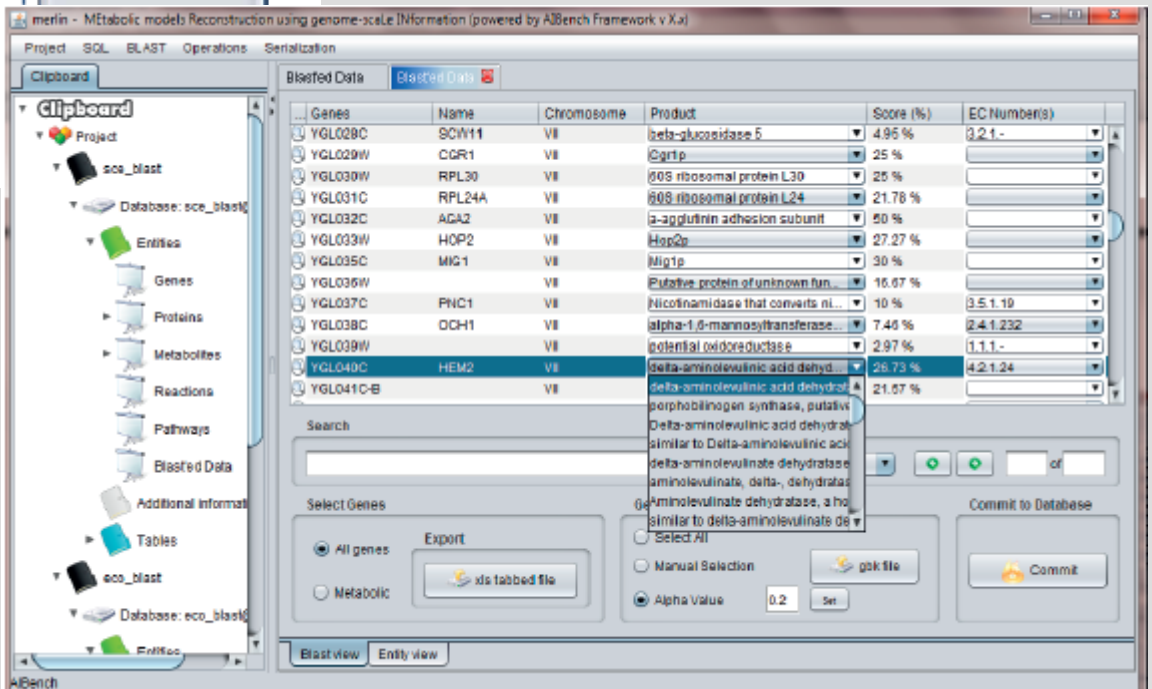
- Design of industrial strains for industrial biotechnology
- Growth medium design
- Discovery of new gene functions
- Better understanding of microbial physiology
- Identification of potential drug targets in pathogens
- ...

METABOLIC MODELS
GENOME ANNOTATION IN RECONSTRUCTION

- Annotation available usually after sequencing
- However, it can be old or incomplete!



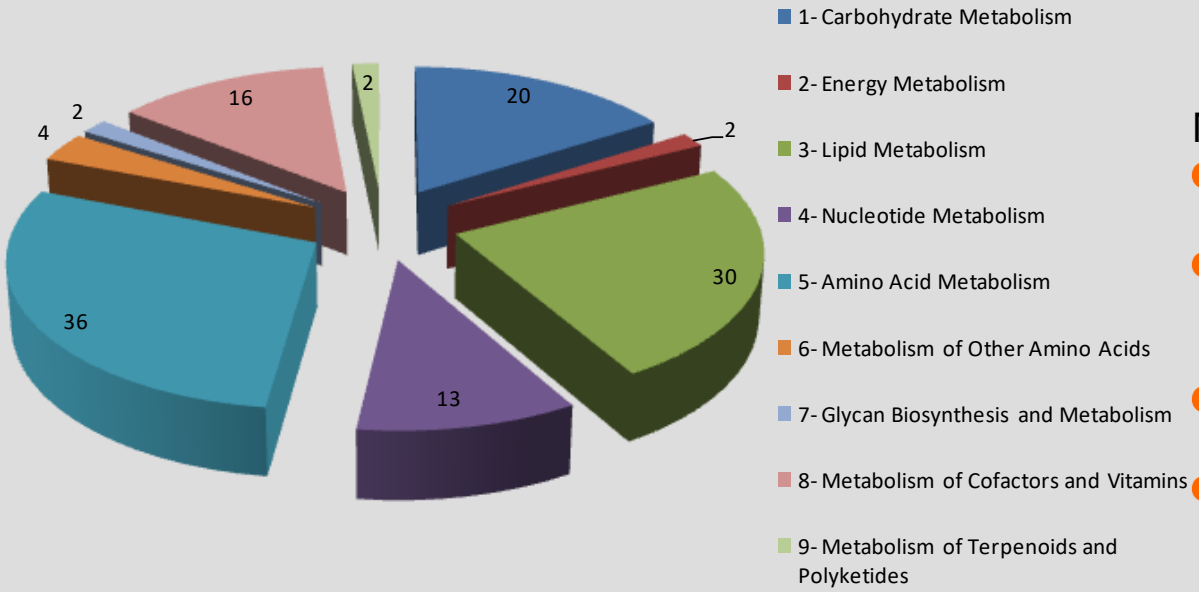
merlin:
Metabolic Models Reconstruction using Genome-Scale Information



Dias *et al*. (2010) Computers Application in Biotechnology

METABOLIC MODELS
K. LACTIS RECONSTRUCTION

Metabolism - new enzymes

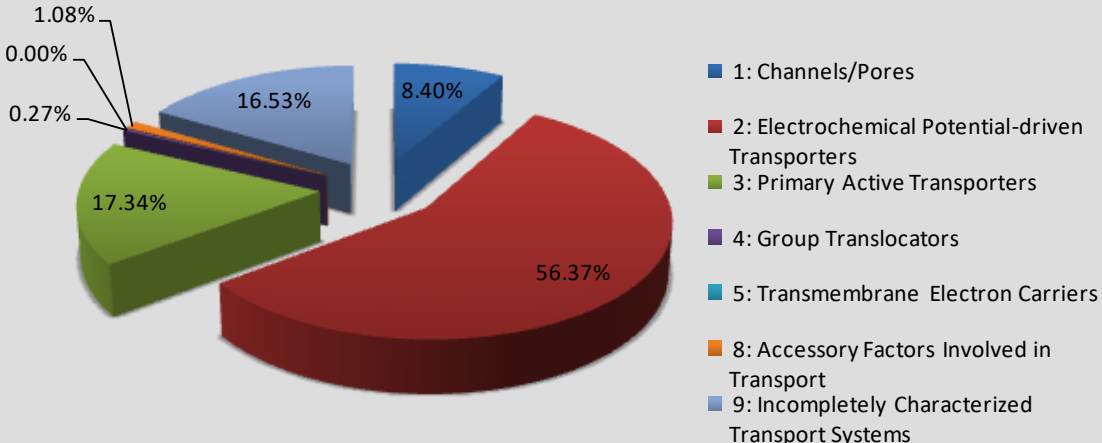


MOTIVATION

- Growth on **lactose** as a sole carbon source
- Various industrial applications, especially in the **dairy industry** but also host for **recombinant proteins**
- Molecular tools that make it amenable to **genetic manipulation**
- Evolutionary distance to *S. cerevisiae* allows to perform **comparative studies** between these two species

# of *K. lactis* genes with:	distinct	total
Yeast metabolic homologues	1627	1725
K. lactis transporter classfication (TC) annotation	6	6
Other metabolic homologues	62	70

TC numbers



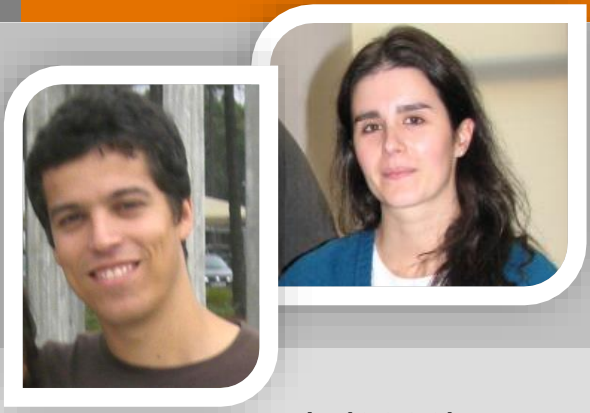
Dias *et al*. (2011) (submitted)

METABOLIC MODELS
DATA INTEGRATION DURING RECONSTRUCTION

Matching Problems for compounds in *E. coli* in different databases

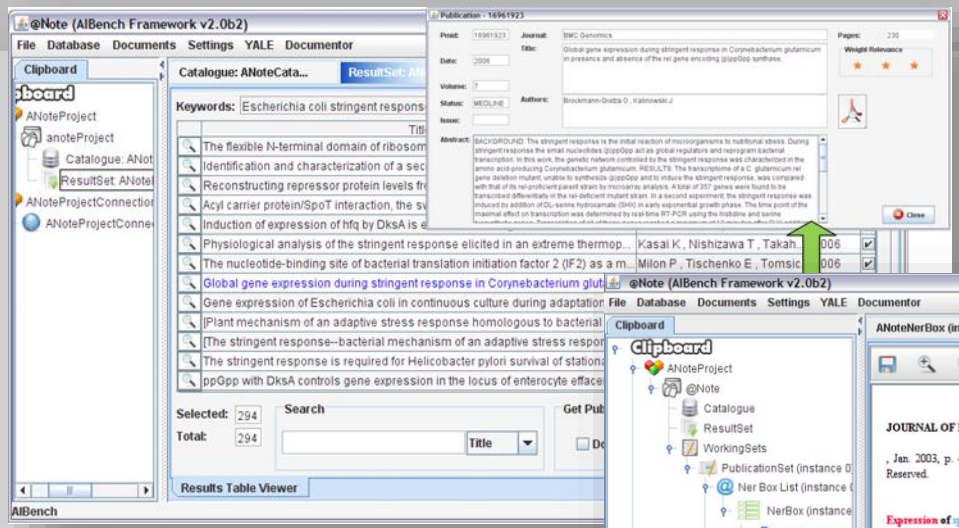
Lourenço et al. (2011) Briefing in Bioinformatics

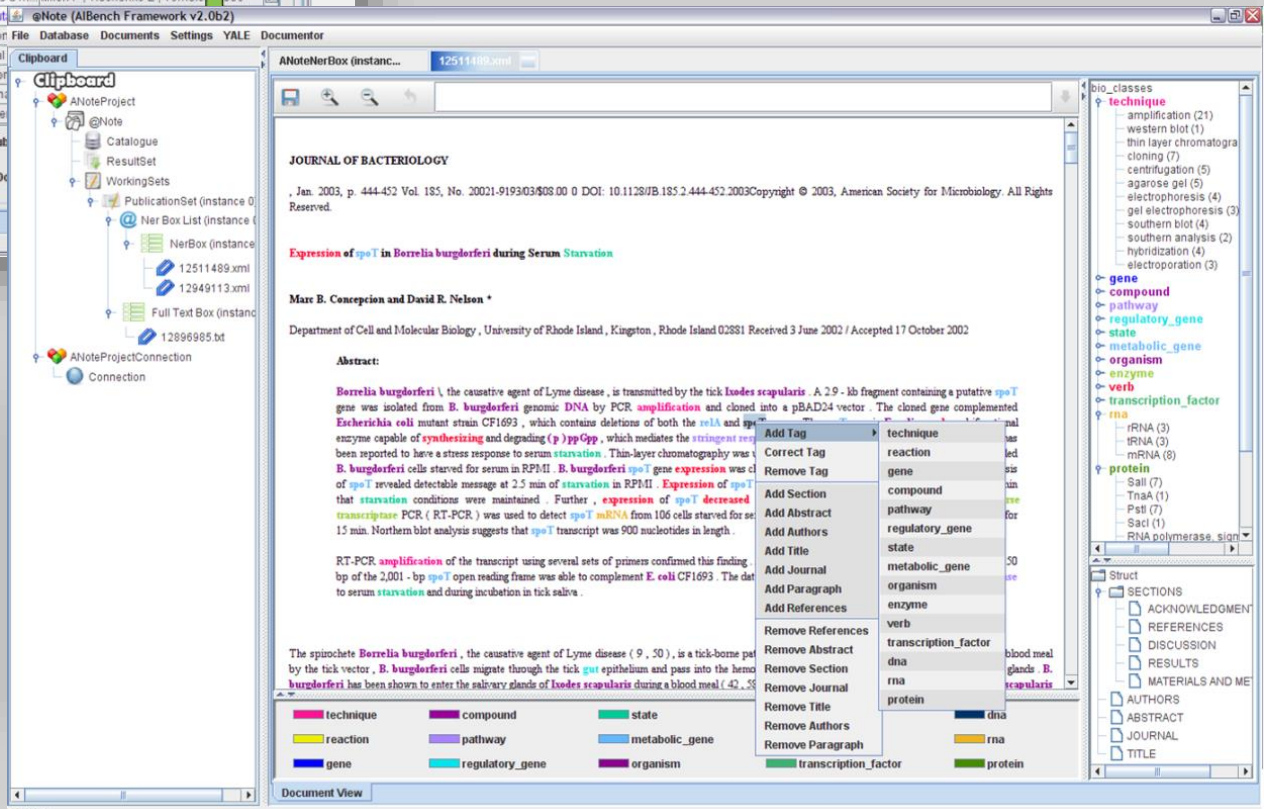
METABOLIC MODELS
TEXT MINING FOR AIDING RECONSTRUCTION

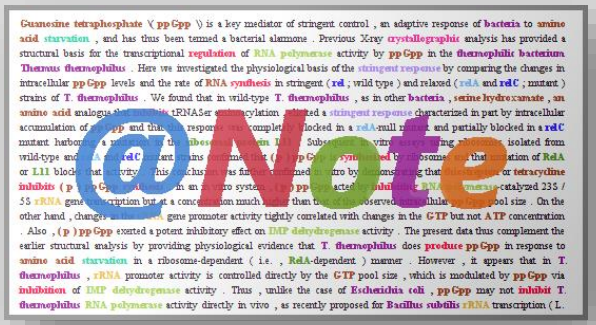


Tools for automatically inferring metabolic and regulatory networks from literature data - @**Note**

Lourenço et al. (2009),
J Biomedical Informatics
42, 710-720



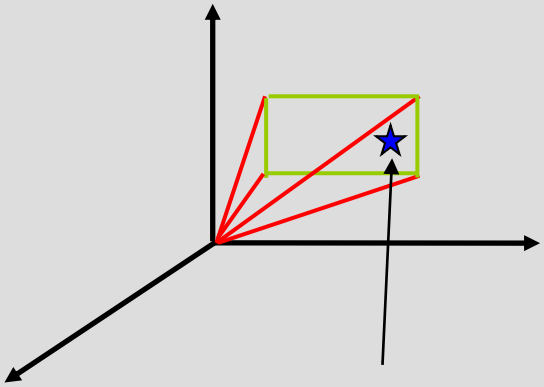




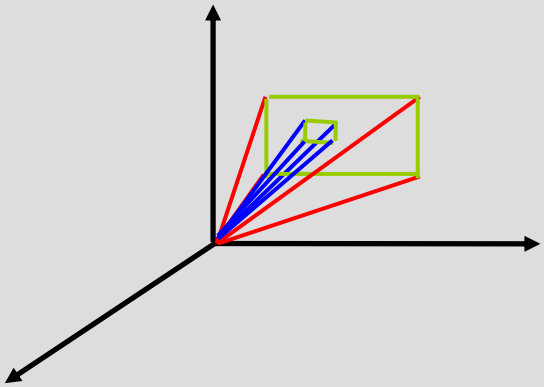
Ways to reduce the cone of solutions given by the stoichiometric model

- By optimizing a given criterion – FBA, MOMA, ROOM…
- By the introduction of regulatory information (ex: Gene Networks)

FBA: Flux Balance Analysis
ROOM: Regulatory On/Off *Minimization*
MOMA: Minimization of Metabolic Adjustment

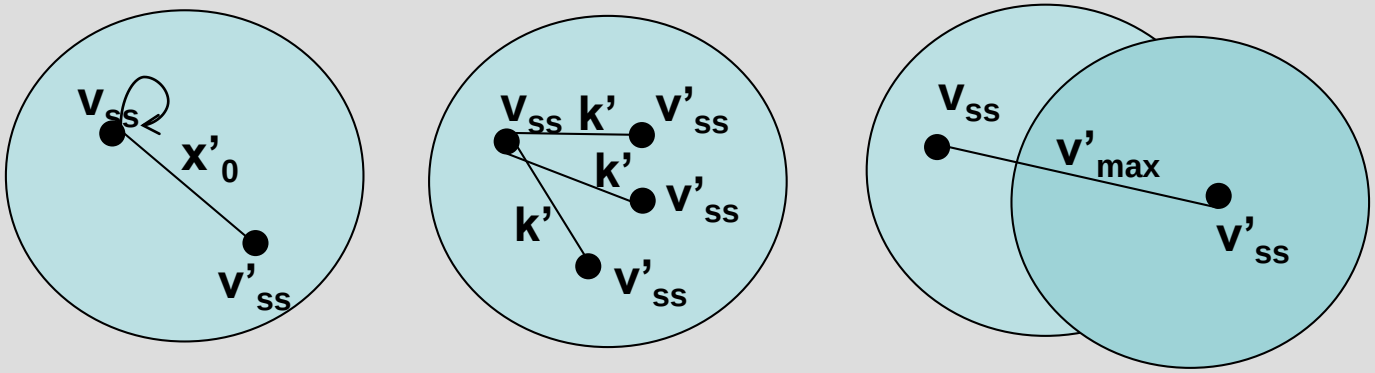


Particular Solution



Reduction of the solutions space

- Often, we have kinetic information for parts of the network
- What can we learn from the kinetics of those parts of the network that we know?
  - If reaction rates in the stoichiometric models are constrained by v_{max}, then the flux space given by the stoichiometry should be reachable by changing the kinetic parameters...
  - Or not?

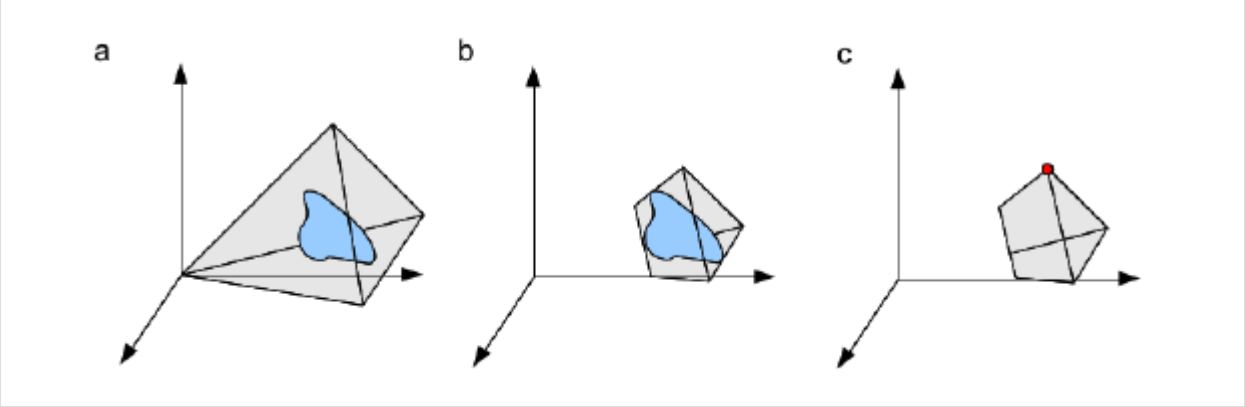


- Construction of a kinetically feasible flux cone:

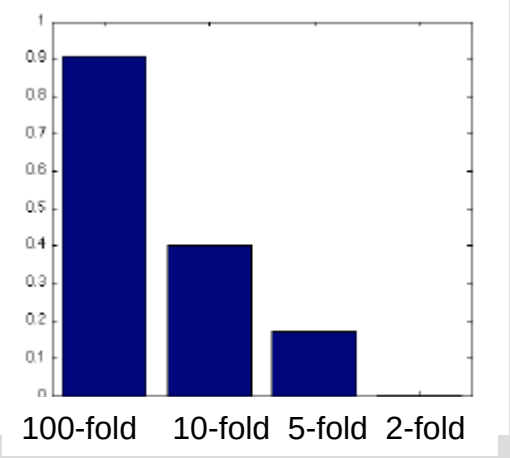
a) Limiting the range of the kinetic constants results in a smaller feasible space.

b) The flux cone can be adjusted to fit the feasible space.

c) Simulation methods such as FBA can use the reduced flux cone to search for optimal solutions.

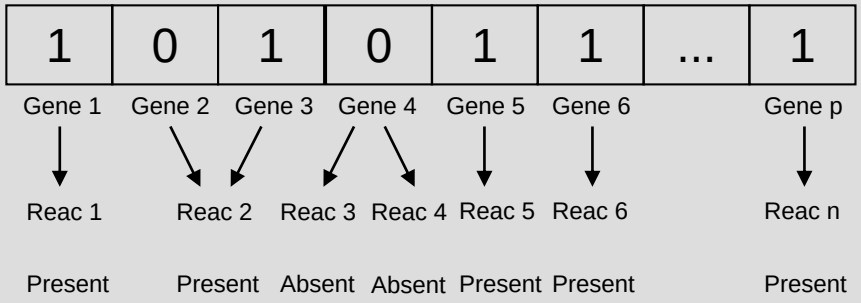


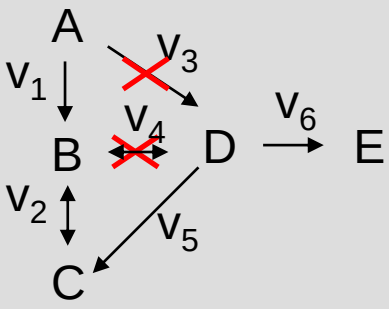
- Effect of constraining the range of variation of the kinetic constants of the dynamic model of the central carbon metabolism in the volume of the solution spaces



IN SILICO METABOLIC ENGINEERING
OPTGENE

Modified “Genome”





$$\begin{array}{c} v_1\ v_2\ v_3\ v_4\ v_5\ v_6\ \ldots\ flux n \\ \left[\begin{array}{cccccccc} -1 & 0 & -1 & 0 & 0 & 0 & \ldots & 0 \\ 1 & -1 & 0 & -1 & 0 & 0 & \ldots & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 1 & -1 & -1 & \ldots & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \end{array} \begin{array}{l} A \\ B \\ C \\ D \\ E \\ \ldots \\ metabolite m \end{array} \qquad \begin{array}{l} 0 \le v_1 \le +\infty \\ -\infty \le v_2 \le +\infty \\ 0 \le v_3 \le 0 \\ 0 \le v_4 \le 0 \\ 0 \le v_5 \le +\infty \\ 0 \le v_6 \le +\infty \\ \ldots \\ 0 \le v_n \le +\infty \end{array}$$

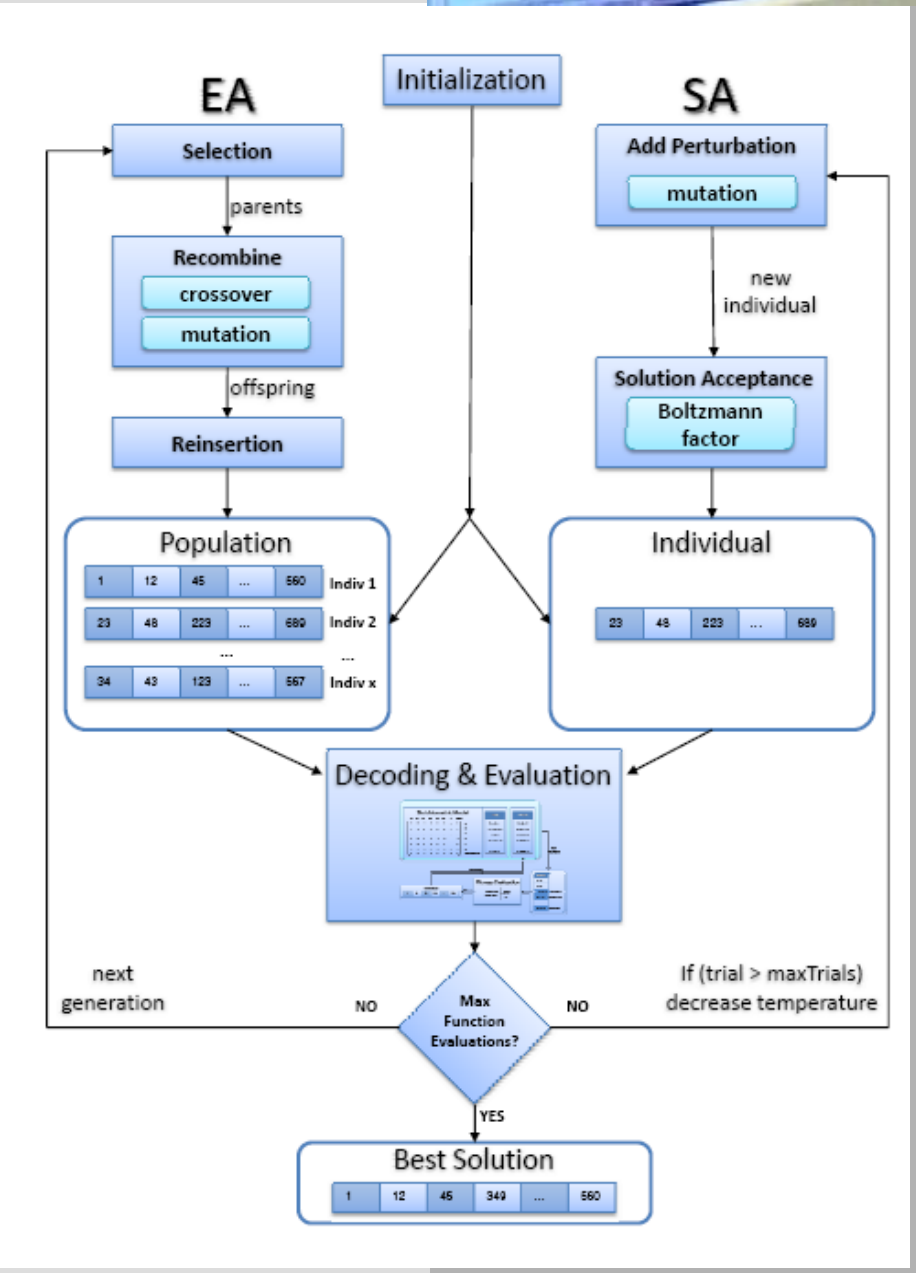
Modified reaction network

Modified metabolic model

Represetation of the metabolic genotype

Tools and Algorithms for Optimization (OptGene Algorithms):

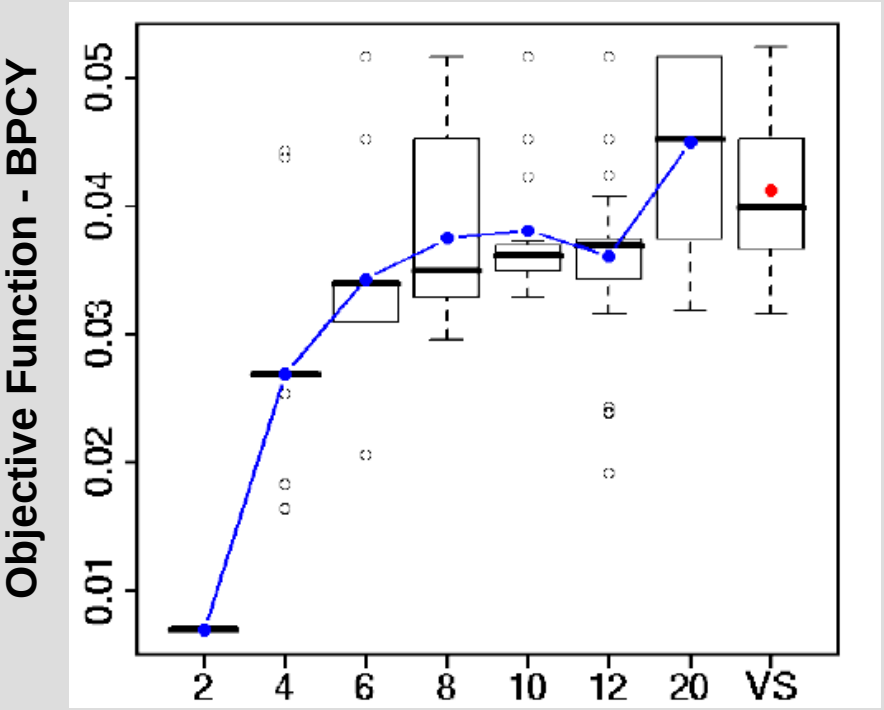
- EAs – Evolutionary Algorithms
- SA – Simulated Annealing algorithms
- Local Search



Patil et al (2005) BMC Bioinf 6

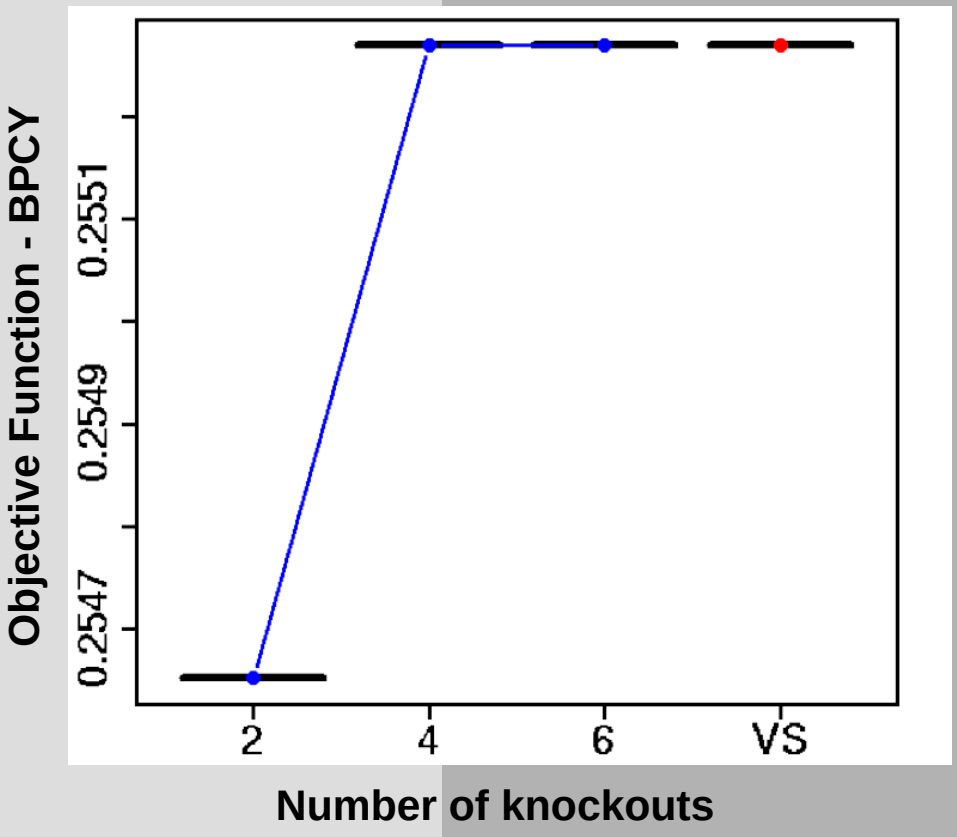
Rocha et al (2008) BMC Bioinf 9

S. cerevisiae, succinate, EA

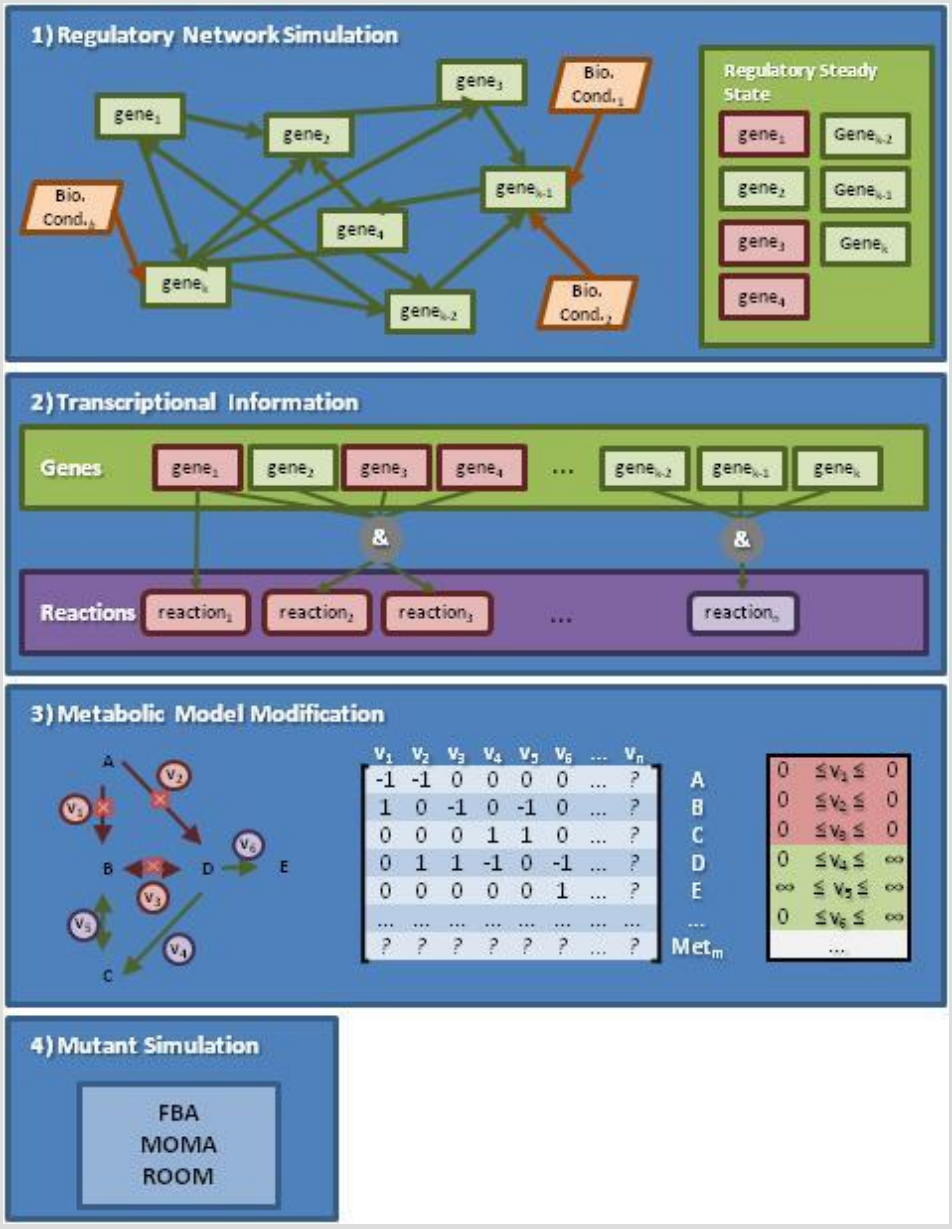


Rocha et al (2008) BMC Bioinf 9

E. coli, anaerobic lactate, SA



IN SILICO METABOLIC ENGINEERING
INCORPORATION OF REGULATORY MODELS



Results showing the differences obtained by deleting genes or reactions for lactate production with *E. coli*

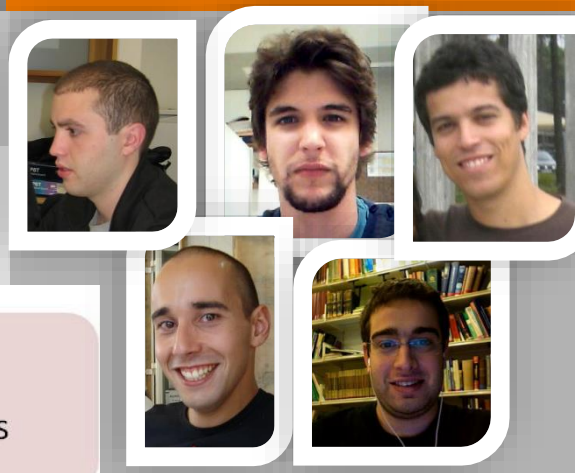
Optimization Type	Algorithm	Yield	Nr knock.
Reactions	SA	0,348	17
Genes	SA	0,293	12

Vilaça et al., Biosystems (in press)

IN SILICO METABOLIC ENGINEERING
SUCCESSFUL APPLICATIONS OF THE ALGORITHMS

- Production of Succinate with *S. cerevisiae*
- Production of Succinate with *E. coli*
- Production of Sesquiterpenes with *S. cerevisiae*
- Production of aminoacids with *E. coli* (ongoing)

IN SILICO METABOLIC ENGINEERING
OPTFLUX



Open-source
- allows all users to use the tool freely
- invites the contribution of other researchers

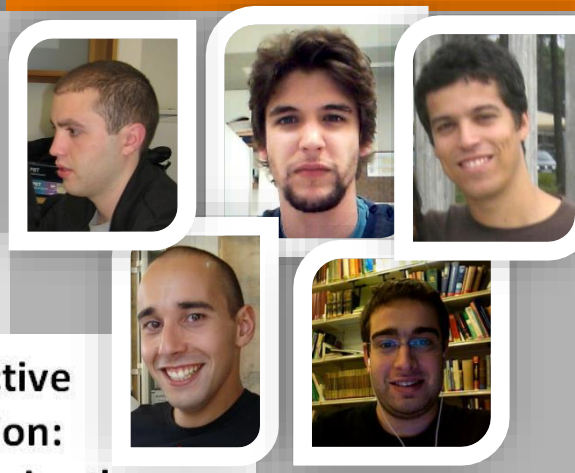
User-friendly
- facilitates its use by users with no/little background in modeling/informatics
- integrates a tool that allows the visualization of the metabolic models and results

Modular
- facilitates the addition of specific features by computer scientists / bioinformaticians
- based on the general-purpose AIBench platform

Compatible with standards
- SBML- Systems Biology Markup Language
- Cell Designer layouts

Rocha et al., BMC Systems Biology, 2010

IN SILICO METABOLIC ENGINEERING
OPTFLUX



Strain Optimization

- ✓ Evolutionary algorithms
- ✓ Simulated Annealing
- ✓ Local optimization

Objective function: Maximize the production of a given metabolite

Phenotype Simulation

- ✓ FBA, MOMA, ROOM
- ✓ Boolean net simulation
- ✓ …

Determine the set of fluxes (in steady state)

Environmental Conditions + **Genetic conditions (e.g. knockouts)**

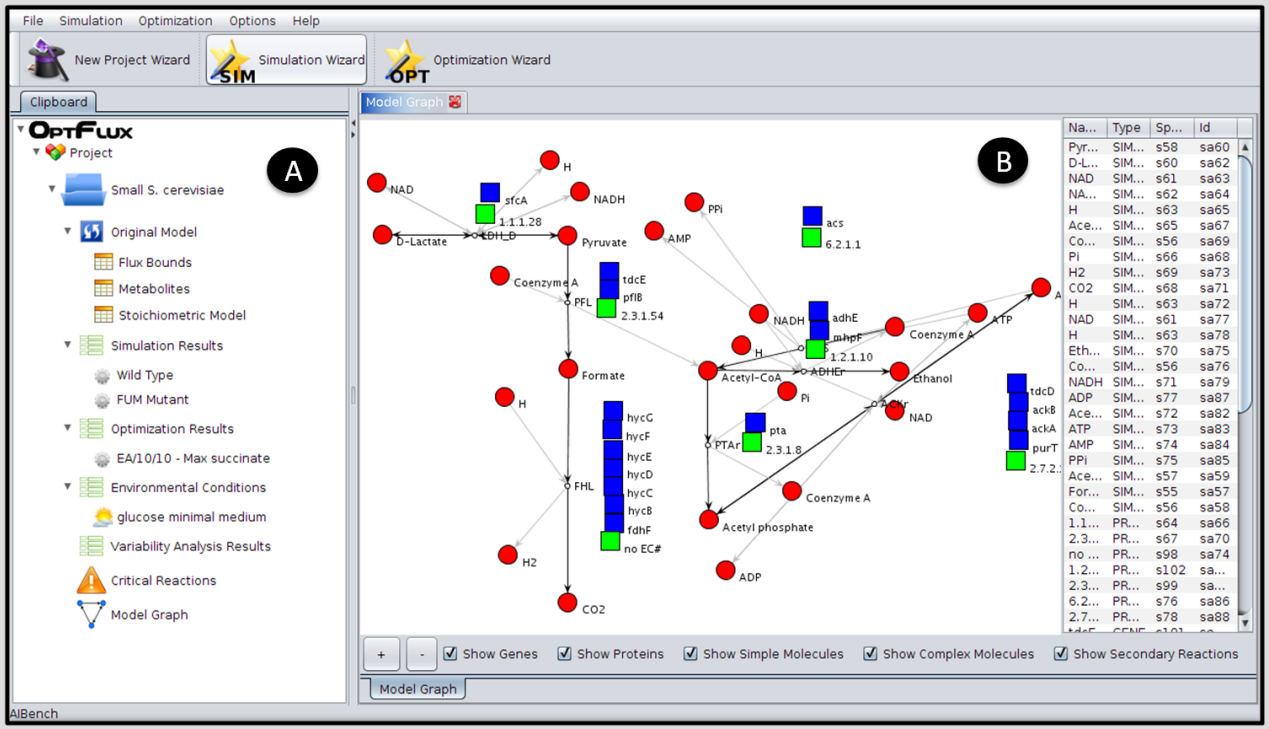
Override Model

Model

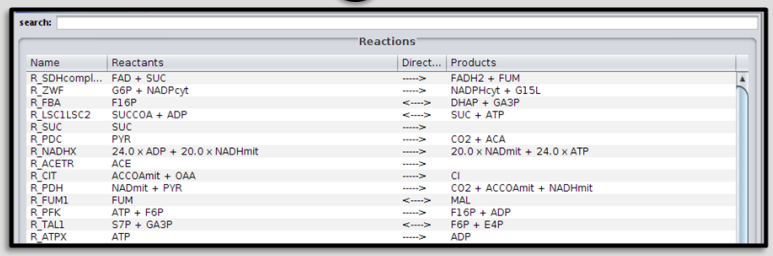
- ✓ Stoichiometric
- ✓ Regulatory
- ✓ …

Original Model

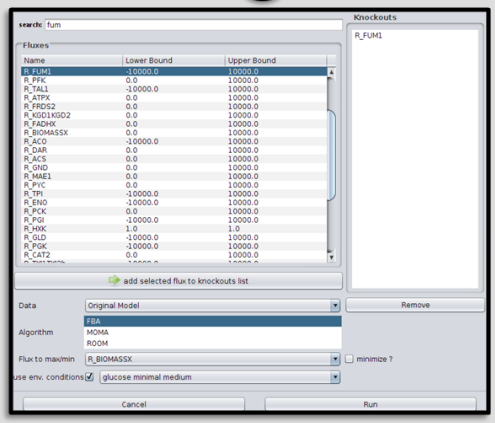
IN SILICO METABOLIC ENGINEERING
OPTFLUX



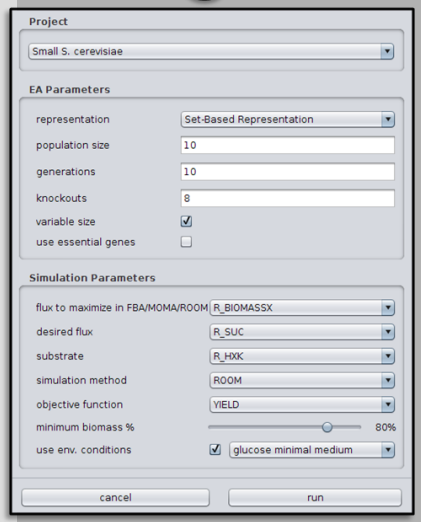
C



D



E



control variables
(Feeding rate)

Fermentation Process

State variables

measured
variables

feedback

+

−

estimated variables

$$F_{in}(t) = \frac{k_1 \mu_{set}(t)\hat{X}(t)W(t)}{S_{in}} - \tau_u \frac{\hat{\mu}(t) - \mu_{set}}{S_{in}} W(t)$$

Control Law

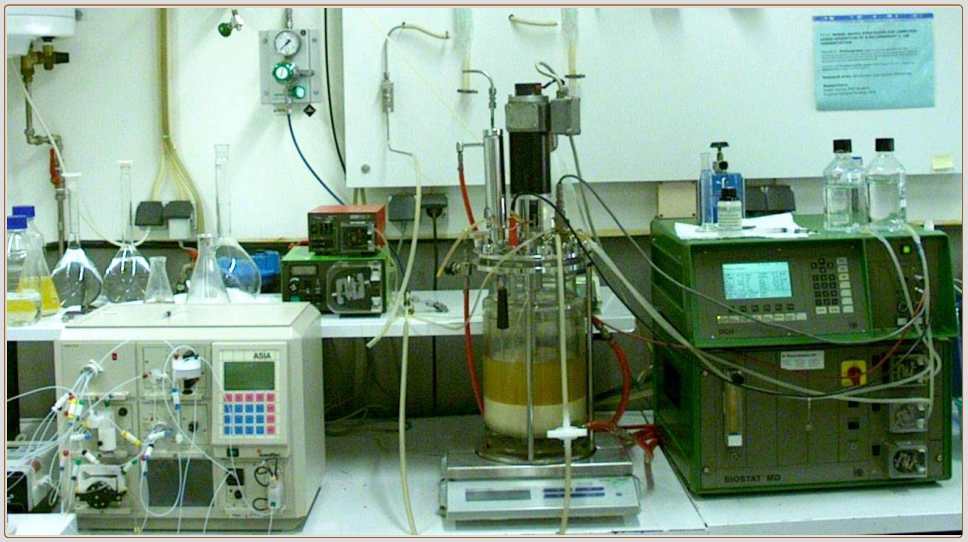
$$\frac{d\hat{\xi}}{dt} = Kr\left(\hat{\xi},t\right) - D\hat{\xi} + F - Q + \Omega\left(\hat{\xi}\right)\left[\xi_1 - \hat{\xi}_1\right]$$

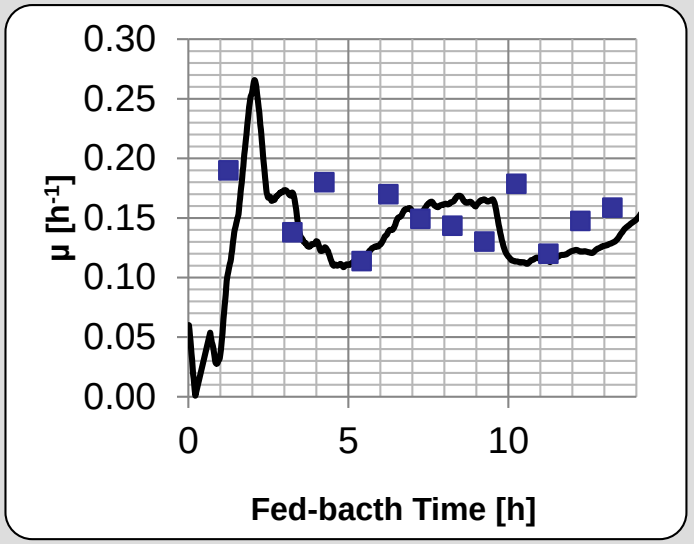
OBSERVER
AND ESTIMATORS

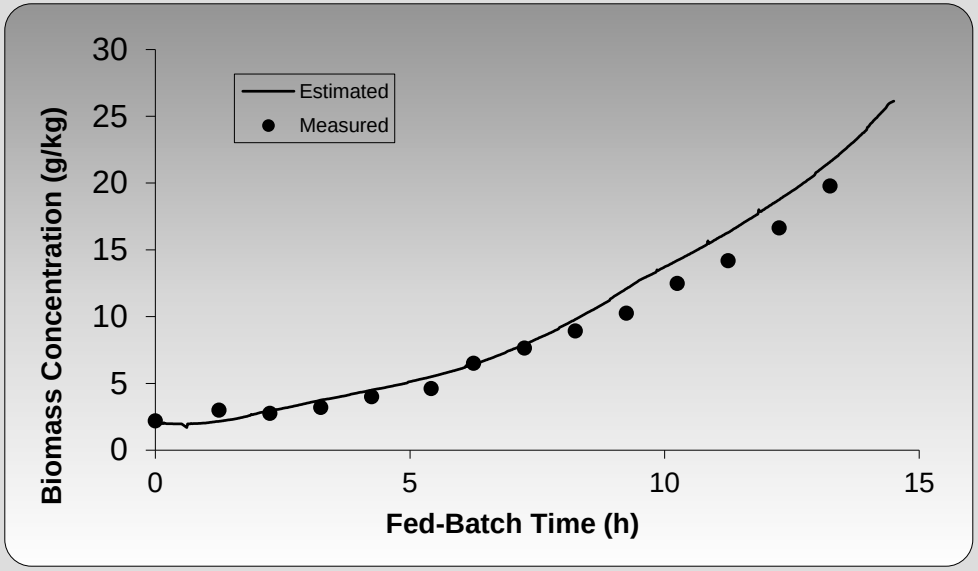
$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ A \\ O \\ C_T \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & 0 \\ 0 & k_3 & -k_4 \\ -k_5 & -k_6 & -k_7 \\ k_8 & k_9 & k_{10} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} X - D \begin{bmatrix} X \\ S \\ A \\ O \\ C_T \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{F_{in,S}}{W}\right) S_{in} \\ 0 \\ OTR \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ CTR \end{bmatrix}$$

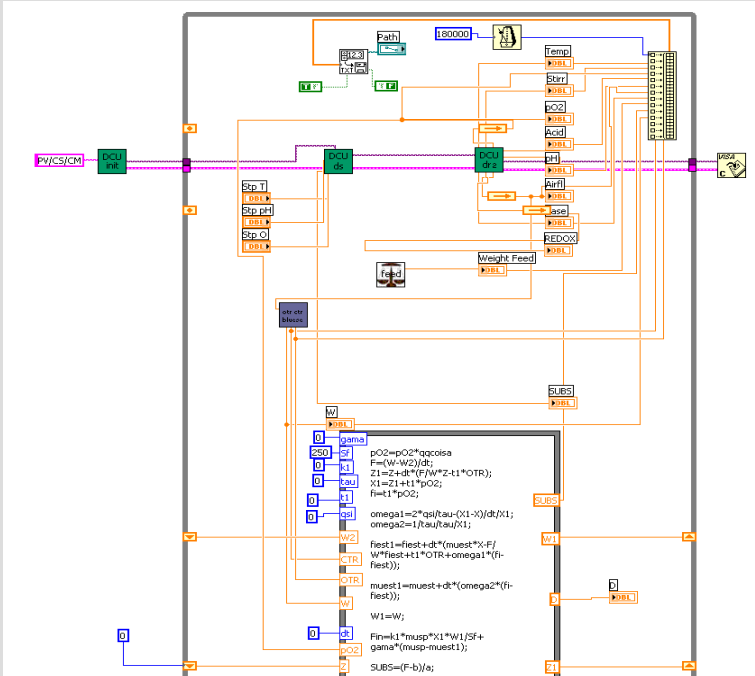
Process Model

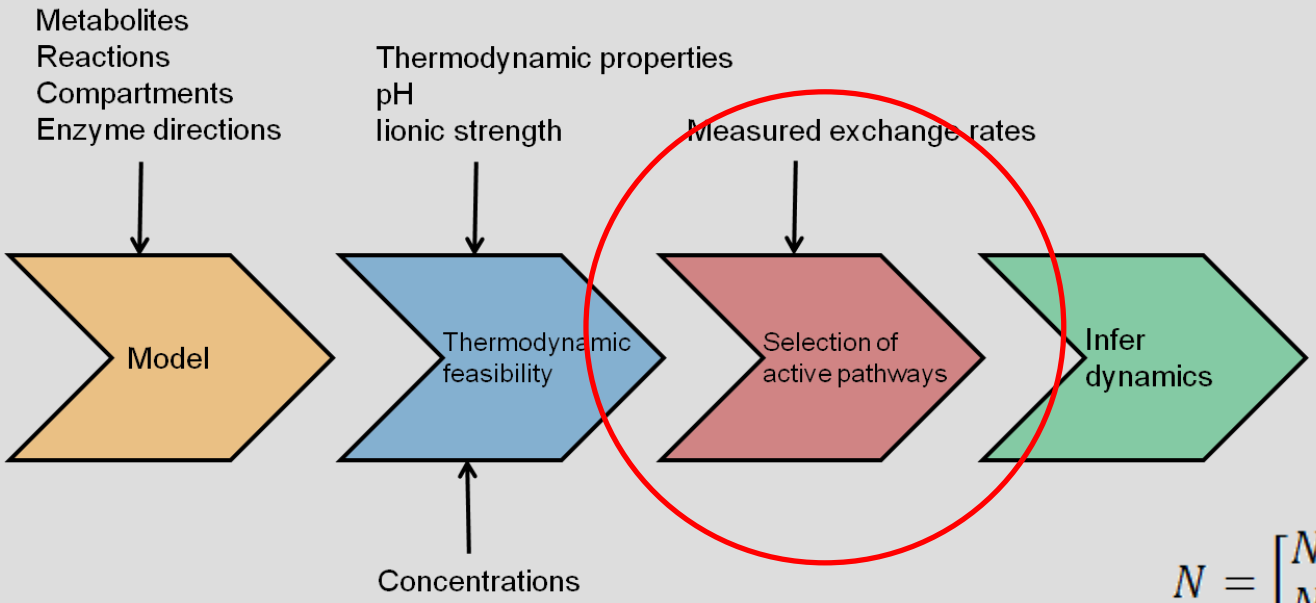
PROCESS ENGINEERING
MODEL-BASED CONTROL











Soons et al, JPC, 2011

Selection based on objective function:

$J = RMSE + c_1 \cdot P + c_2 \cdot MS$

weighted sum of squared errors

penalty for inefficiency of each elementary mode

penalty for model size

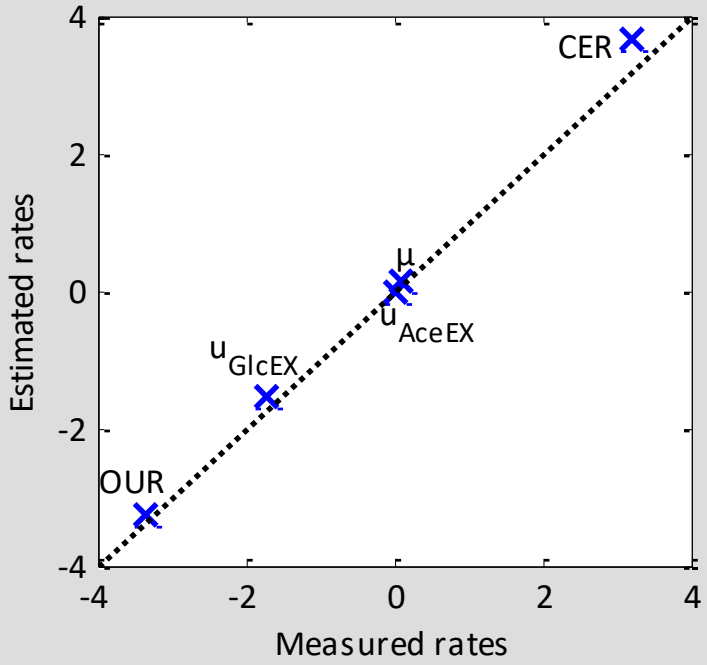
$$N = \begin{bmatrix} N_s \\ N_\xi \end{bmatrix}$$

$$\frac{ds}{dt} \cong 0 \implies \begin{bmatrix} N_s \\ N_\xi \end{bmatrix} v - \begin{bmatrix} 0 \\ v_m \end{bmatrix} = 0$$

$$\frac{d\xi}{dt} = Kr(t) + u(t)$$

$$K = N_\xi \cdot E$$

- Reduction of the original model from 2706 to 3 elementary modes:
  - EM1 : 0.21 GlcEX + 0.67 O2 → 0.015 Biomass + 0.67 CO2
  - EM2 : 0.056 GlcEX + 0.22 O2 → 0.056 Acetate + 0.22 CO2
  - EM3 : 0.040 GlcEX + 0.24 O2 → 0.024 CO2
- Good match with measured data from literature



- Initially: 2706 elementary modes

Soons et al, JPC, 2011

CONCLUSIONS

- So far, rational metabolic engineering design has only been performed with **stoichiometric models** and indicate only **knockout and gene additions**
- Nevertheless, it is already possible to **improve *in silico* the production of targeted compounds**
- Predictions are enhanced if **regulatory information** is added to the models
- Additional constrains maybe derived if **kinetic information** is available for part of the network
- Model reconstruction is far away from being made in a standard way
- The bridge between process engineering and optimization and large scale models is still not there...

ACKNOWLEDGMENTS

Funding Agencies

Portugal - FCT
EC – SysInBio CA
Dupont – Science & Engineering Grant
Dupont – R&D contract
MIT-Portugal Program

International Contributors

Kiran Patil – EMBL, Germany
Jens Nielsen – Chalmers – Sweden
Bruce Tidor – MIT
Silas Villas-Boas – Univ. of Auckland, New Zealand
Andreas Gombert – University of São Paulo, Brazil
Jean-Francois Tomb and **Marcellinus Pont** – Dupont, USA

Contributors – IBB – UMinho

Isabel Rocha(PI)
Anália Lourenço (PI)
Zita Soons
Sonia Carneiro
Carla Portela
Óscar Dias
Daniel Machado
Paulo Maia
Pedro Tiago
Paulo Vilaça
Rafael Carreira
Rafael Costa
Hugo Costa
Simão Soares
Marisa Cunha

Contributors – CCTC– UMinho

Miguel Rocha (PI)
José Pedro Pinto