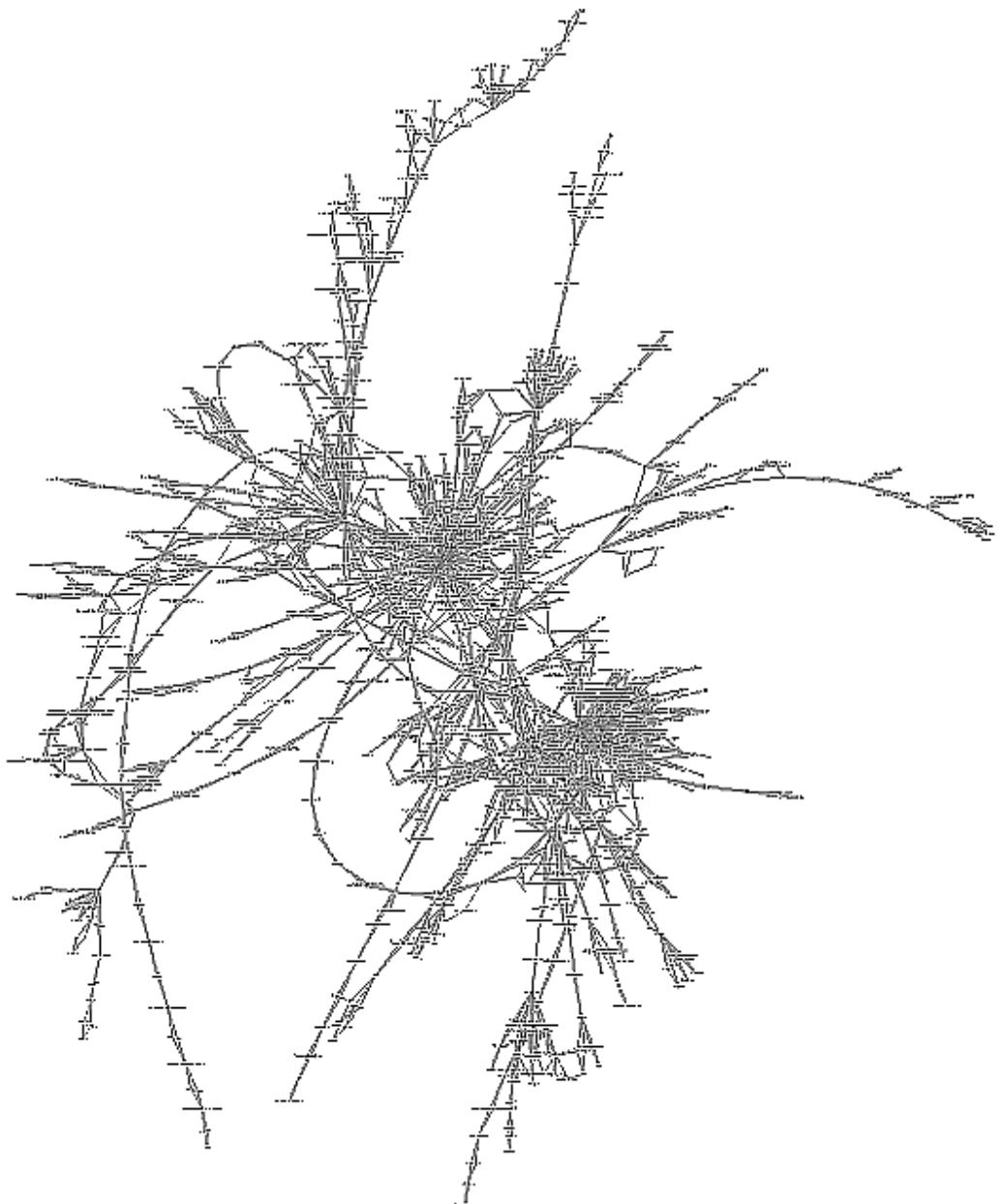


# Metabolic network reconstruction and flux balance analysis

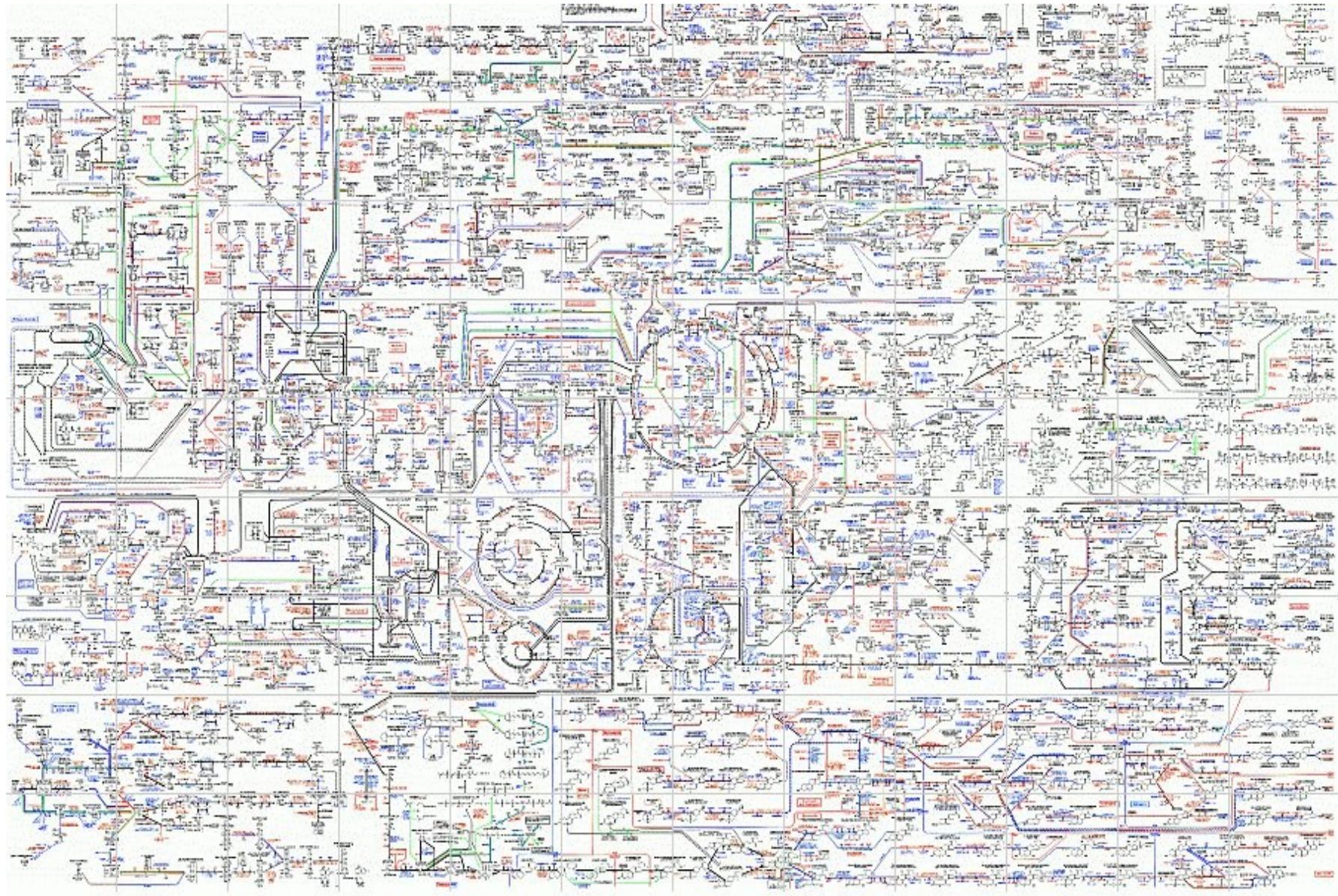
**Ludovic Cottret**  
**[ludovic.cottret@toulouse.inra.fr](mailto:ludovic.cottret@toulouse.inra.fr)**

# M1 MABS Toulouse

## Mars 2013



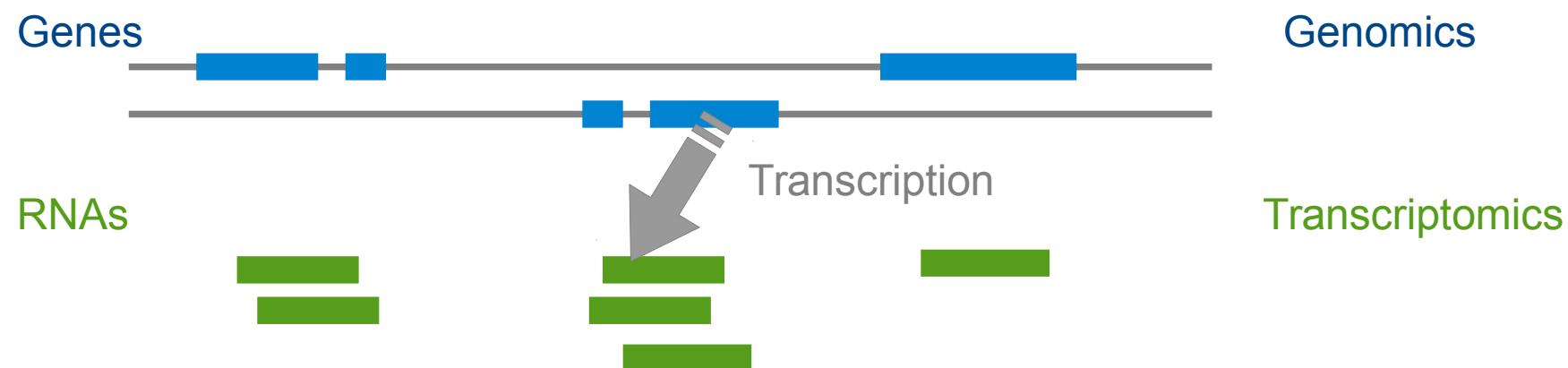
# What is a metabolic network ?



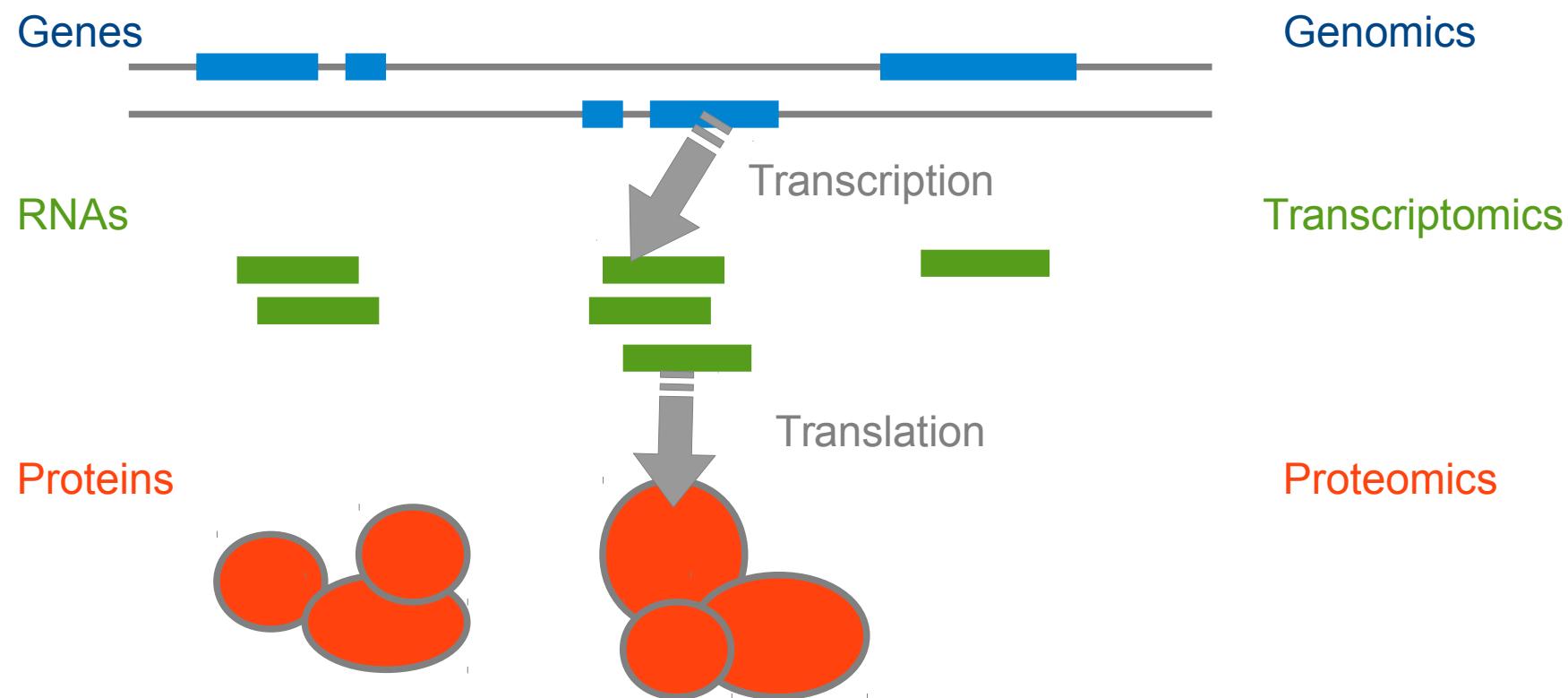
# From genes to reactions



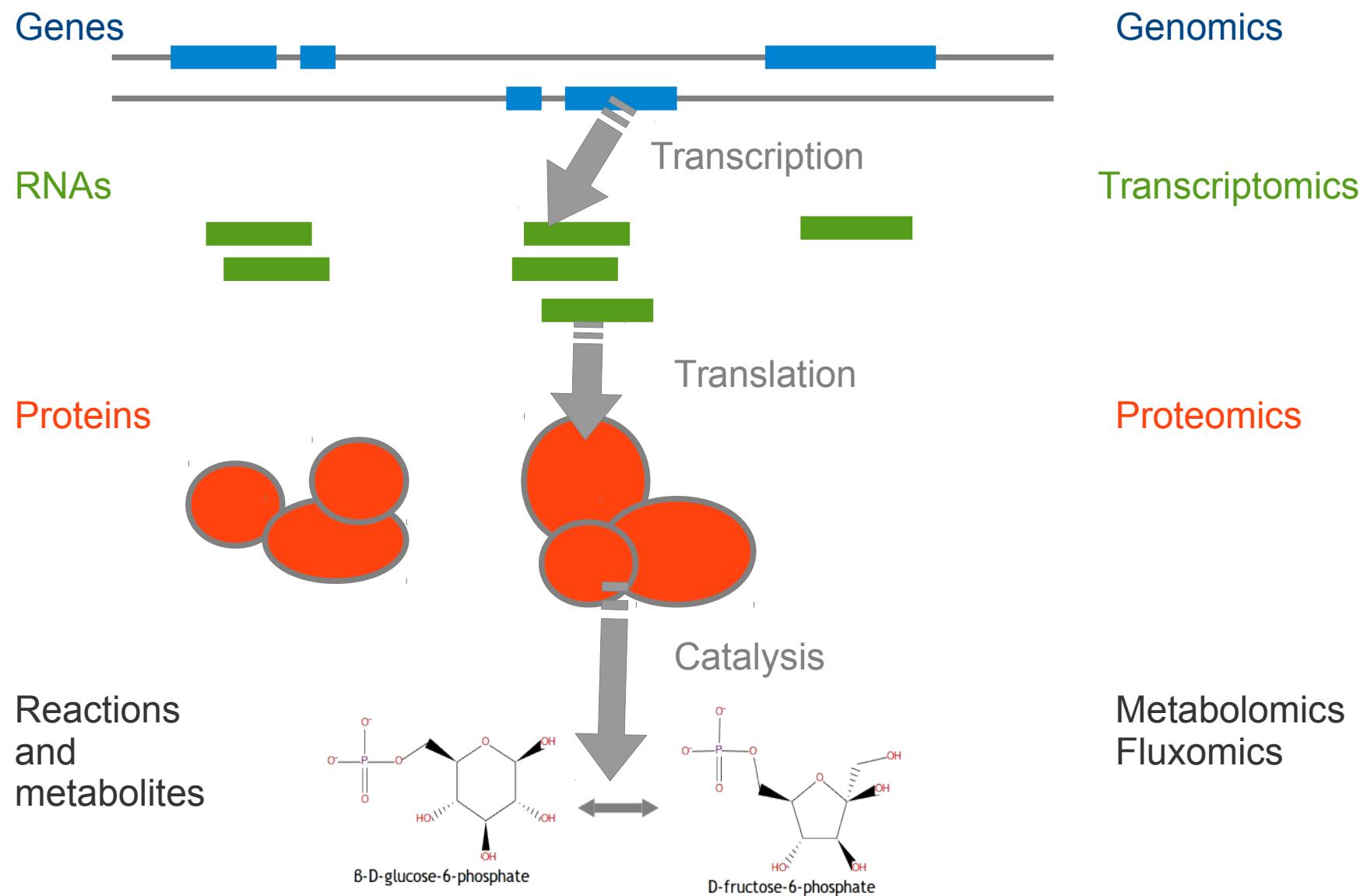
# From genes to reactions



# From genes to reactions



# From genes to reactions



# From genes to reactions

Genes

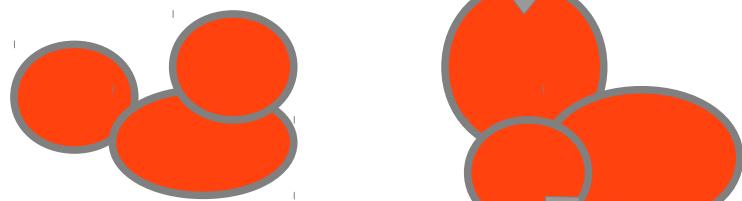


RNAs



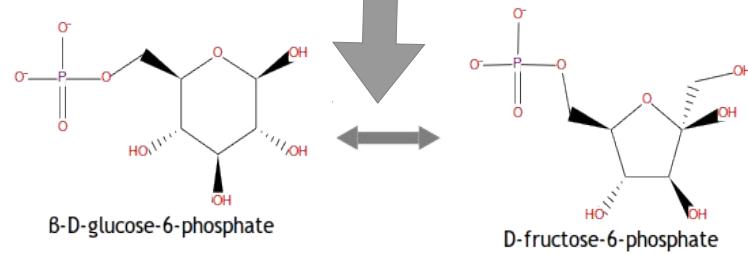
Transcription

Proteins



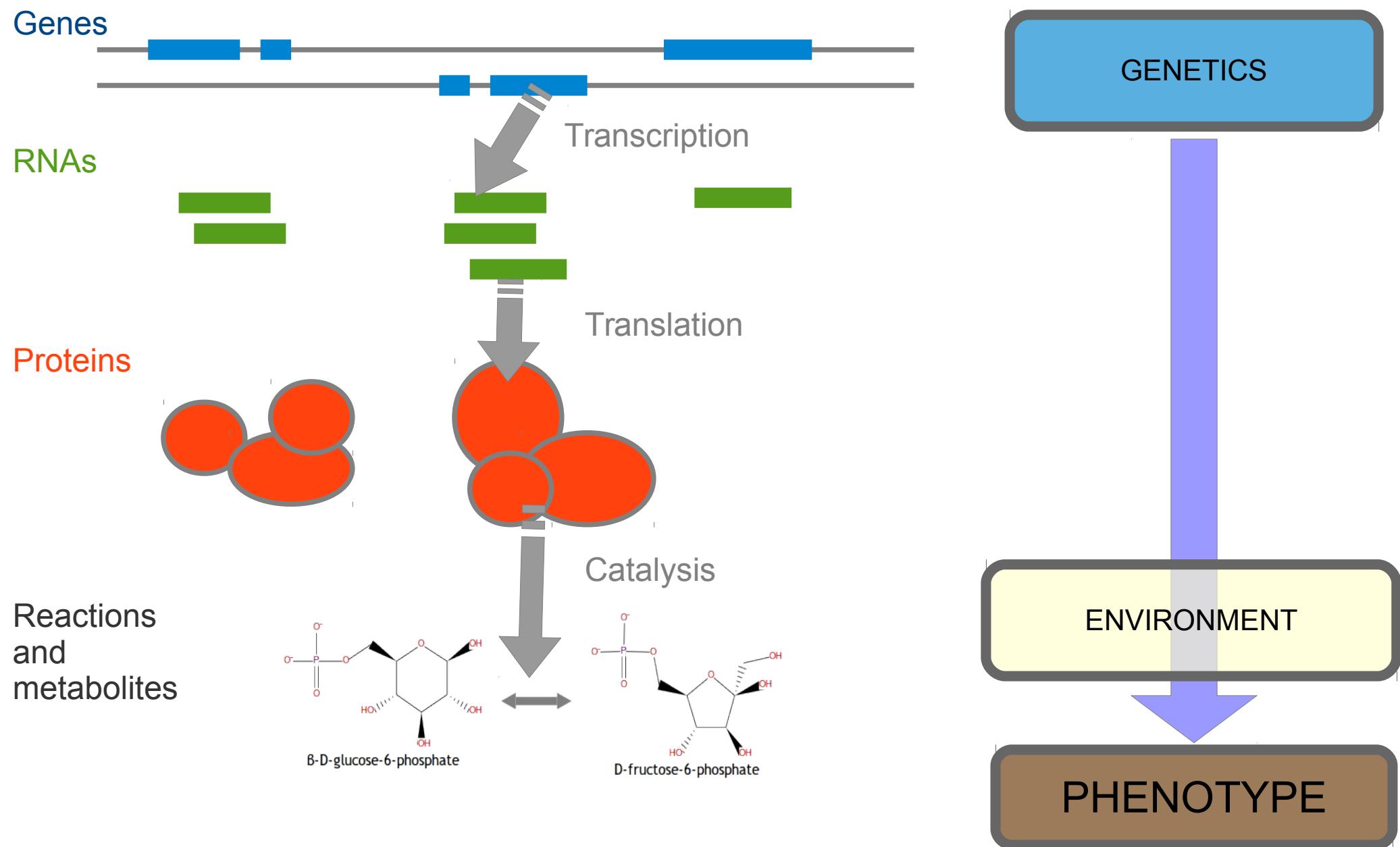
Translation

Reactions  
and  
metabolites

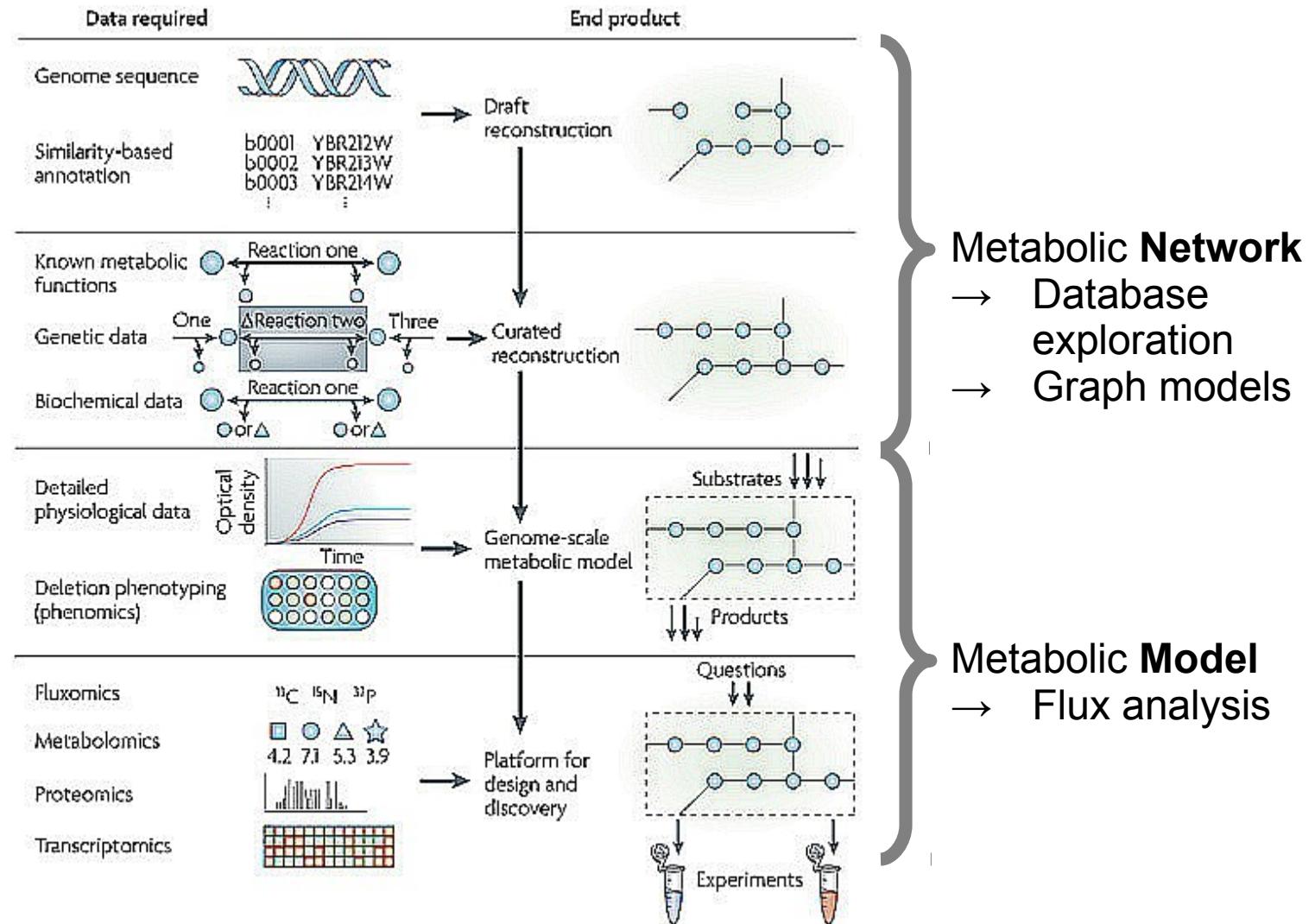


PHENOTYPE

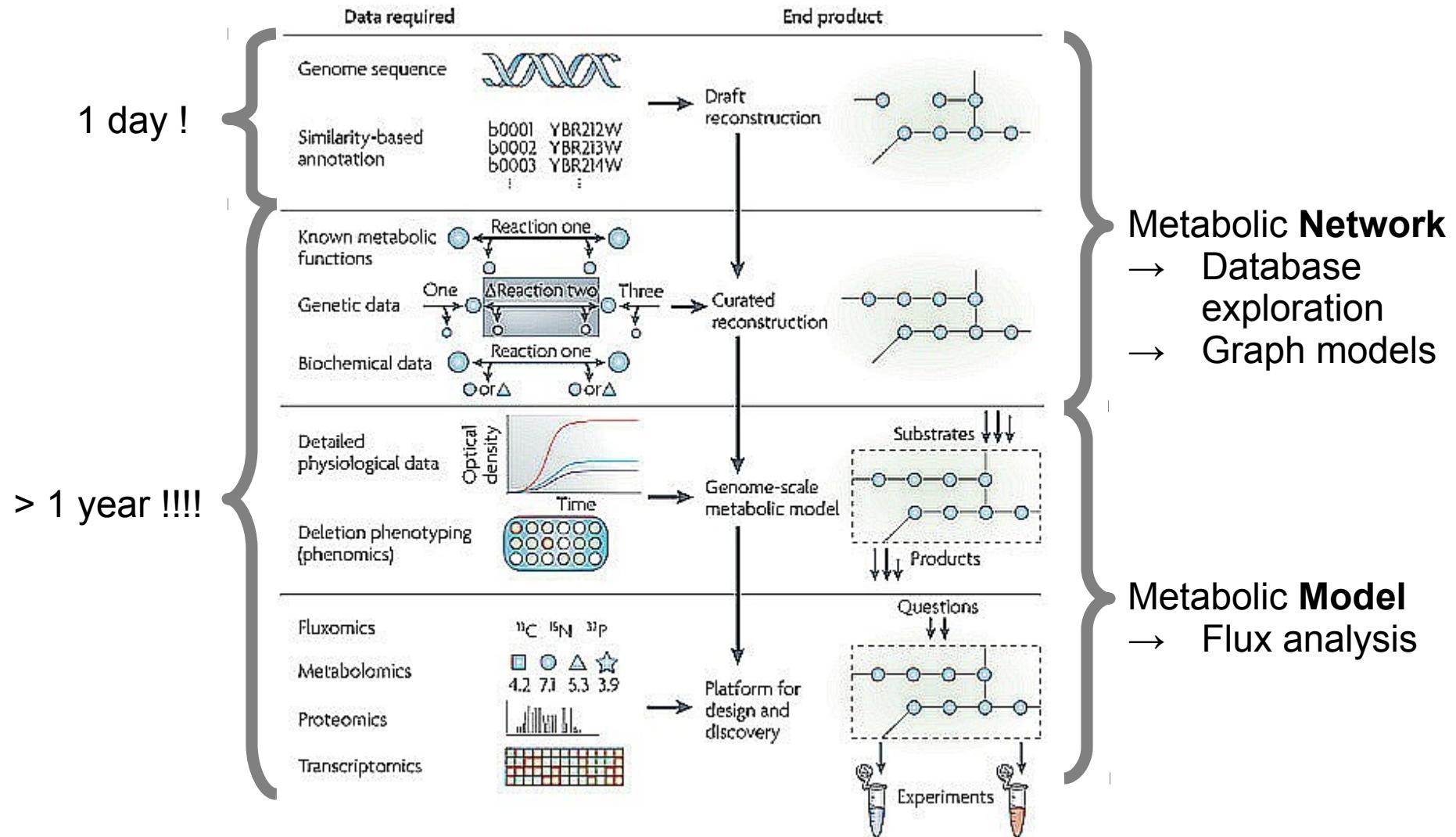
# From genes to reactions



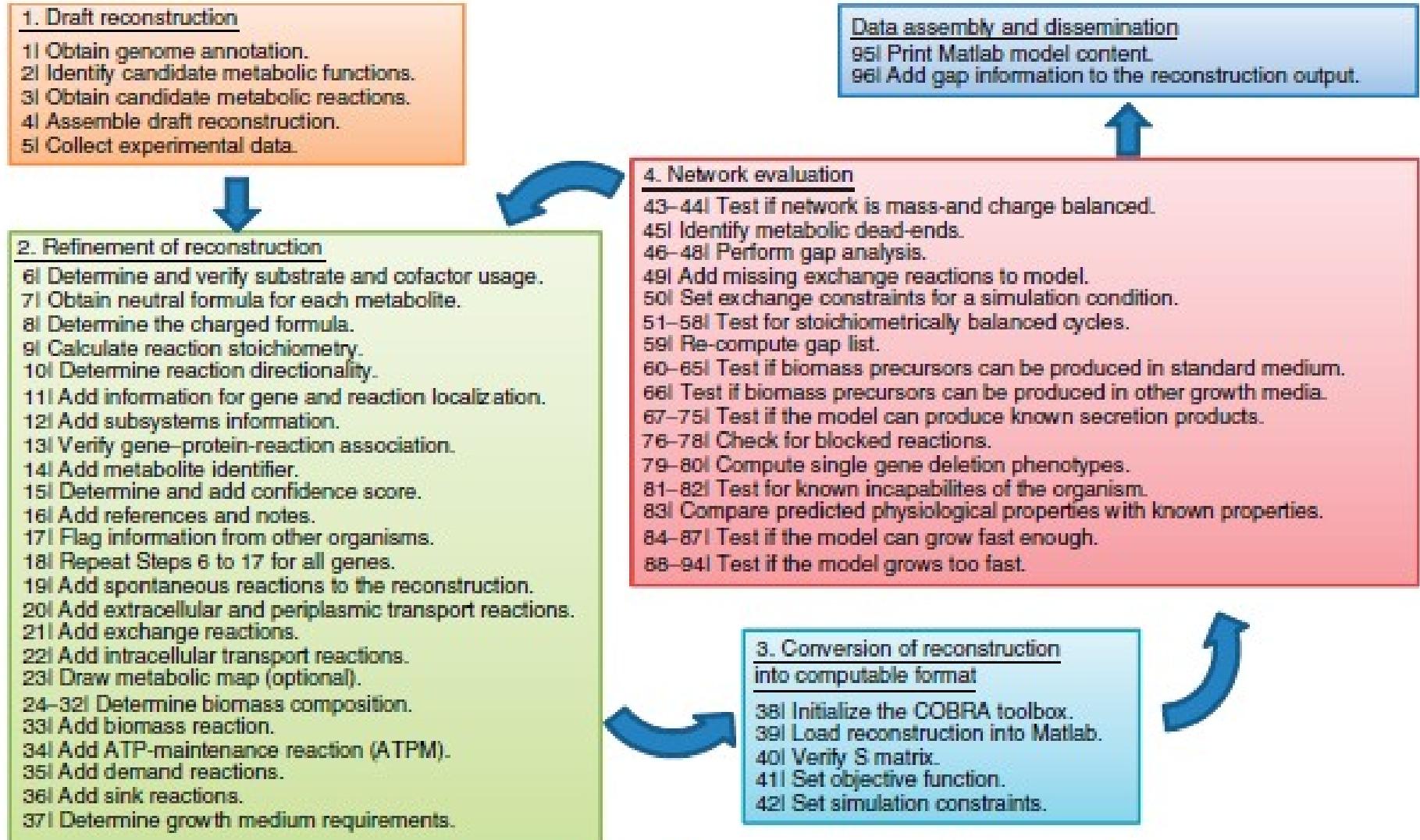
# Metabolic network and model reconstruction



# Metabolic network and model reconstruction



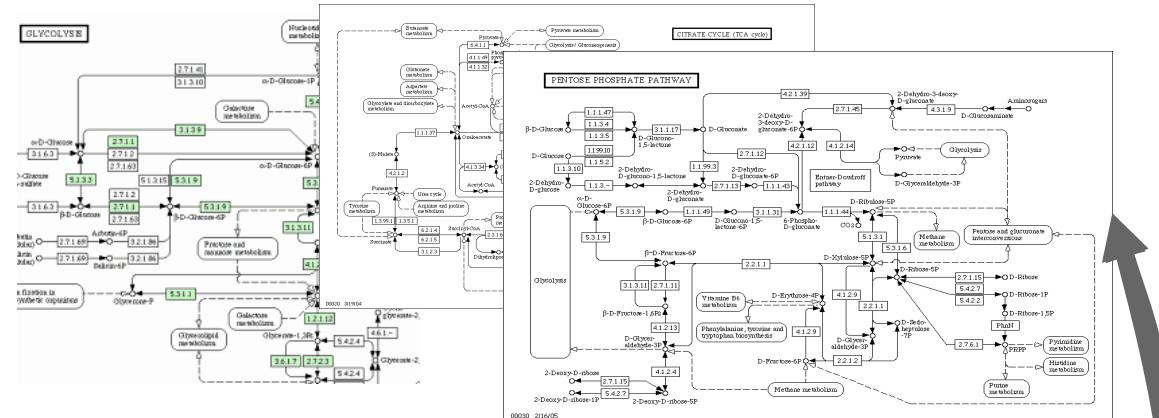
# A long task...



First draft reconstruction

# Automatic draft reconstruction

>gi|49175990|ref|  
 AGCTTTCAATTCTGACTC  
 TTCTGAACGGTTACCTC  
 TATAGGCATAGGCACAC  
 ATTACCACCACCATCAC  
 CCCGACCTGACAGTGC  
 GTTCGGCGGTACATCAG  
 AGGCAGGGGCAGGTGGC  
 AAAAAACCATAGCGGC  
 GATCCCCACTCCCTCCCT



## ANNOTATION

gene

```

/db_xref="taxon:38
1..1317
/locus_tag="CRP_00
/db_xref="GeneID:4
1..1317
/locus_tag="CRP_00
/codon_start=1
/transl_table=11
/product="tRNA_mod

```

CDS

**List of metabolic functions**  
 EC number: X.X.X.X  
 Name of the enzymatic activity

## Metabolic pathways

## Reactions

[reduced flavodoxin + a ribonucleoside triphosphate = a nucleoside diphosphate + H<sub>2</sub>O → phosphate + a nu](#)  
[ATP + a 1,2-diacylglycerol = ADP + an L-phosphatida](#)  
[a ribonucleoside diphosphate + ATP → a ribonucleosi](#)  
[ATP = cyclic-AMP + diphosphate](#),  
[ATP + a fatty acid + acyl carrier protein = AMP + di](#)



BIOCYC



# Point de départ : l'annotation fonctionnelle du génome



*“Ce qui est vrai pour le colibacille est vrai pour l’éléphant”*

# *Jacques Monod*

L'annotation fonctionnelle se fait principalement par comparaison avec les annotations d'organismes modèles.

# Rappels : homologie, orthologie, paralogie et analogie

- Les gènes **homologues** ont un ancêtre commun
- Les gènes **orthologues** sont des homologues qui ont divergé après un évènement de spéciation
- Les gènes **paralogues** sont des homologues qui ont divergé après un évènement de duplication
- Les gènes **analogues** ne sont pas homologues mais ont la même fonction

# Assignements par homologie : hypothèses

- Deux gènes (protéines) dont la séquence est proche sont susceptibles de coder pour la même fonction
- Les orthologues sont supposés avoir la même fonction
- La fonction des paralogues est supposé avoir divergé

# Annotation fonctionnelle des gènes

## Limitations

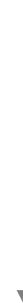
- Les orthologues n'ont pas toujours la même fonction
- Certains paralogues récents peuvent partager la même fonction
- Les groupes d'orthologues de référence sont le plus souvent inférés automatiquement : propagation d'erreurs
- Impossible de détecter les analogues
- 40 % des gènes n'ont pas de fonction
- 40 % des enzymes n'ont pas de gène assigné

# Informations additionnelles

- Profiles phylogénétiques
- Synténie
- Fusion de protéines
- Corégulation
- Pour les gènes métaboliques : analyse du réseau métabolique
- ...

# Des annotations aux réactions

Listes de gènes métaboliques



?

Listes de  
réactions

# Les bases de données pour faire le lien entre EC numbers et réactions

- ENZYME
- BRENDA
- KEGG
- BioCyc, MetaCyc
- Rhea

# Définir les voies métaboliques à partir des réactions

- Par comparaison avec les voies métaboliques de référence :
  - KEGG : en allumant les gènes dans les cartes métaboliques
  - Pathway-Tools (BioCyc) : en établissant des règles pour définir la présence ou l'absence d'une voie métabolique
- Par analyse du réseau métabolique
  - Chemins dans les graphes
  - Modes élémentaires

# Outils de reconstruction automatique du réseau métabolique

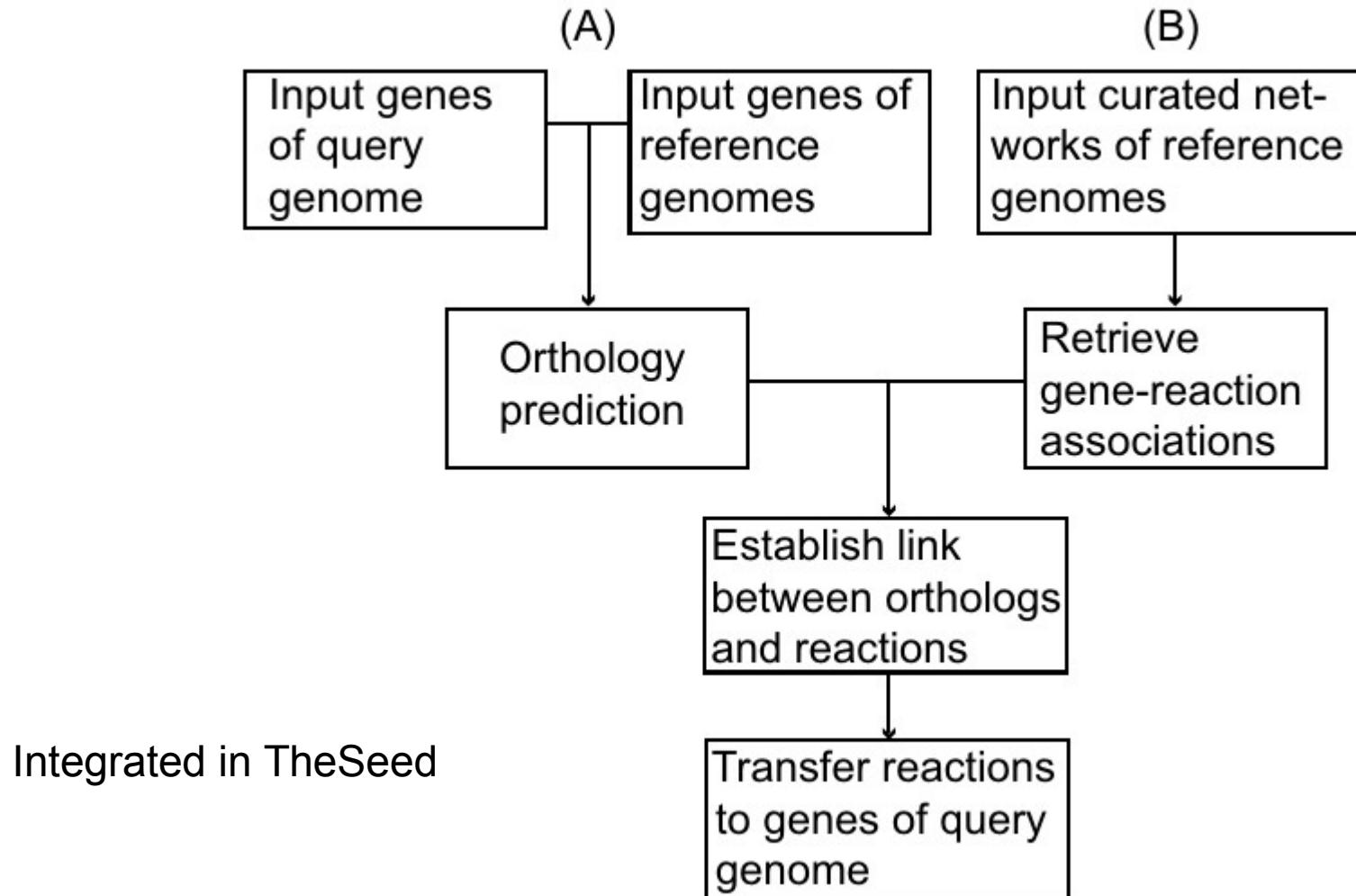
- KAAS (KEGG) : à partir de séquences brutes, classe chaque gène dans un groupe d'orthologues et retourne les cartes métaboliques avec les réactions identifiées
- Pathologic (BioCyc) : à partir d'annotations existantes, retourne une liste de voies métaboliques
- Model Seed : à partir de séquences brutes, retourne une liste de scénarios possibles et un réseau métabolique global
- PRIAM : à partir de séquences brutes, retourne une liste de numéros EC potentiels et les cartes métaboliques KEGG avec les réactions identifiées

# Limitations des reconstructions automatiques

- Erreurs provenant de l'annotation du génome
- Spécificité des enzymes (un même EC peut correspondre à plusieurs réactions)
- Réactions manquantes
- Réactions génériques (ex: 1.1.1.1 : an alcohol + NAD+ → an aldehyde or ketone + NADH)
- Direction des réactions non assignée
- Compartimentation
- Erreur dans la formulation des réactions
- Prédiction incorrecte des voies métaboliques

How to obtain a better first draft ?

# Propagation from curated metabolic models



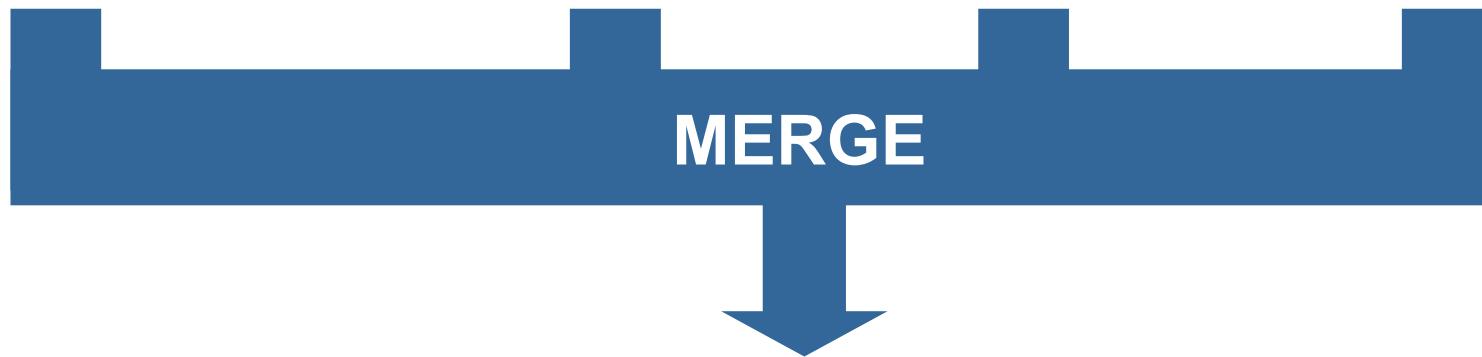
# Merge several draft reconstructions



**BIOCYC**  
<http://biocyc.org/>

**PRIAM**

**The SEED**



Draft reconstruction

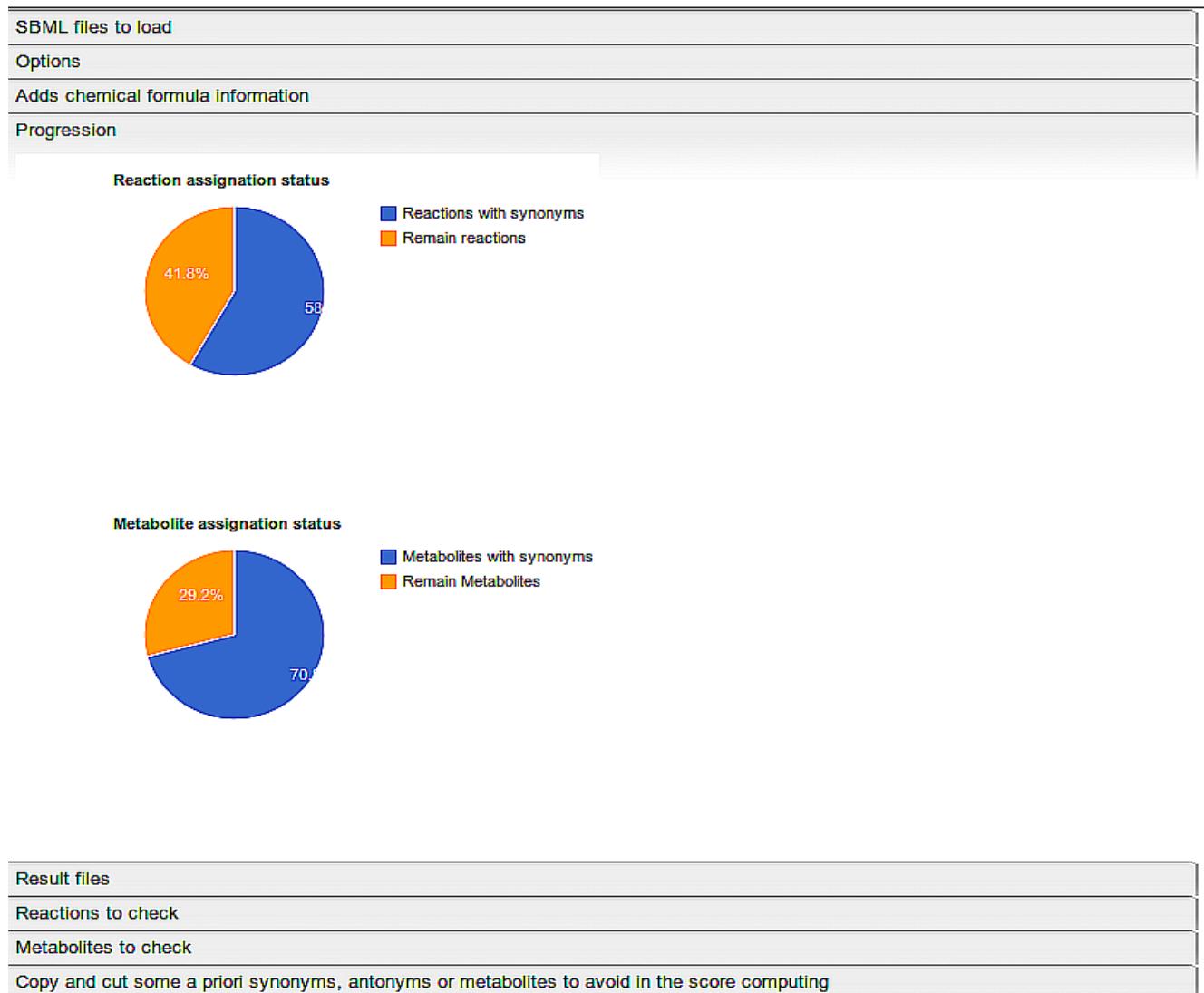
# Reconciliation of metabolic annotations

- **MetRxn** : incorporates data from 8 different metabolic databases and 90 genome-scale metabolic models
- **Rhea** : freely available, manually annotated database of chemical reactions created in collaboration with the Swiss Institute of Bioinformatics (SIB)
- **Chebi** : The database and ontology of Chemical Entities of Biological Interest

# Example of bioinformatics help to reconcile networks

			Reaction in the SBML file to standardize	Reaction in the standard SBML file	Score	F
●	●	U	R_GLXt2	R_GALCTNLT2pp	77.0	R N N
●	●	U	R_GLXt2	R_RATn	50.0	R N N
●	●	U	R_GLXt2	R_MSO3abcpp	65.0	R N N
●	●	U	R_GLXt2	R_ATPtp_H	65.0	R N N

# Example of bioinformatics help to reconcile networks



# Metabolic network curation

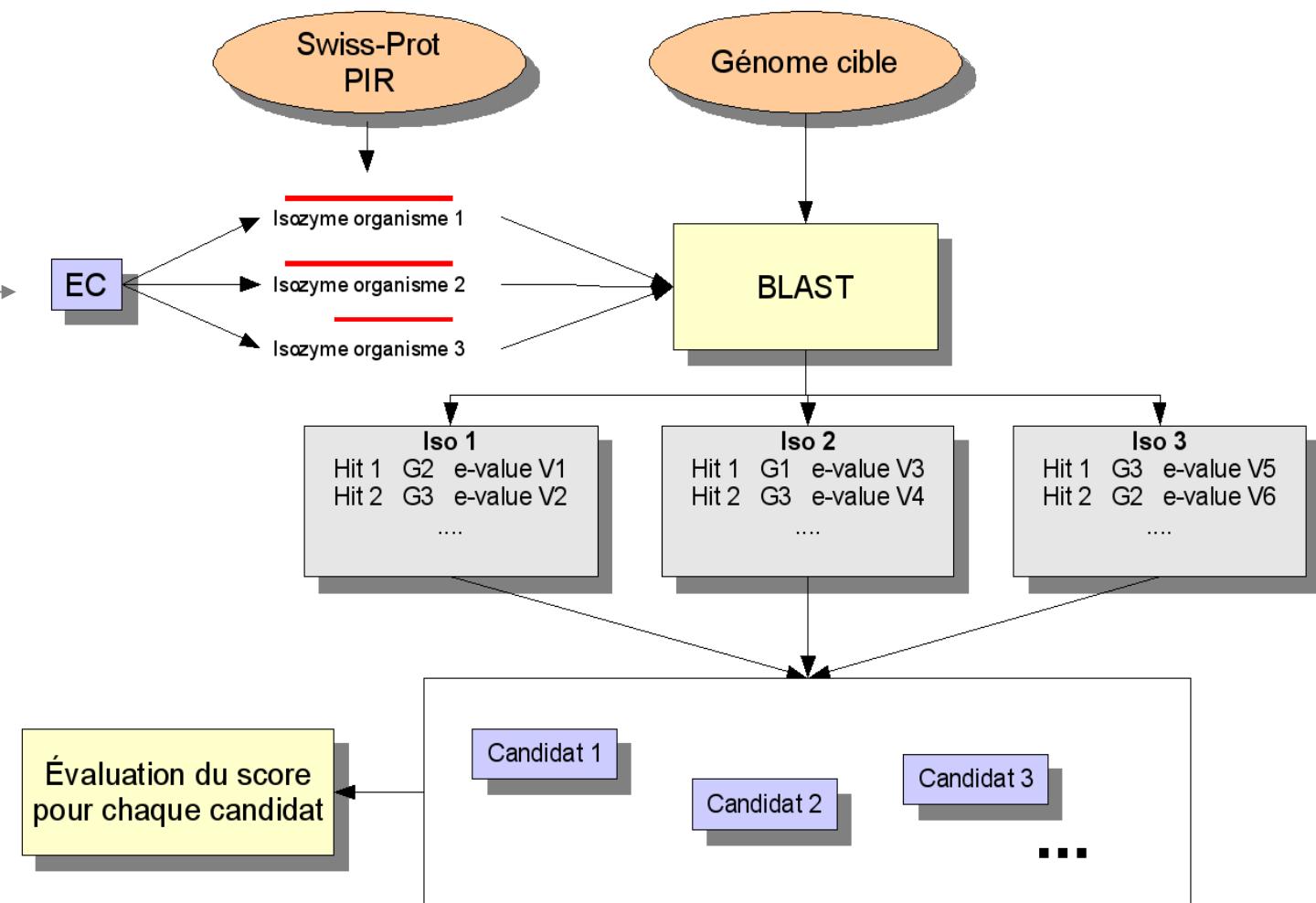
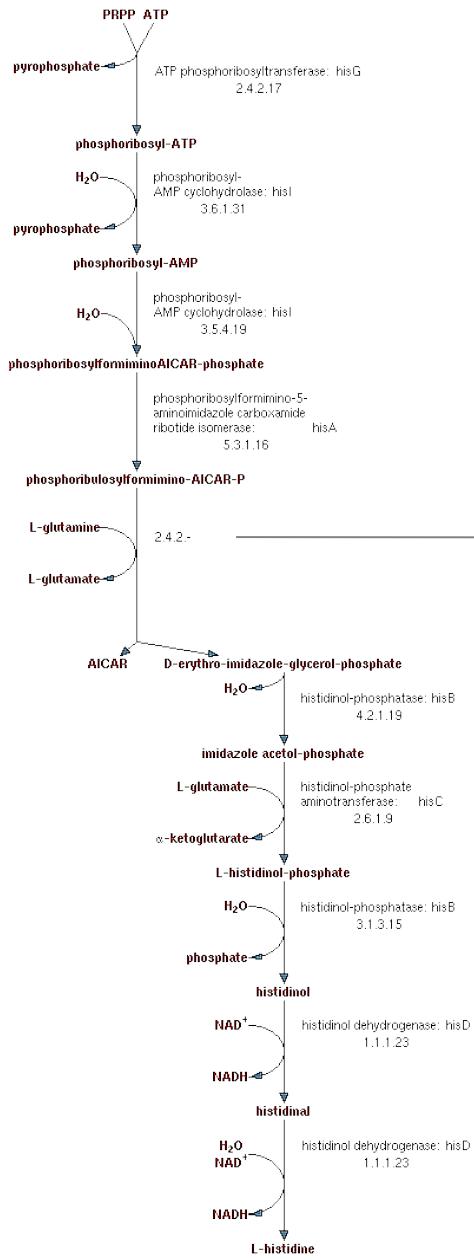
# Correction de la spécificité des réactions

- Utilisation de BRENDA qui contient la spécificité d'enzymes chez de nombreux organismes
- Vérification de la disponibilité de tous les substrats et cofacteurs
- Assignation de la spécificité grâce à un organisme de référence proche de l'organisme étudié

# Identification des réactions manquantes

- Ajouter les réactions spontanées
- Sur la base des voies métaboliques identifiées : Pathway Hole Filler dans PATHOLOGIC
- Test de scénarios : The Seed

# Pathway Hole Filler



# Réactions génériques

- Préciser les substrats et les produits : aucune méthode automatique ne fait pour l'instant cette opération
  - Utiliser les bases de données métaboliques et les métabolites déjà présents dans le réseau

# Direction des réactions

- Sur la base des voies métaboliques identifiées
- Grâce à l'analyse des contraintes topologiques et stoechiométriques du réseau métabolique

(Kümmel, A.; Panke, S. & Heinemann, M. Systematic assignment of thermodynamic constraints in metabolic network models BMC Bioinformatics, 2006, 7, 512)

# Compartimentation

- Utilisation des données de protéomique
- Recherche de signatures dans les séquences
- Prise en compte de réactions *a priori* localisées et des contraintes du réseau global (Mintz-Oron, S.; Aharoni, A.; Ruppin, E. & Shlomi, T. Network-based prediction of metabolic enzymes' subcellular localization. Bioinformatics, 2009, 25, i247-i252)

# Erreurs dans la formulation des réactions

- Coefficients stoechiométriques erronés
- Métabolites manquants (ex : H<sup>+</sup>)

# Metabolic network edition and sharing

# Exchange formats

- SBML : centré sur les réactions et sur la modélisation
- BioPax : contient d'autres informations que les réactions

# Raw SBML

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1" xmlns:html="http://www.w3.org/1999/xhtml">
<model id="Ec_iAF1260" name="Ec_iAF1260">
<listOfCompartments>
  <compartment id="in" outside="out"/>
  <compartment id="out" />
</listOfCompartments>
<listOfSpecies>
  <species id="A_out" name="A_out" boundaryCondition="true" compartment="out">
  <species id="A_in" name="A_in" boundaryCondition="false" compartment="in">
  <species id="B_in" name="B_in" boundaryCondition="false" compartment="in">
  <species id="B_out" name="B_out" boundaryCondition="true" compartment="out">
</listOfSpecies>
<listOfReactions>
  <reaction id="R1" name="R1" reversible="false">
    <listOfReactants>
      <speciesReference species="A_out" stoichiometry="1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="A_in" stoichiometry="1"/>
    </listOfProducts>
  </reaction>
  <reaction id="R2" name="R2" reversible="false">
    <listOfReactants>
      <speciesReference species="A_in" stoichiometry="1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="B_in" stoichiometry="1"/>
    </listOfProducts>
  </reaction>
  <reaction id="R3" name="R3" reversible="false">
    <listOfReactants>
      <speciesReference species="B_in" stoichiometry="1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="B_out" stoichiometry="1"/>
    </listOfProducts>
  </reaction>
</listOfReactions>
</model>
</sbml>
```

# SBML annotations : SBO terms

- Systems Biology Ontology

The diagram illustrates the use of SBO terms in SBML annotations. On the left, a snippet of SBML code is shown, and on the right, a vertical stack of colored boxes maps specific SBO terms to their definitions.

**SBML** logo

```
<reaction id="E1" reversible="false" sboterm="SBO:0000182">
  <listOfReactants>
    <speciesReference species="S" sboterm="SBO:0000015"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="P" sboterm="SBO:0000011"/>
  </listOfProducts>
  <listOfModifiers>
    <modifierSpeciesReference species="E" sboterm="SBO:0000013"/>
  </listOfModifiers>
  <kineticLaw sboterm="SBO:0000029">
    <math xmlns="http://www.w3.org/1998/Math/MathML">[ ... ]</math>
    <listOfParameters>
      <parameter id="K" value="1" sboterm="SBO:0000027"/>
    </listOfParameters>
  </kineticLaw>
</reaction>
```

Conversion

Substrate

Product

Catalyst

Henri-Michaelis-Menten rate law

Michaelis constant

# SBO terms

- SBO:0000000 - swo term
  - ⓘ SBO:0000064 - mathematical expression
    - ⊕ ⓘ SBO:0000355 - conservation law
    - ⊕ ⓘ SBO:0000474 - convenience function
    - ⓘ SBO:0000001 - rate law
      - ⊕ ⓘ SBO:0000268 - enzymatic rate law
      - ⊕ ⓘ SBO:0000192 - Hill-type rate law, generalised form
        - ⓘ SBO:0000195 - Hill-type rate law, microscopic form
        - ⓘ SBO:0000198 - Hill-type rate law, reduced form
      - ⊕ ⓘ SBO:0000012 - mass action rate law
      - ⊕ ⓘ SBO:0000527 - modular rate law
      - ⊕ ⓘ SBO:0000391 - steady state expression
    - ⊕ ⓘ SBO:0000544 - metadata representation
    - ⊕ ⓘ SBO:0000004 - modelling framework
    - ⊕ ⓘ SBO:0000231 - occurring entity representation
    - ⊕ ⓘ SBO:0000003 - participant role
    - ⓘ SBO:0000236 - physical entity representation
      - ⊕ ⓘ SBO:0000241 - functional entity
      - ⓘ SBO:0000409 - interaction outcome
      - ⓘ SBO:0000240 - material entity
        - ⓘ SBO:0000291 - empty set
        - ⊕ ⓘ SBO:0000354 - informational molecule segment
        - ⊕ ⓘ SBO:0000245 - macromolecule
          - ⊕ ⓘ SBO:0000248 - chemical macromolecule
          - ⊕ ⓘ SBO:0000246 - information macromolecule
            - ⓘ SBO:0000251 - **deoxyribonucleic acid**
            - ⓘ SBO:0000252 - polypeptide chain
            - ⓘ SBO:0000250 - ribonucleic acid
        - ⓘ SBO:0000285 - material entity of unspecified nature
      - ⊕ ⓘ SBO:0000253 - non-covalent complex
      - ⓘ SBO:0000406 - observable
      - ⓘ SBO:0000405 - perturbing agent
      - ⊕ ⓘ SBO:0000290 - physical compartment
      - ⊕ ⓘ SBO:0000247 - simple chemical

# Minimum information requested in the annotation of biochemical models (MIRIAM)

- Description complète du modèle (auteurs, date de création, etc...)
- Description des entités grâce à des urls

# Minimum information requested in the annotation of biochemical models (MIRIAM)

```
<speciesType metaid="_metaE_1" id="E_1" name="YBL022C">
  <annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
      xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
      <rdf:Description rdf:about="#_metaE_1">
        <bqbiol:isDescribedBy>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:pubmed:8276800"/>
          </rdf:Bag>
        </bqbiol:isDescribedBy>
        <bqbiol:is>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:uniprot:P36775"/>
          </rdf:Bag>
        </bqbiol:is>
        <bqbiol:isEncodedBy>
          <rdf:Bag>
            <rdf:li rdf:resource="urn:miriam:sgd:S000000118"/>
          </rdf:Bag>
        </bqbiol:isEncodedBy>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</speciesType>
```

# SemanticSBML

semanticSBML

BIOMD0000000005.xml

Model - Tyson1991\_CellCycle\_6var [Tyson1991CellModel\_6]

Compartments (1/1)

[cell]

Species (8/9)

- [EmptySet]
- cdc2k [C2]
- cdc2k-P [CP]
- p-cyclin\_c... [M]
- p-cyclin\_c... [pM]
- cyclin [Y]
- p-cyclin [YP]
- total\_cycl... [YT]
- total\_cdc2 [CT]

Reactions (9/9)

- cyclin\_cdc... [Reaction1]
- cdc2k phos... [Reaction2]
- cdc2k deph... [Reaction3]
- cyclin cdc... [Reaction4]
- deactivati... [Reaction5]
- cyclin bio... [Reaction6]
- default de... [Reaction7]
- cdc2 kinas... [Reaction8]

cdc2k [C2]

MIRIAM SBO

Current

qualifier	resource	id	name
bio:isVersionOf	UniProt	P04551	Mitogen-activated protein kinase kinase 1

delete

Copy/Paste

copy

Add

qualifier	resource	id
bio:is	Brenda Tissue Ontology	

add

Search

-activated protein kinase 1 export to: xml json yaml OR Predict annotations

bio:is use this qualifier

+ Nucleotide Sequence Database	Mitogen-activated protein kinase 1
+ UniProt	Mitogen-activated protein kinase 1
+ Nucleotide Sequence Database	Mitogen-activated protein kinase 11
+ UniProt	Mitogen-activated protein kinase 11
+ Nucleotide Sequence Database	Mitogen-activated protein kinase 9
+ UniProt	Mitogen-activated protein kinase 9
+ Nucleotide Sequence Database	Mitogen-activated protein kinase 8
+ UniProt	Mitogen-activated protein kinase 8

# Edition and sharing

- The raw way : SVN or Github
- MetExplore 2

MetExplore

Welcome Cottret Ludovic

Network Data Network Viz

About Mapping Graph Modelling Test Export Import

Compartments Pathways Reactions Metabolites Enzymes Proteins Genes

Commit Changes Delete Multiple affectation Add Comment Add Biblio Modify Status

	id	dbIdentifier	name	ec
1	263883	R_EX_pnto_R_e_	R_R_Pantothenate_exchange	NA
2	263864	R_EX_btd_RR_e_	R_R_R_2_3_Butanediol_exchange	NA
3	263032	R_BTDD_RR	R_R_R_butanediol_dehydrogenase	1.1
4	263951	R_BTDt_RR	R_R_R_butanediol_transport	NA
5	264068	R_PRMIClI	R_1_5_phosphoribosyl_5_5_phosphoribosylamino_methylen...	5.3
6	262947	R_EX_13BDglcn_e_	R_1_3_beta_D_Glucan_exchange	NA
7	264381	R_13GS	R_1_3_beta_glucan_synthase	2.4
8	263664	R_GBEZ	R_1_4_alpha_glucan_branching_enzyme	2.4
9	263021	R_16GS	R_1_6_beta_glucan_synthase	NA
10	263863	R_AGAT_SC	R_1_Acyl_glycerol_3_phosphate_acyltransferase_yeast_specific	2.3
11	263508	R_GALIGH	R_1_alpha_D_Galactosyl_myo_inositol_galactohydrolase	3.2
12	263915	R_PI35BP5P_SC	R_1_phosphatidylinositol_3_5_bisphosphate_5_phosphatase_ye...	NA
13	263467	R_PIN3K_SC	R_1_phosphatidylinositol_3_kinase_yeast_specific	2.7
14	263811	R_PI45BP5P_SC	R_1_phosphatidylinositol_4_5_bisphosphate_5_phosphatase_ye...	NA
15	263488	R_PI45BPP_SC	R_1_phosphatidylinositol_4_5_bisphosphate_phosphodiesterase_...	3.1
16	264101	R_P5CDm	R_1_pyrroline_5_carboxylate_dehydrogenase_mitochondrial	1.5
17	263408	R_MI145Ptn	R_1D_myo_Inositol_1_4_5_trisphosphate_nuclear_transport_via...	NA
18	263764	R_DKMPPD2	R_2_3_diketo_5_methylthio_1_phosphopentane_degradation_reac...	NA
19	263786	R_DRTPPD	R_2_5_diamino_6_ribitylamino_4_3H_pyrimidinone_5_phosphat...	NA
20	262971	R_DROPPRy	R_2_5_diamino_6_ribosylamino_4_3H_pyrimidinone_5_phospha...	NA
21	264225	R_ACHBSm	R_2_aceto_2_hydroxybutanoate_synthase_mitochondrial	4.1

# Metabolic network exploration

# BioCyc-like pages from the pathway-tools

The screenshot shows the Trypanocyc website. At the top, there's a logo with the word "Trypanocyc" and a blue circular icon. Below the logo is a navigation bar with links for "Home", "Search", "Tools", and "Help". To the right of the navigation bar is a search bar containing the text "Searching Trypanosoma brucei". Further to the right are "Quick Search" and "Gene Search" buttons. The main content area has a light blue header with the text "Searching Trypanosoma brucei".

## Trypanocyc Overview

Trypanocyc is a community annotated Pathway/Genome Database of *Trypanosoma brucei*, the causative agent of African Trypanosomiasis. Trypanocyc is maintained at the French National Institute of Agricultural Research (INRA) to collect information on the *T.brucei* metabolism and make it available to the public. TrypanoCyc was built using a collaborative web platform (TrypAnnot) allowing splitting the annotation efforts between a group of experts each one working on his/her pathways of interest. TrypnoCyc improve automatic metabolic network reconstruction by adding developmental-stage and compartment information on enzyme activity. These metadata on the network will help generate tailor-made metabolic networks (e.g. metabolic network of procyclic mitochondria) and help understand the parasite metabolism to a greater extend.

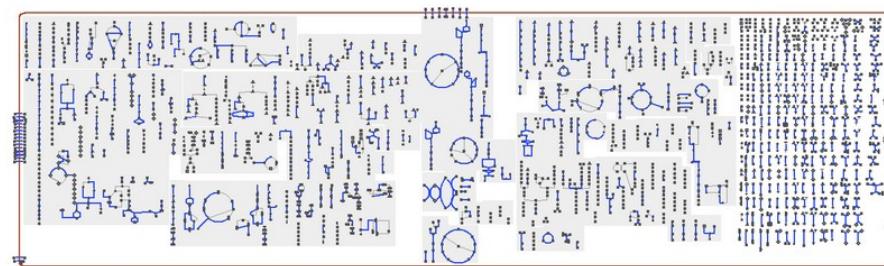
## Trypanocyc Annotation

One of the major objectives of the Trypanocyc project is to make it easy for biologists to annotate reactions and provide information regarding the

- localization of enzyme activity and
- presence/absence of enzyme activity at various developmental stages of the parasite.

## Mapping, Model Analysis and SBML Export

Pathway tools provide a set of tools in mapping Metabolites, Enzymes, Genes, Reactions and Pathways and visualization of experimental data on the model. The [Metexplore](#) tool can provide additional tools for mapping, visualization, model analysis and SBML export of the whole metabolic network or its subnetworks.



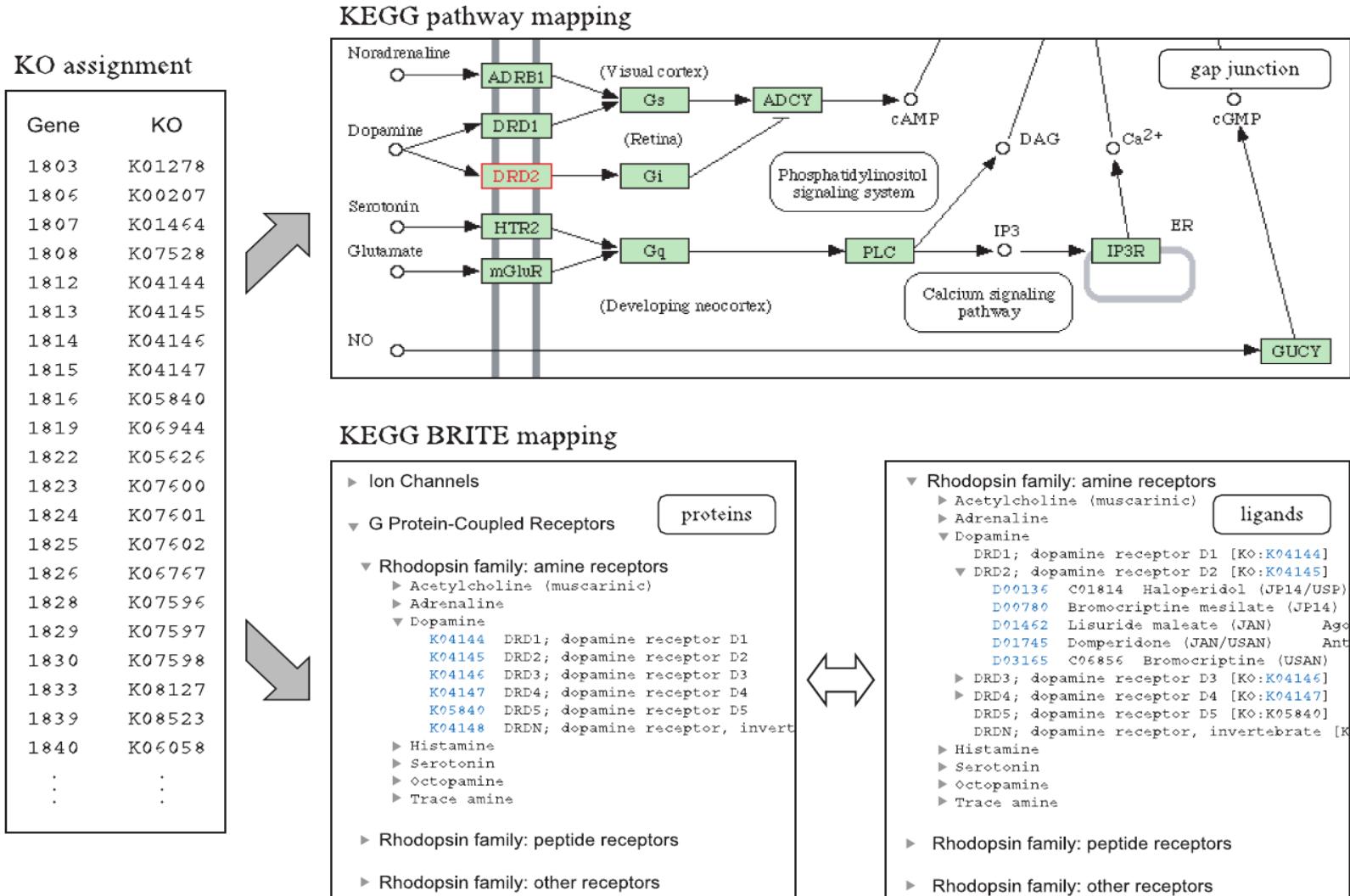
Coordination of the annotation: Flora Logan-Klumpler, Fabien Jourdan and Michael Barrett

Methodological and technical developments: Ludovic Cottret, Florence Vinson and Fabien Jourdan

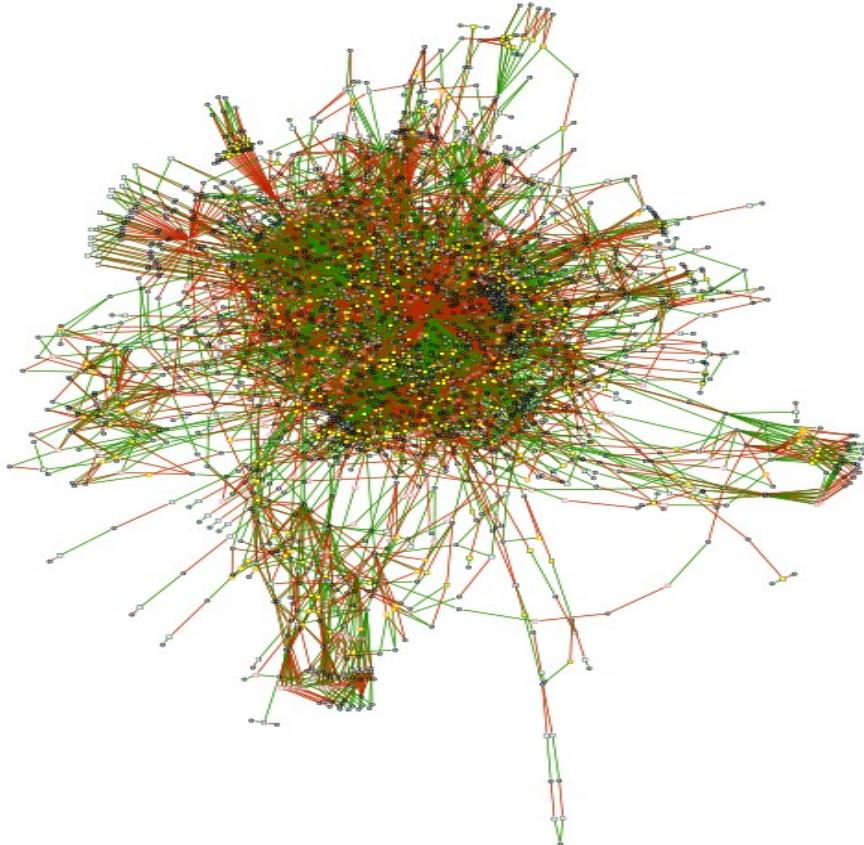
Curator: Sanu Shameer

Annotators: Fiona Achcar, Mike Barrett, Michael Boshart, Frederic Bringaud, Peter Butikofer, Darren Creek, Harry De Koning, Charles Ebikeme, Alan Fairlamb, Mike Ferguson, Michael Ginger, Eduard Kerkhoven, Flora Logan, Pascal Maser, Paul Michels, Archana Nayak, Derek Nolan, Christian Olsen, Fred Oppenhoes, Marc Ouellette, Meg Phillips, Mick Urbaniak, Terry Smith, Martin Taylor, Aloysius Tielens and Jaap van Hellemond

# KEGG pathways

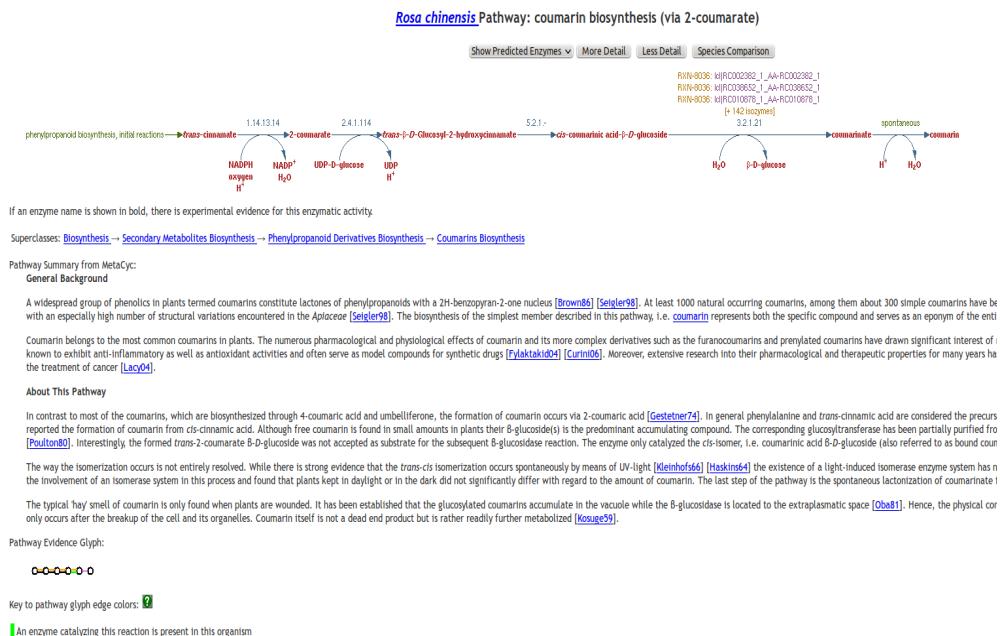
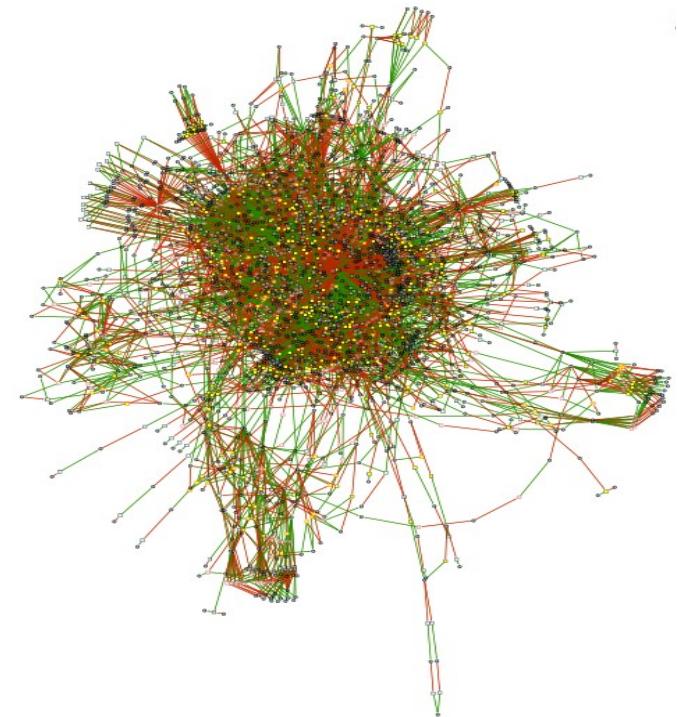
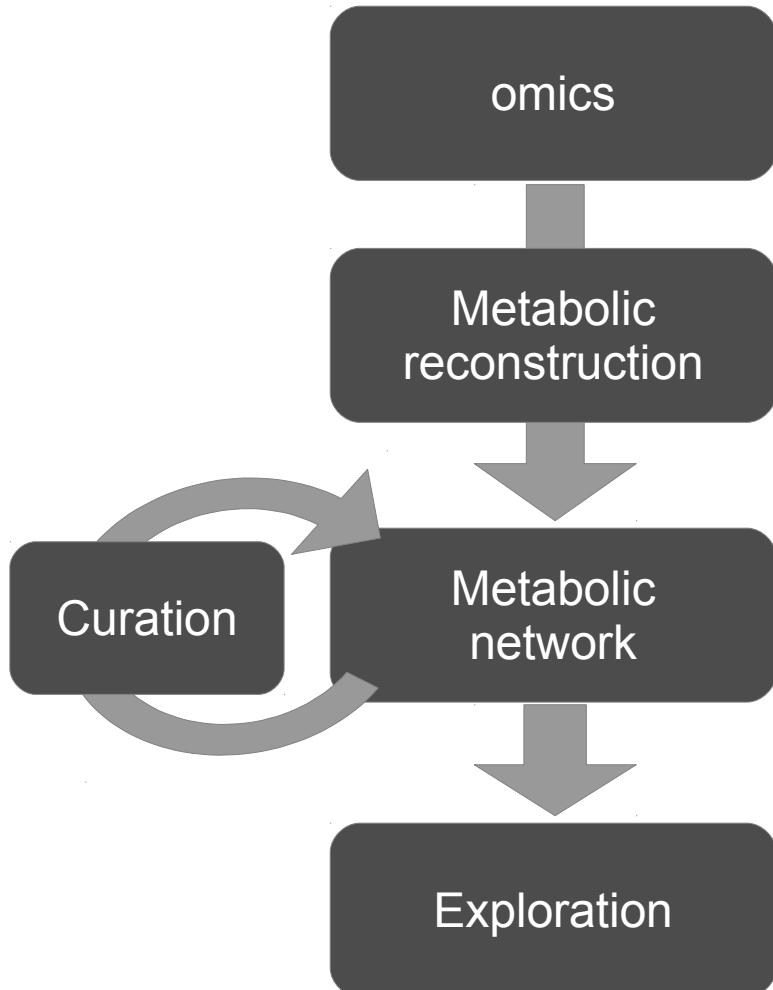


# Network visualisation

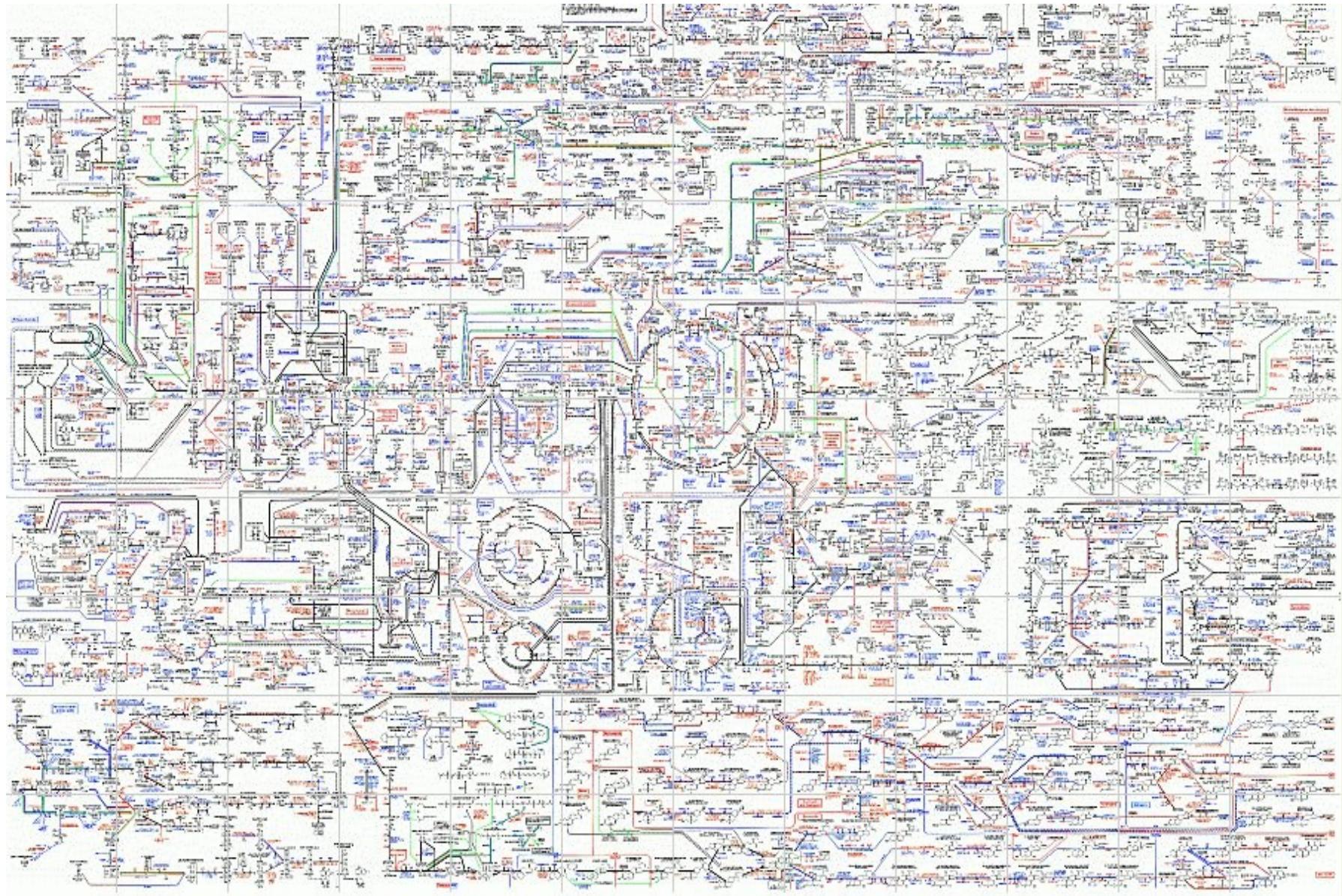


- Cytoscape
- Tulip (systrip plugin)
- Wanted
- Javascript libraries : cytoscape.js, d3.js, etc...
- MetExplore via Cytoscape Java and javascript

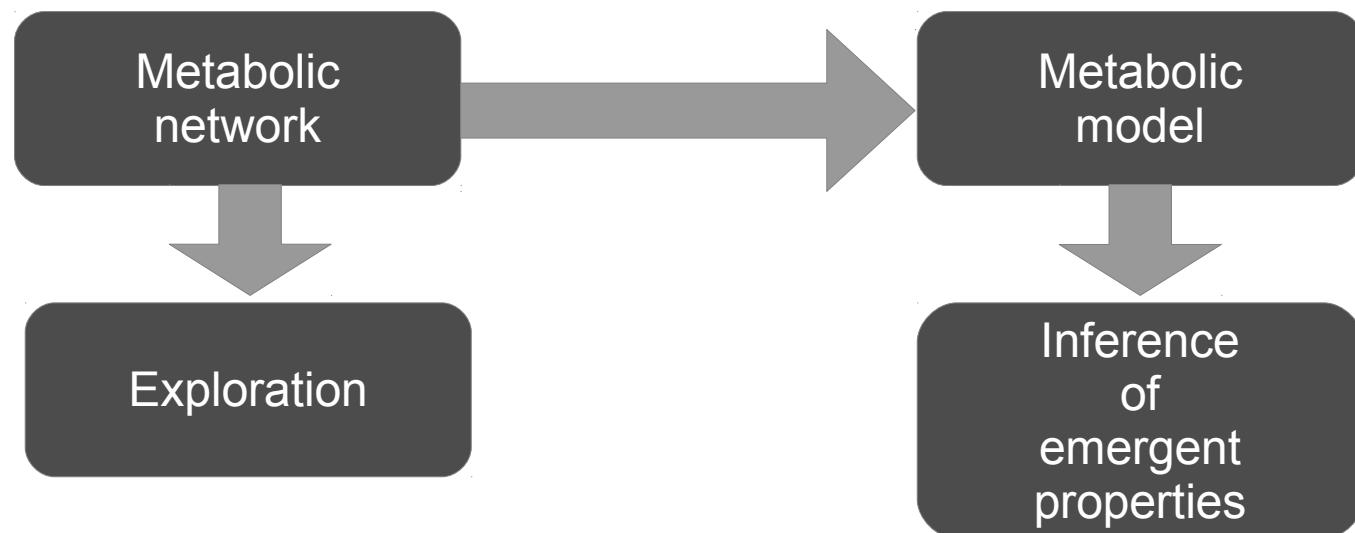
# What else ?



# The metabolic network is a complex system



# From the metabolic network to the metabolic model

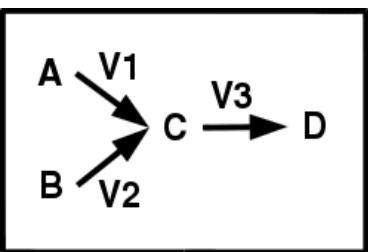


# Two ways to model genome scale networks

- Metabolic Graphs
- **Flux models**

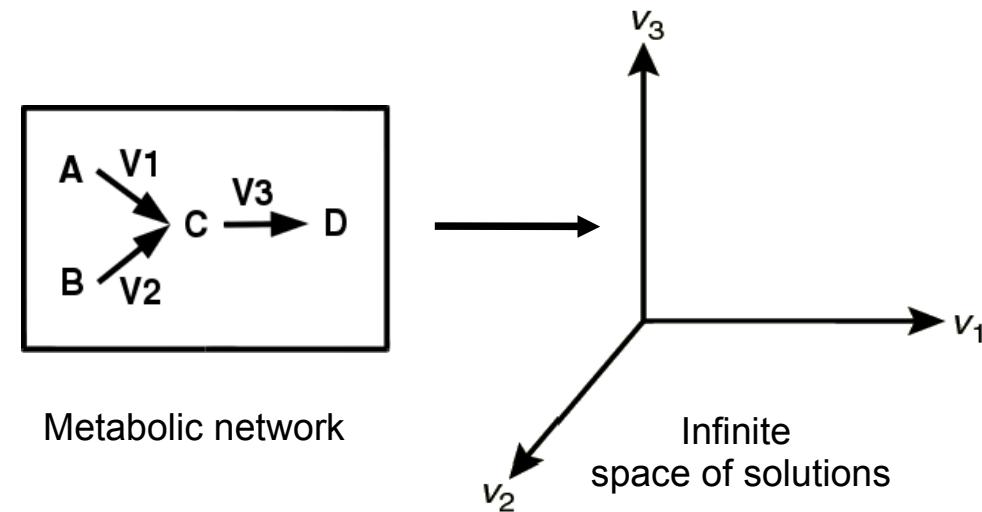
# **Flux analysis principle**

# Flux analysis principle



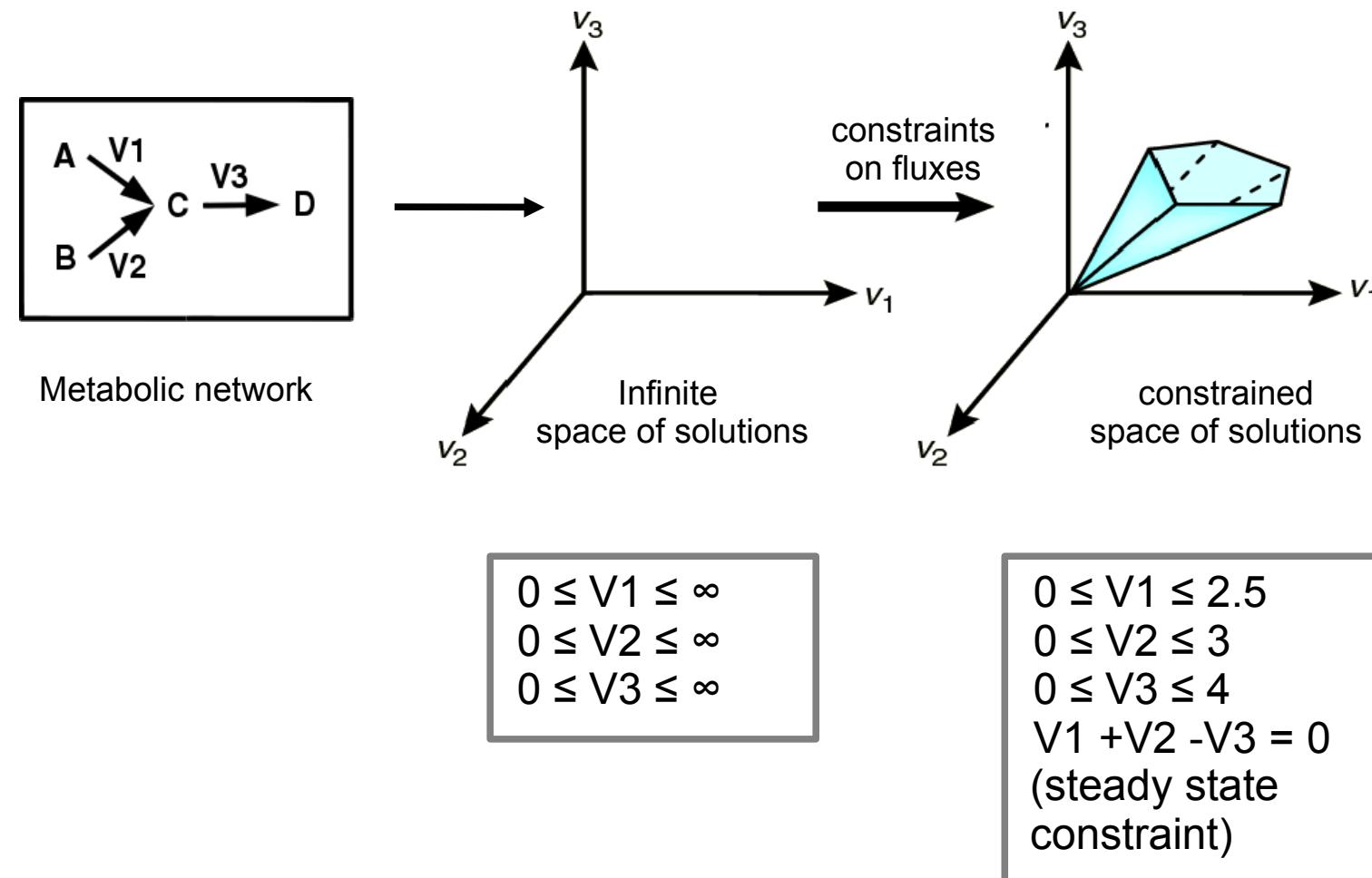
Metabolic network

# Flux analysis principle

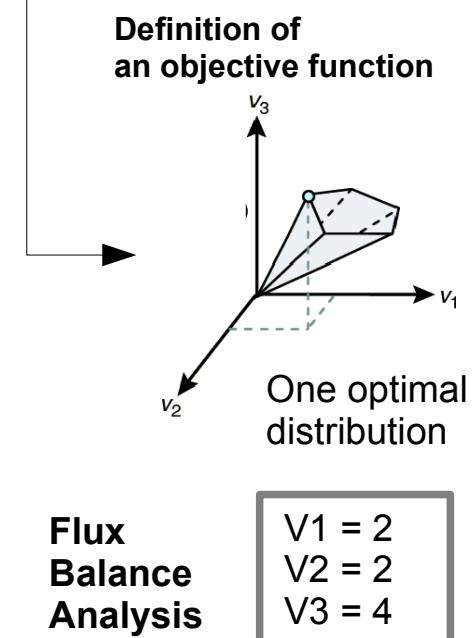
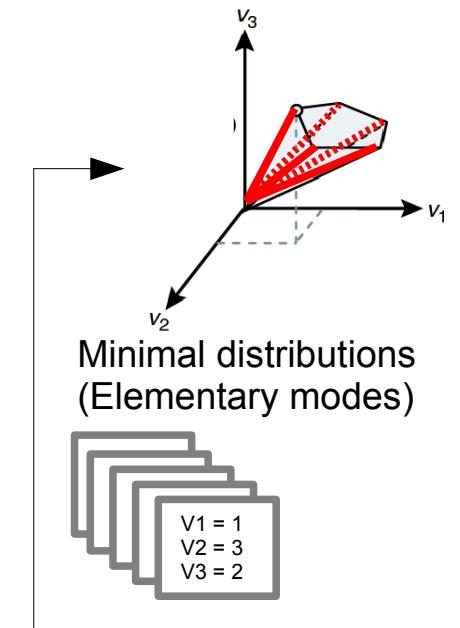
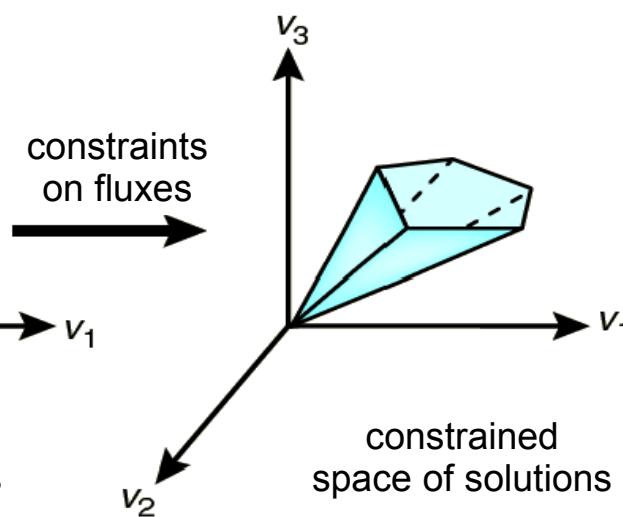
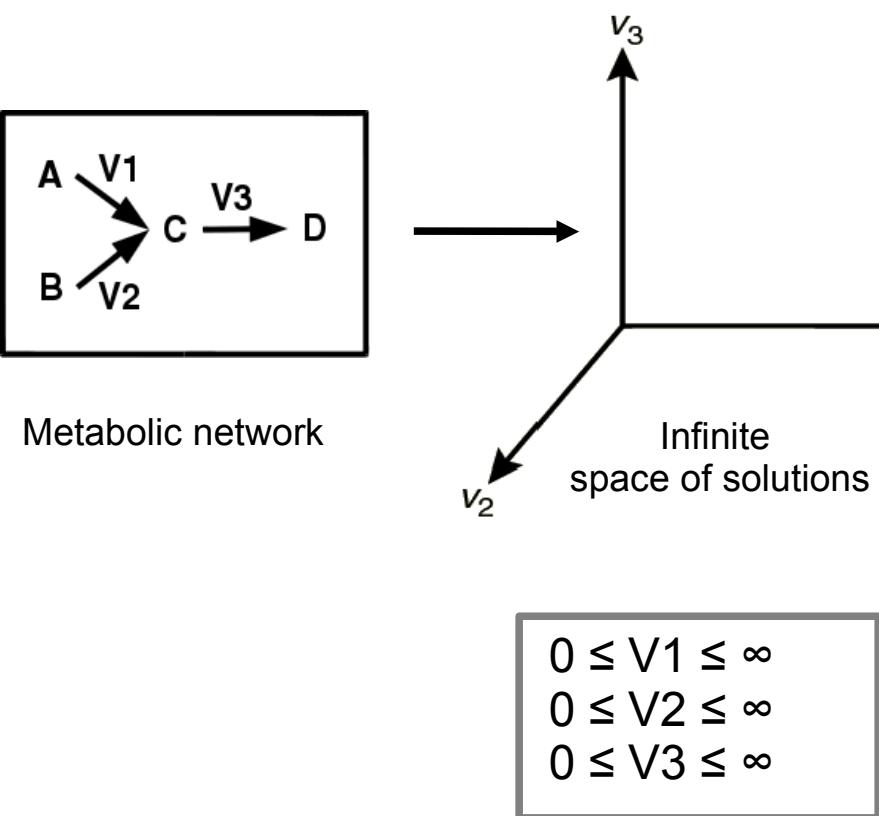


$$\begin{aligned}0 \leq V1 \leq \infty \\ 0 \leq V2 \leq \infty \\ 0 \leq V3 \leq \infty\end{aligned}$$

# Flux analysis principle

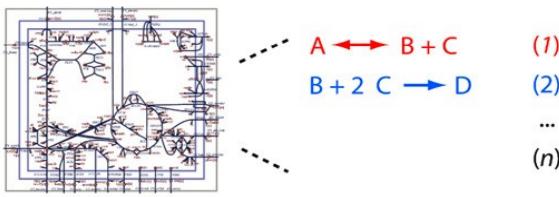


# Flux analysis principle



# Formulation of a FBA problem

## a Curate metabolic reactions



## b Formulate **S** matrix

		Reactions	
Metabolites	A	1 2 ... n	
	B	-1 1 -1 ...	
	C	1 -2 ...	
	D	1 ...	
	...	...	
	m		

**S**

## c Apply mass balance constraints

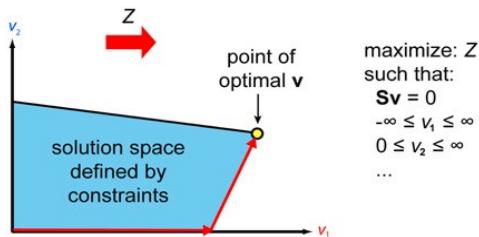
$$\begin{matrix} \mathbf{S} (m \times n) \\ \begin{bmatrix} -1 & & & \\ 1 & -1 & & \\ 1 & -2 & & \\ 1 & & & \end{bmatrix} \end{matrix} * \begin{matrix} \mathbf{v} (n \times 1) \\ \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix} \end{matrix} = 0 \rightarrow \begin{matrix} m \text{ mass balance} \\ \text{equations} \\ -v_1 + \dots = 0 \\ v_1 - v_2 + \dots = 0 \\ v_1 - 2 v_2 + \dots = 0 \\ v_2 + \dots = 0 \\ \vdots \end{matrix}$$

## d Define objective function $Z$

$$Z = \begin{matrix} \mathbf{c}^T (1 \times n) \\ \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \end{matrix} * \begin{matrix} \mathbf{v} (n \times 1) \\ \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \end{matrix}$$

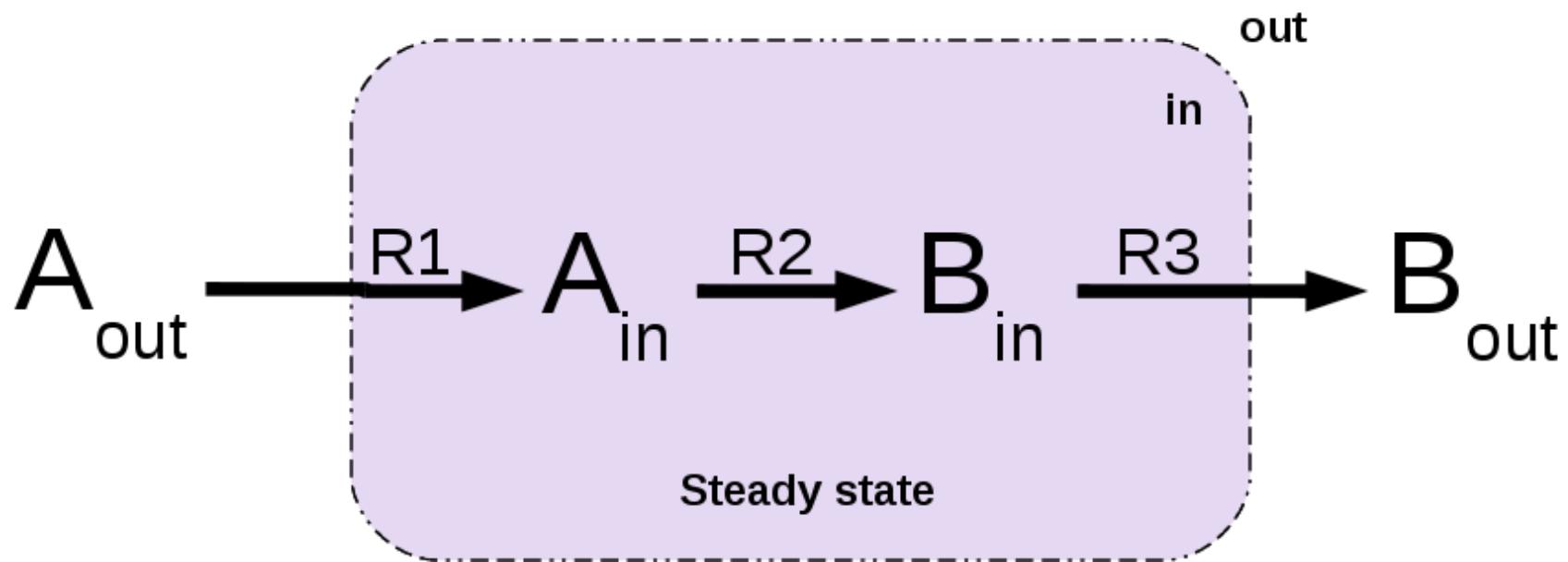
sets reaction 1 as the objective

## e Optimize $Z$ using linear programming

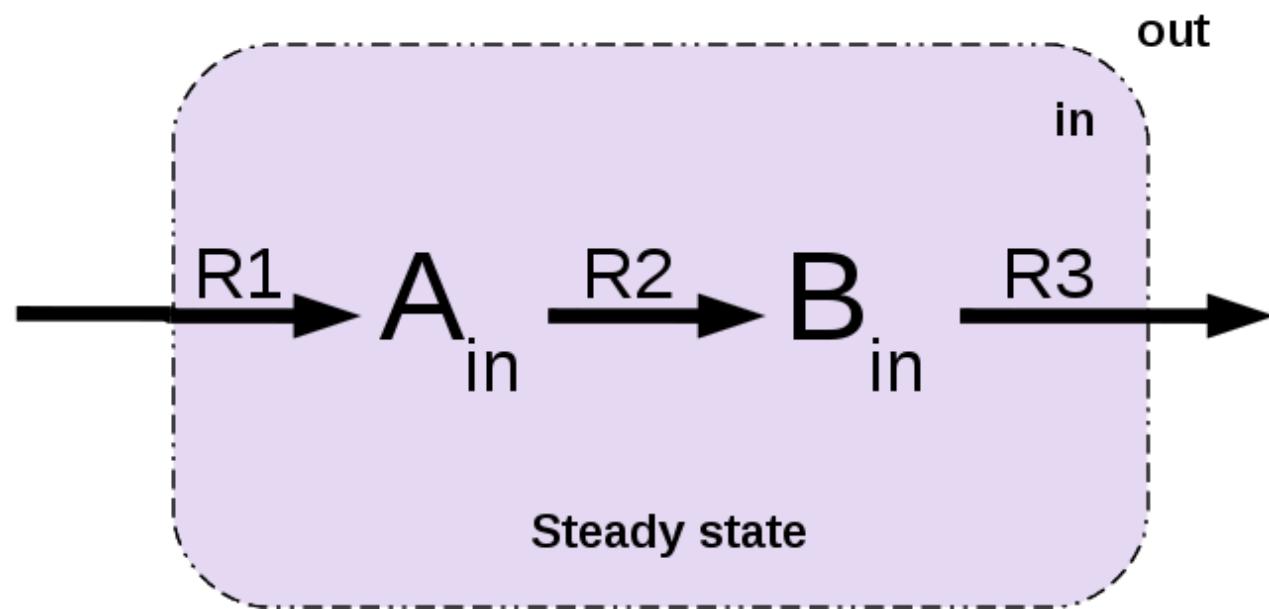


What is flux balance analysis? Jeffrey D. Orth, Ines Thiele, and Bernhard Ø. Palsson. Nat Biotechnol., 2011

# Steady state assumption



# Steady state assumption



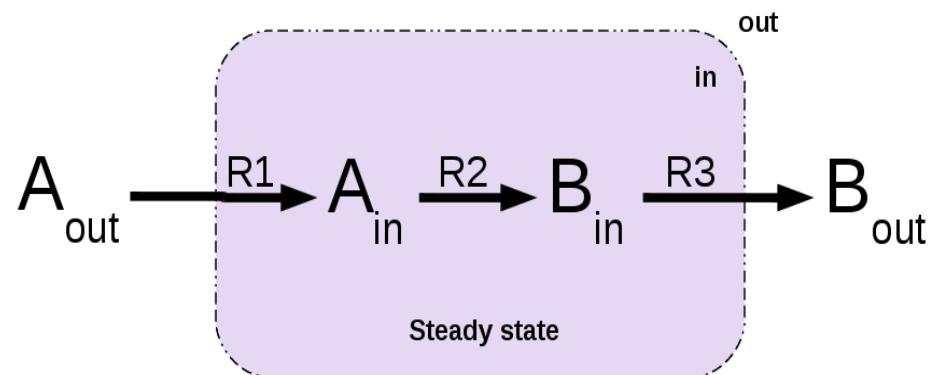
# Biomass reaction

Fake reaction whose the substrates are all the cell components and growth requirements.

The stoichiometric coefficients correspond to the relative proportion of each component in the biomass composition.

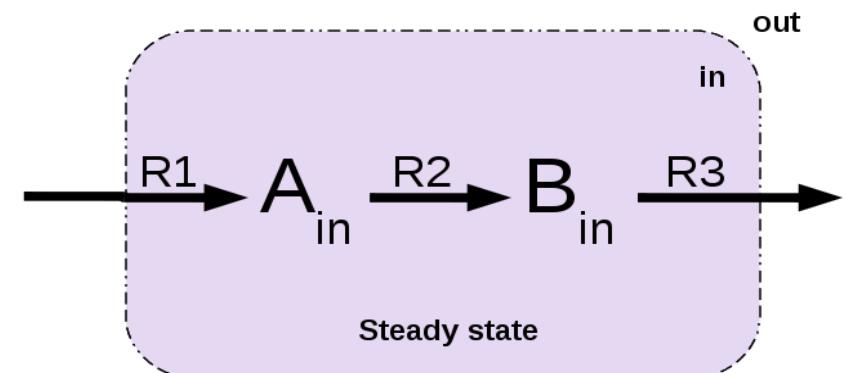
# The starting point : the SBML file

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1" xmlns:html="http://www.w3.org/1999/xhtml">
<model id="Ec_iAF1260" name="Ec_iAF1260">
<listOfCompartments>
  <compartment id="in" outside="out"/>
  <compartment id="out" />
</listOfCompartments>
<listOfSpecies>
  <species id="A_out" name="A_out" boundaryCondition="true" compartment="out">
    <species id="A_in" name="A_in" boundaryCondition="false" compartment="in">
      <species id="B_in" name="B_in" boundaryCondition="false" compartment="in">
        <species id="B_out" name="B_out" boundaryCondition="true" compartment="out">
</listOfSpecies>
<listOfReactions>
  <reaction id="R1" name="R1" reversible="false">
    <listOfReactants>
      <speciesReference species="A_out" stoichiometry="1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="A_in" stoichiometry="1"/>
    </listOfProducts>
  </reaction>
  <reaction id="R2" name="R2" reversible="false">
    <listOfReactants>
      <speciesReference species="A_in" stoichiometry="1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="B_in" stoichiometry="1"/>
    </listOfProducts>
  </reaction>
  <reaction id="R3" name="R3" reversible="false">
    <listOfReactants>
      <speciesReference species="B_in" stoichiometry="1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="B_out" stoichiometry="1"/>
    </listOfProducts>
  </reaction>
</listOfReactions>
</model>
</sbml>
```



# The starting point : the SBML file

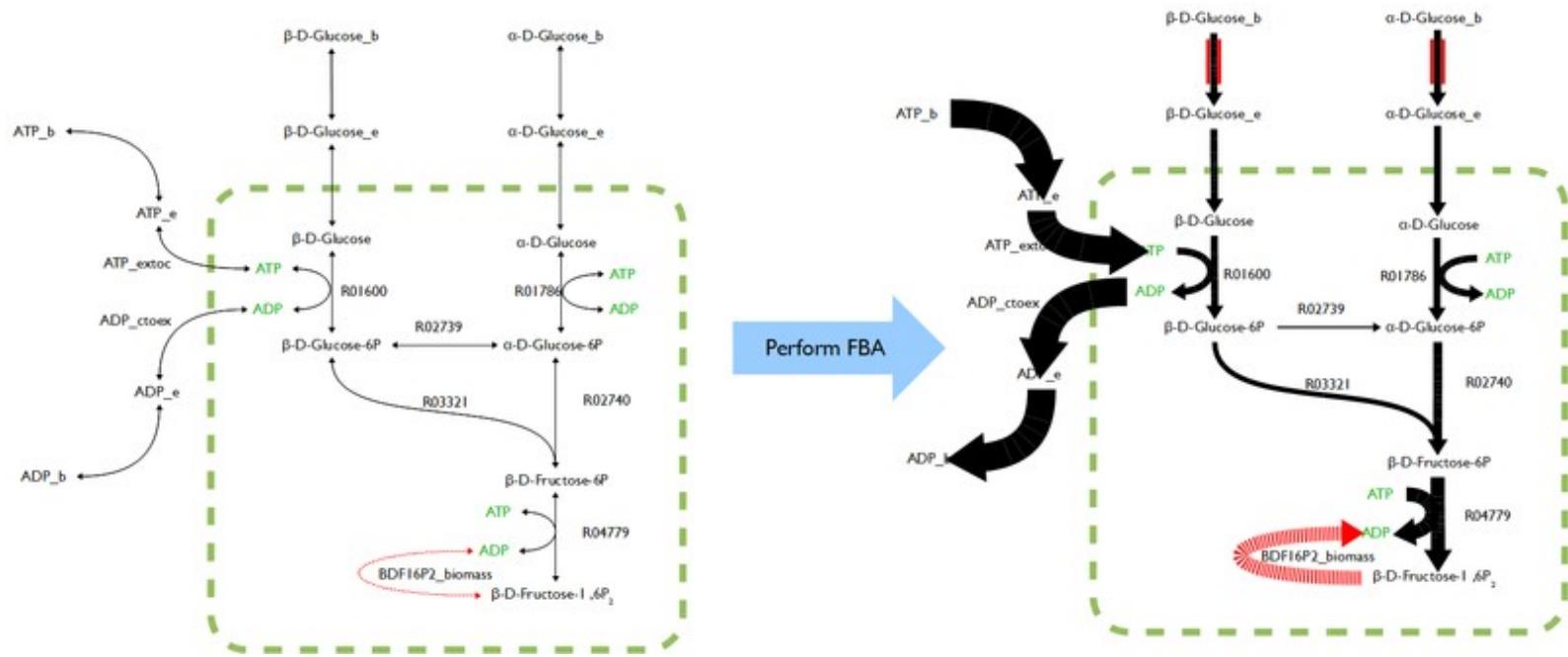
```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2" level="2" version="1" xmlns:html="http://www.w3.org/1999/xhtml">
<model id="Ec_iAF1260" name="Ec_iAF1260">
<listOfCompartments>
  <compartment id="in"/>
</listOfCompartments>
<listOfSpecies>
  <species id="A_in" name="A_in" boundaryCondition="false" compartment="in" />
  <species id="B_in" name="B_in" boundaryCondition="false" compartment="in" />
</listOfSpecies>
<listOfReactions>
  <reaction id="R1" name="R1" reversible="false">
    <listOfProducts>
      <speciesReference species="A_in" stoichiometry="1"/>
    </listOfProducts>
  </reaction>
  <reaction id="R2" name="R2" reversible="false">
    <listOfReactants>
      <speciesReference species="A_in" stoichiometry="1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="B_in" stoichiometry="1"/>
    </listOfProducts>
  </reaction>
  <reaction id="R3" name="R3" reversible="false">
    <listOfReactants>
      <speciesReference species="B_in" stoichiometry="1"/>
    </listOfReactants>
  </reaction>
</listOfReactions>
</model>
</sbml>
```



# Additional requirements for building flux models

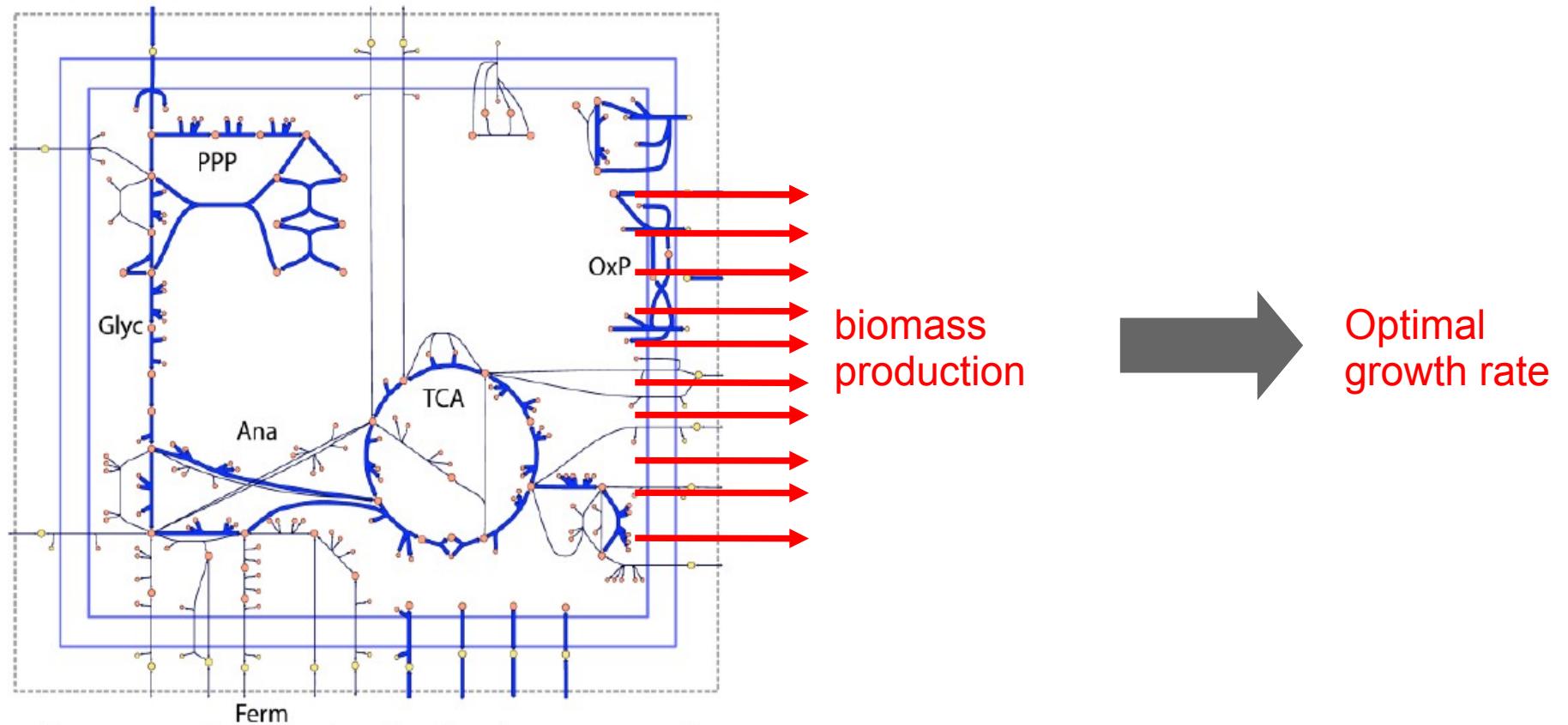
- Reactions must be balanced
- Identification of dead reactions (reactions that are not able to carry any flux)
- Identification of dead metabolites (metabolites that are not balanced) : needs to add exchange reactions
- Lower and upper bounds for each reaction
- Biomass reaction
- Media that enable growth
- Add ATP and growth requirement reactions
- Add Gene Protein Reaction links

# Metabolic flux map

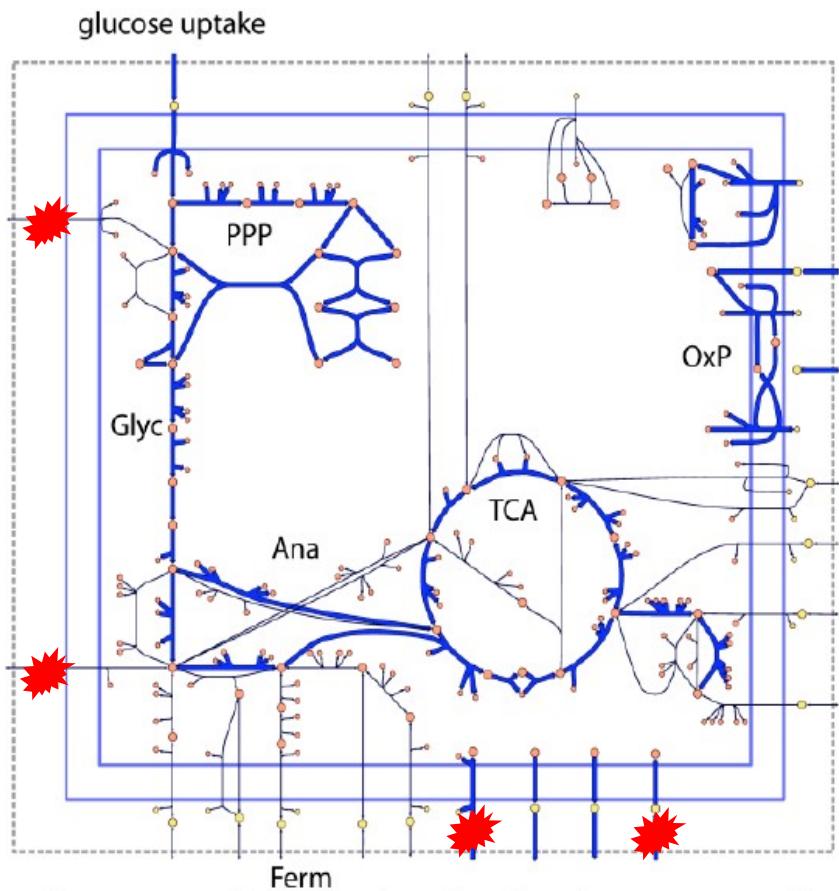


Source : Wikipedia

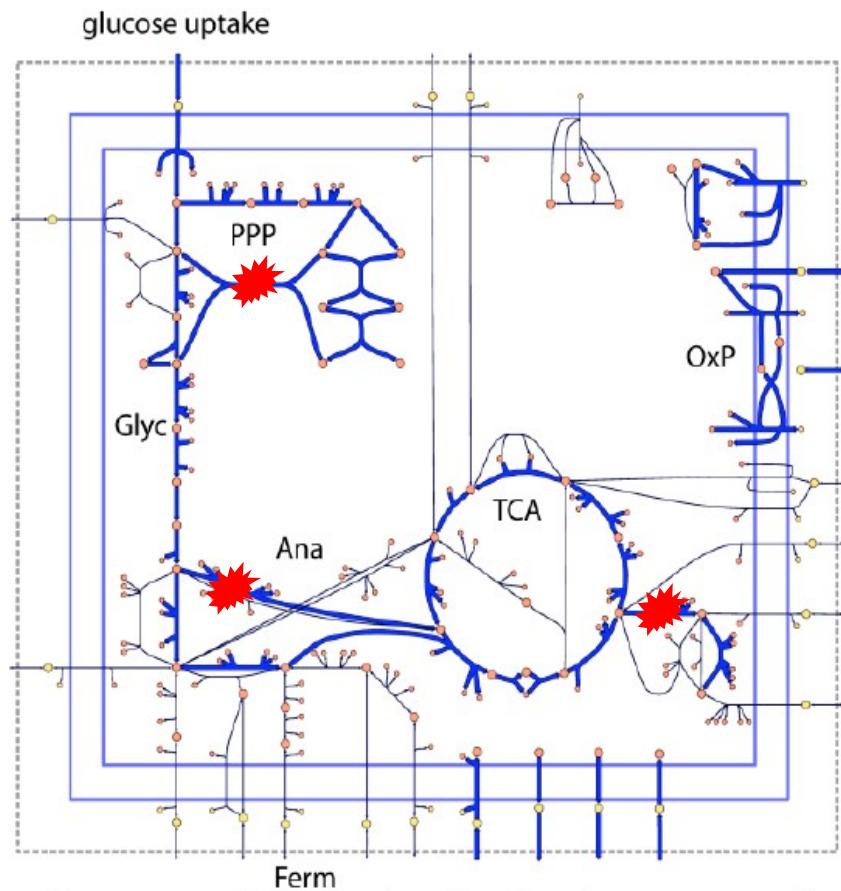
# Computing optimal growth rate



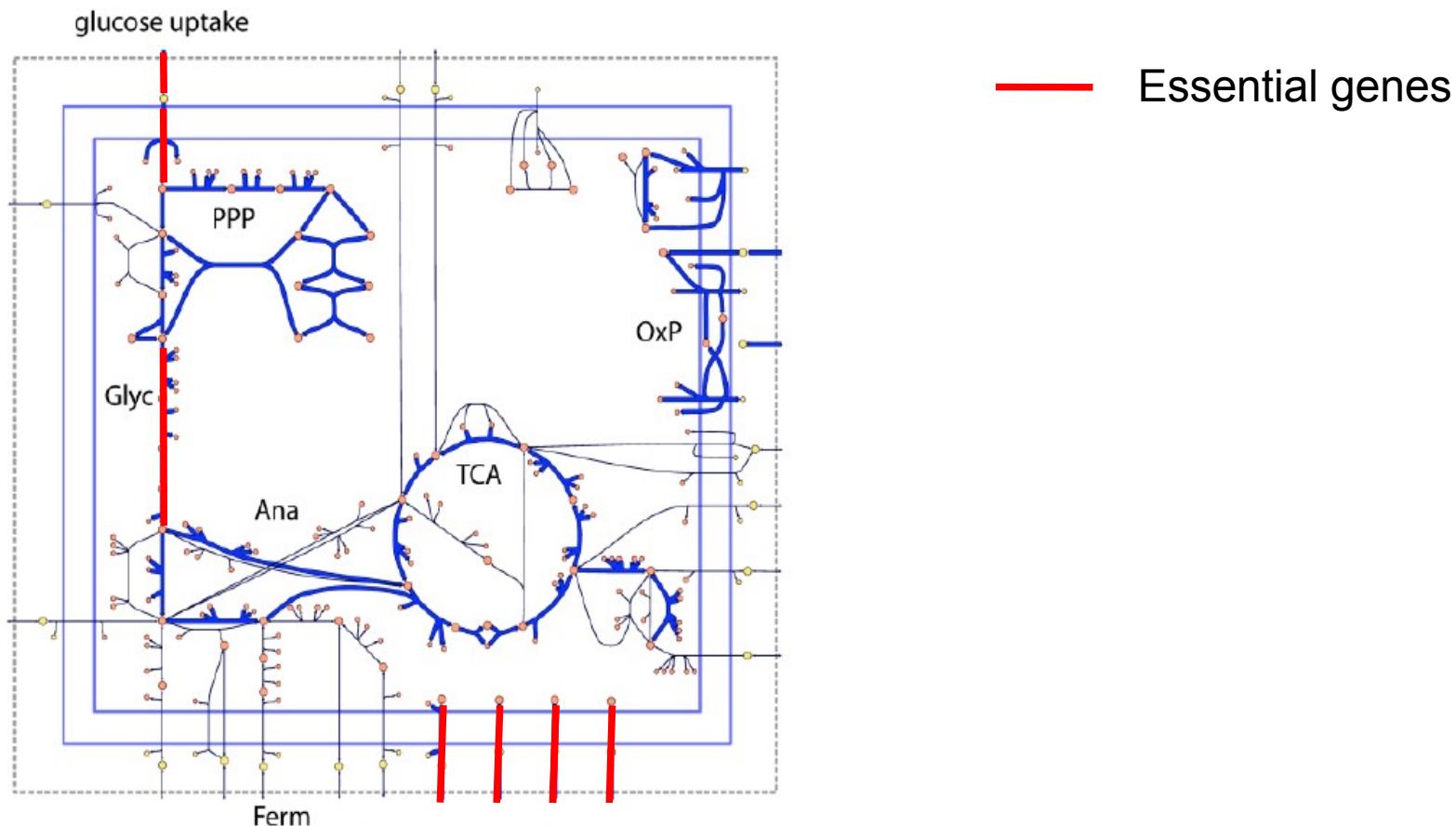
# Environmental constraints



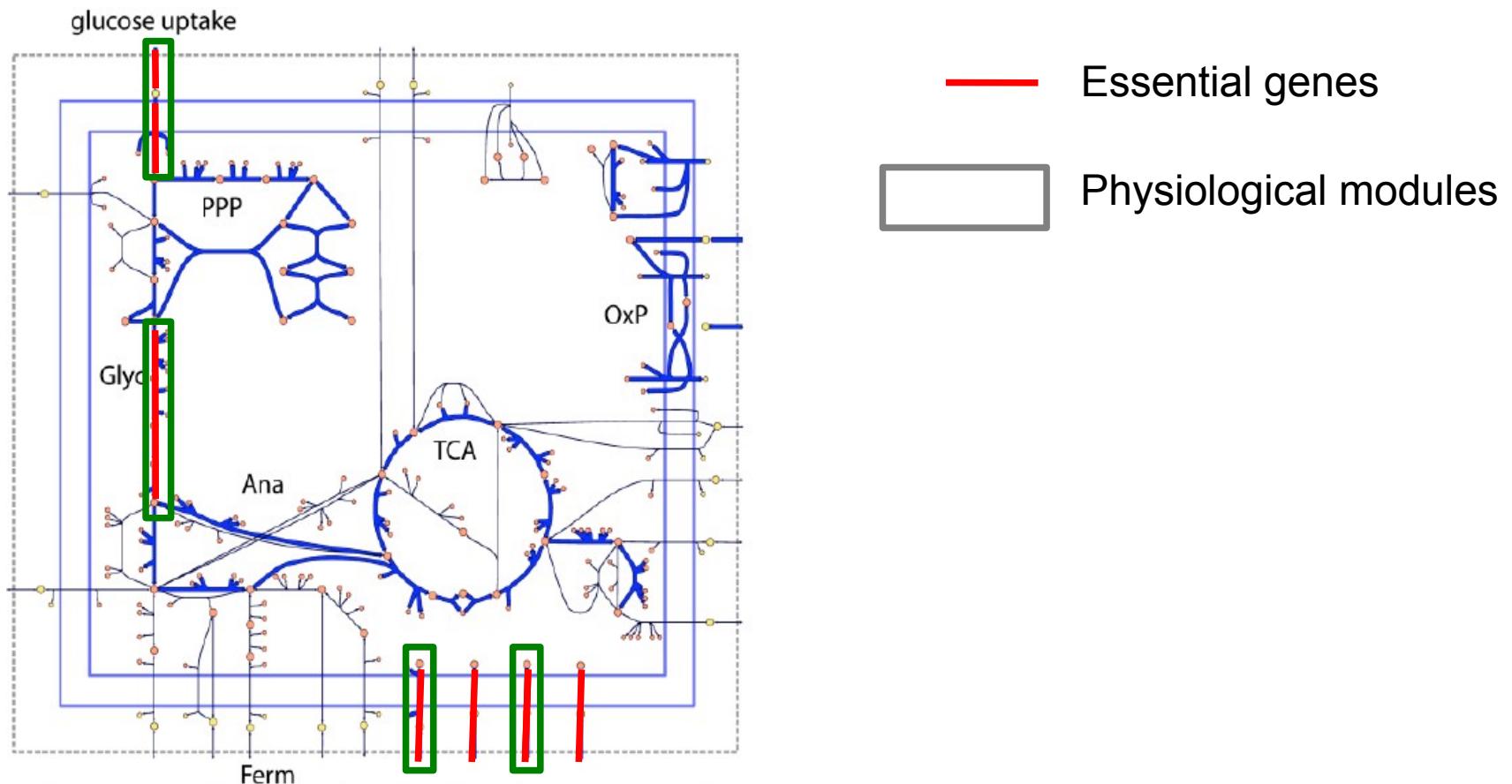
# Genetic constraints



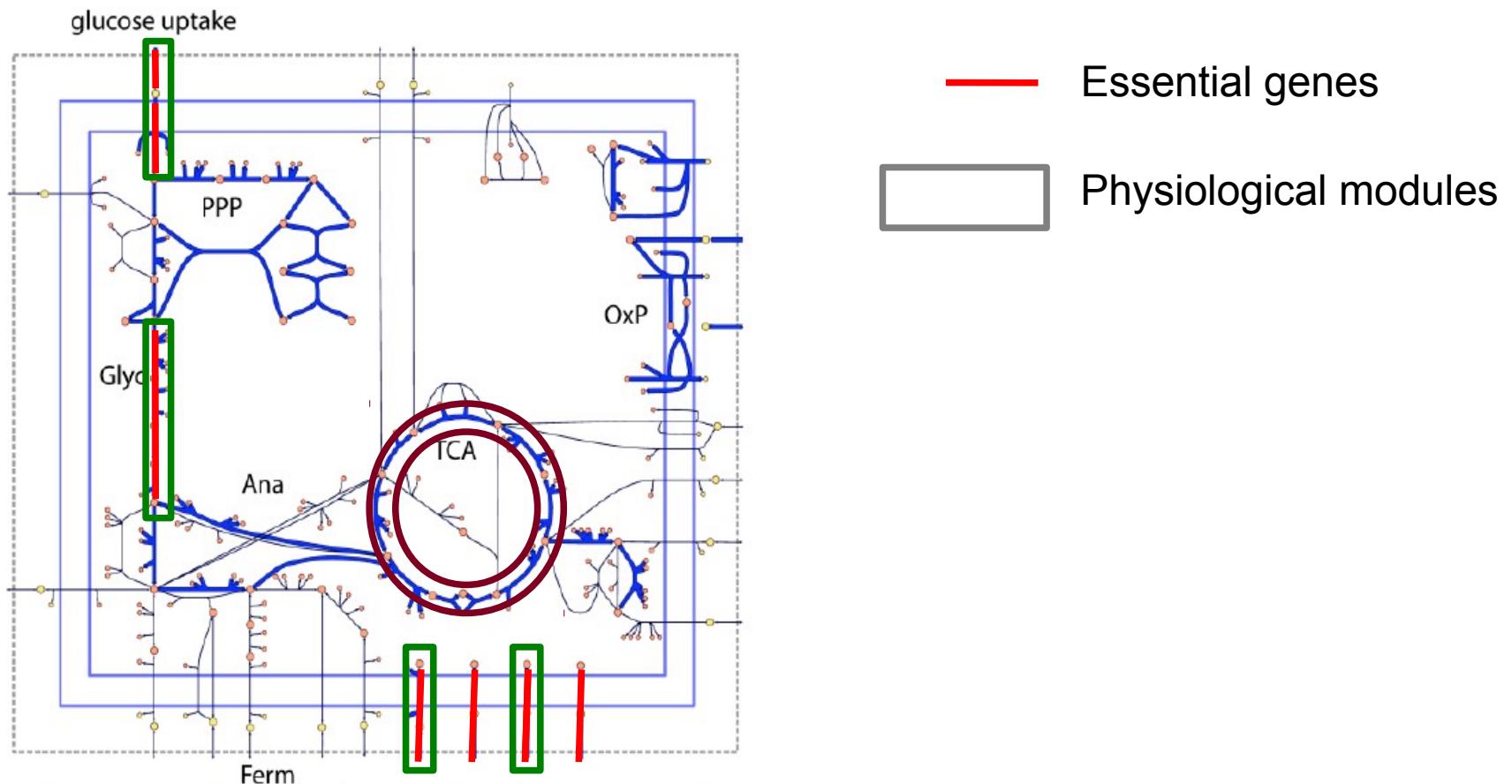
# Essential genes



# Physiological modules

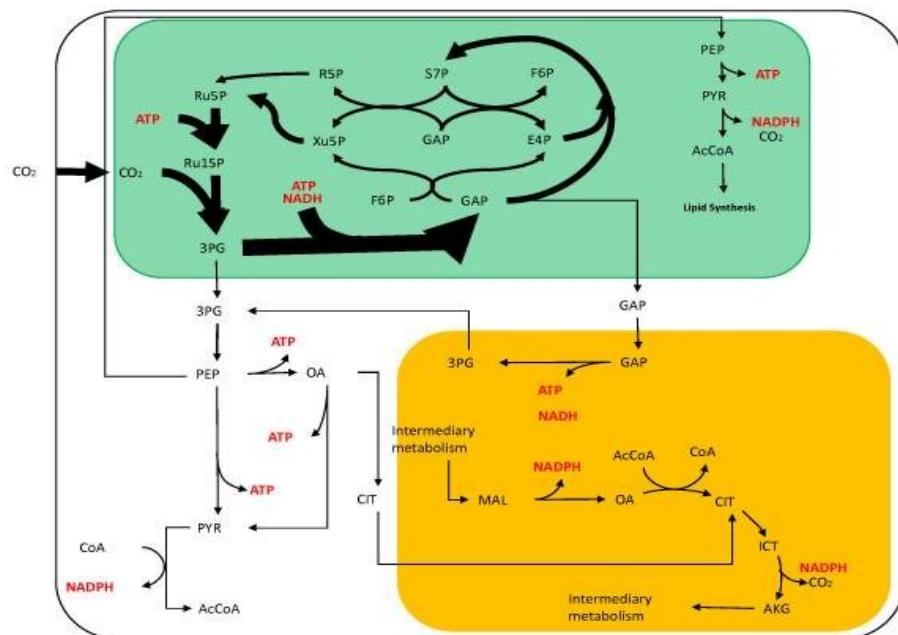


# Physiological modules

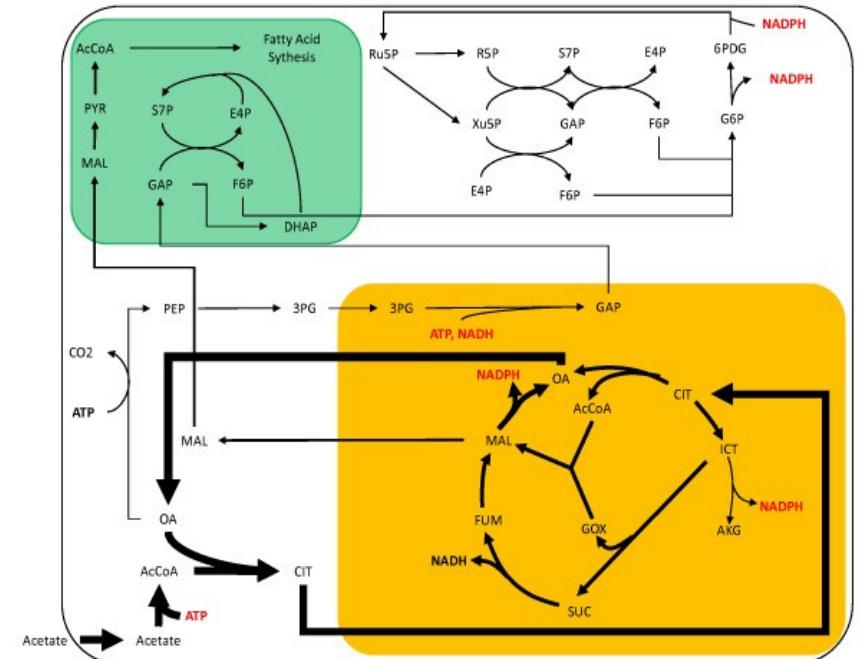


# Redistribution of fluxes

## Autotrophic metabolism



## Heterotrophic metabolism



# Model checking (1)

## Essential genes

		<i>Experimental data</i>	
		Growth	Essential
<i>In silico</i>	Growth		
	Essential	FP	
		FN	

False positives (FP)  
Possible explanation:  
-Missing regulatory rule  
-Falsely included reaction  
-Incomplete biomass reaction

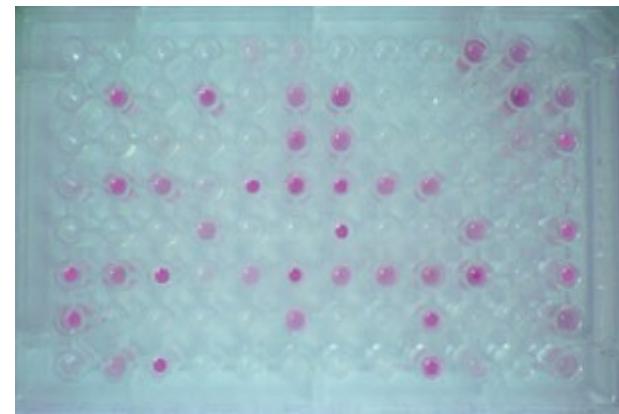
False negatives (FN)  
Possible explanation:  
-Missing metabolic transport reaction  
-Missing enzyme reaction

# Model checking (2)

## Biolog profile

a.

Carbon Source	BIOLOG substrate utilization	in silico growth prediction
no carbon source	X	X
trehalose	X	X
D-cellobiose	X	X
D-mannose	X	X
2,3-butanediol	X	X
D,L-carnitine		X
L-phenylalanine	X	
L-threonine	X	
4-aminobutyrate		
acetate		
citrate		
D-alanine		
D-fructose		
D-glucose		
D-mannitol		
gluconate		
glycerol		
L-alanine		
L-aspartate		
L-histidine		
D,L-lactate		
L-leucine		
L-ornithine		
L-proline		
L-serine		
oxoglutarate		
putrescine		
succinate		
L-asparagine		
malonate		
L-glutamic acid		



b.

Intracellular compounds in iMO1056 that lack transporters:

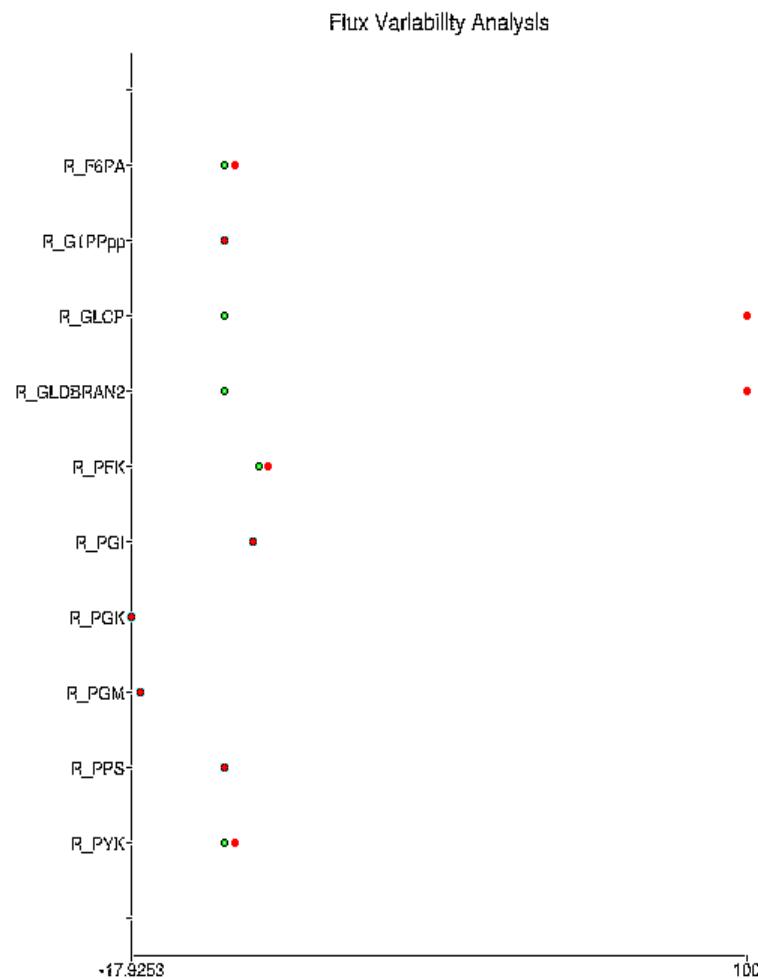
iMO1056 grows with added transporter?

Y                    N

BIOLOG indicates substrate oxidation?	iMO1056 grows with added transporter?	
	Y	N
Y	3 Incorrectly lacks transporter	2 Incorrectly lacks pathway and transporter.
N	6 Correctly lacks transporter	4 Correctly lacks pathway or transporter

# Flux variability analysis

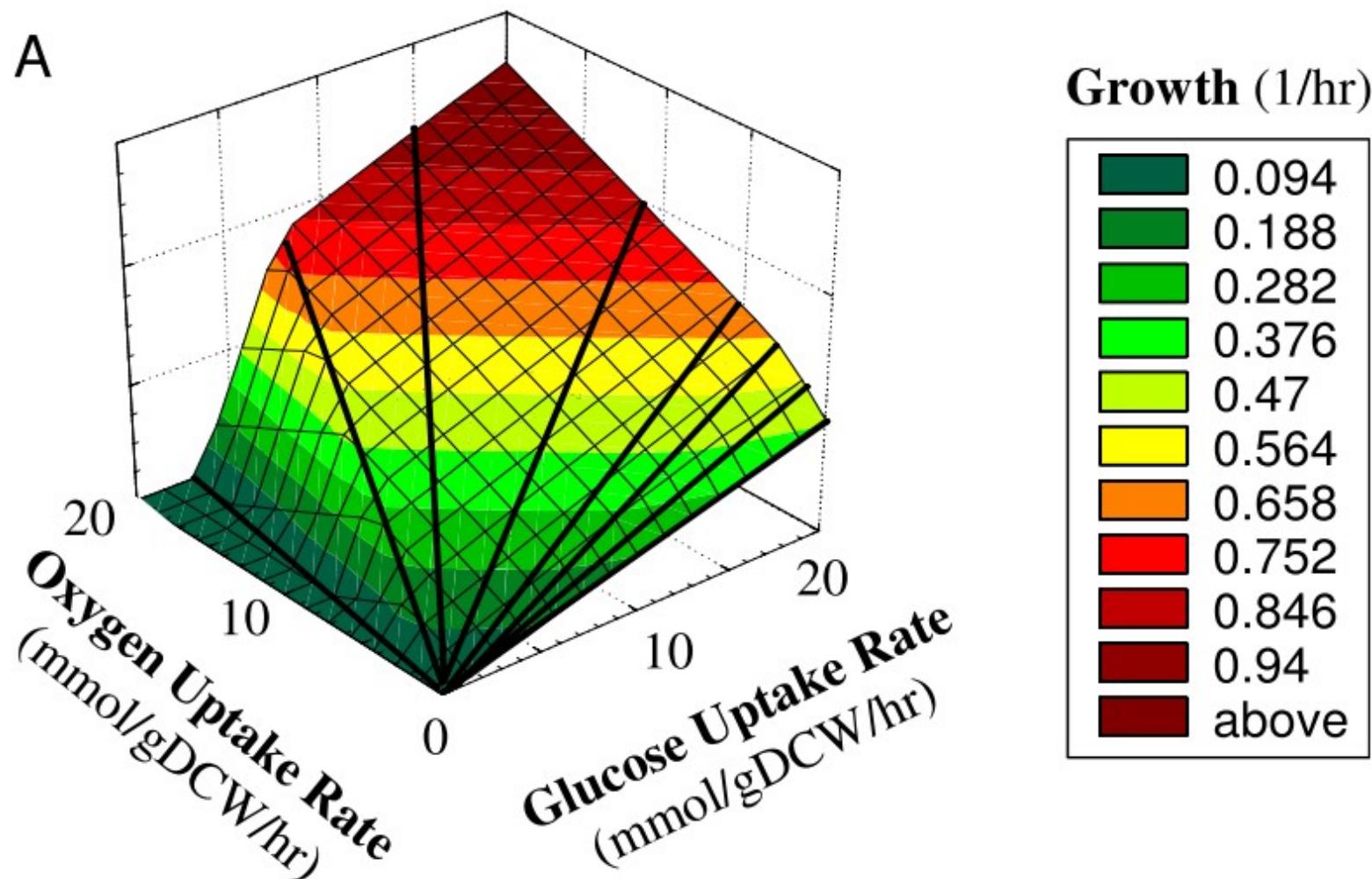
---



For make plot clearer, the max value is set to 100 and the min value is set to -100.

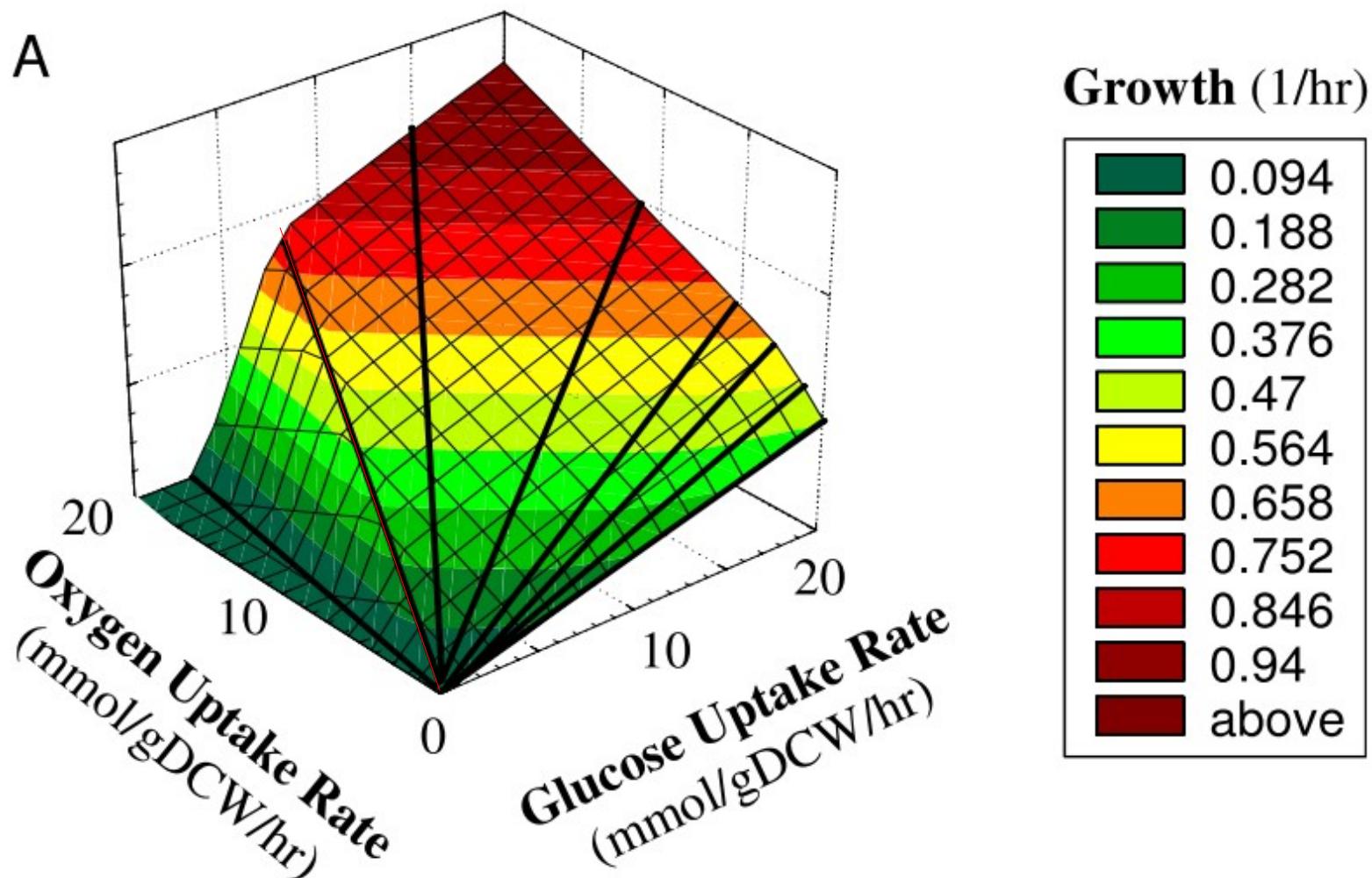
---

# Phenotypic phases



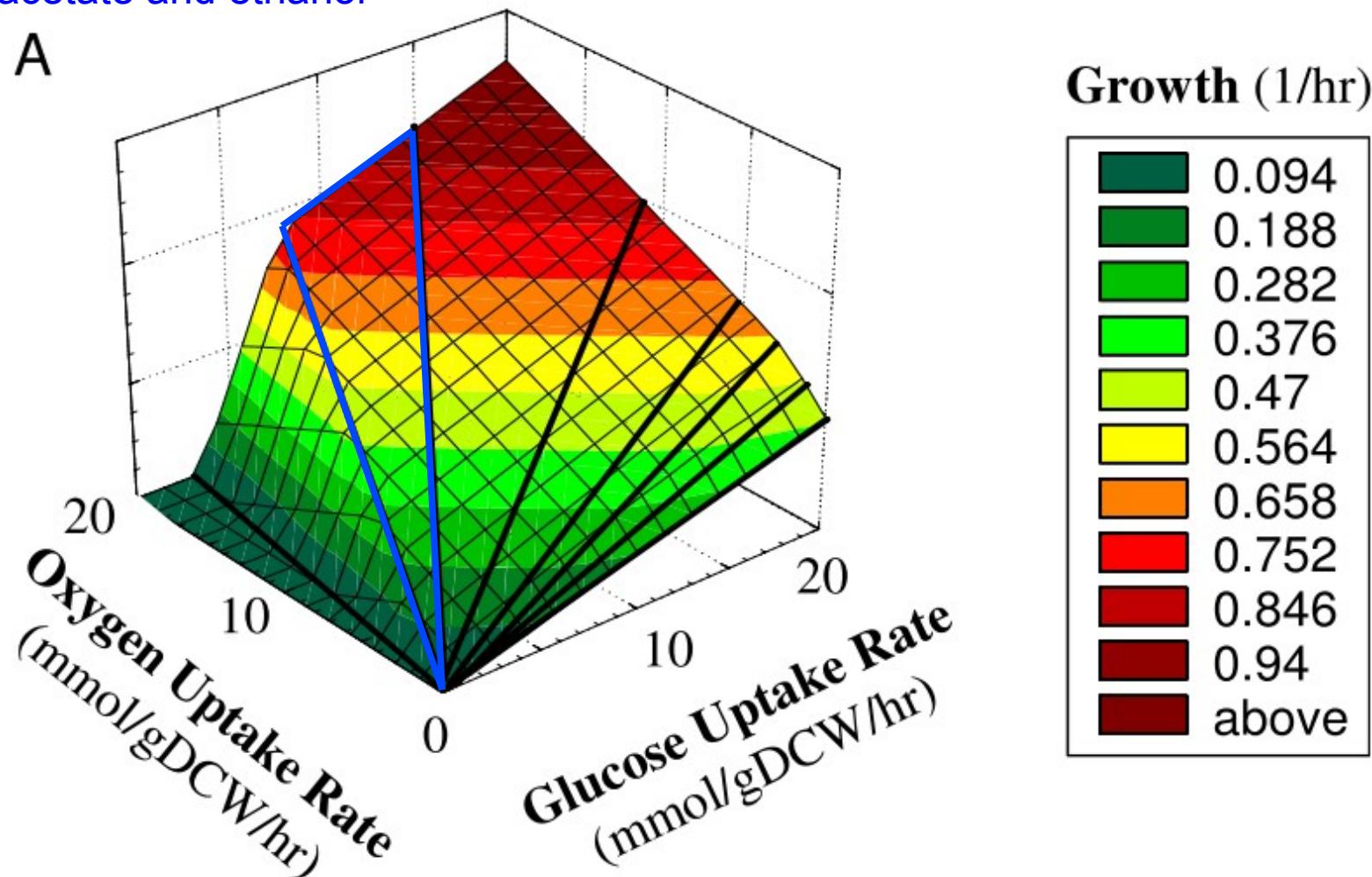
# Phenotypic phases

Glucose is completely oxidized to produce biomass



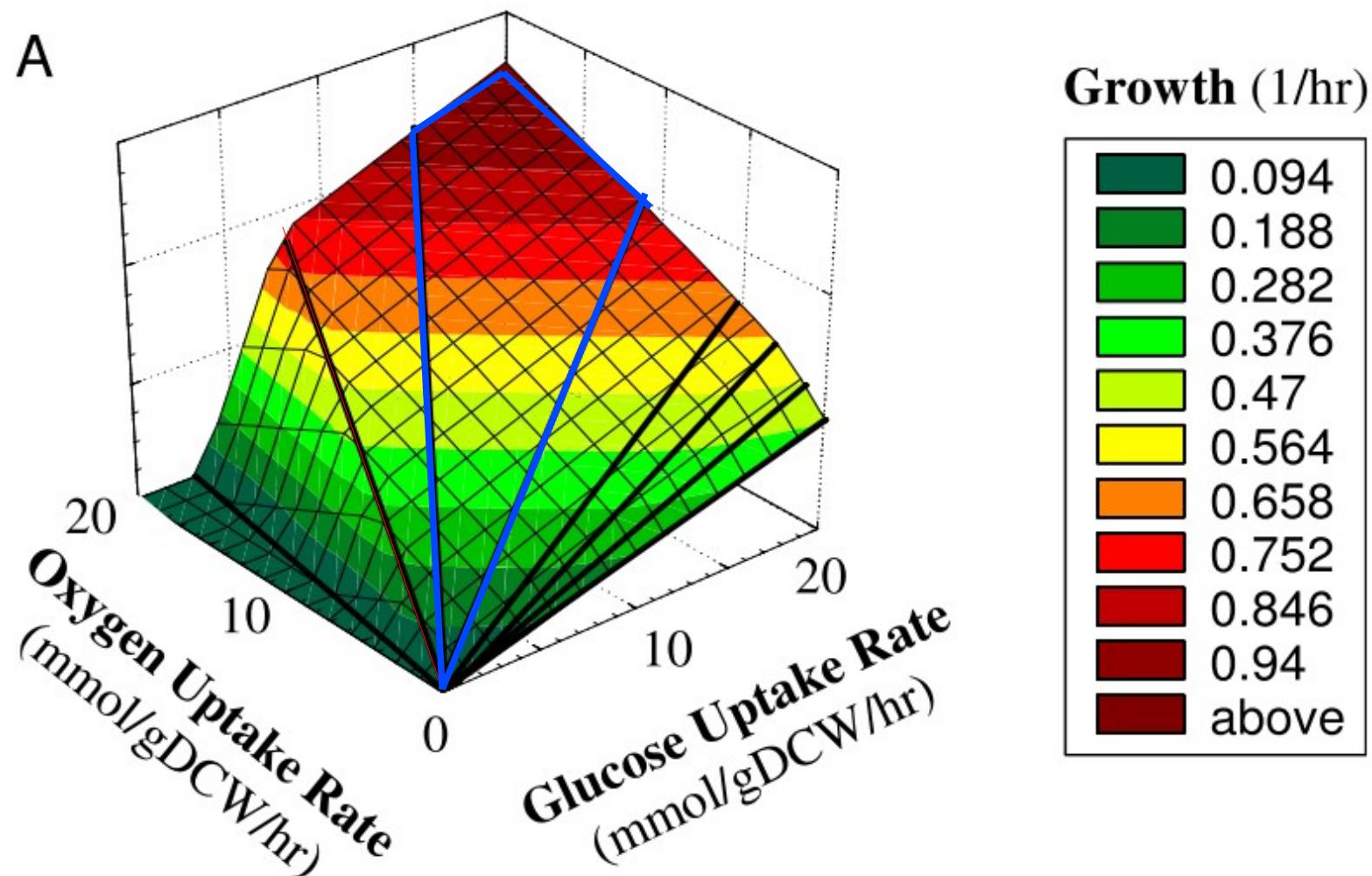
# Phenotypic phases

Oxygen limited. Excess in mitochondrial NAD+ → production of acetate and ethanol



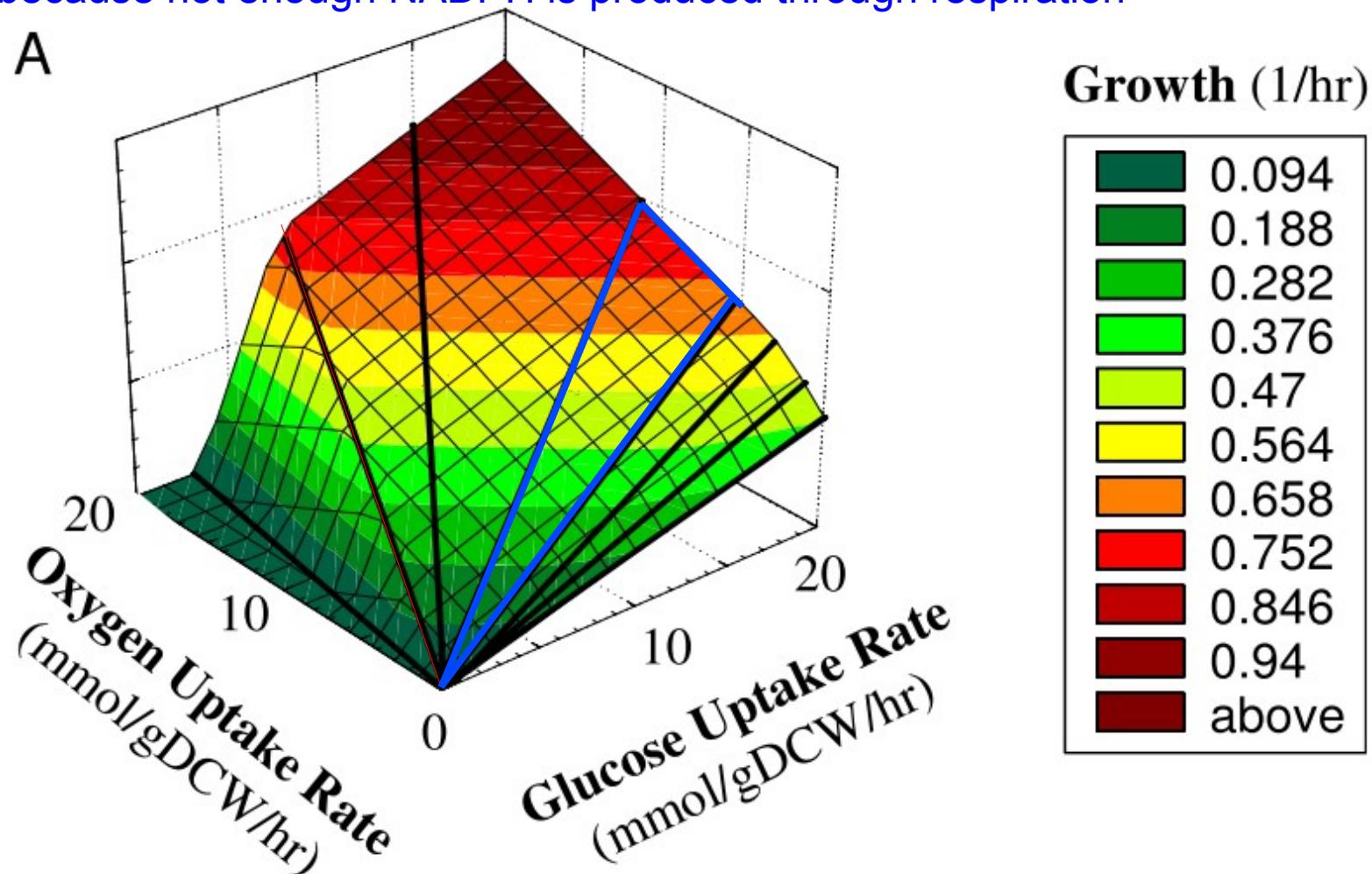
# Phenotypic phases

Three glycolysis reactions become essential



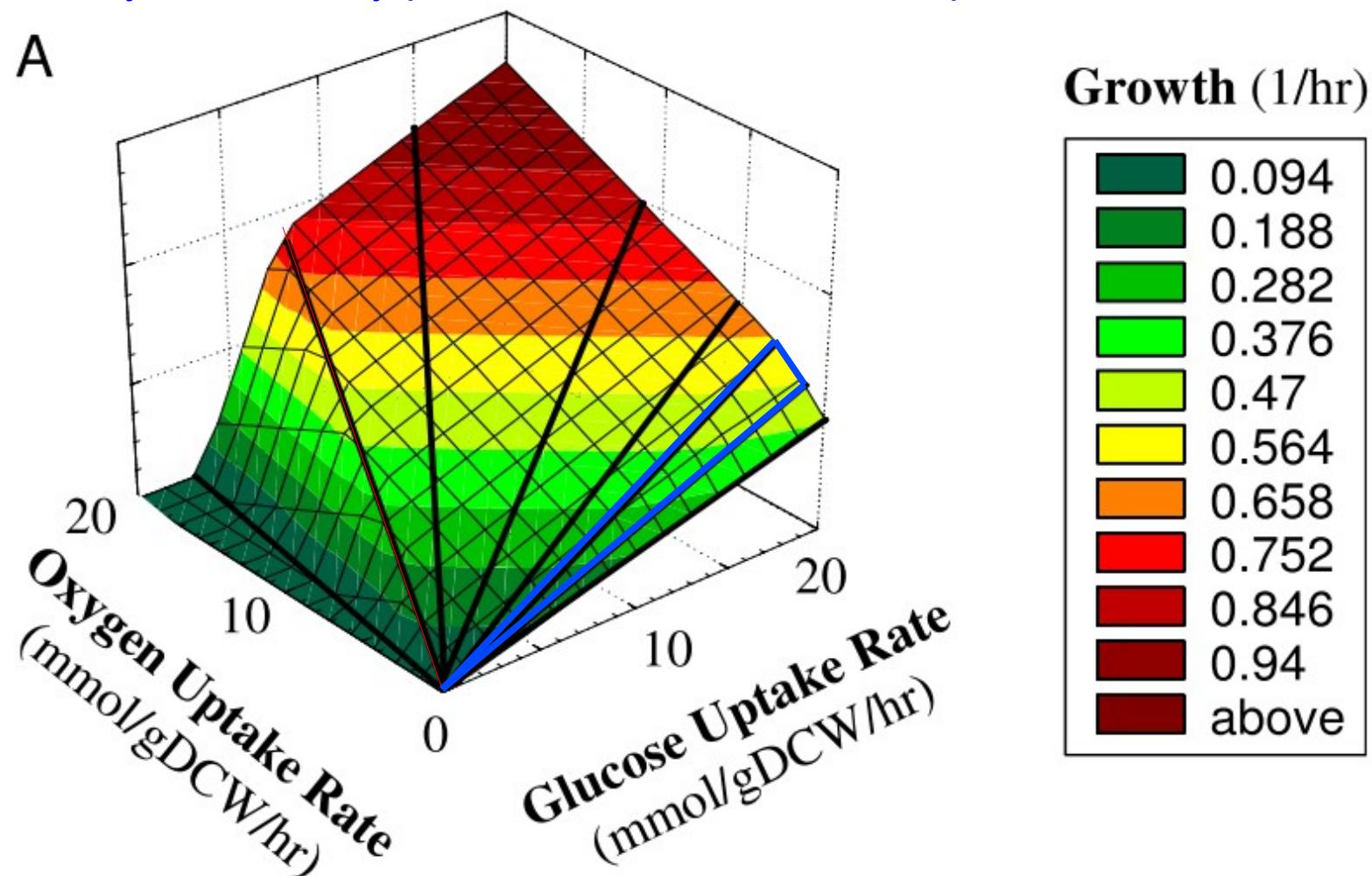
# Phenotypic phases

The pentose phosphate pathway is used to generate NADPH because not enough NADPH is produced through respiration

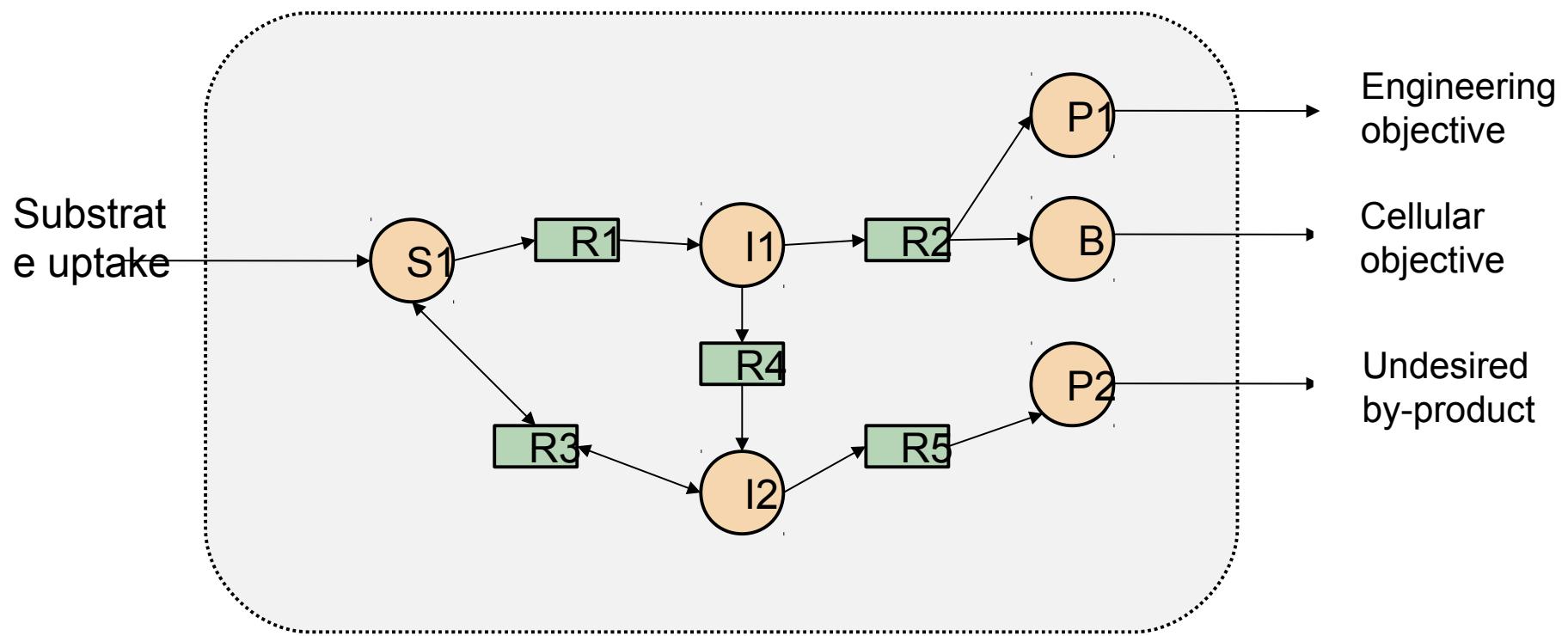


# Phenotypic phases

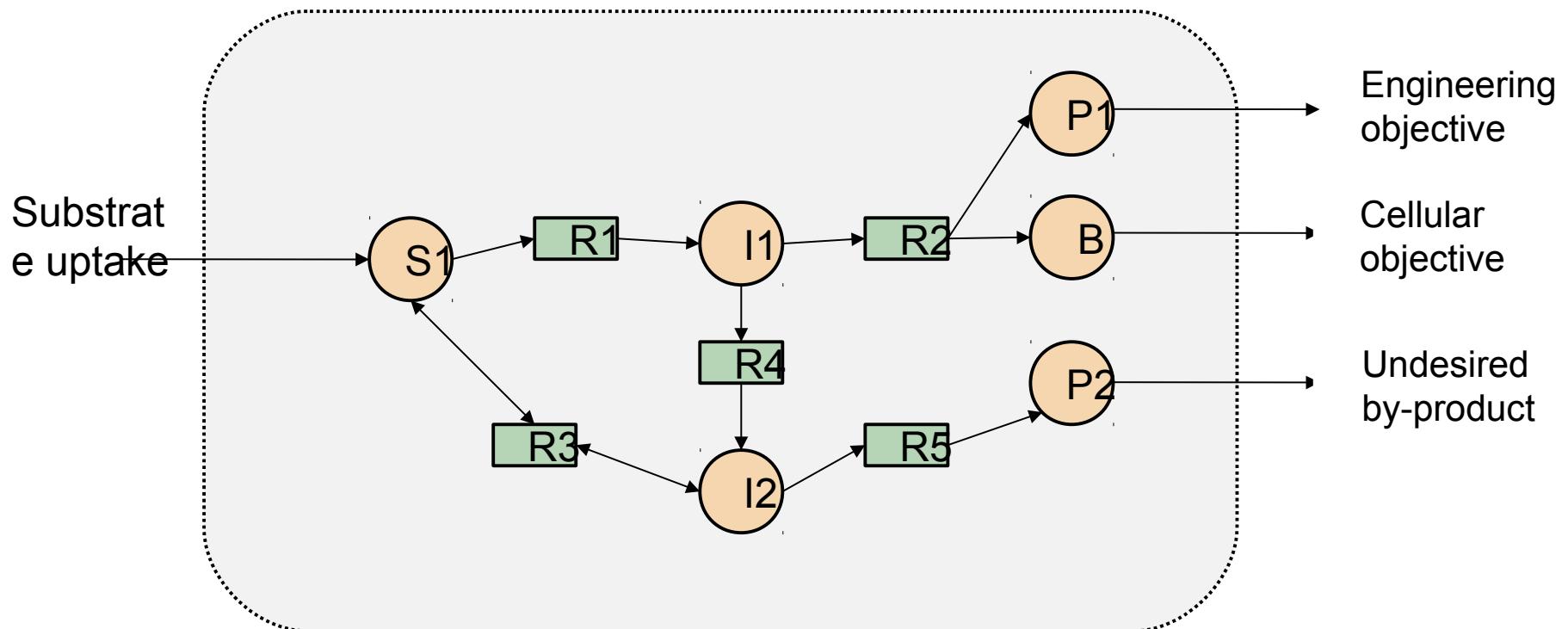
Acetate is not produced anymore. Ethanol is secreted as the only metabolic by-product to balance the redox potential of the cell.



# Optimisation de souches microbiennes par knock-outs

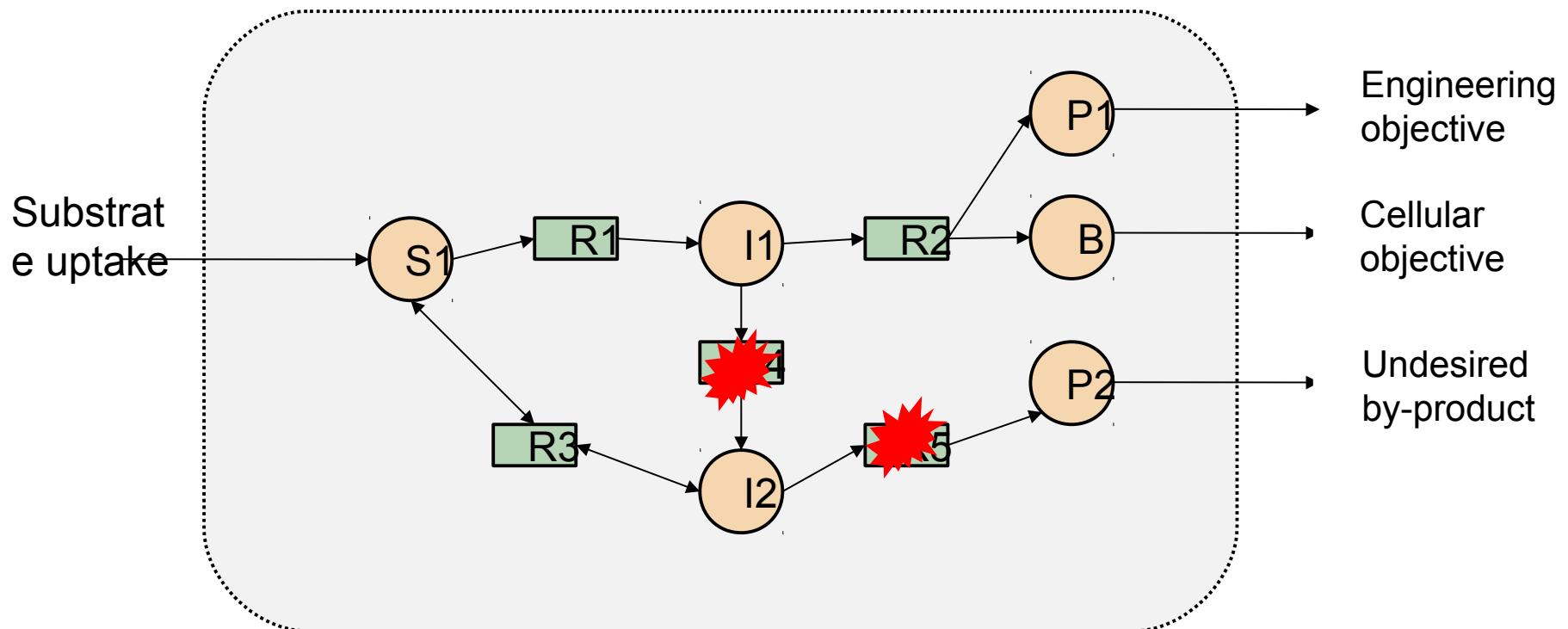


# Optimisation de souches microbiennes par knock-outs



OptKnock method identifies reaction deletions leading to the overproduction of metabolites while ensuring the growth maintenance

# Optimisation de souches microbiennes par knock-outs



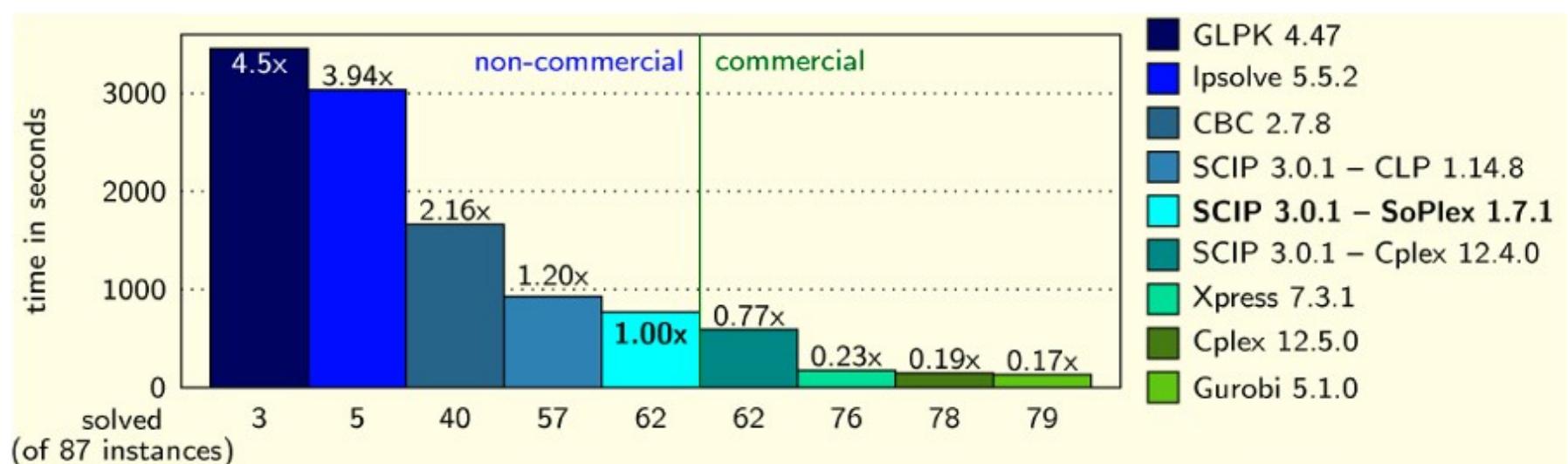
OptKnock method identifies reaction deletions leading to the overproduction of metabolites while ensuring the growth maintenance

# How to make your own FBA tool ?

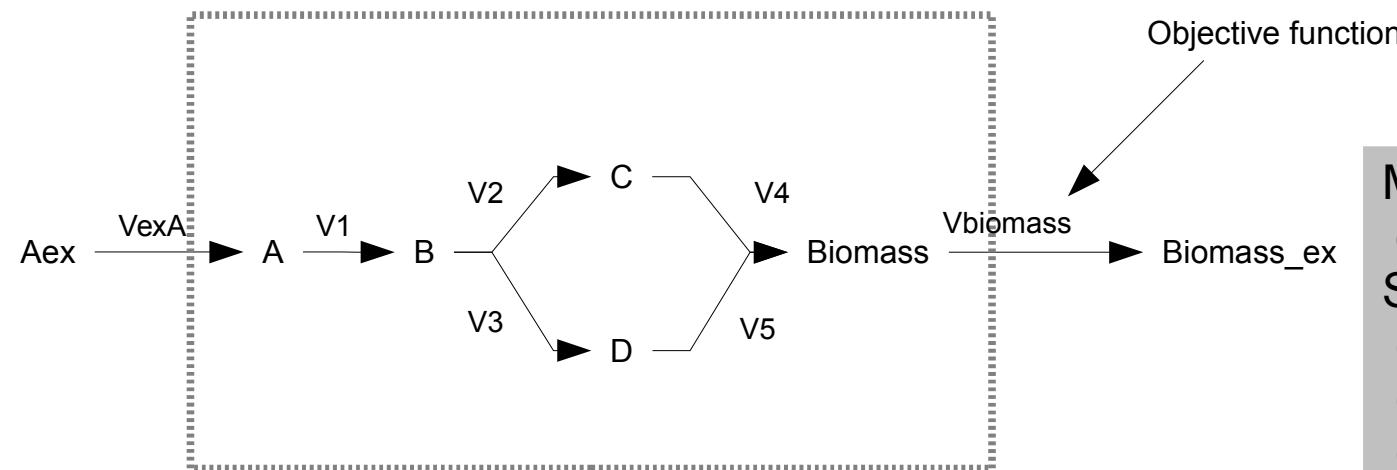
## (1)

# By using optimizers !

- CPLEX
  - GLPK
  - SCIP
  - Gurobi.



# Flux balance analysis formulation problem. Example



Objective function

```
Maximize
obj: v_biomass
Subject To
c1: v_ex_a - v1 = 0
c2: v1 - v2 - v3 = 0
c3: v2 - v4 = 0
c4: v3 - v5 = 0
c5: - v_biomass + v4 + v5 = 0
Bounds
0 <= v_biomass <= 10000
0 <= v_ex_a <= 10
0 <= v1 <= 9
0 <= v2 <= 8
-6 <= v3 <= 6
0 <= v4 <= 10
0 <= v5 <= 100
End
```

# The tools

- Cobra-Toolbox
- FASIMU
- Surrey-FBA
- Opt-Flux
- MetExplore
- FlexFlux

# Cobra-toolbox

- Version Matlab and Python
- Widely used
- Numerous tools
- Sometimes badly coded

# Cobra-Toolbox

```
initCobraToolbox
model=xls2model('central_metabolism_default.xls')
FBAsolution=optimizeCbModel(model)
edit draw_by_rxn
[Involved_mets,Dead_ends]=draw_by_rxn(model,model.rxns,true,'struc',{''},{''},FBAsolution.x)
[Involved_mets,Dead_ends]=draw_by_rxn(model,[],true,'struc',{''},{''},FBAsolution.x)
edit draw_by_met
[Involved_Rxns,Involved_mets,Dead_ends]=draw_by_met(model,['etoh(c)'],true,2,'sub',{''},FBAsolution.x)
clear
initCobraToolbox
model=xls2model('central_metabolism_default.xls')
FBAsolution=optimizeCbModel(model)
[Involved_mets,Dead_ends]=draw_by_rxn(model,model.rxns,true,'struc',{''},{''},FBAsolution.x)
```

OVR

```
>>> import cobra.test
>>> model = cobra.test.create_test_model()
>>> pgi = model.reactions.get_by_id("PGI")
>>> print pgi.name
glucose 6 phosphate isomerase
```

# FASIMU

- Based on batch scripts
- Many functions
- Generates many intermediate files
- Very slow

```
unzip FASIMU_complete.zip

cd FASIMU_Liver_Example

sbml2fa liver.sbml

source fasimu

prune-network

cp MIMES.txt MIPES.txt PIPES.txt sub

cd sub

unzip ../../FASIMU_complete.zip\

FASIMU_Liver_Example/simulations

source fasimu

simulate

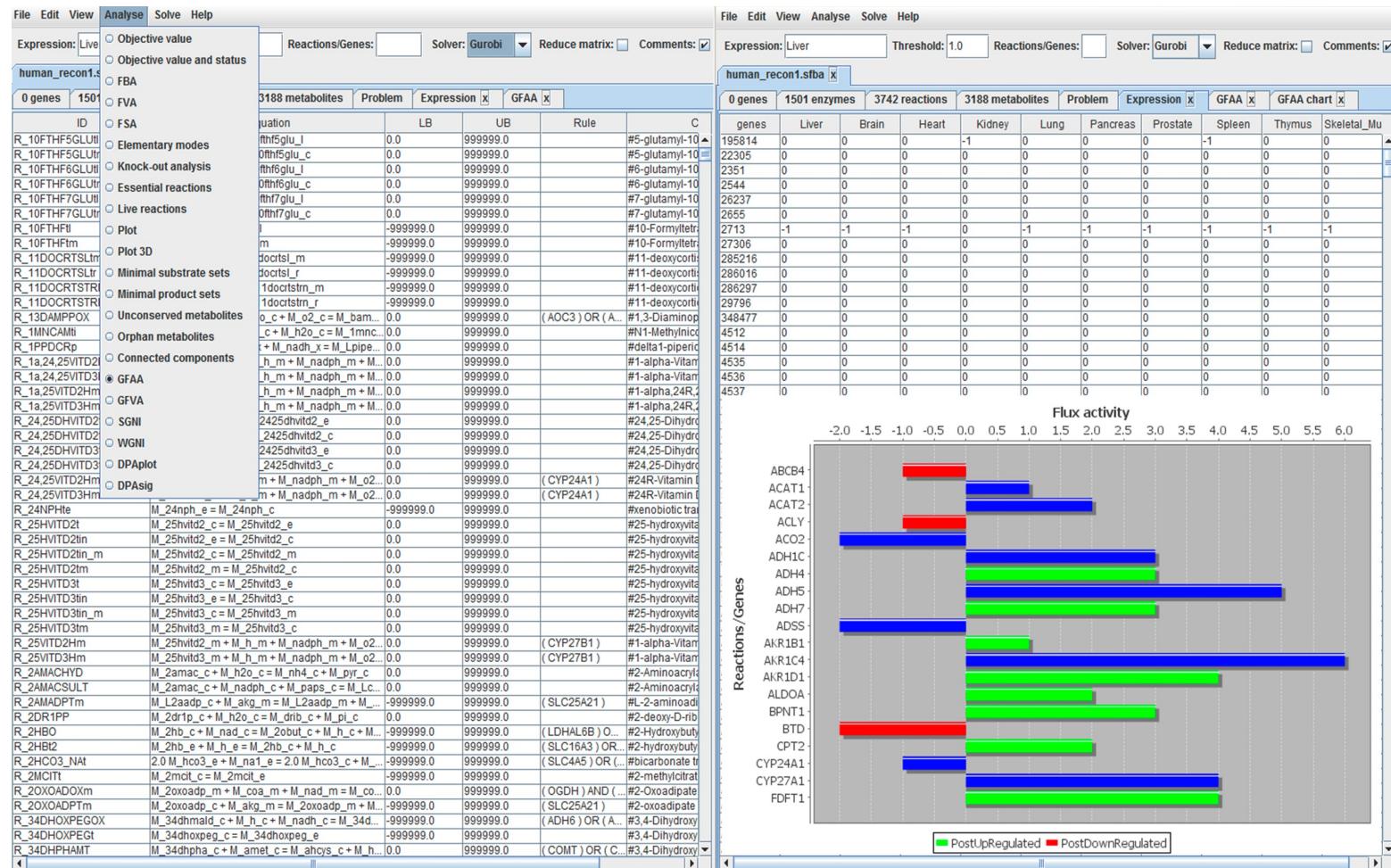
optimization_call="compute-FBA-T-c"

simulate
```

# SurreyFba

- Developed in C
- Java interface JyMet
- Quite slow
- Maintained ?

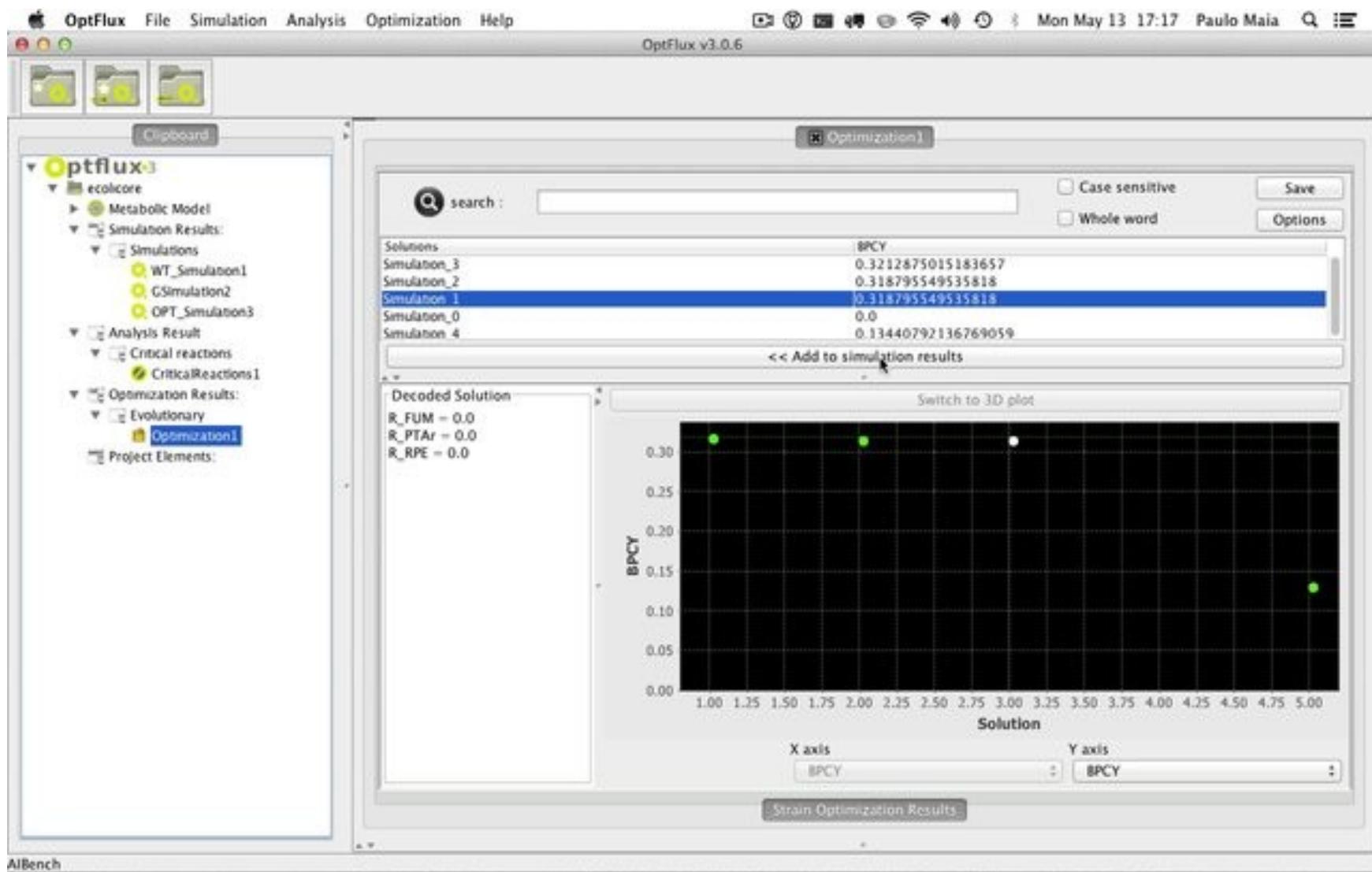
# SurreyFBA



# OptFlux

- User friendly graphical interface
- Workspace concept
- Stores the environmental and genetic constraints

# OptFlux



# MetExplore

- Integrates Surrey-FBA functions
- Facilitates the tuning of the flux constraints
- Facilitates the visualisation via Cytoscape

# MetExplore

## Flux analysis parameters

### Reaction parameters and objective function building

Search:  Clear

Reaction names	Reaction identifiers	Lower bound	Upper bound	Participation to the objective function	Type of reactions	Reaction formula with metabolite names	Reaction formula with metabolite ids	Gene association	Reversible ?	Pathways	Reactions selected for analysis
R_H_exchange	R_EX_h_e_	-999999.00	999999.000	0	Exchange_in	$1.000000 M\_H\_H = 1.000000 M\_H\_H$	$1.000000 M\_h\_e = 1.000000 M\_h\_e$		T	S_-	<input type="checkbox"/>
R_Calcium_exchange	R_EX_ca2_e_	-999999.00	999999.000	0	Exchange_in	$1.000000 M\_Calcium\_Ca = 1.000000 M\_Calcium\_Ca$	$1.000000 M\_ca2\_e = 1.000000 M\_ca2\_e$		T	S_-	<input type="checkbox"/>
R_Mg_exchange	R_EX_mg2_e_	-999999.00	999999.000	0	Ex						
R_CO2_exchange	R_EX_co2_e_	-999999.00	999999.000	0	Ex						
R_Cu2_exchange	R_EX_cu2_e_	-999999.00	999999.000	0	Ex						
R_Sulfate_exchange	R_EX_so4_e_	-999999.00	999999.000	0	Ex						
R_Molybdate_exchange	R_EX_mobd_e_	-999999.00	999999.000	0	Ex						
R_D_Glucose_exchange	R_EX_glc_e_	-11.000000	999999.000	0	Ex						
R_Cob_I_lamin exchange	R_EX_chl1_e_	.....	.....	.....	Ex						

Cytoscape Desktop (New Session)

Continuous Editor for Edge Line Width

Continuous Mapping for Edge Line Width

Range Setting

Graphical View

Mapping Type

Edge Line Width

Visual Mapping

Edge Color

Node Visual Mapping

Node Label

Node Label Color

Node Shape

Node Size

Unused Properties

Edge Font Face

Edge Font Size

Edge Label

Edge Label Color

Edge Label Opacity

Edge Line Style

Edge Opacity

Data Panel

Welcome to Cytoscape 2.6.2 Right-click + drag to ZOOM Middle-click + drag to PAN

ID GeneAssociation LowerBound MaxFlux MinFlux ReactionName Type

R\_PFK "( b1854 ) or ( b1676 )" "0.000000" "1.84086" "0" "R\_pyruvate\_kinase" "Internal" "99

M\_13\_dpg\_c M\_13\_dpg\_c

M\_f6p\_c M\_f6p\_c

M\_pep\_c M\_pep\_c

R\_PFK "( b1723 ) or ( b3916 )" "0.000000" "8.32459" "-6.48373" "R\_phosphofructokinase" "Internal" "99

M\_py\_c M\_py\_c

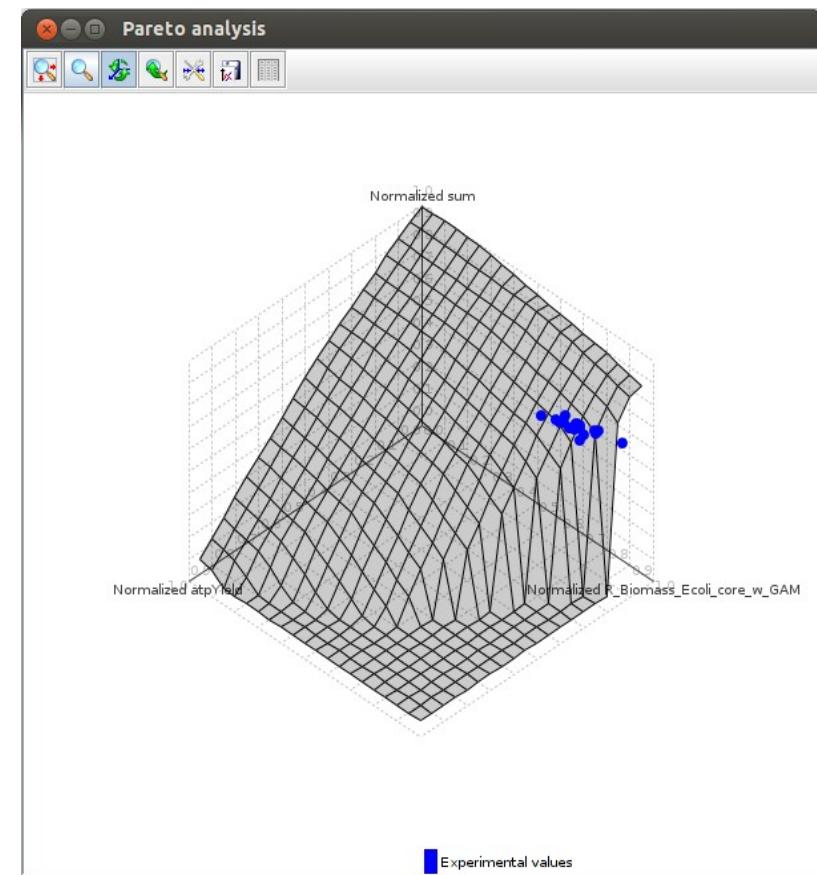
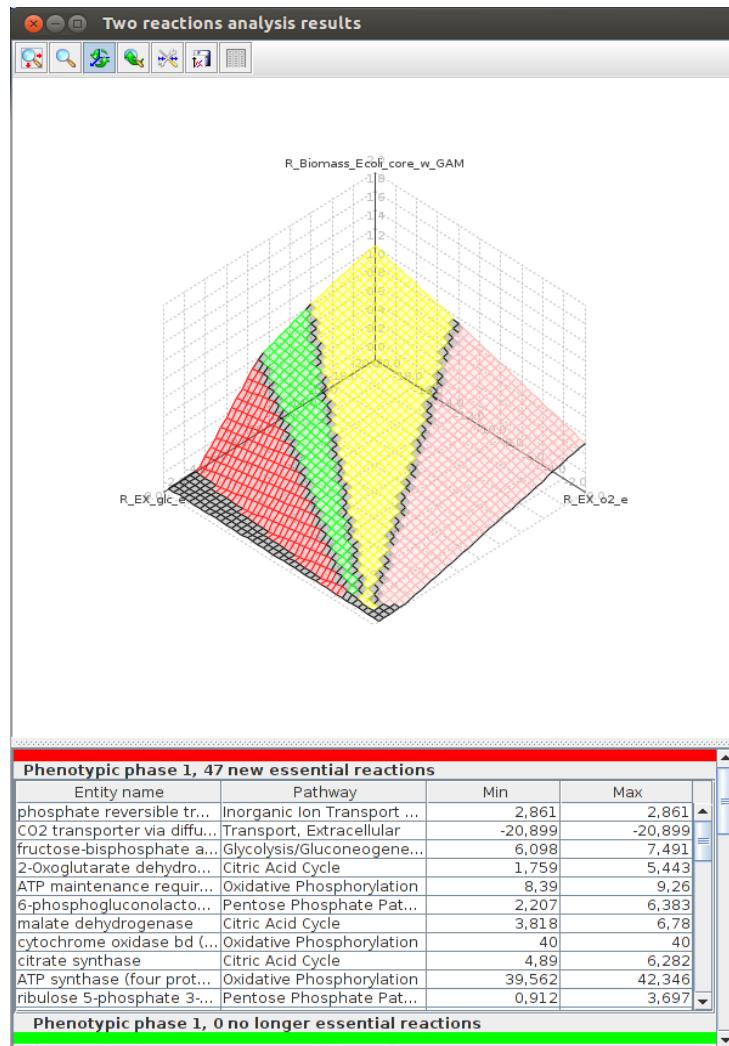
R\_PKG "( b2926 )" "-999999.000000" "-17.9253" "-17.9253" "R\_phosphoglycerate\_kinase" "Internal" "99

M\_h\_c M\_h\_c

# FlexFlux

- Java library
- Developed in a modular way
- Natively integrates regulatory constraints
- Very fast

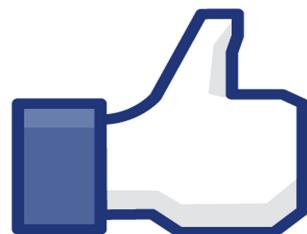
# FlexFlux



# Flux analysis : advantages and drawbacks

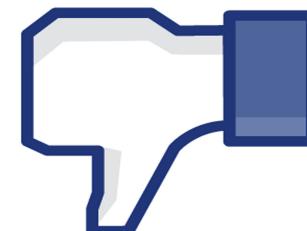
**Semi-quantitative analysis without kinetic parameters**

**Allows to answer complex questions**



**Needs a well curated metabolic network**

**Results highly depend on the *a priori* constraints**



# KEGG

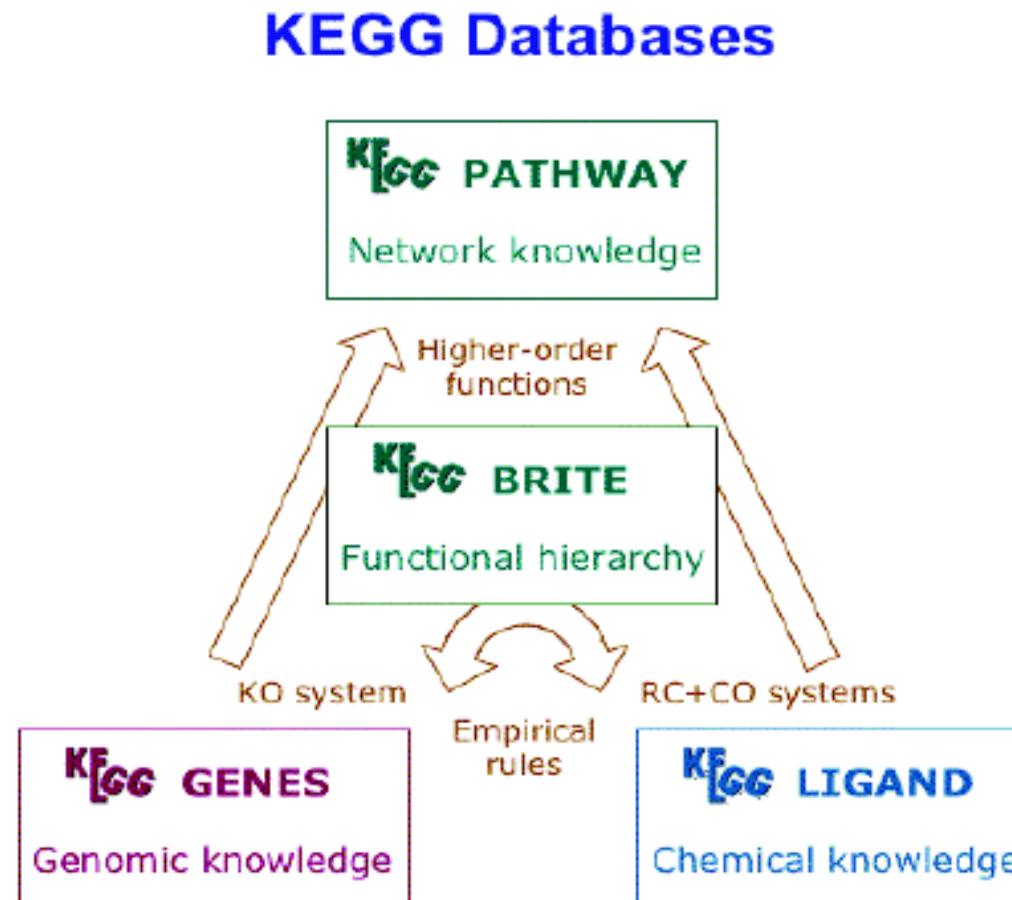


[www.genome.jp/kegg/](http://www.genome.jp/kegg/)

# KEGG en quelques mots

- Projet initié en 1995 à l'université de Kyoto
- Définition (Kanehisa 2006) : base de connaissances pour comprendre les fonctions de haut niveau d'un système biologique, à partir d'informations génomiques et chimiques

# KEGG : une vue très simplifiée



# KEGG ORTHOLOGY (KO)

Chaque gène de KEGG GENES est classé selon la fonction pour laquelle il code.

Ainsi, les gènes orthologues et paralogues qui ont gardé la même fonction se verront attribuer le même numéro KO.

La classification KO se fait en fonction des voies métaboliques présentes dans KEGG.

# KEGG ORTHOLOGY (KO)

---

01100 Metabolism

  01110 Carbohydrate metabolism

  01120 Energy metabolism

  01130 Lipid metabolism

  01140 Nucleotide metabolism

  01150 Amino acid metabolism

    00251 Glutamate metabolism

.....

  00300 Lysine biosynthesis

    K00003 E1.1.1.3, thrA; homoserine dehydrogenase

    K00928 E2.7.2.4, lysC; aspartate kinase

    K00133 E1.2.1.11, asd; aspartate-semialdehyde dehydrogenase

    K01714 E4.2.1.52, dapA; dihydridopicolinate synthase

    K00215 E1.3.1.26, dapB; dihydridopicolinate reductase

    K00674 E2.3.1.117, dapD; 2,3,4,5-tetrahydropyridine-2-carboxylate *N*-succinyltransferase

    K00821 E2.6.1.17; *N*-succinyl-diaminopimelate aminotransferase

    K01439 E3.5.1.18, dapE; succinyl-diaminopimelate desuccinylase

    K01778 E5.1.1.7, dapF; diaminopimelate epimerase

    K01586 E4.1.1.20, lysA; diaminopimelate decarboxylase

.....

  00310 Lysine degradation

.....

01160 Metabolism of other amino acids

.....

01200 Genetic information processing

01300 Environmental information processing

01400 Cellular processes

01500 Human disease

---

# KEGG PATHWAY

Collection de cartes dessinées manuellement représentant des voies métaboliques, des réseaux de gènes ou des voies de transduction du signal. 156,646 cartes générées à partir de 416 voies métaboliques de référence.

# KEGG PATHWAY

**KEGG** Lysine biosynthesis - Reference pathway

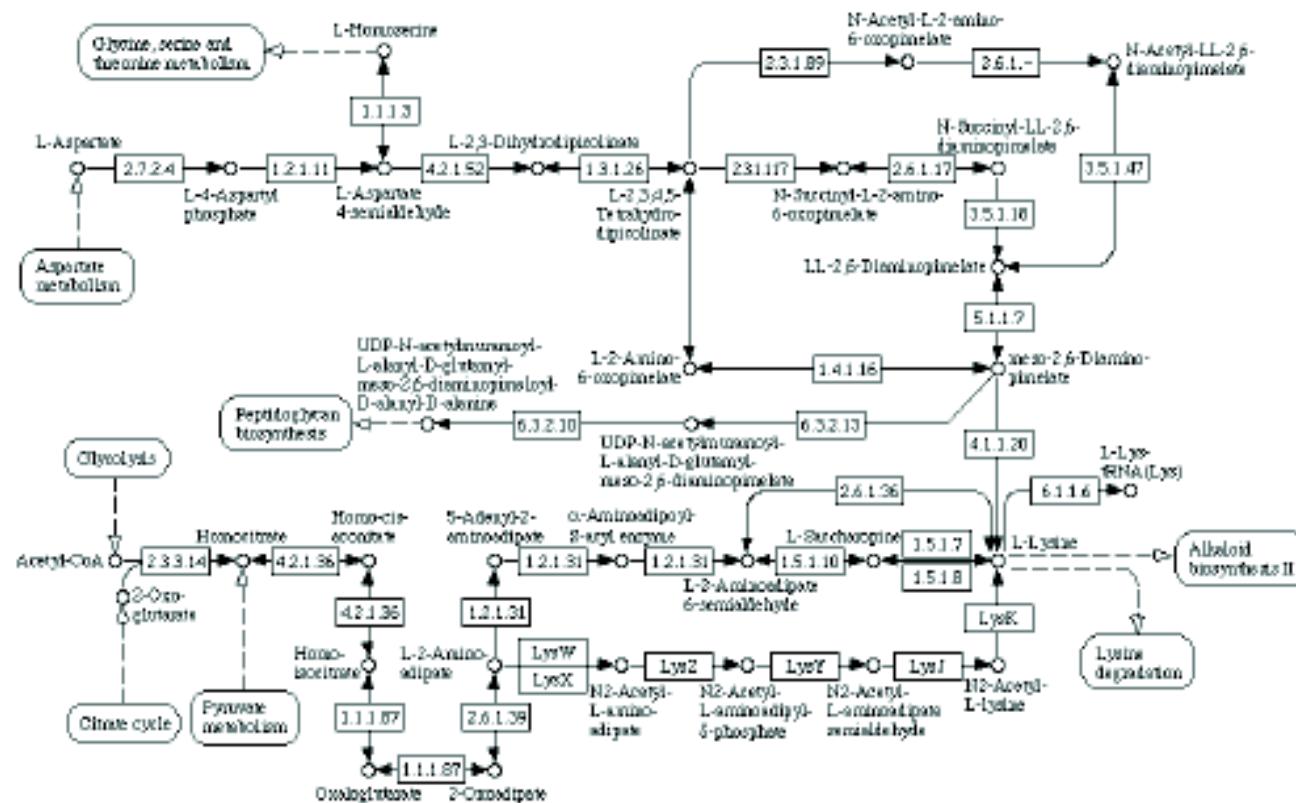
[Pathway menu | Reference list | Ortholog table ]

Reference pathway

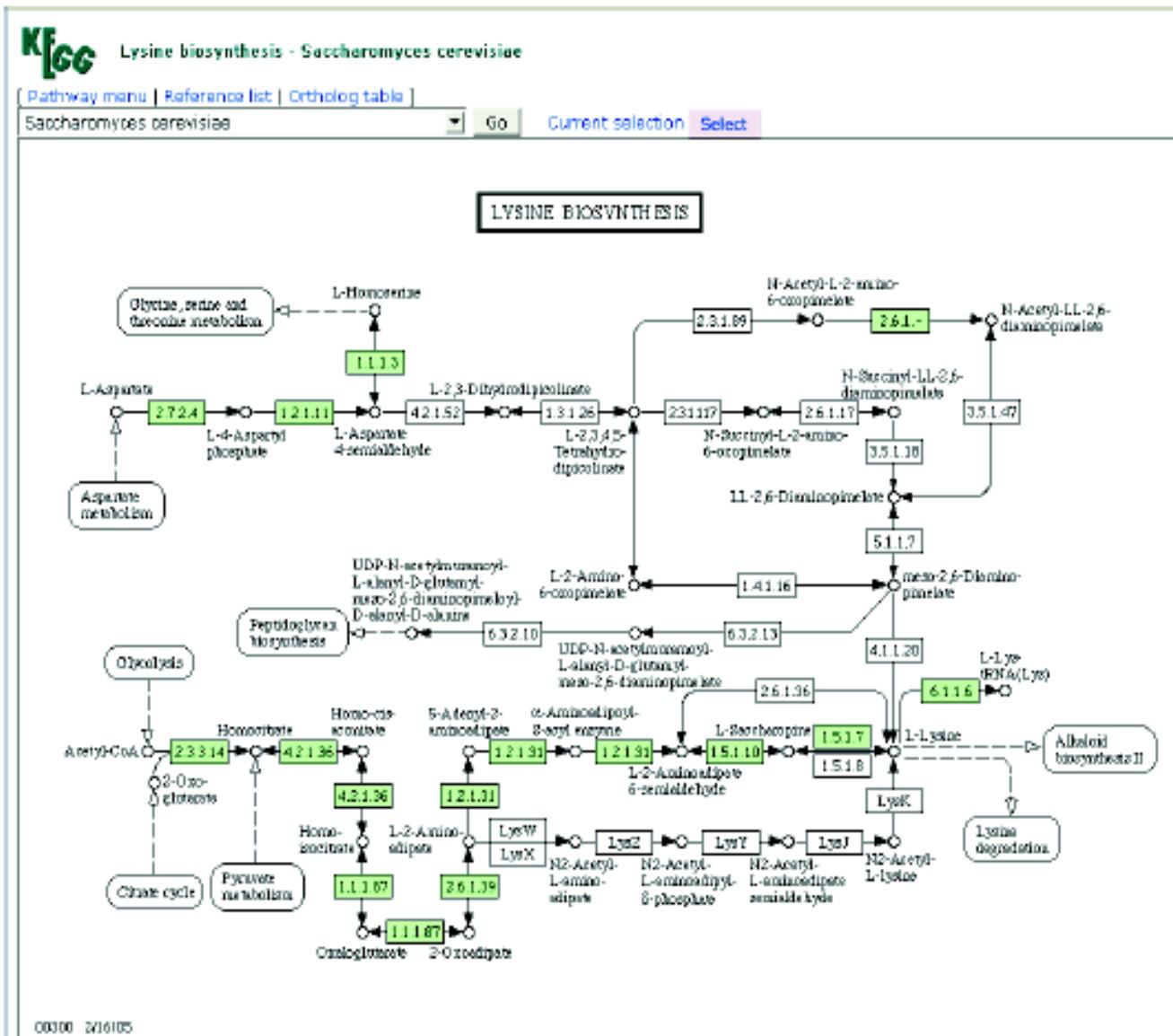
Go

Current selection Select

## LYSINE BIOSYNTHESIS



# KEGG PATHWAY



# KEGG : outils associés

- KEGG API : service WEB utilisant SOAP (Simple Object Access Protocol) et WSDL (Web Services Description Language). Permet de développer des scripts Ruby, Perl, Java, etc...
- KAAS : Annotation fonctionnelle d'un ensemble de gènes par BLAST
- SIMCOMP / SUBCOMP : Recherche de structures de composés similaires
- e-zyme : Assignation de numéros EC à partir d'un ensemble de substrats et de produits.
- Blast, Fasta
- MOTIF : recherche de motifs dans les séquences protéiques (recherche de domaines) et dans les séquences nucléiques (recherche de régions promotrices)
- CLUSTALW / MAFFT / PRRN : alignements multiples
- KGML : Format d'échange (XML)

# KEGG : avantages et limitations

- + Nombreux organismes
- + Informations précises sur les composés
- + Nombreux outils associés
- + Reconstruction des données métaboliques possible à partir d'un génome non annoté
- - Difficile de s'y retrouver dans le site
- - Manque d'expertise

# BIOCYC

<http://biocyc.org/>

# BioCyc en quelques mots

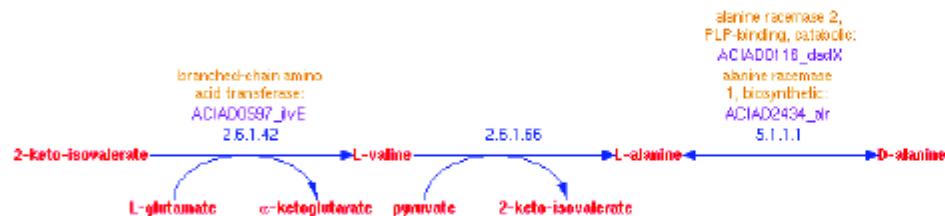
- Université de Stanford (1994)
- Collection de PGDBs (Pathway Genome DataBase) spécifique à un organisme (mis à part Metacyc)
- Trois niveaux d'expertise :
  - 6 bases à haut niveau d'expertise (Ecocyc, Metacyc, HumanCyc, AraCyc, YeastCyc, LeishCyc)
  - 34 bases à moyen niveau d'expertise
  - 1653 bases sans expertise
- Partage des bases et expertise par d'autres groupes (Aracyc, HumanCyc...)

# BioCyc : les voies métaboliques



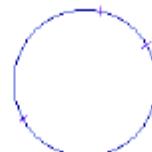
## A. baylyi Pathway: alanine biosynthesis I

[More Detail](#) [Less Detail](#) [Cross-Species Comparison](#)



Locations of Mapped Genes:

48



Product: alanine racemase 2, PLP-binding, catabolic / Alanine racemase); "onmouseout="return nd();">

Superclasses: [Biosynthesis](#) -> [Amino acids](#) -> [Individual amino acids](#) -> [Alanine](#)

Net Reaction Equation: pyruvate + glutamate = L-alanine + 2-ketoglutarate

Superpathways: [superpathway of leucine, valine, alanine, and isoleucine biosynthesis](#), [superpathway of alanine biosynthesis](#)

Variants: [alanine biosynthesis II](#)

Pathway Evidence Glyph:

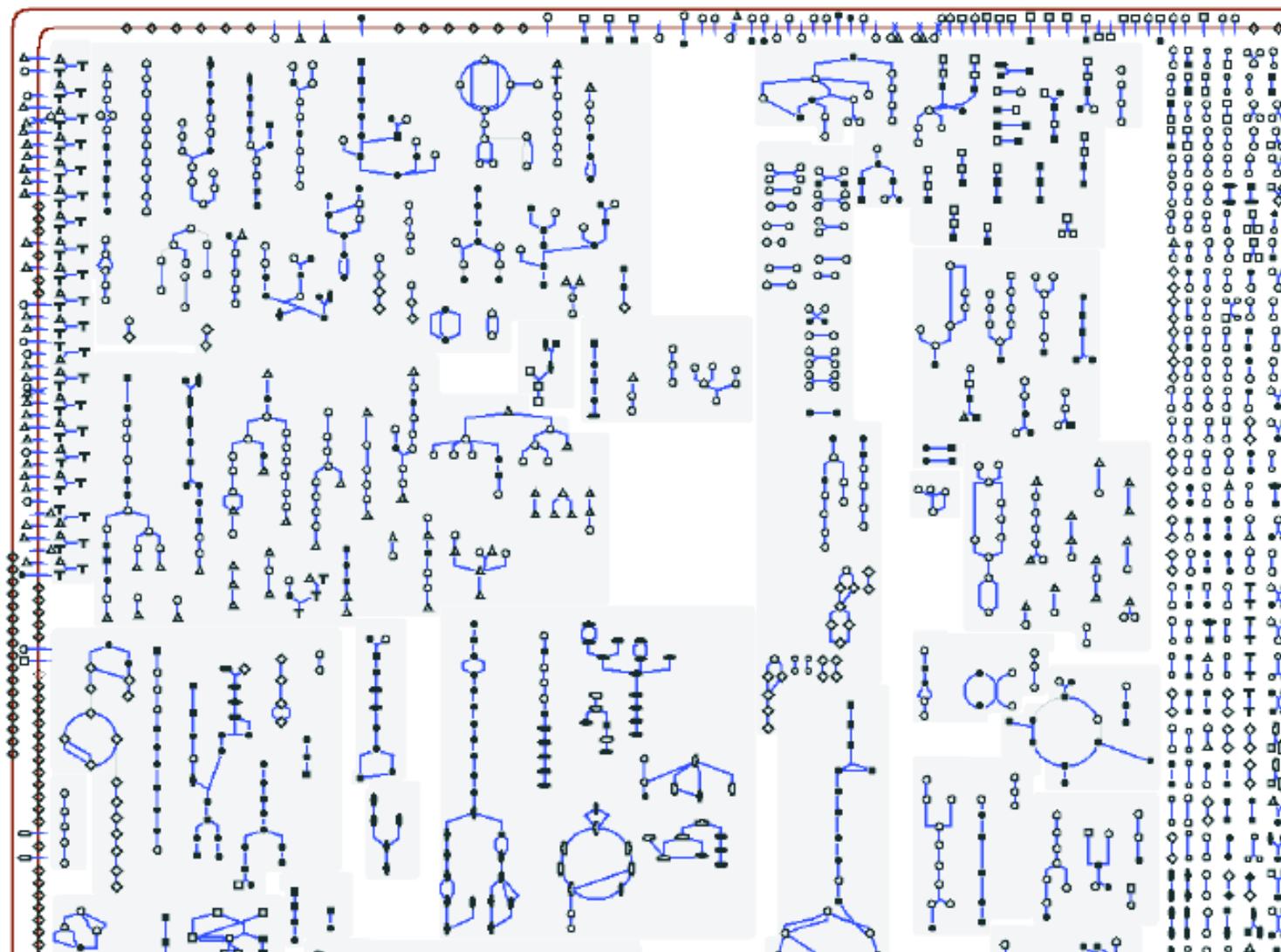


Key to pathway glyph edge colors:

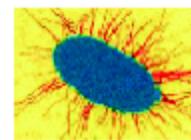
- green: an enzyme catalyzing this reaction is present in this organism
- black: no enzyme catalyzing this reaction has been identified in this organism
- orange: the reaction and any enzyme that catalyzes it (if one has been identified) is unique to this pathway
- magenta: represents spontaneous reactions, or lines that do not represent reactions (e.g. in polymerization pathways)

[Query Page](#) [Advanced Query Page](#) [Report Errors or Provide Feedback](#)

# BioCyc : vue d'ensemble de la cellule



# BioCyc : les réactions



**E. coli K-12 Reaction: 3.1.3.-**

**Cross-Species Comparison**

Superclasses: [EC-Reactions](#) -> [3 -- Hydrolases](#) -> [3.1 -- Acting on ester bonds](#) -> [3.1.3 -- Phosphoric monoester hydrolases](#)

[phosphohistidine phosphatase](#) ; [sixA](#)

In Pathway: [ArcAB Two-Component Signal Transduction System](#)



The reaction direction shown, that is, A + B  $\rightleftharpoons$  C + D versus C + D  $\rightleftharpoons$  A + B, is in accordance with the direction of the reaction within a pathway.

Gene-Reaction Schematic: [?](#)



[Query Page](#)

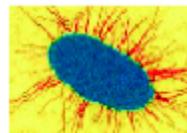
[Advanced Query Page](#)

[BioCyc Home](#)

[Report Errors or Provide Feedback](#)

Please cite the following article in publications resulting from the use of EcoCyc: [Nucleic Acids Research 33:D334-7 2005](#)  
Page generated by SRI International [Pathway Tools version 10.0](#) on Thu Apr 27, 2006.  
EcoCyc version 10.0.

# BioCyc : les composés

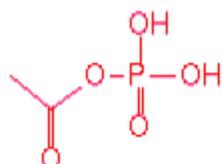


*E. coli K-12 Compound: acetylphosphate*

Synonyms: acetyl-P

Empirical Formula: C<sub>2</sub>H<sub>5</sub>O<sub>5</sub>P

Molecular Weight: 140.03



Smiles: CC(=O)OP(=O)(O)O

Unification Links: CAS:590-54-5 , CAS:19926-71-7 , [LIGAND:c00227](#) , [LIGAND:CD0227](#)

In Pathway Reactions as a Product:

[acetate utilization](#):

[acetate + ATP = acetylphosphate + ADP](#),  
[phosphate + acetyl-CoA = acetylphosphate + coenzyme A](#)

[mixed acid fermentation](#):

[acetate + ATP = acetylphosphate + ADP](#),  
[phosphate + acetyl-CoA = acetylphosphate + coenzyme A](#)

[Query Page](#)

[Advanced Query Page](#)

[BioCyc Home](#)

[Report Errors or Provide Feedback](#)

Please cite the following article in publications resulting from the use of EcoCyc: [Nucleic Acids Research 33:D334-7 2005](#)

Page generated by SRI International [Pathway Tools version 10.0](#) on Thu Apr 27, 2006.

EcoCyc version 10.0.



# BioCyc : recherches avancées

## 1. Enter your query here:

Query database [Escherichia coli K-12 substr. MG1655](#) for Compounds (2318 instances)

Where

Name (2318 values)  ATP

and  in [Appears-In-Left-Side-Of \(758 values\)](#)

we have

Z2   1.1.1

Select an operation to add an additional search component or variable:

## 2. Select attributes to include in the query output:

Column 1	Column 2 <input type="button" value="X"/>	<input type="button" value="add a column"/>
<input checked="" type="radio"/> Sort based on this column Z1 <input type="button" value="Name (up to 2318 values)"/>	<input type="radio"/> Sort based on this column Z1 <input type="button" value="Chemical-Formula (up to 2212 values)"/>	

## 3. Select query output format:

- HTML  Tab Delimited Text (columns are separated by tabs)

# BioCyc : les outils associés

- Pathway Tools Software : ensemble d'outils de navigation, d'édition et de création de PGDBs
  - Développé en LISP, fonctionne sous Linux, Sun et Mac OS X
  - Des librairies Java et Perl (javacyc et perlcyc) permettent d'utiliser les fonctions LISP internes aux Pathway Tools
  - Pathologic : création de PGDBs locales à partir de Metacyc
  - Edition et transformation de l'ontologie de Biocyc
  - Possibilité de construire facilement un portail internet avec ses propres PGDBs
  - Formats d'échange : fichiers plats, SBML, BioPax
- Outils en ligne :
- Blast, Fasta
- Analyse comparative des données de plusieurs organismes
- Visualisation de données expérimentales sur la vue d'ensemble des voies métaboliques