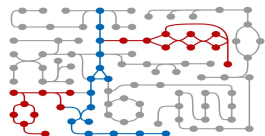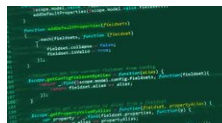# T05 Part 2: Introduction to COBRA
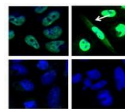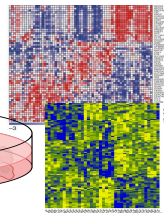
- Miguel Ponce de León - BSC (miguel.ponce@bsc.es)
- Marta Cascante - UB (martacascante@ub.edu)
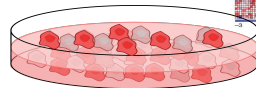
What is a cell made of?

Level (scale) of description

Chemistry

H_2O
C
N P

Building blocks

Phosphate Group

Pentose Sugar

Macromolecules

Cell

Animal Cell
cell membrane

fluid-filled space

organelles

nucleus

chromosomes

cytoplasm

organelle
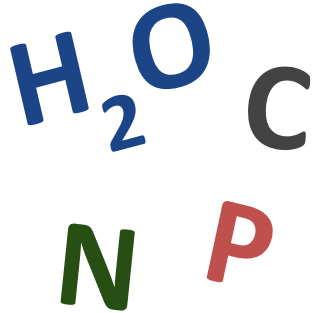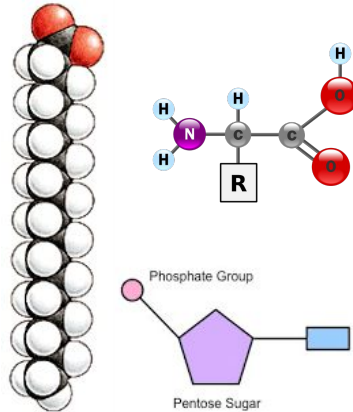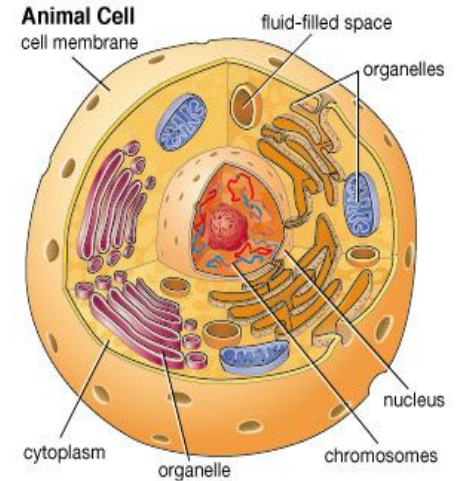
Cell's molecular factory: metabolism → What is metabolism?

Metabolism: the molecular factory of the cell

Is the **network** of biochemical reactions and transport processes that occur within a cell and allow **cell maintenance and growth**

Complex systems
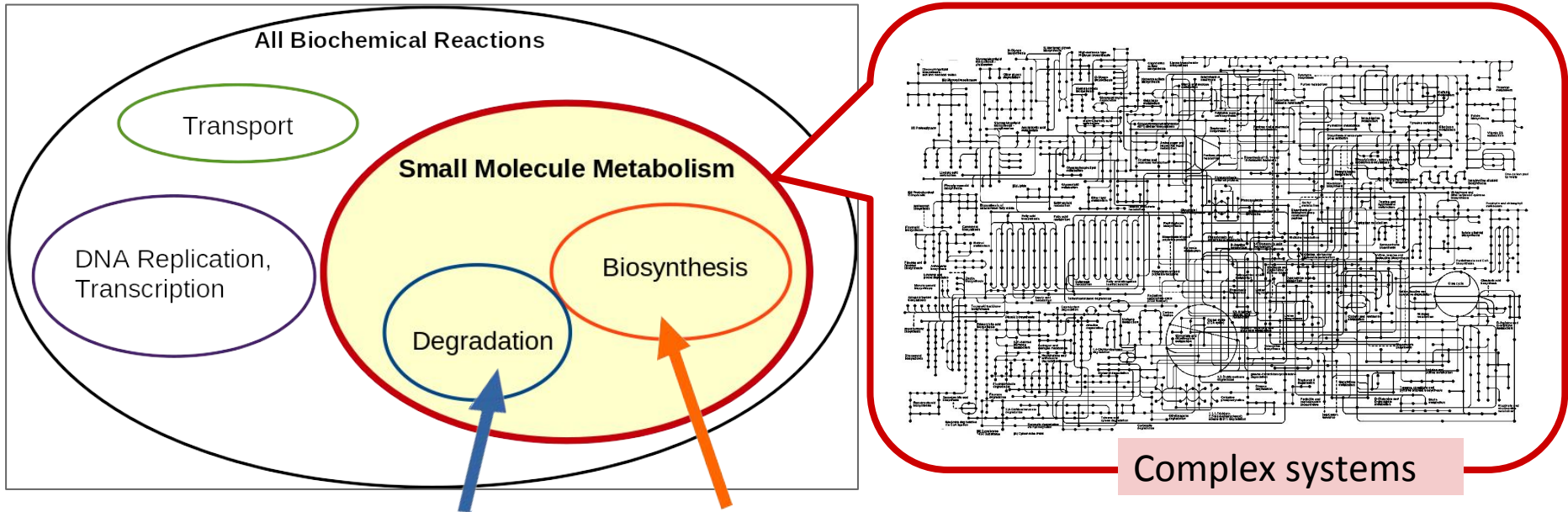
- Generation of **energy (catabolism)** and **building block (anabolism)**
- Include the enzymatic reaction that act over small molecules

Genome-Scale Modeling in Systems Biology
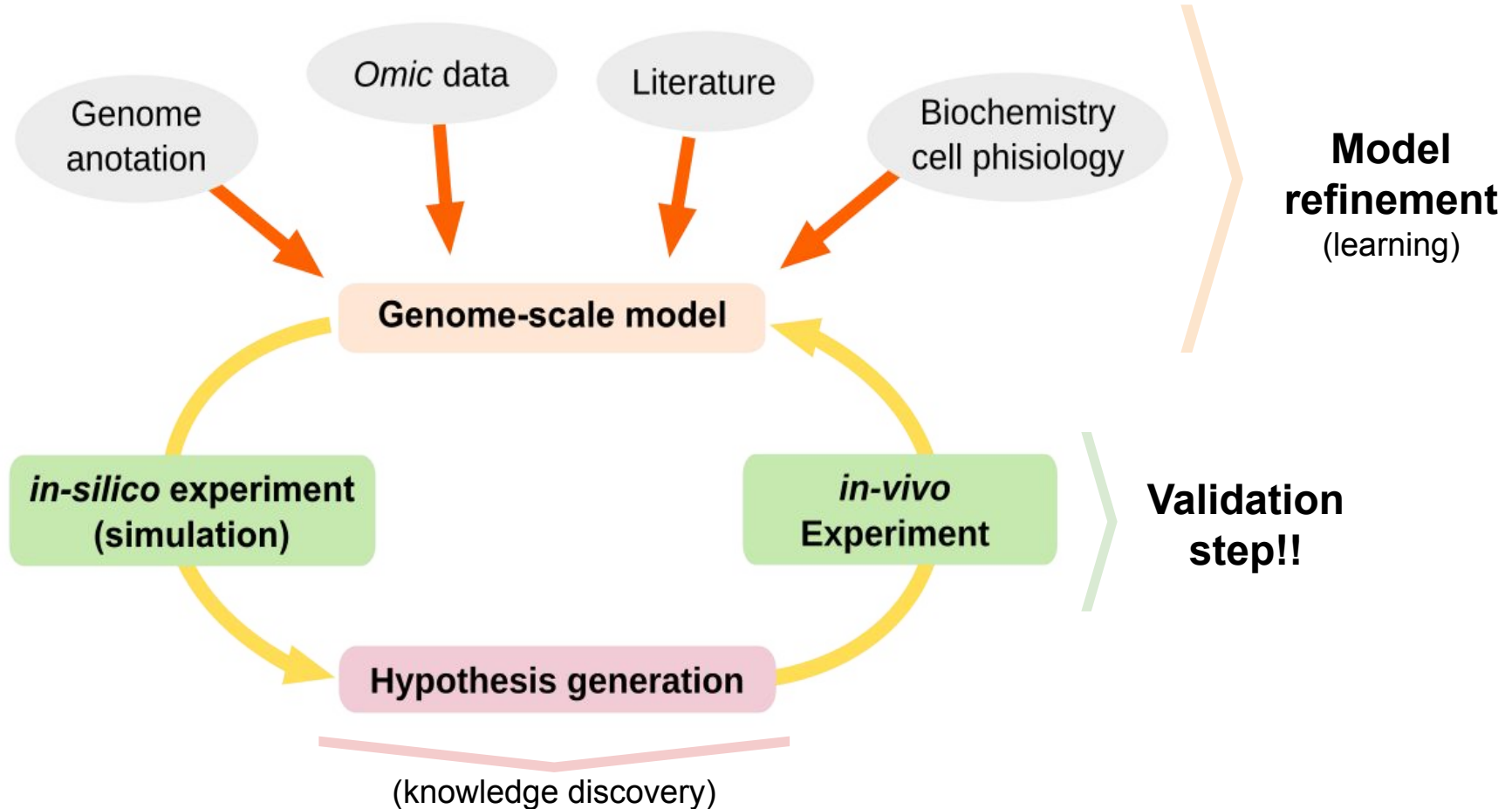
What is a Genome-scale metabolic model?

Is a computational representation the metabolism of a cell

- **Includes:** genes and complexes, biochemical reactions, metabolites, transporters, cell compartments.

- **Uses:** Omic data integration, simulations, *in-silico* predictions.

Genome-scale metabolic model

Gene-Protein-Reaction Rules

Gene A   Gene B

Protein A   Protein B

Enzyme

Byproduct

Nutrient A

Nutrient B

Nutrient C

Amino acids
Nucleotides
Lipids
Cofactos
Others
Energy

Biomass equation

Reaction Network (stoichiometric matrix)
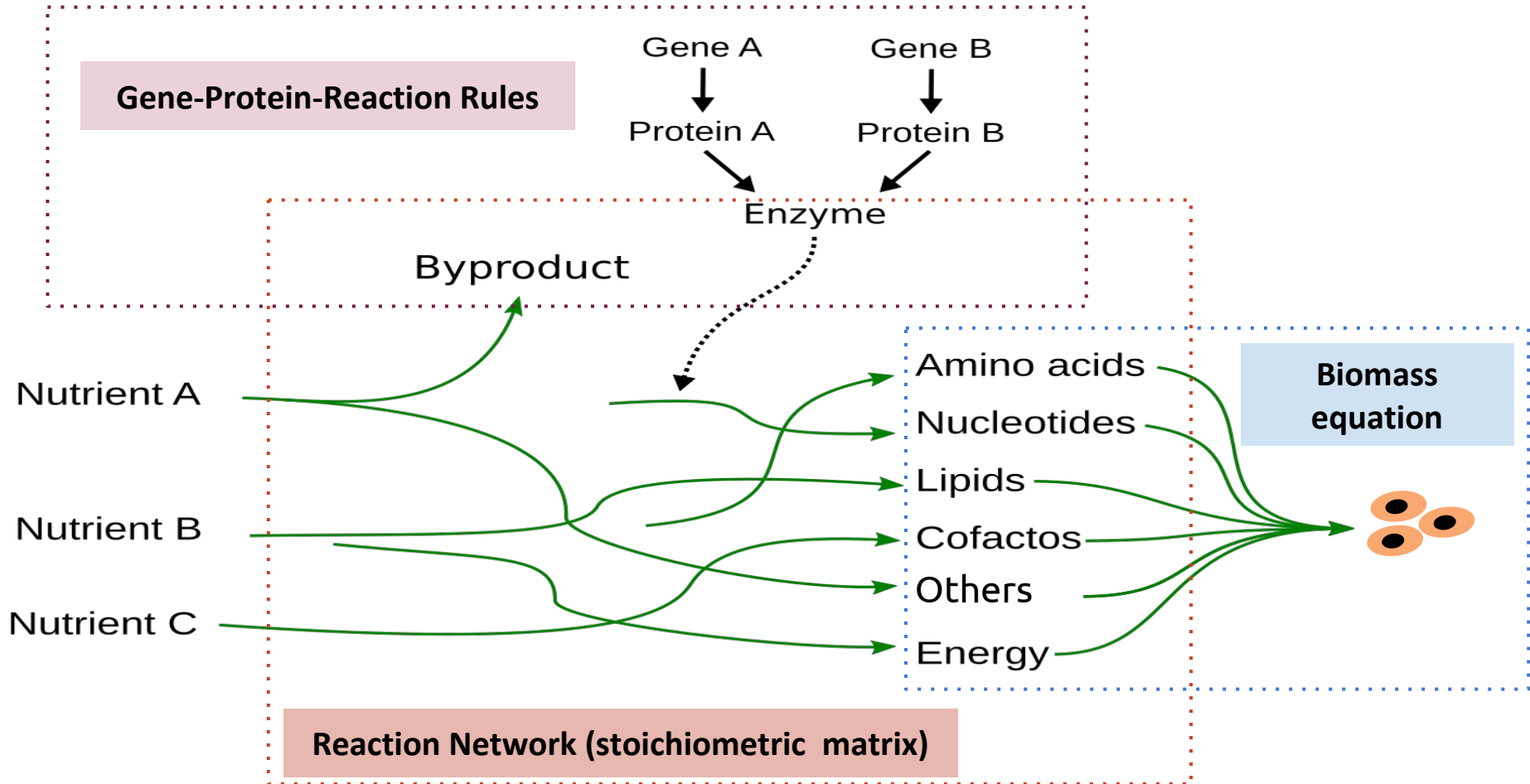
Genome-scale metabolic model

Gene-Protein-Reaction Rules

Gene-Protein Reaction Rules

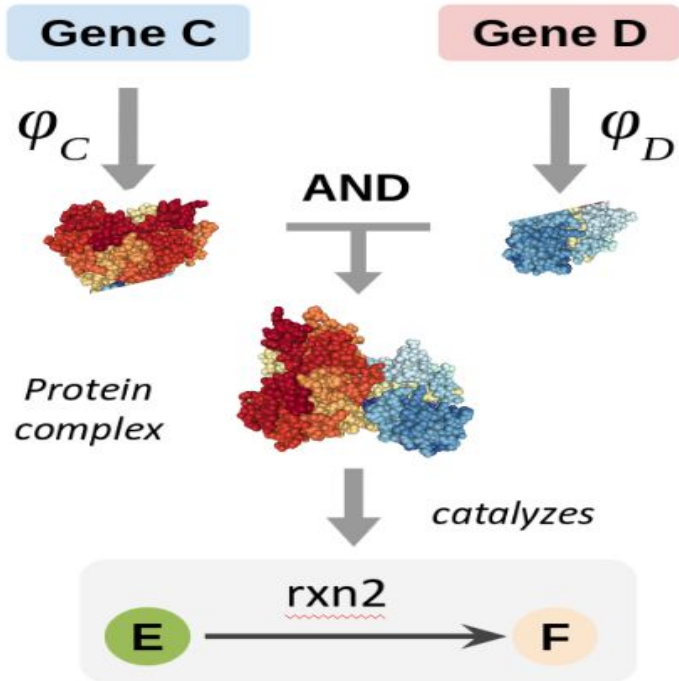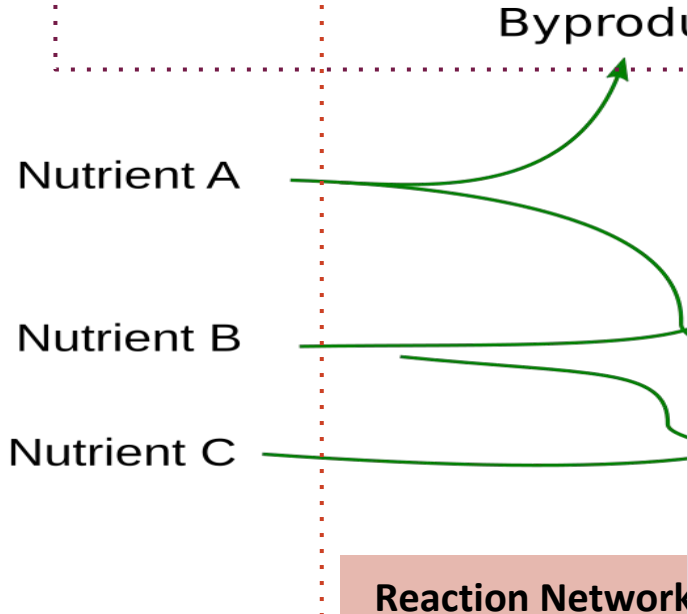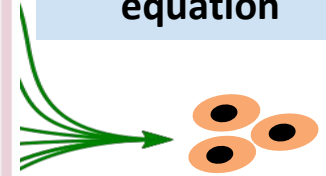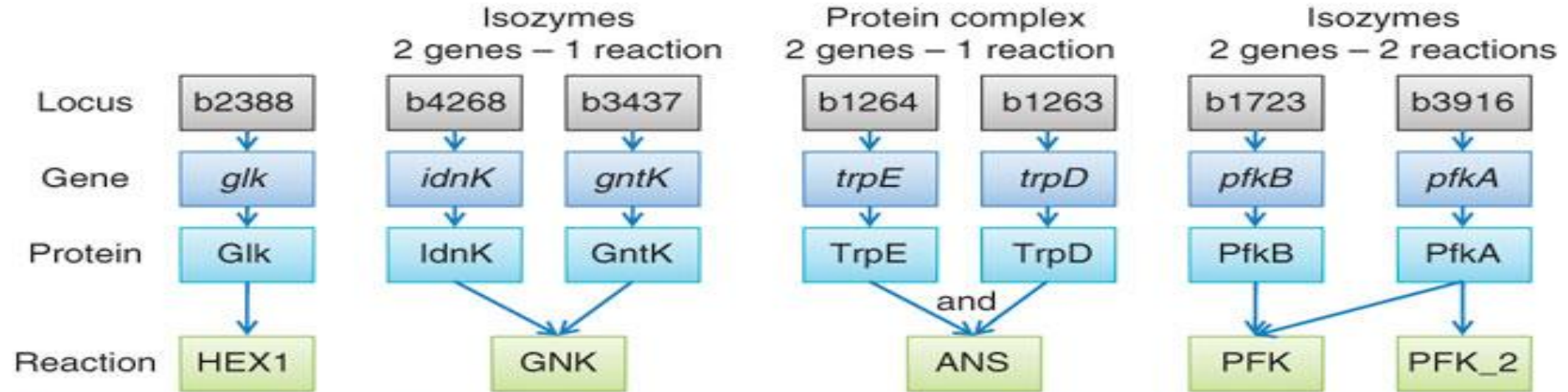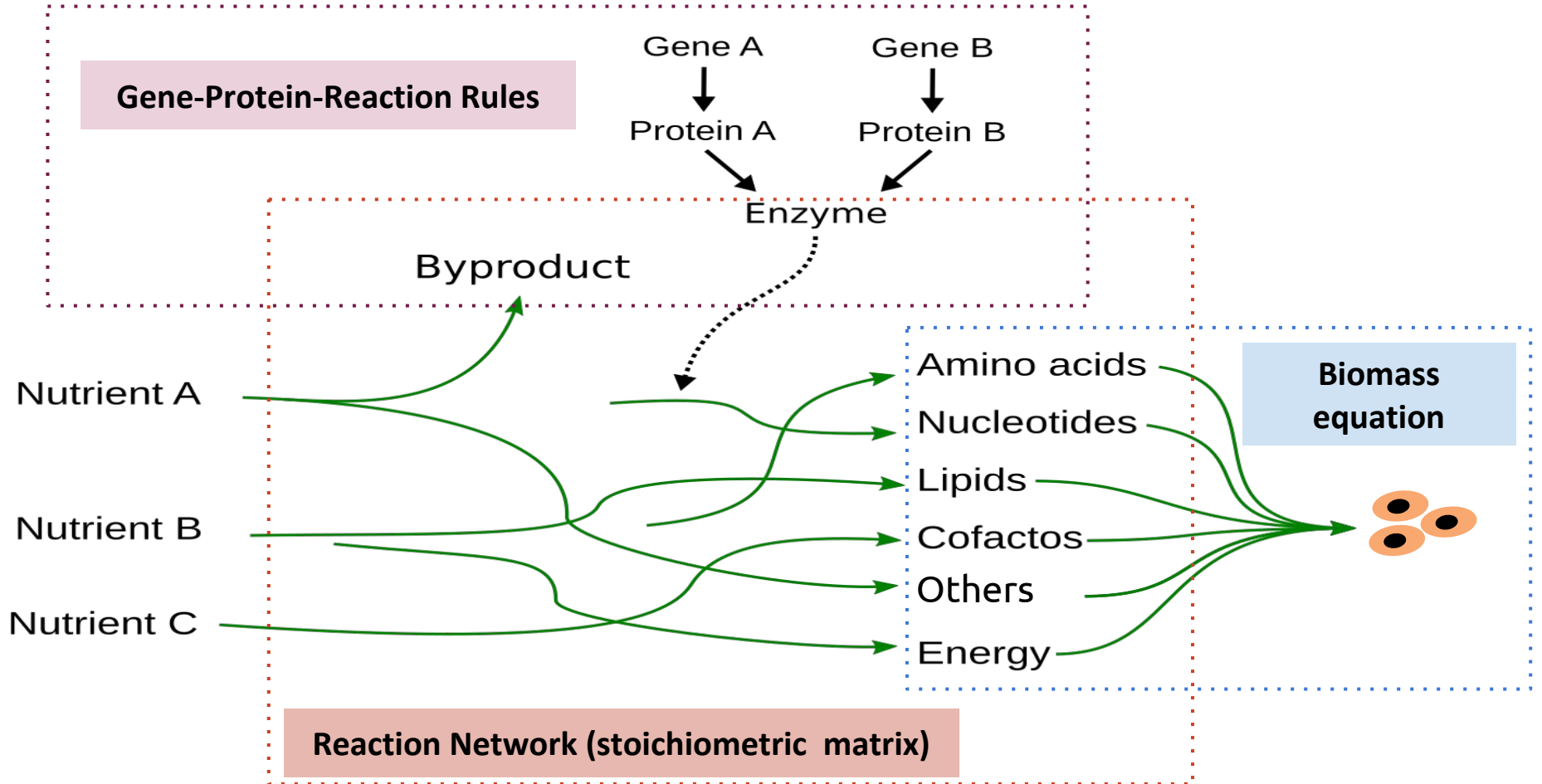*Logical relation between gene complexes and biochemical reactions*

Gene C

φ_C

AND

Gene D

φ_D

Protein complex

catalyzes

E — rxn2 → F

Byprodu

Nutrient A

Nutrient B

Nutrient C

Reaction Network

Biomass equation

Gene Protein Reaction rules: examples

| Isozymes 2 genes – 1 reaction | Protein complex 2 genes – 1 reaction | Isozymes 2 genes – 2 reactions |

	b2388	b4268	b3437	b1264	b1263	b1723	b3916
Locus							
Gene	glk	idnK	gntK	trpE	trpD	pfkB	pfkA
Protein	Glk	IdnK	GntK	TrpE	TrpD	PfkB	PfkA
Reaction	HEX1	GNK		ANS (and)		PFK	PFK_2

Reaction abbreviation	Reaction name	E. C.number	GPR
HEX1	Hexokinase (D-glucose:ATP)	2.7.1.1	(b2388)
GNK	Gluconokinase	2.7.1.12	(b3437) or (b4268)
ANS	Anthranilate synthase	4.1.3.27	(b1264) and (b1263)
PFK	Phosphofructokinase	2.7.1.11	(b1723) or (b3916)
PFK_2	Phosphofructokinase (2)	2.7.1.11	(b3916)

Thiele & Palsson 2010

Genome-scale metabolic model

Gene-Protein-Reaction Rules

Gene A → Protein A

Gene B → Protein B

Protein A → Enzyme ← Protein B

Byproduct

Nutrient A

Nutrient B

Nutrient C

Amino acids
Nucleotides
Lipids
Cofactos
Others
Energy

Biomass equation

Reaction Network (stoichiometric matrix)

Genome-scale metabolic model

Gene A   Gene B

Gene-Protein-Reaction Rules
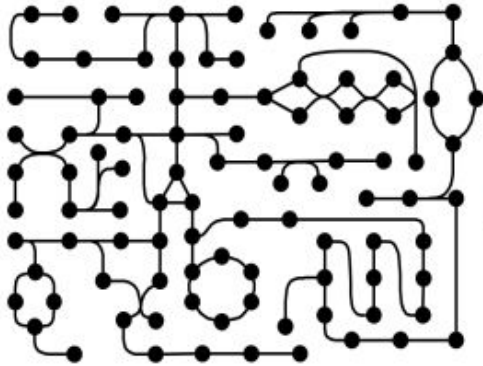
Metabolic Network: Human Metabolic model Recon 2.2.1

Reactions, Transports, Metabolites & Cell compartments (-> stoichiometric matrix)

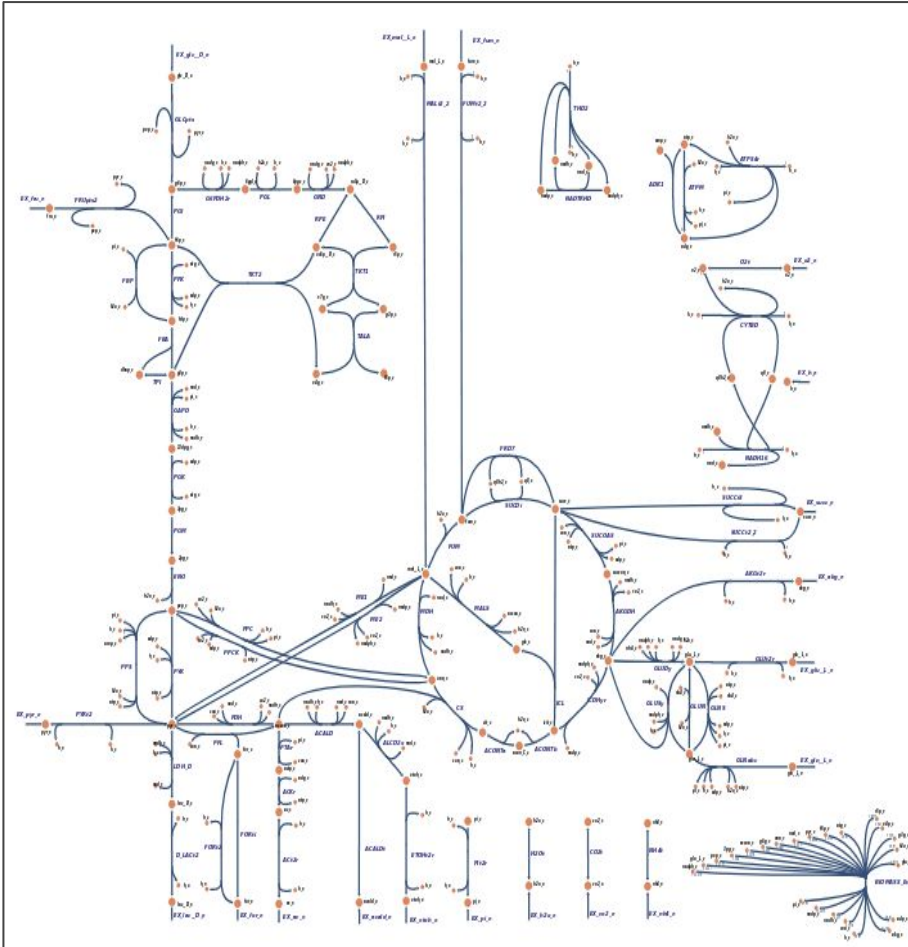$$\ldots \quad rxn2$$

$$\begin{matrix} A \\ B \\ \vdots \\ E \\ F \end{matrix} \begin{pmatrix} 0 \\ 0 \\ \\ -1 \\ +1 \end{pmatrix}$$

	Total
Genes	1675
Metabolites	5324
Reactions	7785

Reaction Network (stoichiometric matrix)

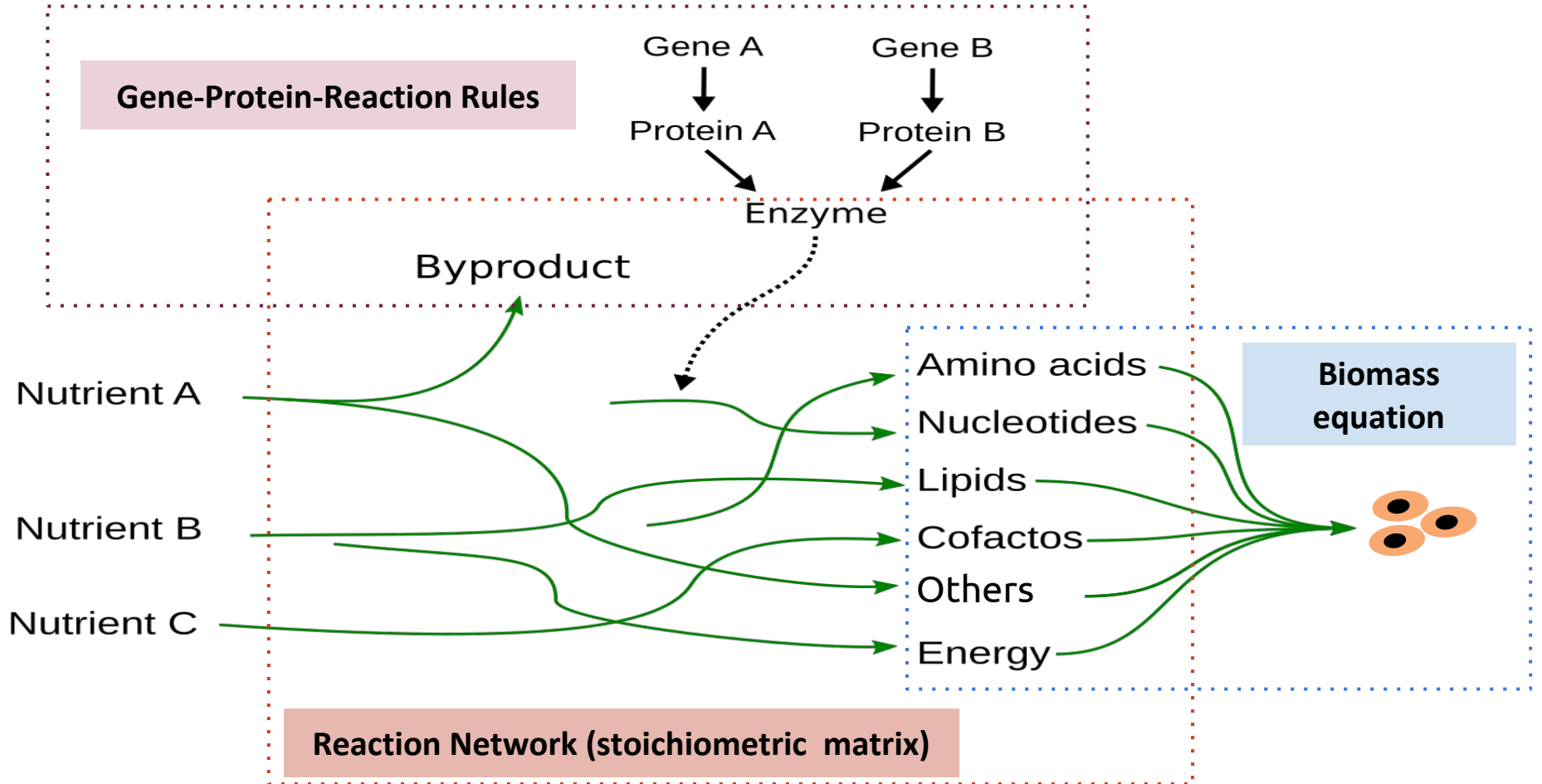# Stochimetric Matrix → Mass Balance Equations (E.coli core)


```
d[13dpg(c)] / dt =  GAPD + PGK          (c): cytosol
d[2pg(c)] / dt =  -ENO - PGM
d[3pg(c)] / dt =  1.496 * Biomass
d[6pgc(c)] / dt =  PGL - GND
d[6pgl(c)] / dt =  G6PDH2r - PGL
d[ac(c)] / dt =  -ACKr + ACt2r
d[acald(c)] / dt =  -ACALD + ACALDt + ALCD2x
d[accoa(c)] / dt =  ACALD + 3.7478 * Biomass
d[acon(C)(c)] / dt =  ACONTa - ACONTb
d[actp(c)] / dt =  ACKr + PTAr
d[akg(c)] / dt =  AKGt2r - 4.1182 * Biomass
d[amp(c)] / dt =  -ADK1 + PPS
d[cit(c)] / dt =  -ACONTa + CS
d[co2(c)] / dt =  CO2t + ICDHyr + AKGDH + GND
                 + ME1 + ME2 + PDH - PPC + PPCK
d[coa(c)] / dt =  -ACALD - 3.7478*Biomass
d[dhap(c)] / dt =  FBA - TPI
...
(continues)
```

Genome-scale metabolic model

Gene-Protein-Reaction Rules

Gene A   Gene B

Protein A   Protein B

Enzyme

Byproduct

Nutrient A
Nutrient B
Nutrient C

Amino acids
Nucleotides
Lipids
Cofactos
Others
Energy

Biomass equation

Reaction Network (stoichiometric matrix)

Genome-scale metabolic model

Biomass Equation

Quantitative molecular composition of a cell

| Biomass	Total
RNA components:	4
DNA components:	4
Protein components:	20
Carbohydrates:	1
Lipid components:	16
Cofactors and vitamins:	10

DNA + RNA
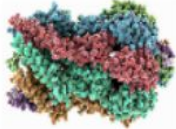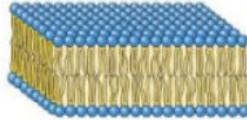
Proteins

Lipids

Vitamins/coenzymes

Carbohydrate

Biomass equation

Reaction Network (stoichiometric matrix)

Biomass Equation of a cell

Description in stoichiometric terms of all components present in a gram (dry weight) of a cell.
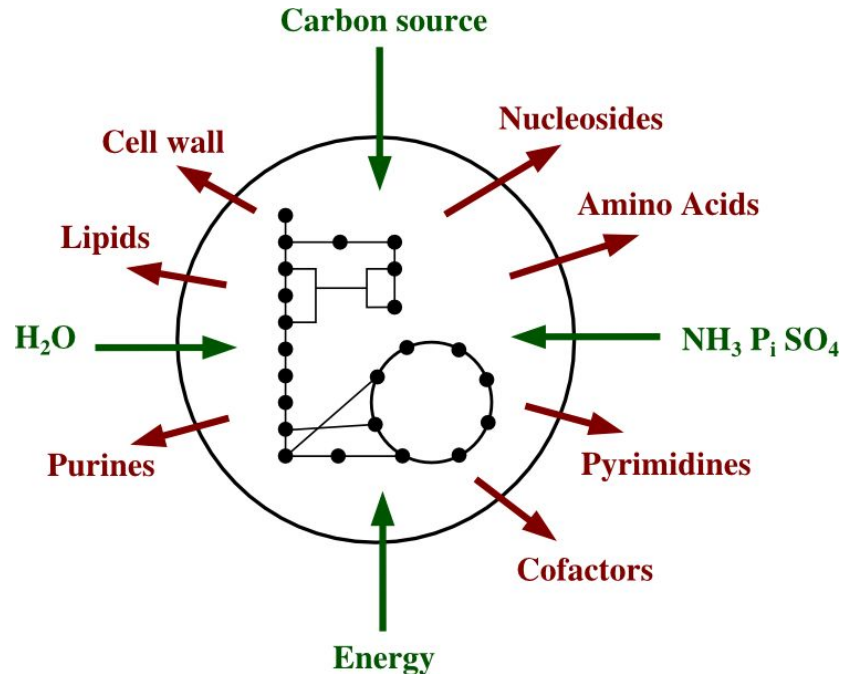
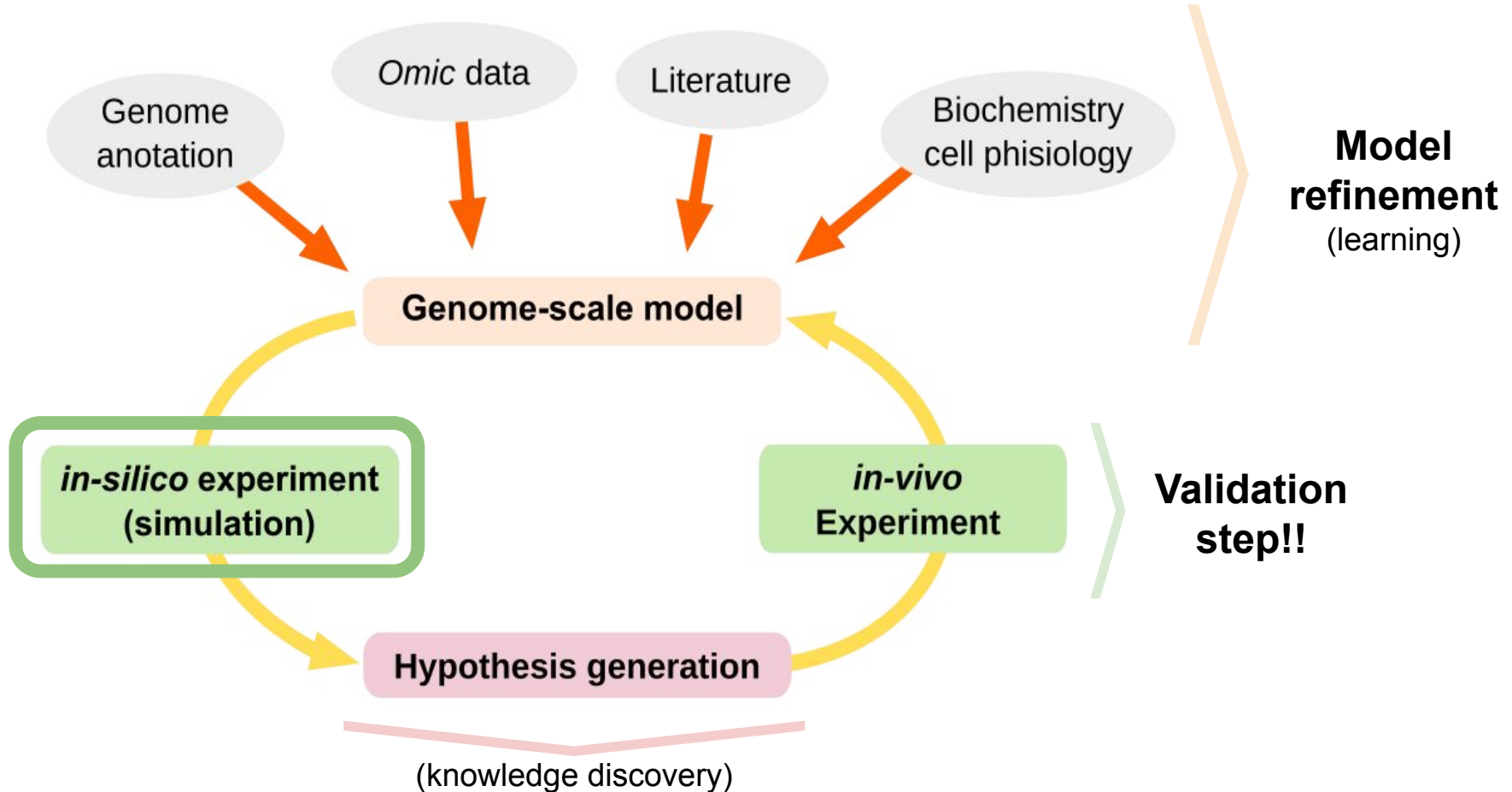| Metabolite	Demand (mmol)
ATP	41.2570
NADH	-3.5470
NADPH	18.2250
G6P	0.2050
F6P	0.0709
R5P	0.8977
E4P	0.3610
T3P	0.1290
3PG	1.4960
PEP	0.5191
PYR	2.8328
AcCoA	3.7478
OAA	1.7867
AKG	1.0789

Neidhardt et al. *Physiology of the Bacterial Cell* (1990)

How does it look (under the hood)?

Genome-scale metabolic model naked

```xml
<model id="Recon2.2.1">
  <listOfUnitDefinitions>
    <unitDefinition id="mmol_per_gDW_per_hr"> ... </listOfUnitDefinitions>
  <listOfCompartments>
    <compartment id="g" name="Golgi apparatus" size="1"/>
    <compartment id="c" name="cytoplasm" size="1"/>
    ...

  </listOfCompartments>
  <listOfSpecies>
    <species id="M_10fthf5glu_c" name="10-formyltetrahydrofolate-[Glu](5)" compartment="c" charge="-6">
      <notes>FORMULA: C40H45N11O19</notes>
    </species>
    ...
  <reaction id="R_ENO" name="enolase" reversible="true">
      <notes>
          <p>GENE ASSOCIATION: HGNC:3350 or HGNC:3354 or HGNC:3353</p>
          <p>CONFIDENCE LEVEL: 5</p>
          <p>SUBSYSTEM: Glycolysis/gluconeogenesis</p>
      </notes>
      <listOfReactants>
        <speciesReference species="M_2pg_c" stoichiometry="1"/>
      </listOfReactants>
      <listOfProducts>
        <speciesReference species="M_h2o_c" stoichiometry="1"/>
        <speciesReference species="M_pep_c" stoichiometry="1"/>
      </listOfProducts>
      <kineticLaw>
        <listOfParameters>
          <parameter id="UPPER_BOUND" value="1000" units="mmol_per_gDW_per_hr"/>
          <parameter id="FLUX_VALUE" value="0" units="mmol_per_gDW_per_hr"/>
          <parameter id="OBJECTIVE_COEFFICIENT" value="0" units="dimensionless"/>
          <parameter id="LOWER_BOUND" value="-1000" units="mmol_per_gDW_per_hr"/>
        </listOfParameters>
      </kineticLaw>
    </reaction>
  </listOf...
```
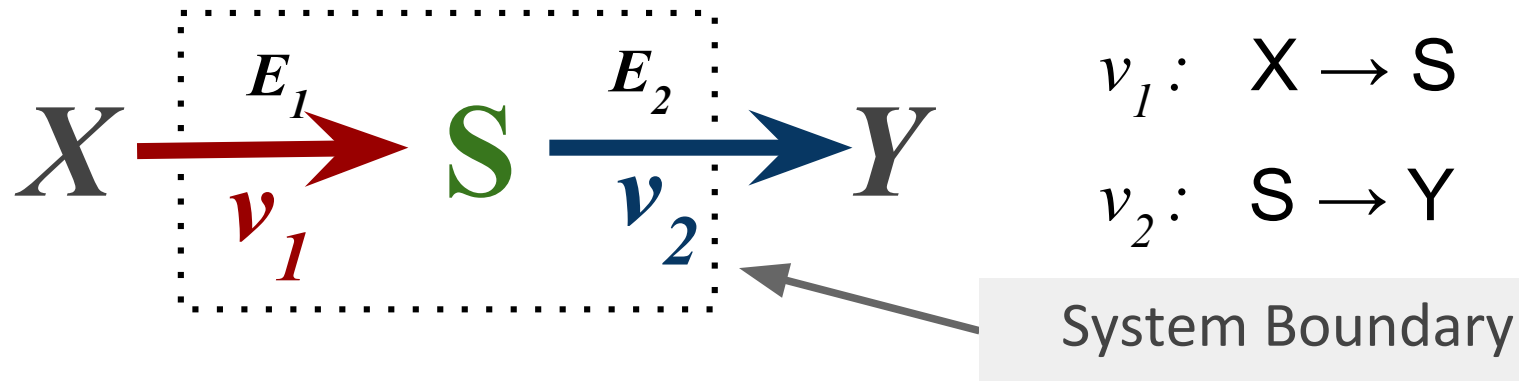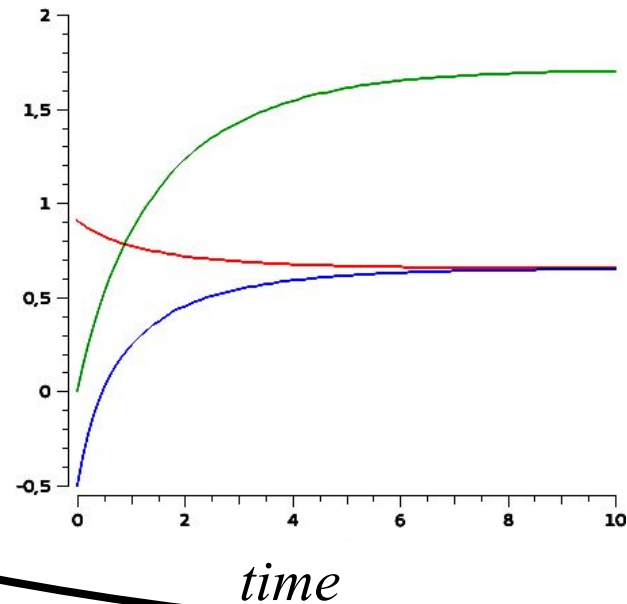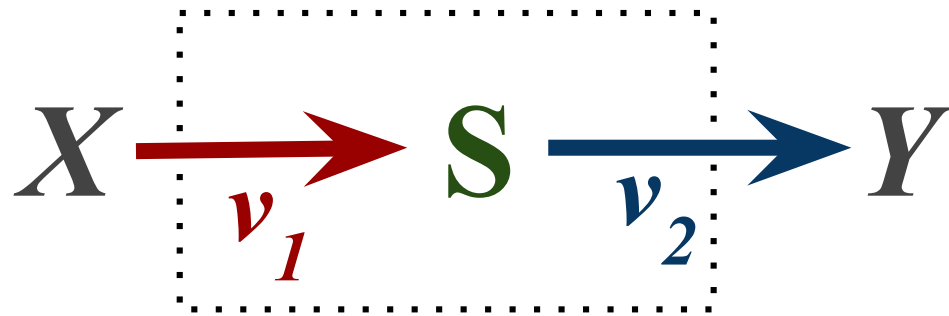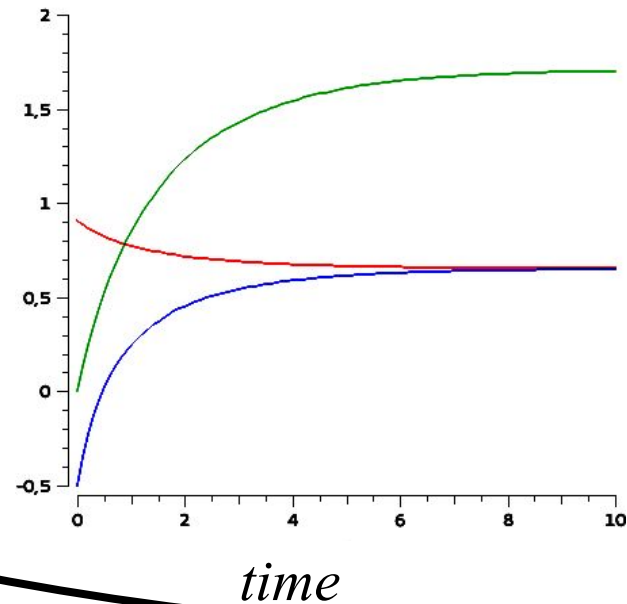
VIRTUAL METABOLIC HUMAN

https://www.vmh.life/

SBML

http://sbml.org/

https://www.ebi.ac.uk/biomodels-main/

BiGG Models

http://bigg.ucsd.edu/

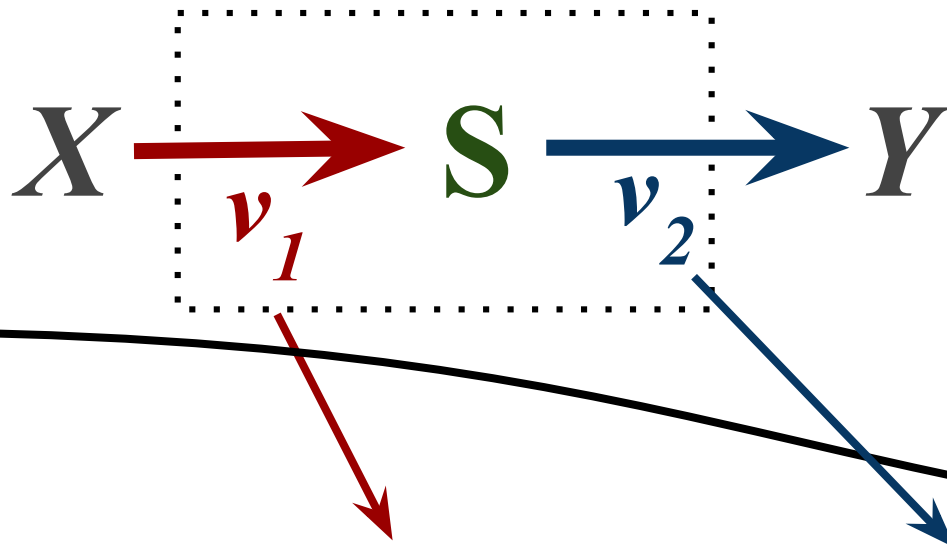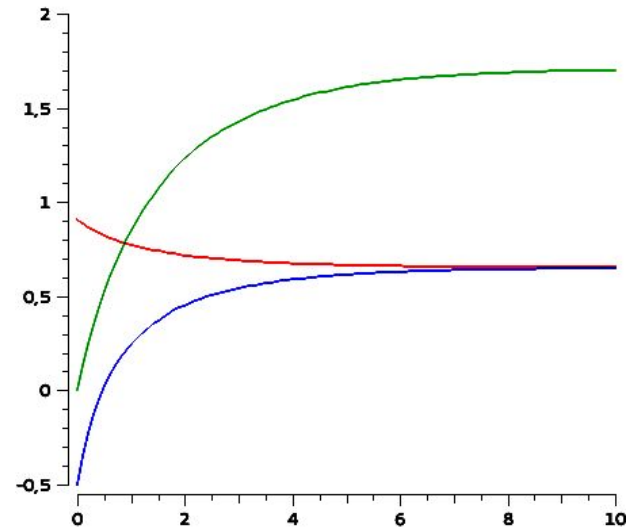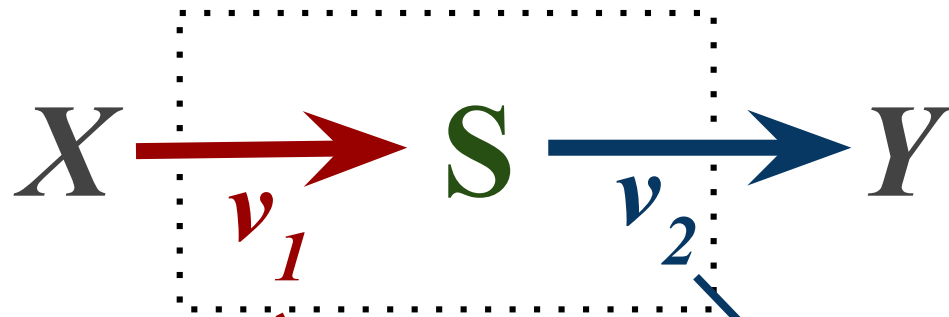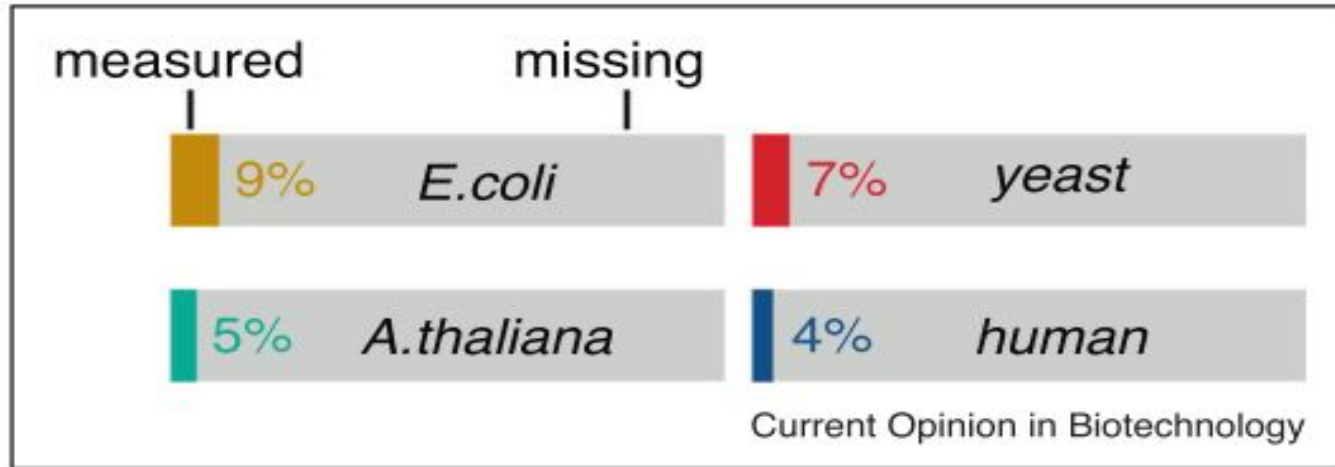# Genome-Scale Modeling in Systems Biology

Modeling metabolic systems

$v_1:$ X $\rightarrow$ S

$v_2:$ S $\rightarrow$ Y

System Boundary

Variables: $[S]$, v_1 , v_2

Parameters: X, Y, E_1, E_2

Kinetic description

$$X \xrightarrow{v_1} S \xrightarrow{v_2} Y$$

time

Kinetic description

$X \xrightarrow{v_1} S \xrightarrow{v_2} Y$

time

Kinetic Mechanisms:

- Mass action
- <u>Michaelis-Menten</u>
- Others

Kinetic description

$$X \xrightarrow{v_1} S \xrightarrow{v_2} Y$$

Many parameters!

$$\frac{d\,([S])}{d\,t} = \left(\frac{Vf_{(V1)} \cdot [X]}{Kms_{(V1)}} \cdot \frac{Vr_{(V1)} \cdot [S]}{Kmp_{(V1)}} \over 1 + \frac{[X]}{Kms_{(V1)}} + \frac{[S]}{Kmp_{(V1)}}} \right) + \left(\frac{Vf_{(V2)} \cdot [S]}{Kms_{(V2)}} \cdot \frac{Vr_{(V2)} \cdot [Y]}{Kmp_{(V2)}} \over 1 + \frac{[S]}{Kms_{(V2)}} + \frac{[Y]}{Kmp_{(V2)}}} \right)$$

time

Kinetic constants: the state of the art

Current Opinion in Biotechnology

Nº of reactions from GEMs:

- *E. coli* (*i*JO1266): 2251
- Budding yeast (*i*ND750): 1149
- Arabidopsis (--): 1363
- Human (Recon1): 7785

Davidi, Dan and Ron Milo. 2017. Curr. Op Biotech

Constraint-based approach: intuición

Theory

- Full information
- Solution is a unique point (or limit cycle) in the system's phase space
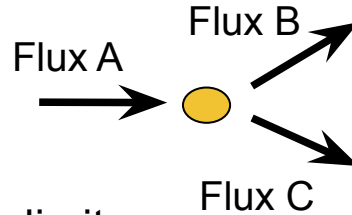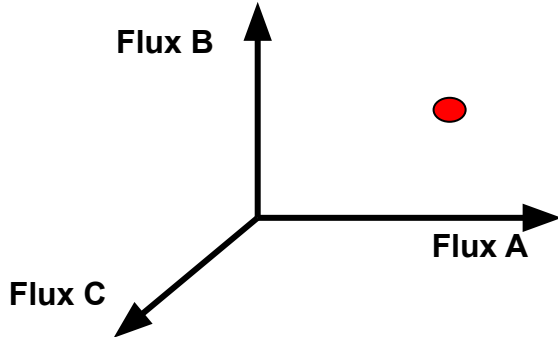
Genome-scale

- Incomplete information
- **Solution (flux) space**
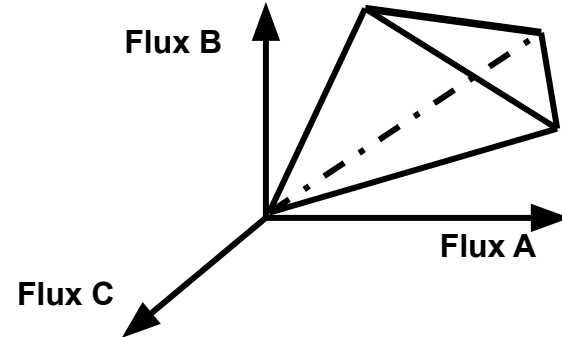
For genome-scale reconstructed metabolic network there are still not enough information to create the full kinetic description → too many unknown parameters!

Modeling metabolic systems

Kinetic Modelling (Differential equations)

○ Time evolution of system variables (+)

○ Sensitivity Analysis (+)

○ Unknown Kinetic Parameter (-)

○ Unknown Enzymatic Mechanisms (-)

Constraint-Based Modeling (CBM)

○ Only need stoichiometry (+)

○ Structure is an invariant property (+)

○ Computationally tractable using **genome-scale models** (+)

○ No information of metabolite concentrations (-)

○ Only valid under steady-state (-)

Constraint-Based Modeling

Glucose + ATP

Glucose-6-P + ADP

Glucokinase (single reaction)

Glucokinase

$$
\begin{array}{c}
\text{Glucose} \\
\text{ATP} \\
\\
\text{G-6-P} \\
\text{ADP}
\end{array}
\left[
\begin{array}{ccc}
& -1 & \\
& -1 & \\
& & \\
& +1 & \\
& +1 &
\end{array}
\right] \longrightarrow N
$$

Stoichiometric matrix N (metabolic network)

Constraint-Based Modeling
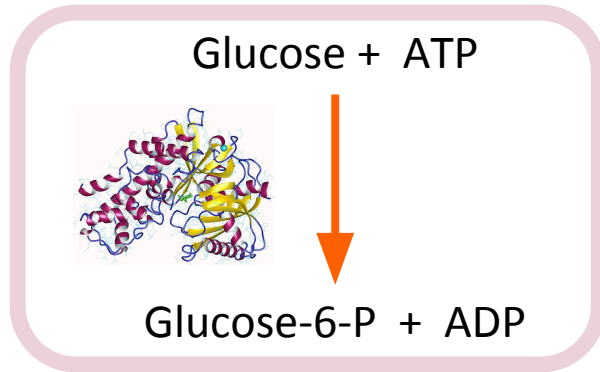
Glucose + ATP

↓

Glucose-6-P + ADP

Glucokinase (single reaction)

Glucokinase

$$\begin{array}{c} \text{Glucose} \\ \text{ATP} \\ \\ \text{G-6-P} \\ \text{ADP} \end{array} \begin{bmatrix} -1 & & \\ -1 & & \\ & & \\ +1 & & \\ +1 & & \end{bmatrix} \rightarrow N$$

Stoichiometric matrix N (metabolic network)

The Constraints

Constraint-Based Modeling
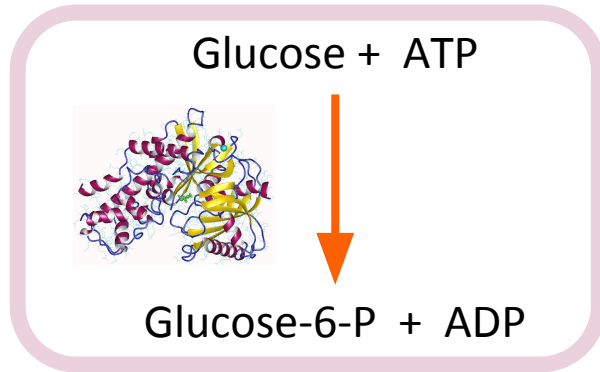
Glucose + ATP

↓

Glucose-6-P + ADP

Glucokinase (single reaction)

Glucokinase

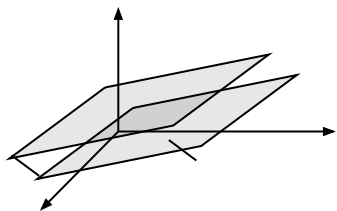|		
Glucose	-1	
ATP	-1	
G-6-P	+1	
ADP	+1	

$\rightarrow N$

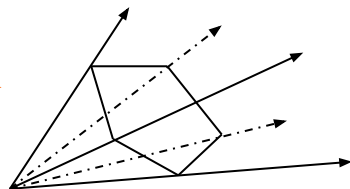Stoichiometric matrix N (metabolic network)

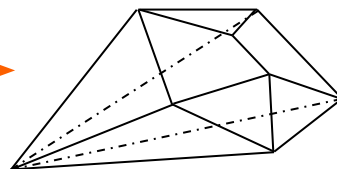## The Constraints

Mass Balance

$$N \cdot v = 0$$
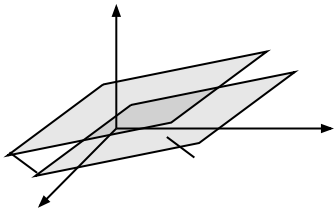
Constraint-Based Modeling
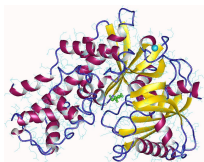
Glucose +

Glucose-6-P

Glucokinase (sing

The Constraints

Mass Balance

$$N \cdot v = 0$$

N
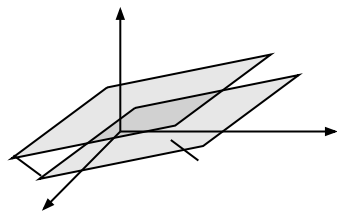
```
d[13dpg(c)] / dt =  GAPD + PGK = 0

d[2pg(c)] / dt =  -ENO - PGM = 0

d[3pg(c)] / dt =  1.496 x Biomass

d[6pgc(c)] / dt =  PGL - GND = 0

d[6pgl(c)] / dt =  G6PDH2r - PGL = 0

d[ac(c)] / dt =  -ACKr + ACt2r = 0

d[acald(c)] / dt =  -ACALD + ACALDt + ALCD2x = 0

d[accoa(c)] / dt =  ACALD + 3.7478 x Biomass

d[acon(C)(c)] / dt =  ACONTa - ACONTb = 0

d[actp(c)] / dt =  ACKr + PTAr = 0

d[akg(c)] / dt =  AKGt2r - 4.1182 x Biomass

d[amp(c)] / dt =  -ADK1 + PPS = 0

d[cit(c)] / dt =  -ACONTa + CS = 0

...

(continues)
```
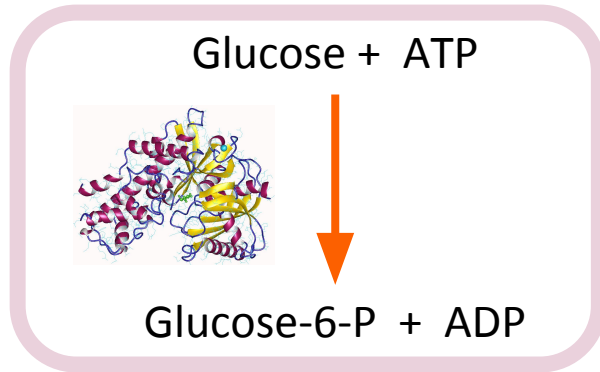
Constraint-Based Modeling

Glucose + ATP

Glucose-6-P + ADP

Glucokinase (single reaction)

Glucokinase

|	-1	
Glucose	-1	
ATP		
G-6-P	+1	
ADP	+1	

$\rightarrow N$

Stoichiometric matrix N (metabolic network)

## The Constraints

Mass Balance

$N \cdot v = 0$

Thermodynamics

$v_i > 0$

Constraint-Based Modeling
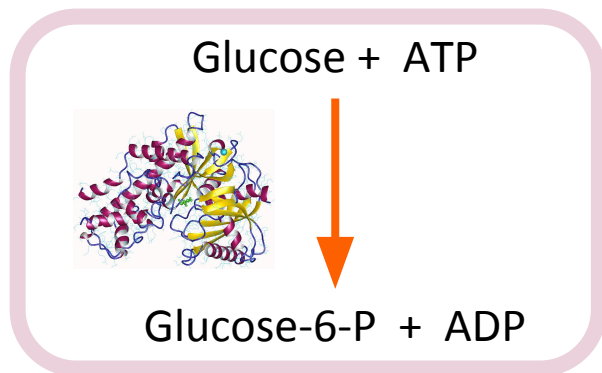
Glucose + ATP

Glucose-6-P + ADP

Glucokinase (single reaction)

Glucokinase

	Glucokinase	
Glucose	-1	
ATP	-1	
G-6-P	+1	
ADP	+1	

$\rightarrow N$

Stoichiometric matrix N (metabolic network)

The Constraints

Mass Balance

$$N \cdot v = 0$$

Thermodynamics

$$v_i > 0$$

Capacities (bounds)

$$v_i < v_{max}$$

Flux Space

Constraint-Based Modeling
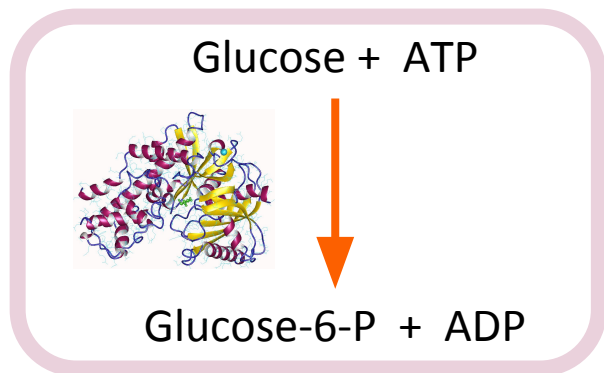
Glucose + ATP

Glucose-6-P + ADP

Glucokinase (single reaction)

Glucokinase

$$\begin{array}{c} \text{Glucose} \\ \text{ATP} \\ \\ \text{G-6-P} \\ \text{ADP} \end{array} \begin{pmatrix} -1 \\ -1 \\ \\ +1 \\ +1 \end{pmatrix} \rightarrow N$$

Stoichiometric matrix N (metabolic network)

The Constraints

Mass Balance

$$N \cdot v = 0$$

Thermodynamics

$$v_i > 0$$
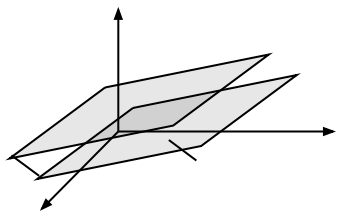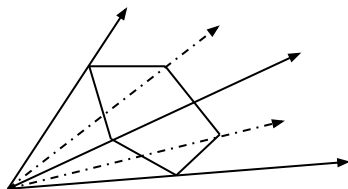
Capacities (bounds)

$$v_i < v_{max}$$

Flux Space
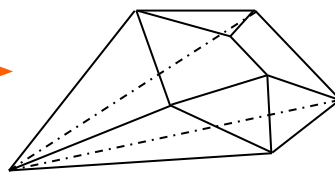
Cell Objective

Flux Balance Analysis

$$Max \ \ c^T \cdot v$$
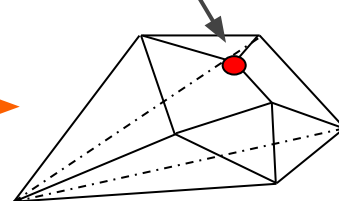
Growth Media

Exchange fluxes: variables that represent the exchanges of nutrients/by-products between the organism and its environment (uptake/secretion)

$$
\begin{array}{c}
 & GLC_{Ex} \quad O_{Ex} \quad CO2_{Ex} \\
\begin{array}{c} Glc \\ O_2 \\ CO_2 \end{array}
\left[
\begin{array}{c|c|c}
-1 & & \\
\hline
 & -1 & \\
\hline
 & & -1
\end{array}
\right]
\end{array}
$$
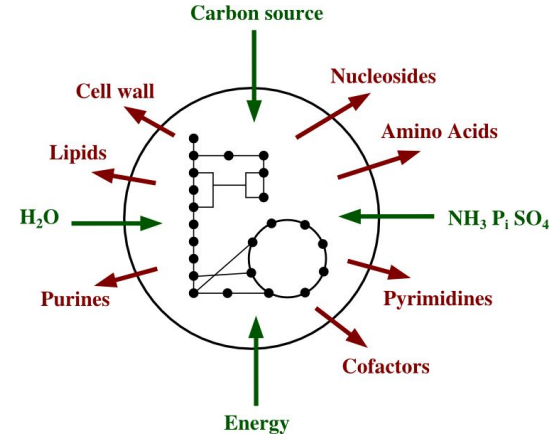
Flux Cone
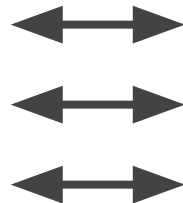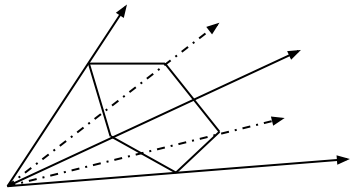
	Min	Max
Glucose	- 6	0
Oxygen	- 20	0
CO2	0	Inf

Flux Space
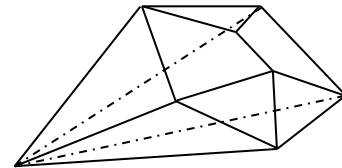
A phylogeny of *CBM* methods

Lewis et al. 2012
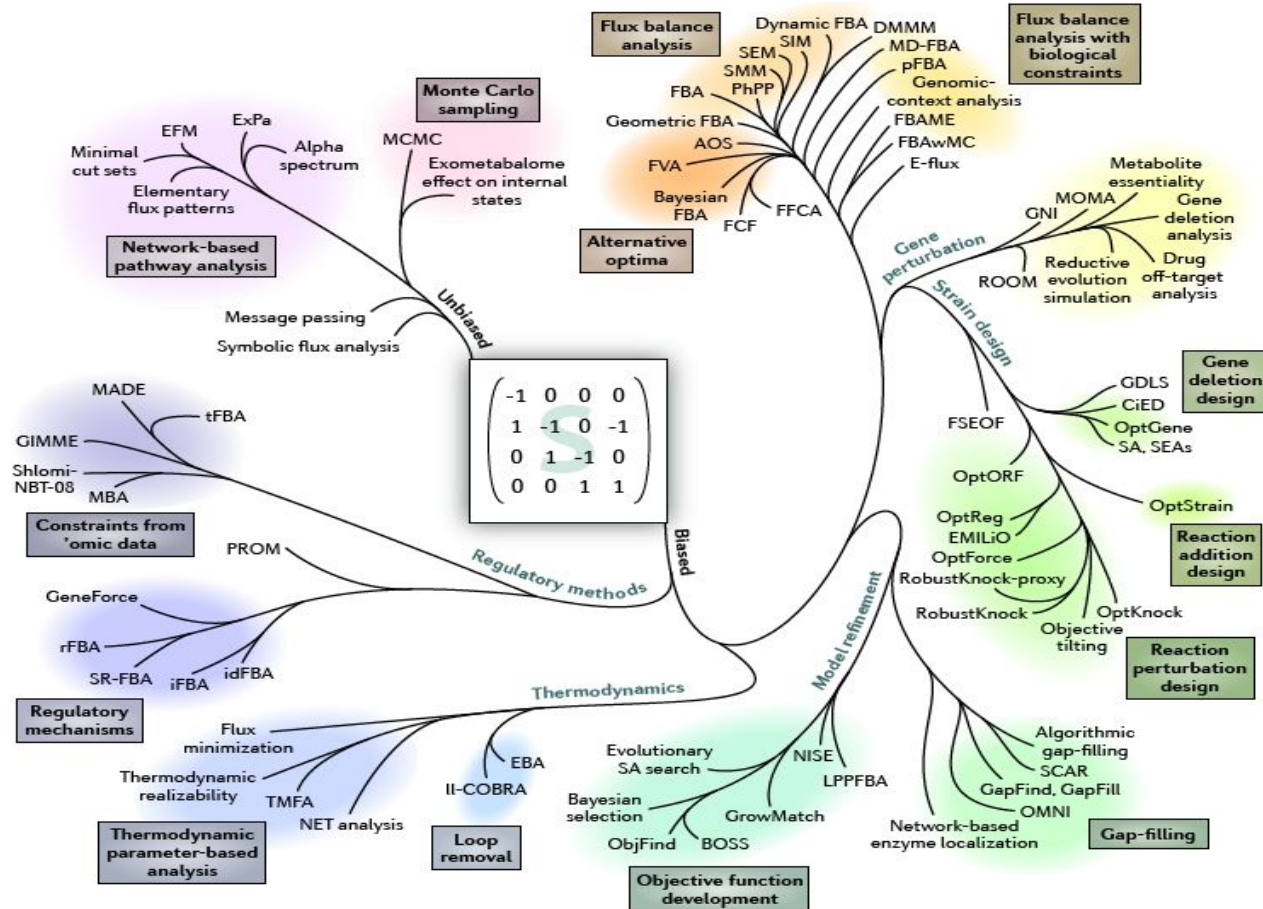
¿How to identify a particular flux distribution?

Using optimization principles

- Ajustando con datos experimentales

- Maximize de Biomass production (growth rate)

- Maximize ATP production

- Minimize metabolic cost

- Multiple criteria (Pareto optimality)

¿phenotype?

Flux space (~genotype)
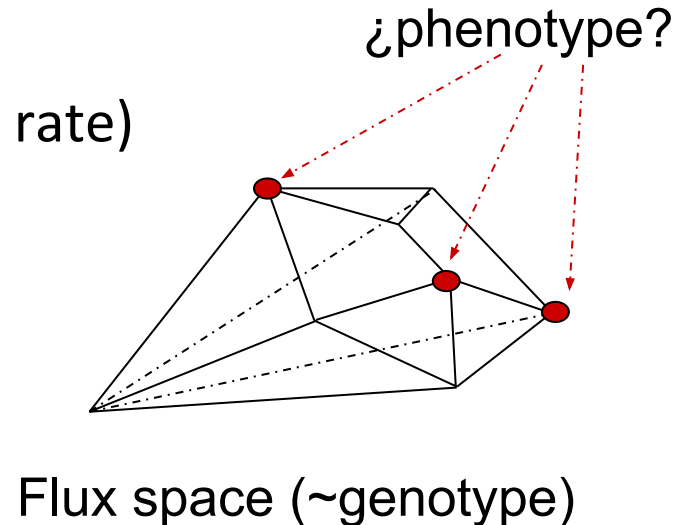
Flux Balance Analysis (FBA)

Definition: computations strategy to calculate the flux distribution of a metabolic network that maximize the growth rate subject to a given set of constraints (e.g. mass balance, thermodynamics, etc)

Constraints:

$$N \cdot \vec{v} = 0$$

$$v_j \geq 0 \quad \forall j \in R_{irrev}$$

$$\alpha_j \leq v_j \leq \beta_j, \quad \forall j \in R$$

Mass balanace

Thermodynamics

Enzyme and transport capacities

Feasible space (flux space)

Optimizations

Flux distribution (phenotype)
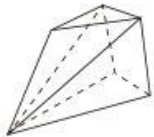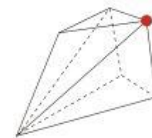
Feasible space (Genotype)

Feasible space (Genotype)

Applications

Gene *knockouts* predictions

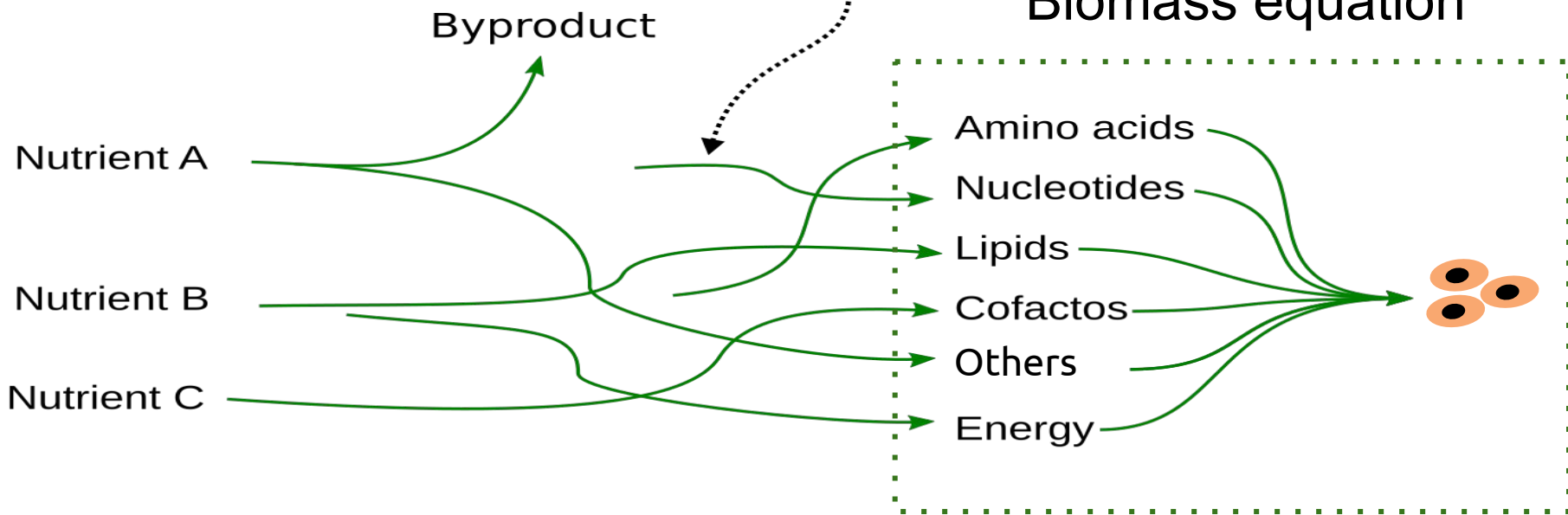# Simulations → *in-silico* predictions of gene KO effect

Simulations → *in-silico* predictions of gene KO effect

What happens if
Gene A is not functional?

Biomass equation

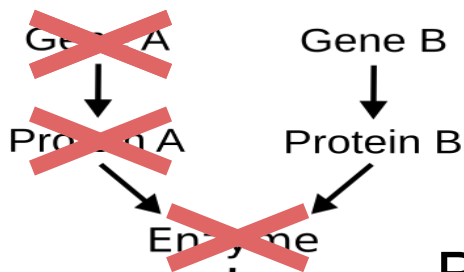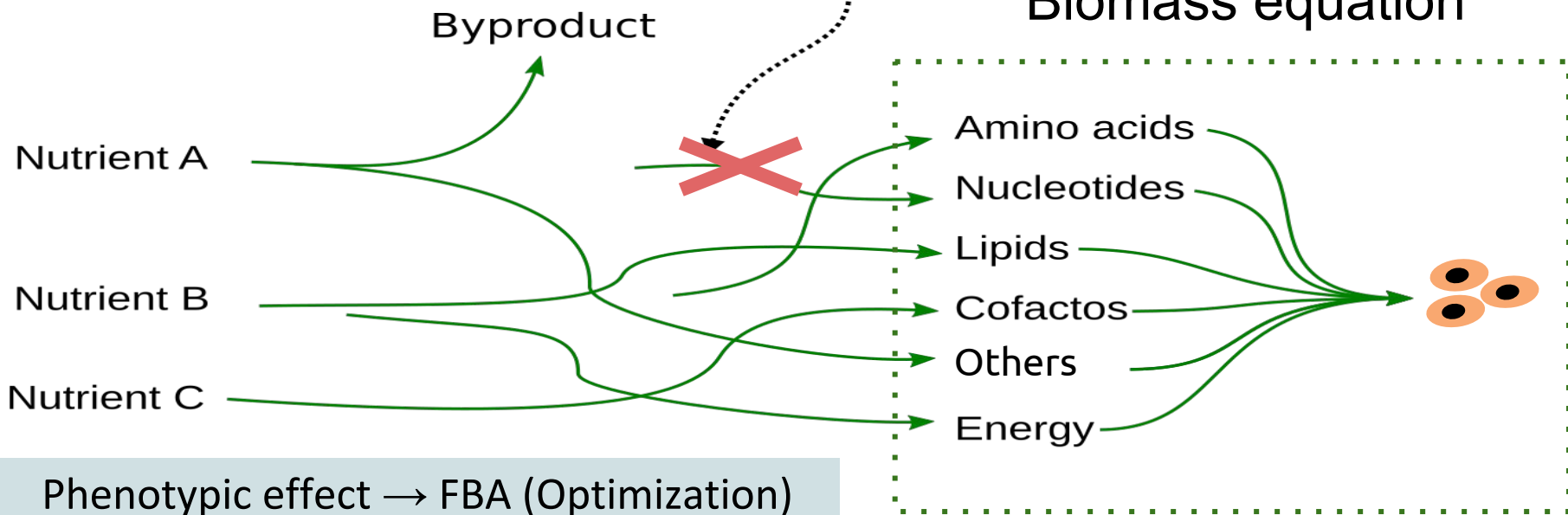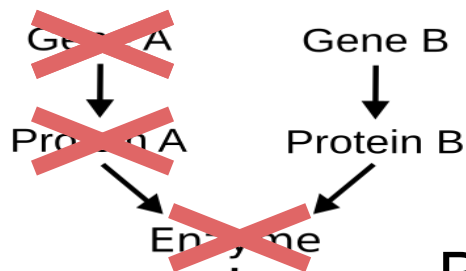# Simulations → *in-silico* predictions of gene KO effect

Simulations → *in-silico* predictions of gene KO effect
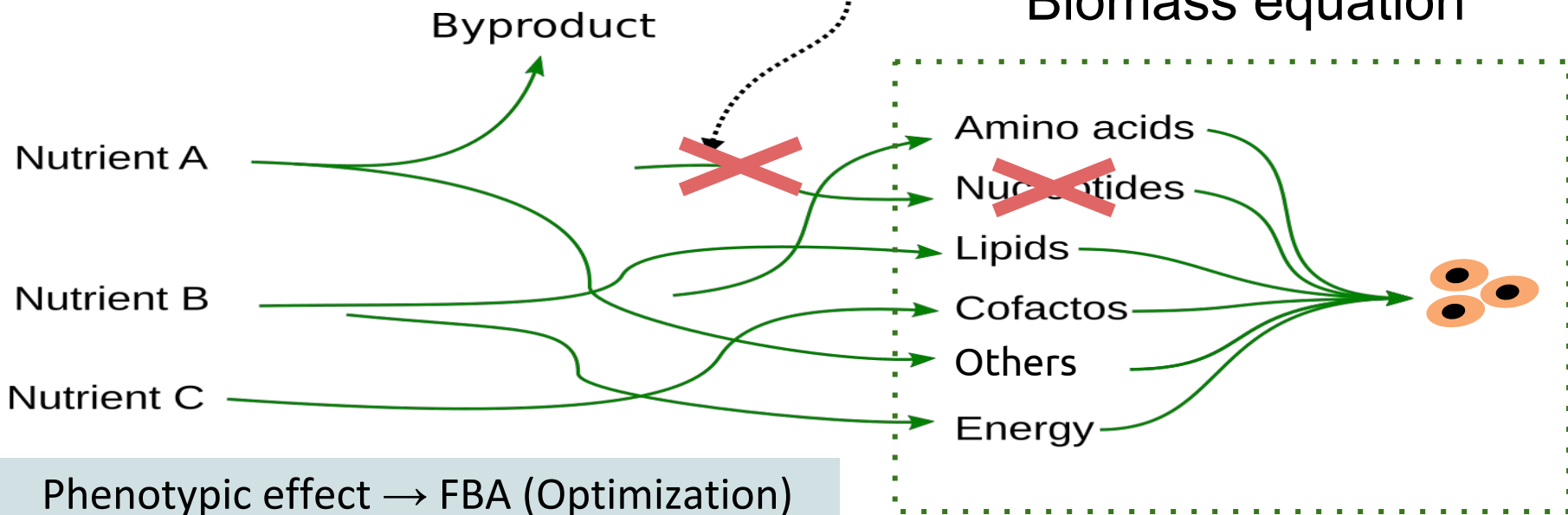
What happens if
Gene A is not functional?

Biomass equation

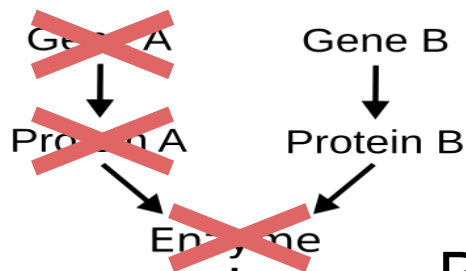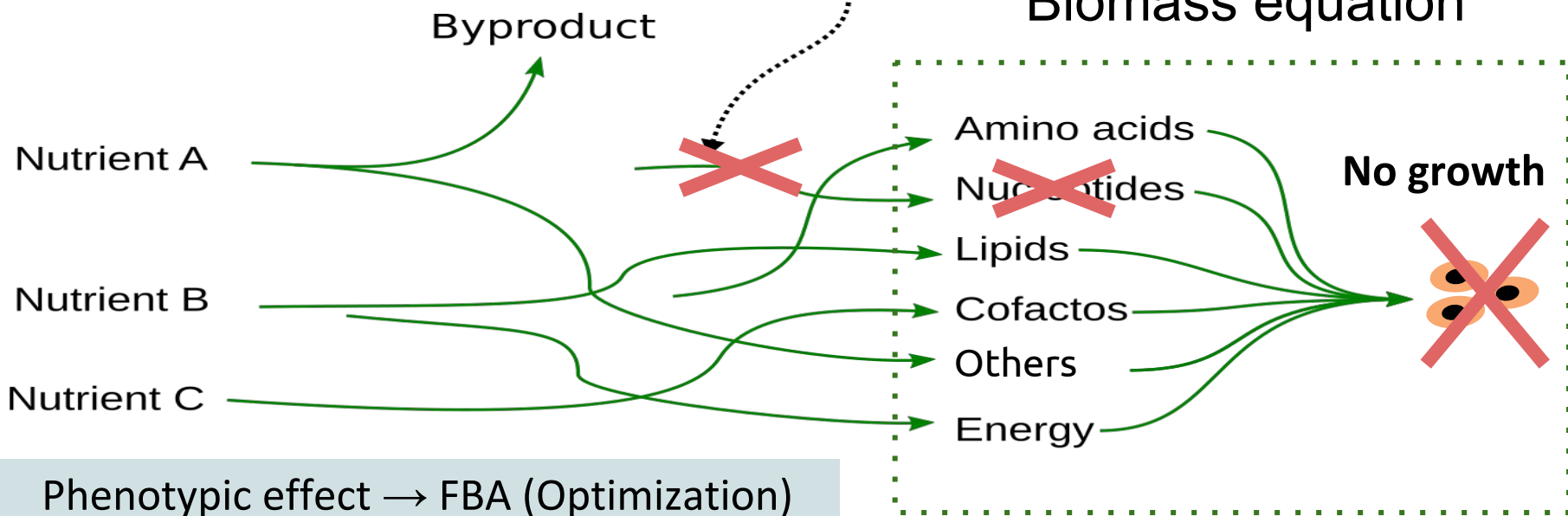# Simulations → *in-silico* predictions of gene KO effect

What happens if
Gene A is not functional?

Biomass equation

Simulations → *in-silico* predictions of gene KO effect

What happens if **Gene A** is not functional?

Gene A   Gene B

Protein A   Protein B

Enzyme

Biomass equation

Byproduct

Nutrient A

Nutrient B

Nutrient C

Amino acids

Nucleotides

Lipids

Cofactos

Others

Energy

Phenotypic effect → FBA (Optimization)

Simulations → *in-silico* predictions of gene KO effect

What happens if
Gene A is not functional?

Gene A   Gene B

Protein A   Protein B

Enzyme

Byproduct

Biomass equation

Nutrient A

Nutrient B

Nutrient C

Amino acids
Nucleotides
Lipids
Cofactos
Others
Energy

Phenotypic effect → FBA (Optimization)

Simulations → *in-silico* predictions of gene KO effect

What happens if
Gene A is not functional?

Byproduct

Gene A

Gene B

Protein A

Protein B

Enzyme

Biomass equation

Nutrient A

Nutrient B

Nutrient C

Amino acids

Nucleotides

Lipids

Cofactos

Others

Energy

No growth

Phenotypic effect → FBA (Optimization)

Simulations → *in-silico* predictions of gene KO effect

What happens if ***Gene A*** is not functional?

Simulation → no growth
Hypothesis: **A is essential**

Biomass equation

No growth

Phenotypic effect → FBA (Optimization)
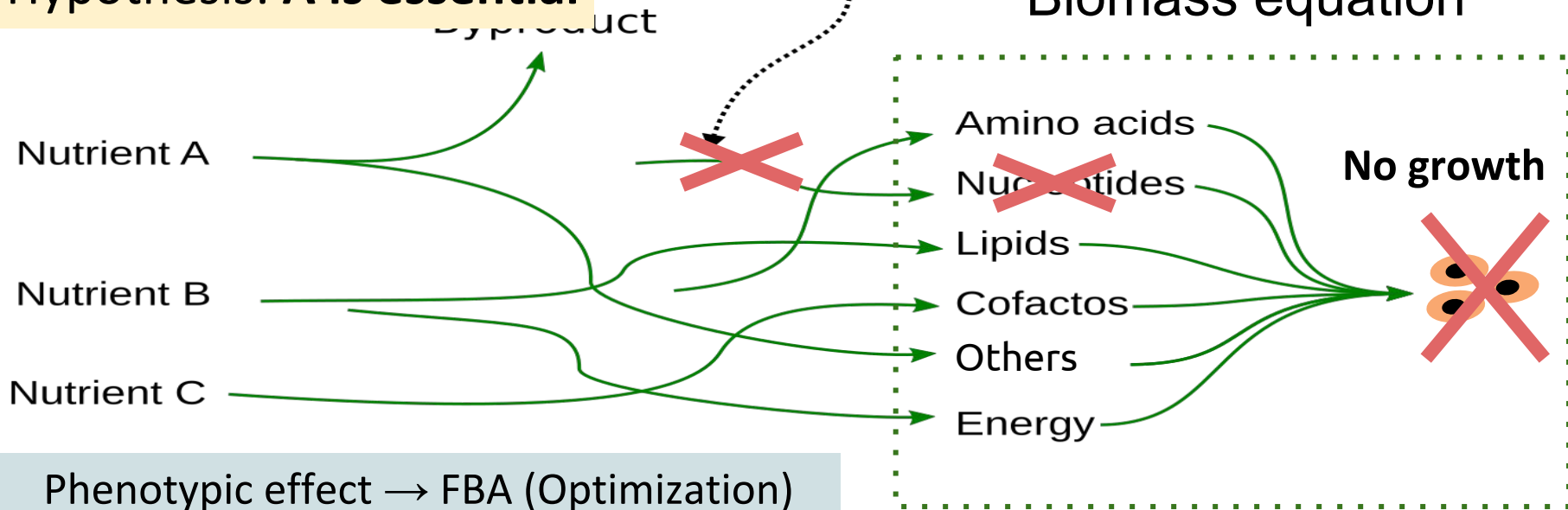
Applications

Gene *knockouts* predictions

Does it really work?

Rerouting of metabolic fluxes

- ▬ **Wild type**
- ▬ **zwf⁻ mutant**
- ▬ **zwf⁻ pnt⁻ double mutant**

In-silico knockout predictions

Rerouting of metabolic fluxes

- Wild type
- zwf⁻ mutant
- zwf⁻ pnt⁻ double mutant

Edwards JS, Palsson BO. PNAS (2000)

Predicción de *knockouts:* estudio original

In-silico **knockout predictions**

86% (68 of 79 cases) of the *in-silico* predictions were **consistent with the experimental data.**

Rerouting of metabolic fluxes
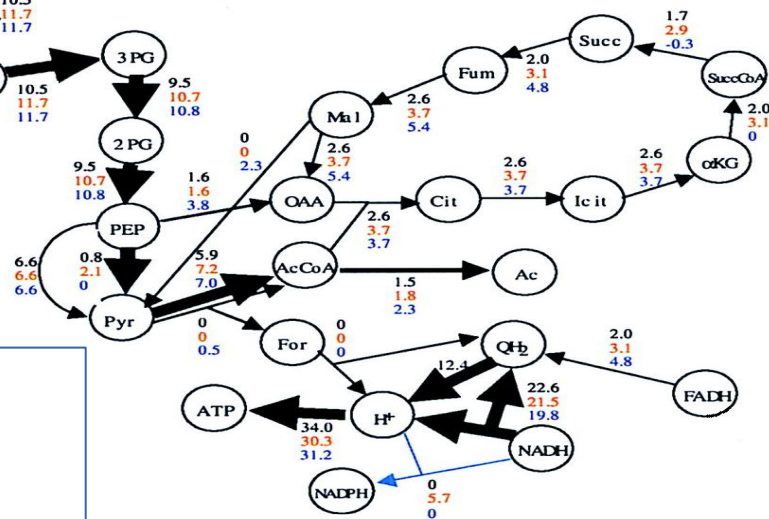
Wild type
zwf⁻ mutant
zwf⁻ pnt⁻ double mutant

Edwards JS, Palsson BO. PNAS (2000)

Predicción de *knockouts:* estudio original

In-silico knockout predictions

Rerouting of metabolic fluxes

- ■ Wild type
- ■ zwf⁻ mutant
- ■ zwf⁻ pnt⁻ double mutant

86% (68 of 79 cases) of the *in-silico* predictions were **consistent with the experimental data.**

Edwards JS, Palsson BO. PNAS (2000)
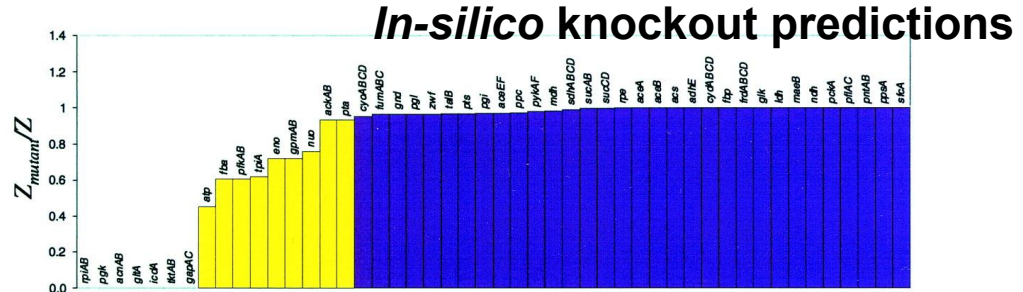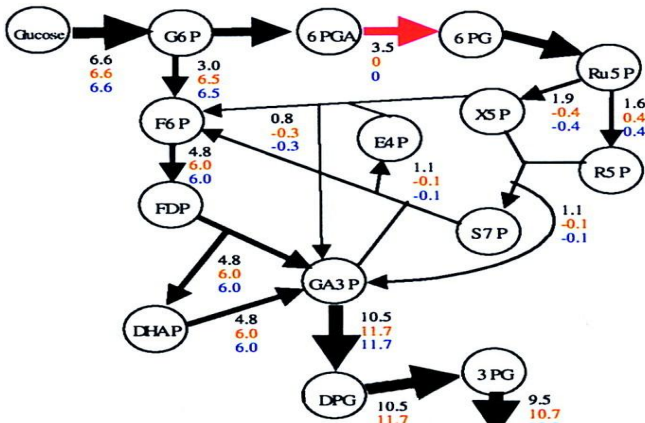
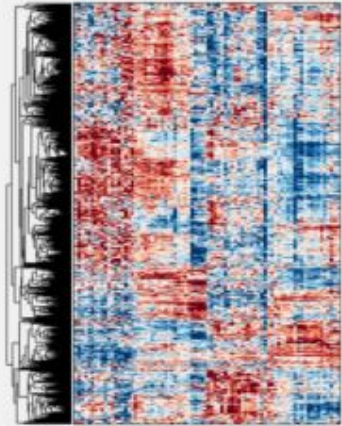# Metabolic modeling in humans

Metabolic modeling in humans
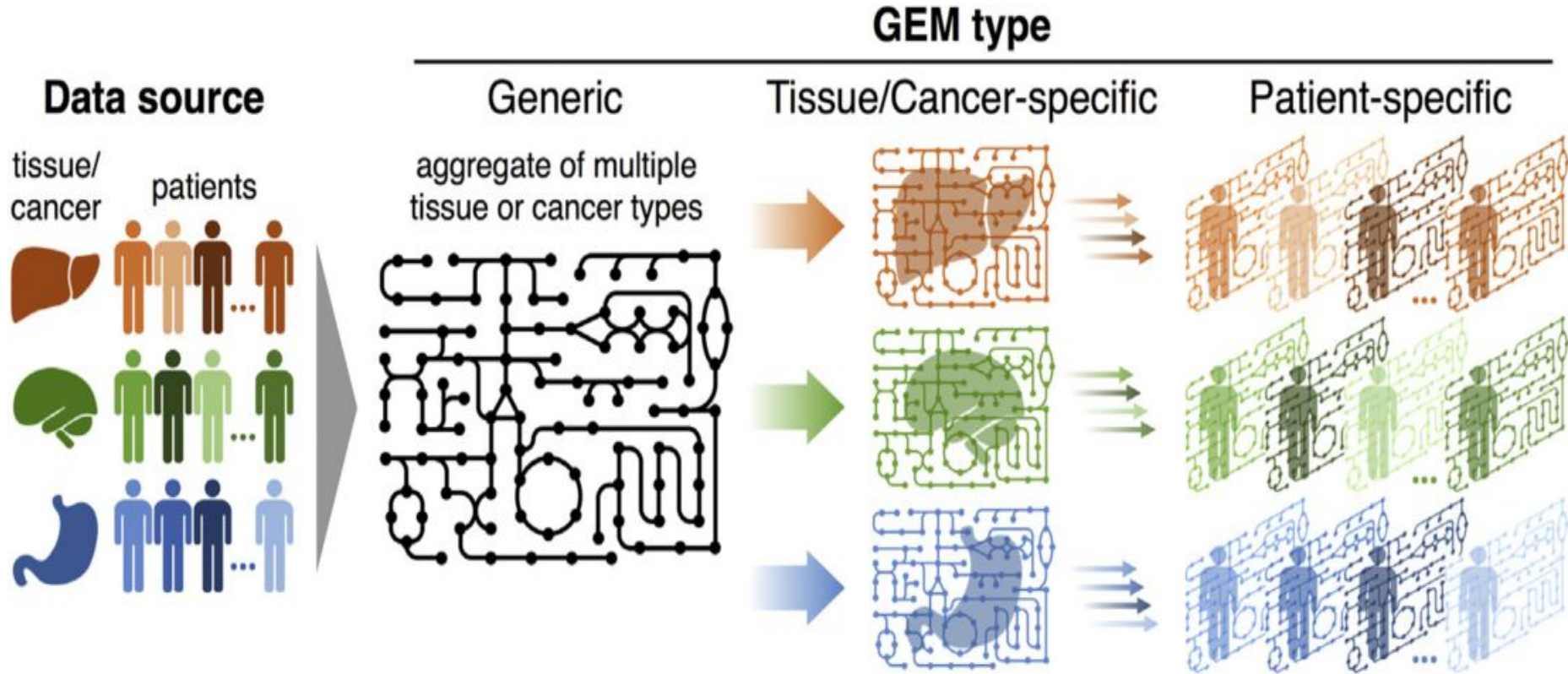
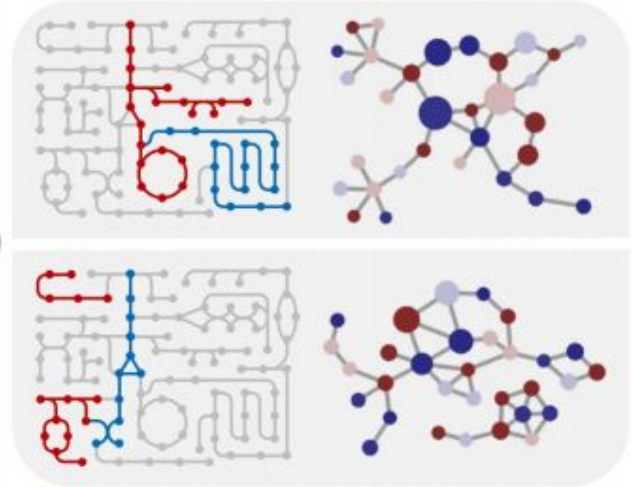# Context-Specific Metabolic Modeling

collect omics data

map to GEM

identify subnetworks, reporters

Cell Context (omics)
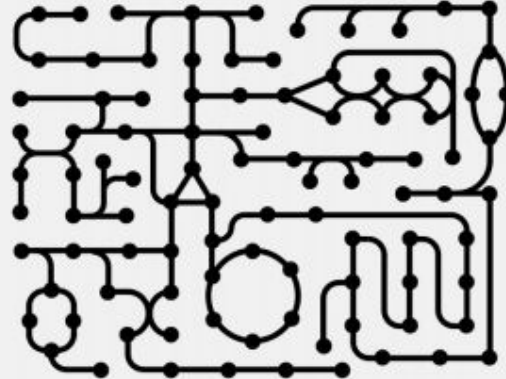
Universal Human metabolism

Cell-type specific metabolic models
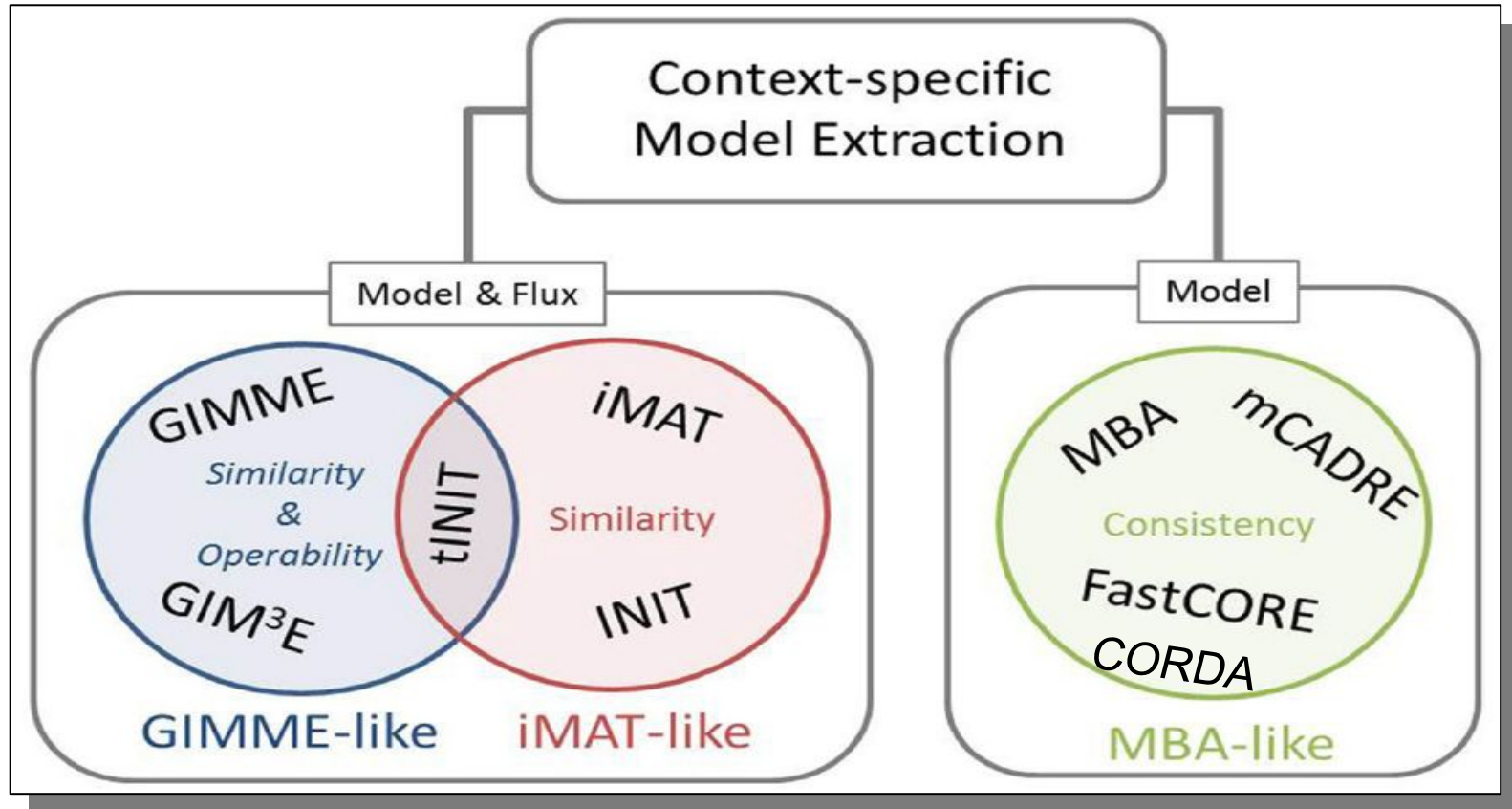
Genome-Scale Model of Human Metabolism

Table 1

Statistics of currently published generic human GEMs.

Generic GEMs	Genes	Metabolites[a]	Reactions[a]	Features
RECON1	1496	1509	3744	Manually reconstructed from bibliomics data
EHMN	2322	2671	2823	Manually reconstructed from bibliomics data
RECON2	1789	2626	7440	Merging EHMN and HepatoNet1 with RECON1
RECON 2.2	1675	5324	7785	Reconstructed by integrating previous versions, with emphasis on mass and charge balance
HMR1.0	1512	3397	4144	Reconstructed based on RECON1, EHMN, HumanCyc and KEGG
HMR2.0	3765	3160	8181	Reconstructed based on HMR1, with additional emphasis on lipid metabolism by integrating iAdipocytes1809, KEGG, Lipidomics Gateway
Recon3D	2248	5835	10600	Reconstructed based on RECON2 and includes mapping to 3D structure of proteins through PDB ids

Swainston, N., et al (2016). Metabolomics, 12(7), 109.

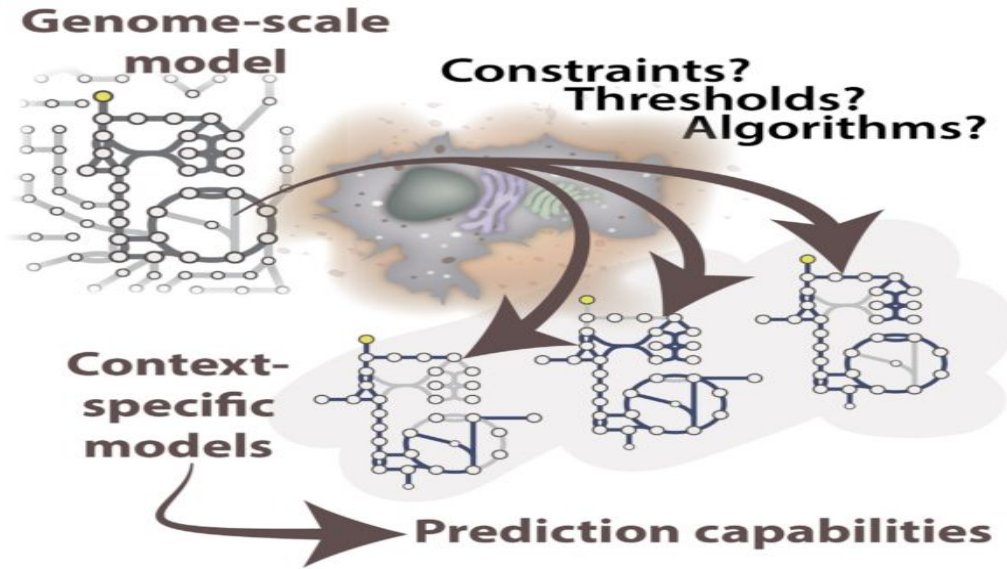- ○ Several options available (all derived from RECON1)
- ○ Recon3D is most recent version

Classes of Model Extraction Methods

Estevez, S. R. and Z. Nikoloski (2014). "Generalized framework for context-specific metabolic model extraction methods."

Algorithm and parameters

CSM depend on key decisions on methodology and data processing

Genome-scale model

Constraints?
Thresholds?
Algorithms?

Context-specific models

Prediction capabilities

- No strong evidence that one MEM universally gives the most accurate models
- Each method has different underlying assumptions that affect the resulting model
- **Gene thresholding** seams to be the most determinant decision