Introduction to Flux Balance Analysis

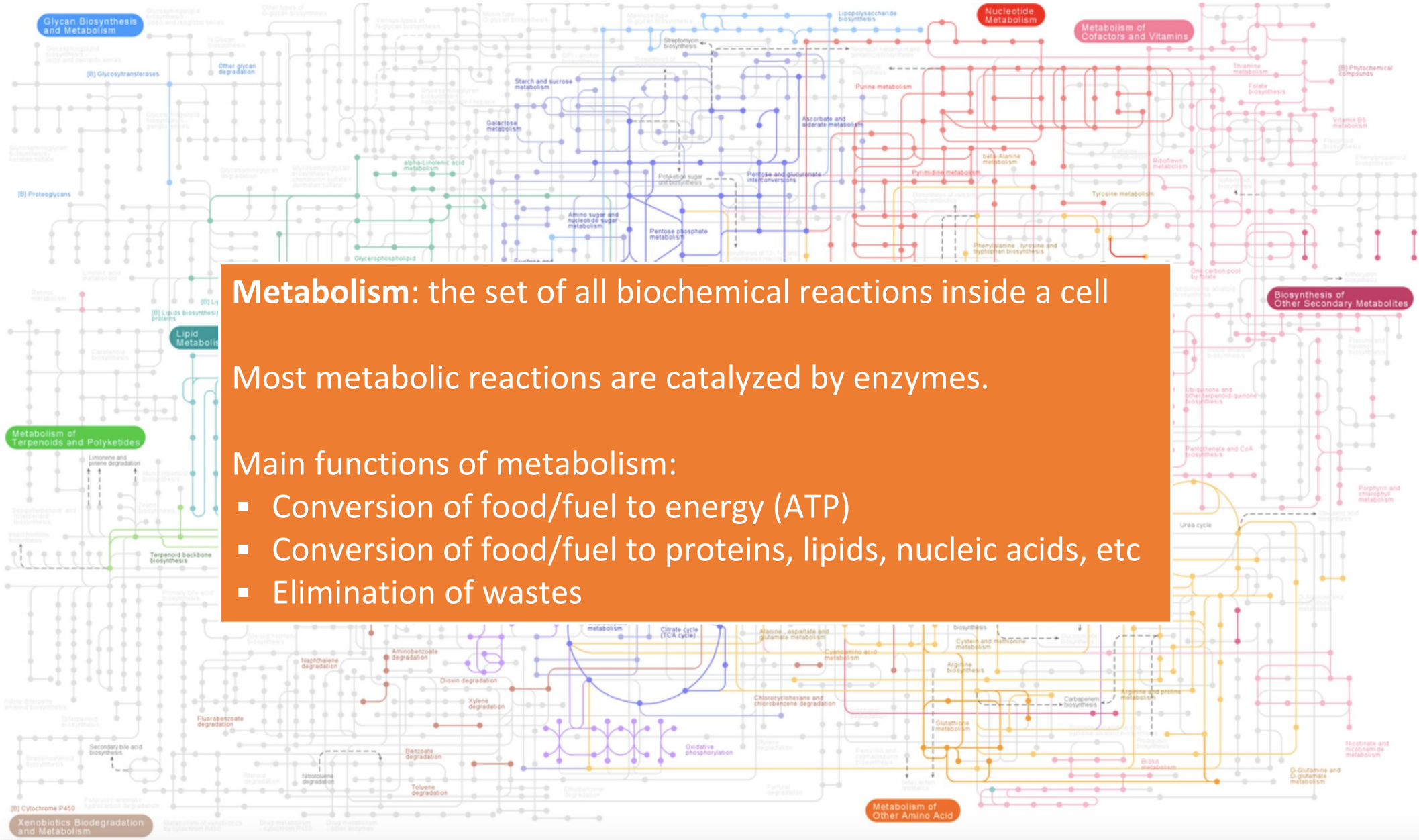
Keesha Erickson

keeshae@lanl.gov

qBio Summer School

June 21, 2018

Escherichia coli metabolic network



Metabolism: the set of all biochemical reactions inside a cell

Most metabolic reactions are catalyzed by enzymes.

Main functions of metabolism:

- Conversion of food/fuel to energy (ATP)
- Conversion of food/fuel to proteins, lipids, nucleic acids, etc
- Elimination of wastes

From the Kyoto Encycolopedia of Genes and Genomes,
http://www.genome.jp/kegg/

Metabolism provides important insights

Studying evolution

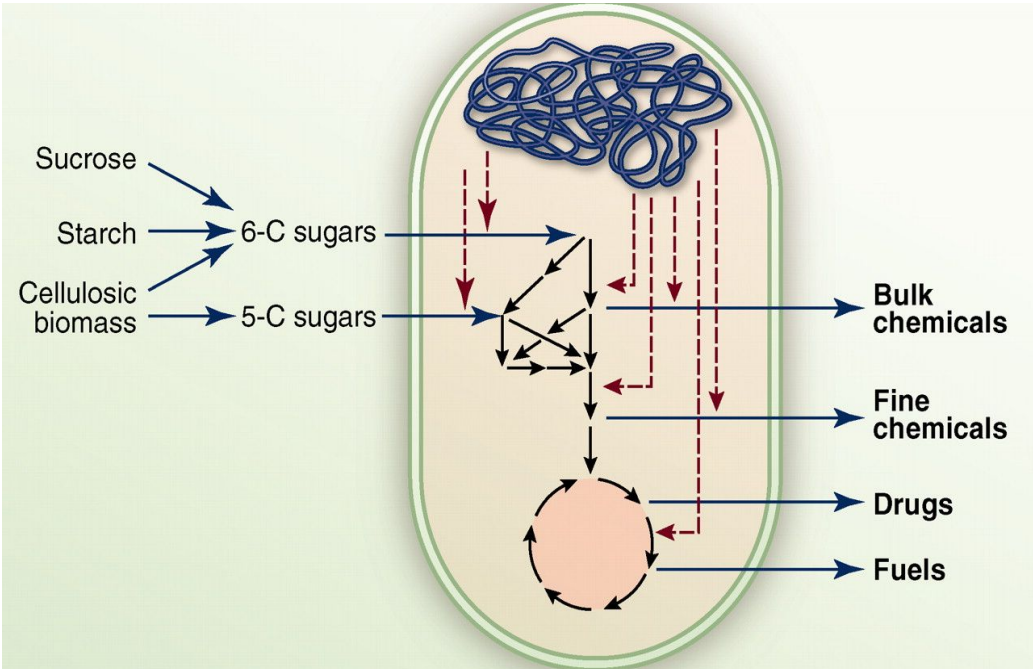
- Effects of horizontal gene transfer
- Effects of gene deletion

Prediction of essential genes

- Minimal genome

Metabolic engineering

- Optimal overproduction of metabolites
- Production coupled to growth



Orth, Fleming, Palsson (2010) EcoSal Plus.

Keasling. (2010) Science.

Flux Balance Analysis

FBA can be used to calculate the flow of metabolites through a metabolic network

FBA calculates **rates**

- Growth rate of an organism
- Production rate of a metabolite
- Yield of a product (production rate of product / consumption rate of substrate)

Typical FBA does not:

- Use kinetic parameters, so cannot predict metabolite concentrations or estimate changes over time
- Consider regulatory effects (gene expression, enzyme cascades, etc)

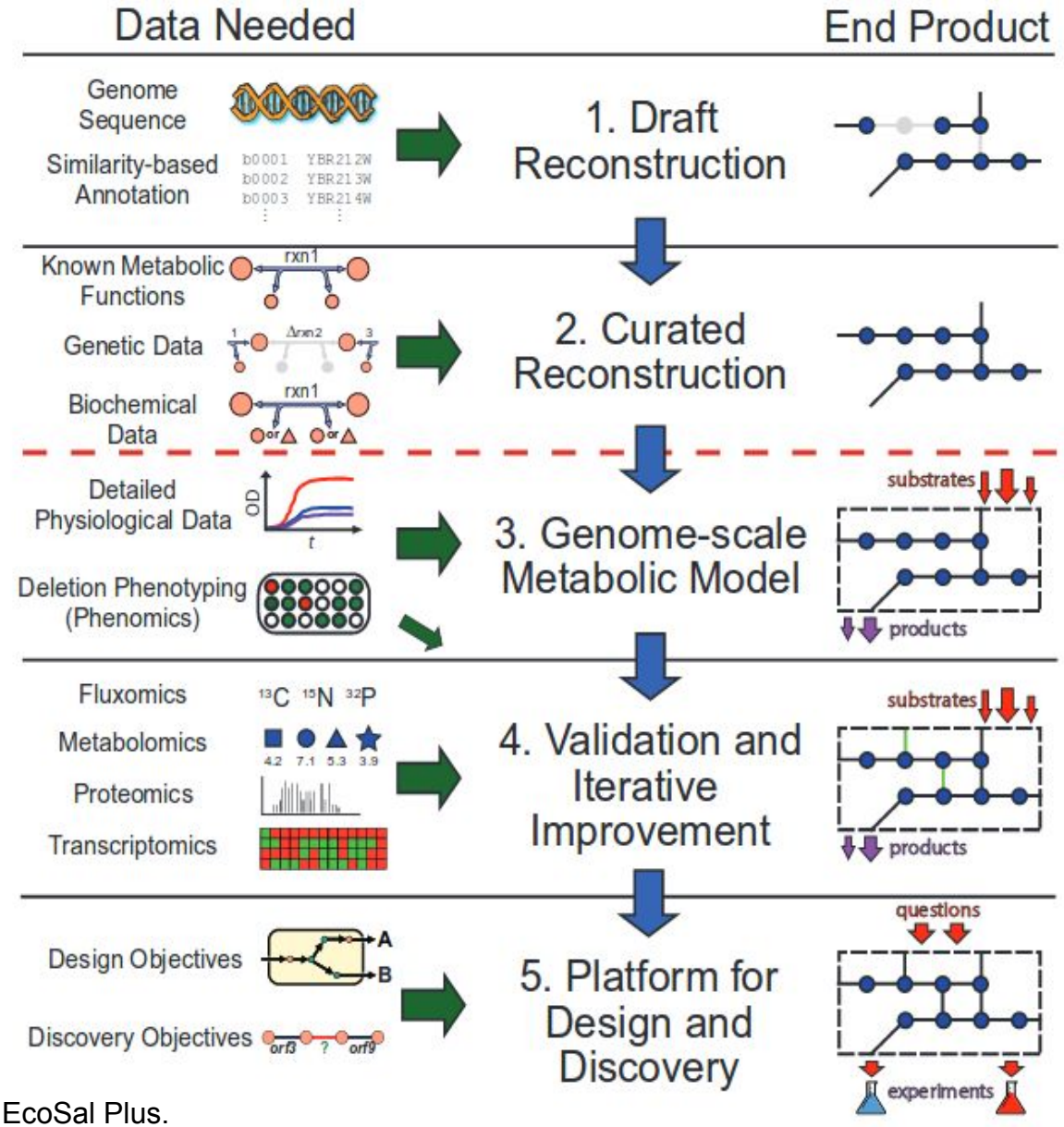
Orth et al. (2010) Nature Biotech.

Outline

- Metabolic network models
- Accounting for growth requirements
- Obtaining fluxes
- Tools for FBA

Metabolic Network Models

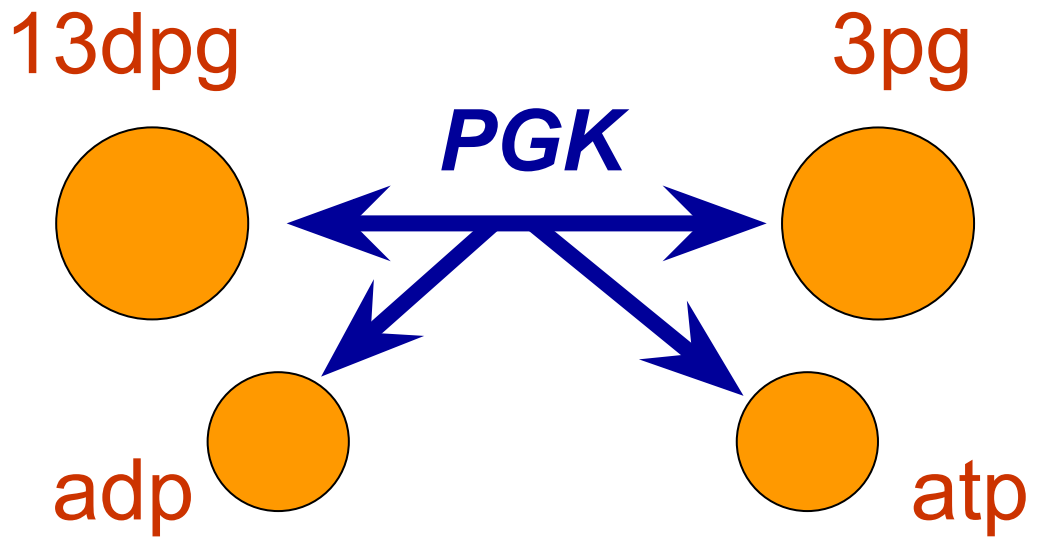
Building a genome-scale metabolic network model



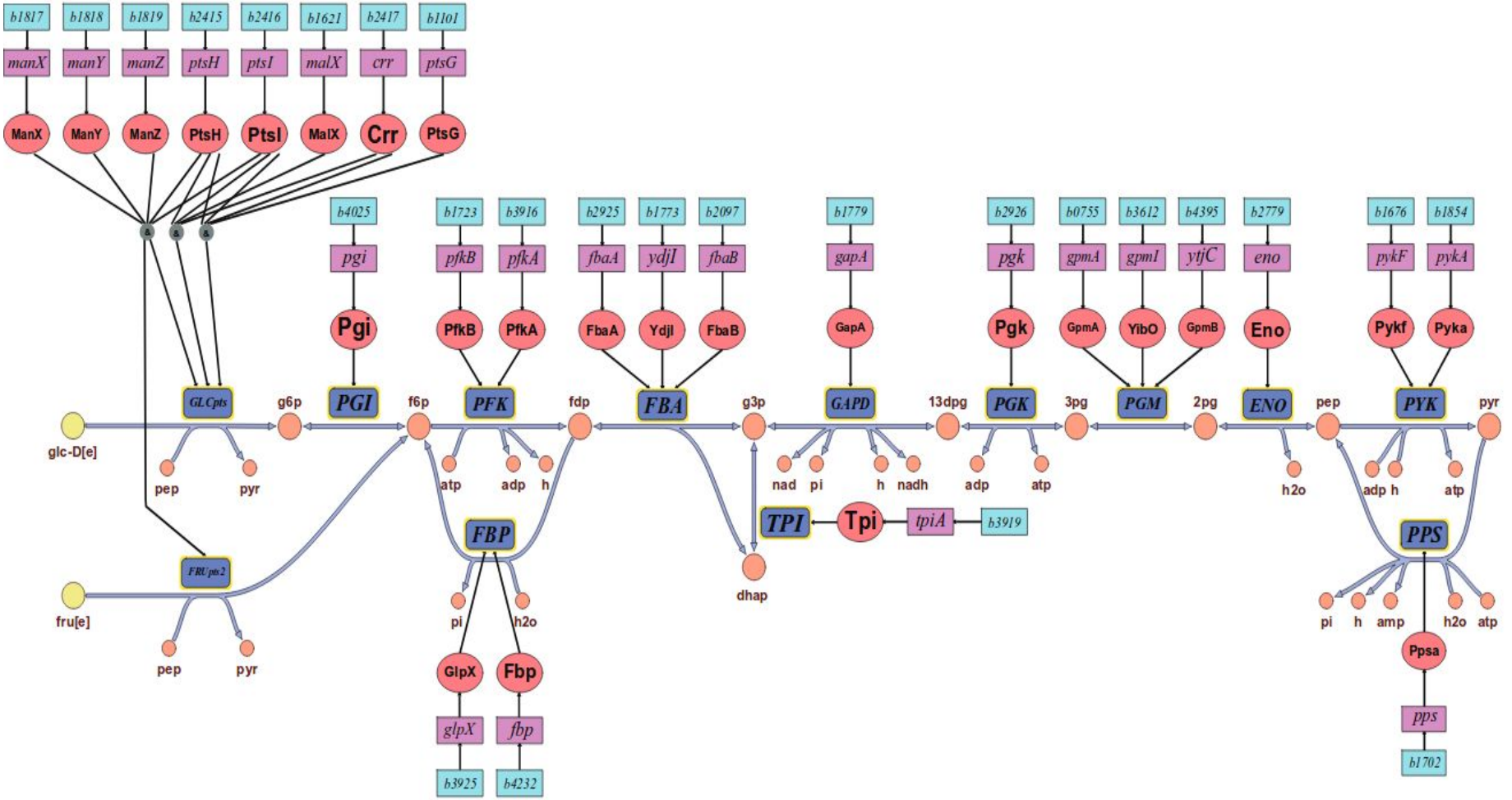
Orth, Fleming, Palsson (2010) EcoSal Plus.

Reactions are associated with specific genes

Gene:	*pgk*
Protein:	Pgk (Phosphoglycerate kinase)
Description:	In glycolysis, catalyzes the transfer of a phosphoryl group from 1,3-bisphospho-D-glycerate to ADP, forming ATP and 3-phospho-D-glycerate
Reaction:	13dpg[c] + adp[c] <=> 3pg[c] + atp[c]

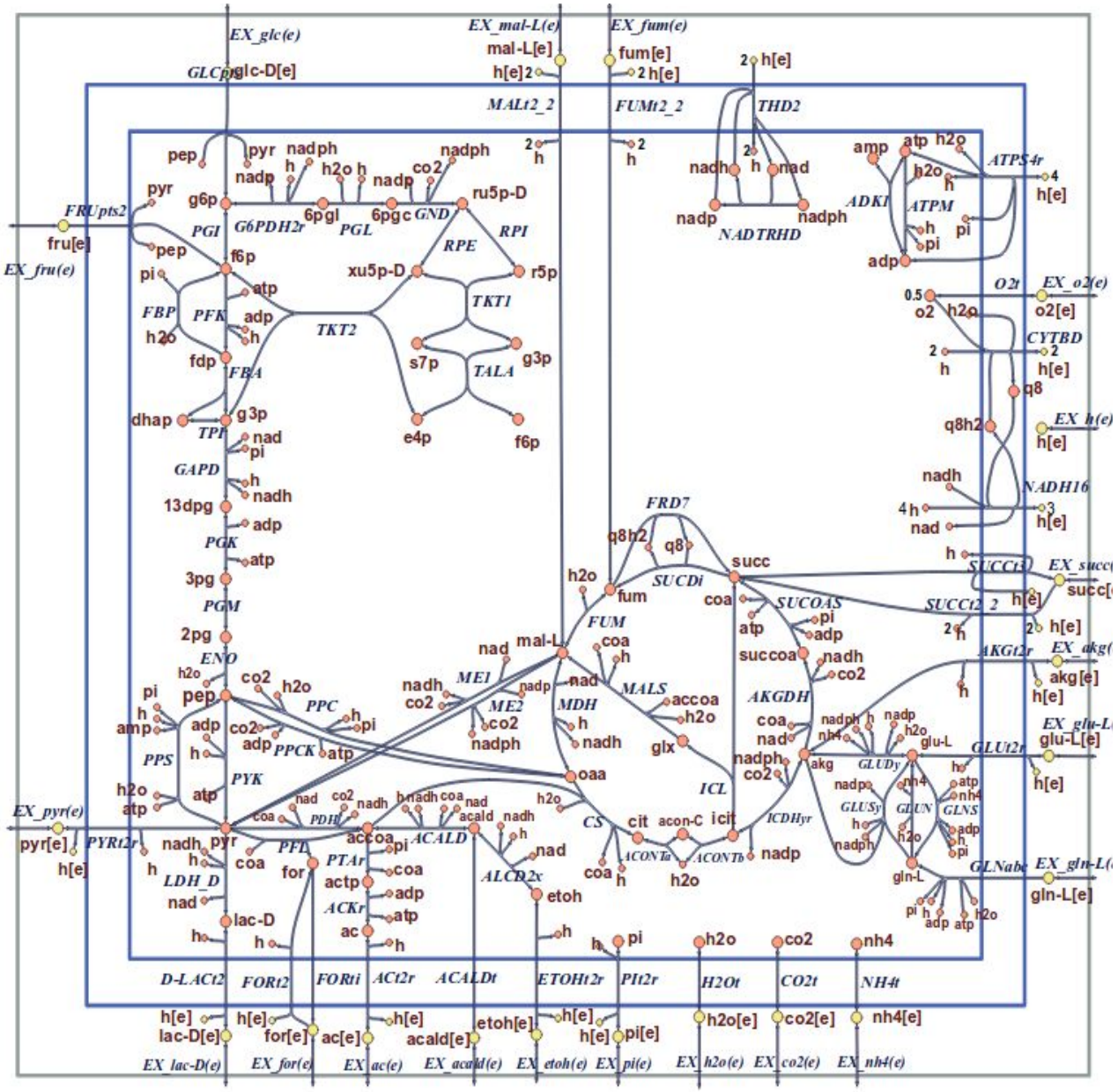


Sets of reactions can be assembled into subsystems



Orth, Fleming, Palsson (2010) EcoSal Plus.

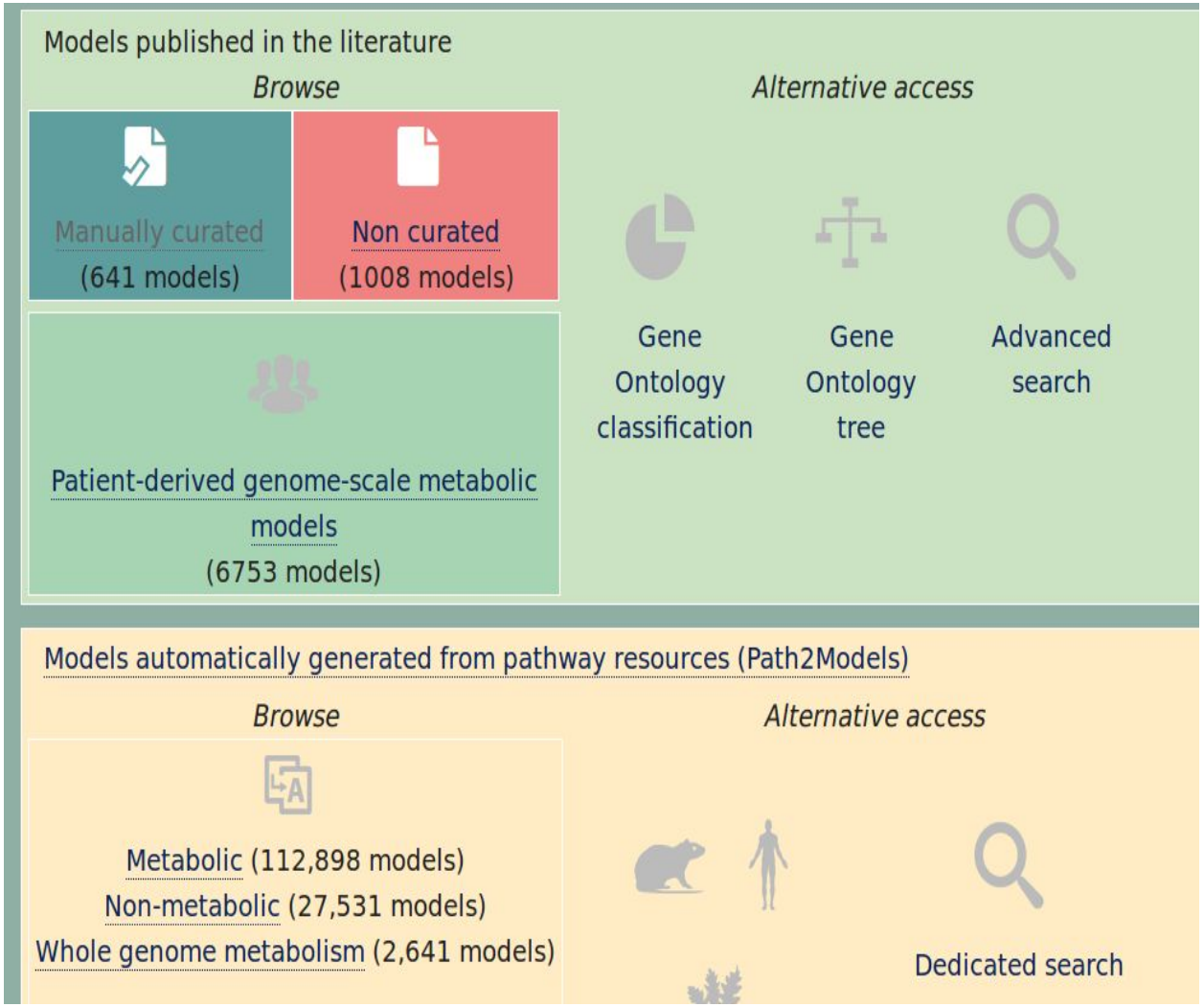
Metabolic network models



E. coli iJO1366:
- 2251 metabolic reactions
- 1136 unique metabolites

Orth, Fleming, Palsson (2010) EcoSal Plus.
Orth et al (2011) Mol Syst Biol.

BioModels Database



1990: first metabolic network model for *E. coli* (Majewski and Domach)
14 metabolic reactions

1997: *E. coli* genome sequenced

2000: first genome-scale network model for *E. coli* (Edwards & Palsson)
~ 720 metabolic reactions

2003: human genome sequenced

2007: first two metabolic network models for humans were published by Palsson group and Goryanin group
~ 3000 metabolic reactions

2011: *E. coli* model iJO1366 published (Orth et al)
~2000 metabolic reactions

2018: Rencon3D model of human metabolism published by Palsson group
~13000 metabolic reactions

https://www.ebi.ac.uk/biomodels-main/ Accessed May 23, 2018

Metabolites

Required attributes

Name
Description
Charged Formula
Charge
Compartment

Metabolite Abbreviation	Metabolite Name	Neutral Formula	Charged Formula	Charge	Compartment	KEGG ID	CAS Number	Alternate Names
10fthf[c]	10-Formyltetrahydrofolate	C20H23N7O7	C20H21N7O7	-2	Cytosol	C00234		10-Formyl-THF
12dgr120[c]	1,2-Diacyl-sn-glycerol (didodecanoyl, n-C12:0)	C27H52O5	C27H52O5	0	Cytosol	C00641		1,2-Diacylglycerol/ D-1,2-Diacylglycerol
12dgr140[c]	1,2-Diacyl-sn-glycerol (ditetradecanoyl, n-C14:0)	C31H60O5	C31H60O5	0	Cytosol	C00641		D-1,2-Diacylglycerol/ 1,2-Diacylglycerol
12dgr141[c]	1,2-Diacyl-sn-glycerol (ditetradec-7-enoyl, n-C14	C31H56O5	C31H56O5	0	Cytosol	C00641		D-1,2-Diacylglycerol/ 1,2-Diacylglycerol
12dgr160[c]	1,2-Diacyl-sn-glycerol (dihexadecanoyl, n-C16:0)	C35H68O5	C35H68O5	0	Cytosol	C00641		D-1,2-Diacylglycerol/ 1,2-Diacylglycerol
12dgr161[c]	1,2-Diacyl-sn-glycerol (dihexadec-9-enoyl, n-C16	C35H64O5	C35H64O5	0	Cytosol	C00641		D-1,2-Diacylglycerol/ 1,2-Diacylglycerol
12dgr180[c]	1,2-Diacyl-sn-glycerol (dioctadecanoyl, n-C18:0)	C39H76O5	C39H76O5	0	Cytosol	C00641		D-1,2-Diacylglycerol/ 1,2-Diacylglycerol
12dgr181[c]	1,2-Diacyl-sn-glycerol (dioctadec-11-enoyl, n-C1	C39H72O5	C39H72O5	0	Cytosol	C00641		1,2-Diacylglycerol/ D-1,2-Diacylglycerol
12ppd-R[c]	(R)-Propane-1,2-diol	C3H8O2	C3H8O2	0	Cytosol	C02912	4254-14-2	(R)-1,2-Propanediol/ (R)-Propylene glycol/ D
12ppd-S[c]	(S)-Propane-1,2-diol	C3H8O2	C3H8O2	0	Cytosol	C02917	4254-14-3	(S)-1,2-Propanediol/ (S)-Propylene glycol/ L-
13dpg[c]	3-Phospho-D-glyceroyl phosphate	C3H8O10P2	C3H4O10P2	-4	Cytosol	C00236		1,3-bis-phosphoglycerate/ 3-Phospho-D-gly
14dhncoa[c]	1,4-dihydroxy-2-napthoyl-CoA	C32H38N7O19P3S	C32H38N7O19P3S	-4	Cytosol	C15547		
14glucan[c]	1,4-alpha-D-glucan	C36H62O31	C36H62O31	0	Cytosol	C00718		
15dap[c]	1,5-Diaminopentane	C5H14N2	C5H16N2	2	Cytosol	C01672	462-94-2	Cadaverine/ 1,5-Pentanediamine/ Pentame

Reactions

Required attributes

Name
Description
Formula
Gene-reaction association
Gene(s)

Protein(s)
Cellular subsystem
Flux upper and lower bounds

Name	Description	Formula	Gene-Protein-Reaction Association	Gene-Reaction Association	Protein-Reaction Asso
ENO	enolase	2pg[c] <=> h2o[c] + pep[c]	Eno (b2779)	b2779	Eno
F6PA	fructose 6-phosphate aldolase	f6p[c] <=> dha[c] + g3p[c]	(Fsa (b0825)) or (TalC (b3946))	(b0825 or b3946)	(Fsa) or (TalC)
FBA	fructose-bisphosphate aldolase	fdp[c] <=> dhap[c] + g3p[c]	(FbaB (b2097)) or (B1773 (b1773)) or	(b2097 or b1773 or b2925)	(FbaB) or (B1773) or (
FBP	fructose-bisphosphatase	fdp[c] + h2o[c] -> f6p[c] + pi[c]	(GlpX (b3925)) or (Fbp (b4232)) or (Y	(b3925 or b4232 or b2930)	(GlpX) or (Fbp) or (Yg
G1PPpp	Glucose-1-phosphatase	g1p[p] + h2o[p] -> glc-D[p] + pi[p]	Agp (b1002)	b1002	Agp
G6PP	glucose-6-phosphate phosphatase	g6p[c] + h2o[c] -> glc-D[c] + pi[c]	YbiV (b0822)	b0822	YbiV
GAPD	glyceraldehyde-3-phosphate dehydrogenase	g3p[c] + nad[c] + pi[c] <=> 13dpg[c] + h[c] + nadh[c]	GapA (b1779)	b1779	GapA
GLBRAN2	1,4-alpha-glucan branching enzyme (glycogen	glycogen[c] -> bglycogen[c]	GlgB (b3432)	b3432	GlgB
GLCP	glycogen phosphorylase	glycogen[c] + pi[c] -> g1p[c]	(GlgP (b3428)) or (MalP (b3417))	(b3428 or b3417)	(GlgP) or (MalP)
GLCP2	glycogen phosphorylase	bglycogen[c] + pi[c] -> g1p[c]	(MalP (b3417)) or (GlgP (b3428))	(b3417 or b3428)	(MalP) or (GlgP)
GLCS1	glycogen synthase (ADPGlc)	adpglc[c] -> adp[c] + glycogen[c] + h[c]	GlgA (b3429)	b3429	GlgA
GLDBRAN2	glycogen debranching enzyme (bglycogen -> g	bglycogen[c] -> glycogen[c]	GlgX (b3431)	b3431	GlgX
GLGC	glucose-1-phosphate adenylyltransferase	atp[c] + g1p[c] + h[c] -> adpglc[c] + ppi[c]	GlgC (b3430)	b3430	GlgC
HEX1	hexokinase (D-glucose:ATP)	atp[c] + glc-D[c] -> adp[c] + g6p[c] + h[c]	Glk (b2388)	b2388	Glk
PDH	pyruvate dehydrogenase	coa[c] + nad[c] + pyr[c] -> accoa[c] + co2[c] + nadh[c]	(AceEec (b0114) and AceEec (b0115) and	(b0114 and b0115 and b0116)	(AceEec and AceEec and

Accounting for growth requirements

The Biomass Reaction

To predict growth rate, we need to estimate the rate at which metabolites are converted to biomass constituents (e.g., nucleic acids, proteins, lipids)

The “biomass reaction” predicts the exponential growth rate (μ) of the organism.

Coefficients on metabolites are experimentally determined.

Biomass reaction for *E. coli* iJO1366:

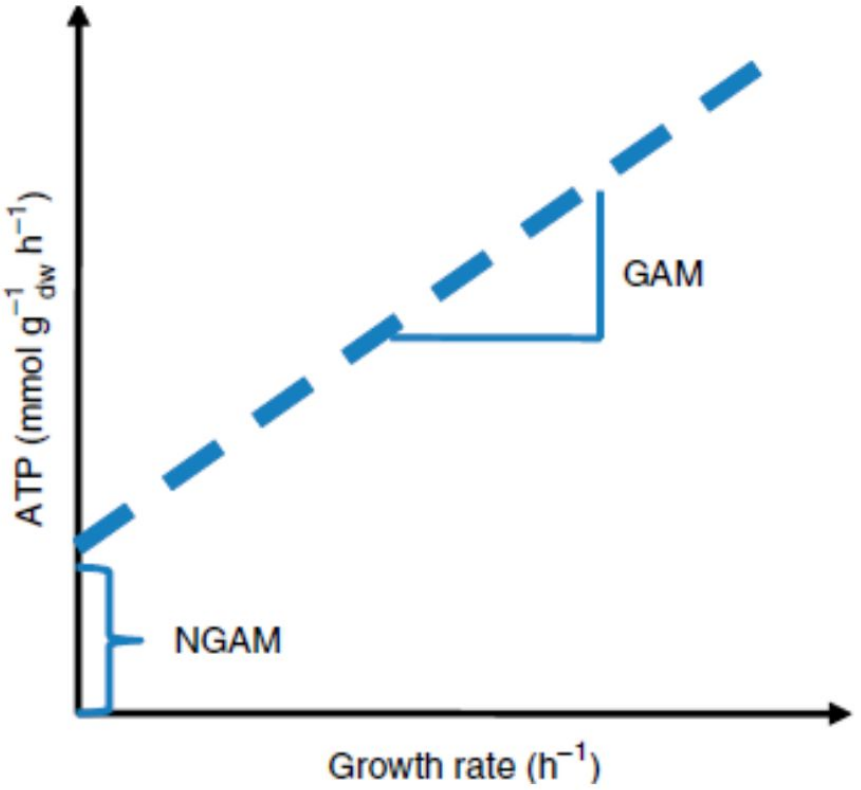
0.000223 10fthf[c] + 0.000223 2dmmql8[c] + 2.5e-005 2fe2s[c] + 0.000248 4fe4s[c] + 0.000223 5mthf[c] + 0.000279 accoa[c] + 0.000223 adocbl[c] + 0.49915 ala-L[c] + 0.000223 amet[c] + 0.28742 arg-L[c] + 0.23423 asn-L[c] + 0.23423 asp-L[c] + 54.12 atp[c] + 0.000116 bmocogdp[c] + 2e-006 btn[c] + 0.004952 ca2[c] + 0.000223 chor[c] + 0.004952 cl[c] + 0.000168 coa[c] + 2.4e-005 cobalt2[c] + 0.1298 ctp[c] + 0.000674 cu2[c] + 0.088988 cys-L[c] + 0.024805 datp[c] + 0.025612 dctp[c] + 0.025612 dgtp[c] + 0.024805 dttp[c] + 0.000223 enter[c] + 0.000223 fad[c] + 0.006388 fe2[c] + 0.007428 fe3[c] + 0.25571 gln-L[c] + 0.25571 glu-L[c] + 0.5953 gly[c] + 0.15419 glycogen[c] + 0.000223 gthrd[c] + 0.20912 gtp[c] + 48.7529 h2o[c] + 0.000223 hemeO[c] + 0.092056 his-L[c] + 0.28231 ile-L[c] + 0.18569 k[c] + 0.43778 leu-L[c] + 3e-006 lipopb[c] + 0.33345 lys-L[c] + 3.1e-005 malcoa[c] + 0.14934 met-L[c] + 0.008253 mg2[c] + 0.000223 mlthf[c] + 0.000658 mn2[c] + 7e-006 mobd[c] + 7e-006 mococdp[c] + 7e-006 mocogdp[c] + 0.000223 mql8[c] + 0.001787 nad[c] + 4.5e-005 nadh[c] + 0.000112 nadp[c] + 0.000335 nadph[c] + 0.012379 nh4[c] + 0.000307 ni2[c] + 0.012366 pe160[c] + 0.009618 pe161[c] + 0.004957 pe181[c] + 0.005707 pg160[c] + 0.004439 pg161[c] + 0.002288 pg181[c] + 0.18002 phe-L[c] + 0.000223 pheme[c] + 0.2148 pro-L[c] + 0.03327 ptrc[c] + 0.000223 pydx5p[c] + 0.000223 q8h2[c] + 0.000223 ribflv[c] + 0.20968 ser-L[c] + 0.000223 sheme[c] + 0.004126 so4[c] + 0.006744 spmd[c] + 9.8e-005 succoa[c] + 0.000223 thf[c] + 0.000223 thmpp[c] + 0.24651 thr-L[c] + 0.055234 trp-L[c] + 0.13399 tyr-L[c] + 5.5e-005 udcpdp[c] + 0.1401 utp[c] + 0.41118 val-L[c] + 0.000324 zn2[c] + 0.008151 colipa[e] + 0.002944 clpn160[p] + .00229 clpn161[p] + 0.00118 clpn181[p] + 0.001345 murein3p3p[p] + 0.000605 murein3px4p[p] + 0.005381 murein4p4p[p] + 0.005448 murein4px4p[p] + 0.000673 murein4px4px4p[p] + 0.031798 pe160[p] + 0.024732 pe161[p] + 0.012747 pe181[p] + 0.004892 pg160[p] + 0.003805 pg161[p] + 0.001961 pg181[p]
-> 53.95 adp[c] + 53.95 h[c] + 53.9459 pi[c] + 0.74983 ppi[c]

Energy requirements

There are two reactions that account for energy required to maintain viability

Growth associated maintenance (GAM) represents ATP needed for replication.
It is included as part of the biomass reaction.

Non-growth associated maintenance (NGAM) accounts for
all other energy needs, and the constraint on this reaction is experimentally determined.



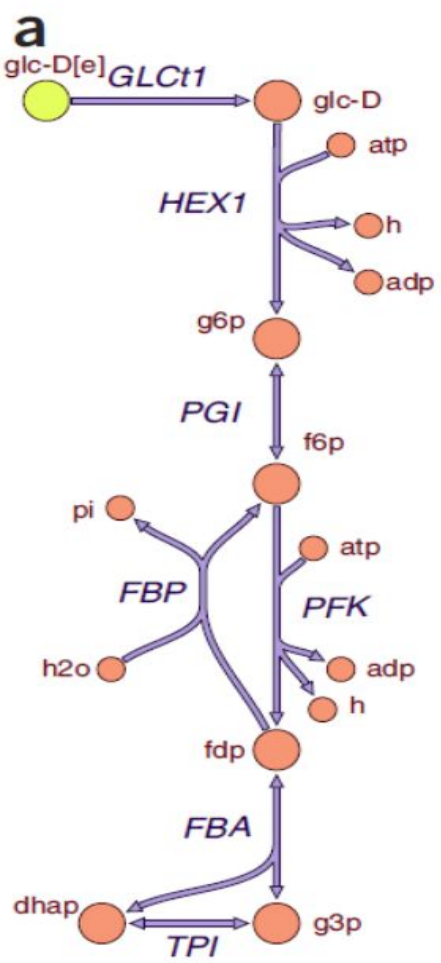
atp[c] + h2o[c] -> adp[c] + h[c] + pi[c]

(Lower bound of 3.15 mmol/gdcw/hr in iJO1366)

Thiele & Palsson (2010) Nature Protocols.

FBA: Calculating fluxes for all reactions in the network, given an objective and constraints

Stoichiometric matrix



b

	GLCt1	HEX1	PGI	PFK	FBP	FBA	TPI	EX_glc
glc-D[e]	−1	0	0	0	0	0	0	−1
glc-D	1	−1	0	0	0	0	0	0
atp	0	−1	0	−1	0	0	0	0
H	0	1	0	1	0	0	0	0
adp	0	1	0	1	0	0	0	0
g6p	0	1	−1	0	0	0	0	0
f6p	0	0	1	−1	1	0	0	0
fdp	0	0	0	1	−1	−1	0	0
pi	0	0	0	0	1	0	0	0
h2o	0	0	0	0	−1	0	0	0
g3p	0	0	0	0	0	1	1	0
dhap	0	0	0	0	0	1	−1	0

***S* matrix:**

Reactions in columns
Rows are metabolites

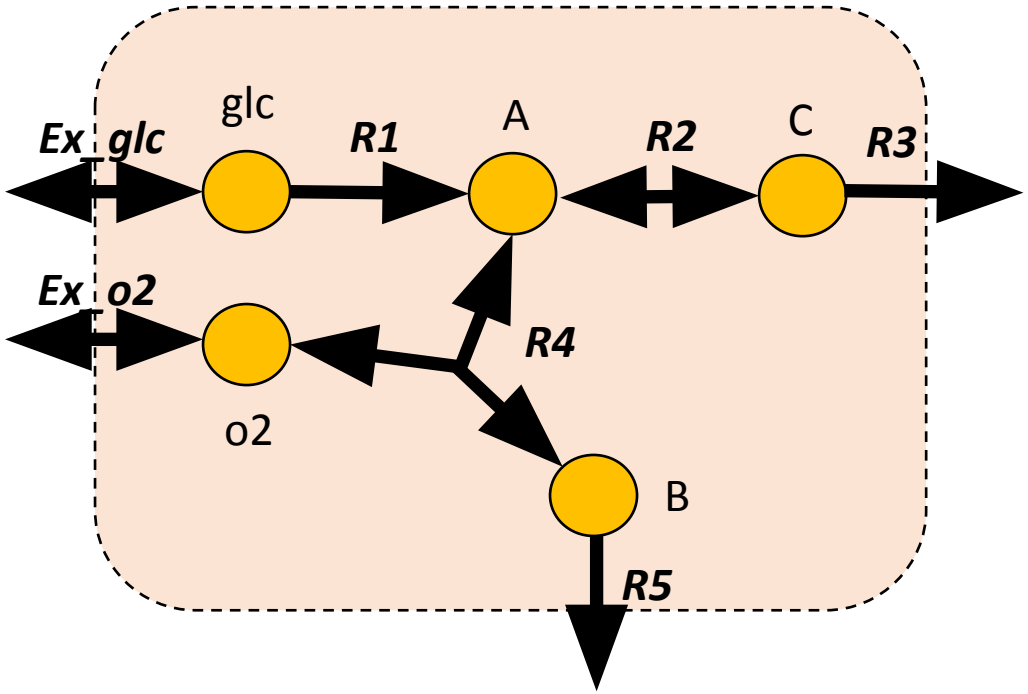
Negative indicates consumption (reactant)

Positive indicates production (product)

Becker et al (2007) Nature Protocols.

Exercise 1

Write the *S* for the following system:



Reactions

Ex_glc	glc <=> 0
Ex_o2	o2 <=> 0
R1	glc -> A
R2	A <=> C
R3	C -> 0
R4	A + o2 <=> B
R5	B -> 0

The steady state assumption

The inner product of the stoichiometric matrix *S* (size m x r) and
the flux vector ***v*** (length r) gives the change in metabolite concentrations
over time (d**x**/dt), where **x** is a vector of metabolite concentrations (length m).

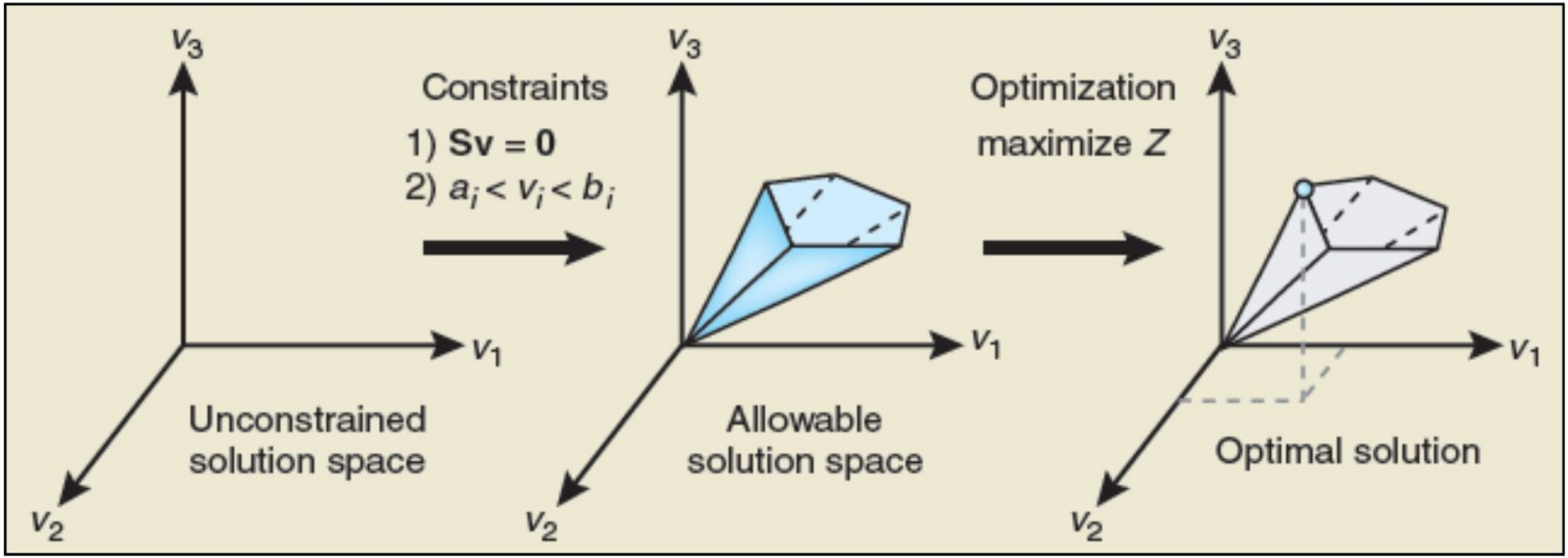
$$\frac{d\mathbf{x}}{dt} = S \cdot \mathbf{v}$$

We are interested in solving for ***v***.
Assuming the cell is in one phenotype for a time longer than it takes for metabolite concentrations to change dramatically, we make the steady state assumption:

$$\frac{d\mathbf{x}}{dt} = 0 = S \cdot \mathbf{v}$$

Now we can solve for ***v***.
However, as there are many more reactions (unknown variables) than metabolites, there will not be one unique solution.
Thus, it is helpful to impose constraints.

Constraining the solution space



Orth & Thiele (2010) Nature Biotech.
Becker et al (2007) Nature Protocols.

The linear programming problem

Linear programming: optimizing a linear function subject to various constraints

Canonical form:

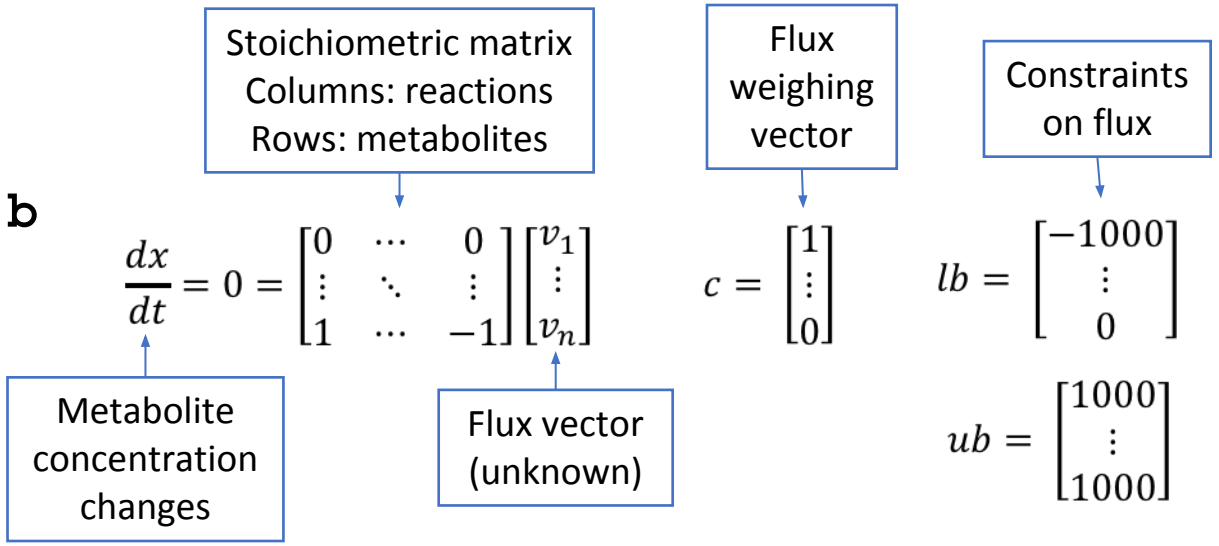
maximize $\mathbf{c}^T\mathbf{x}$
subject to $A\mathbf{x} \leq \mathbf{b}$
and $\mathbf{x} \geq 0$

Determine $\mathbf{x}$ given A and $\mathbf{b}$

For FBA:

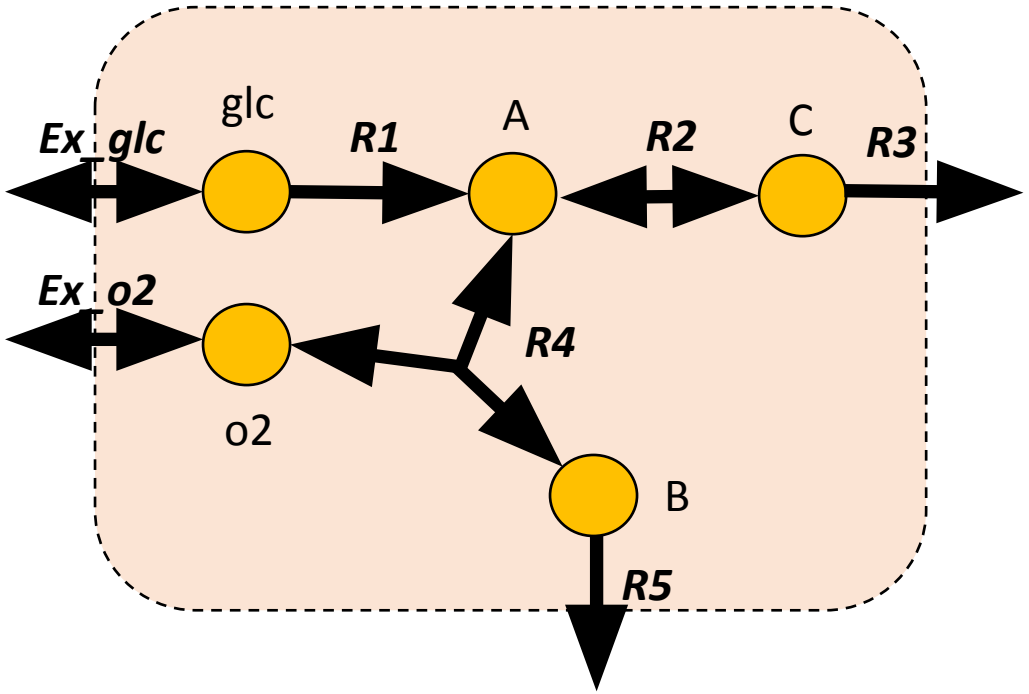
maximize $\mathbf{c}^T\mathbf{v}$
subject to $S\mathbf{v} = 0$
and $\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub}$

Determine $\mathbf{v}$ given S and $\mathbf{lb}$, $\mathbf{ub}$



Exercise 2

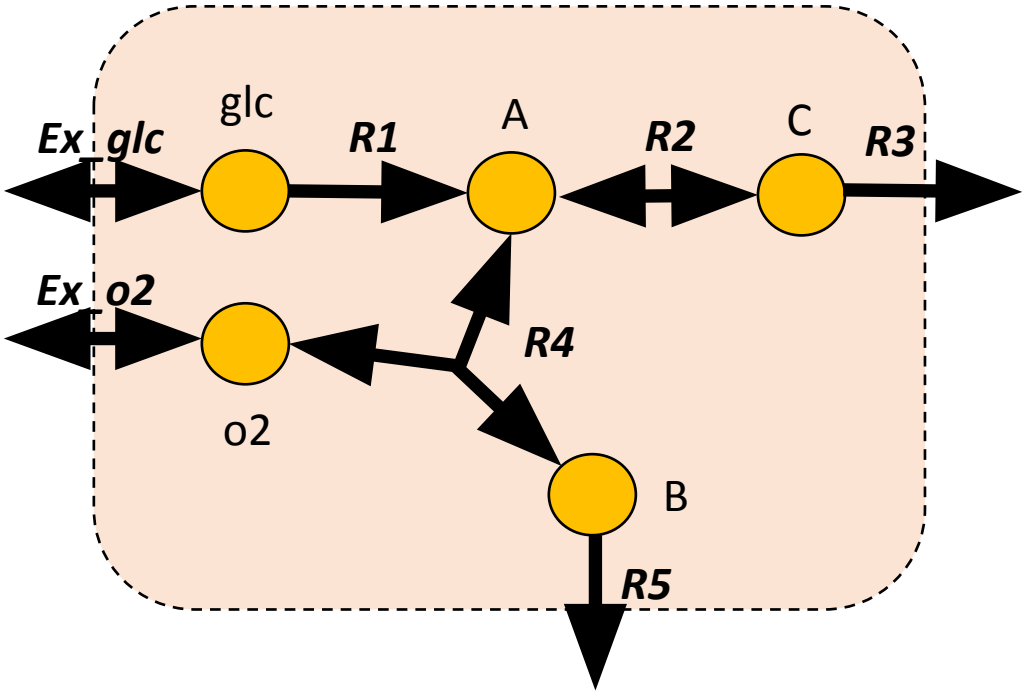
Write *S*, **c**, **lb**, and **ub** for the following system, where you want to maximize production of species C.



Reactions

Ex_glc	glc <=> 0
Ex_o2	o2 <=> 0
R1	glc -> A
R2	A <=> C
R3	C -> 0
R4	A + o2 <=> B
R5	B -> 0

Exercise 2



Reactions

Ex_glc	glc <=> 0
Ex_o2	o2 <=> 0
R1	glc -> A
R2	A <=> C
R3	C -> 0
R4	A + o2 <=> B
R5	B -> 0

$$S = \begin{bmatrix} -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix} \begin{matrix} \text{glc} \\ \text{o2} \\ \text{A} \\ \text{B} \\ \text{C} \end{matrix}$$

Important constraints

Reversibility

Substrates/media conditions

- Carbon source
- Nitrogen source

Varma & Palsson (1994) Applied Environmental Biology:

Flux balance models of metabolism use stoichiometry of metabolic pathways, metabolic demands of growth, and optimality principles to predict metabolic flux distribution and cellular growth under specified environmental conditions. These models have provided a mechanistic interpretation of systemic metabolic physiology, and they are also useful as a quantitative tool for metabolic pathway design. Quantitative predictions of cell growth and metabolic by-product secretion that are experimentally testable can be obtained from these models. In the present report, we used independent measurements to determine the model parameters for the wild-type *Escherichia coli* strain W3110. We experimentally determined the maximum oxygen utilization rate (15 mmol of O_2 per g [dry weight] per h), the maximum aerobic glucose utilization rate (10.5 mmol of Glc per g [dry weight] per h), the maximum anaerobic glucose utilization rate (18.5 mmol of Glc per g [dry weight] per h),

Minimum growth rate

Selecting the objective

What do you want to maximize/minimize?

Common objectives:

Objective	**Rationale**	**Example reaction**
Biomass reaction	Biologically relevant – safe to assume organism tries to optimize growth	biomass
Transport reaction	Calculate maximum theoretical yield or production rate	EX_etoh(e)

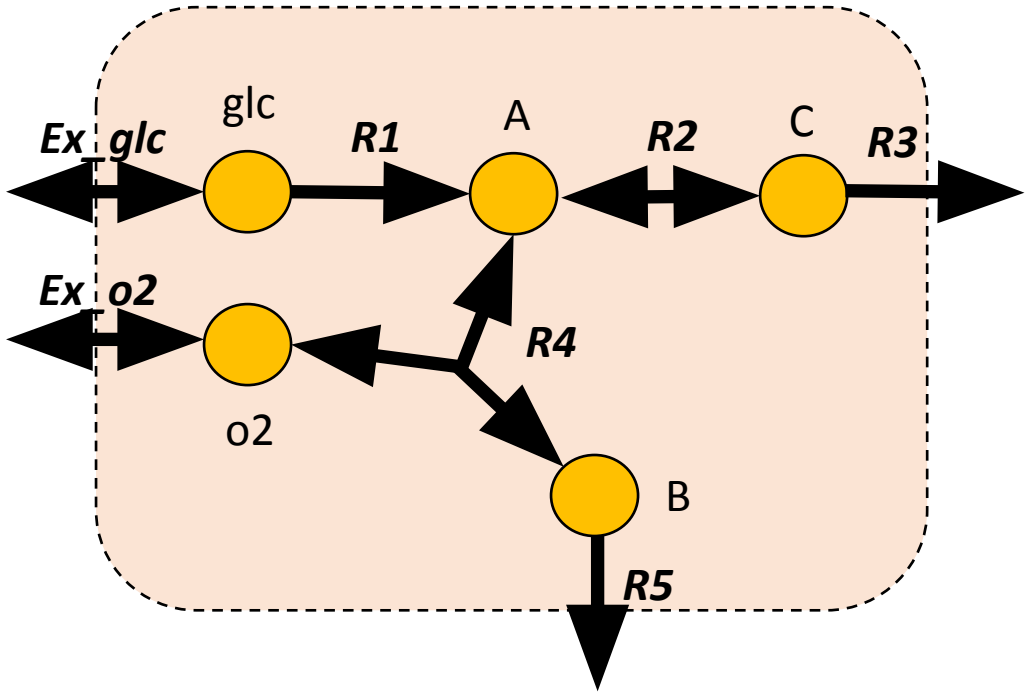
Exercise 3

Adjust *S*, **c**, **lb**, and **ub** to simulate

a. anaerobic conditions

b. aerobic conditions

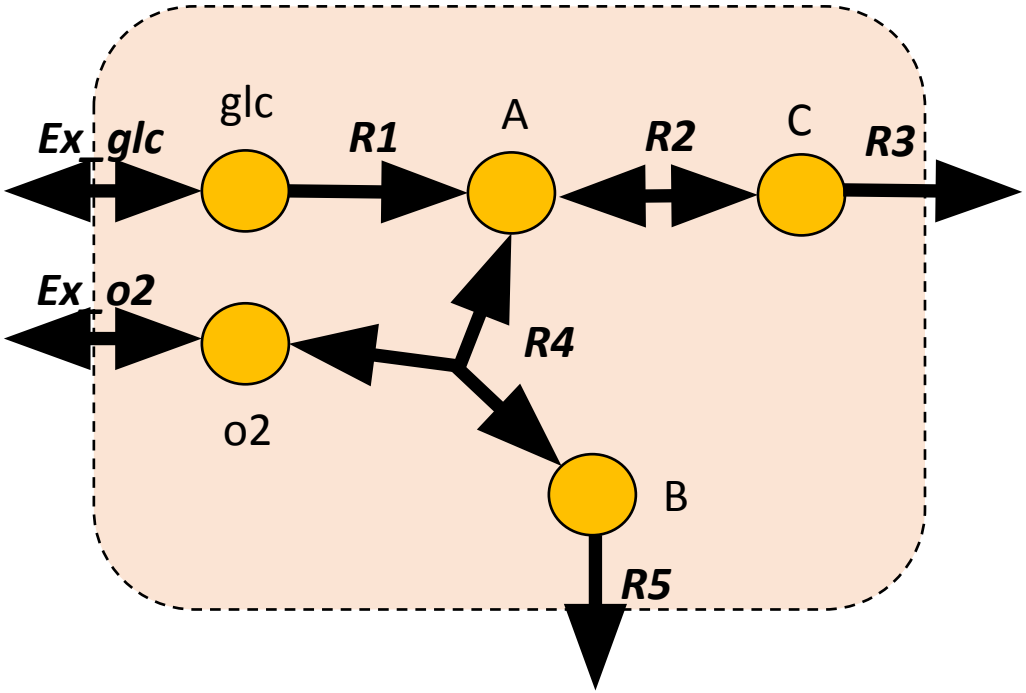
maximizing production of species C.



Reactions

Ex_glc	glc <=> 0
Ex_o2	o2 <=> 0
R1	glc -> A
R2	A <=> C
R3	C -> 0
R4	A + o2 <=> B
R5	B -> 0

Exercise 3



Reactions

Ex_glc glc <=> 0
Ex_o2 o2 <=> 0
R1 glc -> A
R2 A <=> C
R3 C -> 0
R4 A + o2 <=> B
R5 B -> 0

$$S = \begin{bmatrix} -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix} \begin{matrix} \text{glc} \\ \text{o2} \\ \text{A} \\ \text{B} \\ \text{C} \end{matrix}$$

Duality

PRIMAL PROBLEM

$$\text{Maximize} \quad Z = \boldsymbol{c}^T \boldsymbol{v} \qquad (8)$$

$$\text{Subject to} \quad \boldsymbol{S}\boldsymbol{v} = 0$$

$$\boldsymbol{v}^{LB} \geq \boldsymbol{v} \geq \boldsymbol{v}^{LB}$$

Primal solution gives a set of optimal fluxes

DUAL PROBLEM

$$\text{Minimize} \quad \boldsymbol{q}_1{}^T \boldsymbol{v}^{LB} + \boldsymbol{q}_2{}^T \boldsymbol{v}^{UB} \qquad (9)$$

$$\text{Subject to} \quad \boldsymbol{c}^T = \boldsymbol{\lambda}^T \boldsymbol{S} + \boldsymbol{q}_1{}^T + \boldsymbol{q}_2{}^T$$

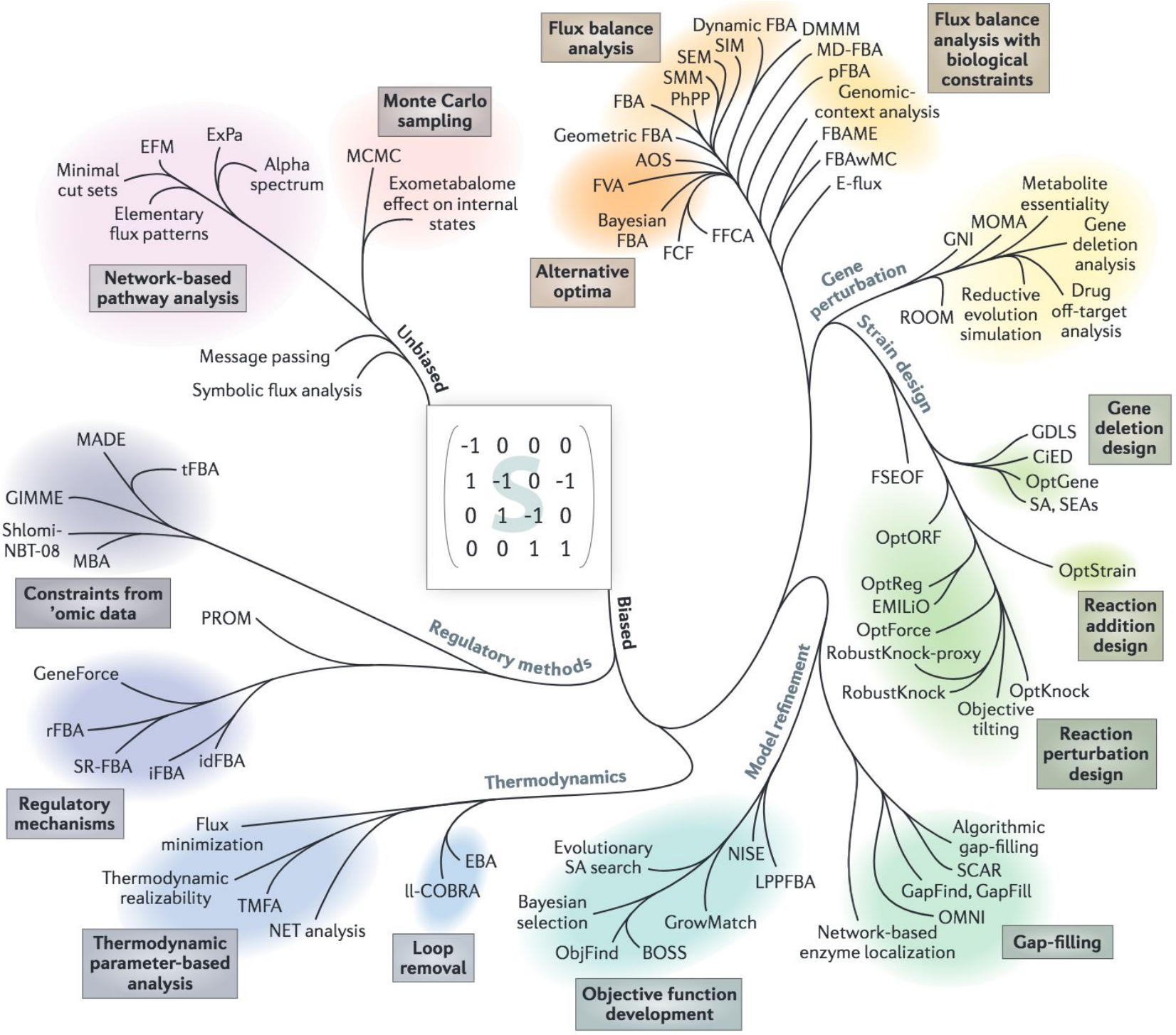
$$\boldsymbol{q}_1 \leq 0, \boldsymbol{q}_2 \geq 0$$

Dual solution gives the *shadow price* for each metabolite

- **"Sensitivity of the objective function to each steady state metabolite constraint"**
- **In economic terms, the marginal cost / marginal utility of relaxing a constraint**
- **Provides a way to find which metabolites have greatest impact on the solution**
  - **Very negative shadow prices influence objective function more**

Reznik et al. (2013) PLOS Comp. Biol.

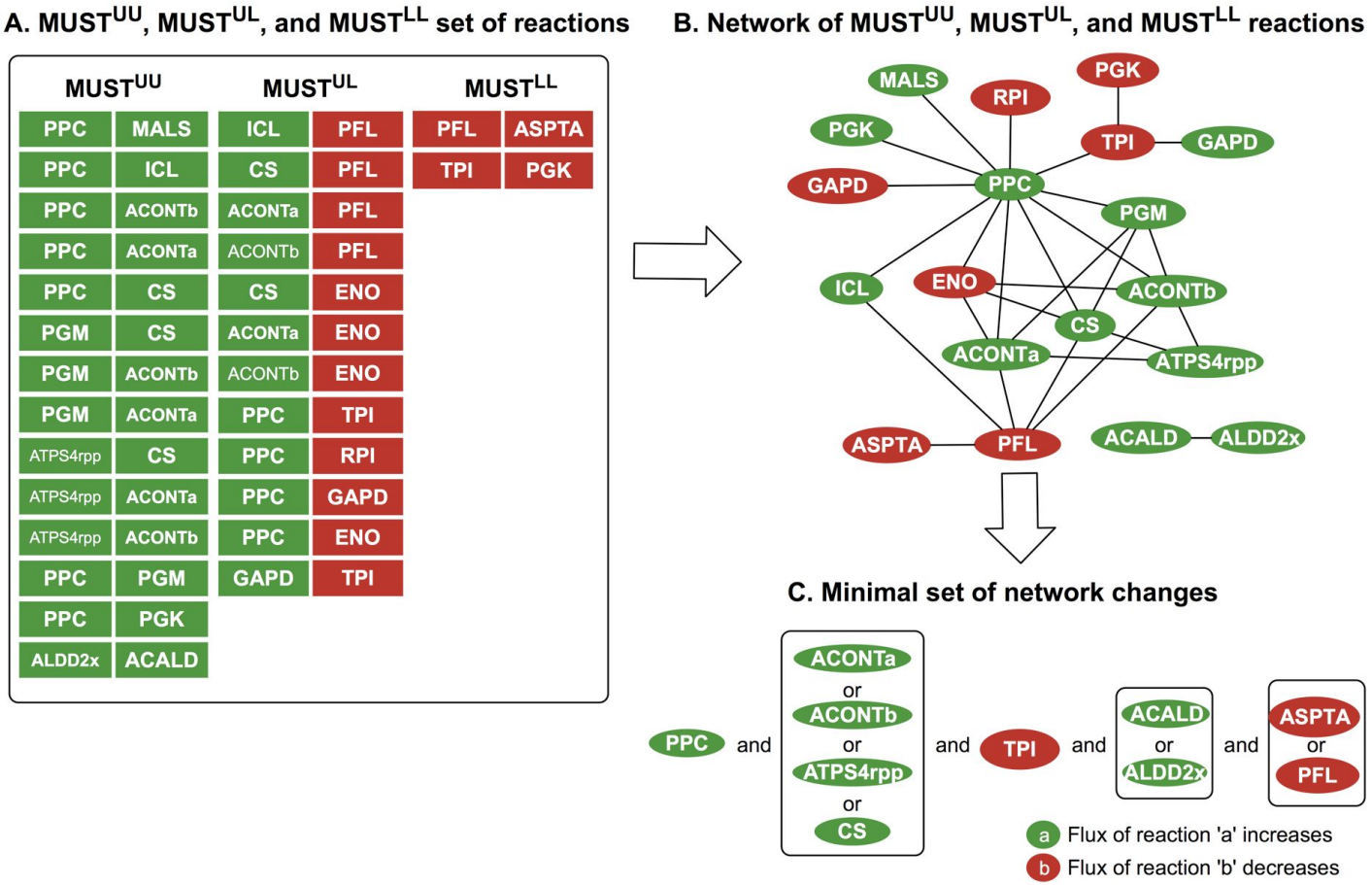
Tools for FBA



Lewis et al (2012) Nature Reviews Microbiology.

Identifying many genetic interventions

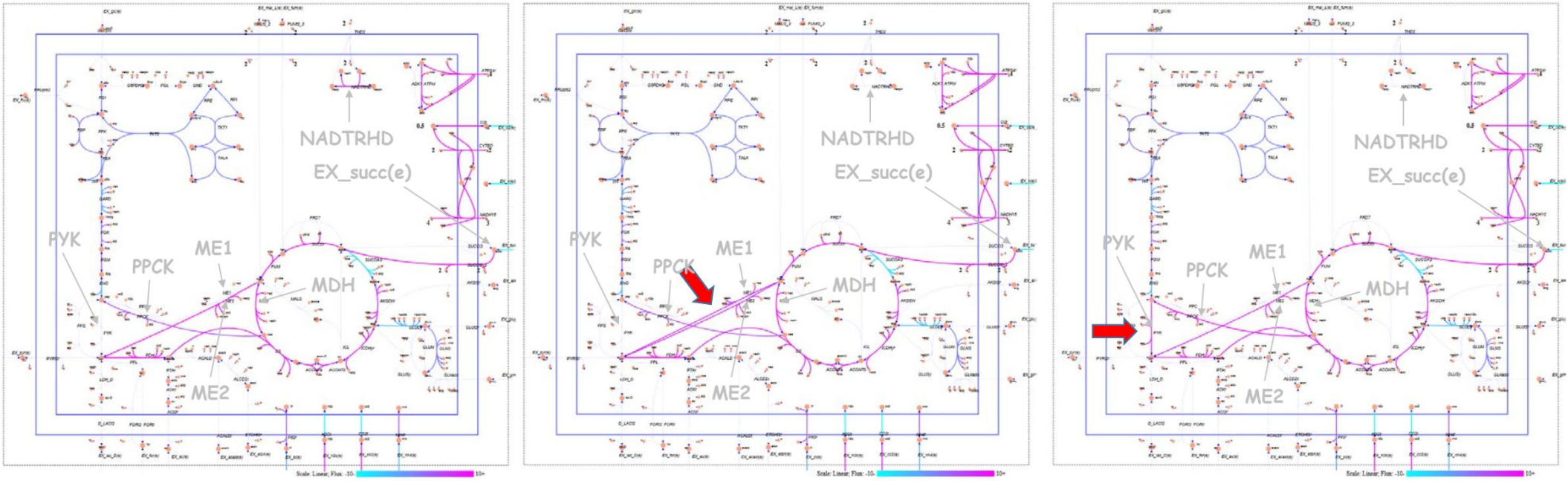
- OptKnock
- OptForce



Burgard et al. (2003) Biotechnol Bioeng.
Ranganathan et al. (2010) PLOS Comp Biol.

Flux variability analysis

- Even with constraints, there is not necessarily one unique solution
- Flux variability analysis (FVA) is a method to identify min/max flux values for every reaction that allow objective function to be satisfied



Hinton (2015) BIE5500/6500 lecture notes