COMP 150 CSB –
Computational Systems Biology

Constraint-Based Modeling

Soha Hassoun
Department of Computer Science (primary)
Department of Chemical and biological Engineering
Department of Electrical and Computer Engineering

Tufts UNIVERSITY

Reading

- S Matrix, and FBA basics (reference)

Keupfer chapter:

Metabolic Flux Analysis - Methods and Protocols, Edited by Jens O. Krömer, Chapter 1, "Stoichiometric Modelling of Microbial Metabolism" by Lars Kuepfer

- Applications of Constrained Based modeling – read the supplementary material:

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis?. Nature biotechnology, 28(3), 245.

- Hands on COBRA Toolbox Tutorial

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., ... & Magnusdottir, S. (2017). Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3. 0. arXiv preprint arXiv:1710.04038.

Slides/Figures from

- Scott Hinton
- Roded Sharan
- Tomer Shlomi
- Elhanan Borenstein
- Eran Eden
- Palsson, B. Ø. (2015). Systems biology: constraint-based reconstruction and analysis. Chapter 9

Outline (key slides have underlined titles)

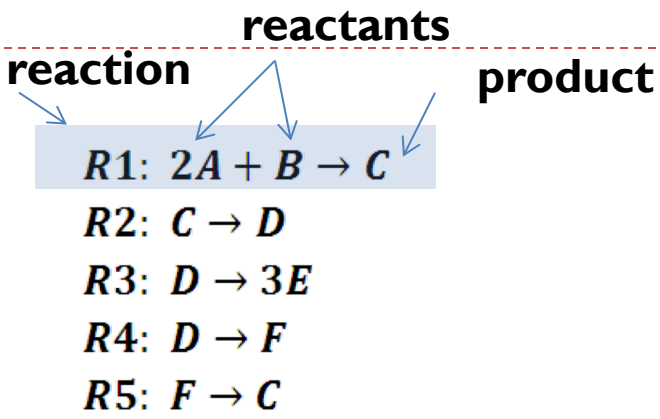
- **Systems Biology Paradigm – revisited**
- **S matrix**
- **The model and its boundaries**
- **Dynamic mass balance**
- **Dynamic mass balance at steady state**
- **Adding constraints**
- **Flux cone**
- **Calculating optimal flux distribution**
- **Choosing the objective function**
- **Solving linear program**
- **Example applications of Constraint-Based Modeling**
- ***E. coli* Model we will use**
- **COBRA Toolbox overview**

Systems Biology Paradigm - revisited

- Network Reconstruction:
  - Define and enumerate list of biological components
    - E.g. from gene sequencing
  - Create an organism/sample model that captures interactions between components
    - Complex – multiple steps, deal with incomplete information
    - Formalized 96 steps to this process
- Constructed networks are converted to mathematical models
- Models are analyzed, queried, and interpreted using constraint-based modeling
- Models are used in many applications
  - Prediction, hypothesis testing, ..
  - Re-engineering cellular behavior

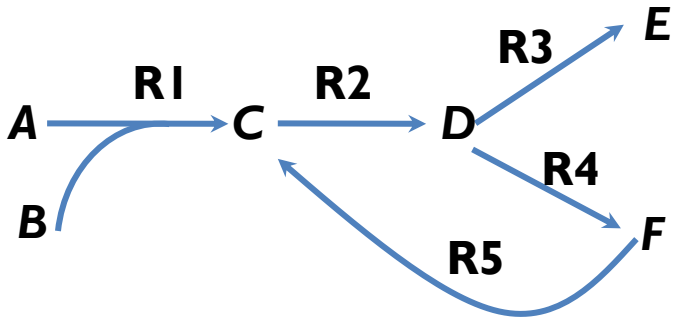
S Matrix

- A metabolic network



- Stoichiometric matrix represents a biochemical network

	R1	R2	R3	R4	R5
A	-2				
B	-1				
C	1	-1			1
D		1	-1	-1	
E			3		
F				1	-1

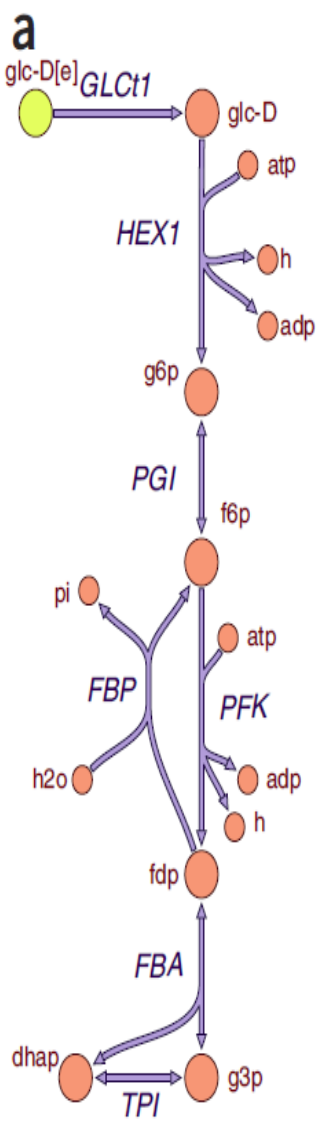


Graph representation

- Reactions can be reversible: thermodynamics dictate direction

Example - Creating A Stoichiometric Matrix

The stoichiometric matrix, S, is the centerpiece of a mathematical representation of genome-scale metabolic networks. This matrix represents each reaction as a column and each metabolite as a row, where each numerical element is the corresponding stoichiometric coefficient.



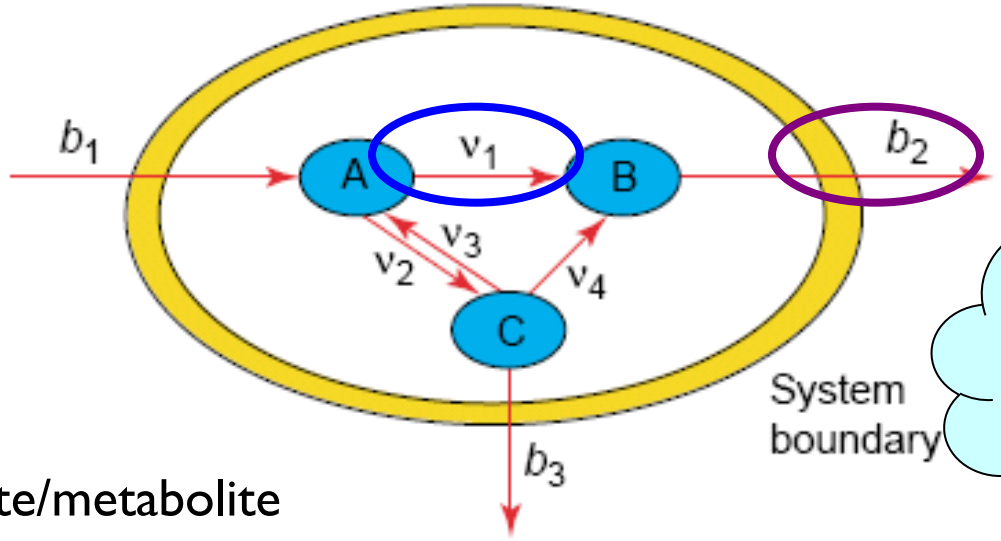
b

	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

Becker, S. A., A. M. Feist, et al. (2007). "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox." Nature protocols **2(3): 727-738.**

The model and its boundaries



How do we define a biologically significant system boundary?

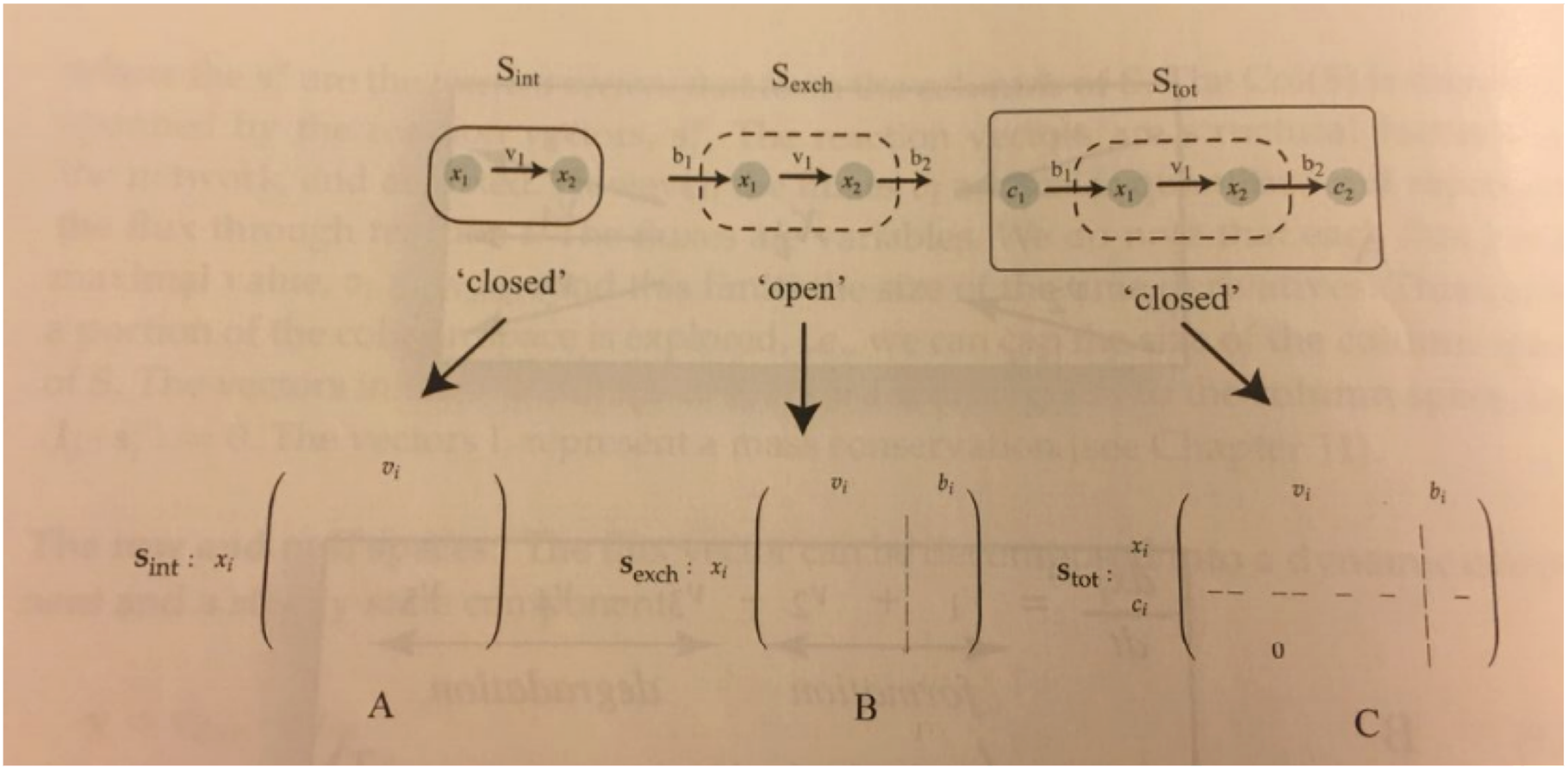
Vertex - substrate/metabolite concentration.

Edge - flux (conversion mediated by enzymes of one substrate into the other)

- **Internal flux edge**
- **External flux edge**

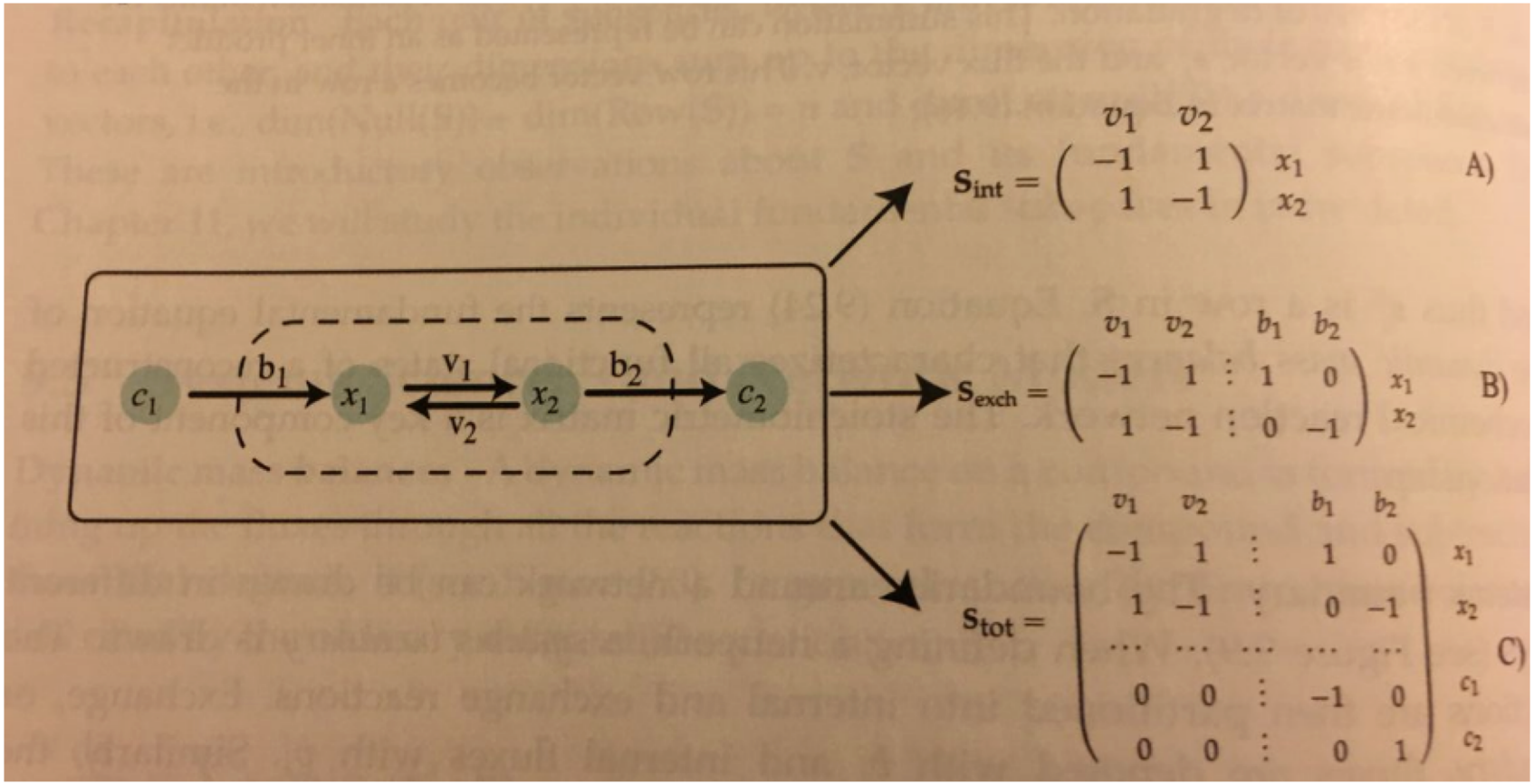
Boundary Scenarios

(fig. 9.9 Palsson Book)



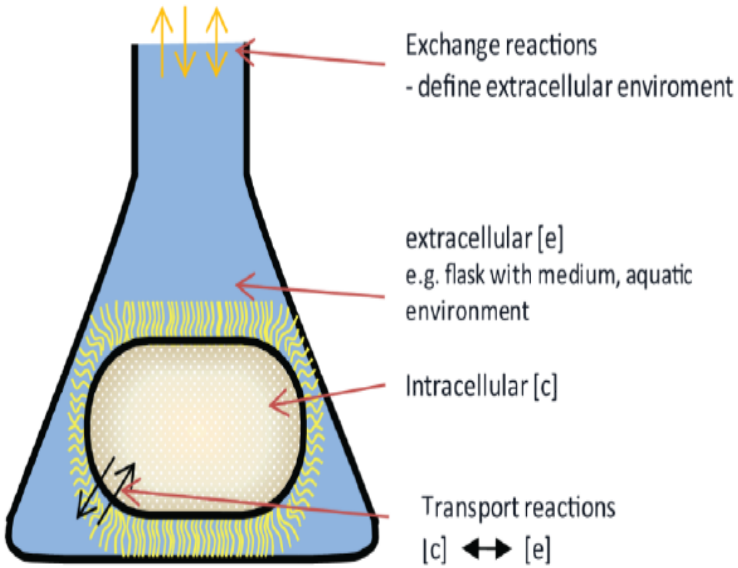
Boundary Conditions – Example

(fig. 9.10 Palsson Book)

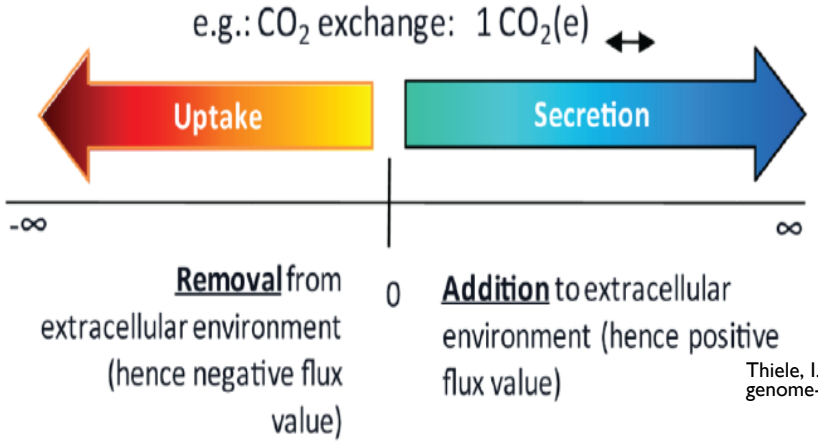


System Boundaries:

Exchange & Transport Reactions



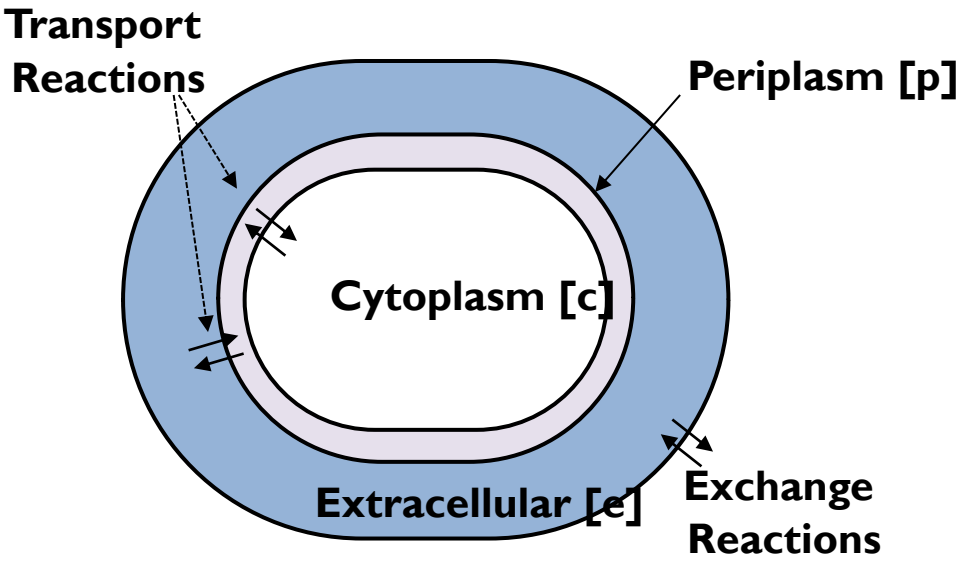
Exchange reactions are defined as follows:



Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.

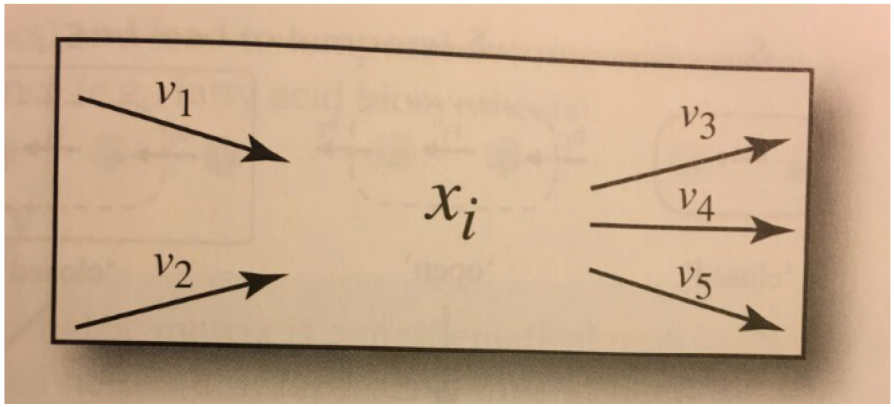
System Boundaries:

Cellular Exchange & Transport Reactions



Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.

Dynamic mass balance



$$\frac{dx_i}{dt} = \underbrace{v_1 + v_2}_{\text{formation}} - \underbrace{v_3 - v_4 - v_5}_{\text{degradation}}$$

$$= \left\langle (1,1,-1,-1,-1)\,,\,(v_1,v_2,v_3,v_4,v_5)^{\mathrm{T}} \right\rangle$$

$$= \left\langle \mathbf{s}_i^x \cdot \mathbf{v} \right\rangle$$

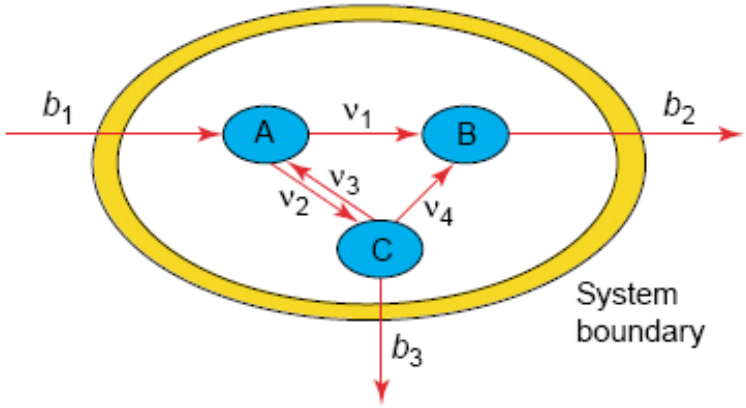
A row in S that involves reactions acting on x_i

Dynamic mass balance

$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$

$$\frac{dB}{dt} = v_1 + v_4 - b_2$$

$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$



$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (7 columns) · V (7 rows)

Concentration vector — Stoichiometry Matrix — Flux vector

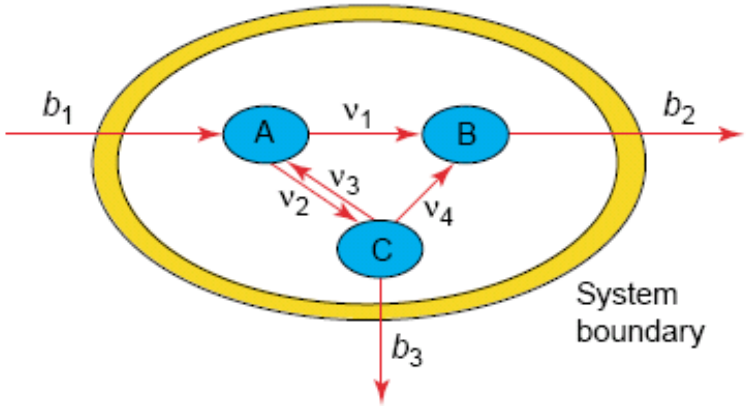
$$\frac{dx}{dt} = S \cdot v$$

Dynamic mass balance

$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$

$$\frac{dB}{dt} = v_1 + v_4 - b_2$$

$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$



$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (7 columns) · V (7 rows)

Concentration vector — Stoichiometry Matrix — Flux vector

$$\frac{dx}{dt} = S \cdot v$$

Dynamic mass balance

Problem …

V=V(k1, k2,k3…) is actually a function of concentration as well as several kinetic parameters

Difficult determine kinetic parameters experimentally.

In general, there is not enough kinetic information the literature to construct the model.

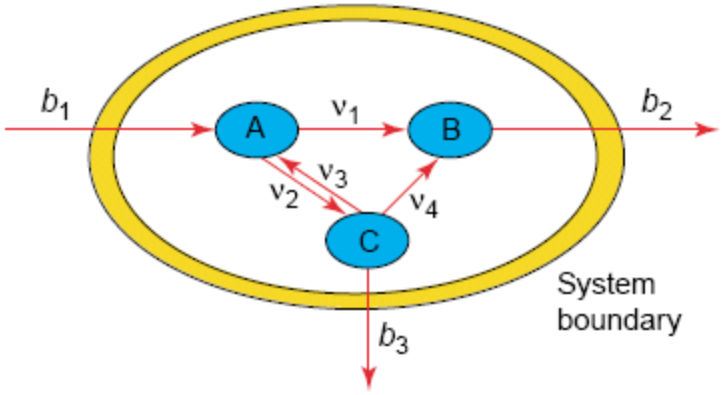
Solution !

In order to identify invariant characteristics of the network, assume the network is in **steady state**.

Dynamic mass balance at steady state

1. What does "steady state" mean?

$$\frac{dx}{dt} = 0$$



2. Is it biologically justifiable to assume it?

"The steady state approximation is generally valid because of fast equilibration of metabolite concentrations (**seconds**) with respect to the time scale of genetic regulation (**minutes**)" – Segre 2002

Why does the steady state assumption help us solve our problem?

$$\frac{dx}{dt} = S \cdot v \longrightarrow 0 = S \cdot v$$

Steady state assumption

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (7 columns) × V (7 rows)

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (7 columns) × V (7 rows)

Adding constraints

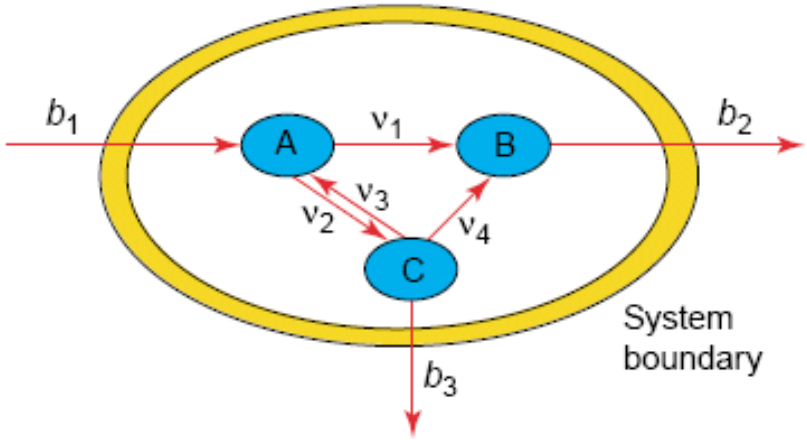
Constraints on internal fluxes: $v_i \geq 0, \forall i$

Constraints on external fluxes:

Source → $b_j \leq 0$

Sink → $b_j \geq 0$

Sink/source → b_j is unconstrained

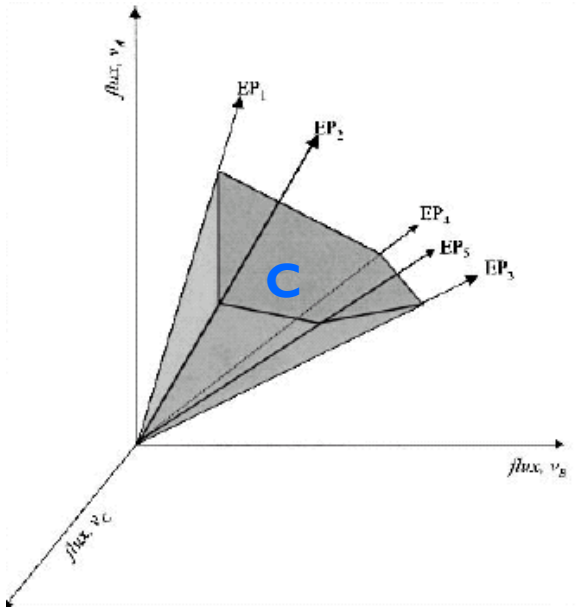


Flux cone

$$0 = S \cdot v$$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (columns) × v (rows)



Observation: the number of reactions considerably exceeds the number of metabolites

↓

The S matrix will have more columns than rows

↓

The null space of viable solutions to our linear set of equations contains an infinite number of solutions.

What about the constraints?

↓

"The solution space for any system of linear homogeneous equations and inequalities is a convex polyhedral cone." - Schilling 2000

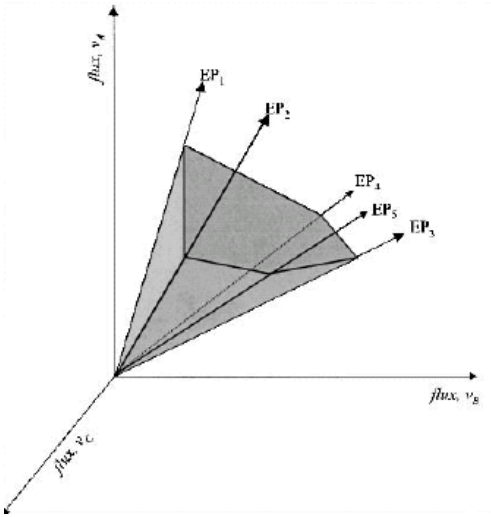
The edges of the cone are called the Extreme Pathways. "extreme rays correspond to edges of the cone. They are said to generate the cone and cannot be decomposed into non-trivial combinations of any other vector in the cone."- schilling 2000

Narrowing the steady state flux cone

- The steady state flux cone contains an infinite flux distributions!
- Only a small portion of them is physiologically feasible.
- More constraints on the external fluxes.

These depend on factors as:

- Organism
- Environment and accessibility substrates
- maximum rates of diffusion mediated transport
- Etc…



Calculating optimal flux distribution

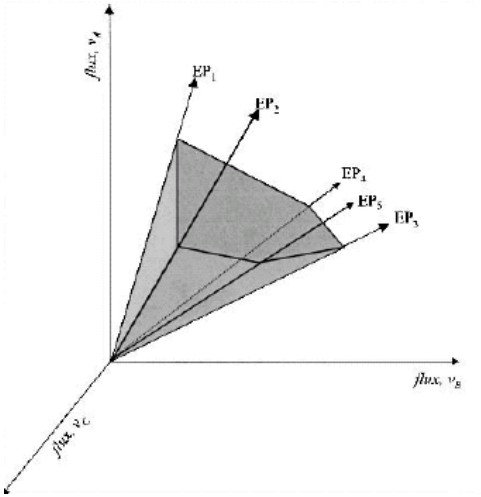
- How can we identify a "biologically meaningful" flux?

Assumption…

the metabolic network is optimized with respect to a certain objective function Z.

Find the vector v in the flux cone which maximizes Z

…this can be can formulated as an optimization problem:



Minimize/Maximize $Z = \sum_j c_j v_j$ s.t. $S \cdot v = 0$ + inequality constraints

This optimization problem is a classical linear programming **(LP)** problem that can be solved using the simplex algorithm.

Choosing the objective function Z

We want to choose a Z that is biologically meaningful.

Reasonable options could be:

1. Z: Cellular growth (maximization)
2. Z: Particular metabolite engineering (maximization)
3. Z: Energy consumption (minimization)

Example:

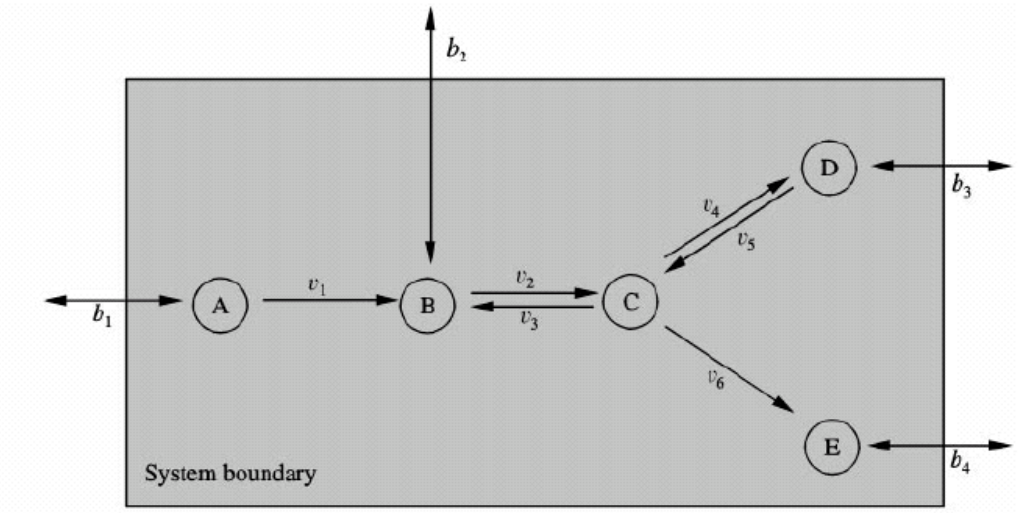
cellular growth is correlated with the production of E and D

We want a v that:

(A) Resides inside the cone.

(B) **maximizes sum of fluxes that produce E and D**

$Z = b_3 + b_4$



Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.

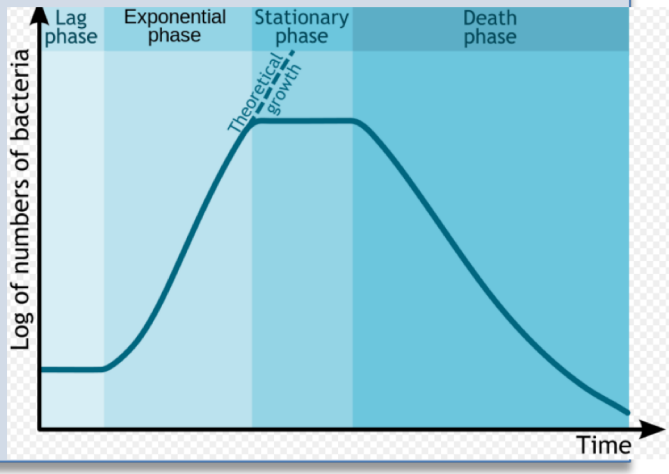
Biomass Precursors

- The biomass reaction accounts for all the fractional contributions from biosynthetic precursors and key cofactors to create 1g of biomass.
- These factional contributions need to be determined experimentally for cells growing in log phase.
- It may not be possible to obtain a detailed biomass composition for the target organism. In this case, one can estimate the relative fraction of each precursor from existing databases.

Cellular component	Cellular content %(wt/wt)
Protein	55
RNA	20.5
DNA	3.1
Lipids	9.1
Lipopolysaccharides	3.4
Peptidoglycan	2.5
Glycogen	2.5
Polyamines	0.4
Other	3.5
Total	100.00

What is "log phase"?

The log phase (sometimes called the logarithmic phase or the exponential phase) is a period characterized by cell doubling.



Maintenance Energy Requirements

- To simulate growth, the energy required to maintain the cell growth must be accounted for.
- Two forms of energy are required; growth associated maintenance (GAM) energy and non-growth associated maintenance (NGAM) energy.
- GAM reaction accounts for the energy (ATP) necessary to replicate a cell. It is represented in the model by

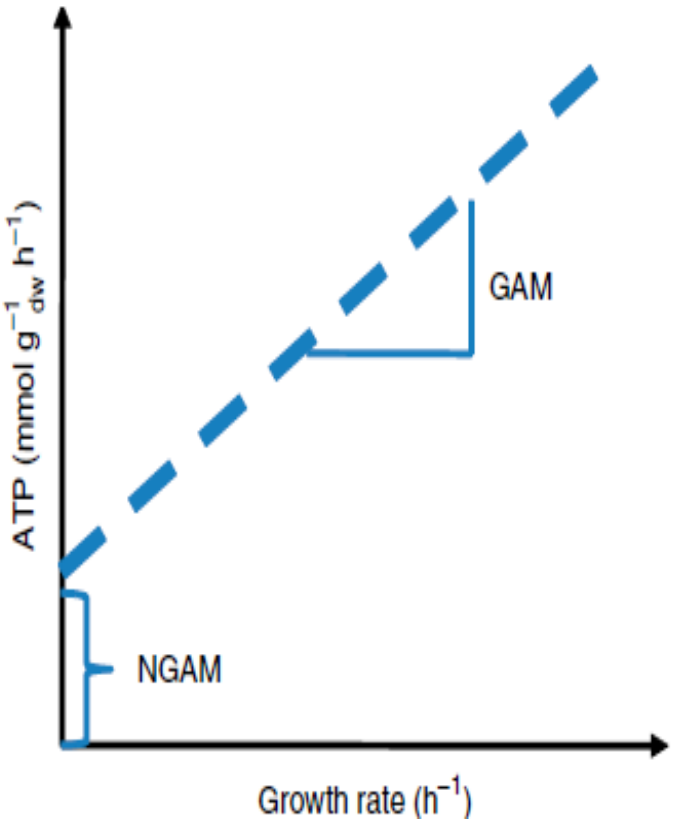
$$x\,ATP + x\,H_20 \rightarrow x\,ADP + x\,P_i + x\,H^+$$

Where x is the number of required phosphate bonds (59.81 in core model). This will be included in the biomass reaction

- The NGAM reaction (ATPM) is given by

$$1\,ATP + 1H_2O \rightarrow 1\,ADP + 1\,P_i + 1\,H^+$$

where the flux through this reaction is constrained by experimental data to 8.39 mmol $g_{DW}^{-1}h^{-1}$

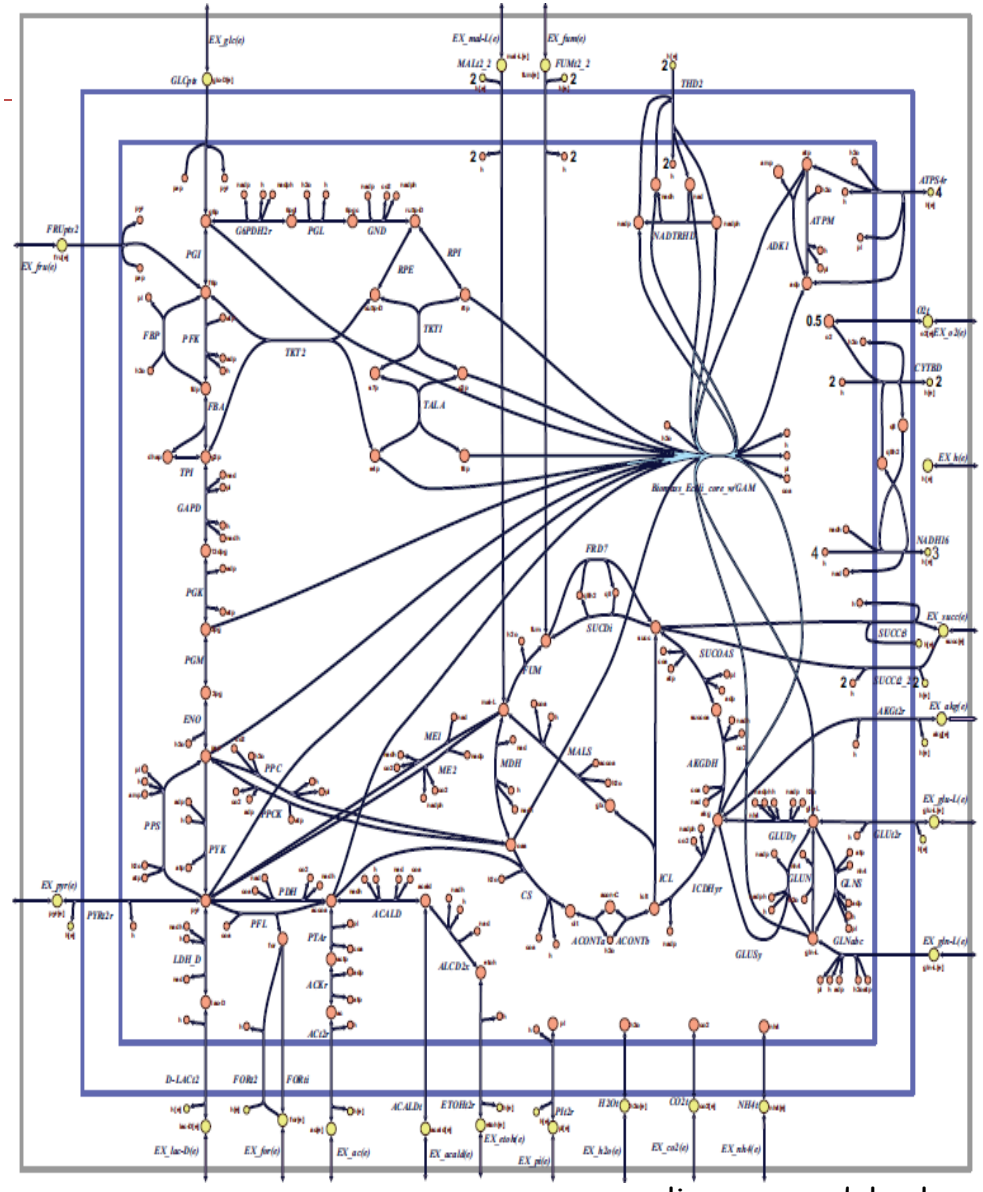


Thiele, I. and B. O. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." Nature protocols 5(1): 93-121.

Biomass Reaction For *E.coli* Core Model

(1.496) 3pg + (3.7478) accoa + (59.8100) atp + (0.3610) e4p + (0.0709) f6p + (0.1290) g3p + (0.2050) g6p + (0.2557) gln-L + (4.9414) glu-L + (59.8100) h2o + (3.5470) nad + (13.0279) nadph + (1.7867) oaa + (0.5191) pep + (2.8328) pyr + (0.8977) r5p --> (59.8100) adp + (4.1182) akg + (3.7478) coa + (59.8100) h + (3.5470) nadh + (13.0279) nadp + (59.8100) pi

* Key Cofactors



ecoli_core_models.xls

iaf1260 BIOMASS OBJECTIVE FUNCTION

(Ec_biomass_iAF1260_core_59p81M)

Z = 0.000223 10fthf[c] + 0.000223 2ohph[c] + 0.5137 ala-L[c] + 0.000223 amet[c] + 0.2958 arg-L[c] + 0.2411 asn-L[c] + 0.2411 asp-L[c] + 59.984 atp[c] + 0.004737 ca2[c] + 0.004737 cl[c] + 0.000576 coa[c] + 0.003158 cobalt2[c] + 0.1335 ctp[c] + 0.003158 cu2[c] + 0.09158 cys-L[c] + 0.02617 datp[c] + 0.02702 dctp[c] + 0.02702 dgtp[c] + 0.02617 dttp[c] + 0.000223 fad[c] + 0.007106 fe2[c] + 0.007106 fe3[c] + 0.2632 gln-L[c] + 0.2632 glu-L[c] + 0.6126 gly[c] + 0.2151 gtp[c] + 54.462 h2o[c] + 0.09474 his-L[c] + 0.2905 ile-L[c] + 0.1776 k[c] + 0.01945 kdo2lipid4[e] + 0.4505 leu-L[c] + 0.3432 lys-L[c] + 0.1537 met-L[c] + 0.007895 mg2[c] + 0.000223 mlthf[c] + 0.003158 mn2[c] + 0.003158 mobd[c] + 0.01389 murein5px4p[p] + 0.001831 nad[c] + 0.000447 nadp[c] + 0.011843 nh4[c] + 0.02233 pe160[c] + 0.04148 pe160[p] + 0.02632 pe161[c] + 0.04889 pe161[p] + 0.1759 phe-L[c] + 0.000223 pheme[c] + 0.2211 pro-L[c] + 0.000223 pydx5p[c] + 0.000223 ribflv[c] + 0.2158 ser-L[c] + 0.000223 sheme[c] + 0.003948 so4[c] + 0.000223 thf[c] + 0.000223 thmpp[c] + 0.2537 thr-L[c] + 0.05684 trp-L[c] + 0.1379 tyr-L[c] + 5.5e-005 udcpdp[c] + 0.1441 utp[c] + 0.4232 val-L[c] + 0.003158 zn2[c] -> 59.81 adp[c] + 59.81 h[c] + 59.806 pi[c] + 0.7739 ppi[c]

Solving Linear Programs

Linear Programming Basics (1)

Consider a system that has two metabolites A and B.

The production constraints on them are

$0 < A < 60$, and $0 < B < 50$

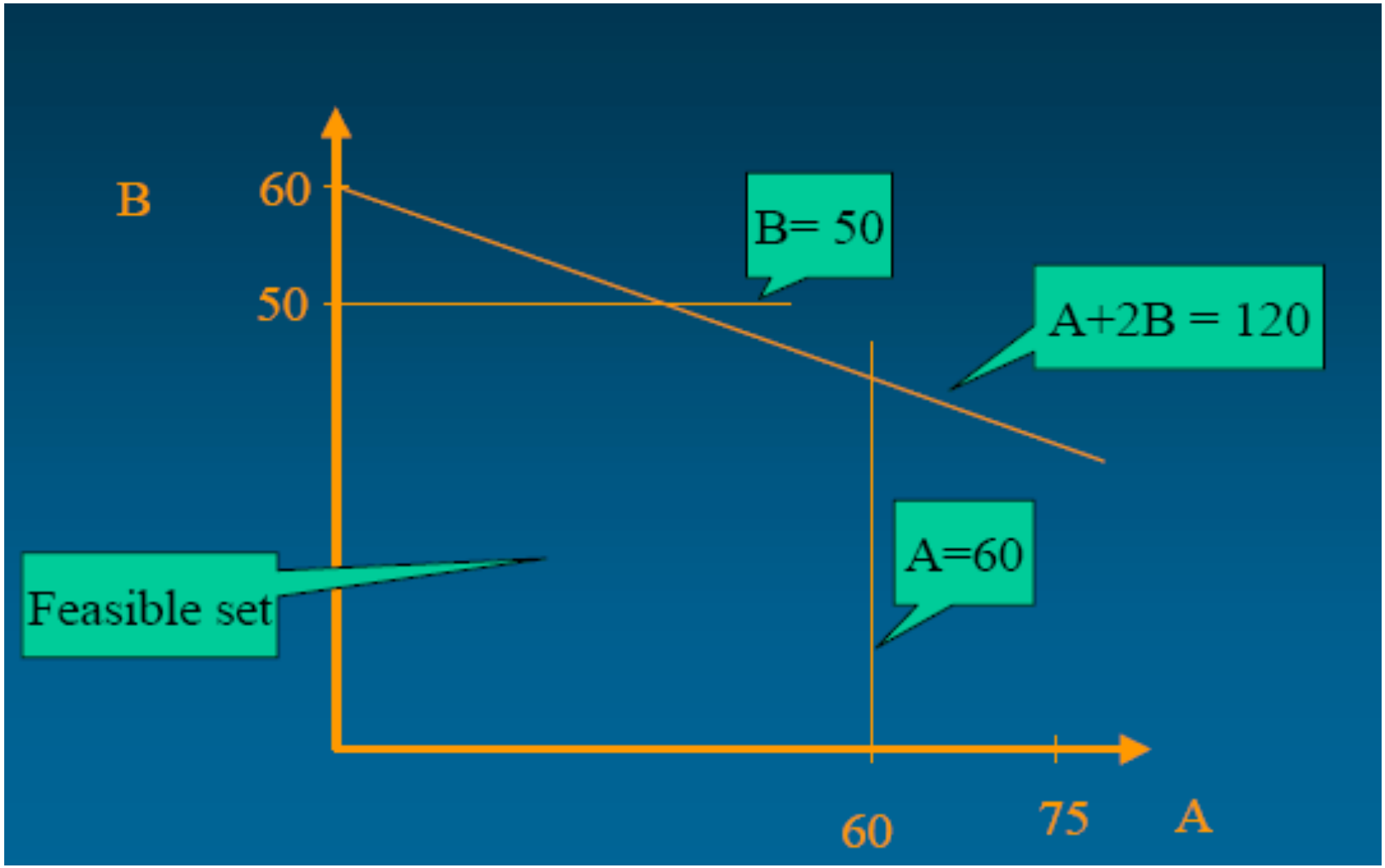
Additionally the capacity for producing them simultaneously is limited by:

$A + 2B < 120$

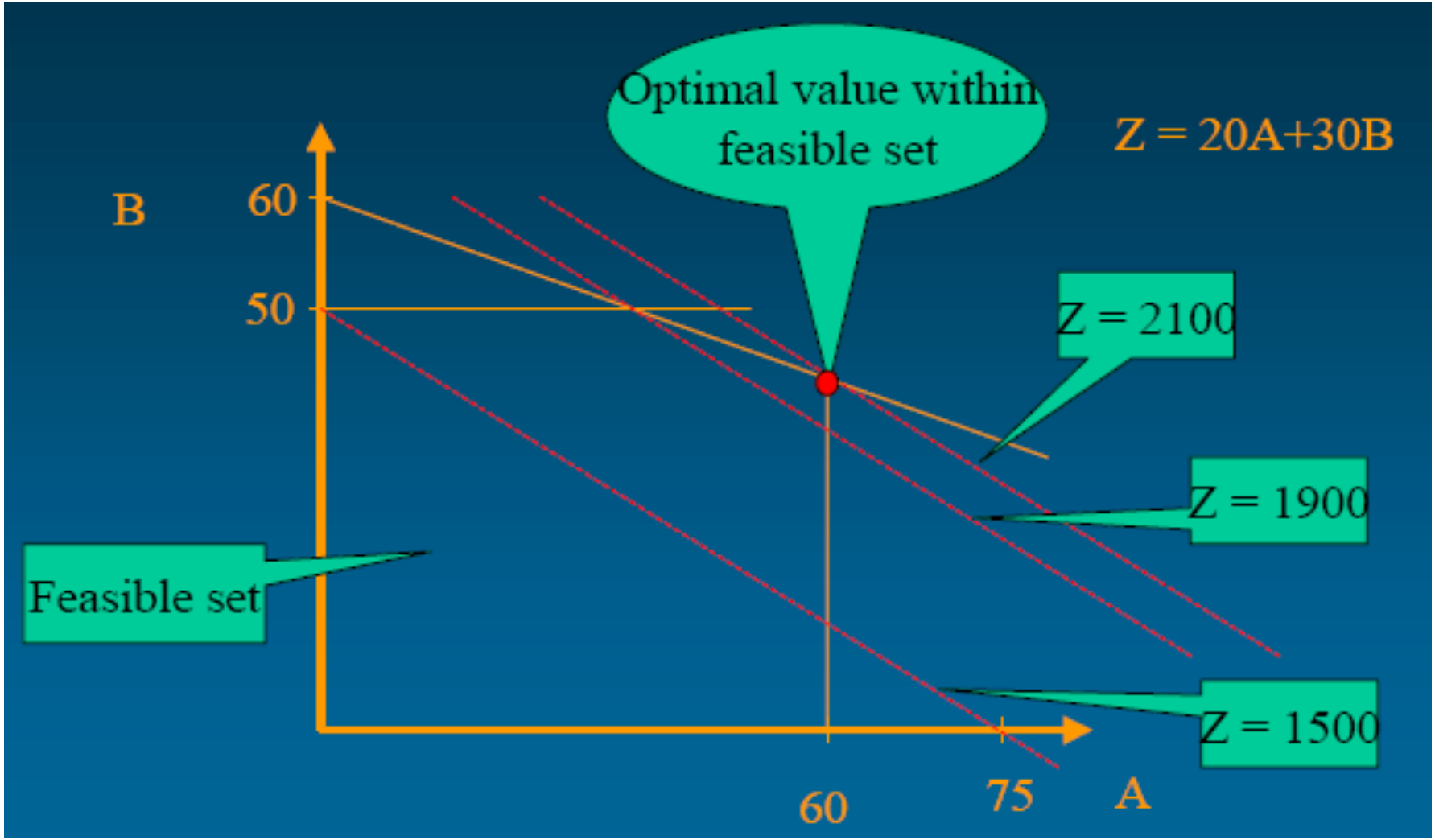
The objective function is

$Z = 20A + 30B$

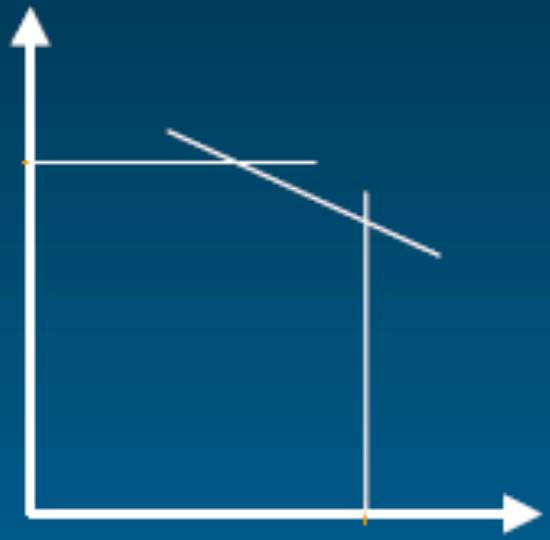
Linear Programming Basics (2)



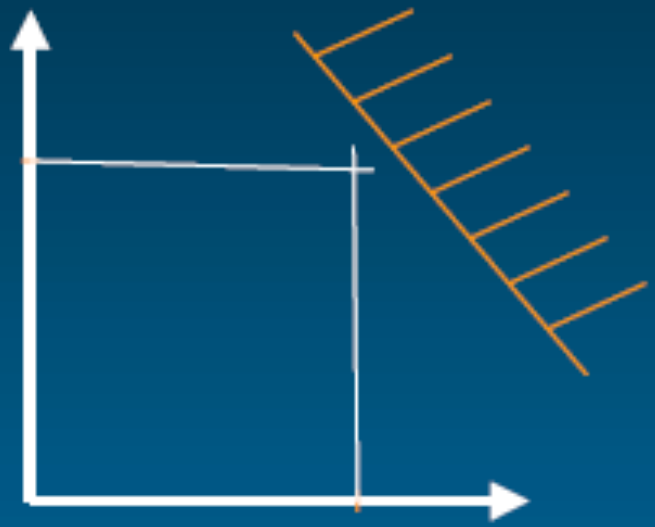
Linear Programming Basics (3)



Linear Programming: Types of Solutions (1)

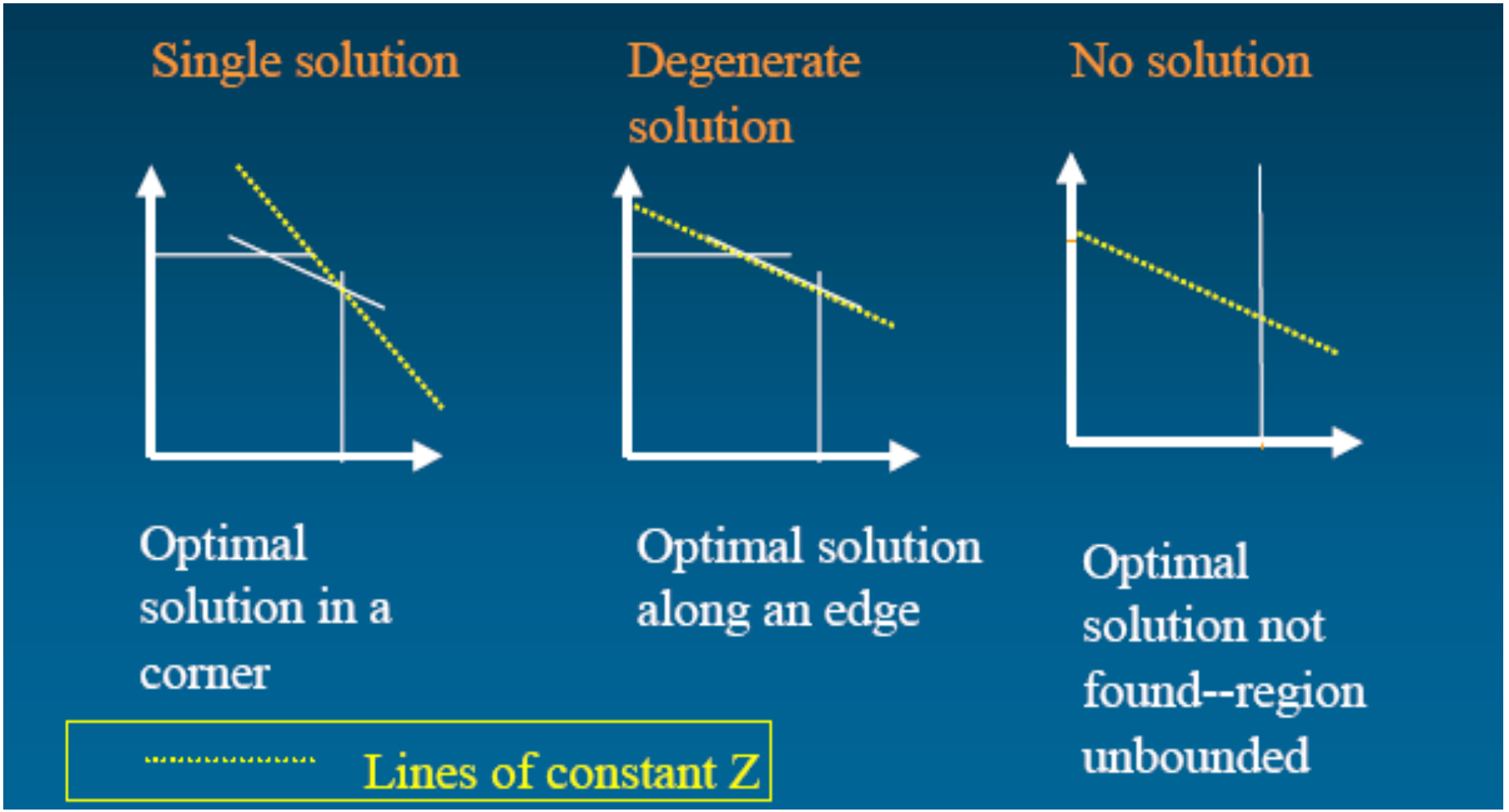


Feasible: solutions possible within all stated constrains



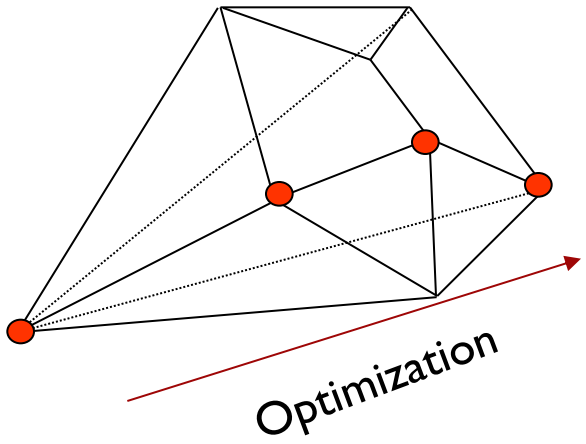
Not feasible: solutions not possible within all stated constrains

Linear Programming: Types of Solutions (2)



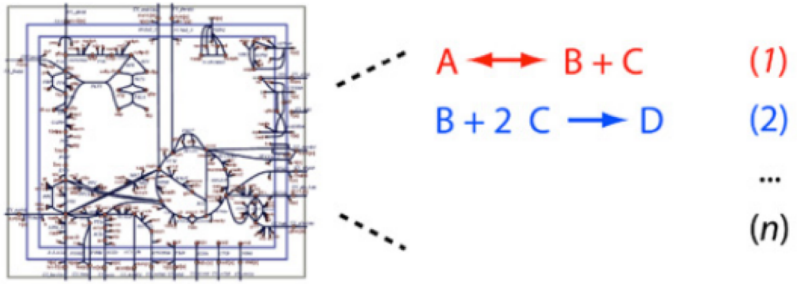
Linear Programming Algorithms

- Simplex algorithm
  - Travels through polytope vertices in the optimization direction
  - Guaranteed to find an optimal solution
  - Exponential running time in worse case
  - Used in practice (takes less than a second)
- When using the COBRA toolbox, use commercial grade LP solvers

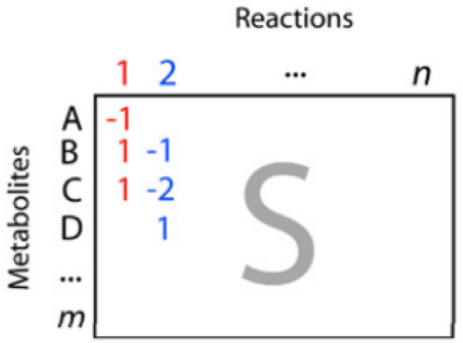


Summary of Flux Balance Analysis

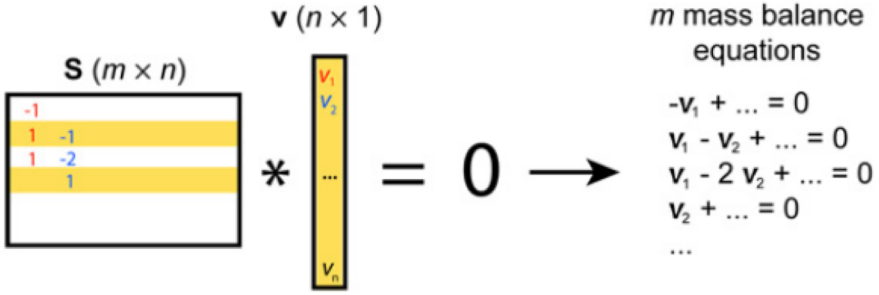
a Curate metabolic reactions



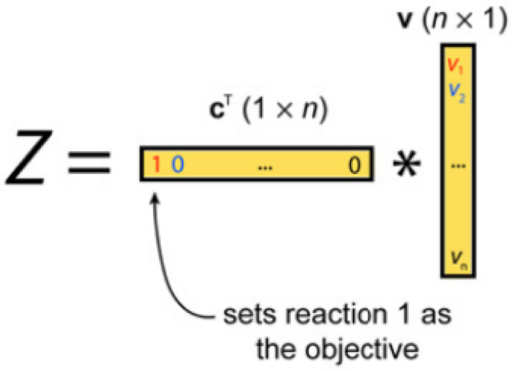
b Formulate **S** matrix



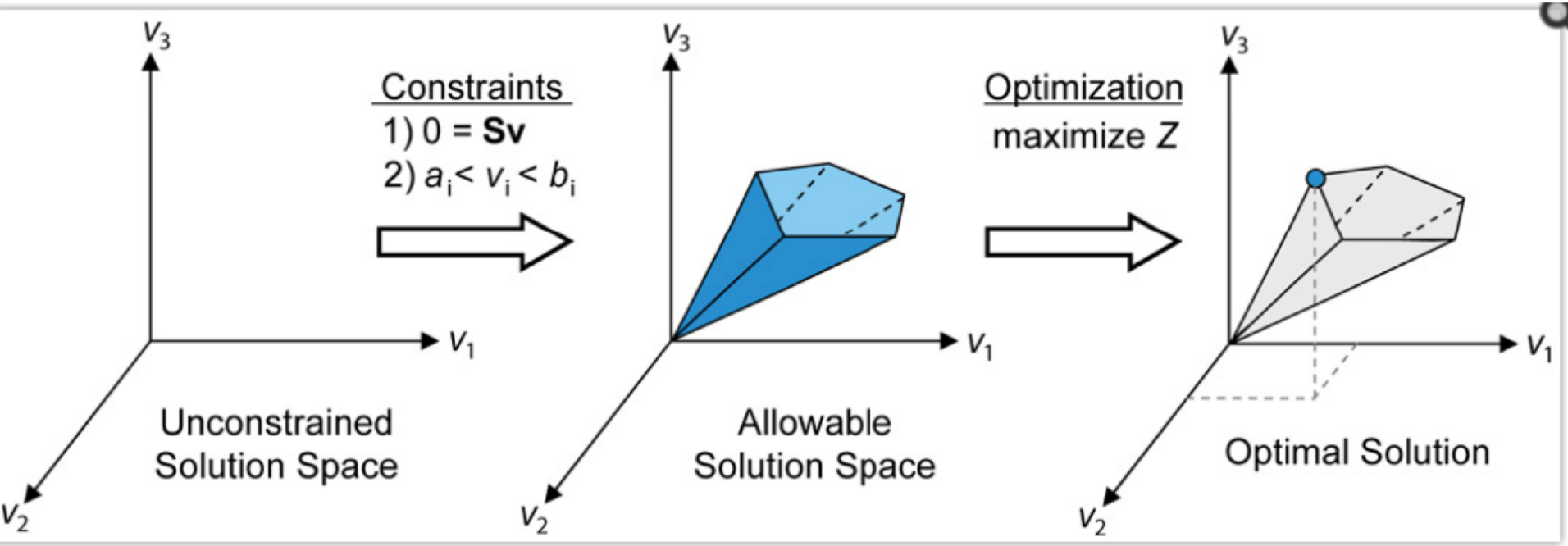
c Apply mass balance constraints



d Define objective function *Z*



Summary of constrained solution space



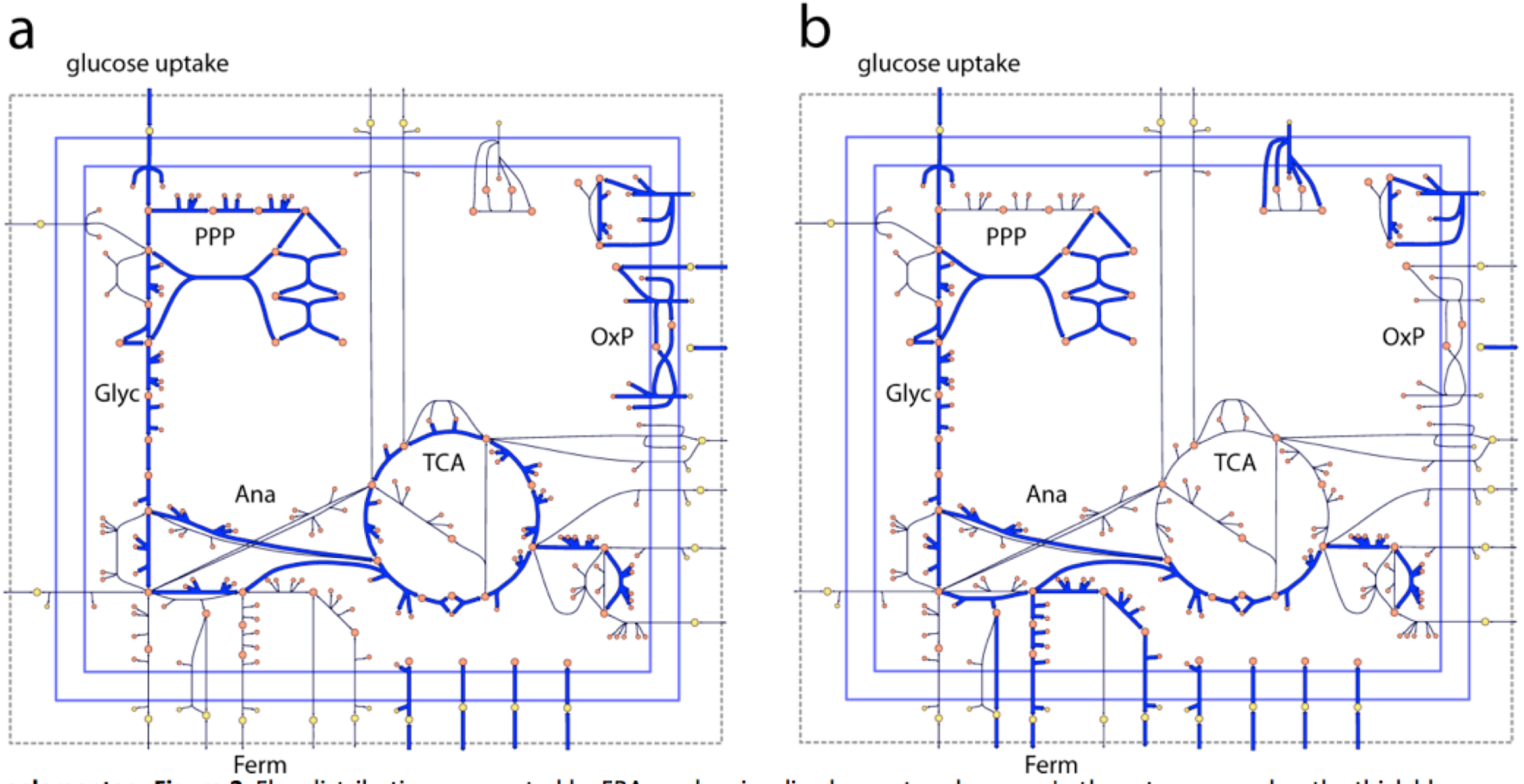
Example Applications

- Select a model for organism
  - Choice of model depends on what needs to be modeled
  - Example: Use *E. coli* core model vs larger more complete models
- Define the application and translate to modification in one or more of the following:
  - Boundary conditions (new uptake reactions)
  - Network changes (gene deletions, gene down regulation, gene additions, ..)

Example Application 1. Compare maximum growth rate with and without oxygen (aerobic vs anaerobic conditions) for *E. coli* with glucose uptake

- Select *E. coli* Model
- Set the maximum glucose uptake rate to 18.5 mmol gDW^{-1} hr^{-1} (millimoles per gram dry cell weight per hour)
  - Actually set the lower bound to be -18.5 because it is an uptake
- Set maximum bound on oxygen to be very large (-1000)
  - Oxygen is unbounded going into the cell
- Set the objective function: maximize production of biomass
- Run FBA to maximize the growth rate. Z = 1.6531/hr
- Repeat above with Oxygen bound set to zero
- Run FBA to maximize the growth rate. Z = 0.4705/hr

Example Application 1. Compare maximum growth rate with and without oxygen (aerobic vs anaerobic conditions) for *E. coli* with glucose uptake



Supplementary Figure 2 Flux distributions computed by FBA can be visualized on network maps. In these two examples, the thick blue arrows represent reactions carrying flux, and the thin black arrows represent unused reactions. These maps show the state of the *E. coli* core model with maximum growth rate as the objective (*Z*) under aerobic (**a**) and anaerobic (**b**) conditions. Reactions that are in use have thick blue arrows, while reactions that carry 0 flux have thin black arrows. The metabolic pathways shown in these maps are glycolysis (Glyc), pentose phosphate pathway (PPP), TCA cycle (TCA), oxidative phosphorylation (OxP), anaplerotic reactions (Ana), and fermentation pathways (Ferm). These flux maps were drawn using SimPheny and edited for clarity with Adobe Illustrator.

Example Application 2. Compare maximum growth rate for *E. coli* with glucose uptake vs. alternate substrates (succinate)

- Select *E. coli* Model
- Set the maximum succinate uptake rate to -20 mmol gDW^{-1} hr^{-1}
  - Set the glucose update to 0!!
- Set maximum bound on oxygen to be very large (-1000)
- Set the objective function: maximize production of biomass
- Run FBA to maximize the growth rate. Z = 0.8401 /hr instead of 1.6531/hr on glucose!

- Can repeat above for anaerobic conditions. Z = 0/hr instead of 0.4705/hr for glucose!
  - Growth is not possible under anaerobic conditions with succinate
  - The maximum amount of ATP that can be produced from this amount of succinate is less than the minimum bound of 8.39 mmol gDW^{-1} hr^{-1} of the ATP maintenance reaction, ATPM, there is no feasible solution

Example Application 2. Compare maximum growth rate for *E. coli* with glucose uptake vs. alternate substrates

- The maximum growth rate of the core *E. coli* model on its 13 different organic substrates, computed by FBA. Growth rate was calculated for both aerobic and anaerobic conditions for each substrate, and the maximum substrate uptake rate was set to 20 mmol gDW^{-1} hr^{-1} for every substrate.

Substrate	Aerobic	Anaerobic
acetate	0.3893	0
acetaldehyde	0.6073	0
2-oxoglutarate	1.0982	0
ethanol	0.6996	0
D-fructose	1.7906	0.5163
fumarate	0.7865	0
D-glucose	1.7906	0.5163
L-glutamine	1.1636	0
L-glutamate	1.2425	0
D-lactate	0.7403	0
L-malate	0.7865	0
pyruvate	0.6221	0.0655
succinate	0.8401	0

Example Application 3. Determine maximum yields of important molecules (cofactors and biosynthetic precursor)

- Calculate maximum yields of the cofactors ATP, NADH, and NADPH from glucose under aerobic conditions
- Set lower bound for glucose to be -1 mmol gDW^{-1} hr^{-1} by setting both the lower and upper bounds to -1
- Set the ATP maintenance reaction as the objective to be maximized.
  - Set the lower bound on this reaction to be zero
  - By default, this reaction has a lower bound of 8.39 mmol gDW^{-1} hr^{-1} to simulate non-growth associated maintenance costs.
- Run FBA
- Repeat above for NADH, NADPH, etc, setting both ATP lower and upper bounds to zero (not requiring cell to produce ATP)
- Results: units of ATP per units of glucose

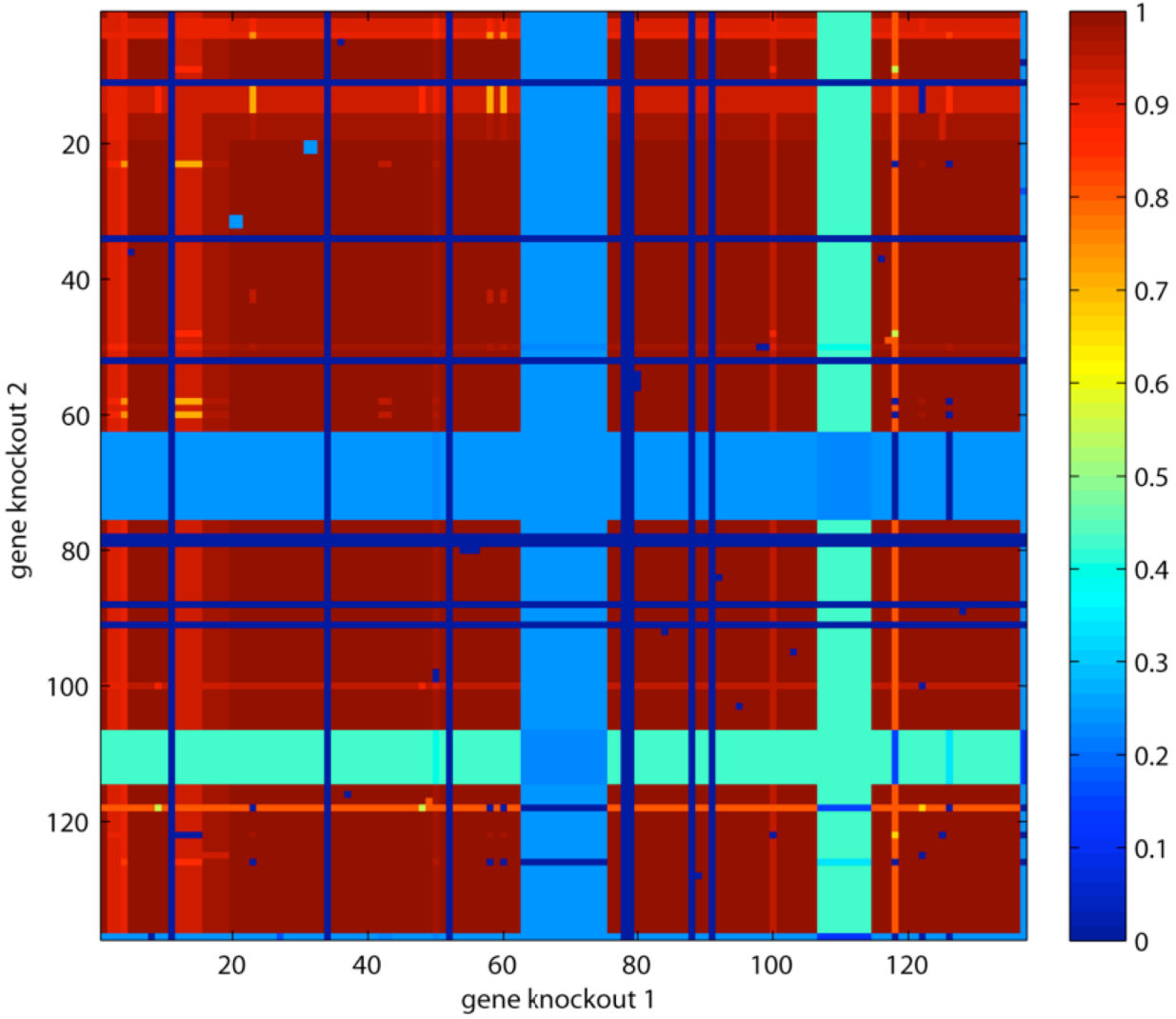
Cofactor	Yield (mol/mol glc)
ATP	17.5
NADH	10
NADPH	8.778

Example Application 4. Simulating Gene Knockouts

- What happens if a gene is silenced? Will the cell die? How is cellular behavior disrupted?
- To simulate the knockout of any gene, its associated reaction or reactions can simply be constrained to not carry flux. Lower and upper bounds are set to zero.
- Most models have a list of gene-protein-reaction interactions (GPRs), a list of Boolean rules that dictate which genes are connected with each reaction in the model.
  - Example: the GPR for phosphofructokinase (PFK) is "b1723 (*pfkB*) or b3916 (*pfkA*)," so according to this Boolean rule, both *pfkB* and *pfkA* must be knocked out to restrict this reaction
- Predict growth for *E. coli* growing aerobically on glucose with the gene b1852 (*zwf*), corresponding to the reaction glucose-6-phospahte dehydrogenase (G6PDH2r), knocked out. The FBA predicted growth rate of this strain is 1.6329 hr^{-1}, which is slightly lower than the growth rate of 1.6531 hr^{-1} for wild-type *E. coli* because the cell can no longer use the oxidative branch of the pentose phosphate pathway to generate NADPH.

Example Application 4. Simulating Gene Knockouts

- Simulate effect of genes knocked out in pairs, and plot growth rates normalized to the maximum growth rate



Supplementary Figure 8 Gene knockout screen on glucose under aerobic conditions. Each of the 136 genes in the core *E. coli* model were knocked out in pairs, and the resulting relative growth rates were plotted. In this figure, genes are ordered by their b number. **Some genes are always essential, and result in a growth rate of 0 when knocked out no matter which other gene is also knocked out**. **Other genes form synthetic lethal pairs**, where knocking out only one of the genes has no effect on growth rate, but knocking both out is lethal. Growth rates in this figure are relative to wild-type.

Example Application 5. Flux Variability Analysis

- For each reaction, maximize and minimize the flux.
  - If min and max flux = 0, then the reaction cannot carry any flux
  - The larger the range, the more "adaptable" the reaction is

E. Coli Model

- Work with iJO1366 model, with 1366 genes, 2251 reactions, and 1136 unique metabolites
- Updated from prior models using experimental data
- Paper has supplementary material with lots of details

A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011

Jeffrey D Orth[1], Tom M Conrad[1], Jessica Na[1], Joshua A Lerman[2], Hojung Nam[1], Adam M Feist[1] and Bernhard Ø Palsson[1,*]

The COnstraint-Based Reconstruction and Analysis (COBRA) toolbox

- A toolbox that uses constraint-based analysis on metabolic network
- Rich set of functions to perform many mundane tasks
  - model = readCbModel(fileName);
  - [minFlux, maxFlux] = fluxVariability(model);
  - model = changeObjective(model, 'BIOMASS_Ec_iJO1366_core_53p95M')
  - model.rxnNames ((model.lb == -10)) // list all reaction names with lower bound of -10
  - reactionFormulaString = 'alac__S_c -> co2_c + (R)-Acetoin_c';
  - model = addReaction(model,'NewRxn','reactionFormula',char(reactionFormulaString), 'reversible',false);

Summary

- Modeling cellular metabolism using constraint-based models
  - **S matrix, model boundaries**
  - **Steady state analysis**
  - **Flux cone**
  - **Setting objective function to compute optimal flux distribution**
- Several example applications
- *E. coli* model overview
- COBRA Toolbox overview