6.047 / 6.878 Computational Biology: Genomes, Networks, Evolution
Fall 2008

Introduction to Steady State Metabolic Modeling

Systems Biology and Metabolic Modeling

- **Steady State Metabolic Modeling**
- **Expression, Regulation, and Steady State Metabolic Modeling**
- **Advanced Systems Modeling**

What is Metabolism?

"The totality of all chemical reactions that occur in living matter" Matthews & van Holde, Biochemistry

Most commonly, these refer to reactions involved in

1) **The generation and storage of energy and oxidation-reduction products**
   - **ATP, NADH, NADPH**

2) **The creation or destruction of cell structural components**
   - **Proteins, Lipids, Carbohydrates, Nucleic Acids**

But we should also properly include:

3) **The transduction and transmission of information**
   - **More commonly studies as *signaling* and *genetics* today**

Why Model Metabolism?

- Predict the effects of drugs on metabolism
  - e.g. what genes should be disrupted to prevent mycolic acid synthesis
- Interpret gene expression data in the context of metabolism
  - e.g. what metabolic state corresponds to a particular expression profile
- Many infectious disease processes involve microbial metabolic changes
  - e.g. switch from sugar to fatty acid metabolism in TB in macrophages

Enzymes

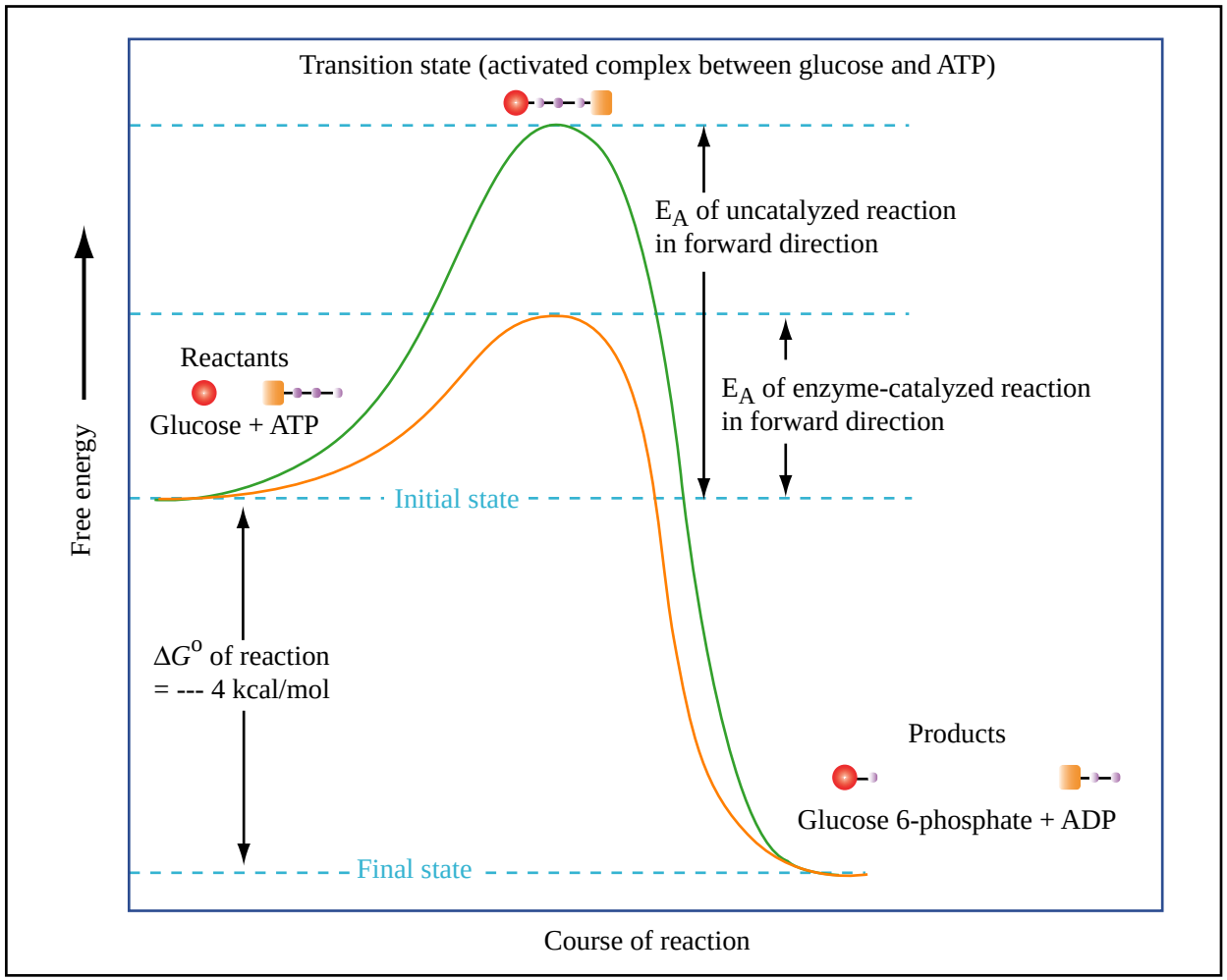


Figure by MIT OpenCourseWare.

Reaction Rates

$$A + 2B \rightarrow 3C$$

Formation rates

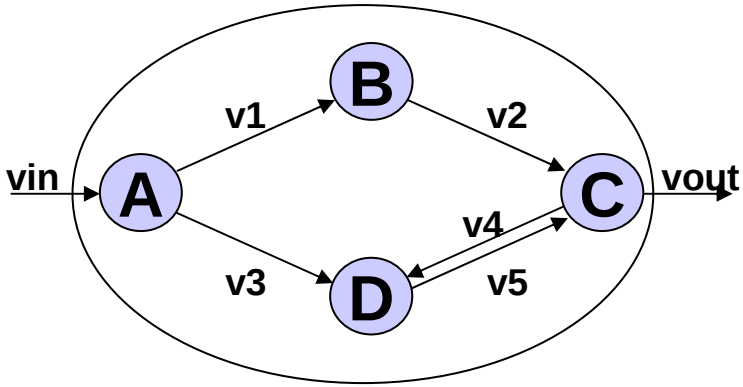
$$v_{fA} = \frac{d[A]}{dt} \qquad v_{fB} = \frac{d[B]}{dt} \qquad v_{fC} = \frac{d[C]}{dt}$$

Reaction Rate = Reaction Velocity = Reaction Flux

$$v = \frac{d[A]}{dt} = \frac{1}{2}\frac{d[B]}{dt} = \frac{1}{3}\frac{d[C]}{dt}$$

Steady State Assumptions

- Dynamics are transient
- At appropriate time-scales and conditions, *metabolism is in steady state*



Two key implications

1. Fluxes are roughly constant
2. Internal metabolite concentrations are constant

$$\frac{d[A]}{dt} = vin - v1 - v3 = 0$$

Metabolic Flux

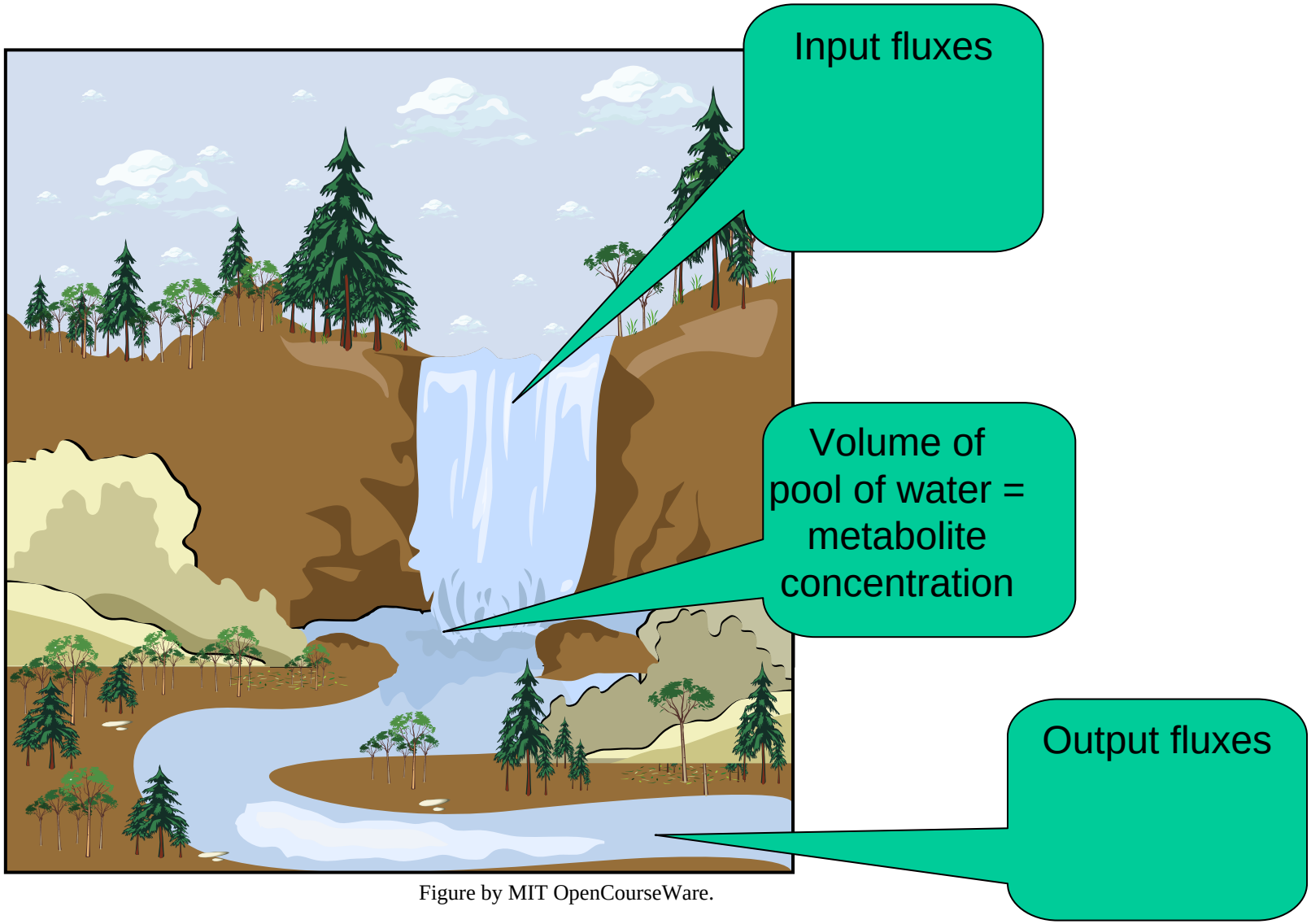


Figure by MIT OpenCourseWare.

Reaction Stoichiometries Are Universal

The conversion of glucose to glucose 6-phosphate always follows this stoichiometry :

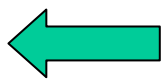
1ATP + 1glucose = 1ADP + 1glucose 6-phosphate

This is <u>chemistry</u> not biology.

Biology => the enzymes catalyzing the reaction

Enzymes influence rates and kinetics

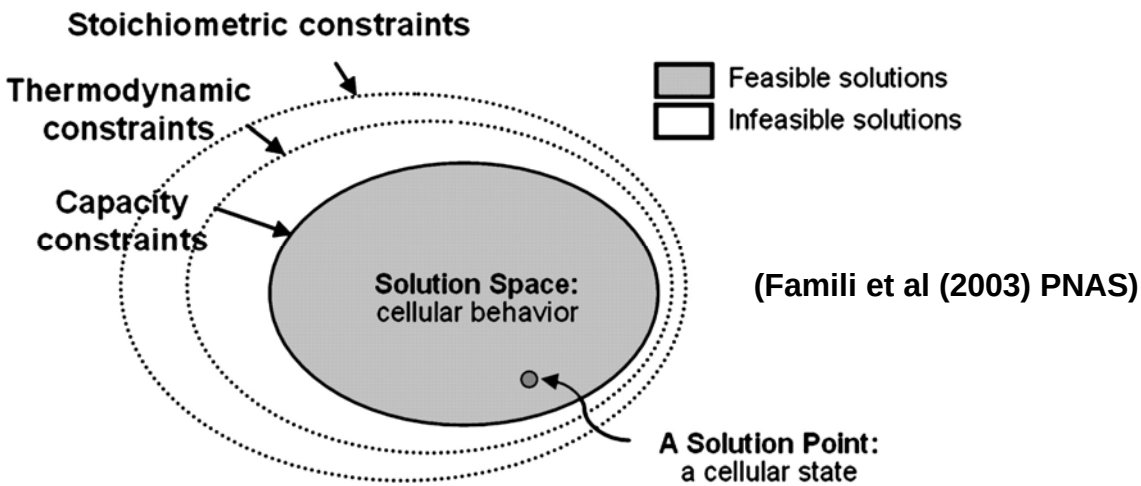
- **Activation energy**
- **Substrate affinity**
- **Rate constants**



Not required for steady state modeling!

Metabolic Flux Analysis

Use universal reaction stoichiometries to predict metabolic network capabilities at steady state*



(Famili et al (2003) PNAS)

Famili, Iman, et al. "Saccharomyces Cerevisiae Phenotypes can be Predicted by Using Constraint-based Analysis of a Genome-scale Reconstructed Metabolic Network.." *PNAS* 100, no. 23 (2003): 13134-13139. Copyright (2003) National Academy of Sciences, U.S.A.

***Not precise, but more precision will come in later slides**

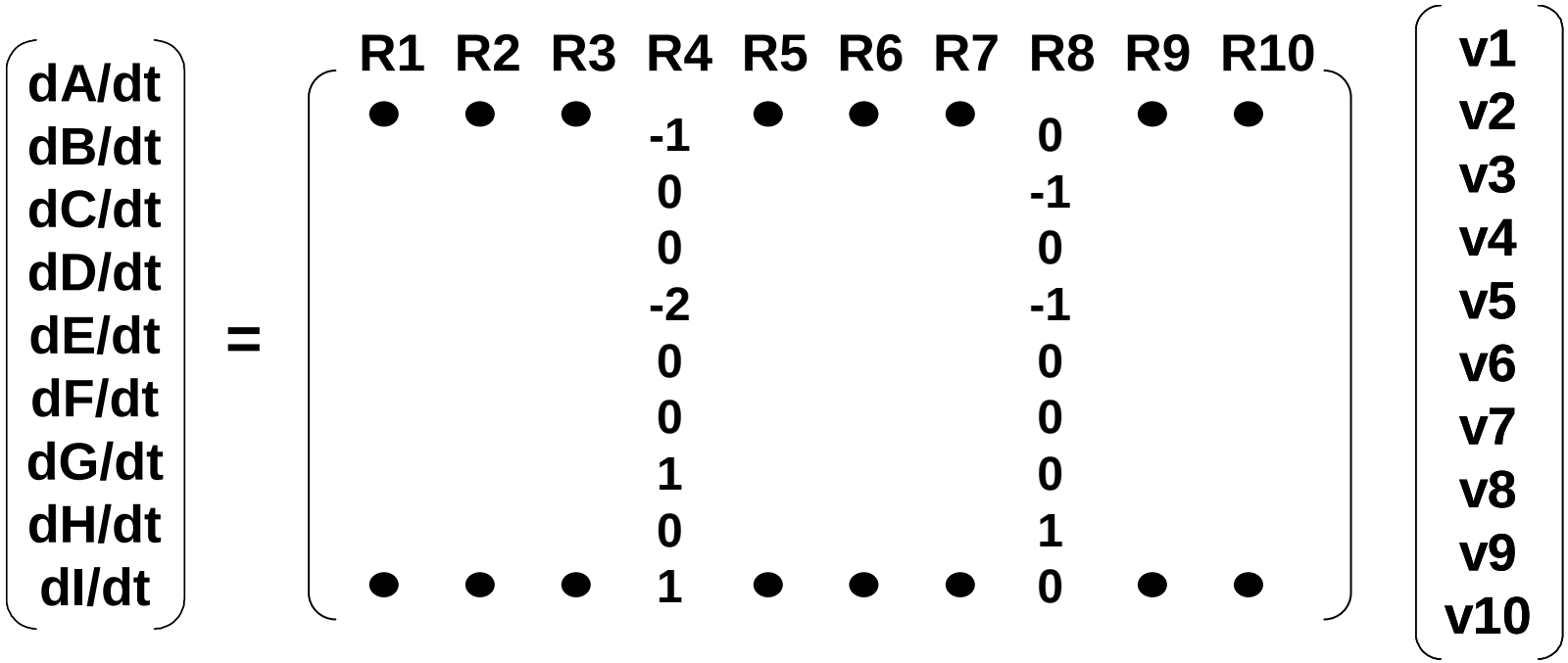
Stoichiometry As Vectors

- We can denote the stoichiometry of a reaction by a vector of coefficients
- One coefficient per metabolite
  - Positive if metabolite is produced
  - Negative if metabolite is consumed

Example:

Metabolites:	**Reactions:**	**Stoichiometry Vectors:**
[A B C D]^T	**2A + B -> C**	**[-2 -1 1 0]^T**
	C -> D	**[0 0 -1 1]^T**

The Stoichiometric Matrix

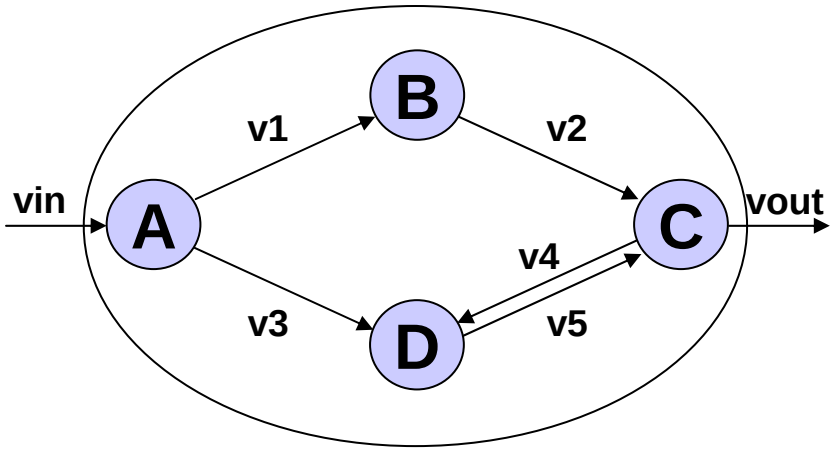


Let V be a vector of fluxes through each reaction

Then S*V is a vector describing the change in concentration of each metabolite per unit time

$$\frac{dx}{dt} = S \bullet V$$

A (Very) Simple System



	v1	v2	v3	v4	v5	vin	vout
A	-1	0	-1	0	0	1	0
B	1	-1	0	0	0	0	0
C	0	1	0	-1	1	0	-1
D	0	0	1	1	-1	0	0

Exchange Reactions

We have introduced two new things

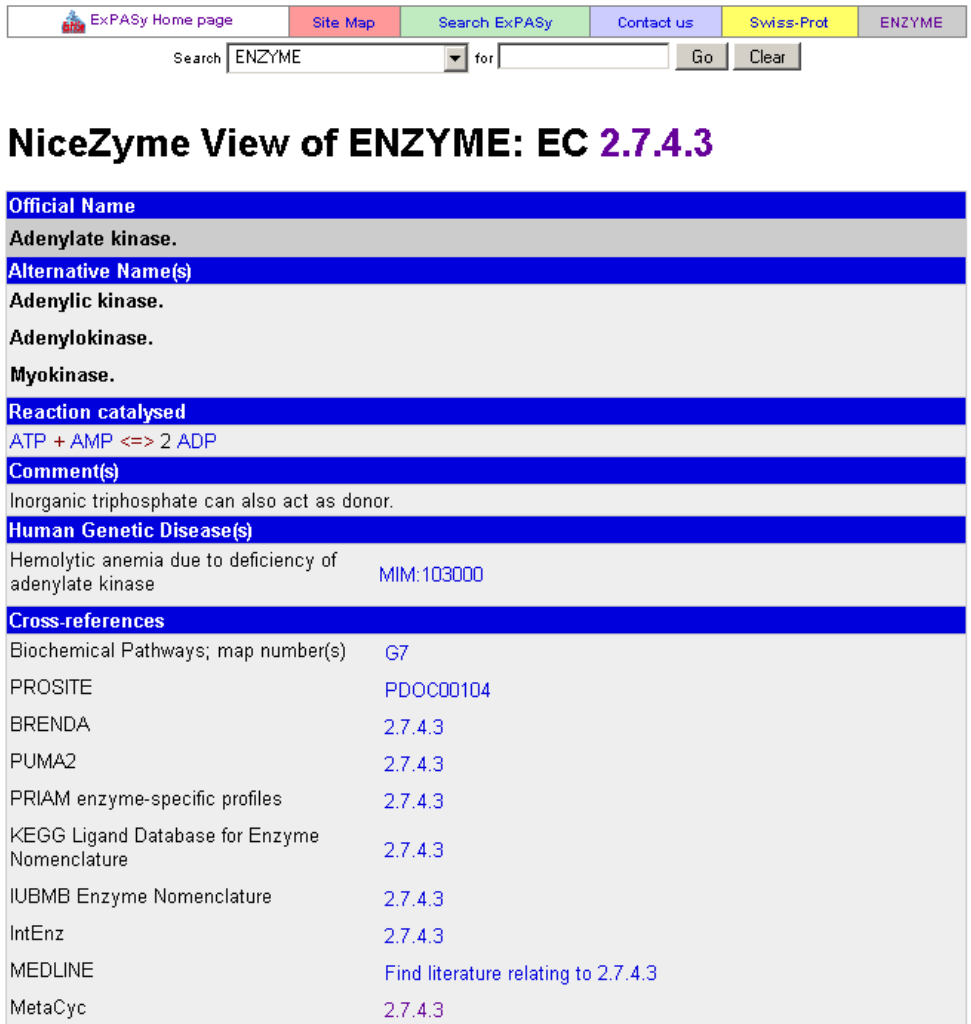
- Reversible reactions – are represented by two reactions that proceed in each direction (e.g. v4, v5)
- Exchange reactions – allow for fluxes from/into an infinite pool outside the system (e.g. vin and vout). *These are frequently the only fluxes experimentally measured.*

Some advantages of S

- Chemistry not Biology: the stoichiometry of a given reaction is preserved across organisms, while the reaction rates may not be preserved
- Does NOT depend on kinetics or reaction rates
- Depends on limited thermodynamic data – only reversibility/irreversibility

Genes to Reactions

- Expasy enzyme database
- Indexed by *EC number*
- EC numbers can be assigned to genes by
  - Blast to known genes
  - PFAM domains



Online Metabolic Databases

There are several online databases with curated and/or automated EC number assignments for sequenced genomes

Kegg

Pathlogic/BioCyc

Images removed due to copyright restrictions. Please see:

http://biocyc.org/intro.shtml

http://www.genome.jp/kegg/

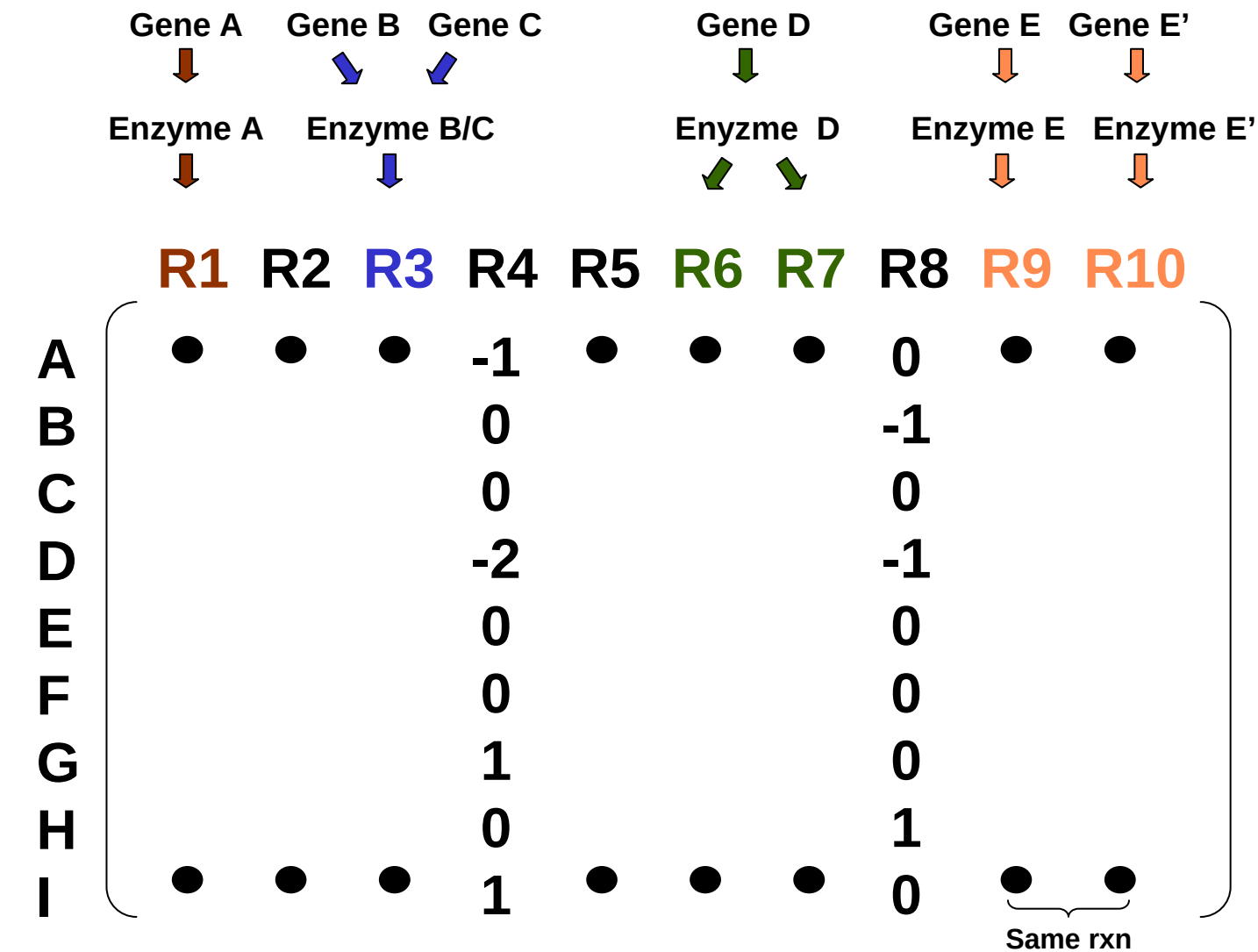
From Genomes to the S Matrix

Examples

Columns encode reactions

Relationships btw genes and rxns
-1 gene 1 rxn
-1 gene 1+ rxns
-1+ genes 1 rxn

The same reaction can be included as multiple roles (paralogs)



What Can We Use S For?

From S we can determine what combination of fluxes are possible in the system and what are not

To get there we need three concepts:

1. Nullspace of S
2. Extreme Pathways
3. Constrained Flux Space

The Steady State Assumption and S

- We have $\frac{dx}{dt} = S \bullet V$
- But also recall that at steady state, metabolite concentrations are constant: *dx/dt*=0

$$\frac{dx}{dt} = S \bullet V = 0$$

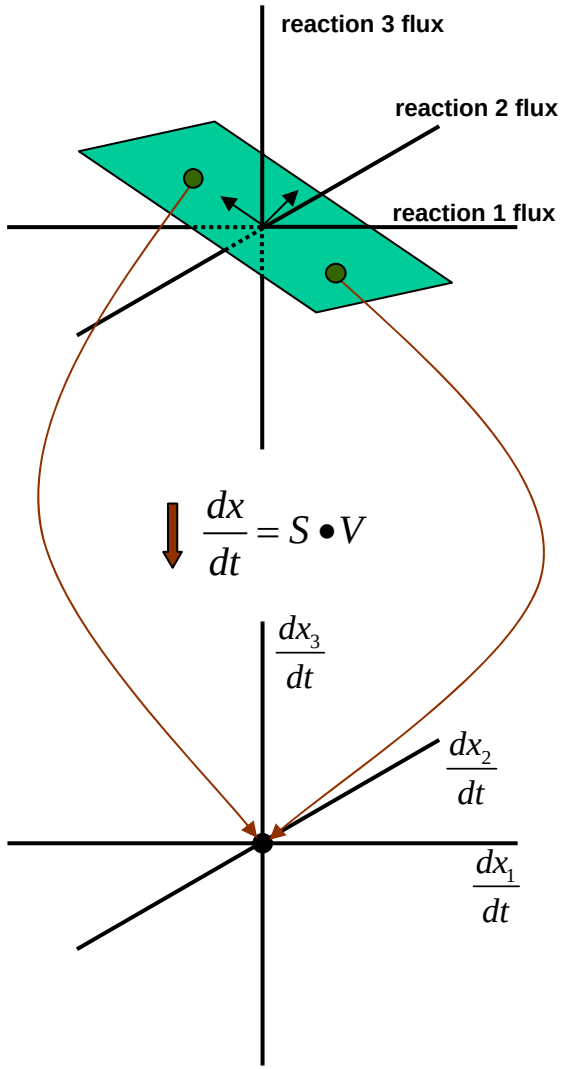
Steady State fluxes are *constrained* to the nullspace of S

The Nullspace of S

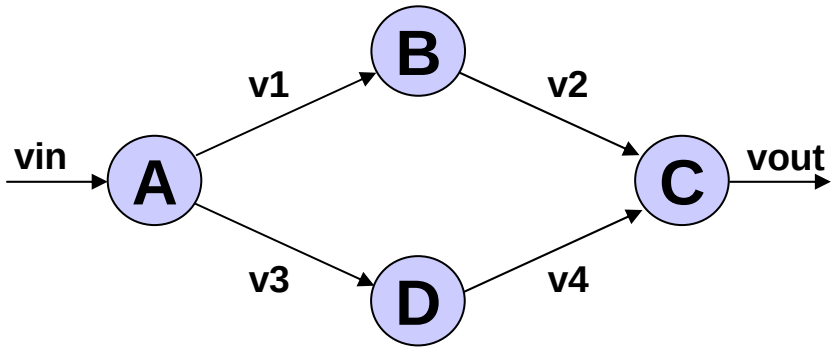
- Subspace of flux vectors that do not change metabolite concentrations
- Can describe nullspace with non-unique basis vectors, b_i
- All nullspace fluxes are linear combinations of this basis:

$$V = \sum_i \alpha_i b_i$$

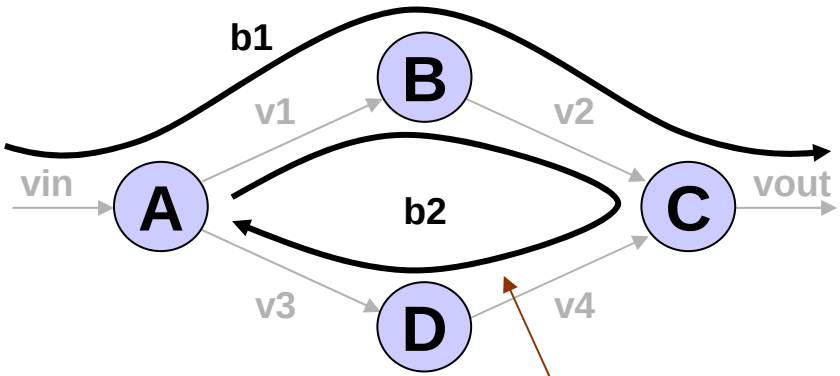
- Can find a basis using standard methods (e.g. SVD)



Example Nullspace Basis



	v1	v2	v3	v4	vin	vout
A	-1	0	-1	0	1	0
B	1	-1	0	0	0	0
C	0	1	0	1	0	-1
D	0	0	1	-1	0	0



b1	b2	
1	1	v1
1	1	v2
0	-1	v3
0	-1	v4
1	0	vin
1	0	vout

b2 includes *negative* fluxes that are not thermodynamically possible

-> Need to *constrain* the nullspace

Extreme Pathways

- The most fundamental constraint is that all fluxes must be *positive**
- In this case, we have the following linear homogeneous equation system:

$$0 = S \cdot V, \qquad v_i \geq 0, \quad i = 1..n$$

- Solution to this set of equations is an exercise in *convex analysis*
- Solution region can be described by a *unique* set of Extreme Pathways

*recall that reversible reactions are represented by two unidirectional fluxes

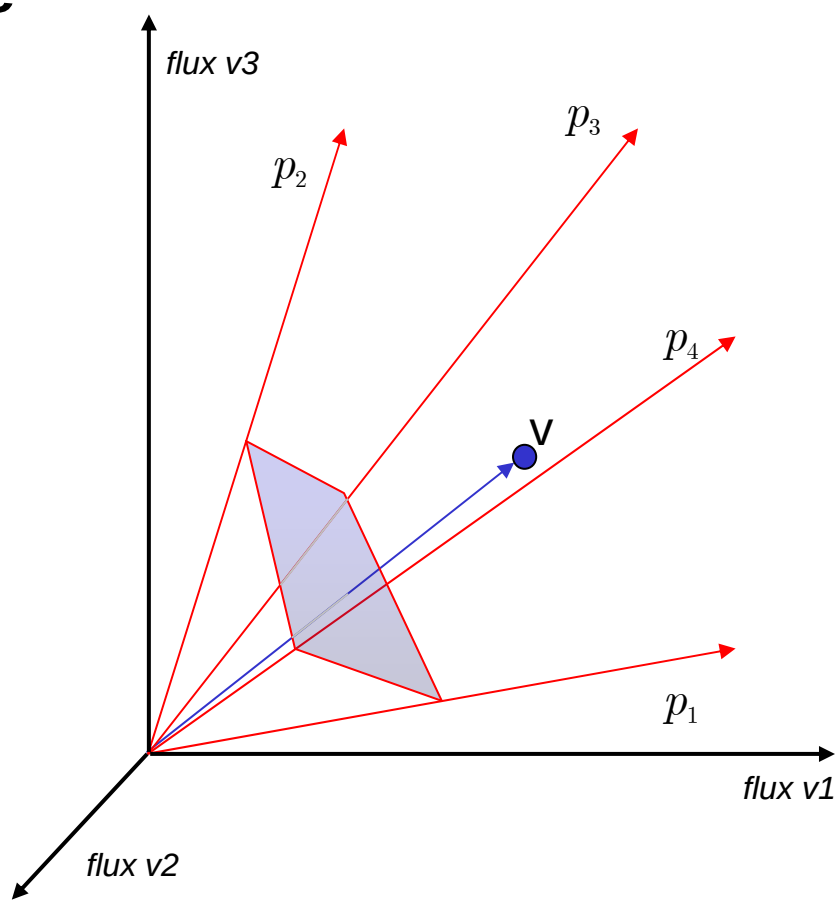
The Flux Cone

Extreme pathways circumscribe a *convex flux cone*

- *Every* steady state flux vector, v, is a *non-negative combination* of these pathways:

$$V = \sum_i \alpha_i p_i \qquad \alpha_i \geq 0$$

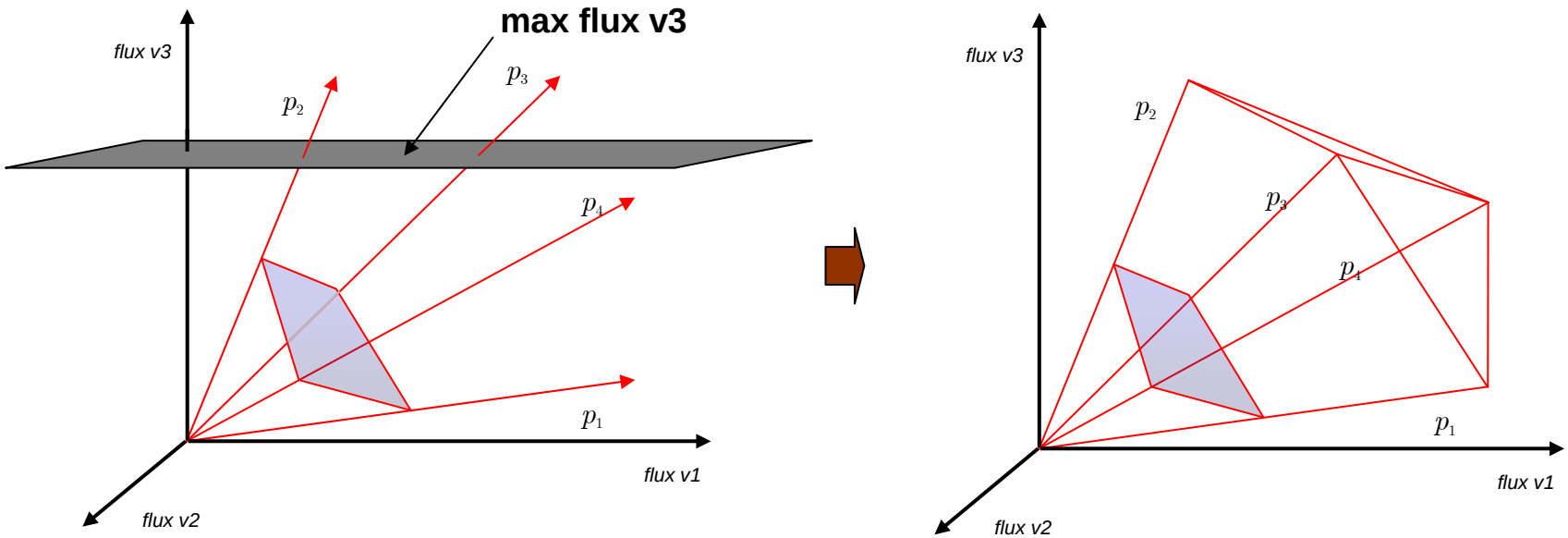
- Extreme pathways represent underlying pathway structure of system



Nullspace	**Flux Cone**
Vector space defined by set of non-unique basis vectors	Vector space defined by set of non-unique basis vectors
Every flux in space uniquely represented as linear combination of basis vectors	Every flux in space non-uniquely represented as non-negative combination of extreme pathways
# Basis vectors = dimension of nullspace	# Extreme pathways $\geq$ dimension of nullspace

Constraining the Solution Space

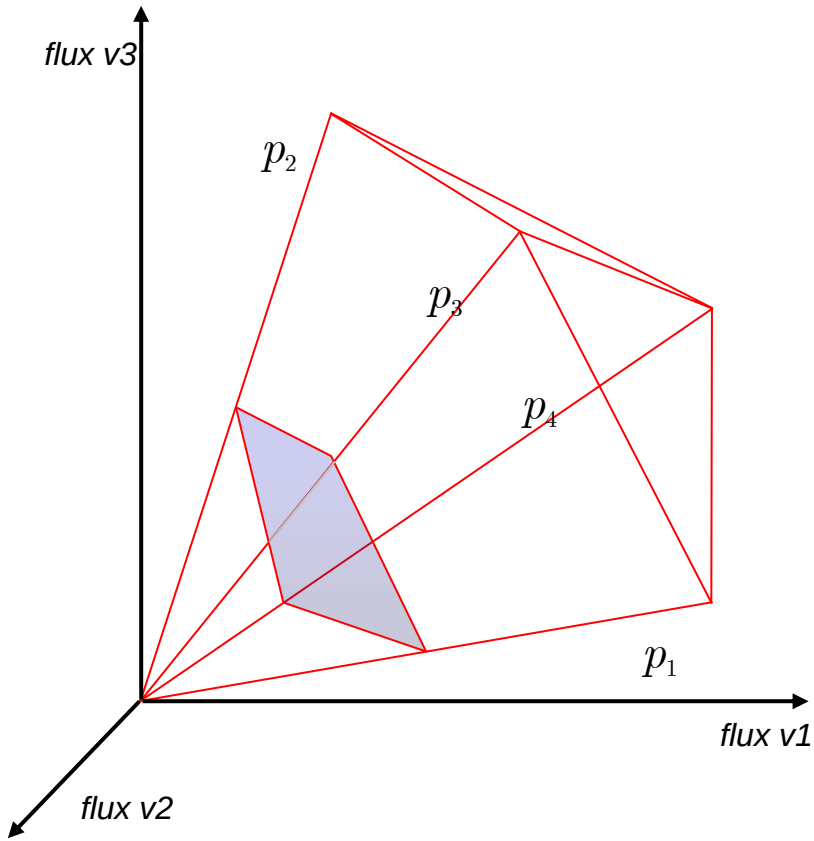
- No reaction has capacity for *infinite* flux
- Often one can estimate constraints on transfer fluxes
  - *Max glucose uptake measured at maximum growth rate*
  - *Max oxygen uptake based on diffusivity equation*
- Flux constraints result in constraints on extreme pathways
  - Need enough constraints to ‘cover’ extreme pathways



The Constrained Flux Cone

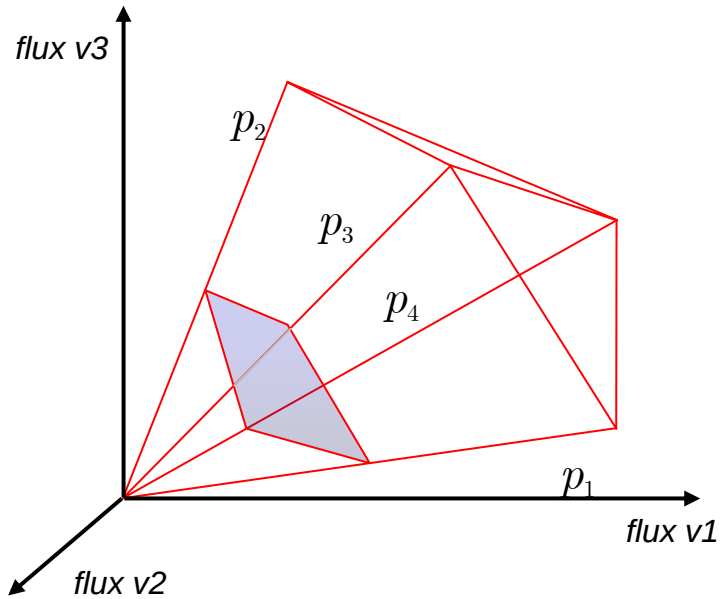
- Contains all achievable flux distributions given the constraints:
  - Stoichiometry
  - Reversibility
  - Max and Min Fluxes

- Only requires:
  - Annotation
  - Stoichiometry
  - Small number of flux constraints (small relative to number of reactions)



Selecting One Flux Distribution

- At any one point in time, organisms have a single flux distribution
- How do we narrow down the range of predicted flux distributions (ideally to one)?



What if we assume organisms are trying to maximize a "fitness" function that is a function of fluxes?

Linear Programming

If we assume the objective function is a *linear function of fluxes*, we can use *linear programming* to find a solution

Linear Programming

Maximize:

$$z = \sum_i c_i v_i = c^T v$$

Subject to:

$$Ax < b$$

$$x \geq 0$$

Solution always lies at boundary of admissible space

Can be found using *simplex algorithm*

Example

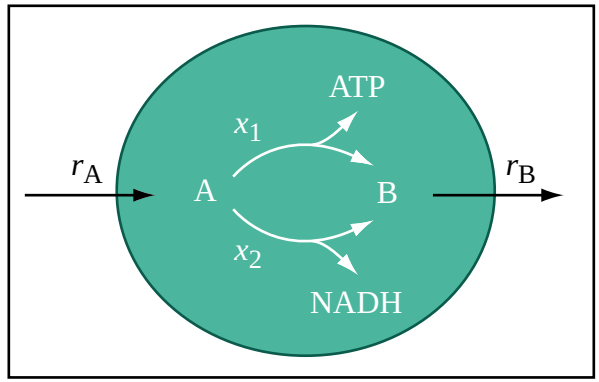
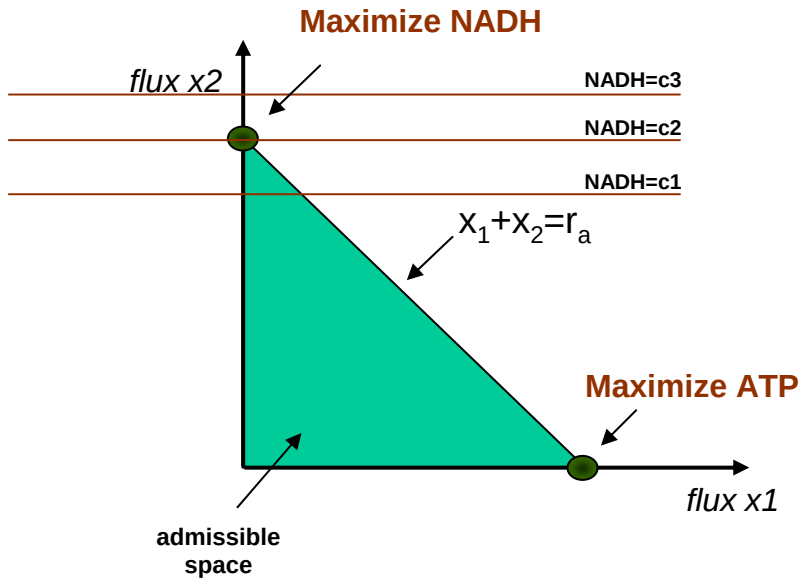


Figure by MIT OpenCourseWare.



Constraints:

$$x_i \geq 0$$

$$r_a \leq const.$$

From Bonarius et al TIBTECH vol 15:308 (1997)

Optimizing *E. coli* Growth

For one gram of E. coli biomass, you need this ratio of metabolites →

Metabolite	(mmol)
ATP	41.257
NADH	-3.547
NADPH	18.225
G6P	0.205
F6P	0.0709
R5P	0.8977
E4P	0.361
T3P	0.129
3PG	1.496
PEP	0.5191
PYR	2.8328
AcCoA	3.7478
OAA	1.7867
AKG	1.0789

Assuming a matched balanced set of metabolite fluxes, you can formulate this objective function ↓

$$Z = 41.257v_{ATP} - 3.547v_{NADH} + 18.225v_{NADPH} + 0.205v_{G6P} + 0.0709v_{F6P} +0.8977v_{R5P} + 0.361v_{E4P} + 0.129v_{T3P} + 1.496v_{3PG} + 0.5191v_{PEP} +2.8328v_{PYR} + 3.7478v_{AcCoA} + 1.7867v_{OAA} + 1.0789v_{AKG}$$

FBA Summary

Stoichiometric Matrix

Gene annotation
Enzyme and reaction catalog

Feasible Space

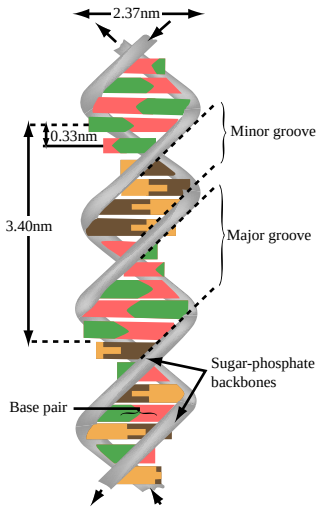
S*v=0

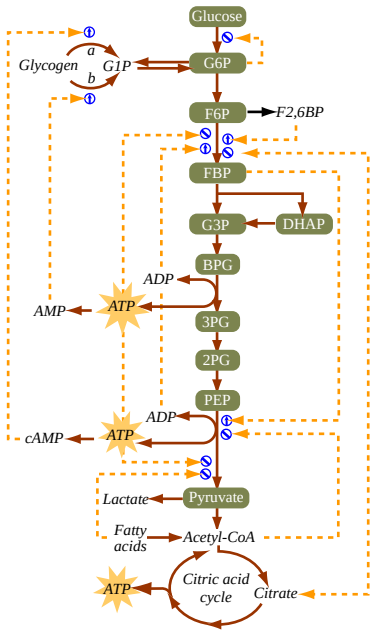
Add constraints:
$v_i>0$
$\alpha_i>v_i>\beta_i$

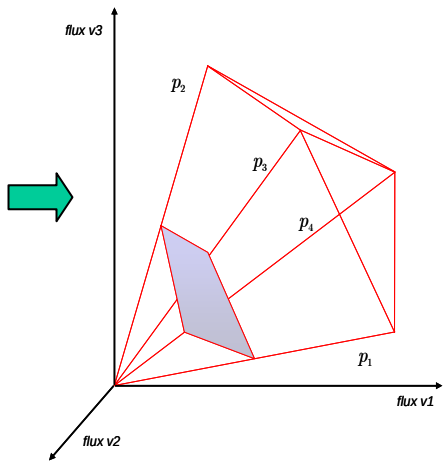
Optimal Flux

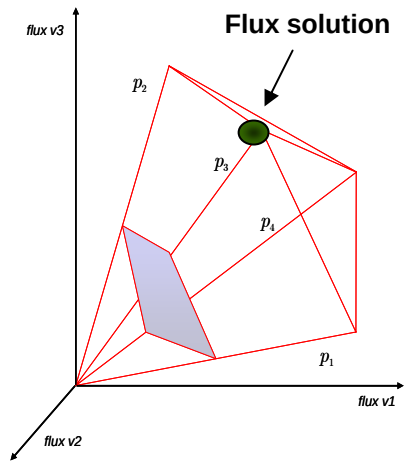
Growth objective
Z=c*v

Solve with linear programming









Next some applications of FBA….

Applications

in silico Deletion Analysis

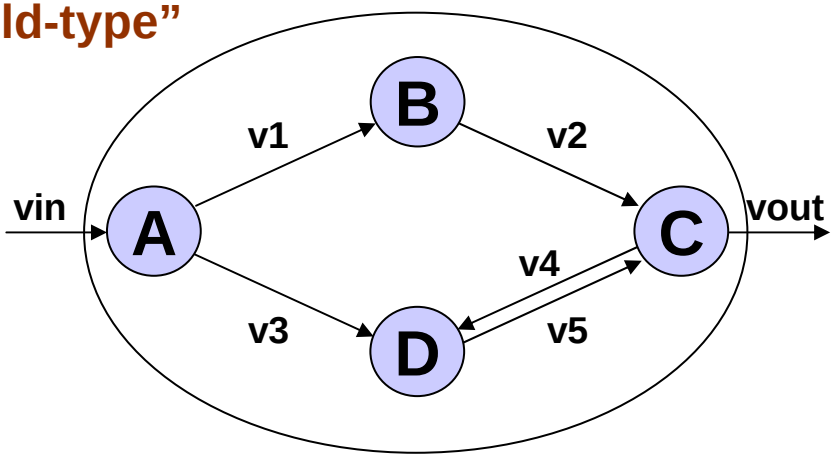
Can we predict gene knockout phenotype based on their simulated effects on metabolism?

Q: Why, given other computational methods exist? (e.g. protein/protein interaction map connectivity)

A: Other methods do not directly consider metabolic flux or specific metabolic conditions

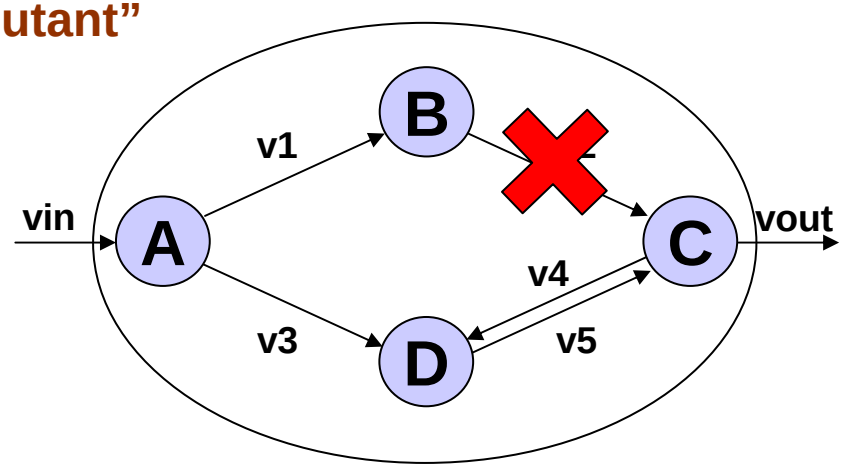
in silico Deletion Analysis

“wild-type”



	v1	v2	v3	v4	v5	vin	vout
A	-1	0	-1	0	0	1	0
B	1	-1	0	0	0	0	0
C	0	1	0	-1	1	0	-1
D	0	0	1	1	-1	0	0

“mutant”



	v1	v2	v3	v4	v5	vin	vout
A	-1		-1	0	0	1	0
B	1		0	0	0	0	0
C	0		0	-1	1	0	-1
D	0		1	1	-1	0	0

Gene knockouts modeled by *removing a reaction*

Mutations Restrict Feasible Space

- KO removes fluxes, and extreme pathways that depend on these fluxes
- Feasible space is constrained
- If original optimal flux is outside new space, new optimal flux is created
- Growth rate at new solution provides a measure of KO phenotype

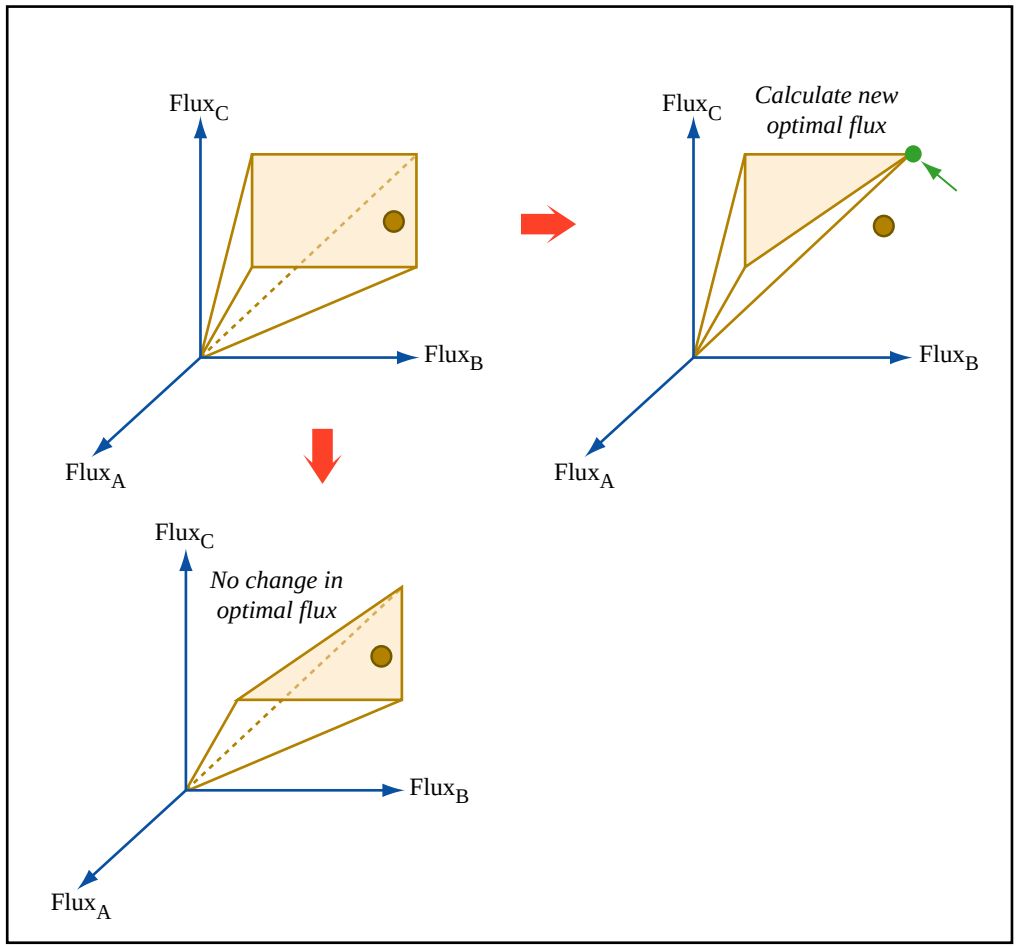


Figure by MIT OpenCourseWare.

Mutant Phenotypes in *E. coli*

Edwards, J.S., and B.O. Palsson. "The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities." *PNAS* 97, no. 10 (2000): 5528-5533.

Model of E. coli central metabolism
436 metabolites
720 reactions

Simulate mutants in glycolysis, pentose phosphate, TCA, electron transport

E. coli KO simulation results

in silico \ *in vivo*	+	-
+	36	2
-	9	32

If Zmutant/Z =0, mutant is no growth (-), growth (+) otherwise

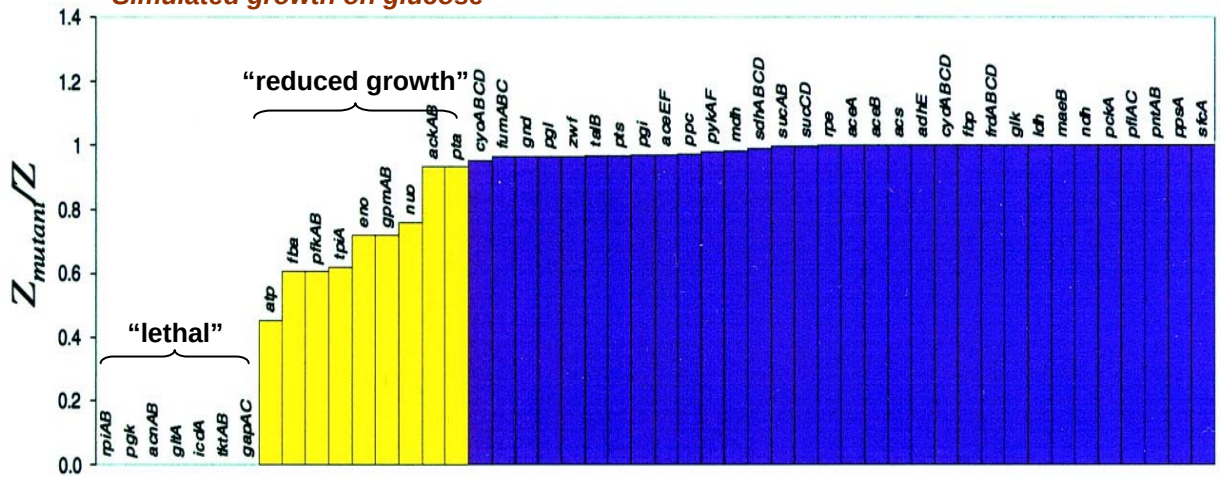
Compare to experiment (in vivo / in silico) 86% agree

Measured optimal growth of mutants (Z_{mutant}) versus non-mutant (Z)

Condition specific prediction

Gene	glc	gl	succ	ac
aceA	+/+		+/+	−/−
aceB				−/−
aceEF[*]	−/+			
ackA				+/+
acn	−/−			−/−
acs				+/+
cyd	+/+			
cyo	+/+			
eno[†]	−/+	−/+	−/−	−/−
fba[‖]	−/+			
fbp	+/+	−/−	−/−	−/−
frd	+/+		+/+	+/+
gap	−/−	−/−	−/−	−/−
glk	+/+			
gltA	−/−			−/−
gnd	+/+			
idh	−/−			−/−
mdh[††]	+/+	+/+	+/+	
ndh	+/+	+/+		
nuo	+/+	+/+		
pfk[†]	−/+			
pgi[‡]	+/+	+/−	+/−	
pgk	−/−	−/−	−/−	−/−
pgl	+/+			
pntAB	+/+	+/+	+/+	
ppc[§]	±/+	−/+	+/+	
pta				+/+
pts	+/+			
pyk	+/+			
rpi	−/−	−/−	−/−	−/−
sdhABCD	+/+		−/−	−/−
sucAB	+/+		−/+	−/+
tktAB	−/−			
tpi[**]	−/+	−/−	−/−	−/−
unc	+/+		±/+	−/−
zwf	+/+	+/+	+/+	

Simulated growth on glucose



Edwards, J. S., and B. O. Palsson. "The Escherichia coli MG1655 in Silico Metabolic Genotype: Its Definition, Characteristics, and Capabilities." *PNAS* 97, no. 10 (2000): 5528-5533. Copyright (2000) National Academy of Sciences, U.S.A.

Edward & Palsson (2000) PNAS

What do the errors tell us?

- Errors indicate gaps in model or knowledge
- Authors discuss 7 errors in prediction
  - *fba* mutants inhibit stable RNA synthesis (not modeled by FBA)
  - *tpi* mutants produce toxic intermediate (not modeled by FBA)
  - 5 cases due to possible regulatory mechanisms (*aceEF, eno, pfk, ppc*)

Edward & Palsson (2000) PNAS

Yeast Metabolic Model

- 1175 Reactions
- 585 Metabolites
- Accounts for 708 (16%) genes
- Includes 140 reactions w/o known genes
- Cytosol and mitochondria compartments
- Palsson group continues to update and improve model

Image removed due to copyright restrictions.

Figure 1, Reconstruction of the metabolic network of *S. cerevisiae*.
Forster, Jochen, et al. "Genome-Scale Reconstruction of the Saccharomyces Cerevisiae Metabolic Network." *Genome Research* 13 (2003): 244-253.

Model available at
http://systemsbiology.ucsd.edu

Forster et al. (2003) Genome Res

Yeast Knockout Analysis

Reported:

81.5% agreement

93 of 114 cases

But broken down by case:

in silico \ *in vivo*	+	-
+	86	20
-	0	3

Many errors of (+/-)

- 7 predict retarded growth

- Others can be explained by unmodeled regulation

Eukaryotic model needs *gene regulation*

Gene	Defined complete Glc (*in silico*/*in vivo*)	Defined minimal Glc (*in silico*/*in vivo*)	Defined minimal Ace (*in silico*/*in vivo*)	Defined minimal Eth (*in silico*/*in vivo*)
ACO1	(+/+)	(−/−)		
CDC19*	(+/−)	(+/−)		
CIT1	(+/+)	(+/+)		
CIT2	(+/+)	(+/+)		
CIT3	(+/+)			
DAL7	(+/+)	(+/+)	(+/+)	(+/+)
ENO1	(+/+)			
ENO2†	(+/−)	(+/−)		
FBA1‡	(+/−)	(+/−)		
FBP1	(+/+)	(+/+)		(−/−)
FUM1	(+/+)			
GLK1	(+/+)			
GND1§	(+/−)	(+/−)		
GND2	(+/+)			
GPM1¶	(+/−)	(+/−)		
GPM2	(+/+)			
GPM3	(+/+)			
HXK1	(+/+)			
HXK2	(+/+)			
ICL1	(+/+)	(+/+)		
IDH1	(+/+)	(+/+)		
IDH2	(+/+)	(+/+)		
IDP1	(+/+)	(+/+)		
IDP2	(+/+)	(+/+)		
IDP3	(+/+)			
KGD1	(+/+)	(+/+)		
KGD2	(+/+)	(+/+)		
LPD1	(+/+)			
LSC1	(+/+)		(+/+)	(+/+)
LSC2	(+/+)		(+/+)	(+/+)
MAE1	(+/+)	(+/+)		(+/+)

Gene	Defined complete Glc (*in silico*/*in vivo*)	Defined minimal Glc (*in silico*/*in vivo*)	Defined minimal Ace (*in silico*/*in vivo*)	Defined minimal Eth (*in silico*/*in vivo*)
MDH1	(+/+)	(+/+)	(+/−)	
MDH2	(+/+)		(+/−)	(+/−)
MDH3	(+/+)			
MLS1	(+/+)	(+/+)	(+/+)	(+/+)
OSM1	(+/+)			
PCK1	(+/+)			
PDC1	(+/+)	(+/+)		
PDC5	(+/+)	(+/+)		
PDC6	(+/+)	(+/+)		
PFK1	(+/+)	(+/+)		
PFK2	(+/+)	(+/+)		
PGI1‡‖	(+/−)	(+/−)		
PGK1‡	(+/−)	(+/−)		
PGM1	(+/+)	(+/+)		
PGM2	(+/+)	(+/+)		
PYC1	(+/+)	(+/+)	(+/−)	(+/−)
PYC2	(+/+)			
PYK2	(+/+)	(+/+)		(+/+)
RKI1	(−/−)			
RPE1	(+/+)			
SOL1	(+/+)			
SOL2	(+/+)			
SOL3	(+/+)			
SOL4	(+/+)			
TAL1	(+/+)	(+/+)		
TDH1	(+/+)			
TKL2	(+/+)			
TPI1‡**	(+/−)			
ZWF1	(+/+)	(+/+)		

Famili et al. (2003) PNAS

Resources

- **Tools and Databases**
  - Kegg
  - BioCyc
  - PathwayExplorer (pathwayexplorer.genome.tugraz.at)

- **Metabolic Modeling**
  - Palsson's group at UCSD (http://gcrg.ucsd.edu/)
  - www.systems-biology.org
  - Biomodels database (www.ebi.ac.uk/biomodels/)
  - JWS Model Database (jjj.biochem.sun.ac.za/database/index.html)