Practical Bioinformatics (Bio334)
Metabolic network analysis

Andreas Wagner

Institute of Evolutionary Biology
and Environmental Studies, UZH
andreas.wagner@ieu.uzh.ch

Two kinds of bioinformatic analyses

Analyses of **genotypes**

- Genome sequences
- Single nucleotide polymorphisms
- Gene families

Analyses of (molecular) **phenotypes**

- Protein structure
- Microarray gene expression data
- Protein interaction data

Most challenging:

Predicting phenotype from genotype

Important to understand

- gene functions
- how mutations alter phenotypes
- genetic risk factors for disease
- **how new phenotypes (adaptations, innovations) originate**
- **how natural selection changes organisms**

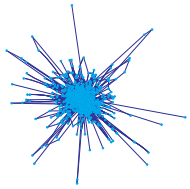
Still impossible for

- whole organisms
- even proteins

Possible for some kinds of biological networks ...

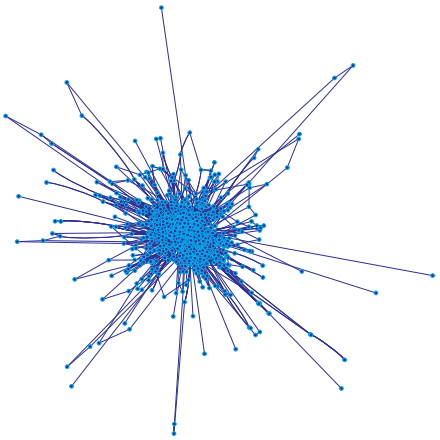
Genome-scale networks comprise
hundreds to thousands of genes and gene products

Protein interaction networks
Transcriptional regulation networks
Metabolic networks

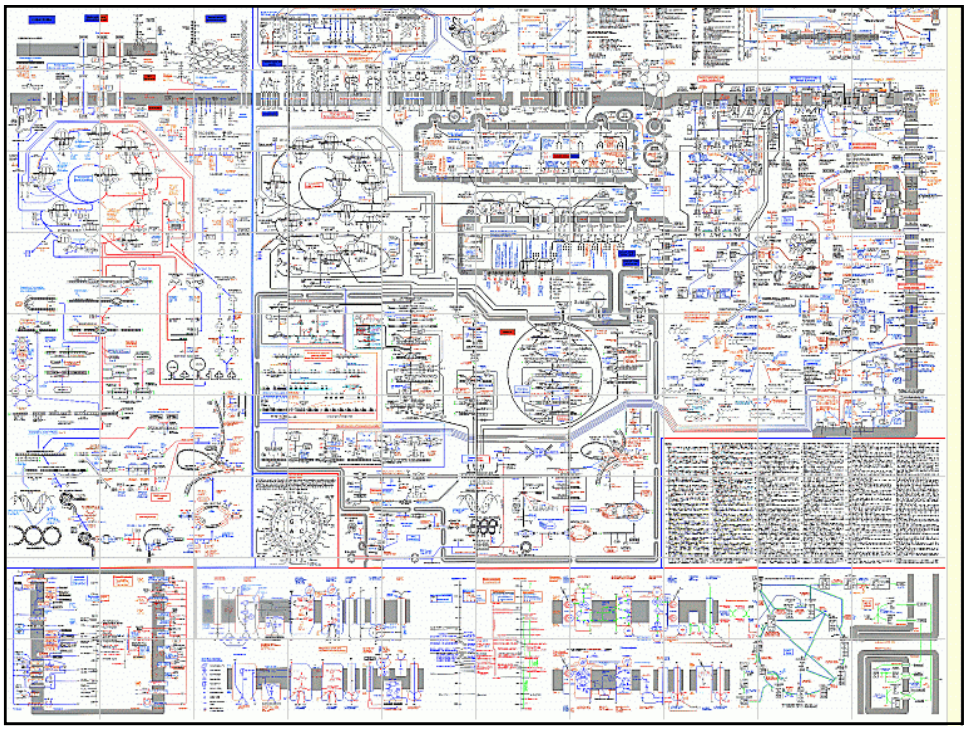


Mathematical characterization based on qualitative understanding of network topology

Metabolic networks



Pajek



A metabolic network is a set of chemical reactions whose two main functions are to produce

chemical energy
(for maintenance of cell functions and for biosyntheses)

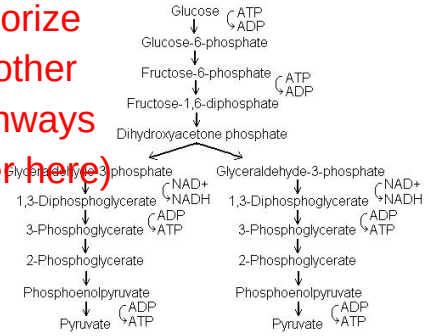
molecular building blocks for biosyntheses

These reactions are catalyzed by enzymes that are encoded by genes.

Even in bacteria like *E.coli*, several hundred such enzymatic reactions are necessary to fulfill these functions.

You already know small parts of metabolic networks: glycolysis

no need to memorize this pathway or other biochemical pathways (just for refresher here)



Converts glucose into pyruvate

Net gain: 2 ATP, 2 NADH per glucose

Provides precursors
3-phosphoglycerate for serine ...

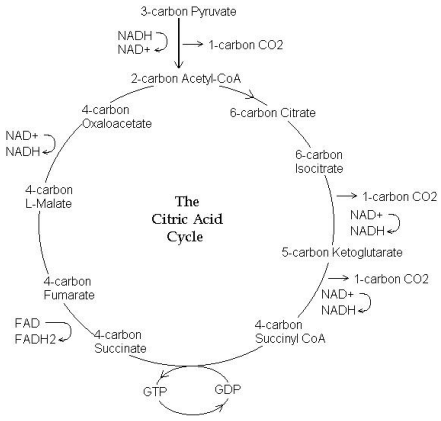
Aerobically: leads into citric acid cycle

Anaerobically: creates molecules like ethanol or lactate to accept electrons from and regenerate NAD from NADH

http://homepages.ius.edu/GKIRCHNE/Glycolysis.htm

Summary reaction:
Glucose + 2ADP + 2NAD + $2P_i$ + $2H^+$-> 2Pyruvate + 2ATP+ 2NADH

You already know small parts of metabolic networks: citric acid cycle



http://homepages.ius.edu/GKIRCHNE/Glycolysis.htm

Starts from decarboxylation of pyruvate into Acetyl-coA

Net energy gain: 3 NADH, 1 FADH2, 1 GTP per pyruvate

3 NADH and 1 FADH2 are transformed into 9 ATP through oxydative phosphorylation

Provides precursors: oxaloacetate for aspartate ...

You already know small parts of metabolic networks: central carbon metabolism

- Glycolysis
- Citric acid cycle
- Pentose phosphate shunt
  - Produces NADPH and precursors for biosyntheses (e.g., ribose-5-phosphate to synthesize RNA and DNA)

The metabolic network of a whole organism comprises many reactions

Organism	Reactions	Metabolites
E. coli	2077	1039
S. cerevisiae	1412	713
A. thaliana	1567	1748
B. aphidicola	263	240

buchnera aphidicola

buchnera aphidicola lives within aphids so it doesnt need so many reactions since it's an endosymbiont (they only provide aphids with amino acids, and also their genome has been greatly reduced)

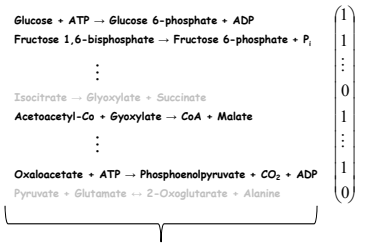
Metabolic genotype

The part of a genome that encodes metabolic enzymes

Less unwieldy:
An organism's set of (enzyme-catalyzed) metabolic reactions

These can be written as follows

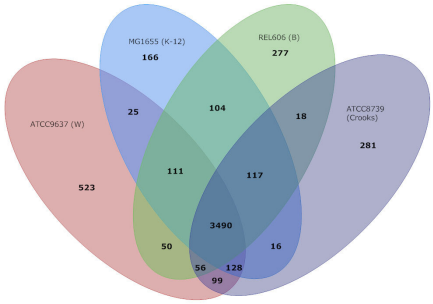
A metabolic genotype



The known "universe" of metabolic reactions (>5000 reactions)

Gene content and metabolic genotypes can evolve rapidly

Numbers of genes in the genomes of four different Escherichia coli strains



E. coli strain	K-12	W
Genes	4493	4482
Unique Genes	403	728
Unique Reactions	**>100**	**>100**

Archer et al. BMC Genomics 2011

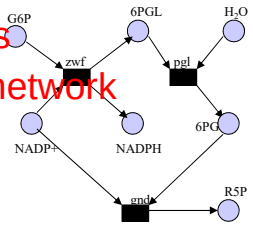
Graphs can (crudely) represent large chemical reaction networks

Stoichiometric Equations

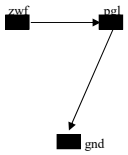
1 Glucose 6-phosphate (G6P) + 1 $NADP^+$	$\overset{zwf}{\Rightarrow}$	1 6-Phosphoglucono δ-lactone (6PGL) + 1 NADPH
1 6-Phosphoglucono δ-lactone + 1 H_2O	$\overset{pgl}{\Rightarrow}$	1 6-Phosphogluconate (6PG)
1 6-Phosphogluconate + 1 $NADP^+$	$\overset{gnd}{\Rightarrow}$	1 Ribulose 5-phosphate (R5P) + 1 NADPH
1 Ribulose 5-phosphate	$\overset{rpe}{\Leftrightarrow}$	1 Xylulose 5-phosphate (X5P)

only rxns

Bipartite graph

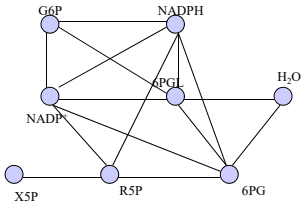


Enzyme graph



Substrate graph

no enzymes here

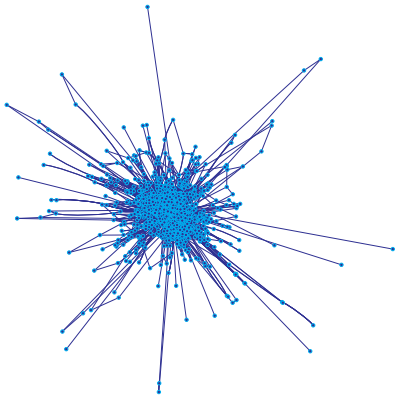


complete repres
of a metabolic network

these two are simpplified representations

bipartite graph:
two node types:
metabolites and enzymes and only they are connected with each other.
no connection between metabolites directly.

An enzyme graph representation of the metabolic network of the yeast Saccharomyces cerevisiae



Fajek

Graph-based analyses of metabolic networks neglect that they have a biological function and phenotype

To understand this phenotype, one needs to understand how matter <u>flows</u> through these networks

Metabolic flux: the rate at which an enzyme converts substrate into product per unit time.

Metabolic phenotype

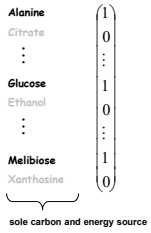
Most general: All the molecules that a metabolism can synthesize in a given chemical environment

The most important of these molecules are biomass precursors (amino acids, DNA and RNA building blocks etc.).

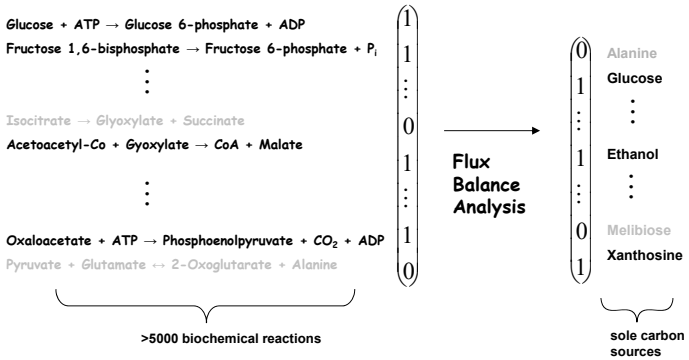
A metabolism is **viable** if it can synthesize all of them.

More specific: the spectrum of nutrients on which a metabolism is viable

A carbon utilization phenotype



One can predict metabolic phenotypes from metabolic genotypes



For optimal cell growth, metabolic networks need to produce biochemical precursors in well-balanced amounts.

This requires a specific distribution of metabolic fluxes through enzymatic reactions in the network

Flux balance analysis can predict this distribution, and the maximally possible rate at which biochemical precursors, and thus biomass, can be synthesized by a metabolism

Flux balance analysis needs

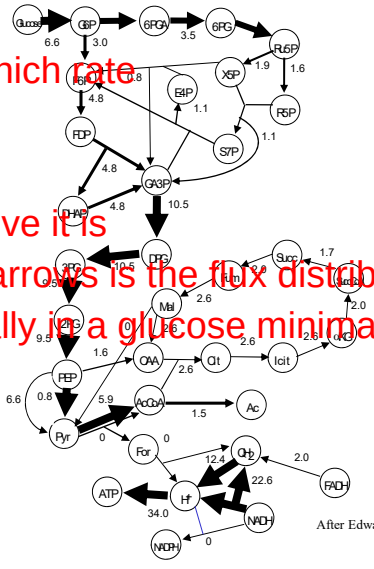
1. a list of chemical reactions known to be catalyzed by enzymes in a given organism
2. Information about nutrients in the chemical environment of a cell and their uptake rate (usually in mol/g dry weight [DW] and hour)

Flux balance analysis computes

1. allowable metabolic fluxes through a metabolic network (fluxes that do not violate the law of mass conservation)

2. within the set of allowable fluxes, those that have desirable properties (e.g., maximal rate of biomass production, maximal biomass yield per unit of carbon source.)

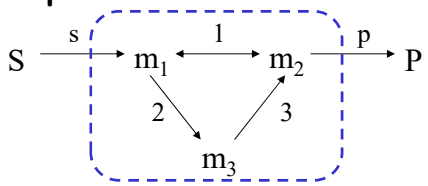
Metabolic flux through central carbon metabolism of *E.coli* growing at a maximally possible rate in a glucose-minimal medium



After Edwards JS, Palsson BO. 2000. *PNAS* 97: 5528-33

thickness of arrows reflects at which rate a chemical reaction proceeds
(thick = high rate)
the smaller the rate, the less active it is
the distribution of the thickness arrows is the flux distribution at which e.coli can grow maximally in a glucose minimal env

A simple chemical reaction network



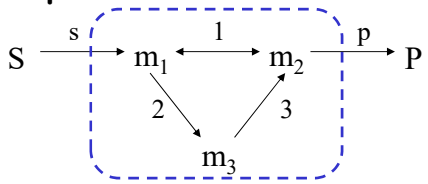
Two external metabolites
- 1 substrate (nutrient) S
- 1 product P

Two transport reactions s,p.
- s determines the uptake rate of S
- p determines the rate at which P is excreted

Three internal metabolites (m_i)

Three internal reactions (1,2,3)

A simple chemical reaction network



Metabolite concentrations m_i change according to the equations

$$\frac{dm_1}{dt} = v_s - v_1 - v_2$$
$$\frac{dm_2}{dt} = v_1 + v_3 - v_p$$
$$\frac{dm_3}{dt} = v_2 - v_3$$

$$\frac{d\vec{m}}{dt} = \mathbf{S}\vec{v}$$

$$\mathbf{S} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix}$$

Stoichiometry matrix

v_i metabolic flux through reaction i

$\vec{v} = (v_s, v_1, v_2, v_3, vP)^\top$

Rows: internal metabolites
Columns: reactions

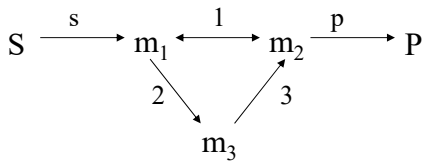
An example stoichiometric matrix for bimolecular reactions

Stoichiometric Equations

1 Glucose 6-phosphate (G6P) + 1 $NADP^+$	$\overset{zwf}{\Rightarrow}$	1 6-Phosphoglucono d-lactone (6PGL) + 1 NADPH
1 6-Phosphoglucono d-lactone + 1 H_2O	$\overset{pgl}{\Rightarrow}$	1 6-Phosphogluconate (6PG)
1 6-Phosphogluconate + 1 $NADP^+$	$\overset{gnd}{\Rightarrow}$	1 Ribulose 5-phosphate (R5P) + 1 NADPH
1 Ribulose 5-phosphate	$\overset{rpe}{\Leftrightarrow}$	1 Xylulose 5-phosphate (X5P)

Stoichiometric Matrix

	zwf	pgl	gnd	rpe
$G6P$	-1	0	0	0
$6PGL$	1	-1	0	0
$6PG$	0	1	-1	0
$R5P$	0	0	1	-1
$X5P$	0	0	0	0
$NADP$	-1	0	-1	0
$NADPH$	1	0	1	0
$H2O$	0	-1	0	0



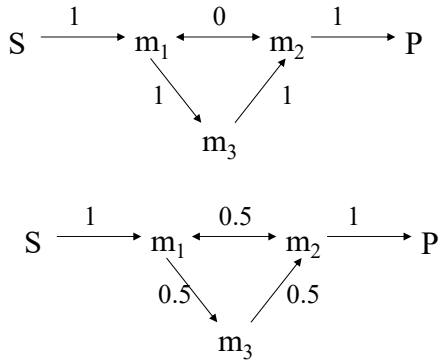
FBA assumes that metabolism is in a steady state where the concentrations of metabolites no longer change

$$\frac{d\vec{m}}{dt} = 0$$

$$\mathbf{S}\vec{v} = 0$$

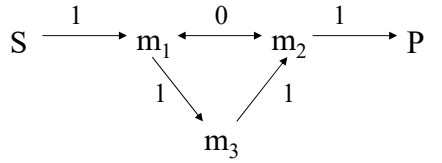
The solutions of these equations are the allowable metabolic fluxes. They form the so-called <u>null space of S</u>

Two allowable flux distributions for our example network

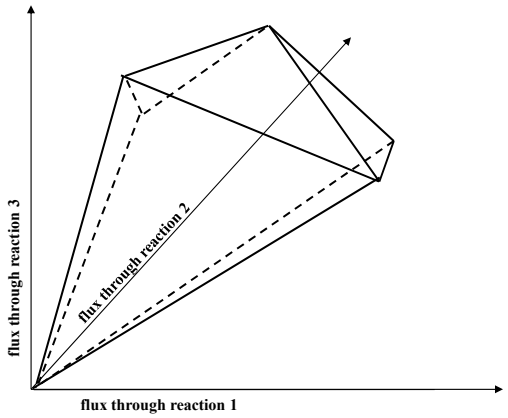


All fluxes of the form (1,x,1-x,1-x,1), 0≤x≤1 are allowable

A flux distribution that maximizes the rate at which reaction 2 (producing m_3) proceeds



The null space of a complex metabolic network forms a high-dimensional "flux cone" (a convex polytope)



Several important properties of a metabolic network can be expressed as a weighted sum of fluxes

$$Z(\vec{v}) = \sum_{i=1}^{m} c_i v_i$$

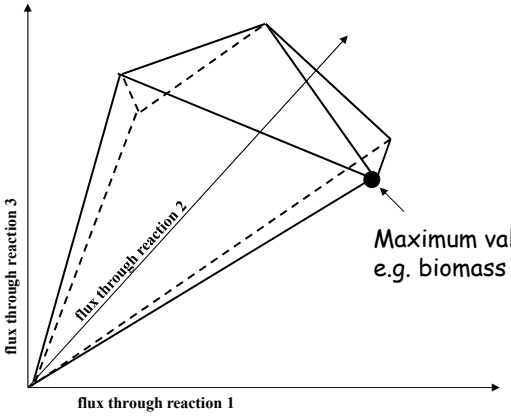
Example:

In the biomass growth reaction,

v_i is the rate at which essential biochemical precursor *i* is produced by a metabolic network.

c_i is a constant that reflects the relative contribution of precursor *i* to biomass (can be estimated from the biomass composition of a cell.)

FBA uses linear programming to determine regions within the flux cone where some flux or a linear function Z of multiple fluxes is maximal.



Z is also sometimes called an objective function

A biomass growth reaction for E.coli

0.513689 ala-L + 0.295792 arg-L + 0.241055 asn-L + 0.241055 asp-L + 0.091580 cys-L + 0.263160 gln-L + 0.263160 glu-L + 0.612638 gly + 0.094738 his-L + 0.290529 ile-L + 0.450531 leu-L + 0.343161 lys-L + 0.153686 met-L + 0.185265 phe-L + 0.221055 pro-L + 0.215792 ser-L + 0.253687 thr-L + 0.056843 trp-L + 0.137896 tyr-L + 0.423162 val-L + 0.026166 datp + 0.027017 dctp + 0.027017 dgtp + 0.026166 dttp + 0.133508 ctp + 0.215096 gtp + 0.144104 utp + 0.013894 murein5px4p[p] + 0.019456 kdo2lipid4[e] + 0.063814 pe160 + 0.075214 pe161 + 0.177645 k + 0.011843 nh4 + 0.007895 mg2 + 0.004737 ca2 + 0.007106 fe2 + 0.007106 fe3 + 0.003158 cu2 + 0.003158 mn2 + 0.003158 mobd + 0.003158 cobalt2 + 0.003158 zn2 + 0.004737 cl + 0.003948 so4 + 0.003948 pi + 0.000576 coa + 0.001831 nad + 0.000447 nadp + 0.000223 fad + 0.000223 thf + 0.000223 mlthf + 0.000223 10fthf + 0.000223 thmpp + 0.000223 pydx5p + 0.000223 pheme + 0.000223 sheme + 0.000055 udcpdp + 0.000223 amet + 0.000223 2ohph + 0.000223 ribflv + 59.984800 atp + 54.462000 h2o

-->

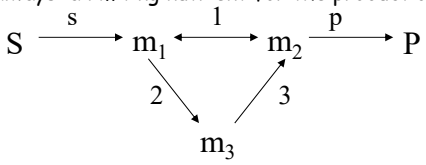
biomass

Limiting nutrient

a nutrient whose uptake rate limits
the synthesis of biomass or of another molecule

Lowering its uptake rate would lower the synthesis rate

In our simple example network
S is always a limiting nutrient for the production of P.



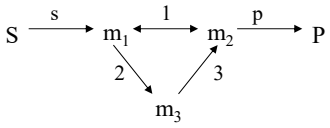
In complex networks and chemically complex environments,
not all available nutrients need to be limiting.

(Relative) Yield

The amount of a metabolic product produced
(relative to a unit of nutrient input).

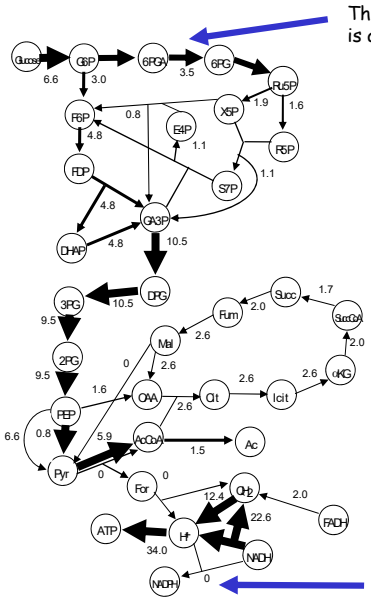
A measure of metabolic efficiency

Example 1: In our simple network, if v_s=1 [mmol / g DW /hr]
then the yield of product P is also v_p=1 (in steady state).



Example 2: ATP yield per mole of glucose in central metabolism
Glycolysis: 2 moles ATP
oxidative phosphorylation: approximately 30 moles ATP

Active, inactive, and essential reactions

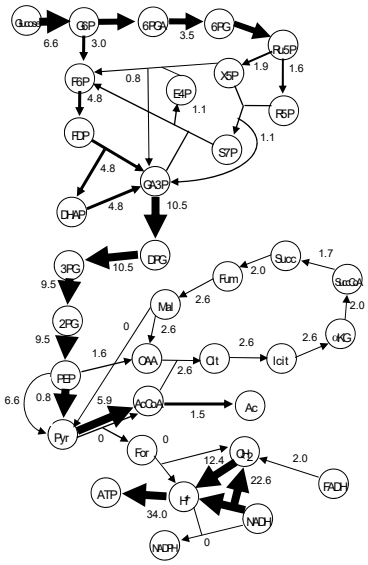


The lactonase reaction creating 6-phosphogluconate is active (flux >0)

The transhydrogenase reaction is inactive (flux zero)

After Edwards JS, Palsson BO. 2000. *PNAS* 97: 5528-33

Active, inactive, and essential reactions



Synthetic null or knock-out mutations of enzyme-coding genes can set the metabolic flux through a reaction to zero

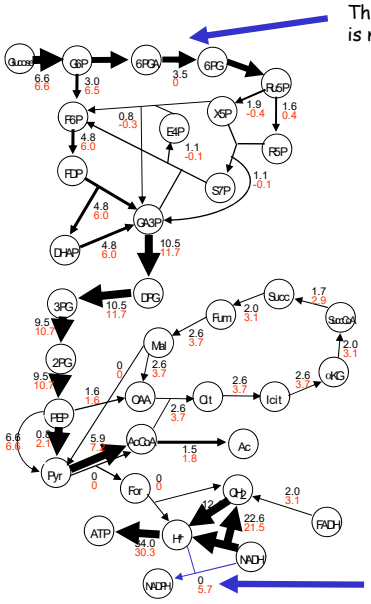
Such mutations have been created for thousands of genes in well-studied organisms

Knock-out of an <u>essential</u> metabolic gene or reaction causes inviability (a zero biomass growth flux)

Knock-out of a <u>non-essential</u> metabolic gene or reaction does not cause inviability

After Edwards JS, Palsson BO. 2000. *PNAS* 97: 5528-33

Active, inactive, and essential reactions

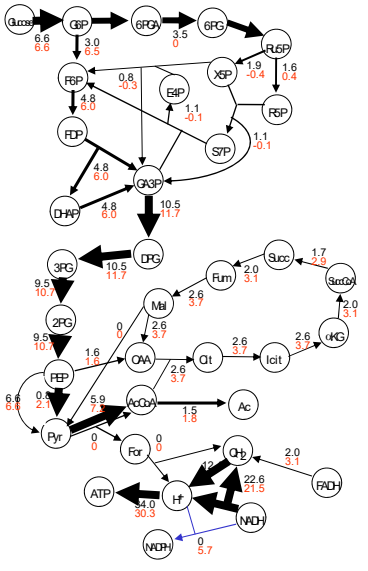


The lactonase reaction creating 6-phosphogluconate is not essential

Knocking out the lactonase reaction requires activation of the transhydrogenase reaction

After Edwards JS, Palsson BO. 2000. *PNAS* 97: 5528-33

Active, inactive, and essential reactions



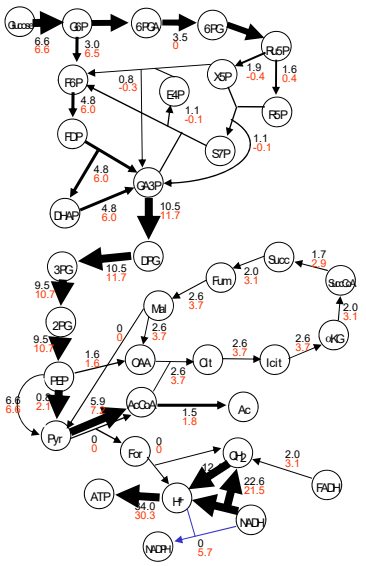
A reaction's essentiality strongly depends on the chemical environment

In complex networks like that of *E.coli* most reactions are non-essential in most chemical environments.

Knocking out non-essential reactions can require large reorganizations of metabolic flux to maintain viability

After Edwards JS, Palsson BO. 2000. *PNAS* 97: 5528-33

Active, inactive, and essential reactions



After Edwards JS, Palsson BO. 2000. *PNAS* 97: 5528-33

all essential reactions are active...

all silent reactions are inessential...

some active reactions are essential, others are inessential...

... in a given environment

Example questions for flux balance analysis

Can a given organism (metabolism) survive in environment X?

How fast could it grow in this environment?

Why are many enzymatic reactions dispensable in any one environment?

Why do some metabolisms have many reactions, while others have few?

Does network function and flux influence network evolution

Is it possible to design "resistance-proof" antimetabolic drugs?

Overview practical exercises

Exercise block 1 ("paper and pencil" analysis of toy networks):
translating reaction diagrams into stoichiometry matrices
dependency of synthesis rates on uptake rates
essential and active reactions
maximal yields

Exercise block 2 (glycolysis and central carbon metabolism):
essential reactions
computing reaction numbers
limiting nutrients, alternative carbon sources, yields

Exercise block 3 (genome-scale metabolic networks):
expensive and cheap metabolites
essential reactions

exam: 1 problem which is similar to one of the problems solved

Overview practical exercises

Exercise block 4 (genome-scale metabolisms):
essential reactions and flux distributions in different environments
rich and minimal environments
alternative carbon sources

Exercise block 5 (genome-scale metabolisms):
metabolic system analysis and antimetabolic drugs