V14 Metabolic Networks - Introduction

There exist different levels of computational methods for describing metabolic networks:

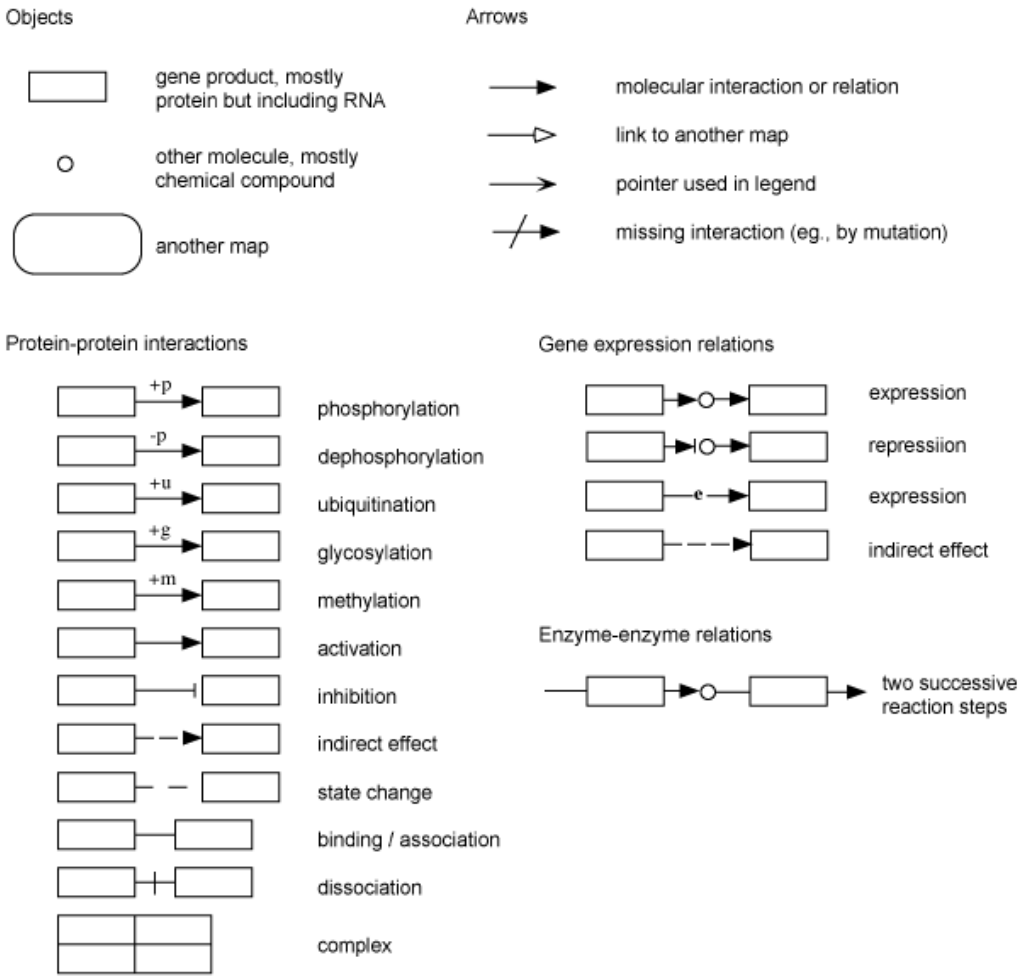
- stoichiometry/kinetics of classical biochemical **pathways** (glycolysis, TCA cycle, ...

- stoichiometric modelling (**flux balance analysis**): theoretical capabilities of an integrated cellular process, feasible metabolic flux distributions

- automatic decomposition of metabolic networks (elementary nodes, extreme pathways ...)

- **kinetic modelling** of coupled cellular pathways (E-Cell ...)
General problem: lack of kinetic information on the dynamics and regulation of cellular metabolism

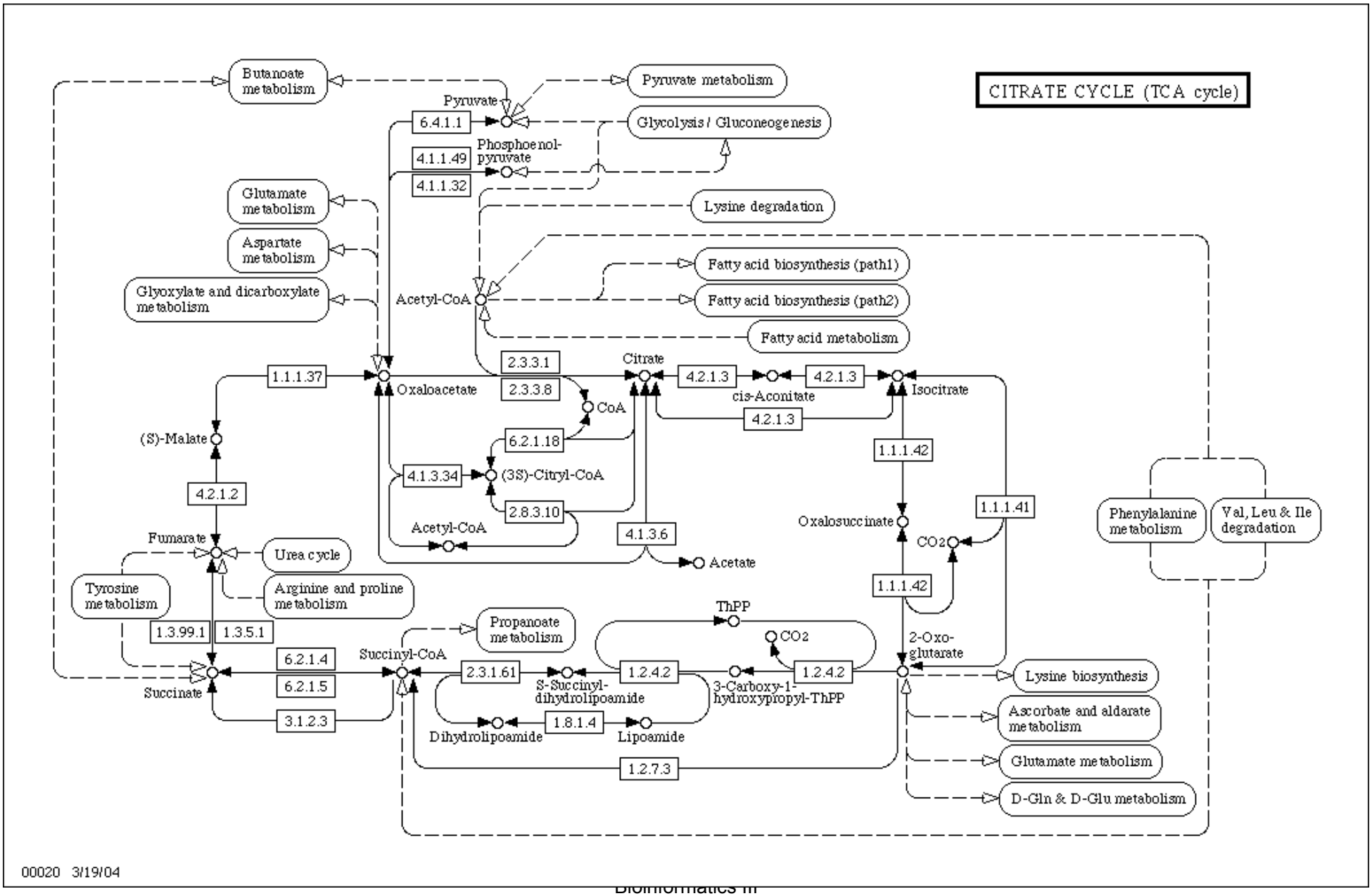
KEGG database



The KEGG PATHWAY database (http://www.genome. jp/kegg/pathway.html) is a collection of graphical diagrams (KEGG pathway maps) representing molecular interaction networks in various cellular processes. Each reference pathway is manually drawn and updated with the notation shown left.

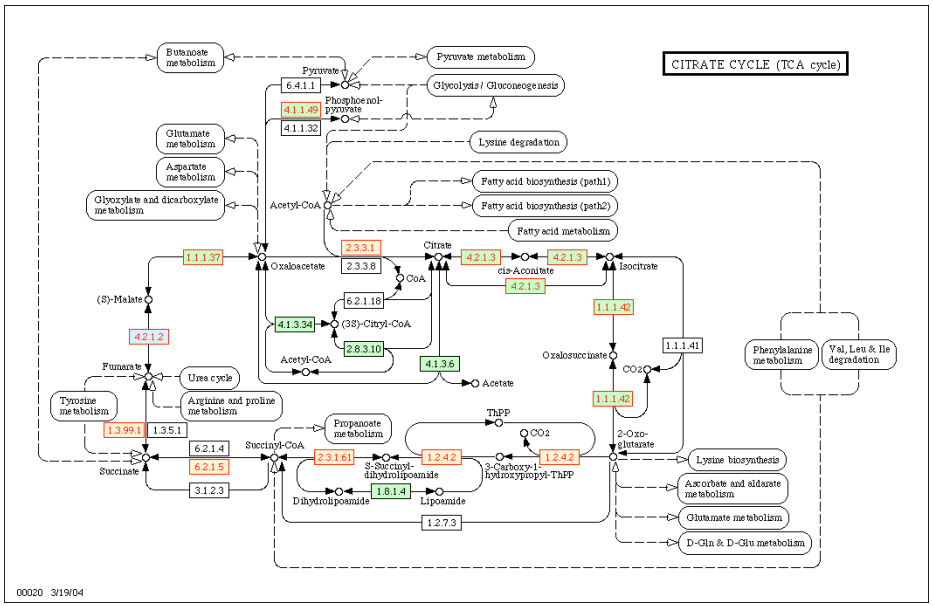
Organism-specific pathways (green-colored pathways) are computationally generated based on the KO assignment in individual genomes.

Citrate Cycle (TCA cycle) in E.coli

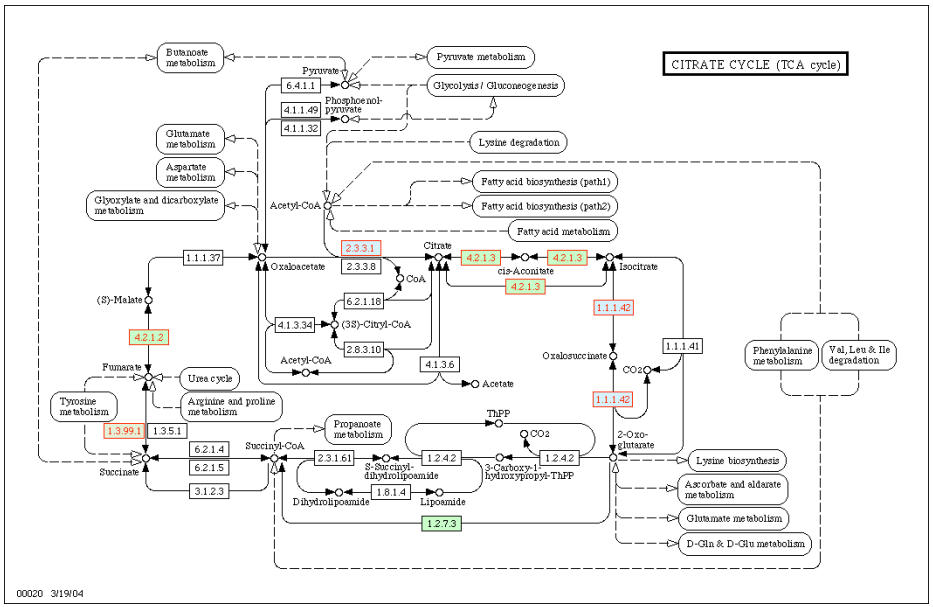


Citrate Cycle (TCA cycle) in different organisms

Citrate cycle (TCA cycle) - Escherichia coli K-12 MG1655



Citrate cycle (TCA cycle) - Helicobacter pylori 26695



Green/red: enzyme annotated in this organism

EcoCyc Database

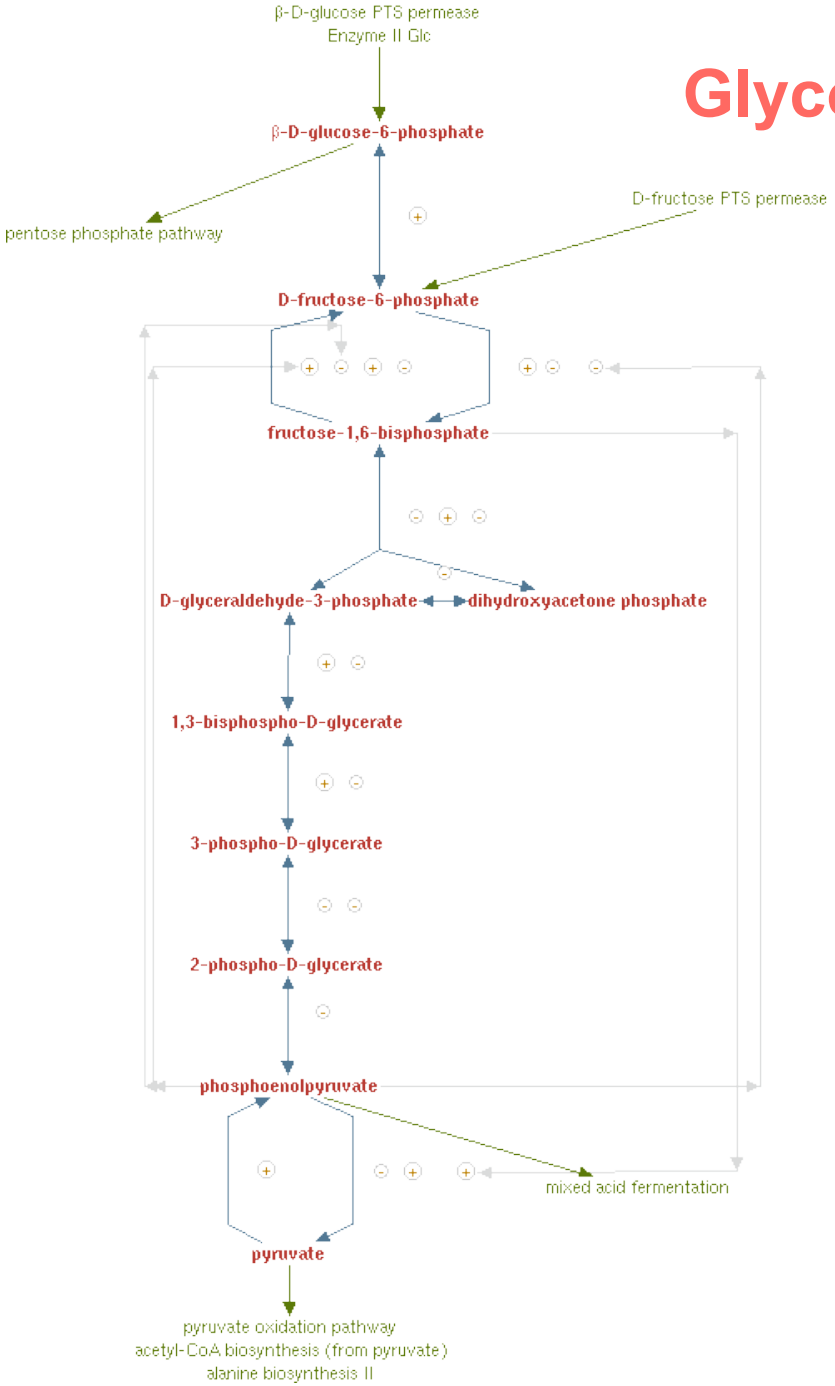
E.coli genome contains 4.7 million DNA bases.
How can we characterize the functional complement of *E.coli* and according to what criteria can we compare the biochemical networks of two organisms?

EcoCyc contains the metabolic map of *E.coli* defined as the set of all known pathways, reactions and enzymes of *E.coli* small-molecule metabolism.

Analyze
- the connectivity relationships of the metabolic network
- its partitioning into pathways
- enzyme activation and inhibition
- repetition and multiplicity of elements such as enzymes, reactions, and substrates.

Ouzonis, Karp, Genome Res. 10, 568 (2000)

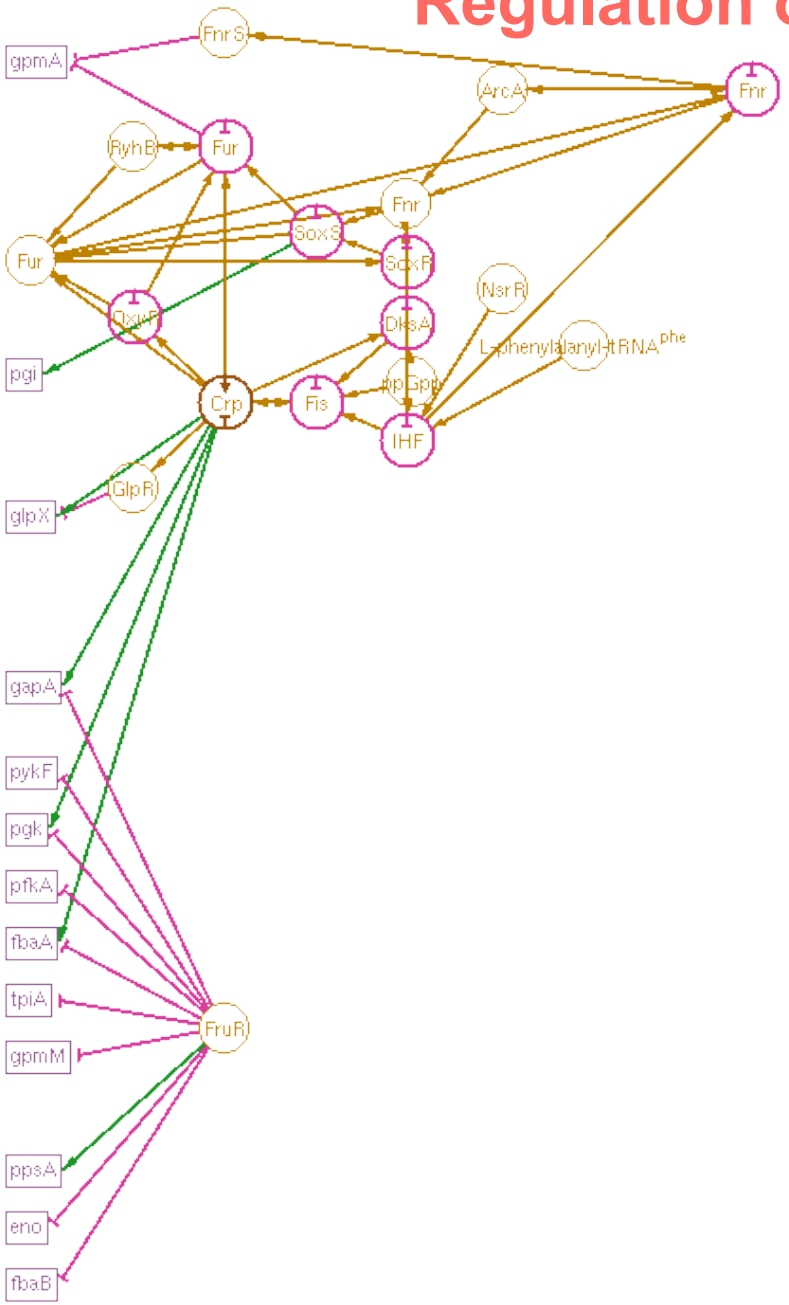
Glycolysis in E.coli



Blue arrows: biochemical reactions
clicking on arrow shows responsible enzyme

\+ and - : activation and inhibition of enzymes

Regulation of Glycolysis in E.coli



Boxed genes on the left are enzymes of glycolysis pathway

pgi: phosphoglucose isomerase
pgk: phosphoglycerate kinase
pfk: 6-phosphofructo kinase ...

Circled FruR, CRP etc. on the right : transcription factors

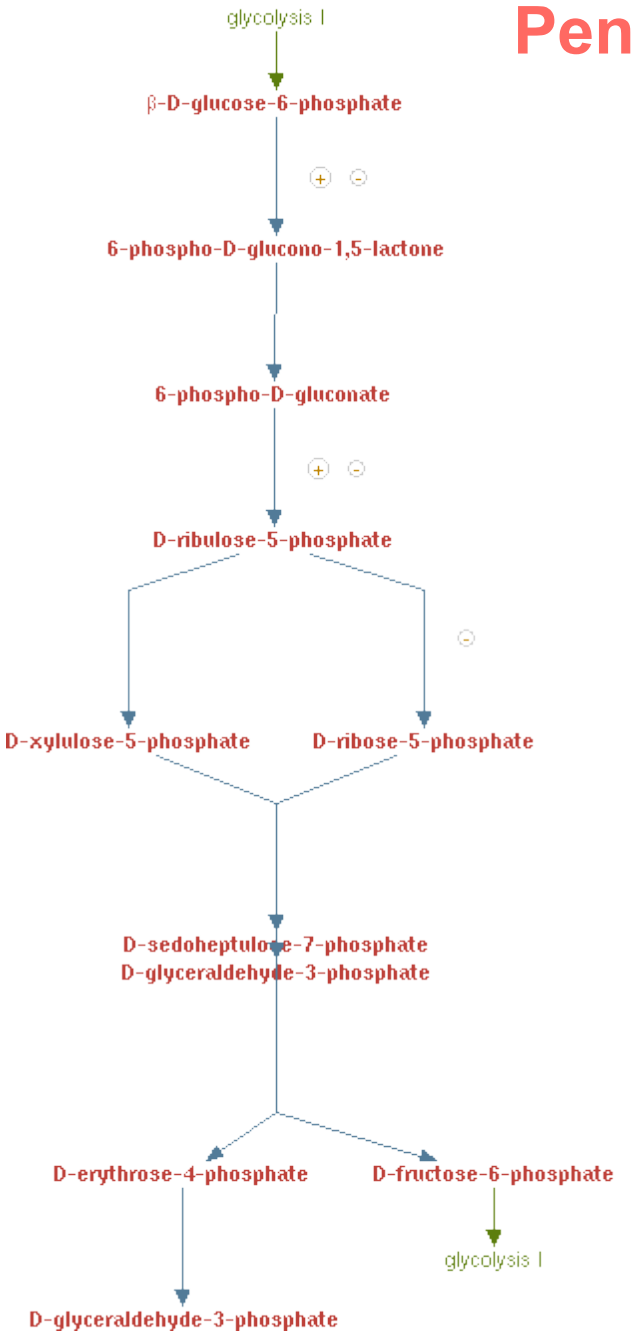
Green pointed arrows: activation of transcription;

Violet blunt arrow : repression;

Brown circle-ended arrow indicates that the factor can activate or repress, depending on circumstances.

www.ecocyc.org

Pentose Phosphate pathway

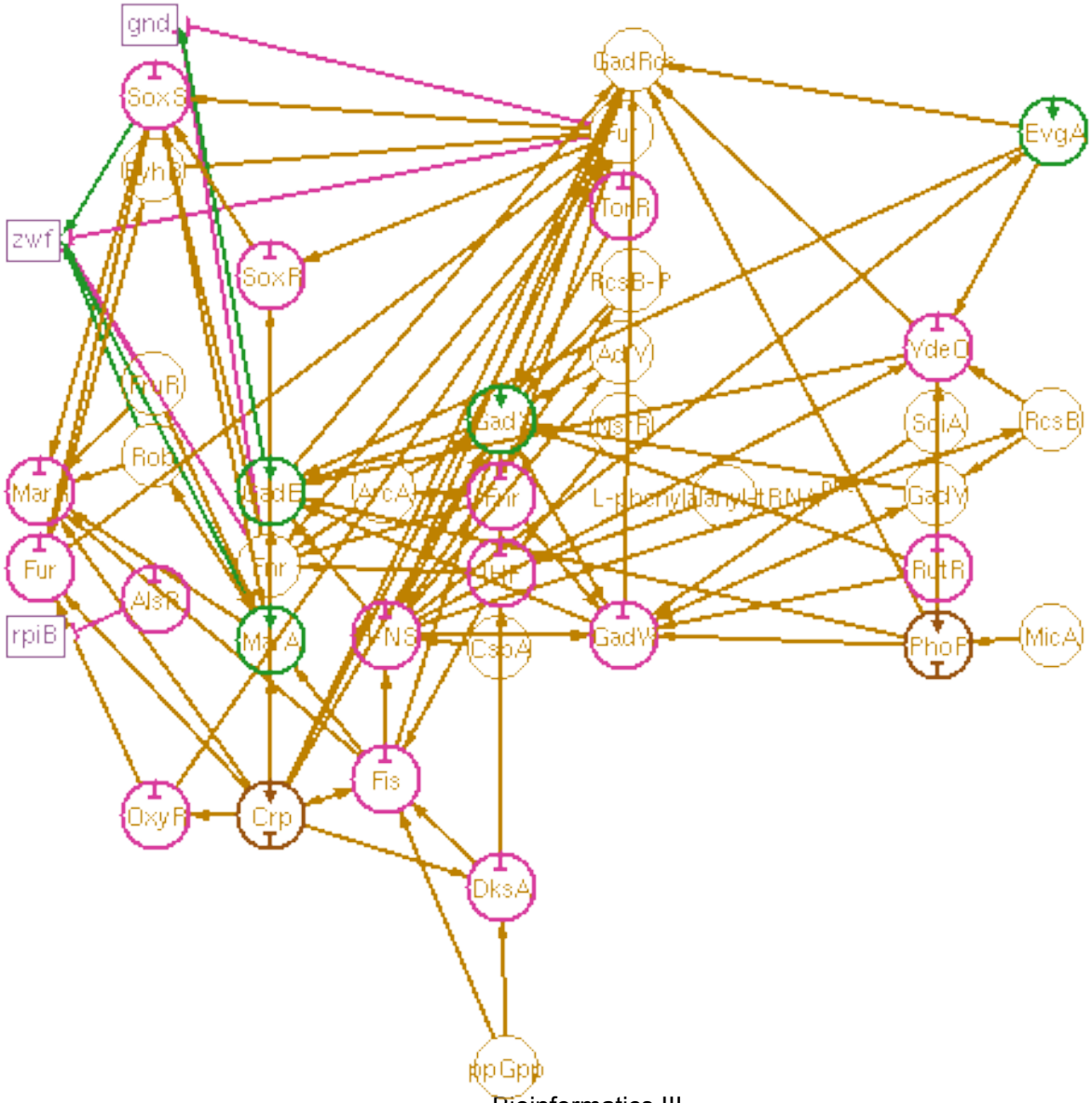


Blue arrows: biochemical reactions
clicking on arrow shows responsible enzyme

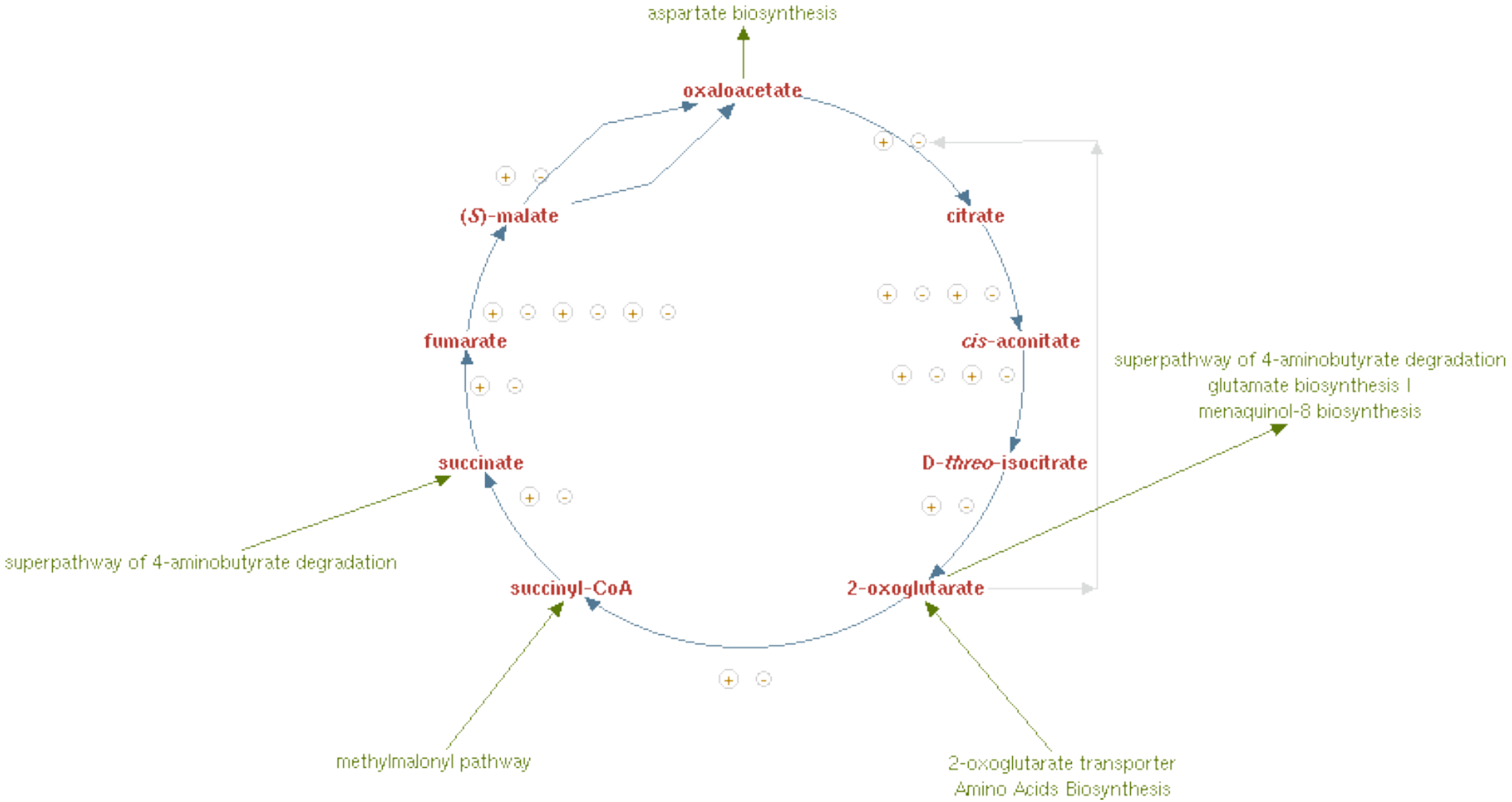
+ and - : activation and inhibition of enzymes

www.ecocyc.org

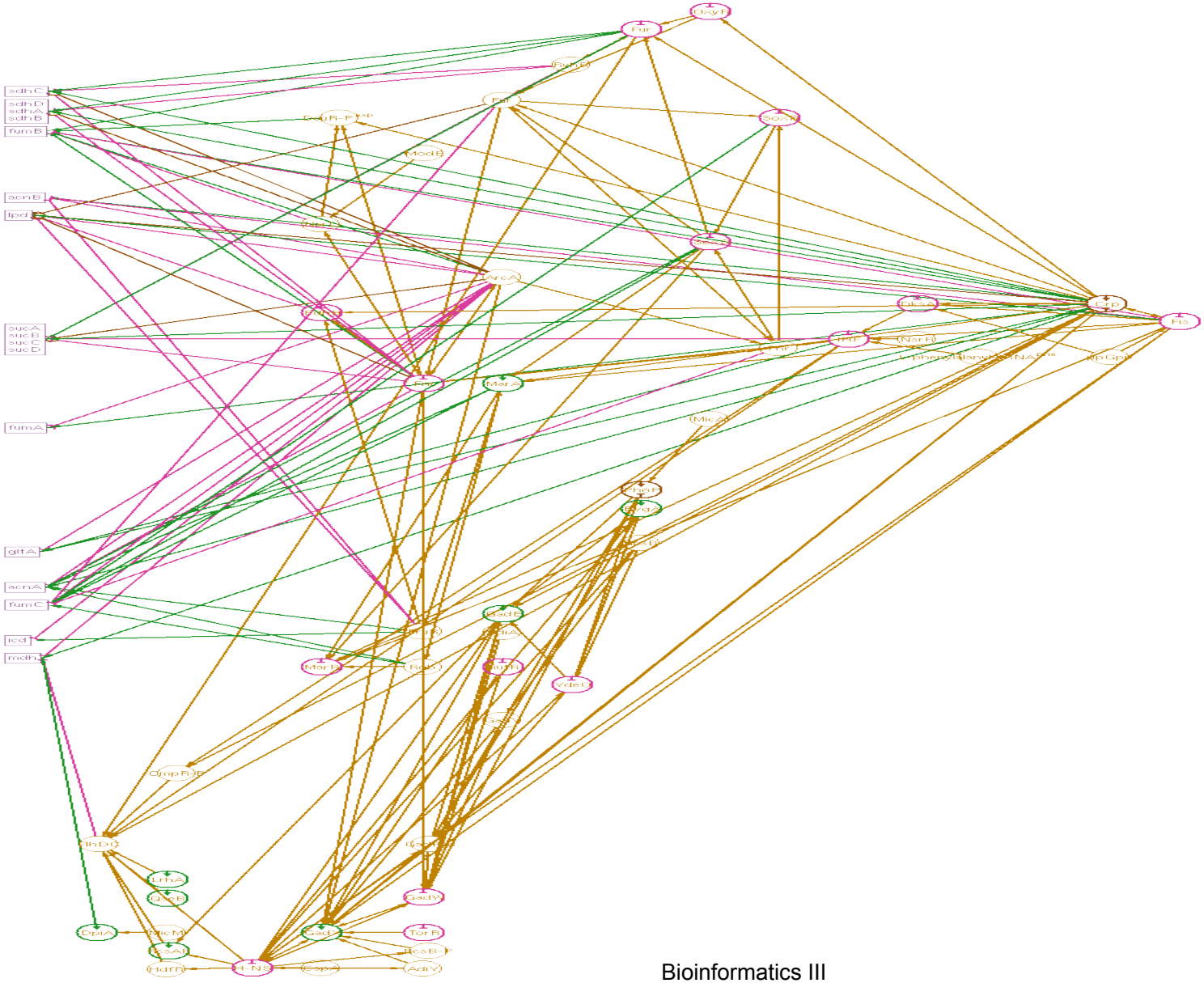
Regulation of Pentose Phosphate Pathway



TCA cycle


aspartate biosynthesis
oxaloacetate
citrate
cis-aconitate
D-threo-isocitrate
2-oxoglutarate
succinyl-CoA
succinate
fumarate
(S)-malate
superpathway of 4-aminobutyrate degradation
glutamate biosynthesis I
menaquinol-8 biosynthesis
superpathway of 4-aminobutyrate degradation
methylmalonyl pathway
2-oxoglutarate transporter
Amino Acids Biosynthesis

Regulation of TCA cycle



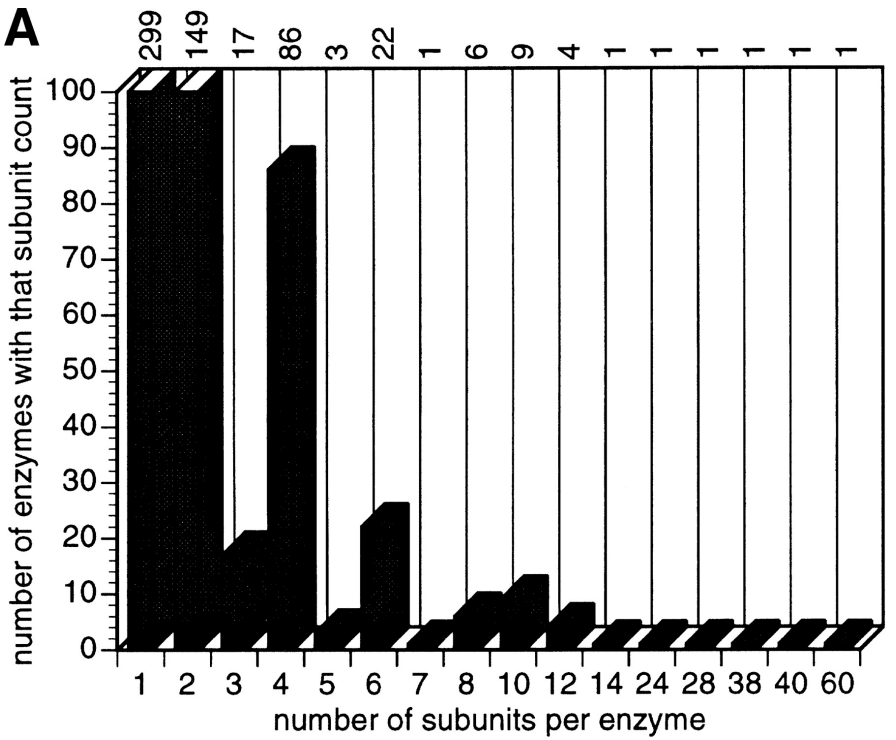
EcoCyc Analysis of *E.coli* Metabolism

In 2000, *E.coli* genome contained 4391 predicted genes, of which 4288 coded for proteins (4503 genes in Dec. 2011, 209 RNAs).

676 of these genes form 607 enzymes of the *E.coli* small-molecule metabolism.

Of those enzymes, 311 are protein complexes, 296 are monomers.

Organization of protein complexes.
Distribution of subunit counts for all EcoCyc protein complexes.
The predominance of monomers, dimers, and tetramers is obvious



Ouzonis, Karp, Genome Res. 10, 568 (2000)

Reactions

EcoCyc describes 905 metabolic reactions that are catalyzed by *E. coli.*
(1991 in Dec. 2011)

Of these reactions, 161 are not involved in small-molecule metabolism,
e.g. they participate in macromolecule metabolism such as DNA replication and tRNA charging.

Of the remaining 744 reactions, 569 have been assigned to at least one pathway.

Ouzonis, Karp, Genome Res. 10, 568 (2000)

Reactions

The number of reactions (744) and the number of enzymes (607) differ ...
WHY??

(1) there is no one-to-one mapping between enzymes and reactions –
some enzymes catalyze multiple reactions, and some reactions are catalyzed by multiple enzymes.

(2) for some reactions known to be catalyzed by *E.coli*, the enzyme has not yet been identified.

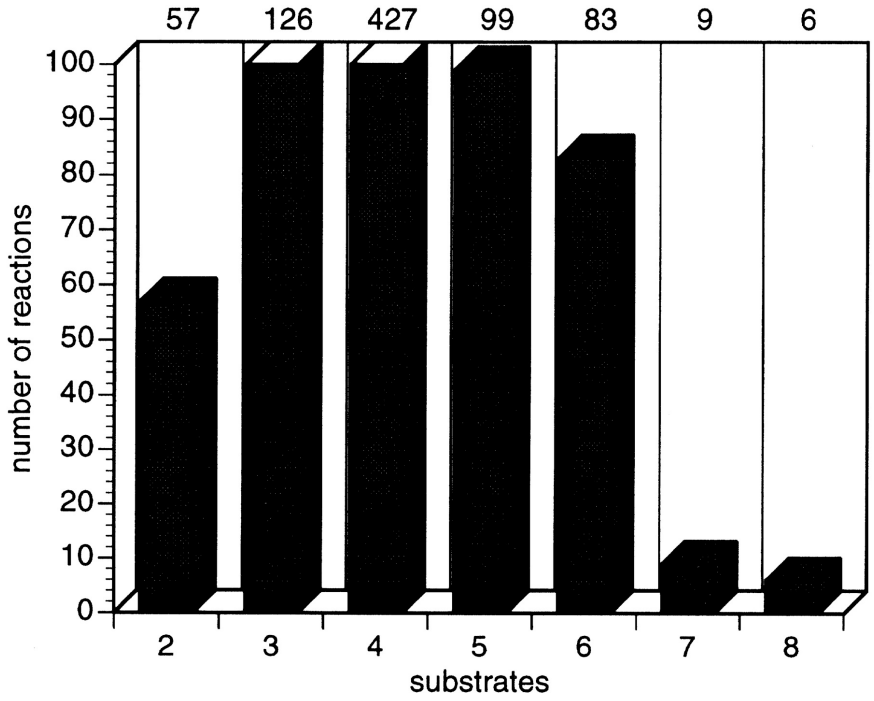
Ouzonis, Karp, Genome Res. 10, 568 (2000)

Compounds

The 744 reactions of *E.coli* small-molecule metabolism involve a total of 791 different substrates.

On average, each reaction contains 4.0 substrates, (think of A + B <-> C + D)

Number of reactions containing varying numbers of substrates (reactants plus products).



Ouzonis, Karp, Genome Res. 10, 568 (2000)

Compounds

Each distinct substrate occurs in an average of 2.1 reactions.

Table 1. Most Frequently Used Metabolites in *E. coli* Central Metabolism

Occurrence	Name of metabolite
205	H_2O
152	ATP
101	ADP
100	phosphate
89	pyrophosphate
66	NAD
60	NADH
54	CO_2
53	H^+
49	AMP
48	NH_3
48	NADP
45	NADPH
44	Coenzyme A
43	L-glutamate
41	pyruvate
29	acetyl-CoA
26	O_2
24	2-oxoglutarate
23	S-adenosyl-L-methionine
18	S-adenosyl-homocysteine
16	L-aspartate
16	L-glutamine
15	H_2O_2
14	glucose
13	glyceraldehyde-3-phosphate
13	THF
13	acetate
12	PRPP
12	[acyl carrier protein]
12	oxaloacetic acid
11	dihydroxy-acetone-phosphate
11	GDP
11	glucose-1-phosphate
11	UMP
10	e^-
10	phosphoenolpyruvate
10	acceptor
10	reduced acceptor
10	GTP
10	L-serine
10	fructose-6-phosphate
9	L-cysteine
9	reduced thioredoxin
9	oxidized thioredoxin
9	reduced glutathione
8	acyl-ACP
8	L-glycine
8	GMP
8	formate

Metabolites were used either as reactants or products.

Ouzonis, Karp, Genome Res. 10, 568 (2000)

11	glucose-1-phosphate

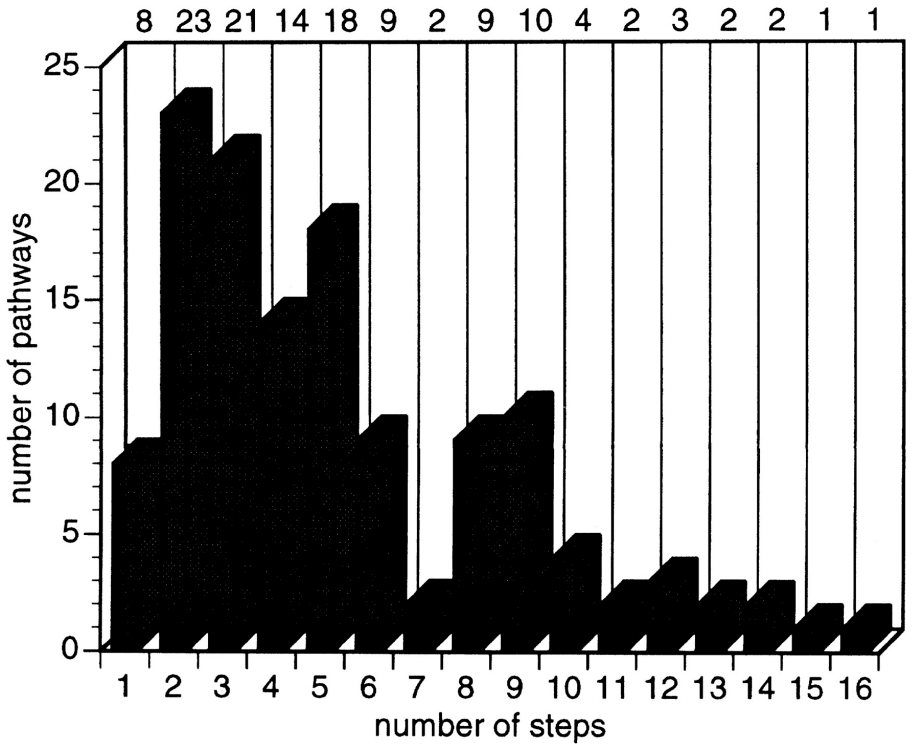
Pathways

EcoCyc describes 131 pathways (347 in Dec. 2011):

- energy metabolism
- nucleotide and amino acid biosynthesis
- secondary metabolism

Pathways vary in length from a single reaction step to 16 steps with an average of 5.4 steps.

Length distribution of EcoCyc pathways



Ouzonis, Karp, Genome Res. 10, 568 (2000)

Pathways

However, there is no precise biological definition of a pathway.

The partitioning of the metabolic network into pathways (including the well-known examples of biochemical pathways) is somehow arbitrary.

These decisions of course also affect the distribution of pathway lengths.

Table 2. List of All Known *E. coli* Metabolic Pathways as Described by EcoCyc

(Deoxy)ribose phosphate metabolism	Isoleucine biosynthesis
3-Phenylpropionate and 3-(3-hydroxyphenyl)propionate degradation	KDO biosynthesis
4-Aminobutyrate degradation	L-alanine degradation
Aerobic electron transfer	L-arabinose catabolism
Aerobic respiration, electron donors reaction list	L-cysteine catabolism
Alanine biosynthesis	L-lyxose metabolism
Anaerobic electron transfer	L-serine degradation
Anaerobic respiration	Lactose degradation
Anaerobic respiration, electron acceptors reaction list	Leucine biosynthesis
Anaerobic respiration, electron donors reaction list	Lipid A precursor biosynthesis
Arginine biosynthesis	Lysine and diaminopimelate biosynthesis
Asparagine biosynthesis and degradation	Mannitol degradation
Aspartate biosynthesis and degradation	Mannose and GDP-mannose metabolism
Betaine biosynthesis	Mannose catabolism
Biosynthesis of proto- and siroheme	Menaquinone biosynthesis
Biotin biosynthesis	Methionine biosynthesis
Carnitine metabolism	Methyl-donor molecule biosynthesis
Carnitine metabolism, CoA-linked	Methylglyoxal metabolism
Cobalamin biosynthesis	NAD phosphorylation and dephosphorylation
Colanic acid biosynthesis	Noncxidative branch of the pentose phosphate pathway
Cyanate catabolism	Nucleotide metabolism
Cysteine biosynthesis	O-antigen biosynthesis
D-arabinose catabolism	Oxidative branch of the pentose phosphate pathway
D-galactarate catabolism	Pantothenate and coenzyme A biosynthesis
D-galacturonate catabolism	Peptidoglycan biosynthesis
D-glucarate catabolism	Phenylalanine biosynthesis
D-glucuronate catabolism	Phenylethylamine degradation
Degradation of short-chain fatty acids	Phosphatidic acid synthesis
Deoxypyrimidine nucleotide/side metabolism	Phospholipid biosynthesis
Deoxyribonucleotide metabolism	Polyamine biosynthesis
dTDP-rhamnose biosynthesis	Polyisoprenoid biosynthesis
Enterobacterial common antigen biosynthesis	ppGpp metabolism
Enterobactin synthesis	Proline biosynthesis
Entner–Doudoroff pathway	Proline utilization
Fatty acid biosynthesis, initial steps	Propionate metabolism, methylmalonyl pathway
Fatty acid elongation, saturated	Purine biosynthesis
Fatty acid elongation, unsaturated	Pyridine nucleotide cycling
Fatty acid oxidation pathway	Pyridine nucleotide synthesis
Fermentation	Pyridoxal 5'-phosphate biosynthesis
Folic acid biosynthesis	Pyridoxal 5'-phosphate salvage pathway
FormylTHF biosynthesis	Pyrimidine biosynthesis
Fucose catabolism	Pyrimidine ribonucleotide/ribonucleoside metabolism
Galactitol catabolism	Pyruvate dehydrogenase
Galactonate catabolism	Pyruvate oxidation pathway
Galactose metabolism	Removal of superoxide radicals
Galactose, galactoside and glucose catabolism	Rhamnose catabolism
Gluconeogenesis	Riboflavin, FMN and FAD biosynthesis
Glucosamine catabolism	Ribose catabolism
Glucose 1-phosphate metabolism	Serine biosynthesis
Glutamate biosynthesis	Sorbitol degradation
Glutamate utilization	Sulfate assimilation pathway
Glutamine biosynthesis	TCA cycle, aerobic respiration
Glutamine utilization	Thiamine biosynthesis
Glutathione biosynthesis	Thioredoxin pathway
Glutathione-glutaredoxin redox reactions	Threonine biosynthesis
Glycerol metabolism	Threonine catabolism
Glycine biosynthesis	Trehalose biosynthesis
Glycine cleavage	Trehalose degradation, low osmolarity
Glycogen biosynthesis	Tryptophan biosynthesis
Glycogen catabolism	Tryptophan utilization
Glycolate metabolism	Tyrosine biosynthesis
Glycolysis	Ubiquinone biosynthesis
Glyoxylate cycle	UDP-*N*-acetylglucosamine biosynthesis
Glyoxylate degradation	Valine biosynthesis
Histidine biosynthesis	Xylose catabolism
Histidine degradation	

The reactions and enzymes within each pathway can be determined using the EcoCyc WWW server that is available at http://ecocyc.DoubleTwist.com/ecocyc/.

Ouzonis, Karp, Genome Res. 10, 568 (2000)

Enzyme Modulation

An `enzymatic reaction` is a type of EcoCyc object that represents the pairing of an enzyme with a reaction catalyzed by that enzyme.

EcoCyc contains extensive information on the modulation of *E.coli* enzymes with respect to particular reactions:
- activators and inhibitors of the enzyme,
- cofactors required by the enzyme
- alternative substrates that the enzyme will accept.

Of the 805 enzymatic-reaction objects within EcoCyc, physiologically relevant activators are known for 22, physiologically relevant inhibitors are known for 80.

327 (almost half) require a cofactor or prosthetic group.

Ouzonis, Karp, Genome Res. 10, 568 (2000)

Enzyme Modulation

Table 3. Most Common Modulators, cofactors, and prosthetic groups of *E. coli* enzymes and Their Frequencies

A. Modulators (activators and inhibitors)

Occurrence	Name of modulator	Activator	Inhibitor
35	Cu^{2+}		•
32	ATP	•	•
30	Zn^{2+}	•	•
29	AMP	•	•
26	ADP	•	•
25	EDTA	•	•
23	*p*-chloromercuribenzoate		•
23	pyrophosphate	•	•
22	K^{+}	•	•
22	phosphate	•	•
20	Hg^{2+}		•
20	Ca^{2+}	•	•
19	*N*-ethylmaleimide	•	•
16	NAD	•	•
16	iodoacetamide		•
16	coenzyme A		•
15	Co^{2+}	•	•
15	Mg^{2+}	•	•
15	phosphoenolpyruvate	•	•
14	Fe^{2+}	•	•
14	GTP	•	•
14	pyruvate	•	•
13	*p*-hydroxymercuribenzoate		•
13	NADP		•
12	Mn^{2+}	•	•

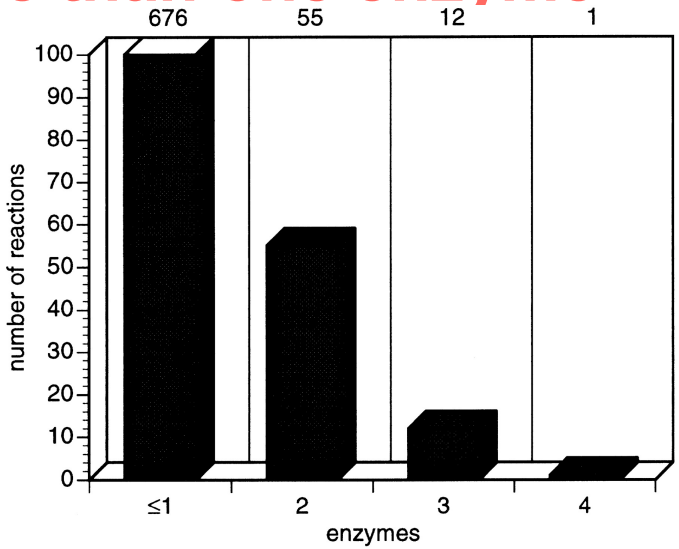
B. Cofactors and prosthetic groups

Occurrence	Name of compound	Cofactor	Prosthetic group
145	Mg^{2+}	•	•
48	pyridoxal 5'-phosphate	•	•
33	Mn^{2+}	•	
31	FAD	•	•
21	Fe^{2+}	•	•
18	Zn^{2+}	•	•
16	thiamine-pyrophosphate		•
11	FMN	•	•
10	Co^{2+}	•	
9	K^{+}	•	
6	Mo^{2+}		•
5	NAD	•	•
4	protoheme		•
4	Ni^{2+}	•	•
4	Ca^{2+}	•	
4	4Fe-4S center		•
3	NH_4^{+}	•	
3	pyruvate		•
3	siroheme		•
3	cytochrome *c*		•
2	heme C		•
2	B_{12}		•
2	NADP	•	
2	Cu^{2+}		•
2	biotin		•
2	Cd^{2+}	•	

Ouzonis, Karp, Genome Res. 10, 568 (2000)

Reactions catalyzed by more than one enzyme

Diagram showing the **number of reactions** that are **catalyzed** by **one or more enzymes**. Most reactions are catalyzed by one enzyme, some by two, and very few by more than two enzymes.



For 84 reactions, the corresponding enzyme is not yet encoded in EcoCyc.

What may be the reasons for isozyme redundancy?

(1) the enzymes that catalyze the same reaction are **paralogs** (homologs) and have duplicated (or were obtained by horizontal gene transfer), acquiring some specificity but retaining the same mechanism (**divergence**)

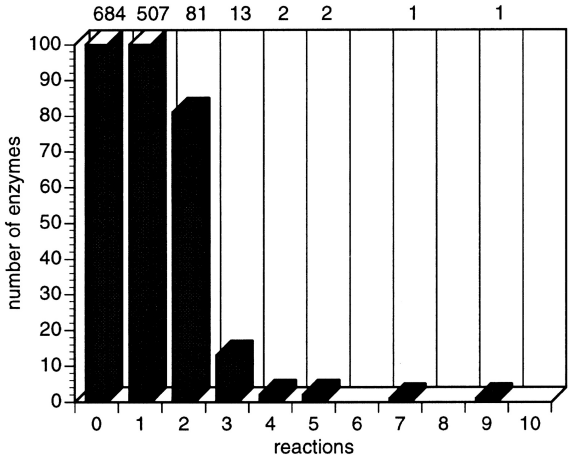
(2) the reaction is easily „invented“; therefore, there is more than one protein family that is independently able to perform the catalysis (**convergence**).

Ouzonis, Karp, Genome Res. 10, 568 (2000)

Enzymes that catalyze more than one reaction

Genome predictions usually assign a single enzymatic function.
However, *E.coli* is known to contain many multifunctional enzymes.
Of the 607 *E.coli* enzymes, 100 are multifunctional, either having the same active site and different substrate specificities or different active sites.

Number of enzymes that catalyze one or more reactions. Most enzymes catalyze one reaction; some are multifunctional.

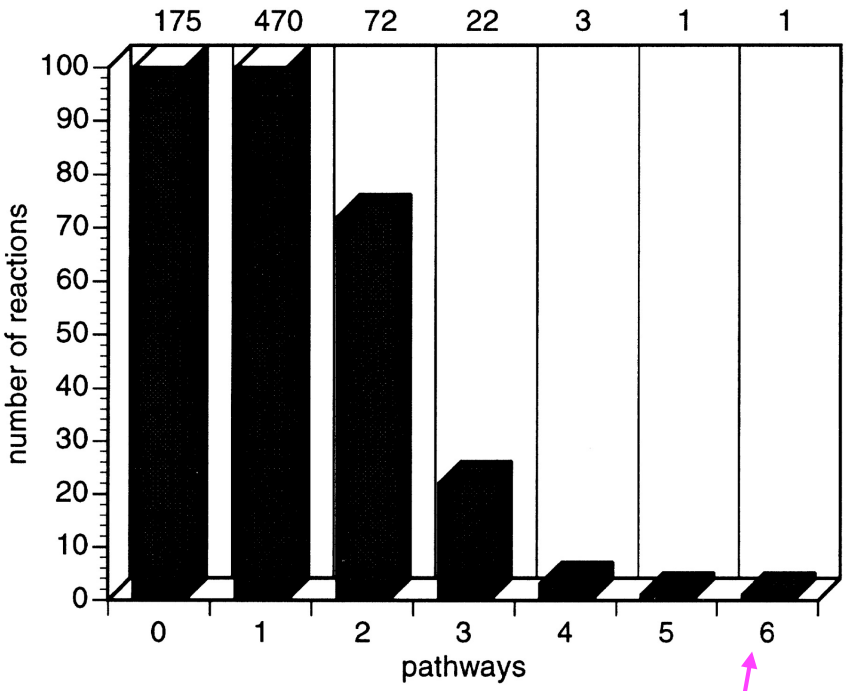


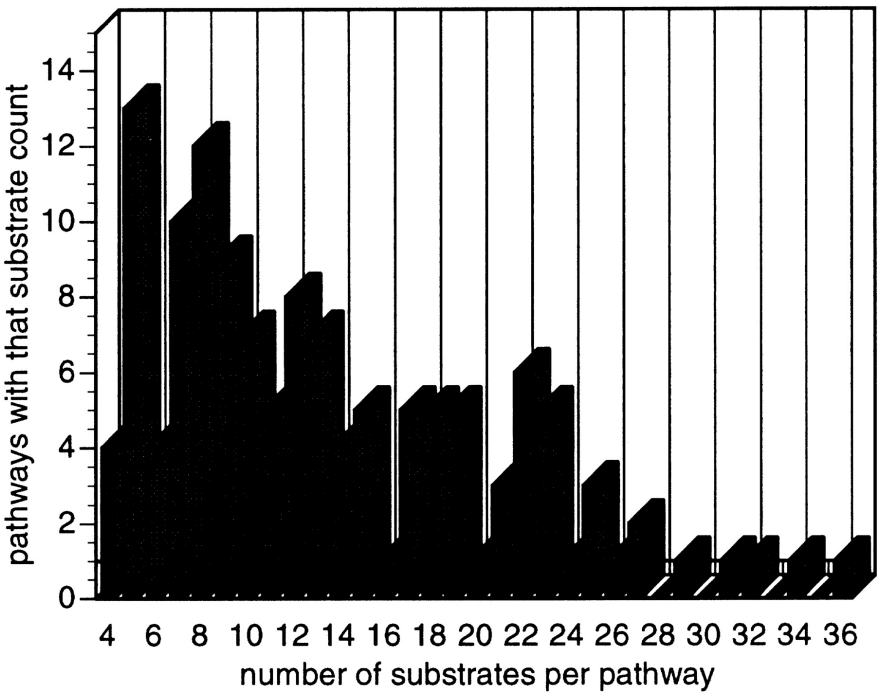
The enzymes that catalyze 7 and 9 reactions are purine nucleoside phosphorylase and nucleoside diphosphate kinase.

The high proportion of multifunctional enzymes implies that the genome projects may significantly underpredict multifunctional enzymes!

Ouzonis, Karp, Genome Res. 10, 568 (2000)

Reactions participating in more than one pathway





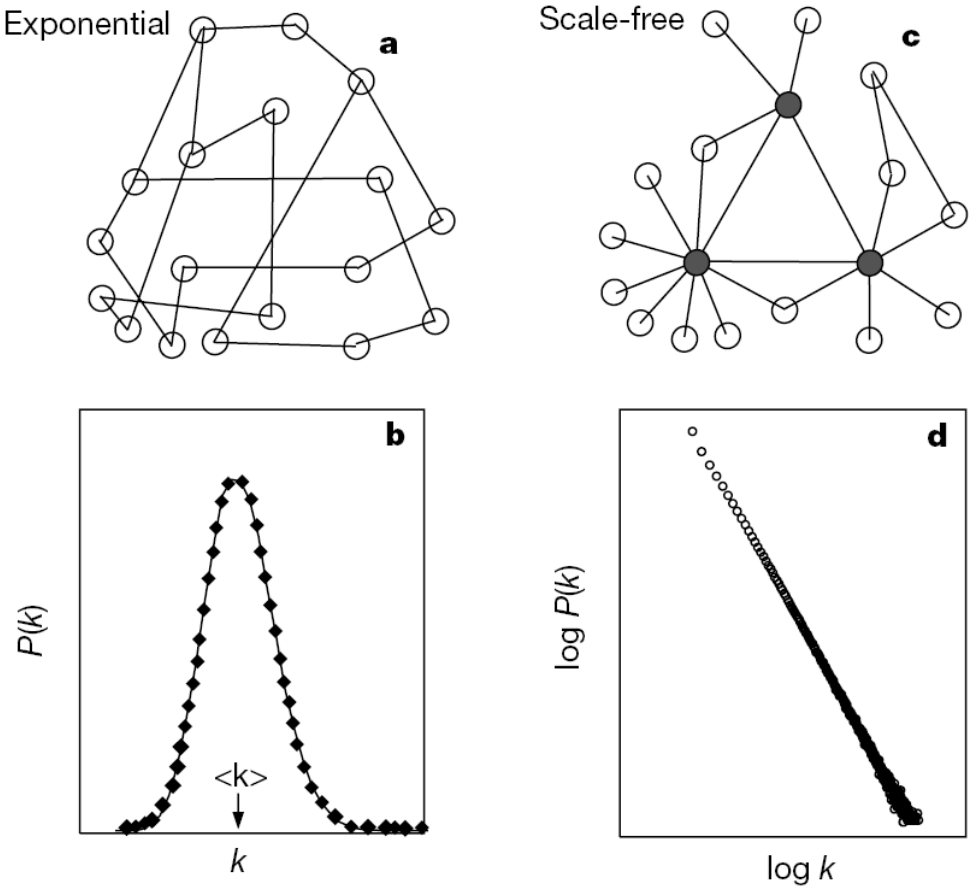
The 99 reactions belonging to multiple pathways appear to be the **intersection points** in the complex network of chemical processes in the cell.

Ouzonis, Karp,
Genome Res. 10, 568 (2000)

E.g. the reaction present in 6 pathways corresponds to the reaction catalyzed by malate dehydrogenase, a central enzyme in cellular metabolism.

Large-scale structure: Metabolic networks are scale-free ☺

Attributes of generic network structures. **a**, Representative structure of the network generated by the Erdös–Rényi network model. **b**, The network connectivity can be characterized by the probability, $P(k)$, that a node has k links. For a random network $P(k)$ peaks strongly at $k = \langle k \rangle$ and decays exponentially for large k (i.e., $P(k) \approx e^{-k}$ for $k >> \langle k \rangle$ and $k << \langle k \rangle$). **c**, In the scale-free network most nodes have only a few links, but a few nodes, called hubs (dark), have a very large number of links. **d**, $P(k)$ for a scale-free network has no well-defined peak, and for large k it decays as a power-law, $P(k) \approx k^{-\gamma}$, appearing as a straight line with slope - on a log–log plot.



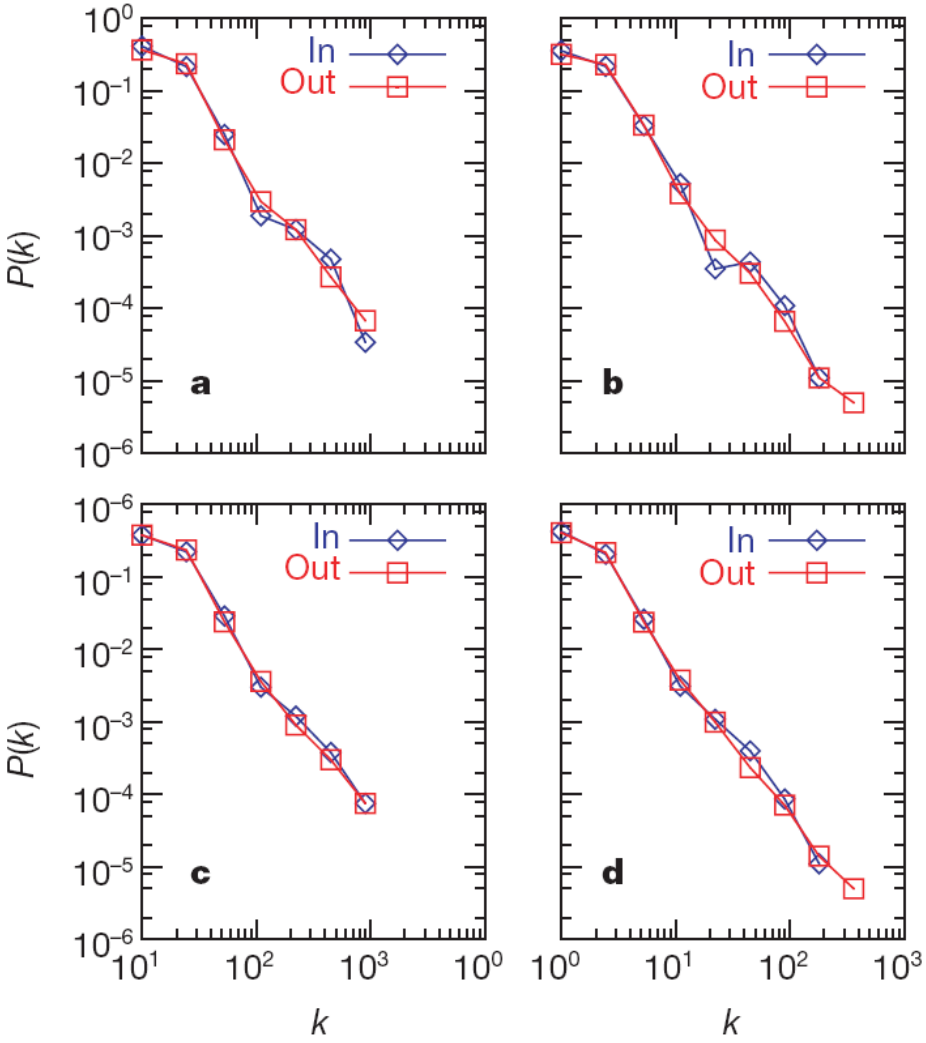
Jeong et al. Nature 407, 651 (2000)

Connectivity distributions *P(k)* for substrates

a, *Archaeoglobus fulgidus* (archae);
b, *E. coli* (bacterium);
c, *Caenorhabditis elegans* (eukaryote)
d, The connectivity distribution averaged over 43 organisms.

x-axis: metabolites participating in *k* reactions
y-axis (*P(k)*): number/frequency of such metabolites

log–log plot, counts separately the incoming (In) and outgoing links (Out) for each substrate. k_{in} (k_{out}) corresponds to the number of reactions in which a substrate participates as a product (educt).



Jeong et al. Nature 407, 651 (2000)

Properties of metabolic networks

a, The histogram of the biochemical pathway lengths, *l*, in *E. coli*.

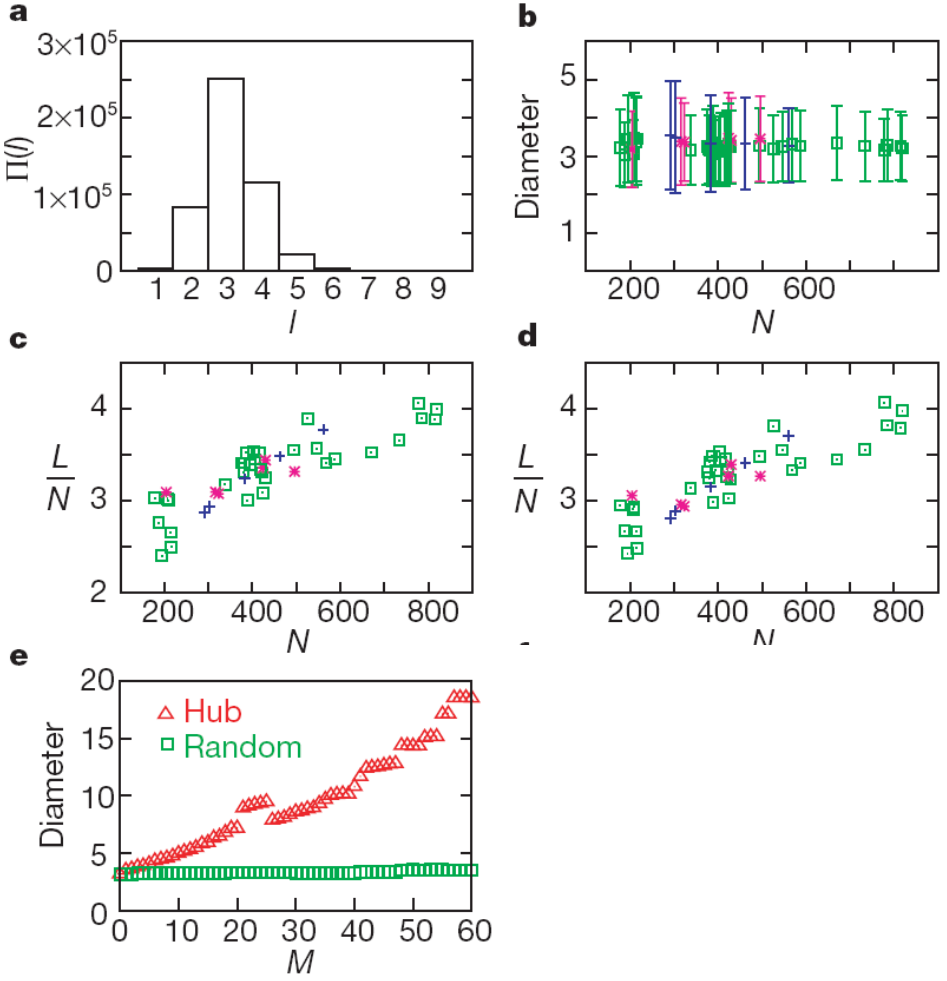
b, The average path length (diameter) for each of the 43 organisms.

N : number of metabolites in each organism

c, **d**, Average number of incoming links (**c**) or outgoing links (**d**) per node for each organism.

e, The effect of substrate removal on the metabolic network diameter of *E. coli*.

In the top curve (red) the most connected substrates are removed first. In the bottom curve (green) nodes are removed randomly. *M* = 60 corresponds to 8% of the total number of substrates in found in *E. coli*.



b–**d**, Archaea (magenta), bacteria (green) and eukaryotes (blue) are shown.

Jeong et al. Nature 407, 651 (2000)

Interpretation of metabolic network connectivity

Despite significant variations in their individual constituents and pathways,

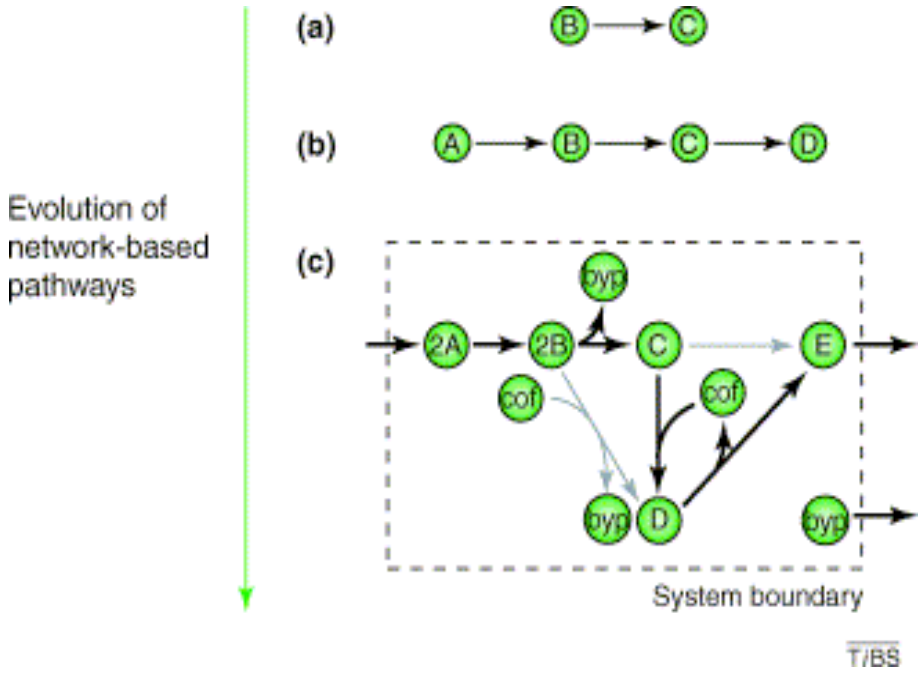
the metabolic networks of 43 organisms representing all 3 domains of life

have the same topological scaling properties and show striking similarities

to the inherent organization of complex non-biological systems.

Jeong et al. Nature 407, 651 (2000)

Development of the network-based pathway paradigm



(a) With advanced biochemical techniques, years of research have led to the precise characterization of individual reactions. As a result, the **complete stoichiometries** of many metabolic reactions have been characterized.
(b) Most of these reactions have been grouped into `**traditional pathways**' (e.g. glycolysis) that do not account for cofactors and byproducts in a way that lends itself to a mathematical description. However, with sequenced and annotated genomes, models can be made that account for many metabolic reactions in an organism.

(c) Subsequently, **network-based**, mathematically defined **pathways** can be analyzed that account for a complete network (black and gray arrows correspond to active and inactive reactions).

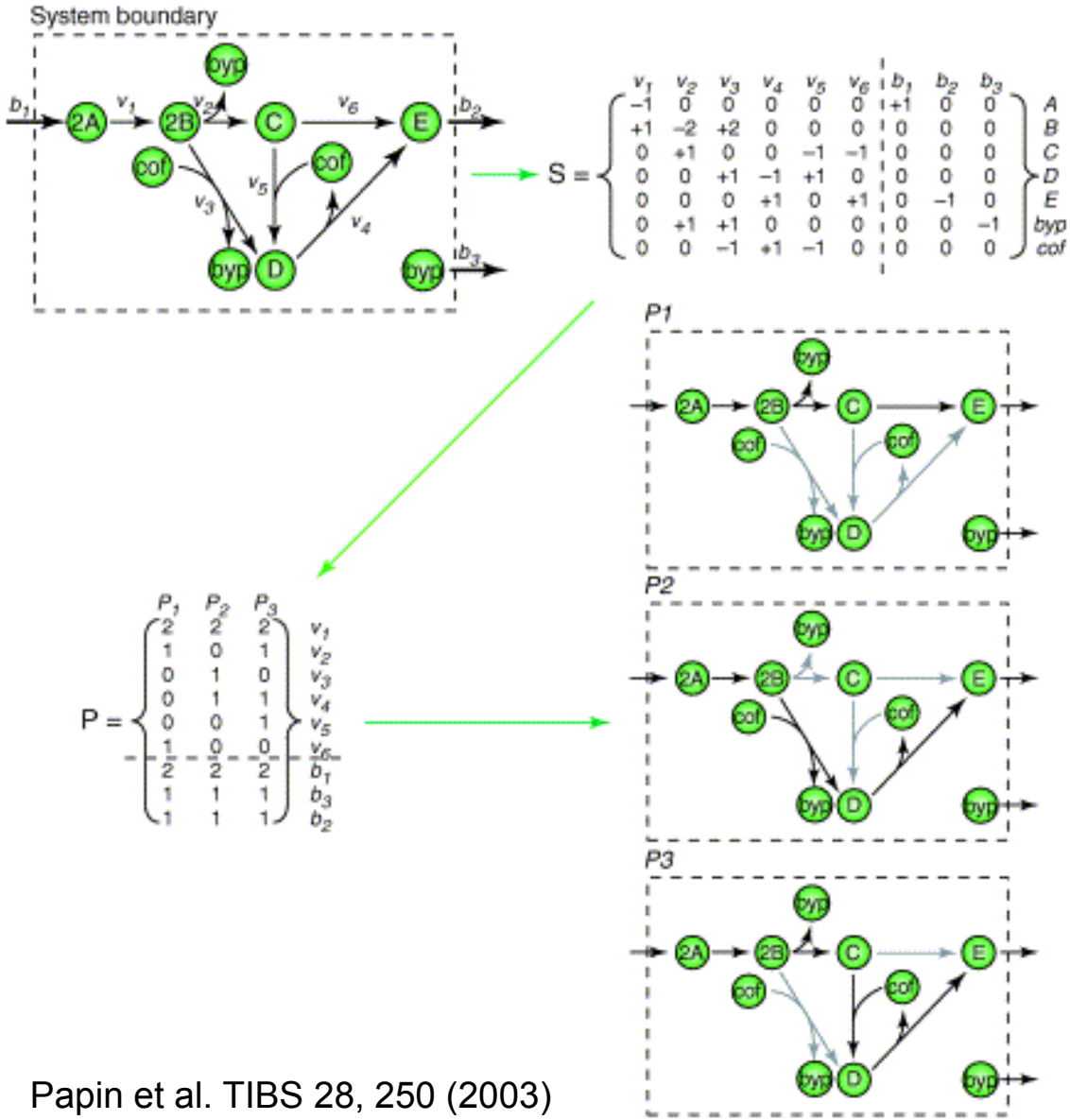
Papin et al. TIBS 28, 250 (2003)

Stoichiometric matrix - Flux Balance Analysis

Stoichiometric matrix S:
m × *n* matrix with stochiometries of the *n* reactions as columns and participations of *m* metabolites as rows.

The stochiometric matrix is an important part of the *in silico* model.

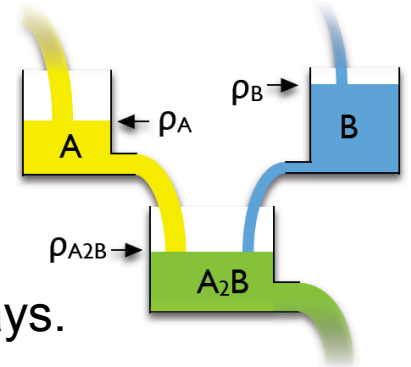
With the matrix, the methods of extreme pathway and elementary mode analyses can be used to generate a unique set of pathways P1, P2, and P3 that allow to express all steady-state fluxes as linear combinations of P1 – P3.



Papin et al. TIBS 28, 250 (2003)

Flux balancing

Any chemical reaction requires **mass conservation**.
Therefore one may analyze metabolic systems
by requiring mass conservation.
Only required: knowledge about stoichiometry of metabolic pathways.



For each metabolite X_i :

$$dX_i/dt = V_{synthesized} - V_{used} + V_{transported\_in} - V_{transported\_out}$$

Steady state: concentrations are constant
=> flux in = flux out

$$\frac{dA_2B(t)}{dt} = G_{A_2B} - L_{A_2B} = 0$$

Flux balancing

Under **steady-state conditions**, the mass balance constraints in a metabolic network can be represented mathematically by the matrix equation:

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

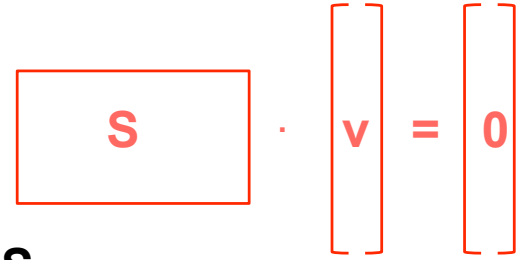
where the matrix **S** is the stoichiometric matrix and the vector *v* represents all fluxes in the metabolic network, including the internal fluxes, transport fluxes and the growth flux.

Flux balance analysis

Since the number of metabolites is generally smaller than the number of reactions ($m < n$) the flux-balance equation is typically **underdetermined**.

Therefore there are generally multiple feasible flux distributions that satisfy the mass balance constraints.

The set of solutions are confined to the **nullspace** of matrix **S**.



Null space: space of feasible solutions

Consider

$$\begin{pmatrix} 0 & 2 & 1 \\ 3 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Corresponds to

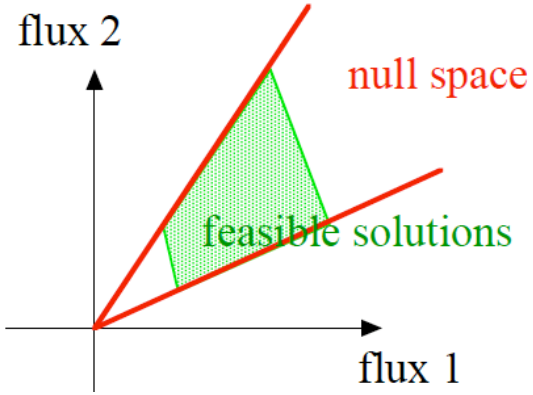
$$\begin{aligned} 2x_2 + x_3 &= 0 \\ 3x_1 - x_2 + x_3 &= 0 \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} 2x_2 &= -x_3 \\ 2x_1 &= -x_3 \end{aligned}$$

=> only one free parameter: x_3

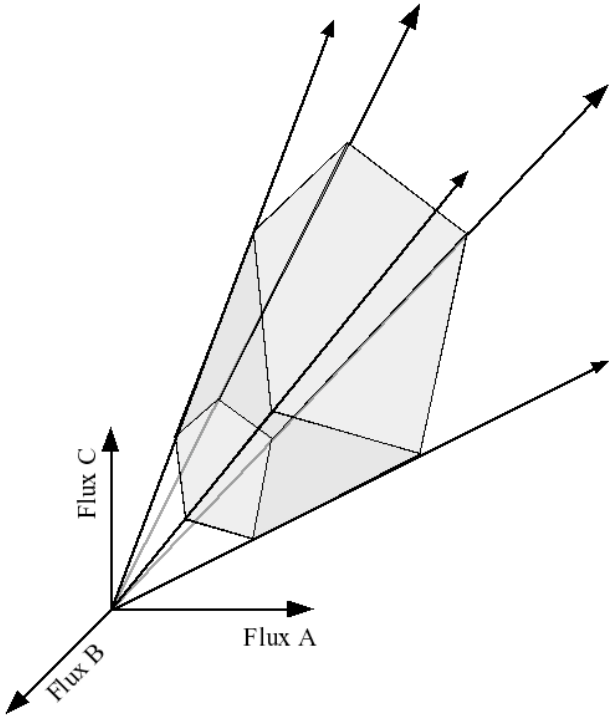
null space: $\vec{x} = \begin{pmatrix} -a \\ -a \\ 2a \end{pmatrix}$

Add inequalities for external fluxes
(here, e.g.: $x_3 \geq 0$)
=> **feasible** solutions for $a \geq 0$

Generally: null space is a cone,
constraints select part of it



Feasible solution set for a metabolic reaction network



The steady-state operation of the metabolic network is restricted to the region within a **pointed cone**, defined as the feasible set.

The feasible set contains all flux vectors that satisfy the physicochemical constrains.

Thus, the feasible set defines the capabilities of the metabolic network. All feasible metabolic flux distributions lie within the feasible set.

Edwards & Palsson PNAS 97, 5528 (2000)

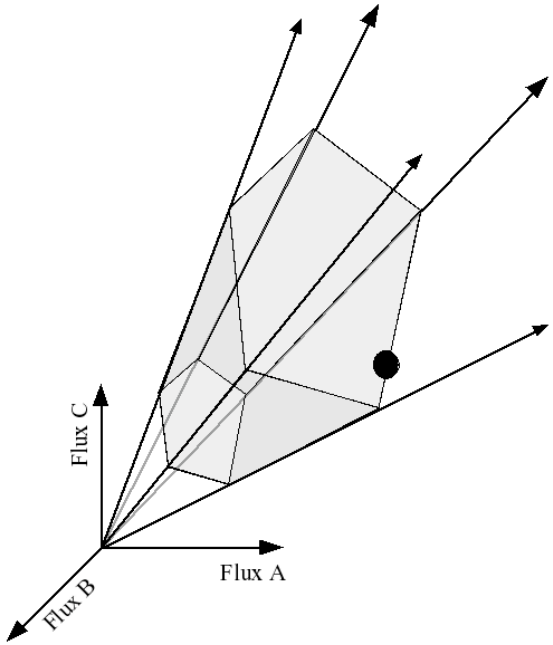
True biological flux

To find the „true“ biological flux in cells (→ e.g. Heinzle, UdS) one needs additional (experimental) information,
or one may impose constraints

$$\alpha_i \leq v_i \leq \beta_i$$

on the magnitude of each individual metabolic flux.

The intersection of the nullspace and the region defined by those linear inequalities defines a region in flux space = the **feasible set of fluxes**.



In the limiting case, where all constraints on the metabolic network are known, such as the enzyme kinetics and gene regulation, the feasible set may be reduced to a single point. This single point must lie within the feasible set.

E.coli in silico

Best studied cellular system: *E. coli*.

In 2000, Edwards & Palsson constructed an *in silico* representation of *E.coli* metabolism.

There were 2 good reasons for this:

(1) genome of *E.coli* MG1655 was already completely sequenced,

(2) Because of long history of *E.coli* research, biochemical literature, genomic information, metabolic databases EcoCyc, KEGG contained biochemical or genetic evidence for every metabolic reaction included in the *in silico* representation. In most cases, there existed both.

Edwards & Palsson

PNAS 97, 5528 (2000)

Genes included in *in silico* model of *E.coli*

Table 1. The genes included in the *E. coli* metabolic genotype (21)

Central metabolism (EMP, PPP, TCA cycle, electron transport)	*aceA, aceB, aceE, aceF, ackA, acnA, acnB, acs, adhE, agp, appB, appC, atpA, atpB, atpC, atpD, atpE, atpF, atpG, atpH, atpI, cydA, cydB, cydC, cydD, cyoA, cyoB, cyoC, cyoD, dld, eno, fba, fbp, fdhF, fdnG, fdnH, fdnI, fdoG, fdoH, fdoI, frdA, frdB, frdC, frdD, fumA, fumB, fumC, galM, gapA, gapC_1, gapC_2, glcB, glgA, glgC, glgP, glk, glpA, glpB, glpC, glpD, gltA, gnd, gpmA, gpmB, hyaA, hyaB, hyaC, hybA, hybC, hycB, hycE, hycF, hycG, icdA, lctD, ldhA, lpdA, malP, mdh, ndh, nuoA, nuoB, nuoE, nuoF, nuoG, nuoH, nuoI, nuoJ, nuoK, nuoL, nuoM, nuoN, pckA, pfkA, pfkB, pflA, pflB, pflC, pflD, pgi, pgk, pntA, pntB, ppc, ppsA, pta, purT, pykA, pykF, rpe, rpiA, rpiB, sdhA, sdhB, sdhC, sdhD, sfcA, sucA, sucB, sucC, sucD, talB, tktA, tktB, tpiA, trxB, zwf, pgl* (30), *maeB* (30)
Alternative carbon source	*adhC, adhE, agaY, agaZ, aldA, aldB, aldH, araA, araB, araD, bglX, cpsG, deoB, fruK, fucA, fucI, fucK, fucO, galE, galK, galT, galU, gatD, gatY, glk, glpK, gntK, gntV, gpsA, lacZ, manA, melA, mtlD, nagA, nagB, nanA, pfkB, pgi, pgm, rbsK, rhaA, rhaB, rhaD, srlD, treC, xylA, xylB*
Amino acid metabolism	*adi, aldH, alr, ansA, ansB, argA, argB, argC, argD, argE, argF, argG, argH, argI, aroA, aroB, aroC, aroD, aroE, aroF, aroG, aroH, aroK, aroL, asd, asnA, asnB, aspA, aspC, avtA, cadA, carA, carB, cysC, cysD, cysE, cysH, cysI, cysJ, cysK, cysM, cysN, dadA, dadX, dapA, dapB, dapD, dapE, dapF, dsdA, gabD, gabT, gadA, gadB, gdhA, glk, glnA, gltB, gltD, glyA, goaG, hisA, hisB, hisC, hisD, hisF, hisG, hisH, hisI, ilvA, ilvB, ilvC, ilvD, ilvE, ilvG_1, ilvG_2, ilvH, ilvI, ilvM, ilvN, kbl, ldcC, leuA, leuB, leuC, leuD, lysA, lysC, metA, metB, metC, metE, metH, metK, metL, pheA, proA, proB, proC, prsA, putA, sdaA, sdaB, setA, serB, serC, speA, speB, speC, speD, speE, speF, tdcB, tdh, thrA, thrB, thrC, tnaA, trpA, trpB, trpC, trpD, trpE, tyrA, tyrA, tyrB, ygjG, ygjH, alaB* (42), *dapC* (43), *pat* (44), *prr* (44), *sad* (45), *methylthioadenosine nucleosidase* (46), *5-methylthioribose kinase* (46), *5-methylthioribose-1-phosphate isomerase* (46), *adenosyl homocysteinase* (47), *L-cysteine desulfhydrase* (44), *glutaminase A* (44), *glutaminase B* (44)
Purine & pyrimidine metabolism	*add, adk, amn, apt, cdd, cmk, codA, dcd, deoA, deoD, dgt, dut, gmk, gpt, gsk, guaA, guaB, guaC, hpt, mutT, ndk, nrdA, nrdB, nrdD, nrdE, nrdF, purA, purB, purC, purD, purE, purF, purH, purK, purL, purM, purN, purT, pyrB, pyrC, pyrD, pyrE, pyrF, pyrG, pyrH, pyrI, tdk, thyA, tmk, udk, udp, upp, ushA, xapA, yicP, CMP glycosylase* (48)
Vitamin & cofactor metabolism	*acpS, bioA, bioB, bioD, bioF, coaA, cyoE, cysG, entA, entB, entC, entD, entE, entF, epd, folA, folC, folD, folE, folK, folP, gcvH, gcvP, gcvT, gltX, glyA, gor, gshA, gshB, hemA, hemB, hemC, hemD, hemE, hemF, hemH, hemK, hemL, hemM, hemX, hemY, ilvC, lig, lpdA, menA, menB, menC, menD, menE, menF, menG, metF, mutT, nadA, nadB, nadC, nadE, ntpA, pabA, pabB, pabC, panB, panC, panD, pdxA, pdxB, pdxH, pdxJ, pdxK, pncB, purU, ribA, ribB, ribD, ribE, ribH, serC, thiC, thiE, thiF, thiG, thiH, thrC, ubiA, ubiB, ubiC, ubiG, ubiH, ubiX, yaaC, ygiG, nadD* (49), *nadF* (49), *nadG* (49), *panE* (50), *pncA* (49), *pncC* (49), *thiB* (51), *thiD* (51), *thiK* (51), *thiL* (51), *thiM* (51), *thiN* (51), *ubiE* (52), *ubiF* (52), *arabinose-5-phosphate isomerase* (22), *phosphopantothenate-cysteine ligase* (50), *phosphopantothenate-cysteine decarboxylase* (50), *phospho-pantetheine adenylyltransferase* (50), *dephosphoCoA kinase* (50), *NMN glycohydrolase* (49)
Lipid metabolism	*accA, accB, accD, atoB, cdh, cdsA, cls, dgkA, fabD, fabH, fadB, gpsA, ispA, ispB, pgpB, pgsA, psd, pssA, pgpA* (53)
Cell wall metabolism	*ddlA, ddlB, galF, galU, glmS, glmU, htrB, kdsA, kdsB, kdtA, lpxA, lpxB, lpxC, lpxD, mraY, msbB, murA, murB, murC, murD, murE, murF, murG, murI, rfaC, rfaD, rfaF, rfaG, rfaI, rfaJ, rfaL, ushA, glmM* (54), *lpoA* (55), *rfaE* (55), *tetraacyldisaccharide 4' kinase* (55), *3-deoxy-D-manno-octulosonic-acid 8-phosphate phosphatase* (55)
Transport processes	*araE, araF, araG, araH, argT, aroP, artI, artJ, artM, artP, artQ, brnQ, cadB, chaA, chaB, chaC, cmtA, cmtB, codB, crr, cycA, cysA, cysP, cysT, cysU, cysW, cysZ, dctA, dcuA, dcuB, dppA, dppB, dppC, dppD, dppF, fadL, focA, fruA, fruB, fucP, gabP, galP, gatA, gatB, gatC, glnH, glnP, glnQ, glpF, glpT, gltJ, gltK, gltL, gltP, glt5, gntT, gpt, hisJ, hisM, hisP, hisQ, hpt, kdpA, kdpB, kdpC, kgtP, lacY, lamB, livF, livG, livH, livJ, livK, livM, lldP, lysP, malE, malF, malG, malK, malX, manX, manY, manZ, melB, mglA, mglB, mglC, mtlA, mtr, nagE, nanT, nhaA, nhaB, nupC, nupG, oppA, oppB, oppC, oppD, oppF, panF, pheP, pitA, pitB, pnuC, potA, potB, potC, potD, potE, potF, potG, potH, potI, proP, proV, proW, proX, pstA, pstB, pstC, pst5, ptsA, ptsG, ptsI, ptsN, ptsP, purB, putP, rbsA, rbsB, rbsC, rbsD, rhaT, sapA, sapB, sapD, sbp, sdaC, srlA_1, srlA_2, srlB, tdcC, tnaB, treA, treB, trkA, trkG, trkH, tsx, tyrP, ugpA, ugpB, ugpC, ugpE, uraA, xapB, xylE, xylF, xylG, xylH, fruF* (56), *gntS* (57), *metD* (43), *pnuE* (49), *scr* (56)

Edwards & Palsson

PNAS 97, 5528 (2000)

E.coli in silico – Flux balance analysis

Define $\alpha_i = 0$ for irreversible internal fluxes,

$\alpha_i = -\infty$ for reversible internal fluxes (use biochemical literature)

Transport fluxes for PO_4^{2-}, NH_3, CO_2, SO_4^{2-}, K^+, Na^+ were unrestrained.

For other metabolites $0 < v_i < v_i^{max}$ except for those that are able to leave the metabolic network (i.e. acetate, ethanol, lactate, succinate, formate, pyruvate etc.)

Find particular metabolic flux distribution in feasible set by **linear programming**. LP finds a solution that **minimizes** a particular metabolic **objective** –*Z* (subject to the imposed constraints) where e.g.

$$Z = \sum c_i \cdot v_i = \langle \mathbf{c} \cdot \mathbf{v} \rangle$$

When written in this way, the **flux balance analysis (FBA)** method finds the solution that maximizes the sum of all fluxes = gives maximal biomass.

Edwards & Palsson, PNAS 97, 5528 (2000)

E.coli in silico

Examine **changes** in the **metabolic capabilities** caused by hypothetical **gene deletions.**

To simulate a gene deletion, the flux through the corresponding enzymatic reaction was restricted to zero.

Compare optimal value of mutant (Z_{mutant}) to the „wild-type“ objective Z

$$\frac{Z_{mutant}}{Z}$$

to determine the systemic effect of the gene deletion.

Edwards & Palsson

PNAS 97, 5528 (2000)

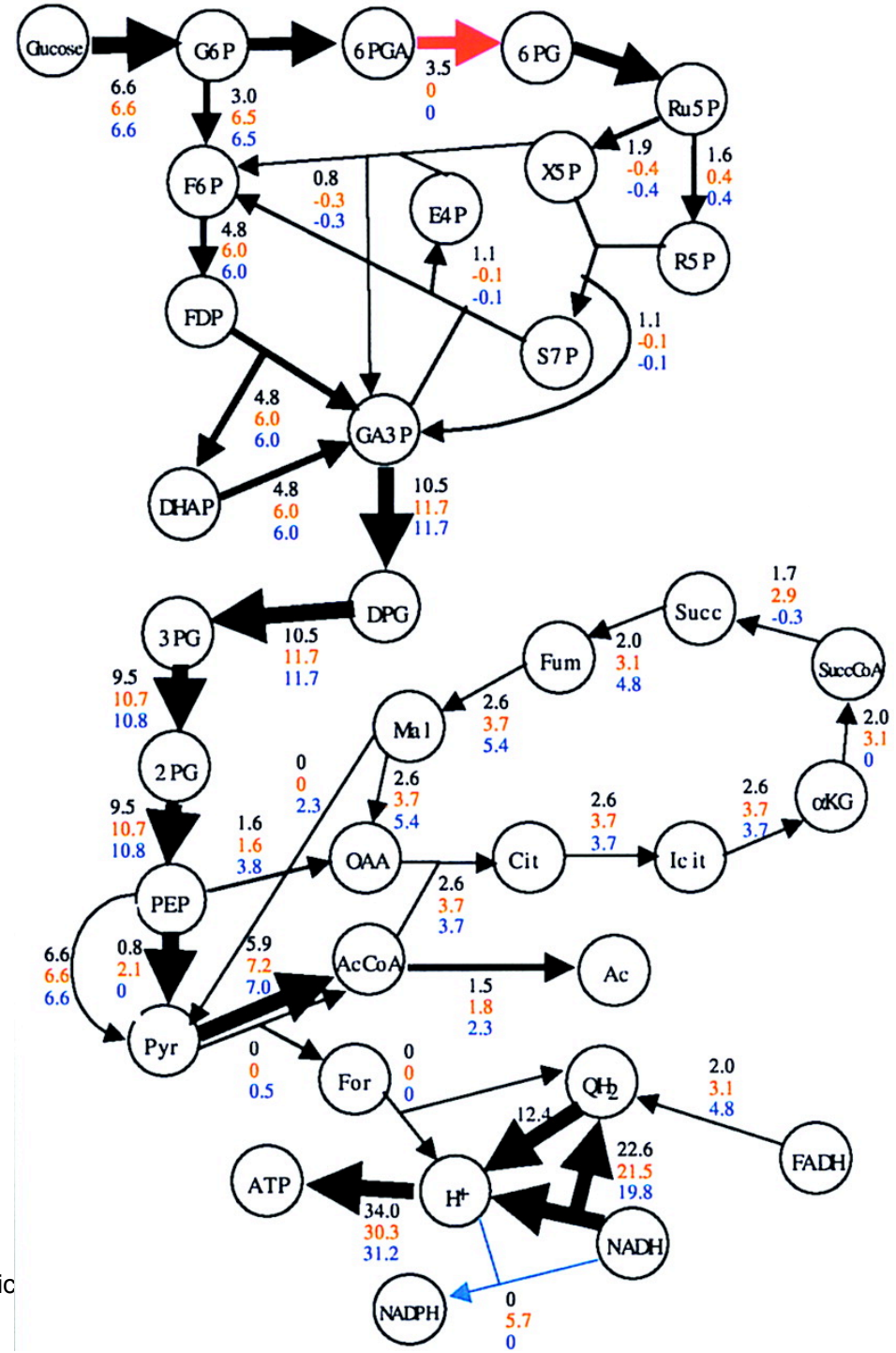
Rerouting of metabolic fluxes

(Black) Flux distribution for the wild-type.

(Red) *zwf*- mutant. Biomass yield is 99% of wild-type result.

(Blue) *zwf*- *pnt*- mutant. Biomass yield is 92% of wildtype result.

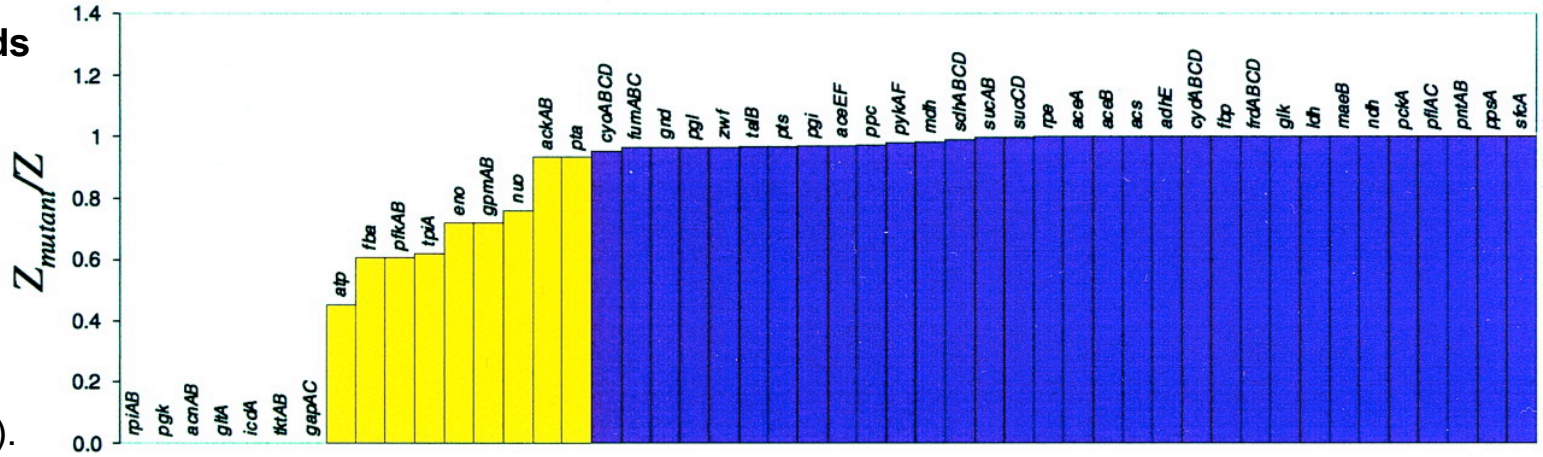
Note how *E.coli in silico* circumvents removal of one critical reaction (red arrow) by increasing the flux through the alternative G6P → P6P reaction.



Edwards & Palsson PNAS 97, 5528 (2000)

Gene deletions in central intermediary metabolism

Maximal biomass yields on glucose for all possible single gene deletions in the central metabolic pathways (gycolysis, pentose phosphate pathway (PPP), TCA, respiration).



The results were generated in a simulated aerobic environment with glucose as the carbon source. The transport fluxes were constrained as follows: glucose = 10 mmol/g-dry weight (DW) per h; oxygen = 15 mmol/g-DW per h.

The maximal yields were calculated by using FBA with the objective of maximizing growth.

Yellow bars: gene deletions that reduced the maximal biomass yield of Z_{mutant} to less than 95% of the *in silico* wild type Z_{wt}.

Edwards & Palsson PNAS 97, 5528 (2000)

Interpretation of gene deletion results

The essential gene products were involved in the 3-carbon stage of glycolysis, 3 reactions of the TCA cycle, and several points within the pentose phosphate pathway (PPP).

The remainder of the central metabolic genes could be removed while *E.coli in silico* maintained the potential to support cellular growth.

This suggests that a large number of the central metabolic genes can be removed without eliminating the capability of the metabolic network to support growth under the conditions considered.

Edwards & Palsson PNAS 97, 5528 (2000)

E.coli in silico – validation

+ and – means growth or no growth.

± means that suppressor mutations have been observed that allow the mutant strain to grow.

4 virtual growth media:

glc: glucose, gl: glycerol, succ: succinate, ac: acetate.

In 68 of 79 cases, the prediction was consistent with exp. predictions.

Red and yellow circles: predicted mutants that eliminate or reduce growth.

Edwards & Palsson

PNAS 97, 5528 (2000)

Table 2. Comparison of the predicted mutant growth characteristics from the gene deletion study to published experimental results with single mutants

Gene	glc	gl	succ	ac
aceA	+/+		+/+	−/−
aceB				−/−
*aceEF**	−/+			
ackA				+/+
acn	−/−			−/−
acs				+/+
cyd	+/+			
cyo	+/+			
eno†	−/+	−/+	−/−	−/−
fbaI	−/+			
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	+/+	+/+	+/+	

Summary - FBA

FBA analysis constructs the **optimal network utilization** simply using the stoichiometry of metabolic reactions and capacity constraints.

For *E.coli* the *in silico* results are mostly **consistent** with experimental data.

FBA shows that the *E.coli* metabolic network contains relatively **few critical gene products** in central metabolism.
However, the ability to adjust to different environments (growth conditions) may be diminished by gene deletions.

FBA identifies „**the best**“ the cell can do, not how the cell actually behaves under a given set of conditions. Here, survival was equated with growth.

FBA does not directly consider **regulation** or regulatory constraints on the metabolic network. This can be treated separately (see future lecture).

Edwards & Palsson PNAS 97, 5528 (2000)