V19 Metabolic Networks - Overview

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

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

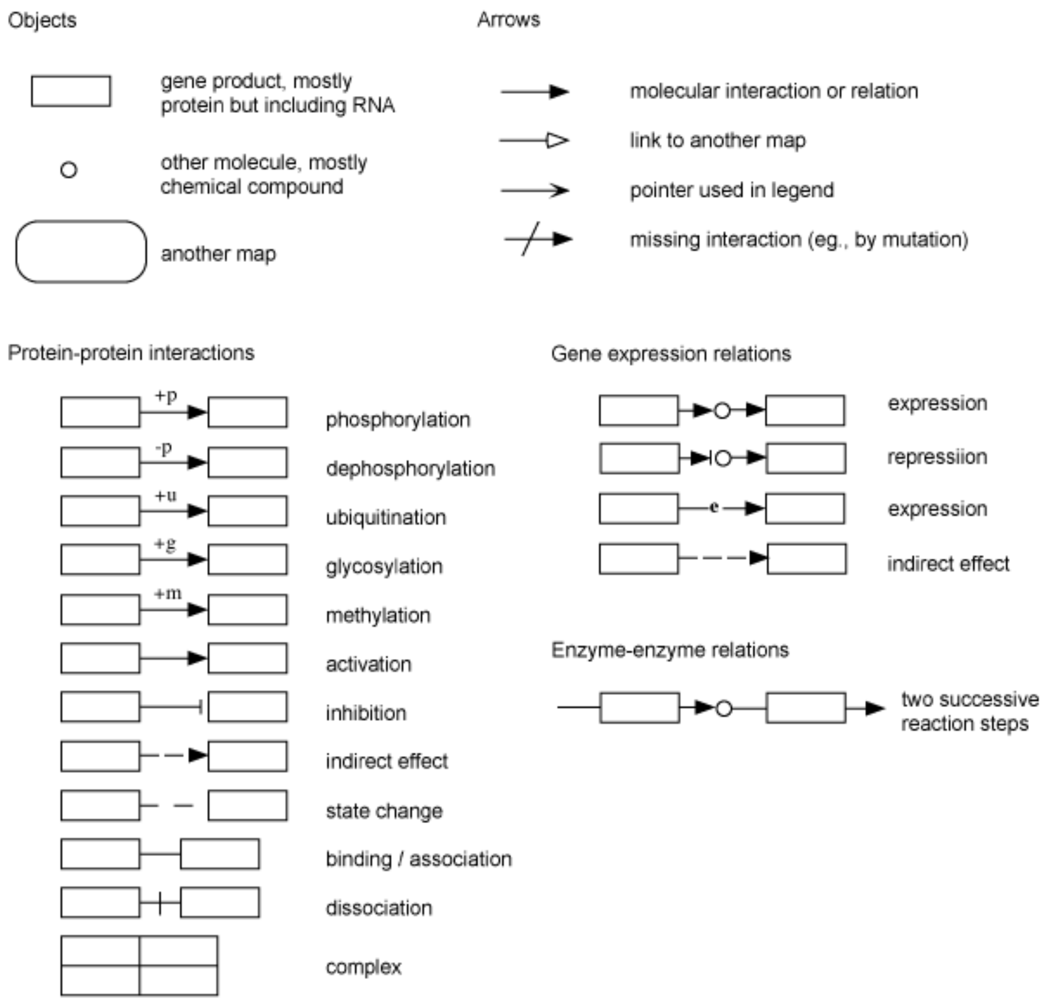
- stoichiometric modelling (**flux balance analysis**): theoretical capabilities of an integrated cellular process, feasible metabolic flux distributions

- 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

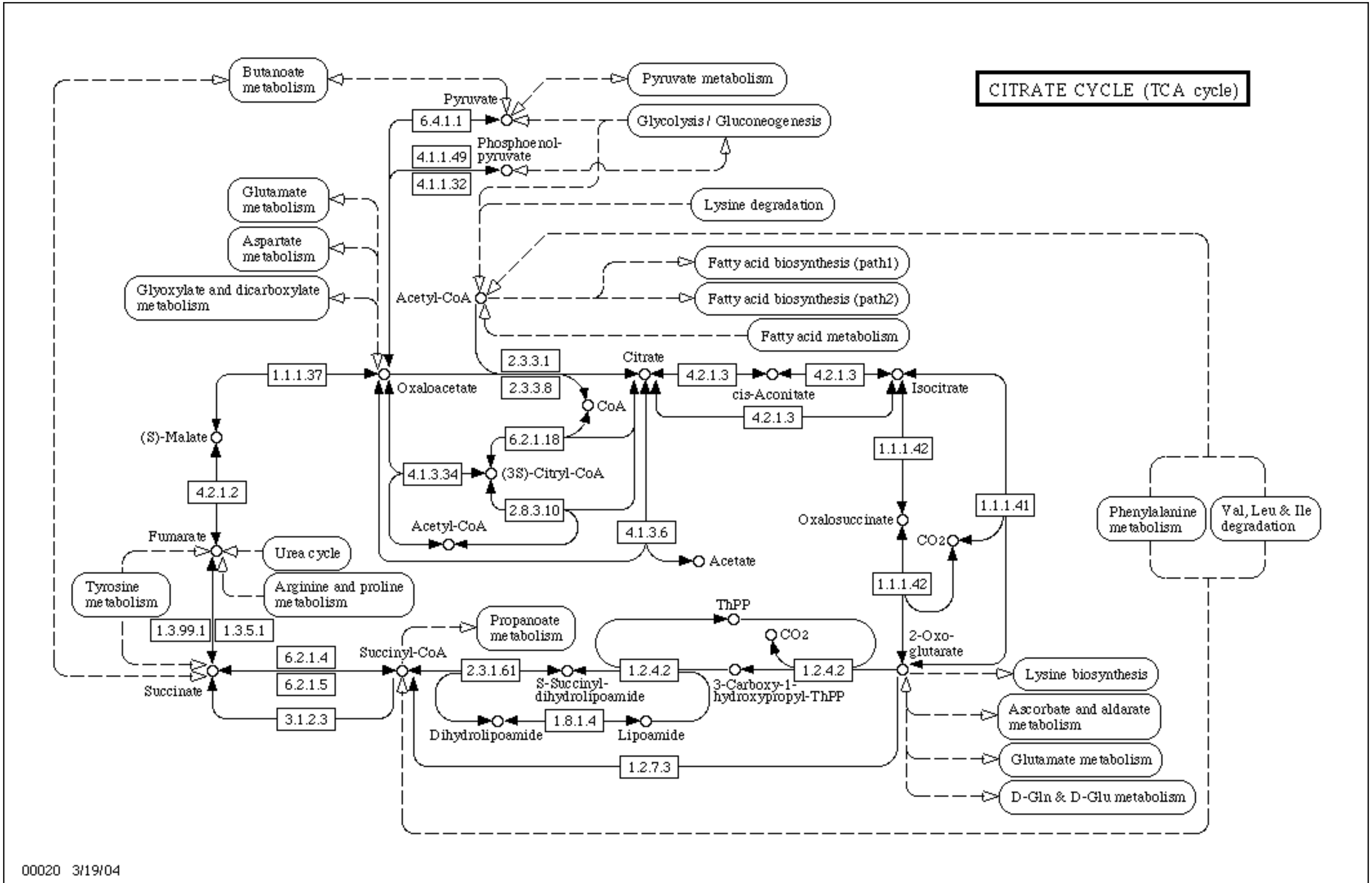
KEGG database



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.

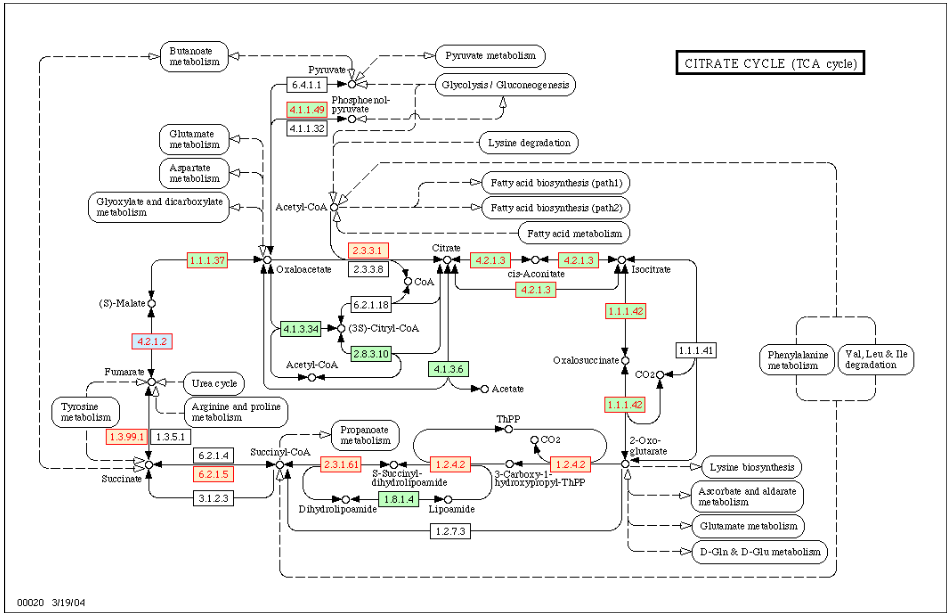
Organism-specific pathways (green-colored pathways) are computationally generated based on the KO assignment in individual genomes.

Citrate Cycle (TCA cycle) in E.coli

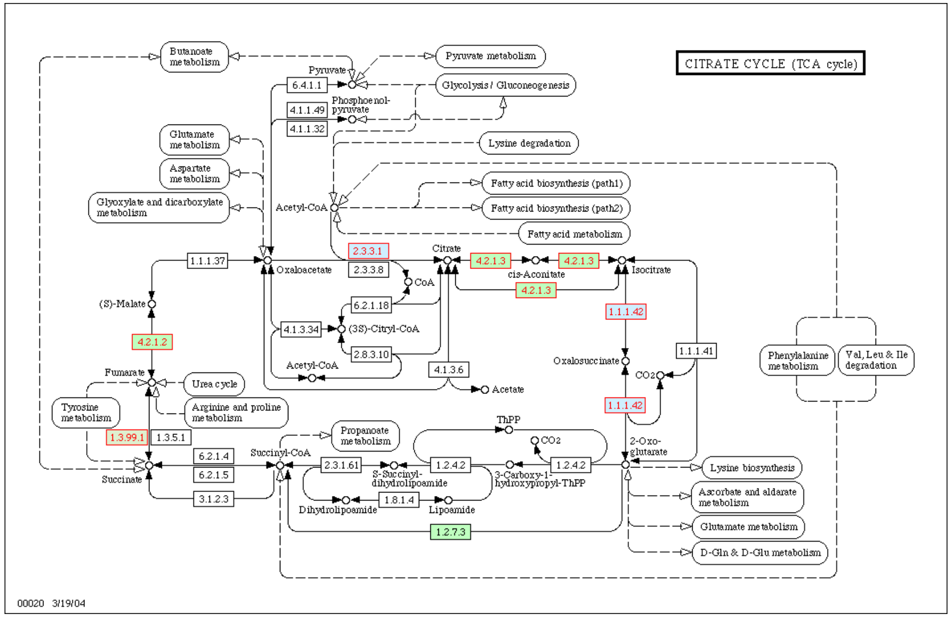


Citrate Cycle (TCA cycle) in different organisms

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.

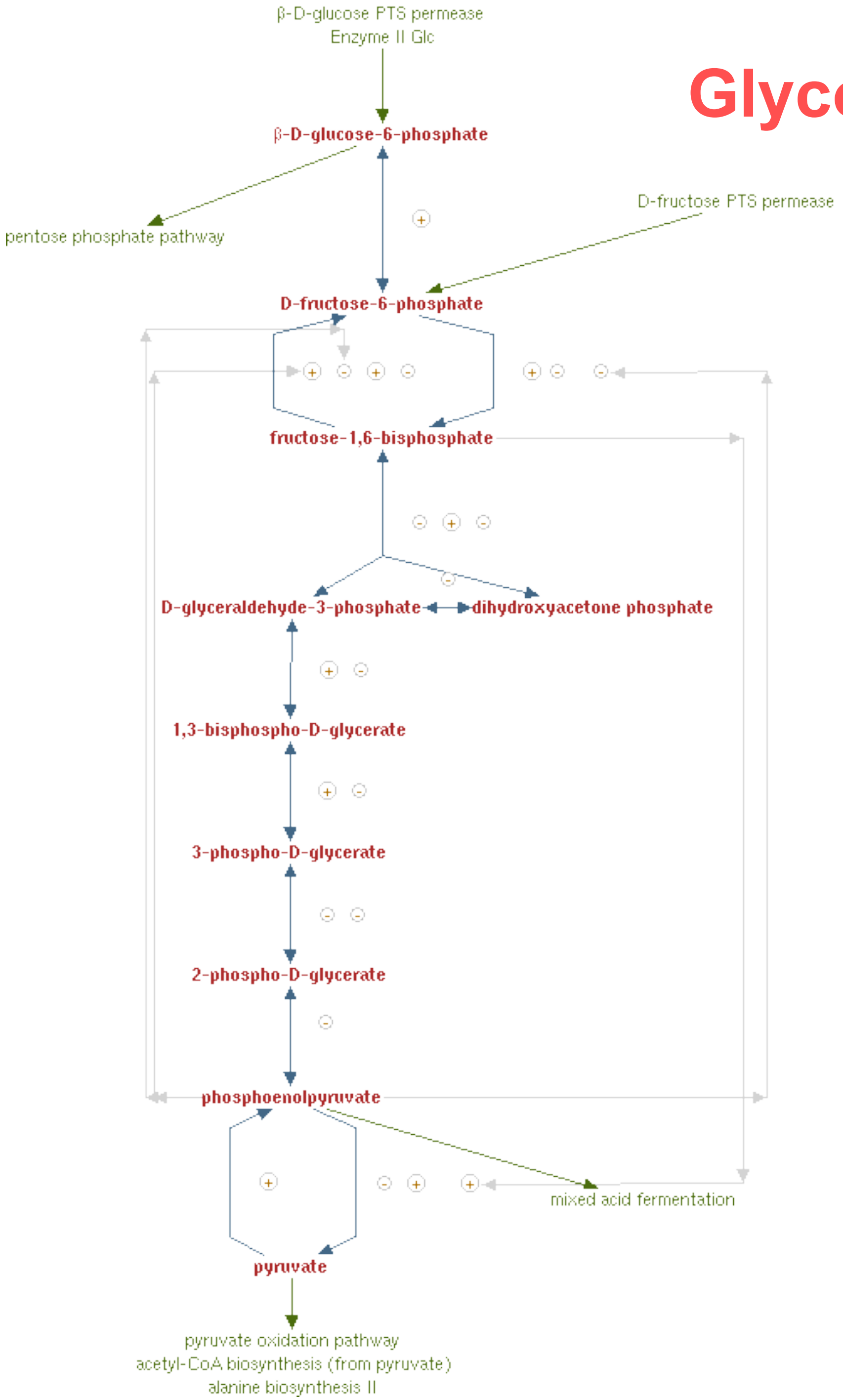
How can we characterize the functional complement of *E.coli* and according to what criteria can we compare the biochemical networks of two organisms?

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)



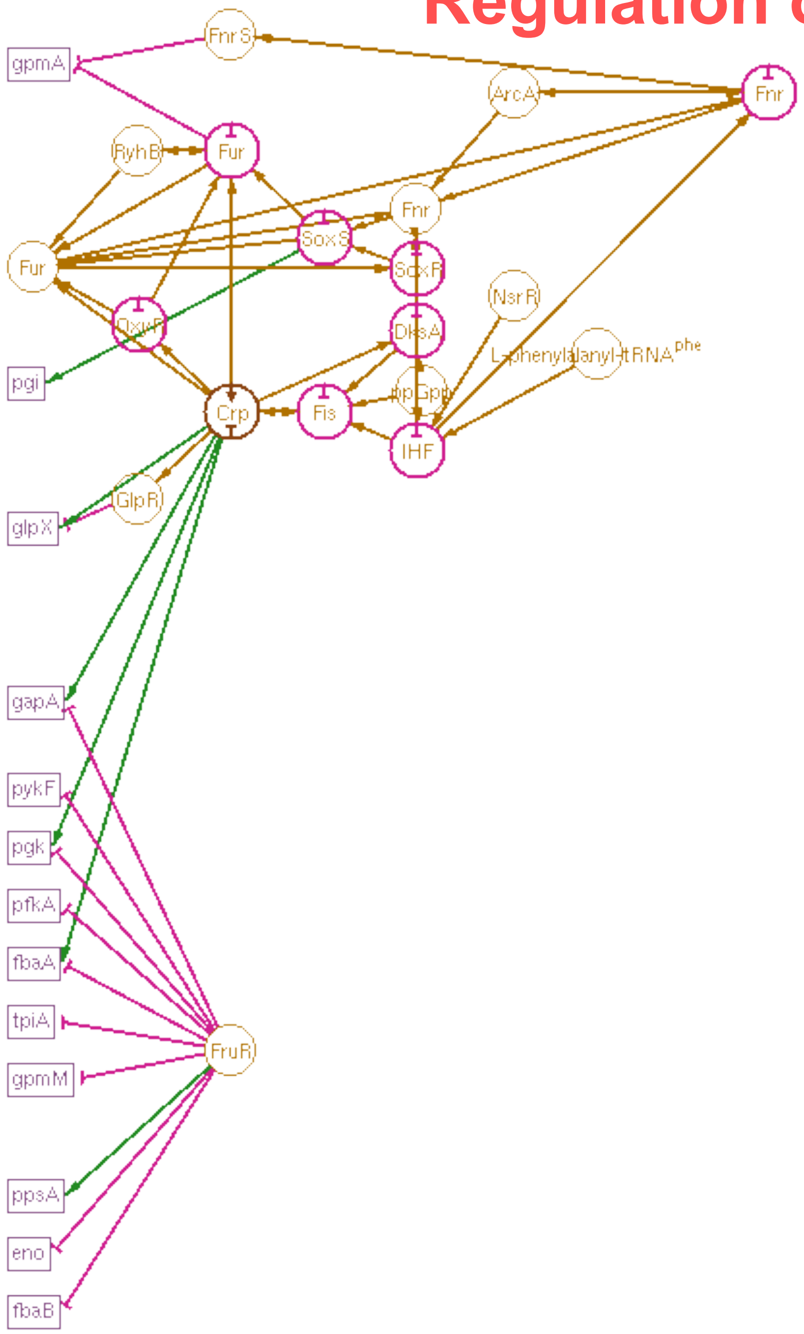
Glycolysis in E.coli

Blue arrows: biochemical reactions
clicking on arrow shows responsible enzyme

+ and - : activation and inhibition of enzymes

www.ecocyc.org

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

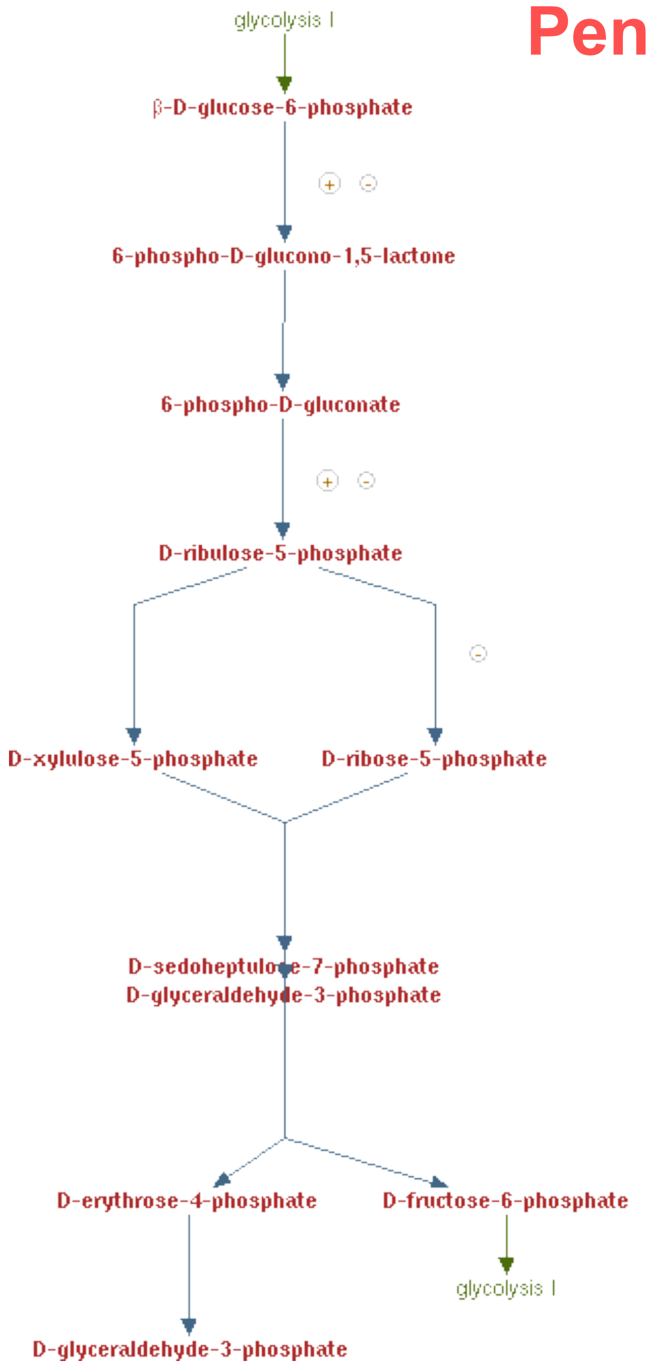
Green pointed arrows: activation of transcription;

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



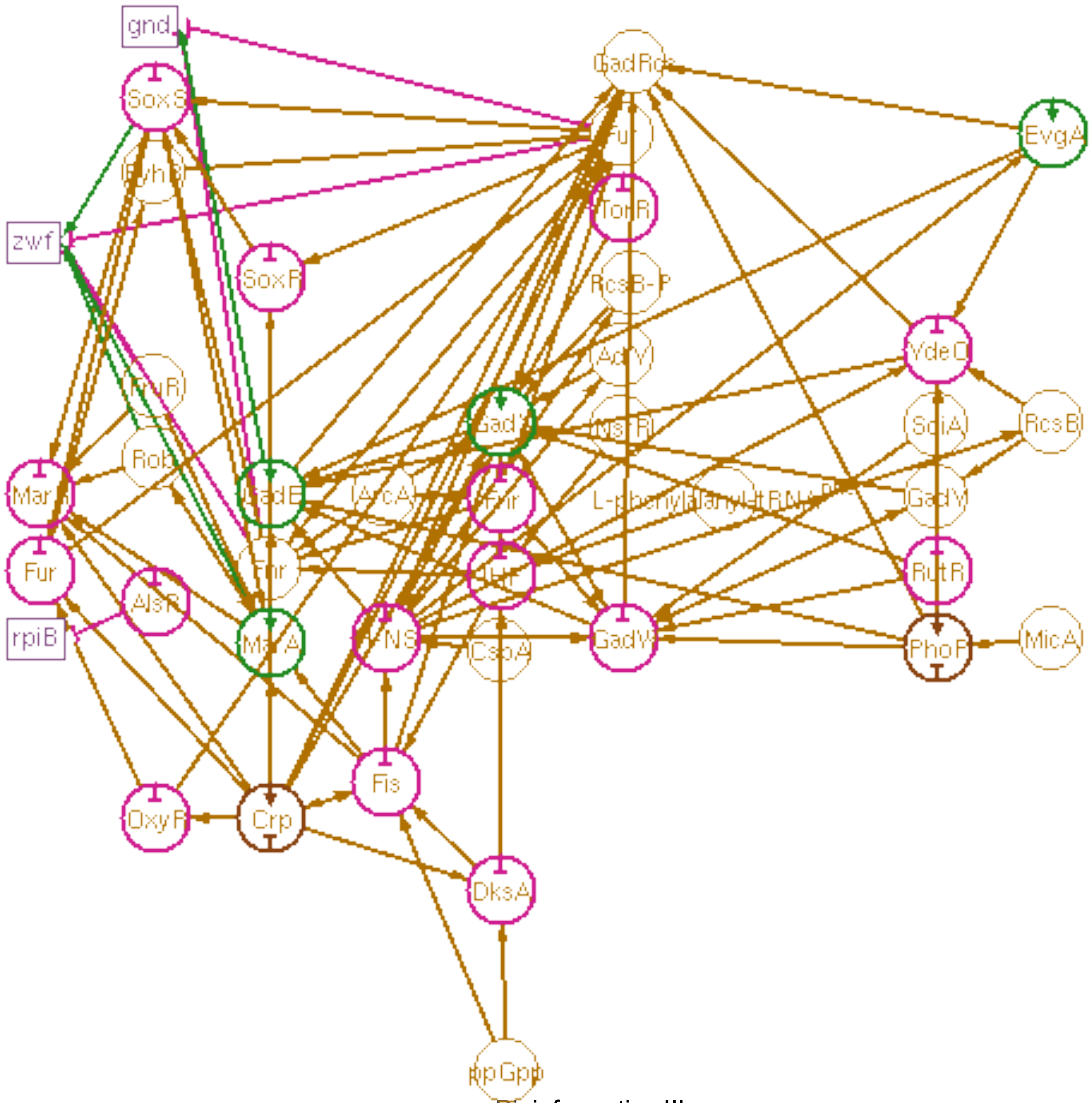
Blue arrows: biochemical reactions

clicking on arrow shows responsible enzyme

+ and - : activation and inhibition of enzymes

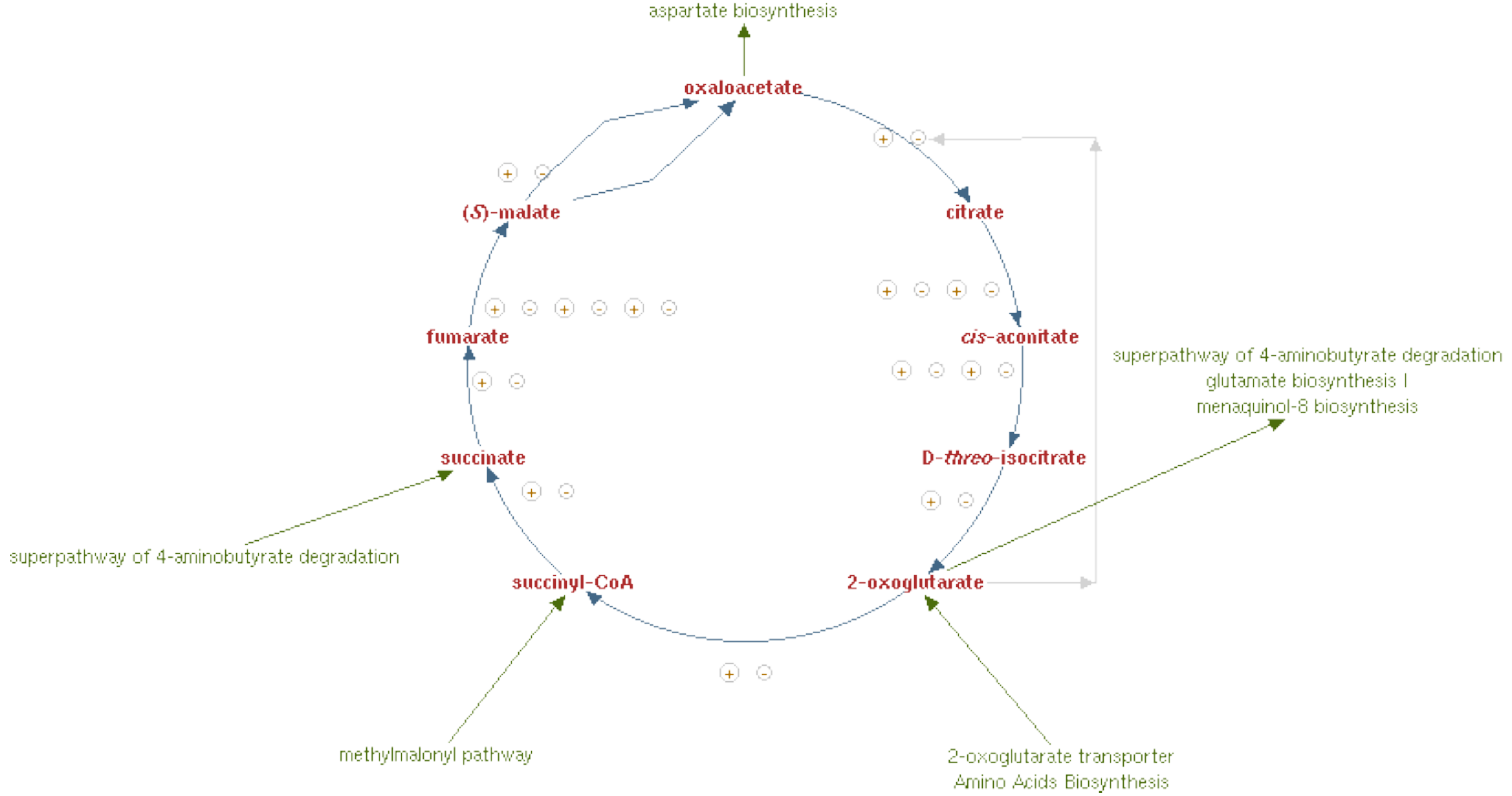
www.ecocyc.org

Regulation of Pentose Phosphate Pathway



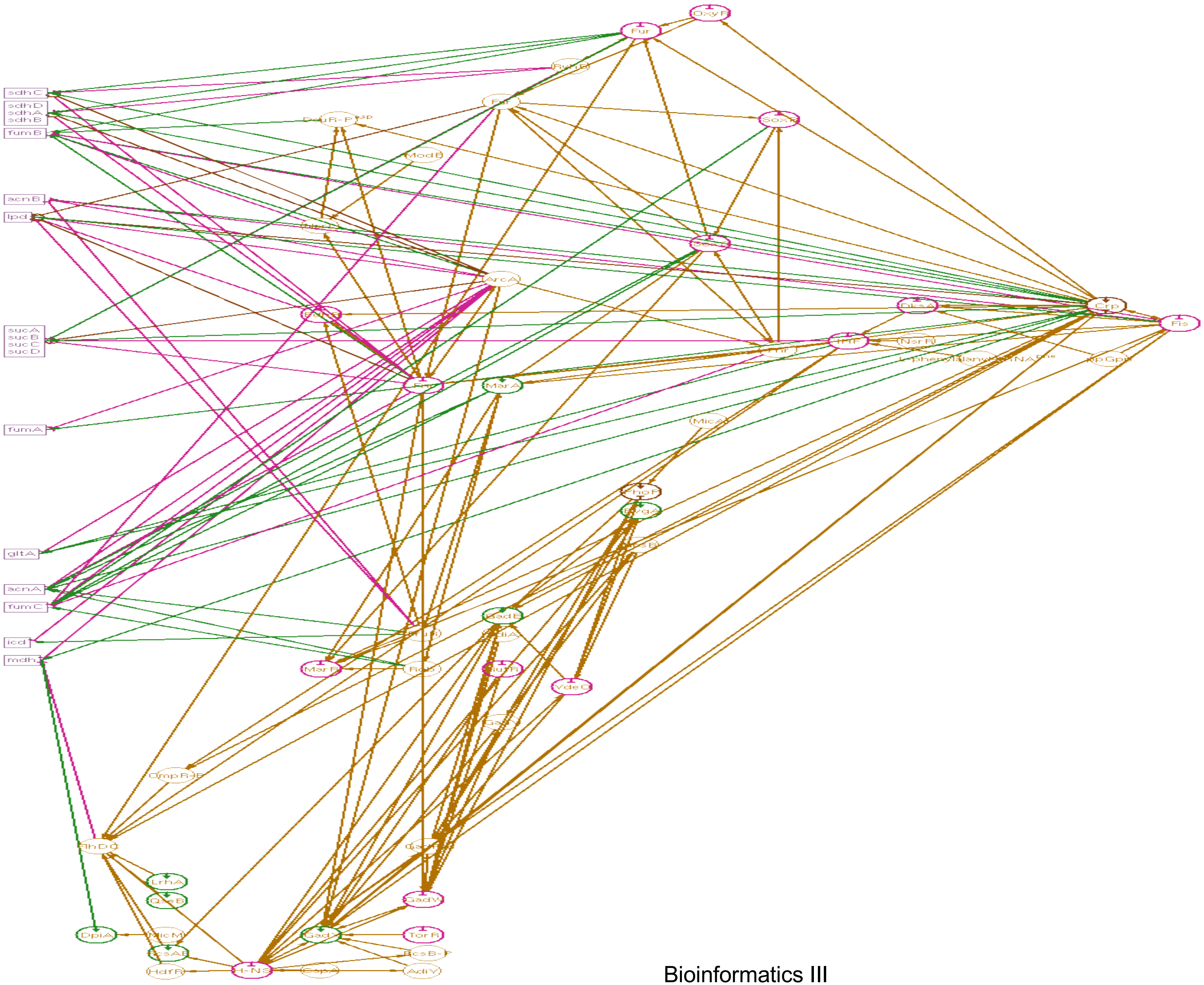
www.ecocyc.org

TCA cycle



www.ecocyc.org

Regulation of TCA cycle



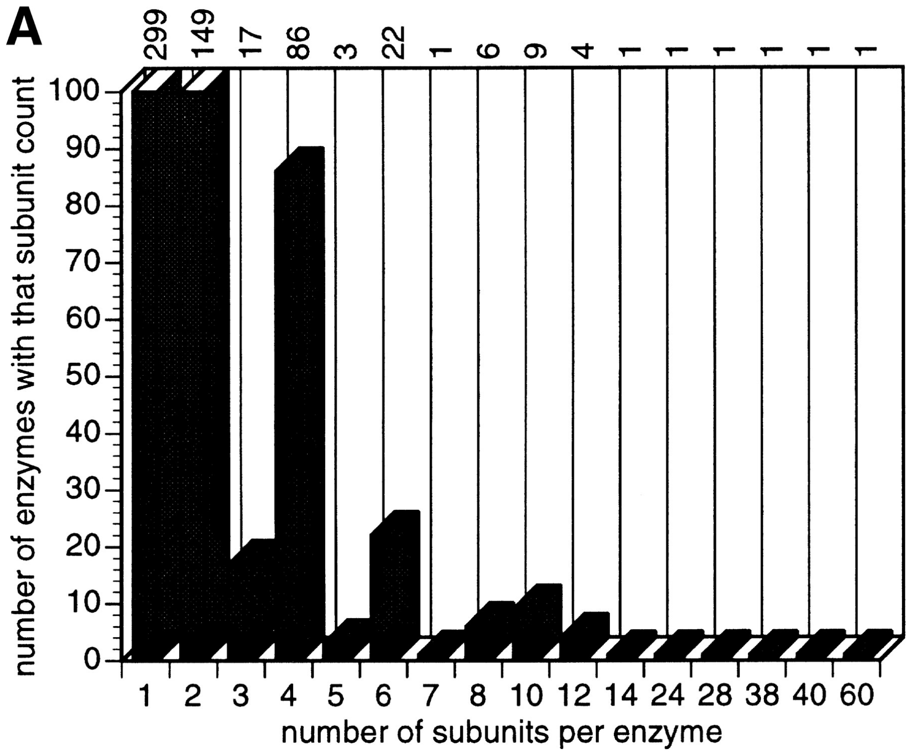
www.ecocyc.org

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.

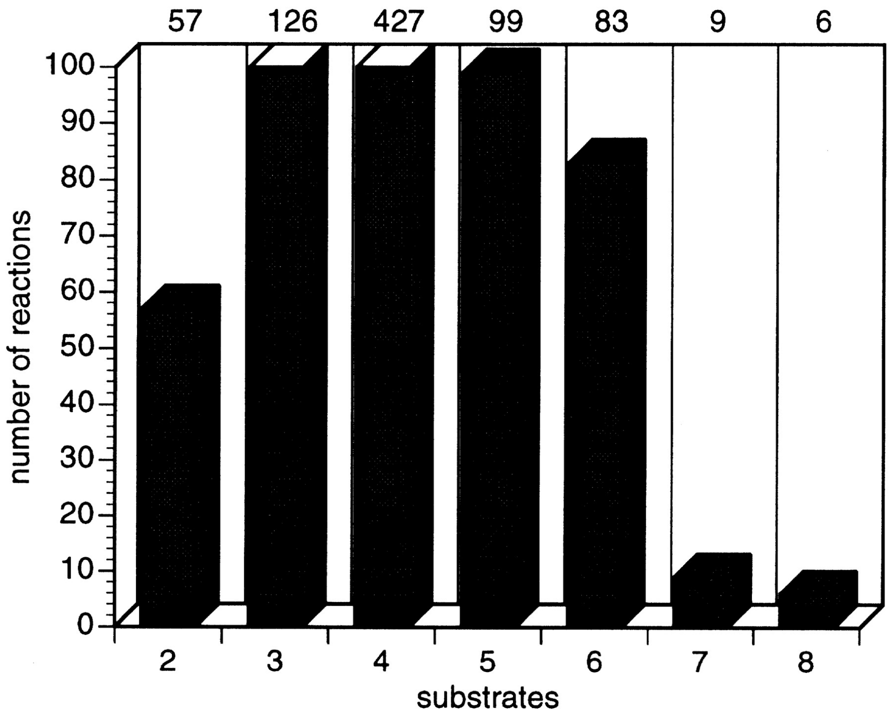
Ouzonis, Karp, Genome Res. 10, 568 (2000)

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)

Pathways

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

energy metabolism

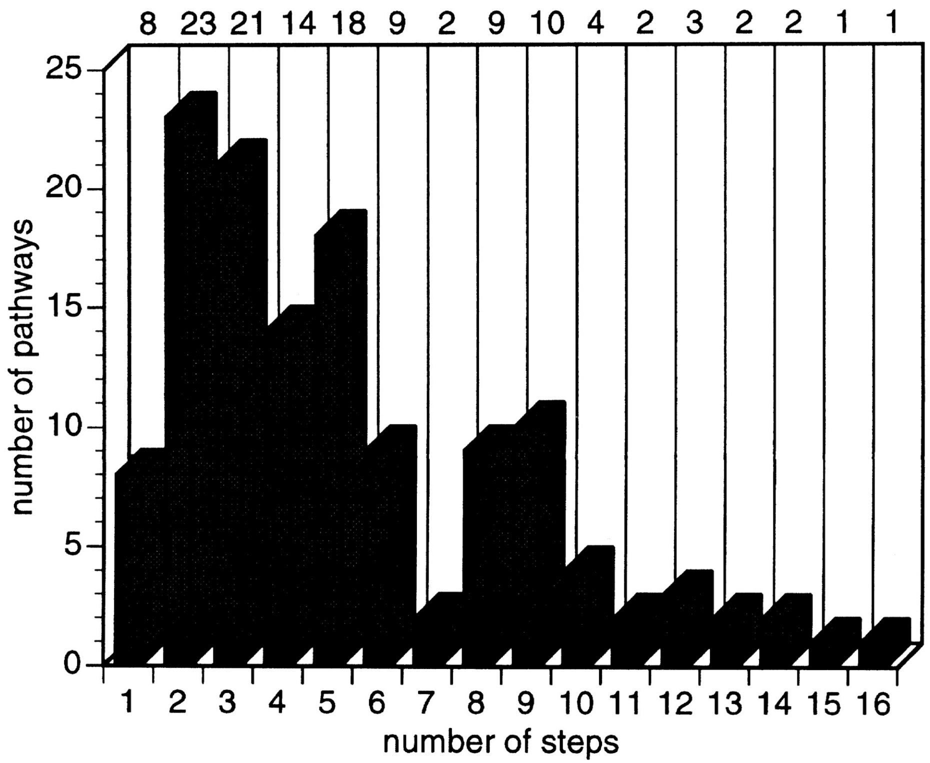
nucleotide and amino acid biosynthesis

secondary metabolism

Length distribution of EcoCyc pathways

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

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



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

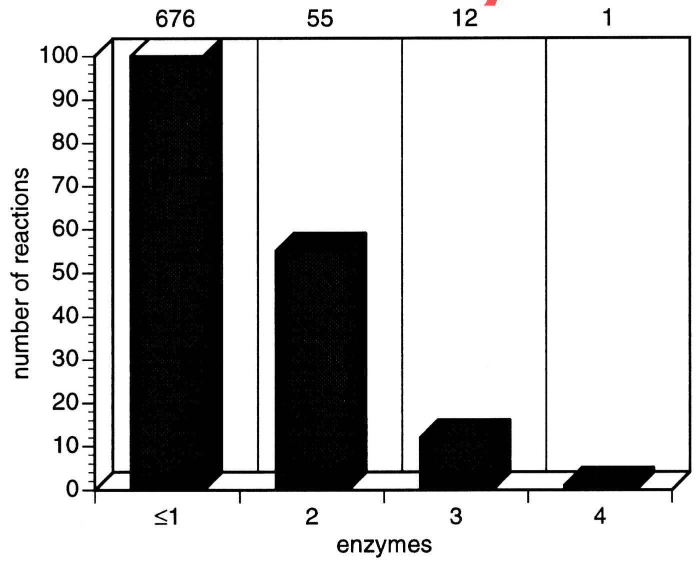
Table 3. Most Common Modulators, cofactors, and prosthetic groups of *E. coli* enzymes and Their Frequencies

A. Modulators (activators and inhibitors)				B. Cofactors and prosthetic groups			
Occurrence	Name of modulator	Activator	Inhibitor	Occurrence	Name of compound	Cofactor	Prosthetic group
35	Cu^{2+}		•	145	Mg^{2+}	•	•
32	ATP	•	•	48	pyridoxal 5′-phosphate	•	•
30	Zn^{2+}	•	•	33	Mn^{2+}	•	
29	AMP	•	•	31	FAD	•	•
26	ADP	•	•	21	Fe^{2+}	•	•
25	EDTA	•	•	18	Zn^{2+}	•	•
23	*p*-chloromercuribenzoate		•	16	thiamine-pyrophosphate		•
23	pyrophosphate	•	•	11	FMN	•	•
22	K^{+}	•	•	10	Co^{2+}	•	
22	phosphate	•	•	9	K^{+}	•	
20	Hg^{2+}		•	6	Mo^{2+}		•
20	Ca^{2+}	•	•	5	NAD	•	•
19	*N*-ethylmaleimide	•	•	4	protoheme		•
16	NAD	•	•	4	Ni^{2+}	•	•
16	iodoacetamide		•	4	Ca^{2+}	•	
16	coenzyme A		•	4	4Fe-4S center		•
15	Co^{2+}	•	•	3	NH_4^{+}	•	
15	Mg^{2+}	•	•	3	pyruvate		•
15	phosphoenolpyruvate	•	•	3	siroheme		•
14	Fe^{2+}	•	•	3	cytochrome *c*		•
14	GTP	•	•	2	heme C		•
14	pyruvate	•	•	2	B_{12}		•
13	*p*-hydroxymercuribenzoate		•	2	NADP	•	
13	NADP		•	2	Cu^{2+}		•
12	Mn^{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?

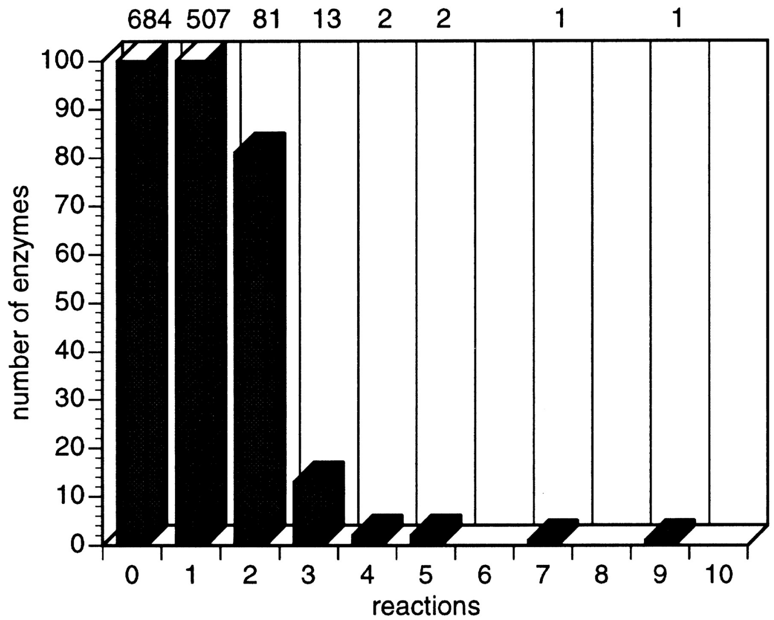
(1) the enzymes that catalyze the same reaction are **paralogs** (homologs) and have duplicated (or were obtained by horizontal gene transfer), acquiring some specificity but retaining the same mechanism (**divergence**)

(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

Of the 607 *E.coli* enzymes, 100 are multifunctional, either having the same active site and different substrate specificities or different active sites.

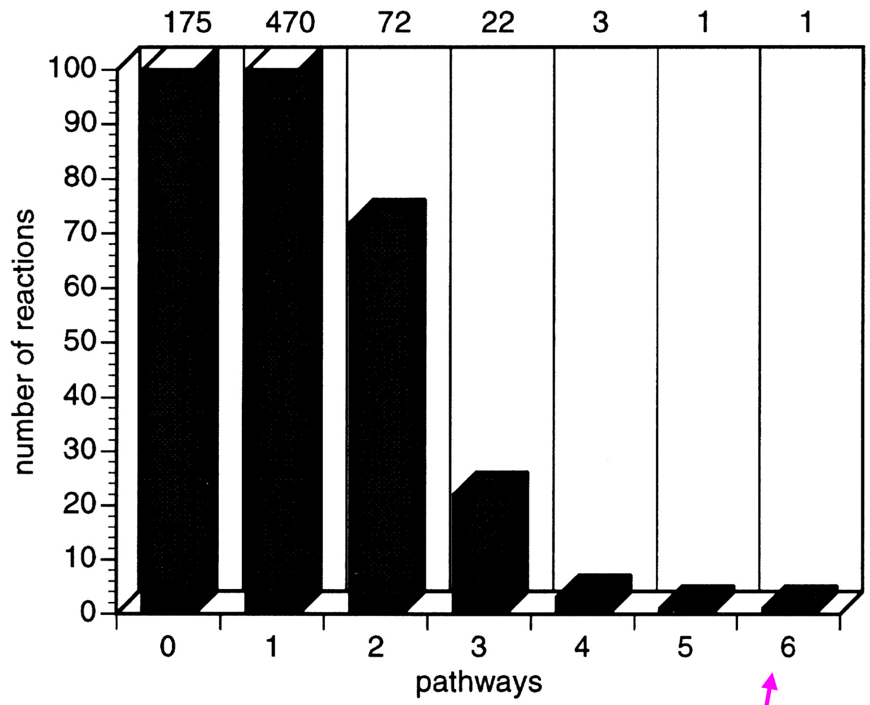


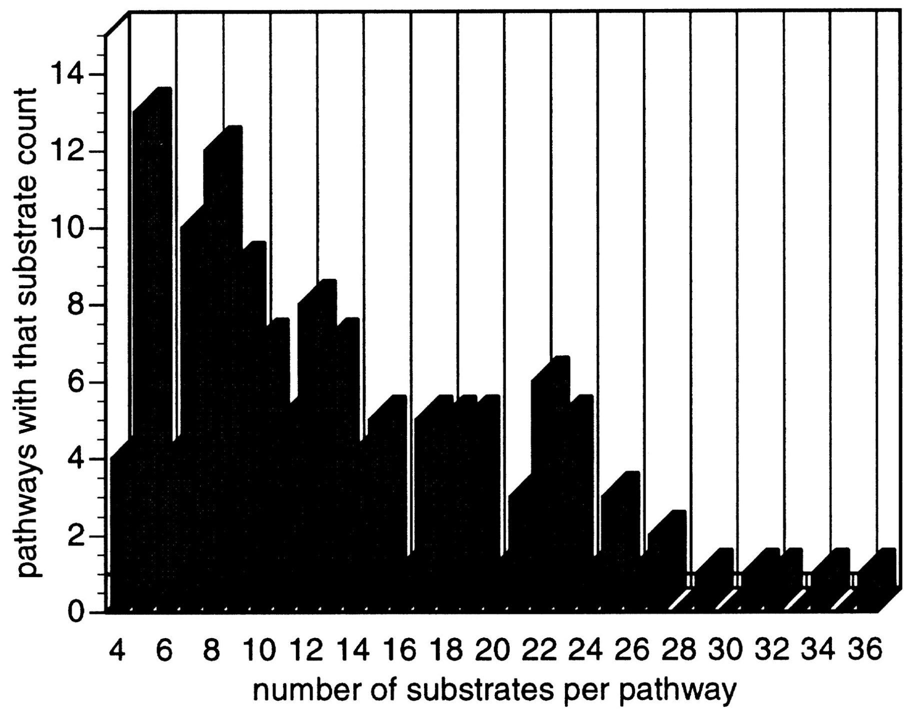
Number of enzymes that catalyze one or more reactions. Most enzymes catalyze one reaction; some are multifunctional.

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

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

Reactions participating in more than one pathway



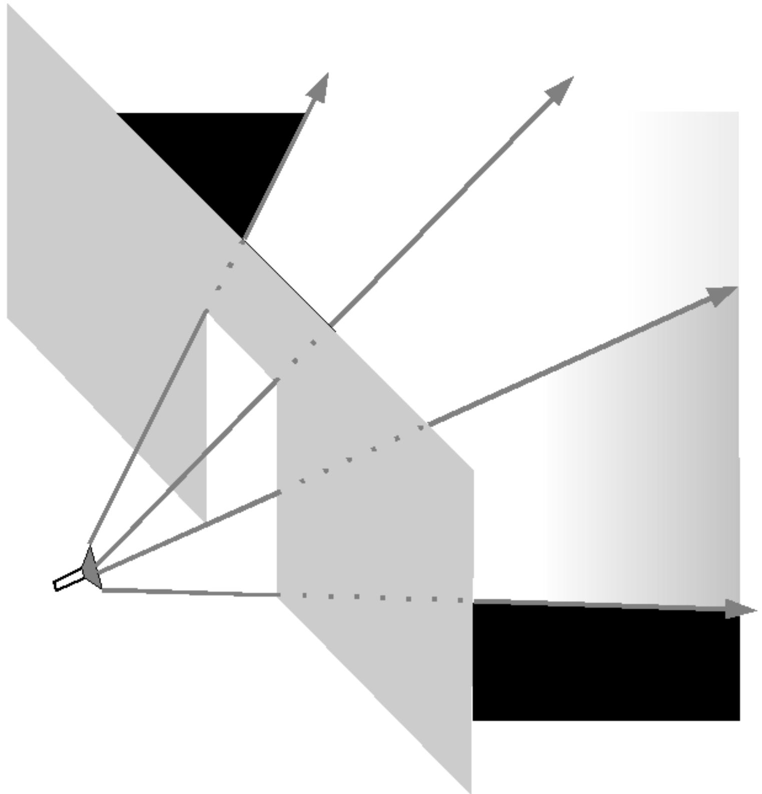


The 99 reactions belonging to multiple pathways appear to be the **intersection points** in the complex network of chemical processes in the cell.

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.

Extreme pathway analysis



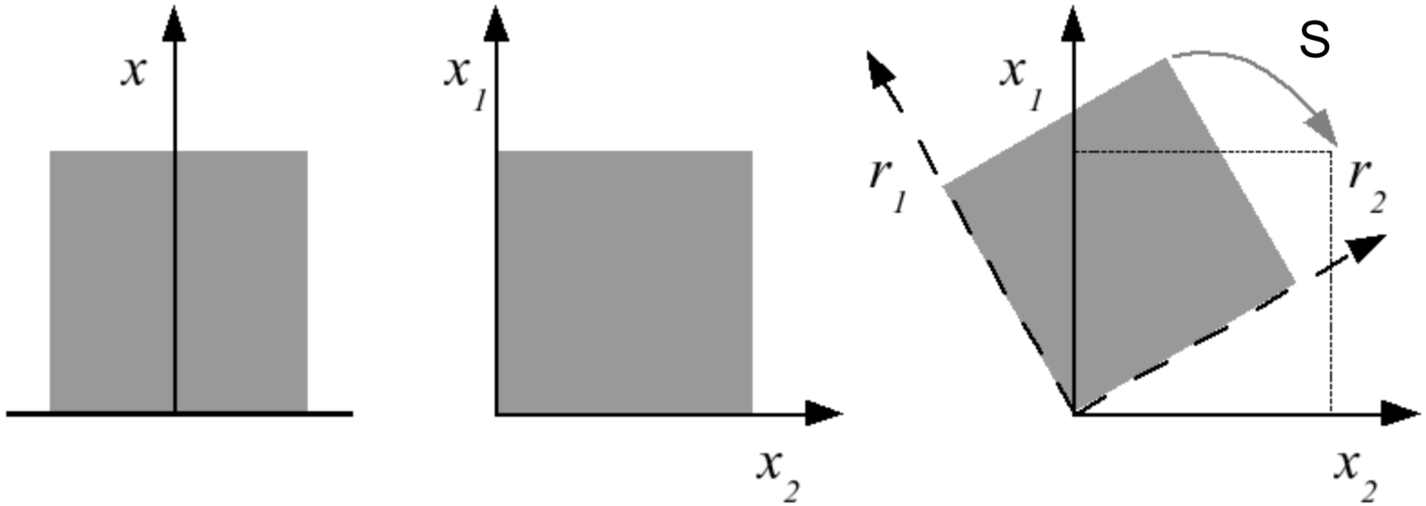
A torch is directed at an open door and shines into a dark room ...

What area is lighted ?

Instead of marking all lighted points individually,
it would be sufficient to characterize the „extreme rays“ that go through the corners of the door.

The lighted area is the area between the extreme rays = linear combinations of the extreme rays.

Idea – extreme pathways



Shaded area:

$x \geq 0$

Shaded area:

$x_1 \geq 0 \wedge x_2 \geq 0$

Either $\mathbf{S} \cdot \mathbf{x} \geq \mathbf{0}$

(**S** acts as rotation matrix)

or find optimal vectors

$\rightarrow$ change coordinate system

from x_1, x_2 to r_1, r_2.

Shaded area:

$r_1 \geq 0 \wedge r_2 \geq 0$

Duality of two matrices
S and R.

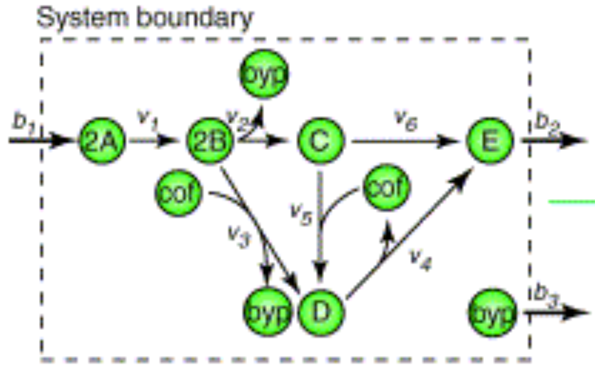
Edwards & Palsson PNAS 97, 5528 (2000)

Stoichiometric matrix

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

The stoichiometric 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.

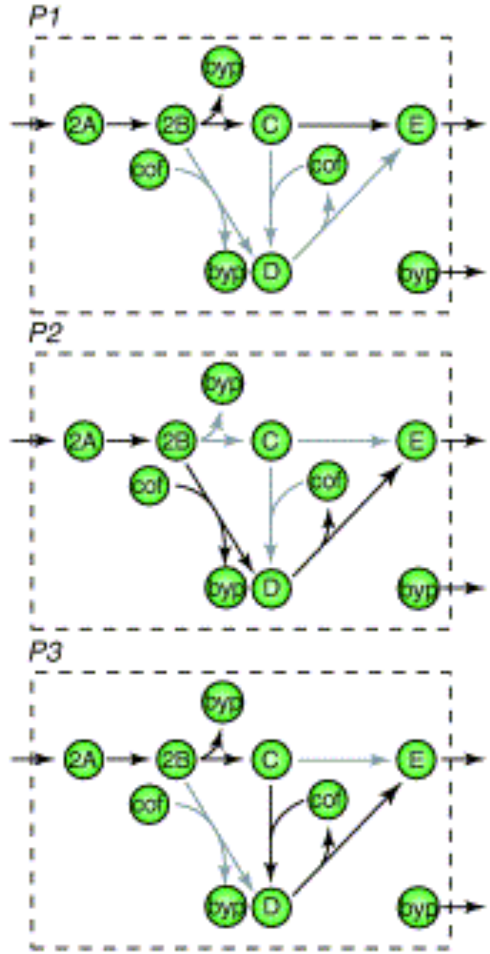


$$S = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{ccccccc|ccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & & b_1 & b_2 & b_3 \end{array}$$

$$S = \left\{\begin{array}{cccccc|ccc} -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ +1 & -2 & +2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & +1 & -1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & 0 & +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & +1 & -1 & 0 & 0 & 0 & 0 \end{array}\right\} \begin{array}{l} A \\ B \\ C \\ D \\ E \\ byp \\ cof \end{array}$$

$$P = \begin{array}{c} \\ \end{array}\begin{array}{ccc} P_1 & P_2 & P_3 \end{array}$$

$$P = \left\{\begin{array}{ccc} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ \hline 2 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{array}\right\} \begin{array}{l} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ b_1 \\ b_3 \\ b_2 \end{array}$$



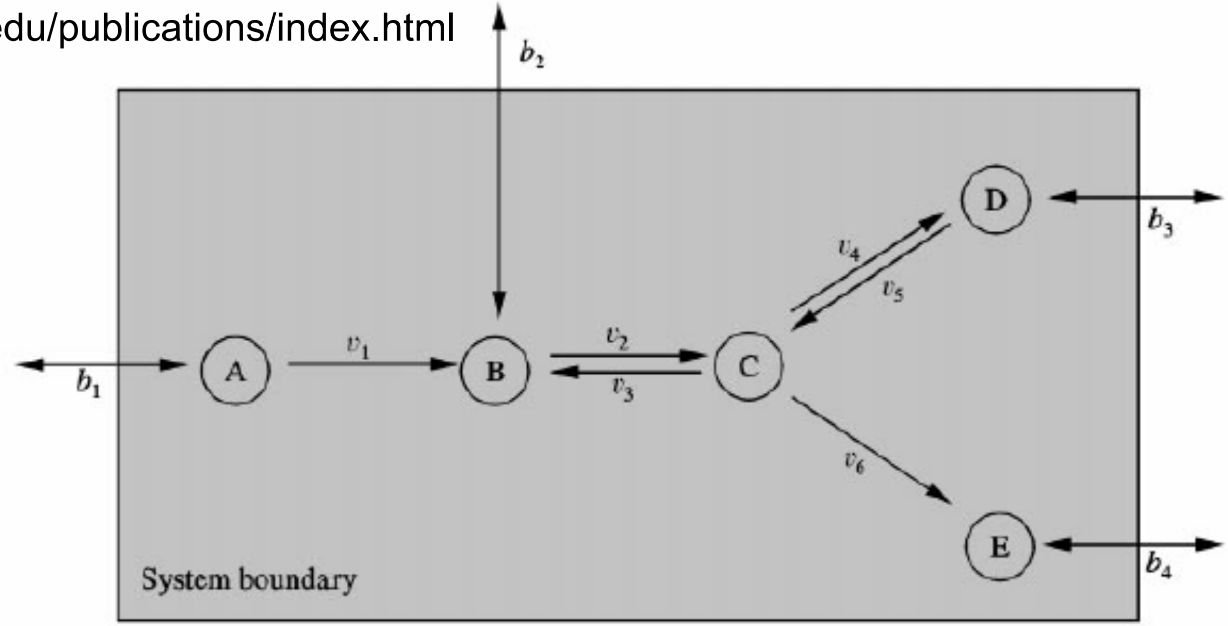
Papin et al. TIBS 28, 250 (2003)

Extreme Pathways

introduced into metabolic analysis by the lab of Bernard Palsson (Dept. of Bioengineering, UC San Diego). The publications of this lab are available at http://gcrg.ucsd.edu/publications/index.html

The extreme pathway technique is based on the stoichiometric matrix representation of metabolic networks.

All external fluxes are defined as pointing outwards.



Mass balance constraints

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

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

Internal flux constraints

$$v_j \geq 0, \quad j = 1, \ldots, 6$$

Exchange flux constraints

$$-\infty \leq b_j \leq +\infty, \quad j = 1, \ldots, 4$$

Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000)

Extreme Pathways – algorithm - setup

The algorithm to determine the set of extreme pathways for a reaction network follows the pinciples of algorithms for finding the extremal rays/ generating vectors of convex polyhedral cones.

Combine $n \times n$ identity matrix (**I**) with the transpose of the stoichiometric matrix $\mathbf{S}^T$. **I** serves for bookkeeping.

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}$$

S

$$\mathbf{T}^{(0)} = \left[\begin{array}{cccccc|ccccc} 1 & & & & & & -1 & 1 & 0 & 0 & 0 \\ & 1 & & & & & 0 & -1 & 1 & 0 & 0 \\ & & 1 & & & & 0 & 1 & -1 & 0 & 0 \\ & & & 1 & & & 0 & 0 & -1 & 1 & 0 \\ & & & & 1 & & 0 & 0 & 1 & -1 & 0 \\ & & & & & 1 & 0 & 0 & -1 & 0 & 1 \end{array}\right],$$

$$\mathbf{T}^{(E)} = \left[\begin{array}{cccccccccc|ccccc} & & & & & & 1 & & & & -1 & 0 & 0 & 0 & 0 \\ & & & & & & & 1 & & & 0 & -1 & 0 & 0 & 0 \\ & & & & & & & & 1 & & 0 & 0 & -0 & -1 & 0 \\ & & & & & & & & & 1 & 0 & 0 & -0 & 0 & -1 \end{array}\right]$$

I **S**^{T}

Schilling, Letscher, Palsson,
J. theor. Biol. 203, 229 (2000)

separate internal and external fluxes

Examine constraints on each of the exchange fluxes as given by

$$\alpha_j \leq b_j \leq \beta_j$$

If the exchange flux is constrained to be positive $\rightarrow$ do nothing.

If the exchange flux is constrained to be negative $\rightarrow$ multiply the corresponding row of the initial matrix by -1.

If the exchange flux is unconstrained $\rightarrow$ move the entire row to a temporary matrix $\mathbf{T}^{(E)}$.

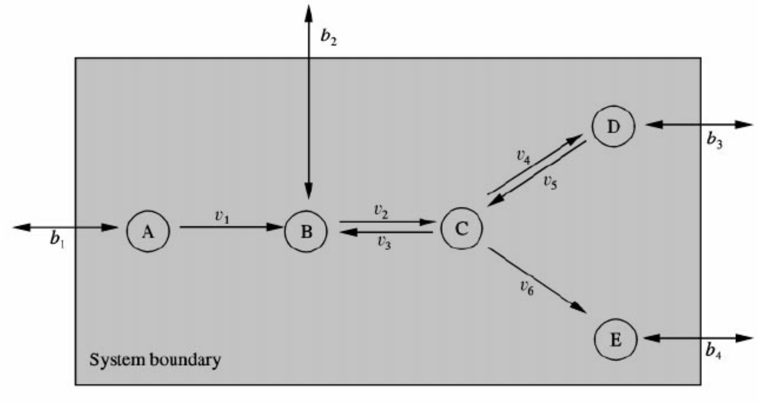
This completes the first tableau $\mathbf{T}^{(0)}$.

Schilling, Letscher, Palsson,
J. theor. Biol. 203, 229 (2000)

idea of algorithm

(1) Identify all metabolites that do not have an unconstrained exchange flux associated with them.
The total number of such metabolites is denoted by μ.
The example system contains only one such metabolite, namely C ($\mu = 1$).

What is the **main idea** of this step?
- We want to find balanced extreme pathways that don‘t change the concentrations of metabolites when flux flows through (input fluxes are channelled to products not to accumulation of intermediates).



- The stochiometrix matrix describes the coupling of each reaction to the concentration of metabolites X.
- Now we need to balance combinations of reactions that leave concentrations unchanged. Pathways applied to metabolites should not change their concentrations $\rightarrow$ the matrix entries need to be brought to 0.

Schilling, Letscher, Palsson,
J. theor. Biol. 203, 229 (2000)

keep pathways that do not change concentrations of internal metabolites

(2) Begin forming the new matrix $\mathbf{T}^{(i)}$ by copying all rows from $\mathbf{T}^{(i-1)}$ which already contain a zero in the column of $\mathbf{S}^T$ that corresponds to the first metabolite identified in step 1, denoted by index *C*. (Here 3rd column of $\mathbf{S}^T$.)

$T^{(0)}$ =

						A	B	**C**	D	E
1						-1	1	**0**	0	0
	1					0	-1	1	0	0
		1				0	1	-1	0	0
			1			0	0	-1	1	0
				1		0	0	1	-1	0
					1	0	0	-1	0	1

$T^{(1)}$ =

1						-1	1	**0**	0	0

+

Schilling, Letscher, Palsson, J. theor. Biol. 203, 229 (2000)

balance combinations of other pathways

(3) Of the remaining rows in **T**$^{(i-1)}$ add together all possible **combinations of rows** which contain values of the opposite sign in column C, such that the addition produces a zero in this column.

$T^{(0)}$ =

1						-1	1	0	0	0
	1					0	-1	1	0	0
		1				0	1	-1	0	0
			1			0	0	-1	1	0
				1		0	0	1	-1	0
					1	0	0	-1	0	1

$T^{(1)}$ =

1	0	0	0	0	0	-1	1	0	0	0
0	1	1	0	0	0	0	0	0	0	0
0	1	0	1	0	0	0	-1	0	1	0
0	1	0	0	0	1	0	-1	0	0	1
0	0	1	0	1	0	0	1	0	-1	0
0	0	0	1	1	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	-1	1
1	2	3	4	5	6	7	8	9	10	11

Schilling, et al.
JTB 203, 229

remove "non-orthogonal" pathways

(4) For all rows added to $\mathbf{T}^{(i)}$ in steps 2 and 3 check that no row exists that is a non-negative combination of any other rows in $\mathbf{T}^{(i)}$.

One method for this works as follows:
let $A(i)$ = set of column indices j for which the elements of row i = 0.

For the example above
A(1) = {2,3,4,5,6,9,10,11}
A(2) = {1,4,5,6,7,8,9,10,11}
A(3) = {1,3,5,6,7,9,11}
A(4) = {1,3,4,5,7,9,10}
A(5) = {1,2,4,6,7,9,11}
A(6) = {1,2,3,6,7,8,9,10,11}
A(7) = {1,2,3,4,7,8,9}

Then check to determine if there exists another row *(h)* for which *A(i)* is a subset of A*(h)*.

If $A(i) \subseteq A(h),\ i \neq h$
where
$A(i) = \{\, j : T_{i,j} = 0,\ 1 \leq j \leq (n+m) \,\}$
then row *i* must be eliminated from $\mathbf{T}^{(i)}$

Schilling et al.
JTB 203, 229

repeat steps for all internal metabolites

(5) With the formation of $\mathbf{T}^{(i)}$ complete steps 2 – 4 for all of the metabolites that do not have an unconstrained exchange flux operating on the metabolite, incrementing *i* by one up to μ. The final tableau will be $T^{(\mu)}$.

Note that the number of rows in $T^{(\mu)}$ will be equal to *k*, the number of extreme pathways.

Schilling et al.
JTB 203, 229

balance external fluxes

(6) Next we append $\mathbf{T}^{(E)}$ to the bottom of $\mathbf{T}^{(\mu)}$. (In the example here $\mu = 1$.) This results in the following tableau:

$T^{(1/E)} =$

1										-1	1	0	0	0
	1	1								0	0	0	0	0
	1		1							0	-1	0	1	0
	1				1					0	-1	0	1	0
		1		1						0	1	0	-1	0
			1	1						0	0	0	0	0
				1	1					0	0	0	-1	1
						1				-1	0	0	0	0
							1			0	-1	0	0	0
								1		0	0	0	-1	0
									1	0	0	0	0	-1

Schilling et al.
JTB 203, 229

balance external fluxes

(7) Starting in the *n+1* column (or the first non-zero column on the right side), if $T_{i,(n+1)} \neq 0$ then add the corresponding non-zero row from $\mathbf{T}^{(E)}$ to row *i* so as to produce 0 in the *n+1*-th column.

This is done by simply multiplying the corresponding row in $\mathbf{T}^{(E)}$ by $T_{i,(n+1)}$ and adding this row to row *i* .

Repeat this procedure for each of the rows in the upper portion of the tableau so as to create zeros in the entire upper portion of the *(n+1)* column.

When finished, remove the row in $\mathbf{T}^{(E)}$ corresponding to the exchange flux for the metabolite just balanced.

Schilling et al.
JTB 203, 229

balance external fluxes

(8) Follow the same procedure as in step (7) for each of the columns on the right side of the tableau containing non-zero entries.

(In our example we need to perform step (7) for every column except the middle column of the right side which correponds to metabolite C.)

The final tableau $\mathbf{T}^{(final)}$ will contain the transpose of the matrix **P** containing the extreme pathways in place of the original identity matrix.

Schilling et al.
JTB 203, 229

pathway matrix

$\mathbf{T}^{(\text{final})} =$

1						-1	1			0	0	0	0	0	0
	1	1								0	0	0	0	0	0
	1		1				-1	1		0	0	0	0	0	0
	1				1		-1		1	0	0	0	0	0	0
		1		1			1	-1		0	0	0	0	0	0
			1	1						0	0	0	0	0	0
				1	1			-1	1	0	0	0	0	0	0

$\mathbf{P}^T =$

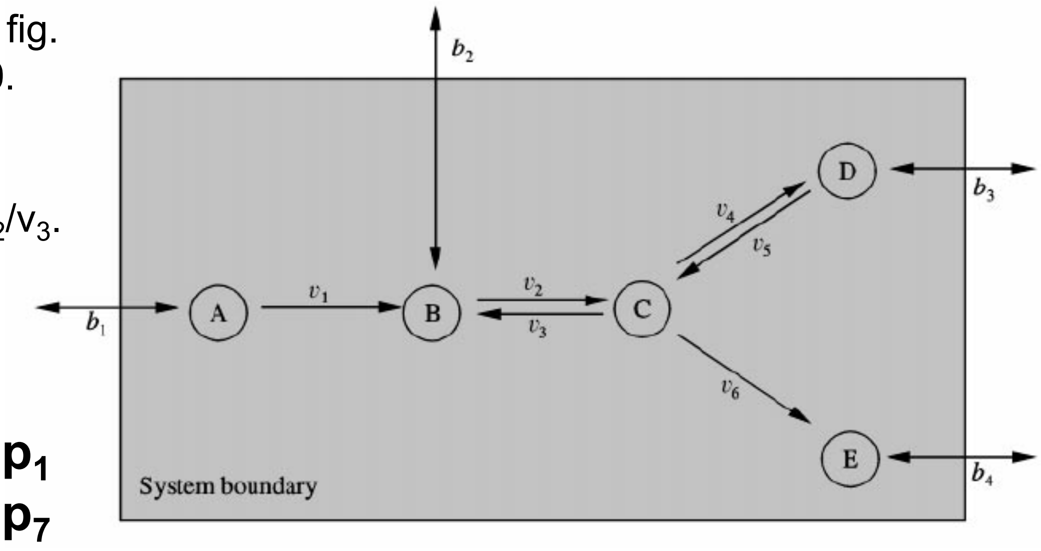
v_1	v_2	v_3	v_4	v_5	v_6	b_1	b_2	b_3	b_4	
1	0	0	0	0	0	-1	1	0	0	$\mathbf{p}_1$
0	1	1	0	0	0	0	0	0	0	$\mathbf{p}_7$
0	1	0	1	0	0	0	-1	1	0	$\mathbf{p}_3$
0	1	0	0	0	1	0	-1	0	1	$\mathbf{p}_2$
0	0	1	0	1	0	0	1	-1	0	$\mathbf{p}_4$
0	0	0	1	1	0	0	0	0	0	$\mathbf{p}_6$
0	0	0	0	1	1	0	0	-1	1	$\mathbf{p}_5$

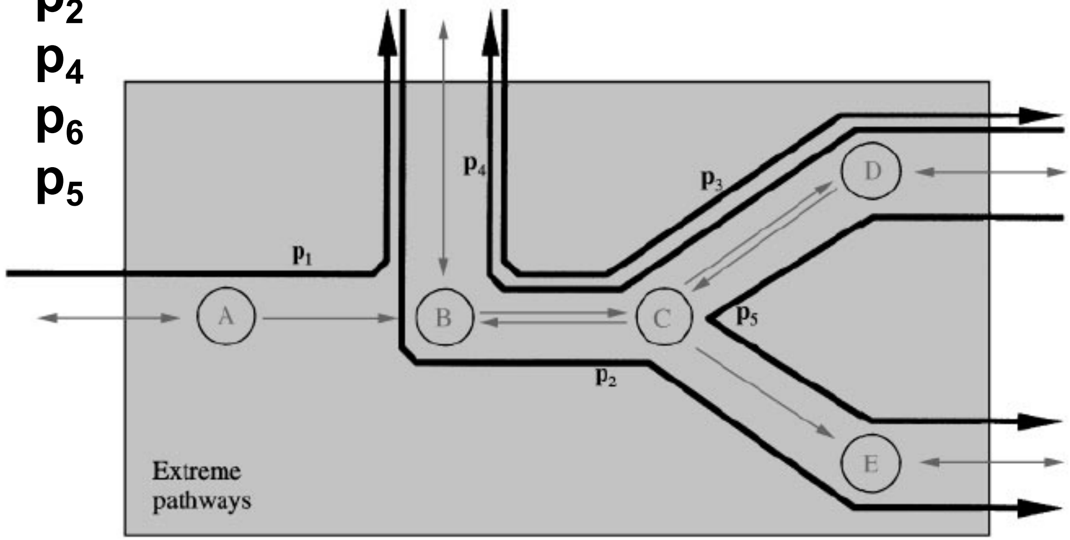
Schilling et al.
JTB 203, 229

Extreme Pathways for model system

2 pathways p_6 and p_7 are not shown in the bottom fig. because all exchange fluxes with the exterior are 0.
Such pathways have no net overall effect on the functional capabilities of the network.
They belong to the cycling of reactions v_4/v_5 and v_2/v_3.

v_1	v_2	v_3	v_4	v_5	v_6	b_1	b_2	b_3	b_4	
1	0	0	0	0	0	-1	1	0	0	p_1
0	1	1	0	0	0	0	0	0	0	p_7
0	1	0	1	0	0	0	-1	1	0	p_3
0	1	0	0	0	1	0	-1	0	1	p_2
0	0	1	0	1	0	0	1	-1	0	p_4
0	0	0	1	1	0	0	0	0	0	p_6
0	0	0	0	1	1	0	0	-1	1	p_5





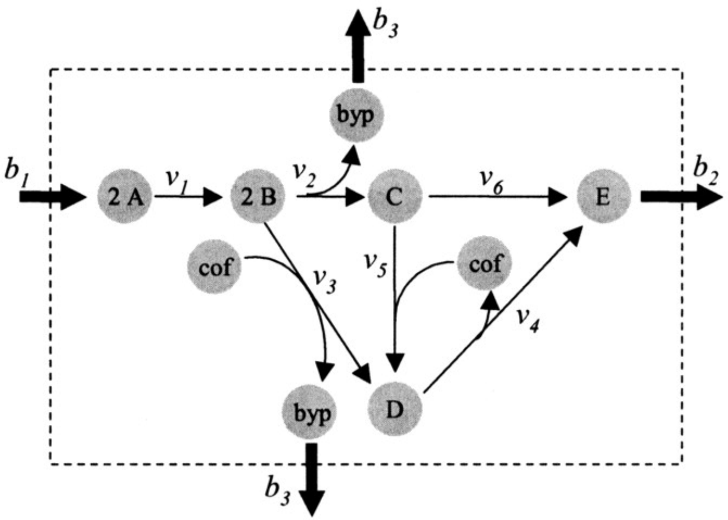
Schilling et al.
JTB 203, 229

How reactions appear in pathway matrix

In the matrix **P** of extreme pathways, each column is an EP and each row corresponds to a reaction in the network.

The numerical value of the *i,j*-th element corresponds to the relative flux level through the *i*-th reaction in the *j*-th EP.

Reaction Network



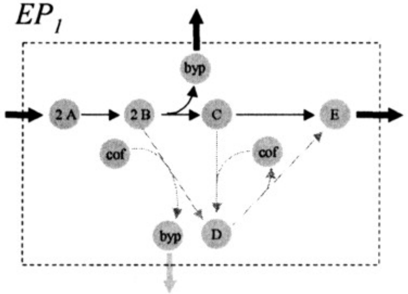
Stoichiometric Matrix

$$\mathbf{S} = \begin{array}{c} \\ \end{array}\begin{array}{c} \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & b_1 & b_2 & b_3 \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & +1 & 0 & 0 \\ +1 & -2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & +1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & +1 & 0 & +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & +1 & -1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{l} \\ A \\ B \\ C \\ D \\ E \\ byp \\ cof \end{array}$$

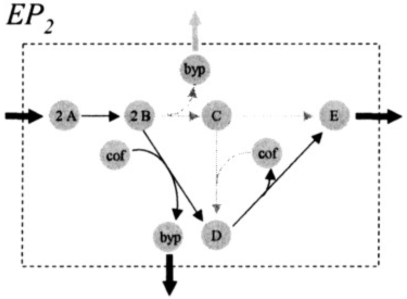
Pathway Matrix

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} EP_1 & EP_2 & EP_3 \end{array} \\ \begin{pmatrix} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array} \begin{array}{l} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ b_1 \\ b_2 \\ b_3 \end{array}$$

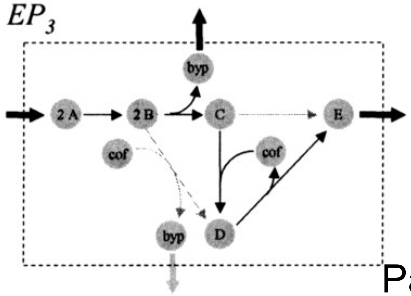
EP_1



EP_2



EP_3



Papin, Price, Palsson, Genome Res. 12, 1889 (2002)

Properties of pathway matrix

After normalizing **P** to a matrix with entries 0 or 1,
the symmetric **Pathway Length Matrix P**$_{LM}$ can be calculated:

$$\mathbf{P}_{LM} = \mathbf{P}^T \cdot \mathbf{P}$$

where the values along the diagonal correspond to the length of the EPs.

Pathway Length

$$\mathbf{P} = \begin{pmatrix} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \Rightarrow \tilde{\mathbf{P}} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \Rightarrow \tilde{\mathbf{P}}^T \bullet \tilde{\mathbf{P}} = \begin{array}{c} \begin{array}{ccc} EP_1 & EP_2 & EP_3 \end{array} \\ \begin{pmatrix} 6 & 4 & 5 \\ & 6 & 5 \\ & & 7 \end{pmatrix} \end{array} \begin{array}{l} \\ EP_1 \\ EP_2 \\ EP_3 \end{array}$$

Comments:

1) The lengths of EP_1, EP_2, and EP_3 are 6, 6, and 7, respectively, the highlighted diagonal elements of the final matrix.

2) EP_2 and EP_3 have a shared length of 5 (indicated by the circle). As seen in the schematics above, they share reactions v_1, v_4, b_1, b_2, and b_3.

The off-diagonal terms of **P**$_{LM}$ are the number of reactions that a pair of extreme pathways have in common.

Papin, Price, Palsson, Genome Res. 12, 1889 (2002)

Properties of pathway matrix

One can also compute a **reaction participation matrix $\mathbf{P}_{PM}$** from **P**:

$$\mathbf{P}_{PM} = \mathbf{P} \cdot \mathbf{P}^T$$

where the diagonal correspond to the number of pathways in which the given reaction participates.

Reaction Participation

C

$$\mathbf{P} = \begin{pmatrix} 2 & 2 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \Rightarrow \tilde{\mathbf{P}} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \Rightarrow \tilde{\mathbf{P}} \bullet \tilde{\mathbf{P}}^T =$$

$$\begin{array}{c} \begin{array}{ccccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & b_1 & b_2 & b_3 \end{array} \\ \left(\begin{array}{ccccccccc} 3 & 2 & 1 & 2 & 1 & 1 & 3 & 3 & 3 \\ & 2 & 0 & 1 & 1 & 1 & 2 & 2 & 2 \\ & & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ & & & 2 & 1 & 0 & 2 & \boxed{2} & 2 \\ & & & & 1 & 0 & 1 & 1 & 1 \\ & & & & & 1 & 1 & 1 & 1 \\ & & & & & & 3 & 3 & 3 \\ & & & & & & & 3 & 3 \\ & & & & & & & & 3 \end{array}\right) \end{array} \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ b_1 \\ b_2 \\ b_3 \end{array}$$

Comments:

1) The number of extreme pathways in which each reaction participates is indicated in the diagonal elements, as highlighted in the final matrix. These can then be expressed as a percentage of the total number of extreme pathways. For example, reaction v_1 has a participation value of 3. Since there are 3 extreme pathways, this can be expressed as 100% reaction participation.

2) The off diagonal terms can indicate correlated groups of reactions. Reactions v_1, b_1, b_2, and b_3 participate in 3 pathways. They also have a shared participation of 3, meaning they act as a correlated group (indicated by circles).

Papin, Price, Palsson, Genome Res. 12, 1889 (2002)

EP Analysis of *H. pylori* and *H. influenza*

Amino acid synthesis in *Heliobacter pylori* vs. *Heliobacter influenza* studied by EP analysis.

Table 4. Summary of the Statistical Analyses of Extreme Pathway Lengths

H. pylori Target product	Number of EPs	Pathway length: average	maximum	minimum	coefficient of variation
Asparagine	340	44	54	28	15%
Aspartic Acid	491	43	52	24	14%
Cysteine	1022	59	71	45	10%
Glutamine	315	41	53	23	18%
Glutamic Acid	493	41	53	25	17%
Glycine	377	51	60	38	10%
Lysine	611	54	66	39	12%
Proline	867	43	56	15	16%
Serine	355	45	54	33	12%
Threonine	469	48	60	31	14%
Tryptophan	1958	64	73	51	6%
Tyrosine	1008	58	68	44	7%
Equimolar Amino Acids	6032	106	112	99	2%
E. coli Ratio Amino Acids	5553	106	112	99	2%

H. influenzae Target product	Number of EPs	Pathway length: average	maximum	minimum	coefficient of variation
Alanine	1739	36	49	18	10%
Asparagine	445	39	52	29	13%
Aspartic Acid	690	35	49	27	14%
Glutamine	690	37	46	28	11%
Glycine	456	39	48	35	7%
Histidine	1507	65	74	61	3%
Isoleucine	1480	47	61	37	9%
Leucine	3884	42	55	31	10%
Lysine	1168	47	61	37	9%
Methionine	1343	48	63	40	8%
Phenylalanine	1758	51	64	43	7%
Proline	2624	38	51	25	11%
Serine	690	37	50	30	10%
Threonine	1318	42	55	32	10%
Tryptophan	3540	58	69	49	6%
Tyrosine	1758	51	64	43	7%
Valine	1739	39	52	23	9%

The coefficient of variation is the standard deviation normalized to the average (expressed as a percent). Equimolar amino acids refers to the set of amino acids in equimolar ratios. *E. coli* ratio amino acids refers to the set of amino acids in ratios analogous to those seen in *E. coli* biomass. EPs, extreme pathways.

Table 1. Number of Reactions Involved in the Production of the Indicated Target Product

H. pylori Target product	Essential reactions	Utilized reactions
Tryptophan	32	105
Tyrosine	28	101
Cysteine	25	102
Glycine	22	97
Lysine	22	102
Serine	16	91
Threonine	14	96
Asparagine	13	91
Aspartic Acid	12	91
Proline	10	91
Glutamic Acid	7	91
Glutamine	6	91
Equimolar Amino Acids	85	140
E. coli Ratio Amino Acids	85	140

H. influenzae Target product	Essential reactions	Utilized reactions
Histidine	51	112
Tryptophan	41	108
Phenylalanine	36	108
Tyrosine	36	108
Methionine	34	106
Isoleucine	31	108
Lysine	31	108
Glycine	29	82
Threonine	26	103
Asparagine	25	98
Serine	25	97
Leucine	23	105
Aspartic Acid	22	97
Glutamine	21	102
Proline	18	103
Valine	17	102
Alanine	12	99

See Fig. 3 for the indicated network inputs and outputs. Essential reactions refers to the number of reactions that were used in every extreme pathway (region I in Fig. 4). Utilized reactions refers to the number of reactions that were used at least once in the set of extreme pathways for the production of the associated product (region II in Fig. 4). The individual amino acids are sorted in descending order according to the number of essential reactions. Equimolar amino acids refers to the set of amino acids in equimolar ratios. *E. coli* ratio amino acids refers to the set of amino acids in ratios analogous to those seen in *E. coli* biomass.

Papin, Price, Palsson, Genome Res. 12, 1889 (2002)

Summary – Extreme Pathways

Extreme Pathway Analysis is a standard technique for analysis of metabolic networks.

Number of EPs can become extremely large – hard to interpret.

EP is an excellent basis for studying systematic effects of reaction cut sets.

It will be very important to consider the interplay of metabolic and regulatory networks.