V14 Graph connectivity – Metabolic networks

In the first half of this lecture section, we use the theory of network flows to give constructive proofs of Menger‘s theorem.

These proofs lead directly to algorithms for determining the edge-connectivity and vertex-connectivity of a graph.

The strategy to prove Menger‘s theorems is based on properties of certain **networks** whose arcs all have **unit capacity**.

These **0-1 networks** are constructed from the original graph.

Determining the connectivity of a graph

Lemma 12.3.1. Let *N* be an *s-t* network such that

outdegree(s) > *indegree(s),*

indegree(t) > *outdegree (t),* and

outdegree(v) = *indegree(v)* for all other vertices *v.*

Then, there exists a directed *s-t* path in network *N.*

Proof. Let *W* be a longest directed trail (trail = walk without repeated edges; path = trail without repeated vertices) in network *N* that starts at source *s*, and let *z* be its terminal vertex.

If vertex *z* were not the sink *t,* then there would be an arc not in trail *W* that is directed from *z* (since *indegree(z) = outdegree(z)).*

But this would contradict the maximality of trail *W.*

Thus, *W* is a directed trail from source *s* to sink *t.*

If *W* has a repeated vertex, then a part of *W* determines a directed cycle, which can be deleted from *W* to obtain a shorter directed *s-t* trail.

This deletion step can be repeated until no repeated vertices remain, at which point, the resulting directed trail is an *s-t* path. □

Determining the connectivity of a graph

Proposition 12.3.2. Let *N* be an *s-t* network such that

outdegree(s) – indegree(s) = m = indegree(t) – outdegree (t),

and *outdegree(v) = indegree(v)* for all vertices $v \neq s,t$.

Then, there exist *m* disjoint directed *s-t* path in network *N.*

Proof. If *m* = 1, then there exists an open eulerian directed trail *T* from source *s* to sink *t* by Theorem 6.1.3.

Review: An eulerian trail in a graph is a trail that visits every edge of that graph exactly once.

Theorem 6.1.3. A connected digraph *D* has an open eulerian trail from vertex *x* to vertex *y* if and only if *indegree(x)* + 1 = *outdegree(x), indegree(y)* = *outdegree(y)* + 1, and all vertices except *x* and *y* have equal indegree and outdegree.
Euler proved that a necessary condition for the existence of Eulerian circuits is that all vertices in the graph have an even degree.

Theorem 1.5.2. Every open *x-y* walk *W* is either an *x-y* path or can be reduced to an *x-y* path.

Therefore, trail *T* is either an *s-t* directed path or can be reduced to an *s-t* path.

Determining the connectivity of a graph

By way of induction, assume that the assertion is true for $m = k$, for some $k \geq 1$, and consider a network N for which the condition holds for $m = k + 1$.
There exists a directed s-t path P by Lemma 12.3.1.

If the arcs of path P are deleted from network N, then the resulting network $N - P$ satisfies the condition of the proposition for $m = k$.

By the induction hypothesis, there exist k arc-disjoint directed s-t paths in network $N - P$. These k paths together with path P form a collection of $k + 1$ arc-disjoint directed s-t paths in network N. □

Basic properties of 0-1 networks

Definition A **0-1 network** is a capacitated network whose arc capacities are either 0 or 1.

Proposition 12.3.3. Let *N* be an *s-t* network such that *cap(e)* = 1 for every arc *e.* Then the value of a maximum flow in network *N* equals the maximum number of arc-disjoint directed *s-t* paths in *N.*

Proof: Let *f** be a maximum flow in network *N*, and let *r* be the maximum number of arc-disjoint directed *s-t* paths in *N.*
Consider the network *N** obtained by deleting from *N* all arcs *e* for which *f*(e)* = 0.
Then *f*(e)* = 1 for all arcs *e* in network *N*.*
It follows from the definition that for every vertex *v* in network *N*,*

$$\sum_{e \in Out(v)} f^*(e) = |Out(v)| = outdegree(v)$$

and

$$\sum_{e \in In(v)} f^*(e) = |In(v)| = indegree(v)$$

Basic properties of 0-1 networks

Thus by the definition of *val(f*)* and by the conservation-of-flow property,

outdegree(s) – indegree (s) = val(f) = indegree(t) – outdegree(t)*

and *outdegree(v) = indegree(v),* for all vertices $v \neq s,t$.

By Proposition 12.3.2., there are *val(f*)* arc-disjoint *s-t* paths in network *N*,* and hence, also in *N*, which implies that $val(f^*) \leq r$.

To obtain the reverse inequality, let $\{P_1, P_2, \ldots, P_r\}$ be the largest collection of arc-disjoint directed *s-t* paths in *N,* and consider the function $f: E_N \rightarrow R^+$ defined by

$$f(e) = \begin{cases} 1, & \text{if some path } P_i \text{ uses arc } e \\ 0, & \text{otherwise} \end{cases} .$$

Then *f* is a feasible flow in network *N,* with *val(f) = r.*

It follows that $val(f^*) \geq r$. □

Separating Sets and Cuts

Review from §5.3

Let *s* and *t* be distinct vertices in a graph *G.* An *s-t* **separating edge set** in *G* is a set of edges whose removal destroys all *s-t* paths in *G.*

Thus, an *s-t* separating edge set in *G* is an edge subset of E_G that contains at least one edge of every *s-t* path in *G.*

Definition: Let *s* and *t* be distinct vertices in a digraph *D.*
An *s-t* **separating arc set** in *D* is a set of arcs whose removal destroys all directed *s-t* paths in *D.*

Thus, an *s-t* separating arc set in *D* is an arc subset of E_D that contains at least one arc of every directed *s-t* path in digraph *D.*

Remark: For the degenerate case in which the original graph or digraph has no *s-t* paths, the empty set is regarded as an *s-t* separating set.

Separating Sets and Cuts

Proposition 12.3.4 Let N be an s-t network such that $cap(e) = 1$ for every arc e. Then the capacity of a minimum s-t cut in network N equals the minimum number of arcs in an s-t separating arc set in N.

Proof: Let $K^* = \langle V_s , V_t \rangle$ be a minimum s-t cut in network N, and let q be the minimum number of arcs in an s-t separating arc set in N.
Since K^* is an s-t cut, it is also an s-t separating arc set. Thus $cap(K^*) \geq q$.

To obtain the reverse inequality, let S be an s-t separating arc set in network N containing q arcs, and let R be the set of all vertices in N that are reachable from source s by a directed path that contains no arc from set S.

Then, by the definitions of arc set S and vertex set R, $t \notin R$, which means that $\langle R, V_N - R \rangle$ is an s-t cut.

Moreover, $\langle R, V_N - R \rangle \subseteq S$. Therefore

Separating Sets and Cuts

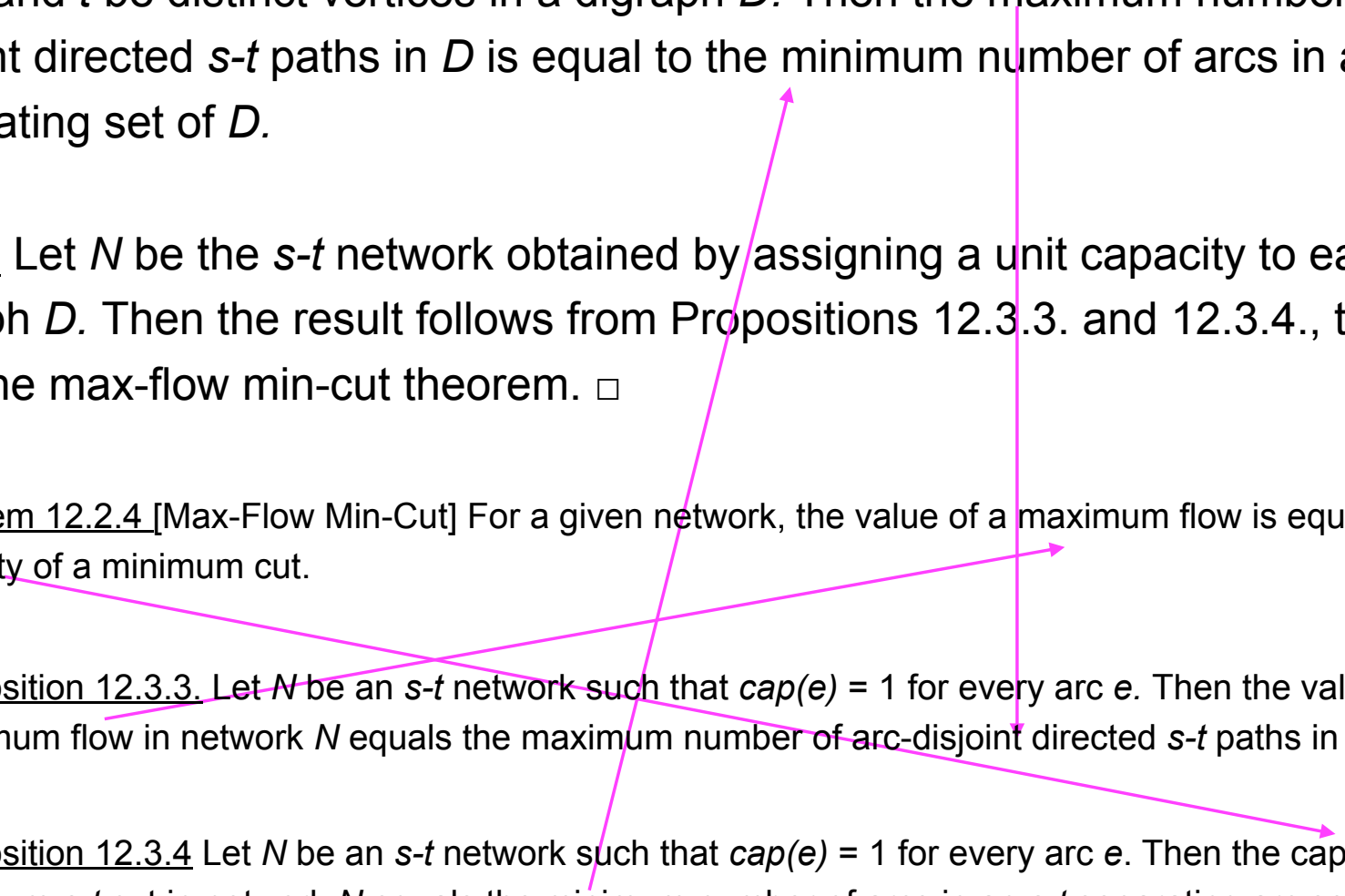
$$
\begin{aligned}
cap(K^*) &\leq cap\langle R, V_N - R\rangle && \text{since } K^* \text{ is a minimum } s-t \text{ cut} \\
&= \left|\langle R, V_N - R\rangle\right| && \text{since all capacities are } 1 \\
&\leq |S| && \text{since } \langle R, V_N - R\rangle \subseteq S \\
&= q
\end{aligned}
$$

which completes the proof. □

Arc and Edge Versions of Menger's Theorem Revisited

Theorem 12.3.5 [Arc form of Menger's theorem]
Let *s* and *t* be distinct vertices in a digraph *D.* Then the maximum number of arc-disjoint directed *s-t* paths in *D* is equal to the minimum number of arcs in an *s-t* separating set of *D.*

Proof: Let *N* be the *s-t* network obtained by assigning a unit capacity to each arc of digraph *D.* Then the result follows from Propositions 12.3.3. and 12.3.4., together with the max-flow min-cut theorem. □



Theorem 12.2.4 [Max-Flow Min-Cut] For a given network, the value of a maximum flow is equal to the capacity of a minimum cut.

Proposition 12.3.3. Let *N* be an *s-t* network such that *cap(e)* = 1 for every arc *e.* Then the value of a maximum flow in network *N* equals the maximum number of arc-disjoint directed *s-t* paths in *N.*

Proposition 12.3.4 Let *N* be an *s-t* network such that *cap(e)* = 1 for every arc *e*. Then the capacity of a minimum *s-t* cut in network *N* equals the minimum number of arcs in an *s-t* separating arc set in *N.*

Metabolic Networks - Introduction

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

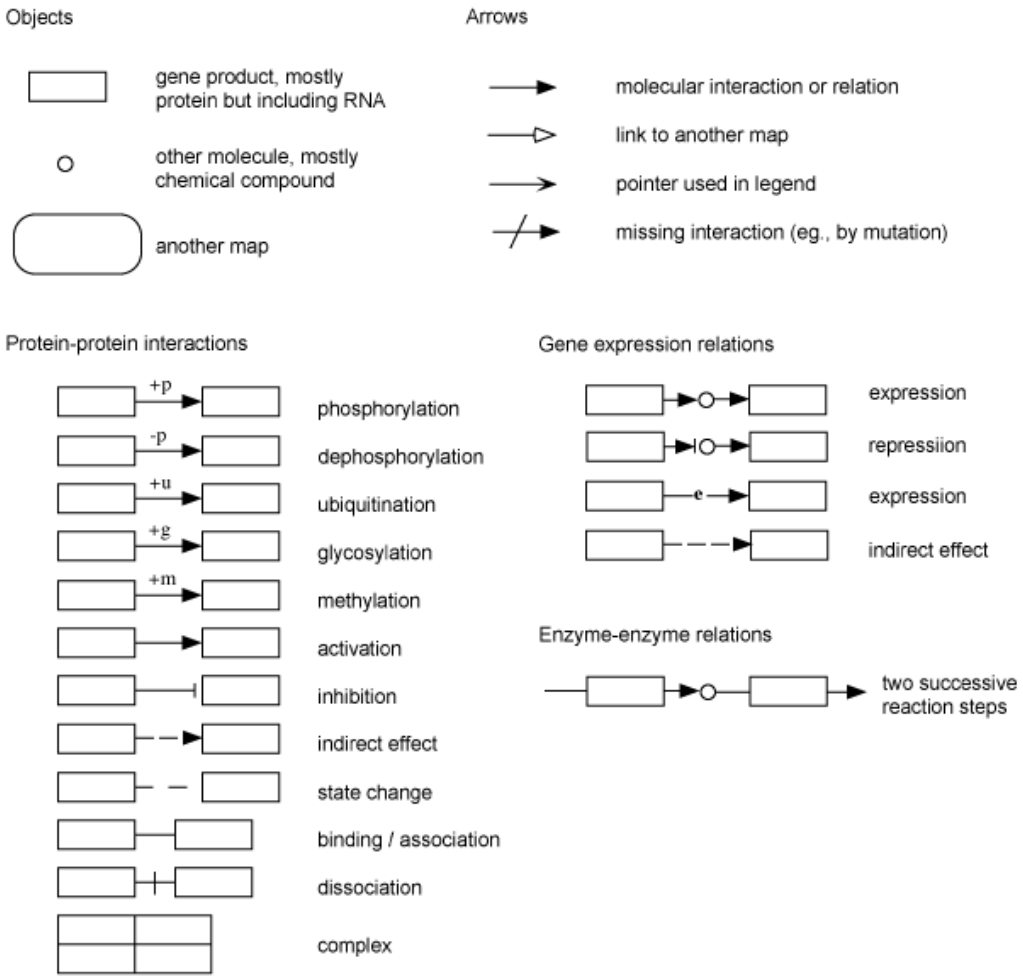
- stoichiometry/kinetics of classical biochemical **pathways** (glycolysis, TCA cycle, ...

- 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

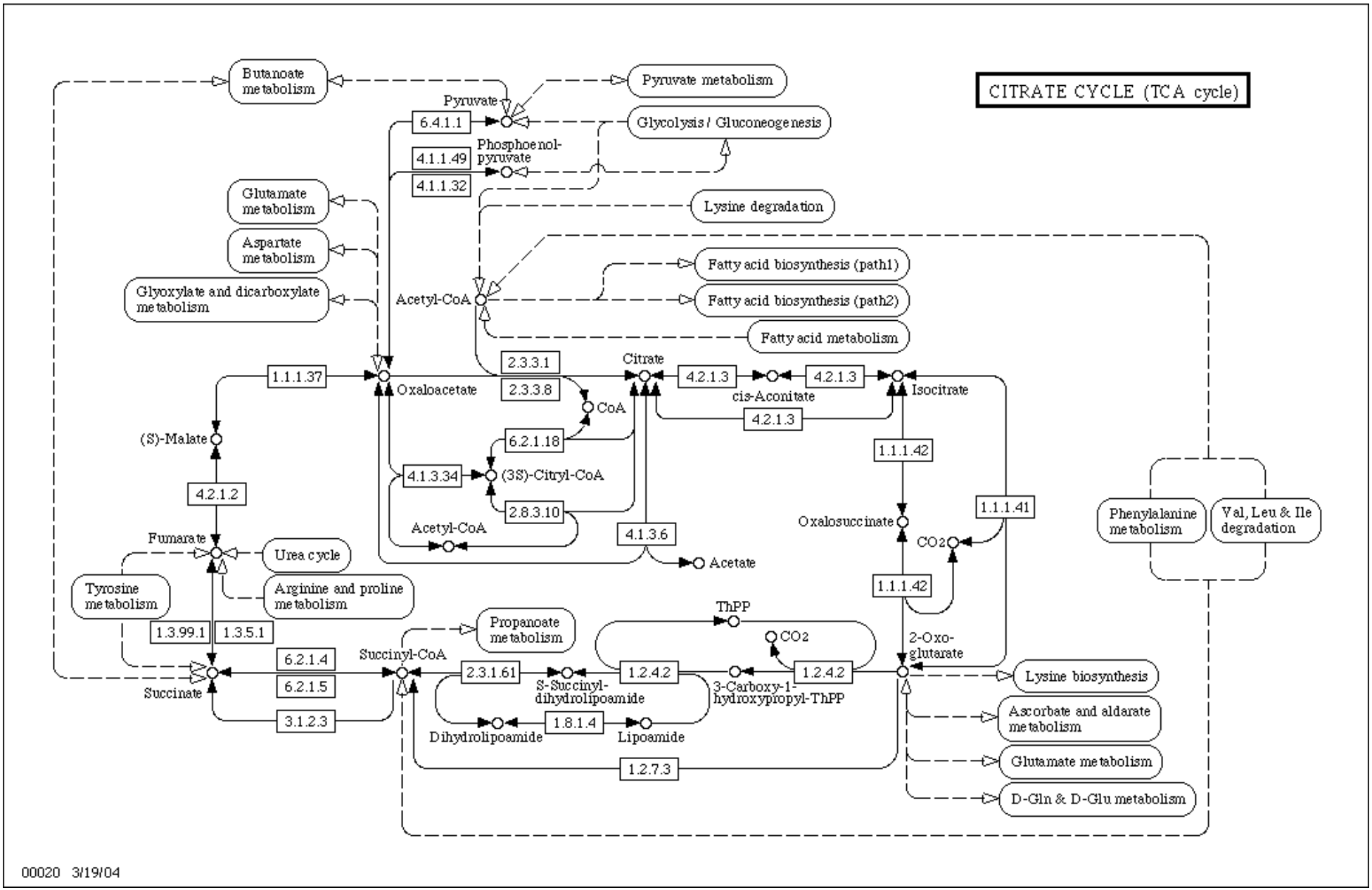
KEGG database



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.

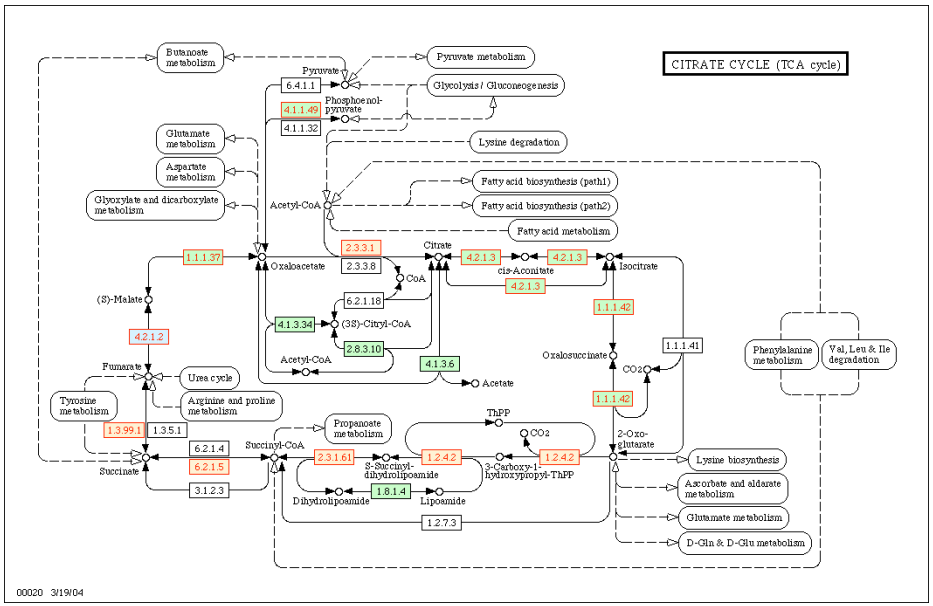
Organism-specific pathways (green-colored pathways) are computationally generated based on the KO assignment in individual genomes.

Citrate Cycle (TCA cycle) in E.coli

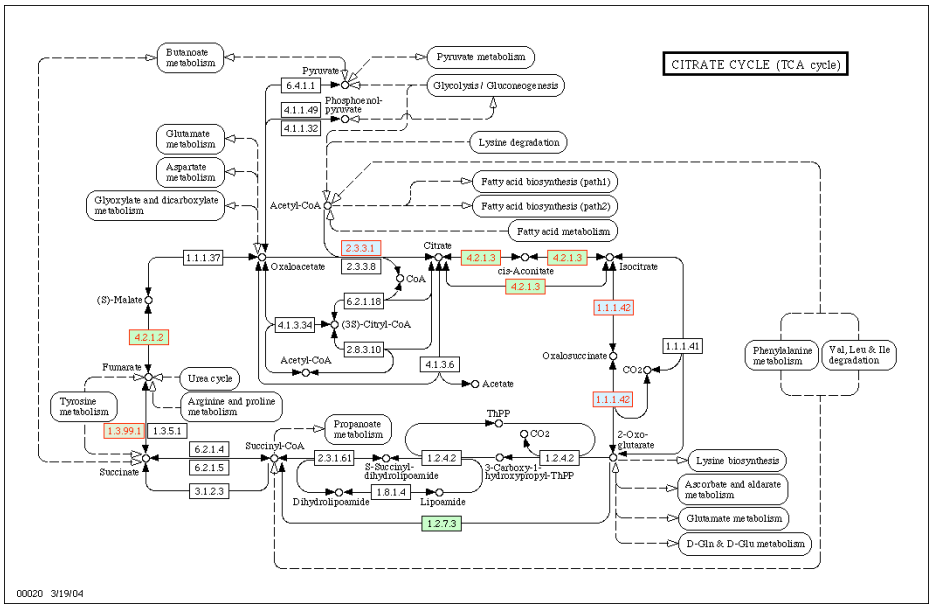


Citrate Cycle (TCA cycle) in different organisms

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



Citrate cycle (TCA cycle) - Helicobacter pylori 26695



Green/red: enzyme annotated in this organism

-KEGG: MAPK pathway

- KEGG: WNT signaling

EcoCyc Database

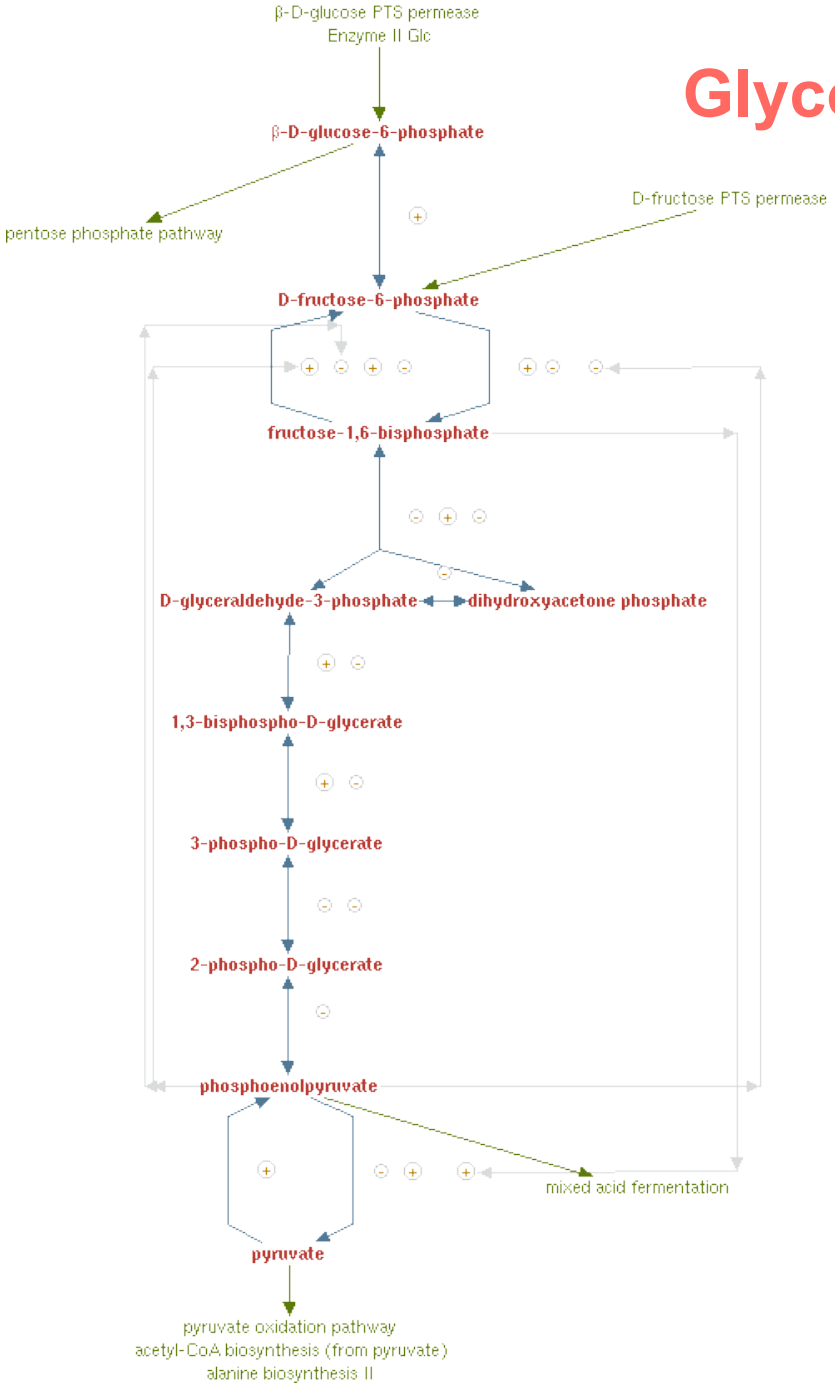
E.coli genome contains 4.7 million DNA bases.
How can we characterize the functional complement of *E.coli* and according to what criteria can we compare the biochemical networks of two organisms?

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

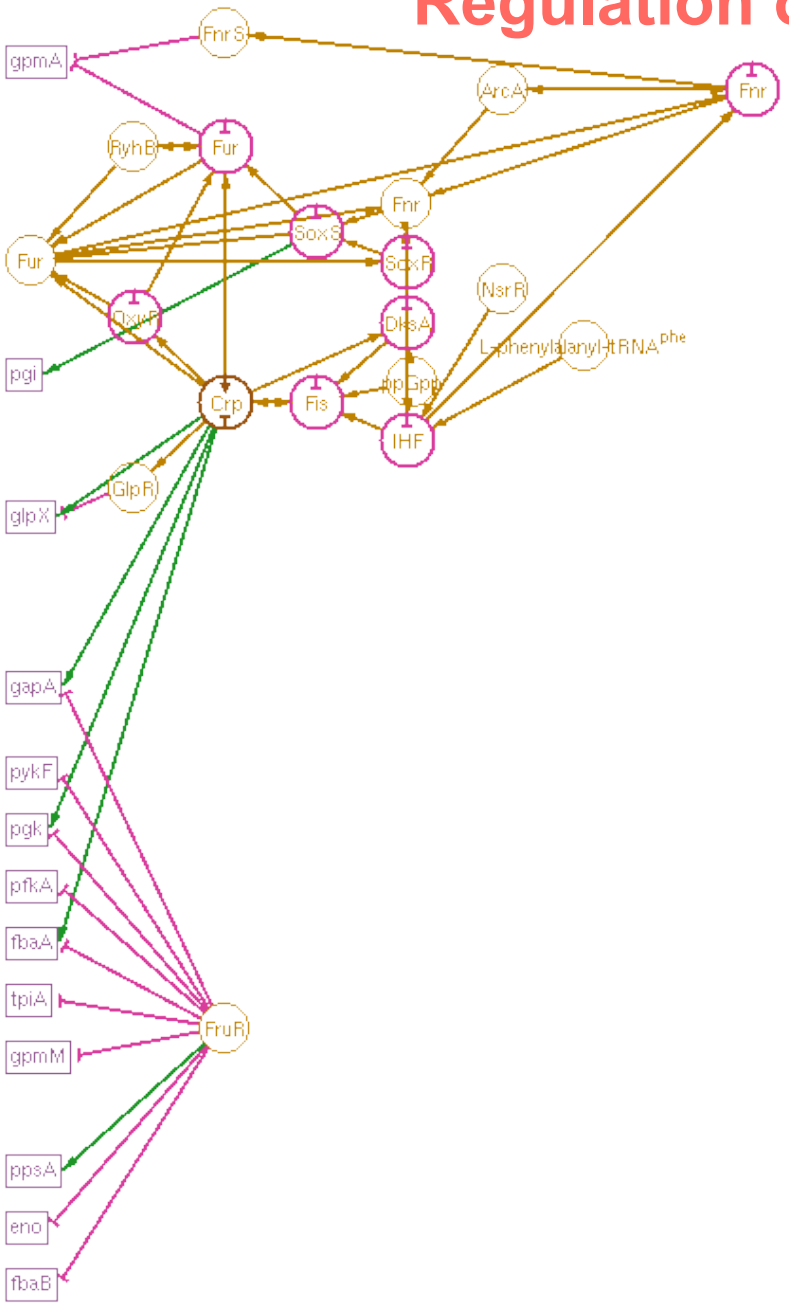


Blue arrows: biochemical reactions
clicking on arrow shows responsible enzyme

\+ 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

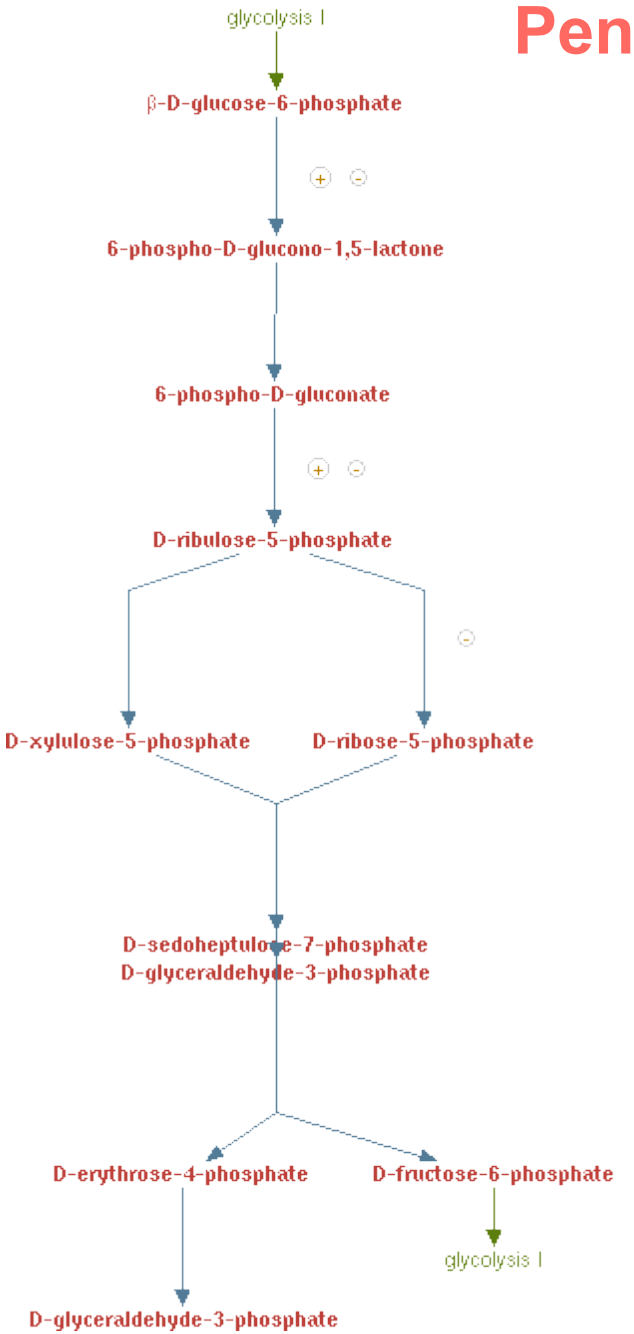
Green pointed arrows: activation of transcription;

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

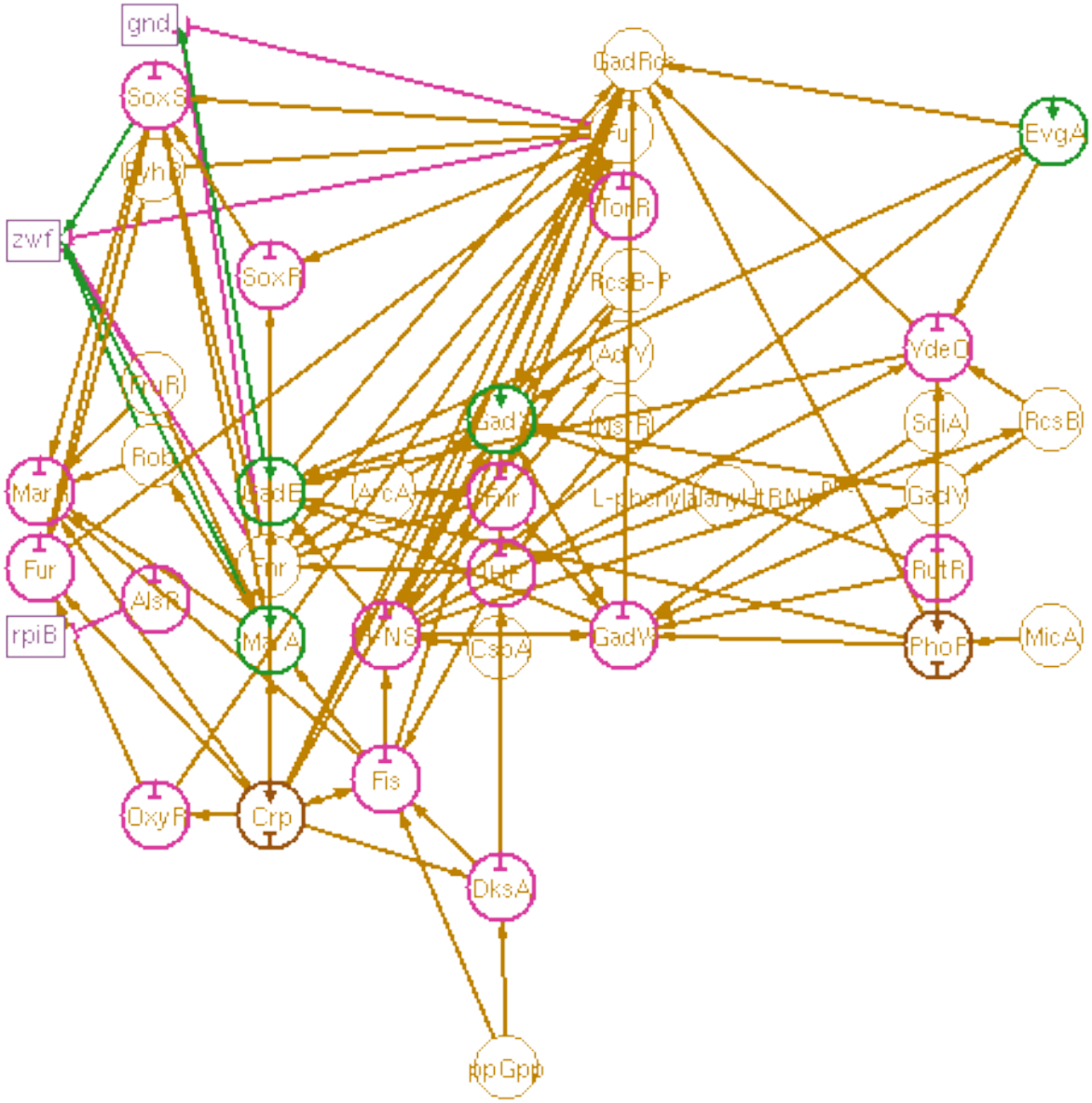


Blue arrows: biochemical reactions
clicking on arrow shows responsible enzyme

+ and - : activation and inhibition of enzymes

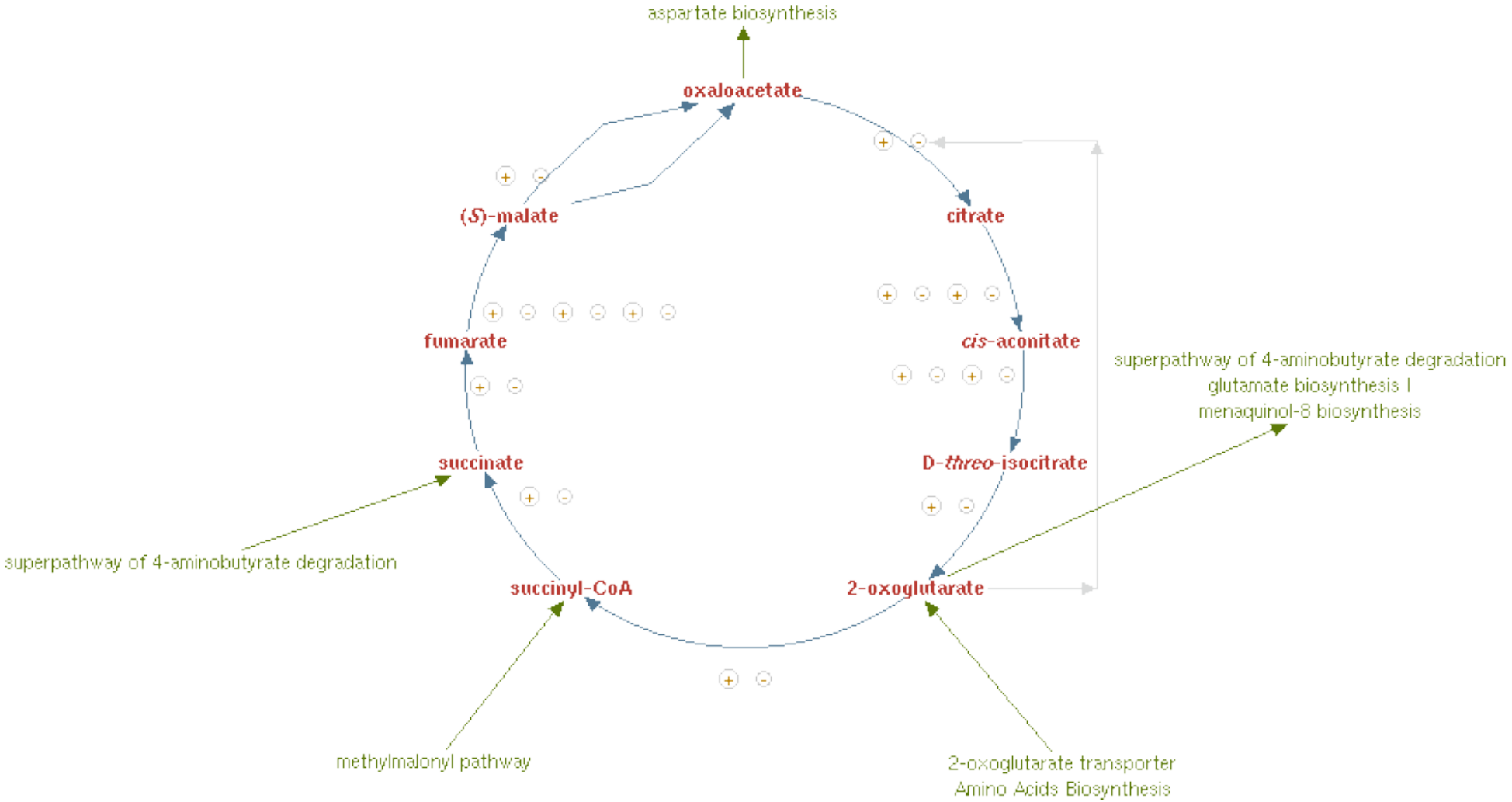
www.ecocyc.org

Regulation of Pentose Phosphate Pathway



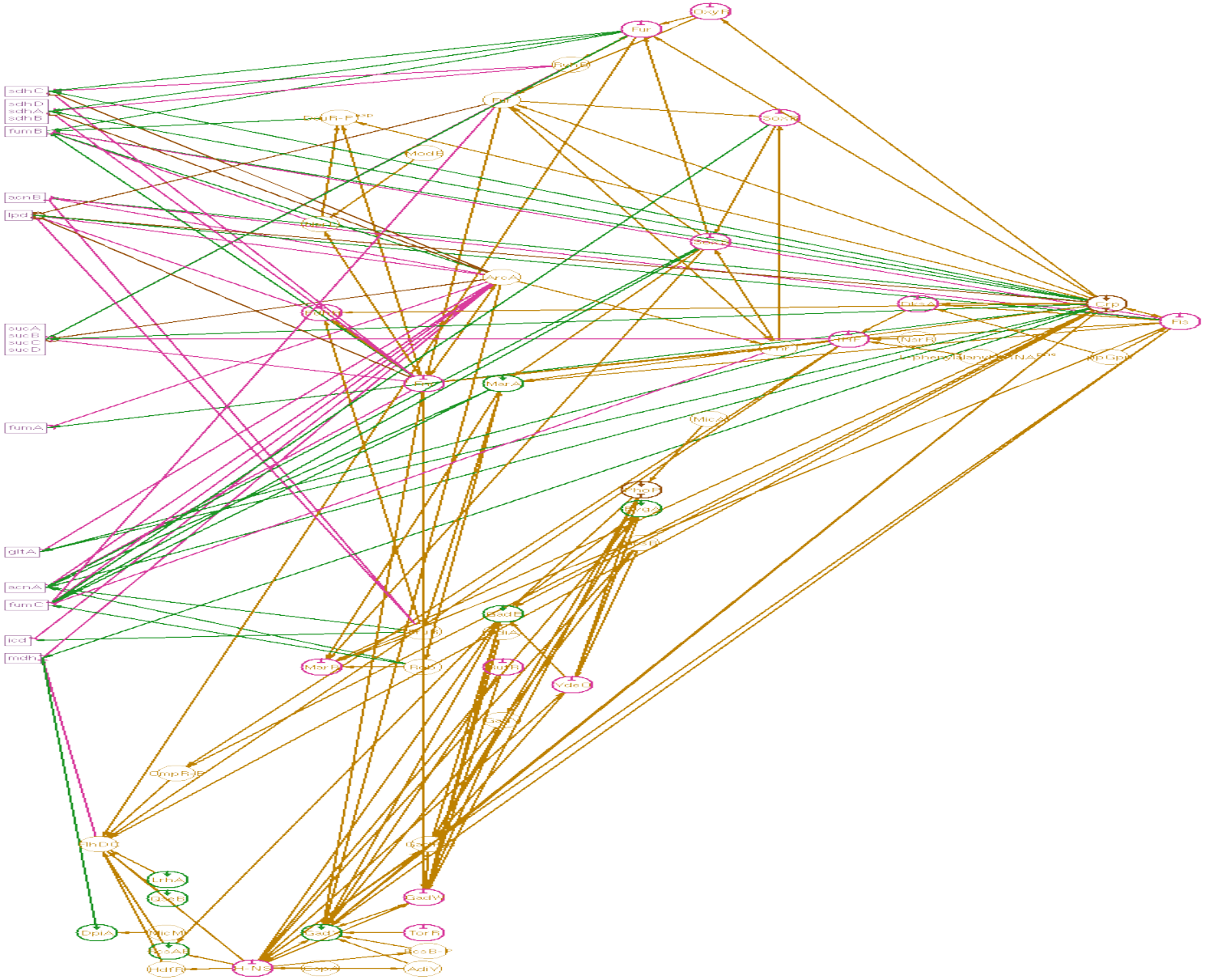
www.ecocyc.org

TCA cycle



www.ecocyc.org

Regulation of TCA cycle



www.ecocyc.org

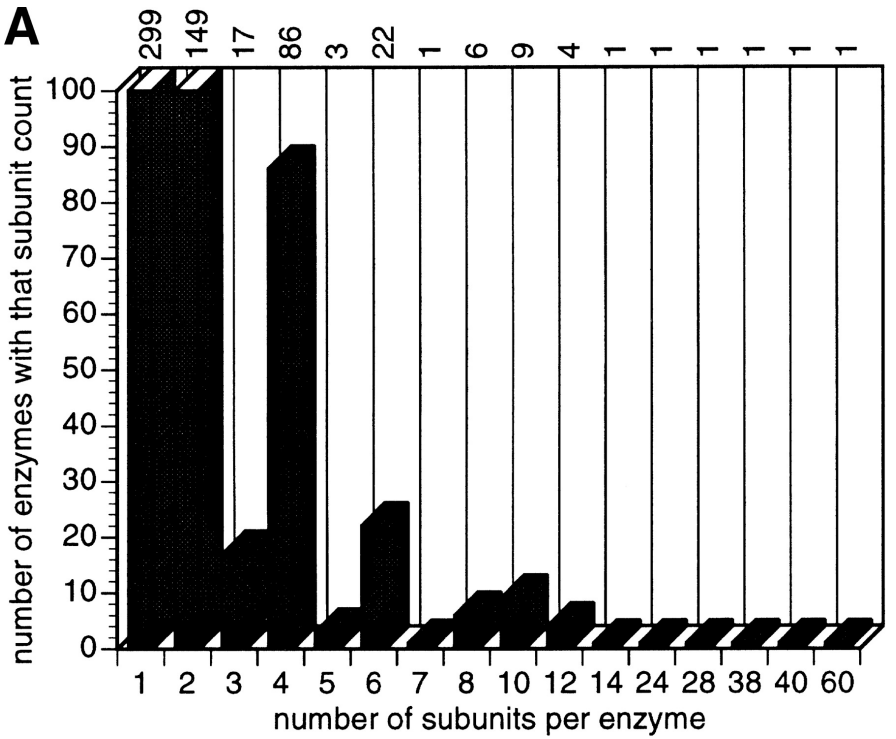
EcoCyc Analysis of E.coli Metabolism

E.coli genome contains 4391 predicted genes, of which 4288 code for proteins (4503 genes in Dec. 2011, 209 RNAs).

676 of these genes form 607 enzymes of *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.

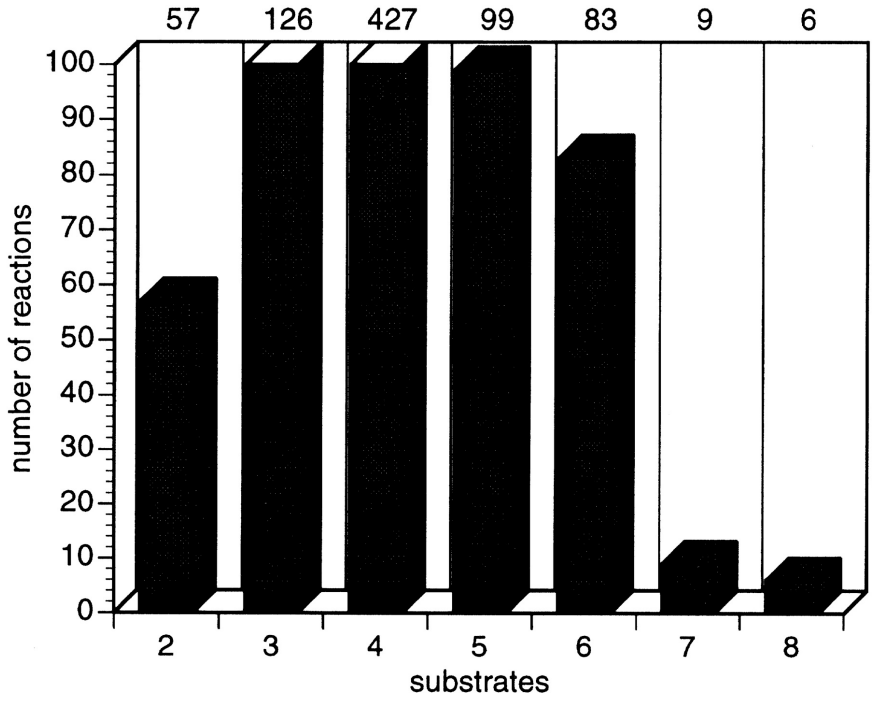
Ouzonis, Karp, Genome Res. 10, 568 (2000)

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

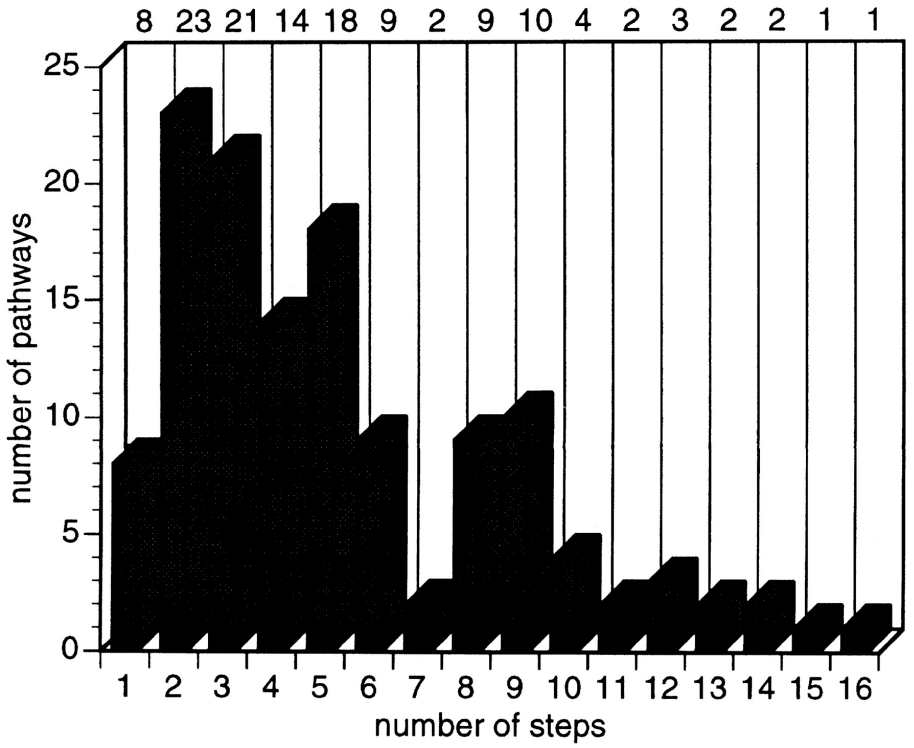
Pathways

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

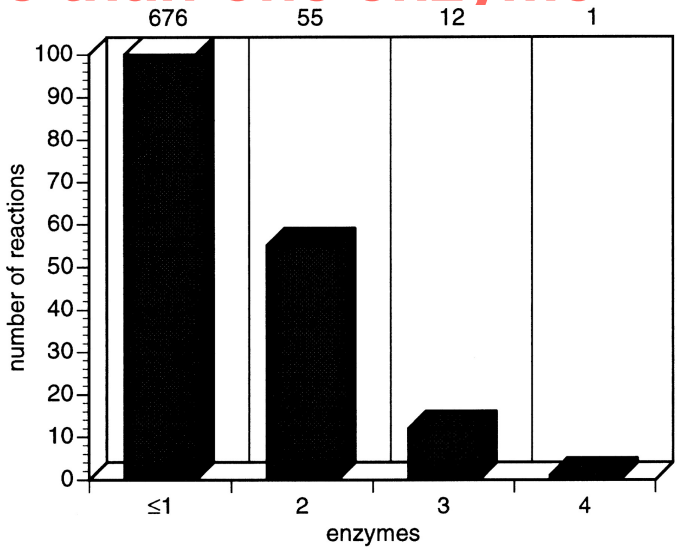
Table 3. Most Common Modulators, cofactors, and prosthetic groups of *E. coli* enzymes and Their Frequencies

A. Modulators (activators and inhibitors)				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**)

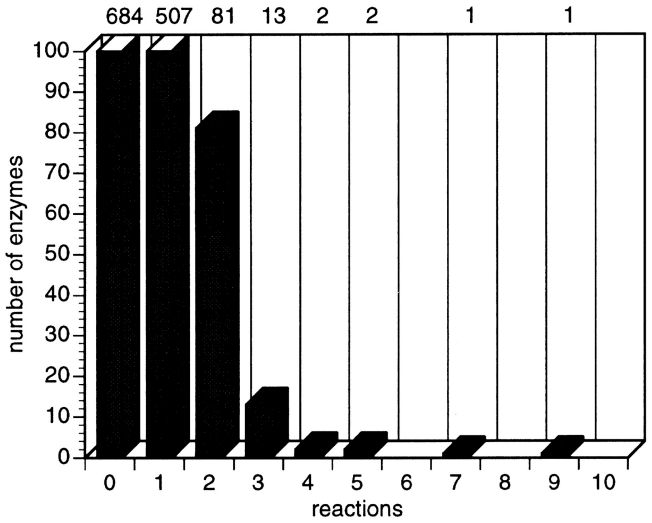
(2) the reaction is easily „invented“; therefore, there is more than one protein family that is independently able to perform the catalysis (**convergence**).

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

Enzymes that catalyze more than one reaction

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.

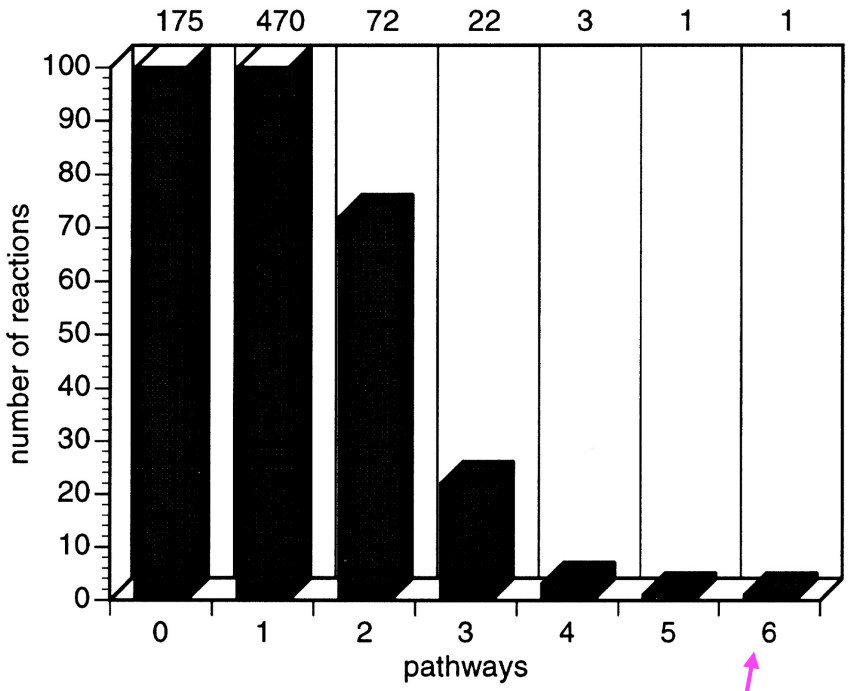


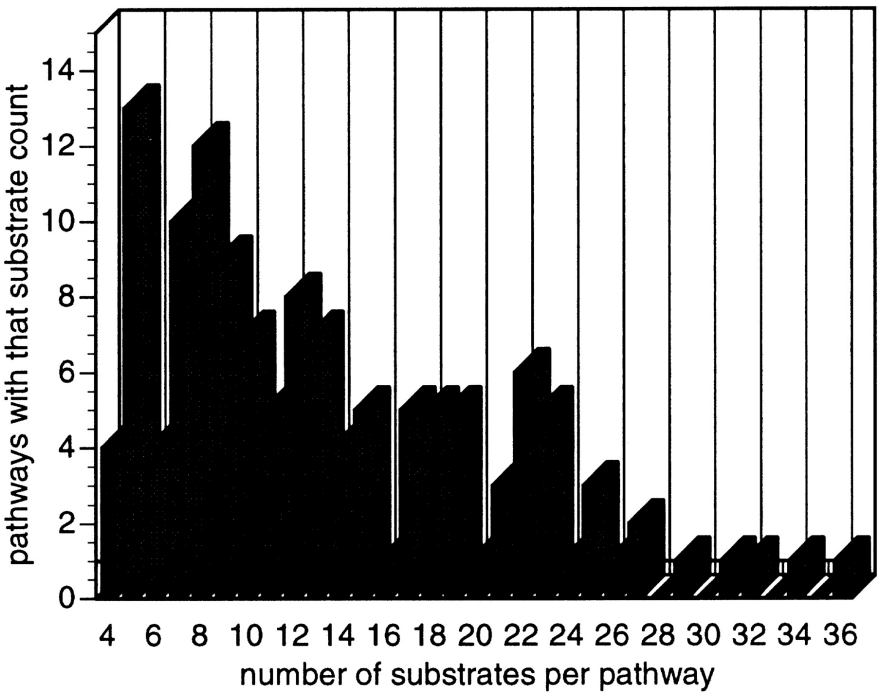
The enzymes that catalyze 7 and 9 reactions are purine nucleoside phosphorylase and nucleoside diphosphate kinase.

Take-home message: *The high proportion of multifunctional enzymes implies that the genome projects significantly underpredict multifunctional enzymes!*

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

Reactions participating in more than one pathway





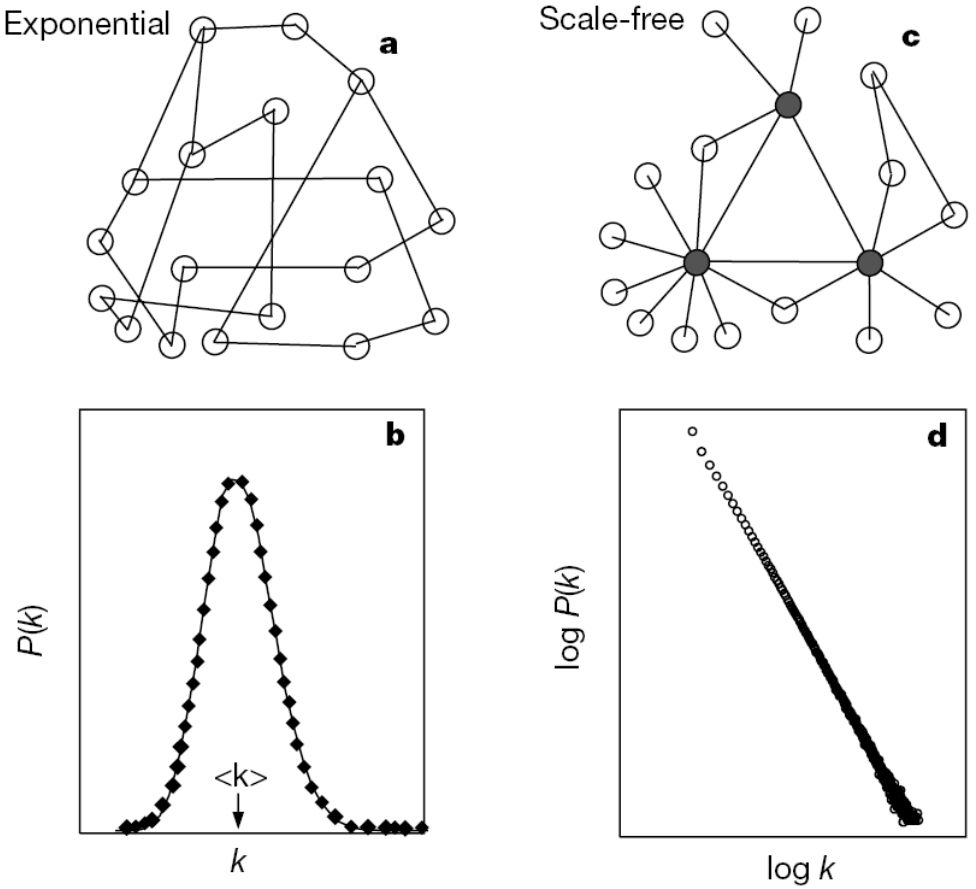
The 99 reactions belonging to multiple pathways appear to be the **intersection points** in the complex network of chemical processes in the cell.

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 ☺

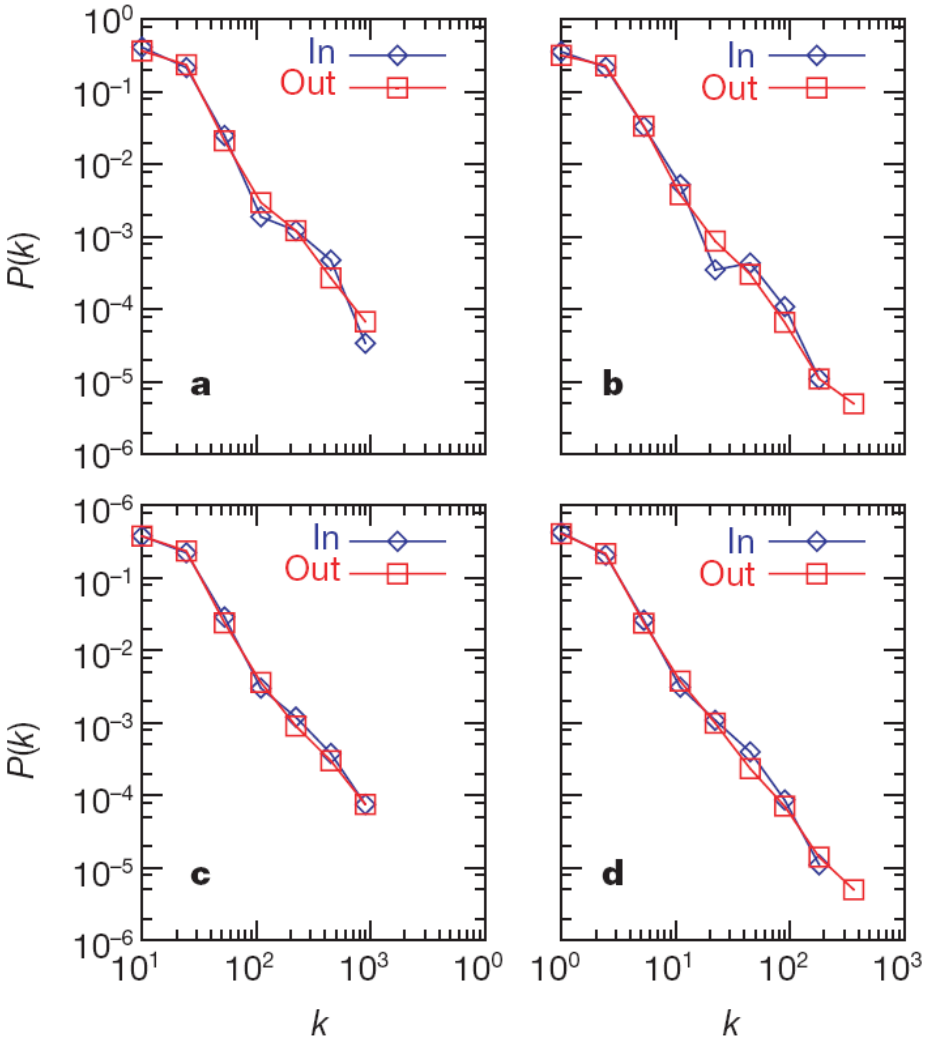
Attributes of generic network structures. **a**, Representative structure of the network generated by the Erdös–Rényi network model. **b**, The network connectivity can be characterized by the probability, $P(k)$, that a node has k links. For a random network $P(k)$ peaks strongly at $k = \langle k \rangle$ 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), shown on a log–log plot, counting 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).
d, The connectivity distribution averaged over 43 organisms.

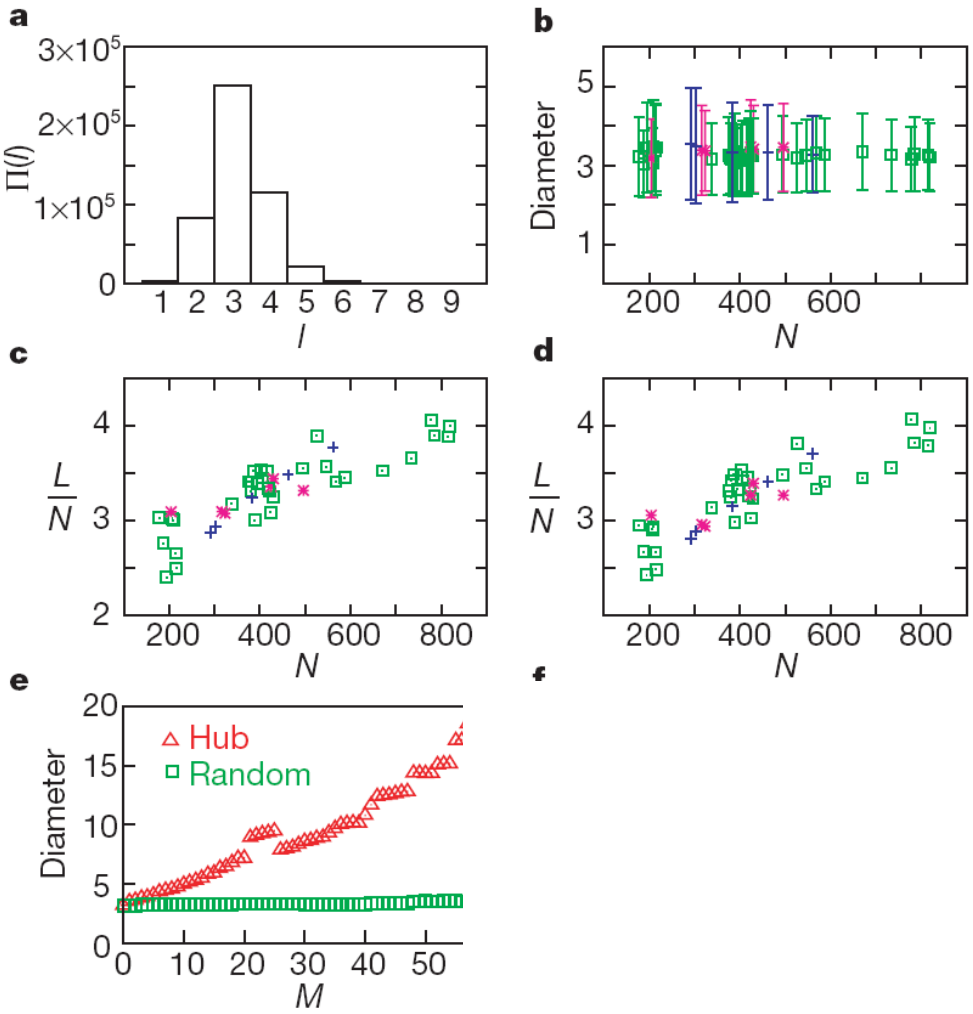


Jeong et al. Nature 407, 651 (2000)

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.
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*.

The horizontal axis in **b**– **d** denotes the number of nodes in each organism. **b**–**d**, Archaea (magenta), bacteria (green) and eukaryotes (blue) are shown.



Jeong et al. Nature 407, 651 (2000)

Conclusions about large-scale structure

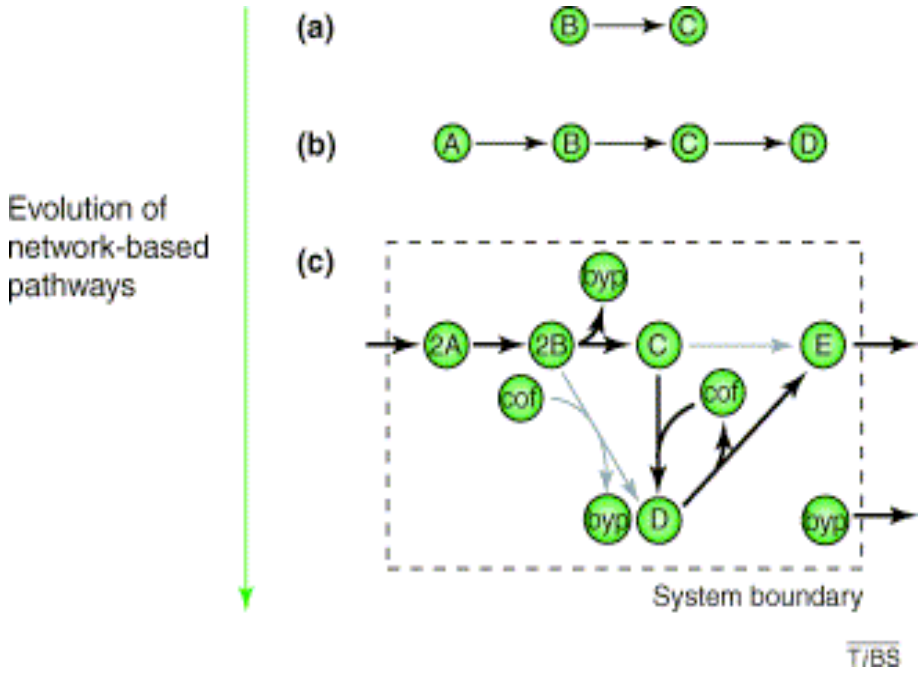
In a cell or microorganism, the processes that generate mass, energy, information transfer and cell-fate specification are seamlessly integrated through a complex network of cellular constituents and reactions.

A systematic comparative mathematical analysis of the metabolic networks of 43 organisms representing all 3 domains of life showed that, despite significant variation in their individual constituents and pathways, these metabolic networks have the same topological scaling properties and show striking similarities to the inherent organization of complex non-biological systems.

This may indicate that metabolic organization is not only identical for all living organisms, but also complies with the design principles of robust and error-tolerant scale-free networks, and may represent a common blueprint for the large-scale organization of interactions among all cellular constituents.

Jeong et al. Nature 407, 651 (2000)

Development of the network-based pathway paradigm



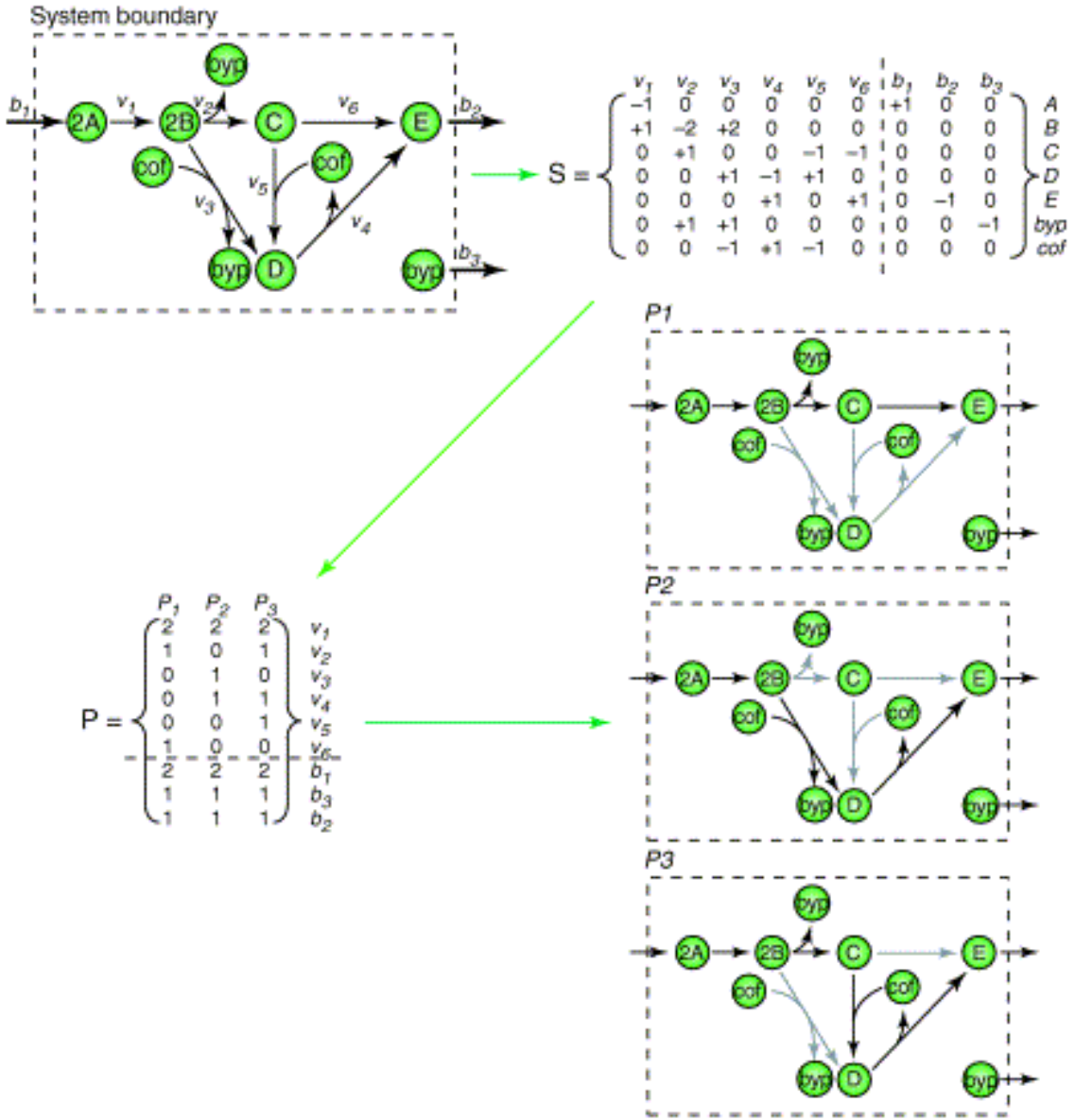
(a) With advanced biochemical techniques, years of research have led to the precise characterization of individual reactions. As a result, the **complete stoichiometries** of many metabolic reactions have been characterized.
(b) Most of these reactions have been grouped into `**traditional pathways**' (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

Stoichiometric matrix:
A matrix with reaction stochiometries as columns and metabolite participations 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 (see future lecture).



Papin et al. TIBS 28, 250 (2003)