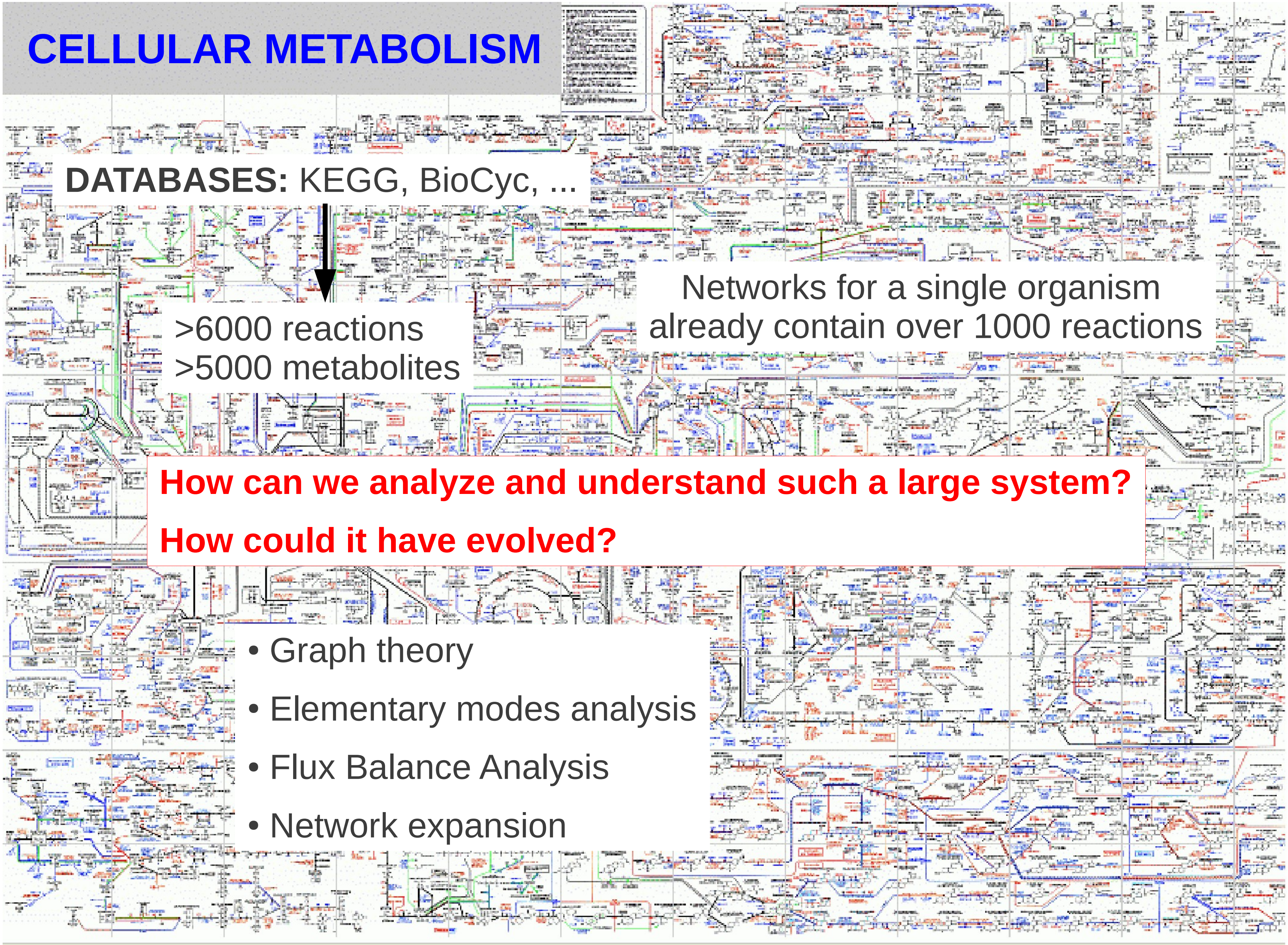
Network Expansion and Metabolic Scope

Oliver Ebenhöh

ICGEB, New Delhi, India, 19 October, 2012

CELLULAR METABOLISM



DATABASES: KEGG, BioCyc, ...

\>6000 reactions
\>5000 metabolites

Networks for a single organism already contain over 1000 reactions

How can we analyze and understand such a large system?

How could it have evolved?

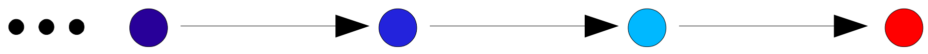
- Graph theory
- Elementary modes analysis
- Flux Balance Analysis
- Network expansion

Pathway Evolution

- No *fossil record* of metabolism available
- The evolutionary history must have left imprints in the present structure

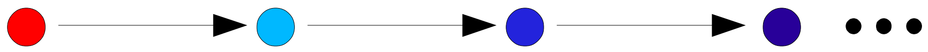
Existing hypotheses on metabolic evolution

retrograde evolution *(Horowitz, 1945)*



depletion of a required substrate

forward evolution

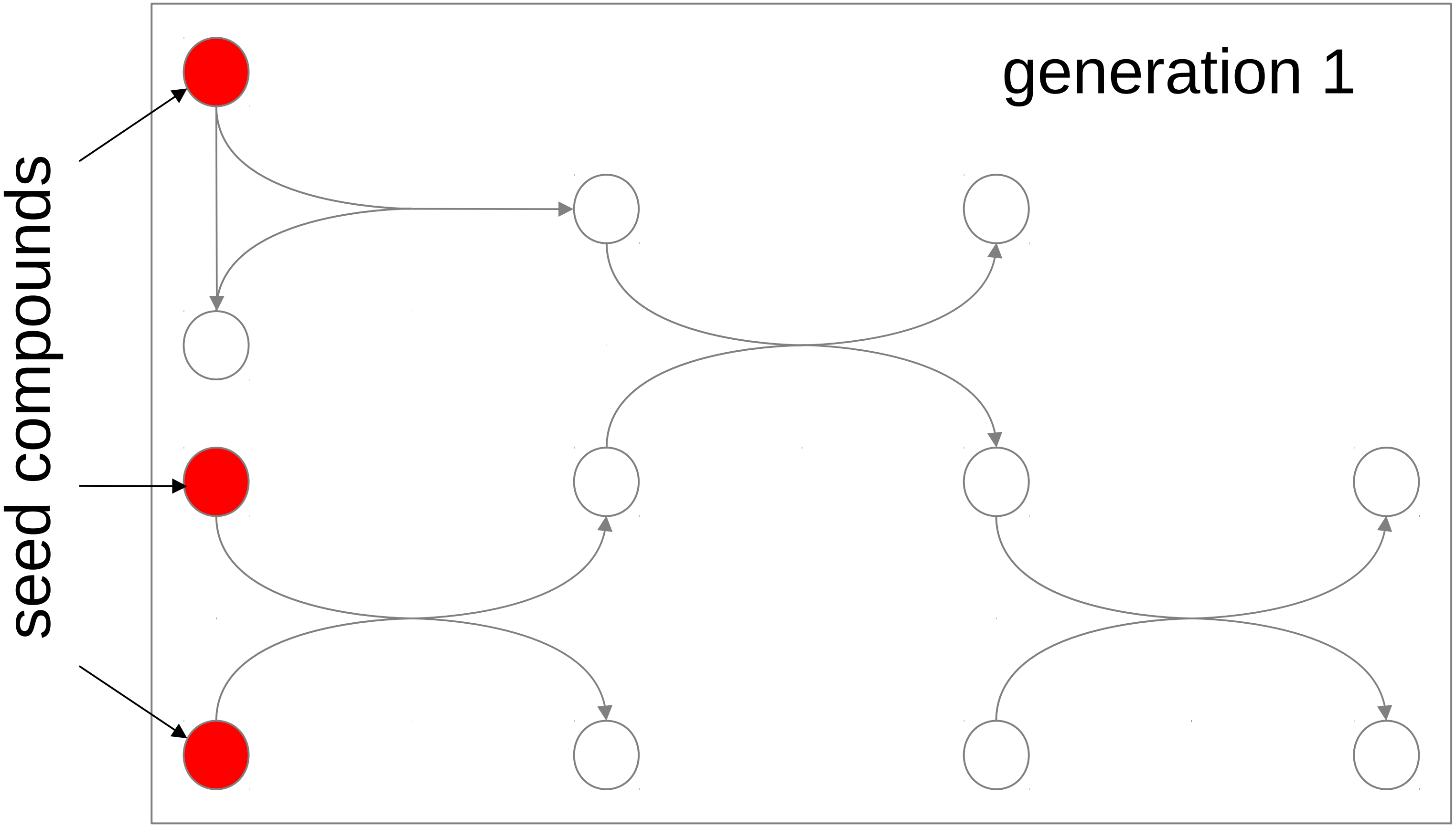


substrates in the environment trigger 'invention' of new metabolites

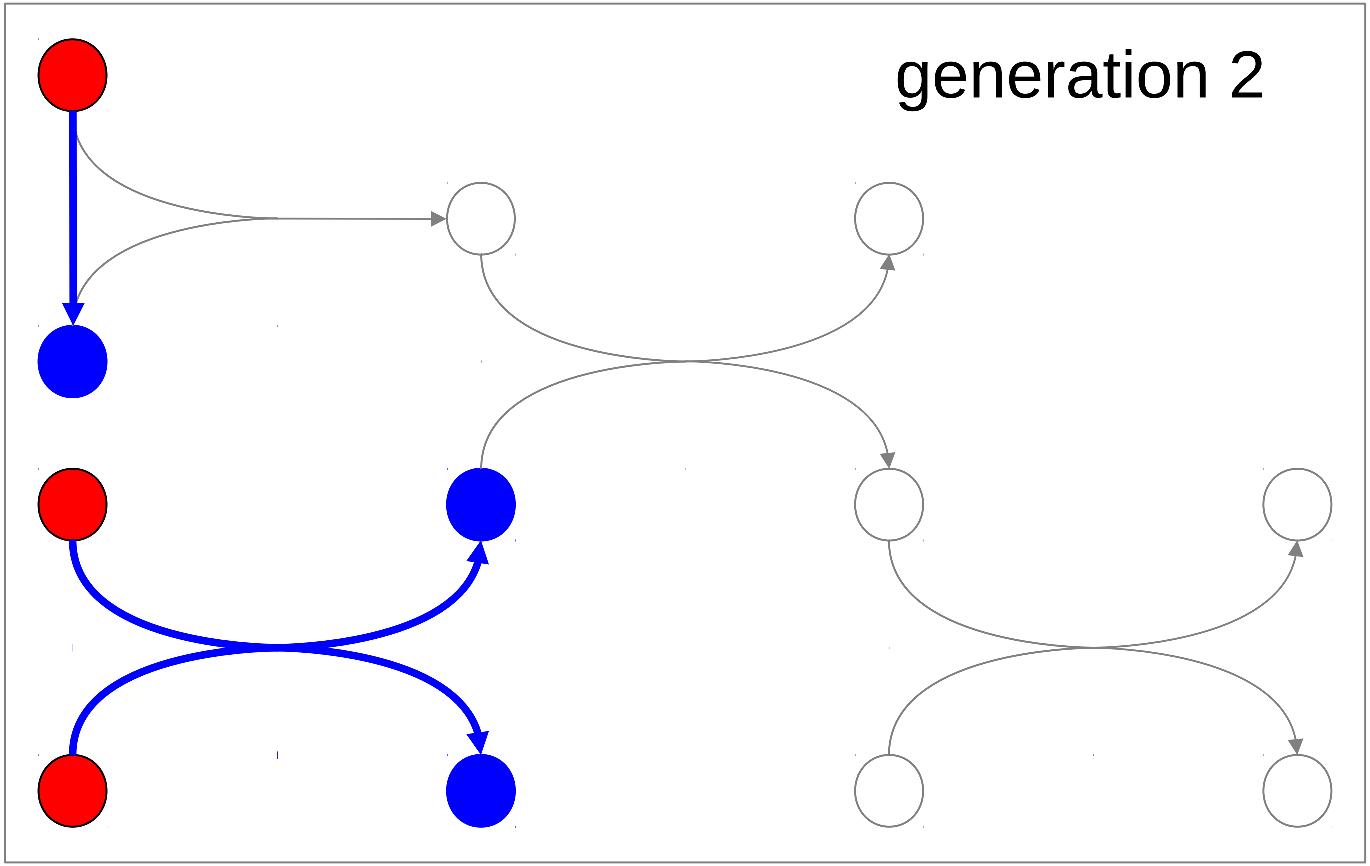
patchwork evolution (evolution to increased specificity)

(Jensen, 1976)

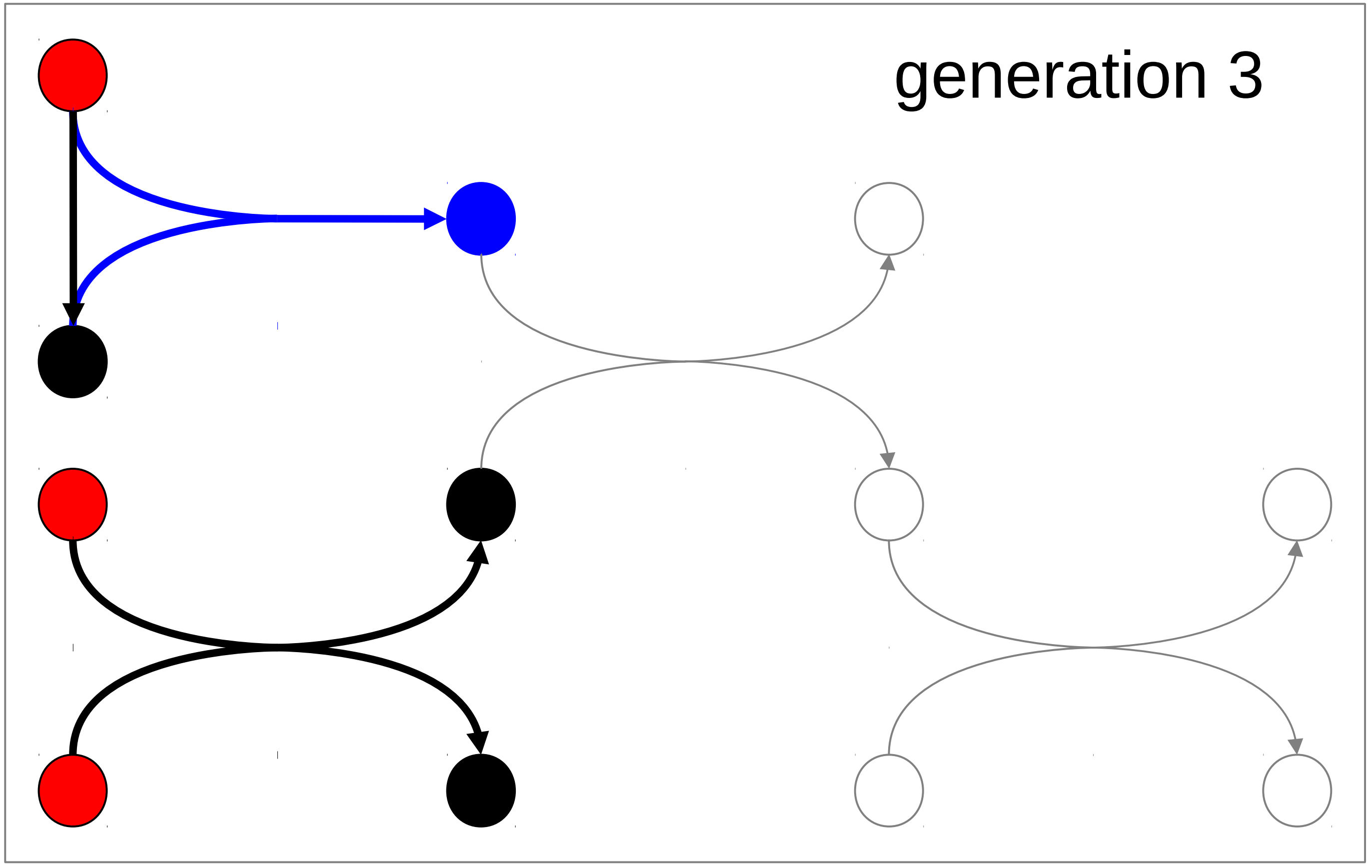
A simple forward approach: Network Expansion



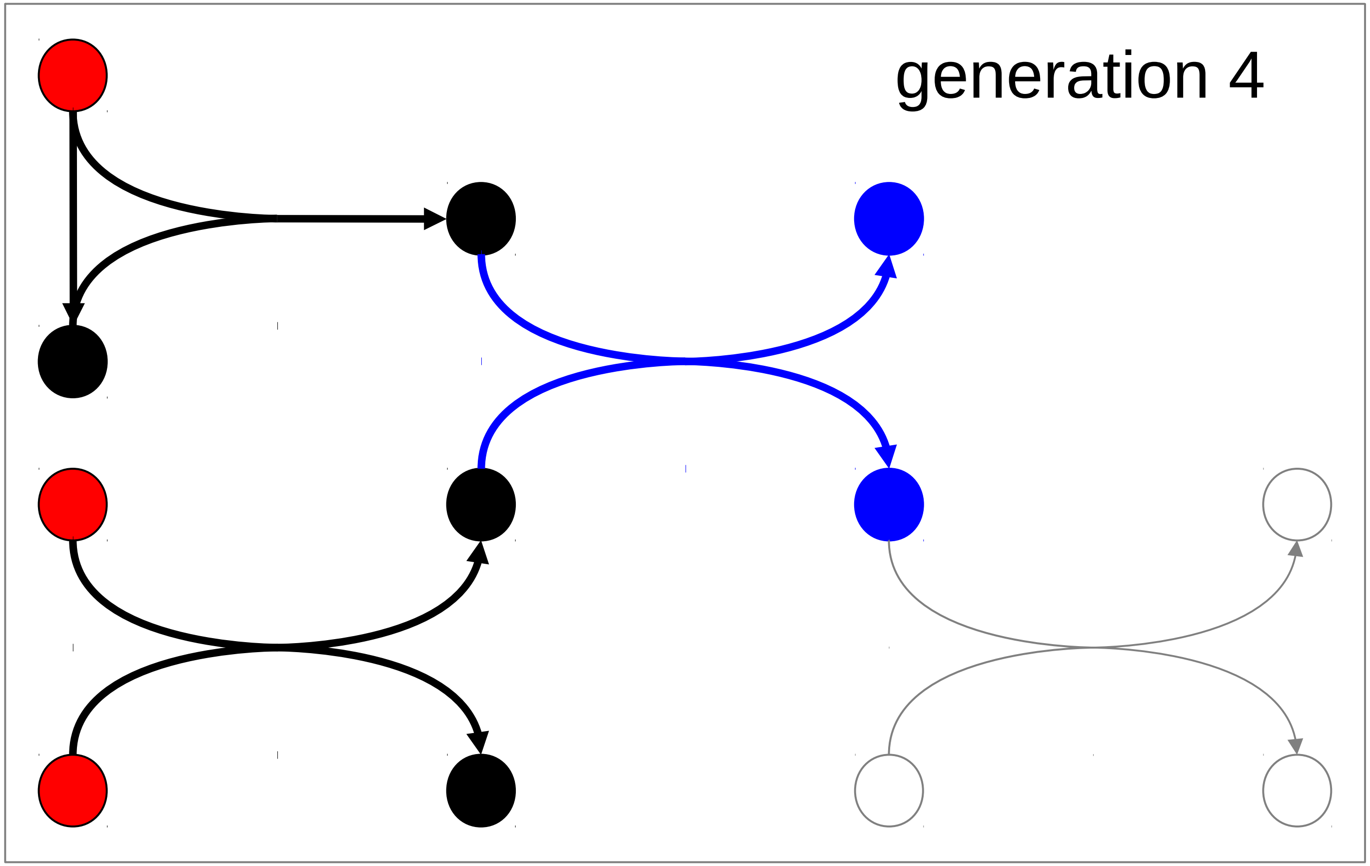
Concept of Network Expansion



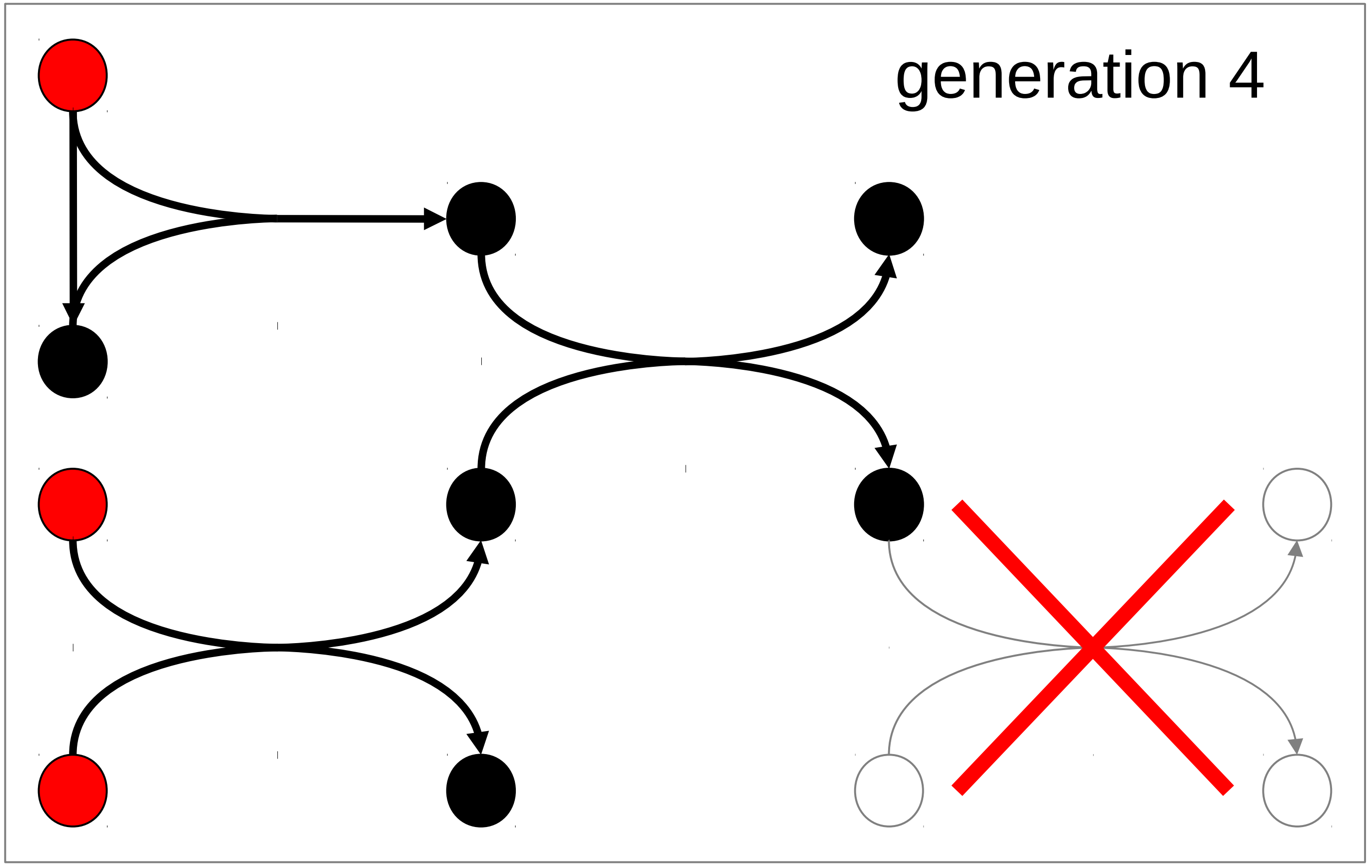
Concept of Network Expansion



Concept of Network Expansion

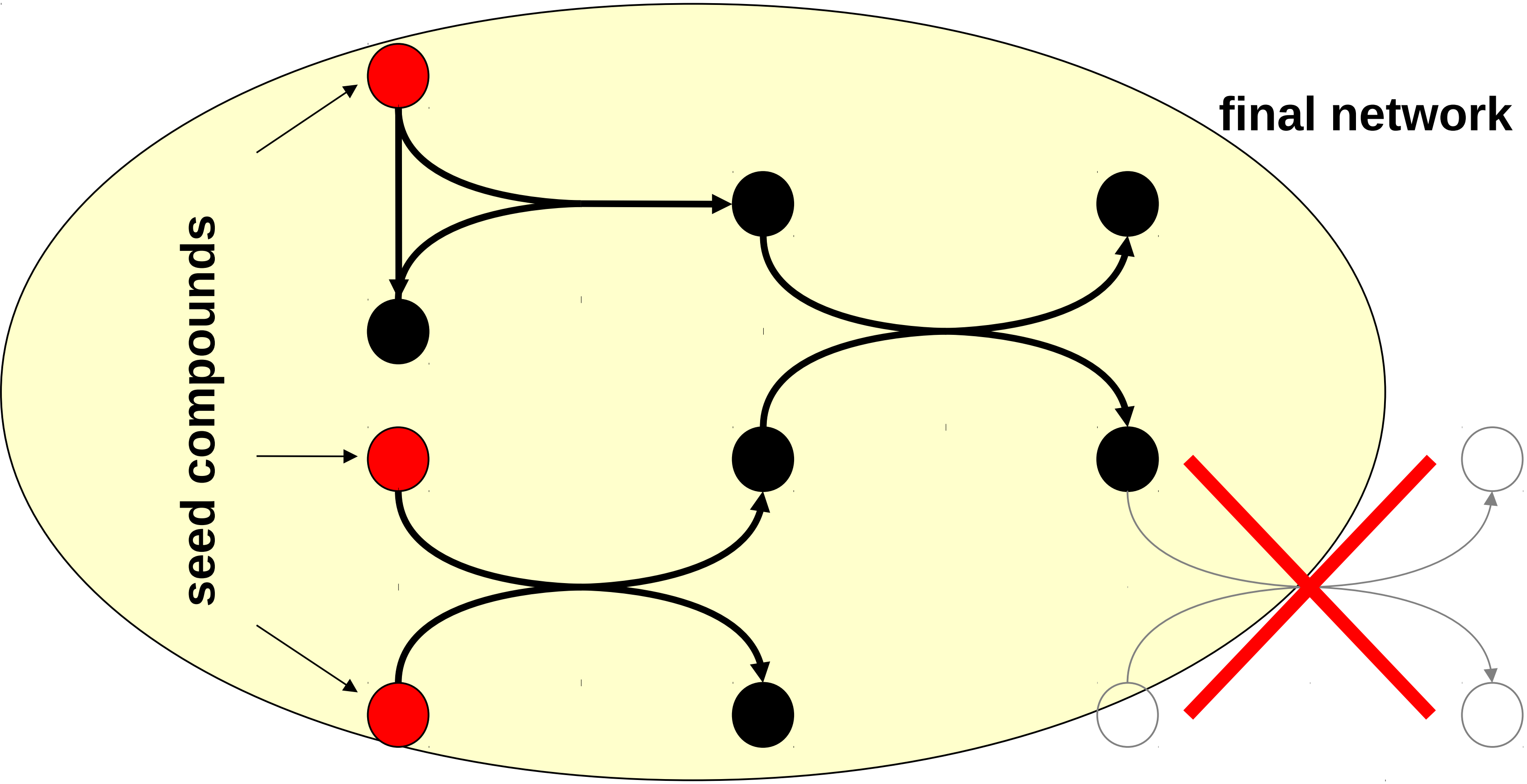


Concept of Network Expansion



Concept of Scopes

Scope: set of compounds that is reached by a network expansion



The Scope describes the synthesizing capacity of the metabolic network,
if it is provided with the seed compounds

Handorf, Ebenhöh & Heinrich, JME, 2005

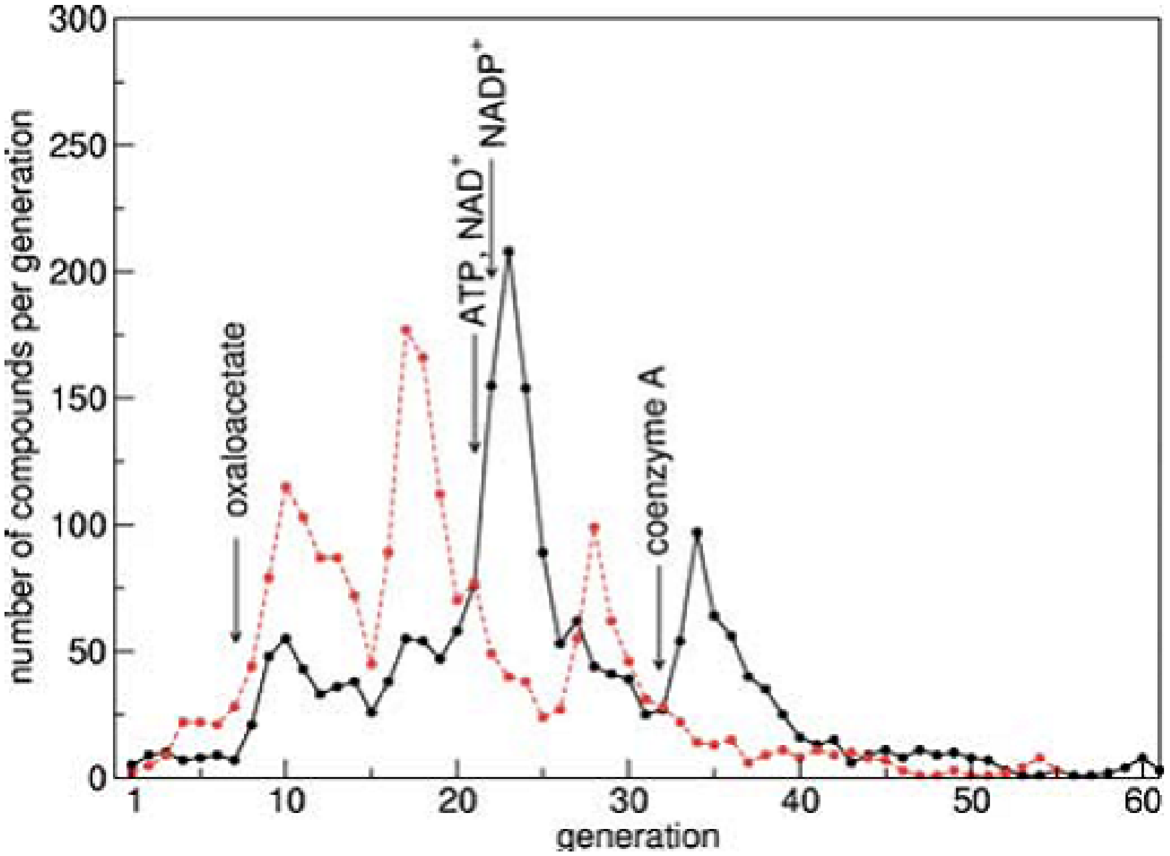
The expansion process

Initial conditions: availability of inorganic, 'prebiotic' compounds

(Martin and Russell, 2003)

carbonic acid:	H_2CO	(carbon)
methanethiol:	CH_3SH	(carbon, sulfur)
ammonia:	NH_3	(nitrogen)
pyrophosphate:	$P_2O_7^{4-}$	(phosphate)

Expansion on the complete KEGG network



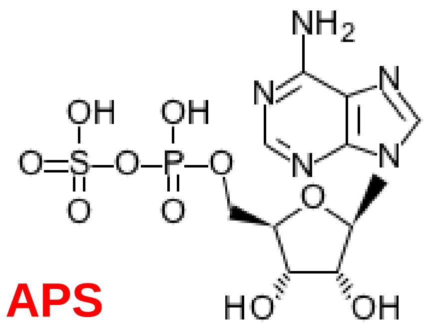
The expansion process

Initial conditions: availability of inorganic, 'prebiotic' compounds

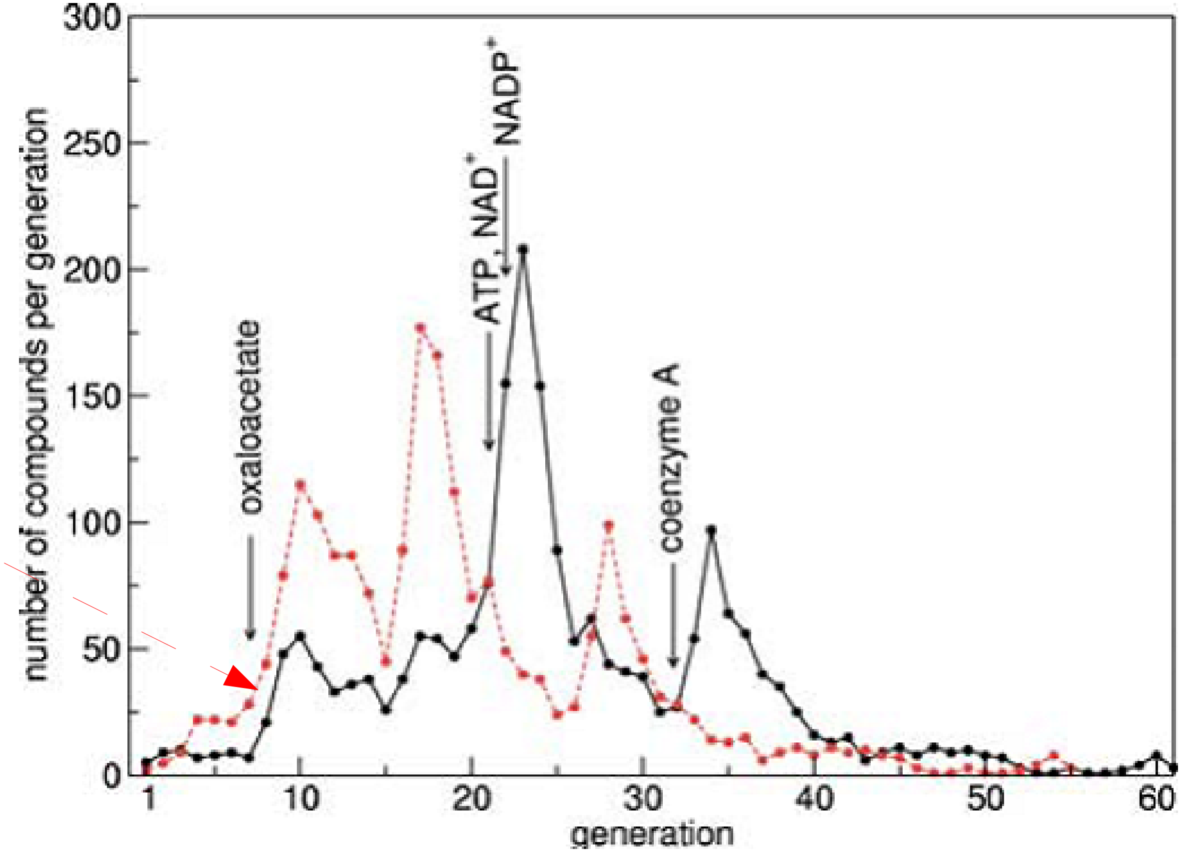
(Martin and Russell, 2003)

carbonic acid:	H_2CO	(carbon)
methanethiol:	CH_3SH	(carbon, sulfur)
ammonia:	NH_3	(nitrogen)
pyrophosphate:	$P_2O_7^{4-}$	(phosphate)

Expansion on the complete KEGG network



APS



Single Organisms

Producibility in the flux language

What are the ***biosynthetic capabilities*** of a network?

Let *U* denote the set of available nutrient metabolites.

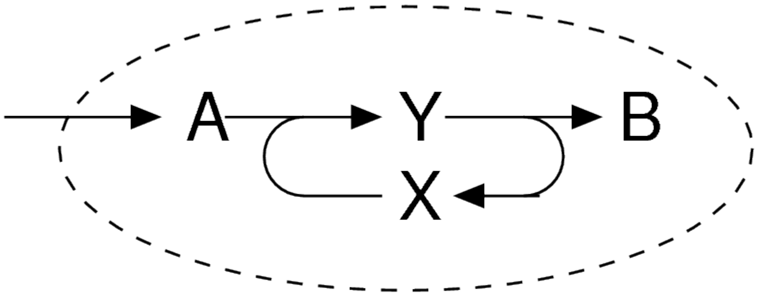
A metabolite is ***producible*** from the nutrients *U* if there exists a flux solution such that

- its own concentration increases
- only nutrients are consumed
- all others are at least balanced

Metabolite *k* is producible if $\exists v: \ [Sv]_k > 0 \ \wedge \ [Sv]_i \geqslant 0 \ \forall \ i \notin U$

Let $P(U)$ denote the set of all metabolites producible from nutrients *U*

Growth and Dilution: Toy models



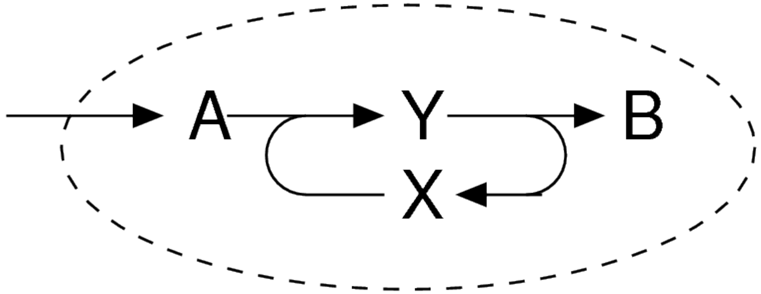
$U = \{A\}$ $P = \{B\}$

X and Y not producible from A!

What if the cell is growing? ⇨ Dilution! $\Rightarrow X, Y \rightarrow 0$

B is not producible under growth!

Growth and Dilution: Toy models

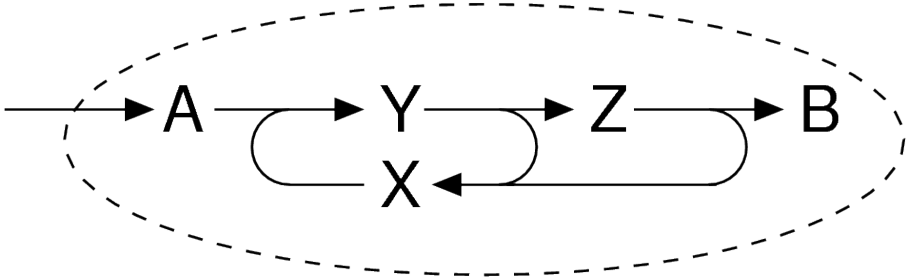


$U = \{A\}$ $P = \{B\}$

X and Y not producible from A!

What if the cell is growing? ⇨ Dilution! $\Rightarrow\ X, Y \rightarrow 0$

B is not producible under growth!



$U = \{A\}$ $P = \{X, Y, Z, B\}$

B is producible under growth!

Sustainability

A metabolite is ***sustainable*** from nutrients *U* if there exists a flux solution such that

- its own concentration increases
- only nutrients are consumed
- all other required intermediates are sustainable

Let *U* denote the set of available nutrient metabolites.

Let $P(U)$ denote the set of all metabolites producible from nutrients *U*

Recursive definition of sustainable metabolites:

Let $P_0 = P(U)$

Define forbidden set of reactions: $F_n = \{ j \mid \exists\ i \notin P_n : S_{ij} < 0 \}$

$$P_{n+1} = \{ k \mid \exists v :\ v_j = 0\ \forall\ j \in F_n \wedge [Sv]_k > 0 \wedge [Sv]_i \geqslant 0\ \forall\ i \notin U \}$$

Let $S(U)$ denote the set of all metabolites sustainable from U, defined by

$$S(U) = \lim_{n \to \infty} P_n$$

Takes a long time to compute!

Relating scopes to flux models

Let $\Sigma(U)$ denote the scope of *U*

It can simply be shown that $\Sigma(U) \subseteq S(U) \subseteq P(U)$

Numerical experiment for the network of *E.coli* (Reed et al., 2003)

determine $\Sigma(U)$ and $S(U)$

for all sets $U = \{k, H_2O\}$

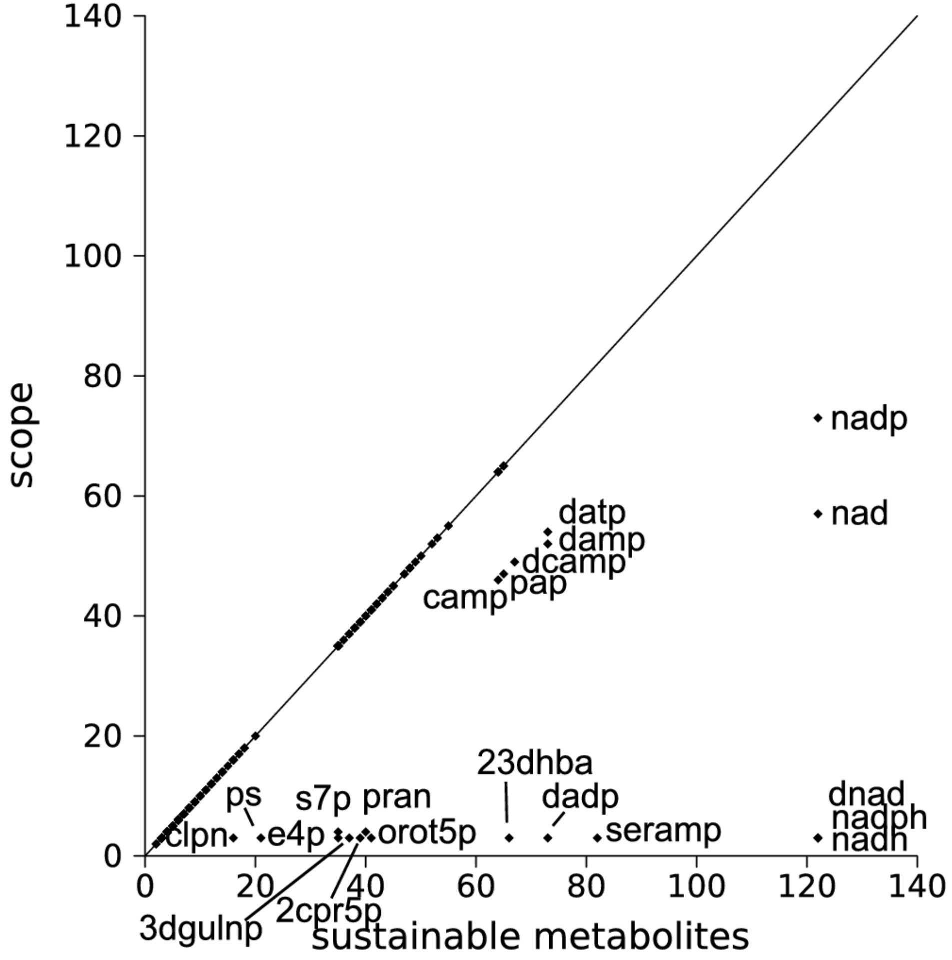
Observation:

for most metabolites:

$$\Sigma(U) = S(U)$$

► Agreement even better for more complex sets *U*

(Kruse and Ebenhöh, GI 2008)



Relating scopes to flux models

Let $\Sigma(U)$ denote the scope of *U*

It can simply be shown that $\Sigma(U) \subseteq S(U) \subseteq P(U)$

Numerical experiment for the network of *E.coli* (Reed et al., 2003)

determine $\Sigma(U)$ and $S(U)$

for all sets $U = \{k, H_2O\}$

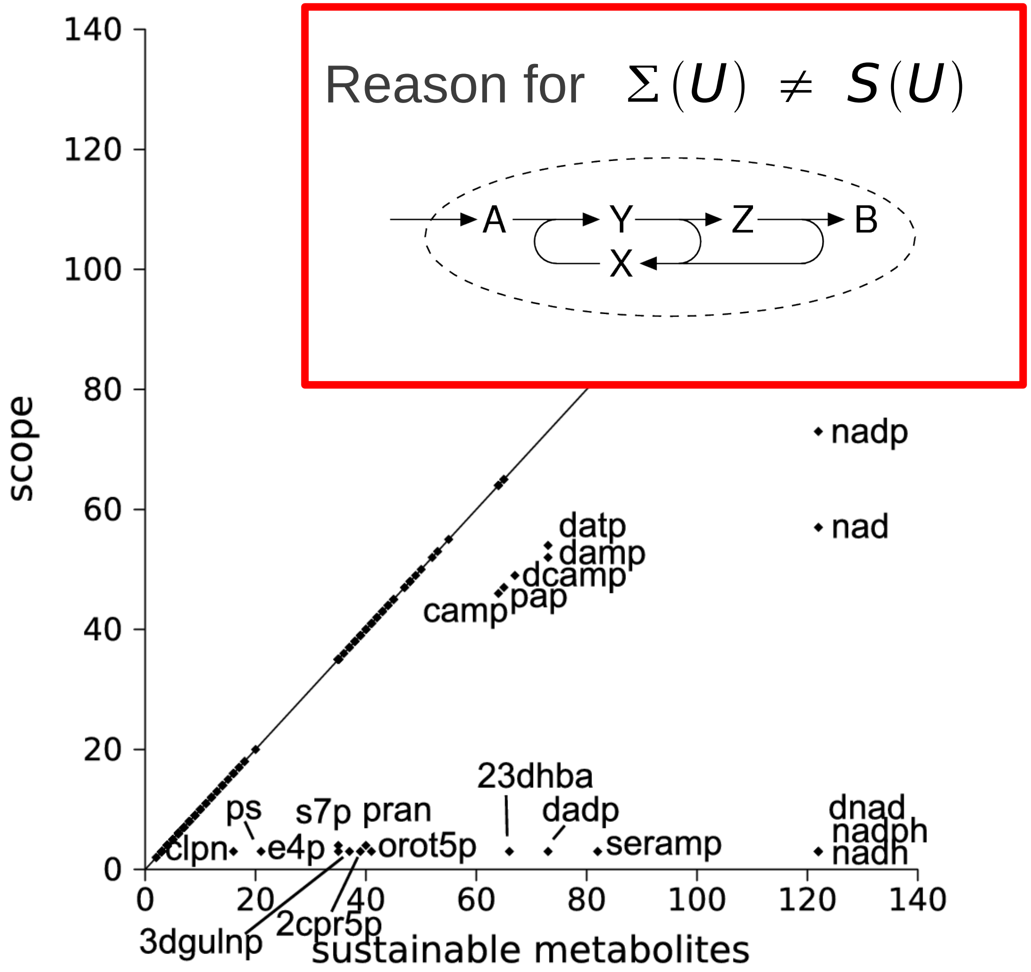
Observation:

for most metabolites:

$$\Sigma(U) = S(U)$$

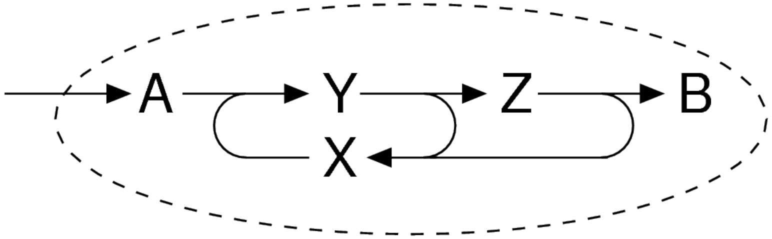
► Agreement even better for more complex sets *U*

(Kruse and Ebenhöh, GI 2008)

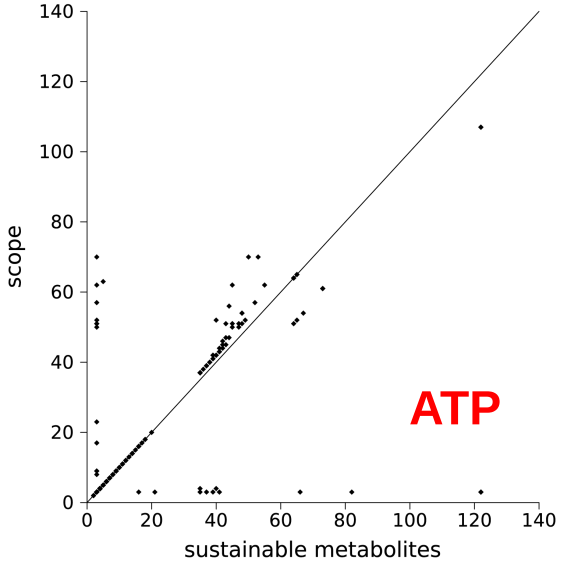

Reason for $\Sigma(U) \neq S(U)$

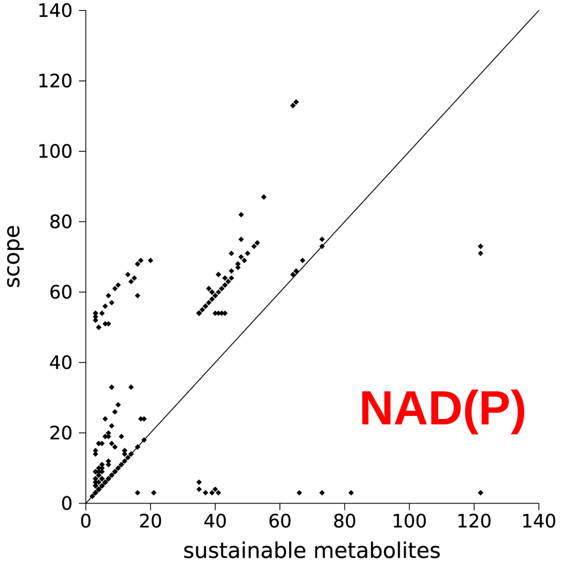
The role of cofactors

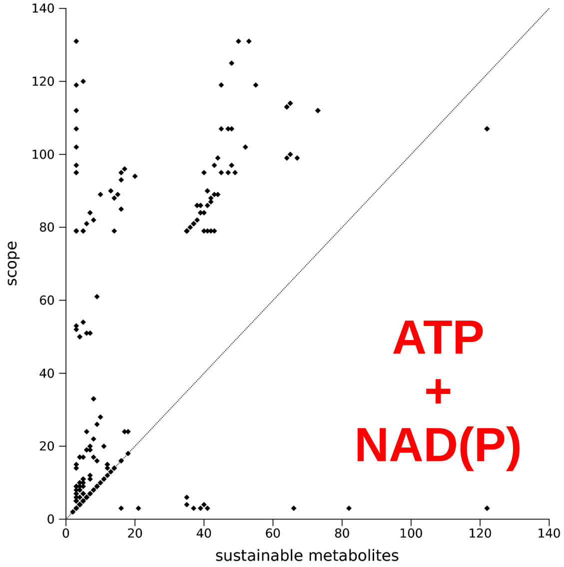
Common cofactors (ATP/NADH) are of the type



We add cofactors to the seed
(ATP does not have to be produced to be used as a cofactor)







We tend to overestimate the 'true' biosynthetic capacity (under constant growth)

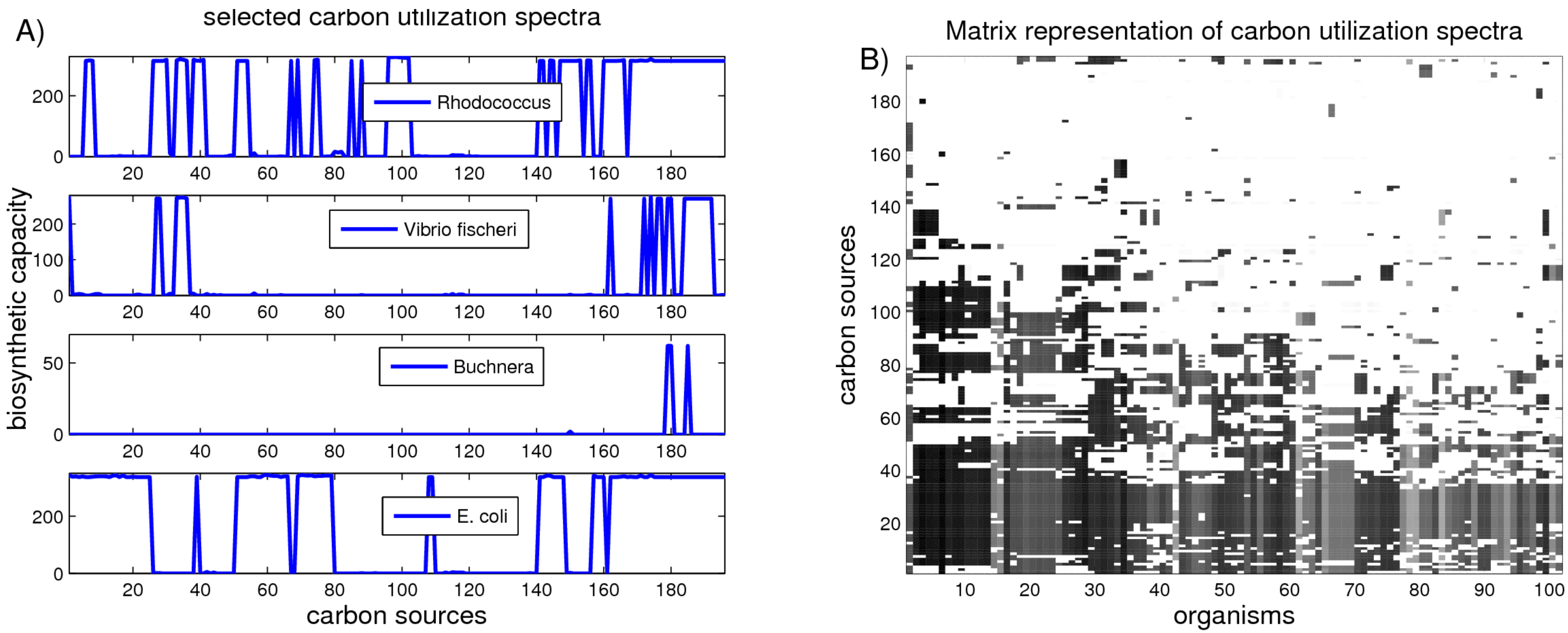
But that's OK to give a meaning to “The scope of glucose”

Single organisms

Investigate biosynthetic capacities of organisms on various carbon sources:

- 447 organism specific networks (KEGG)
- 200 carbon sources

⇨ **CARBON UTILIZATION SPECTRA**

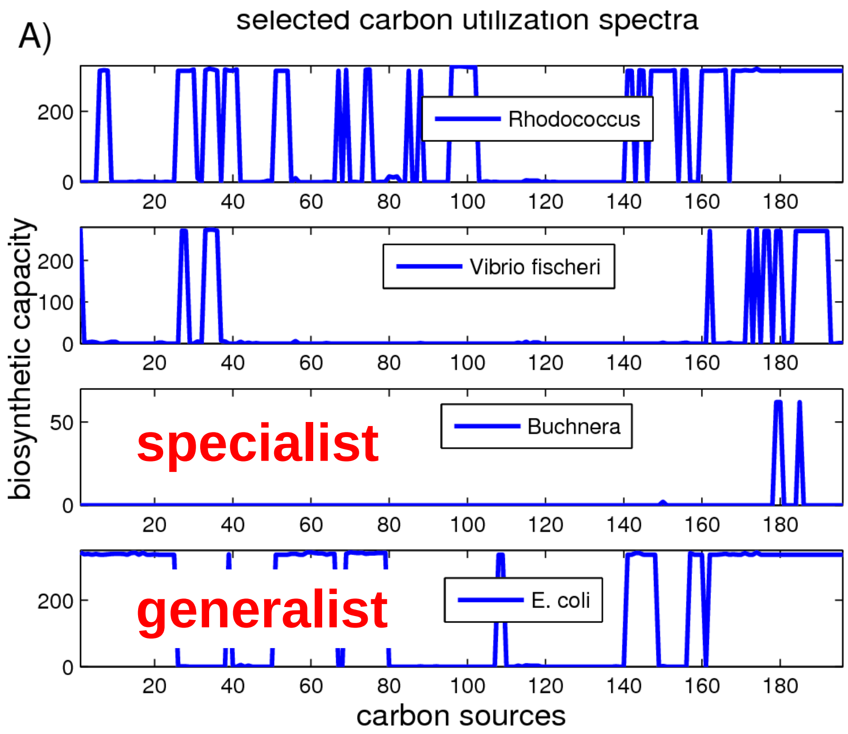


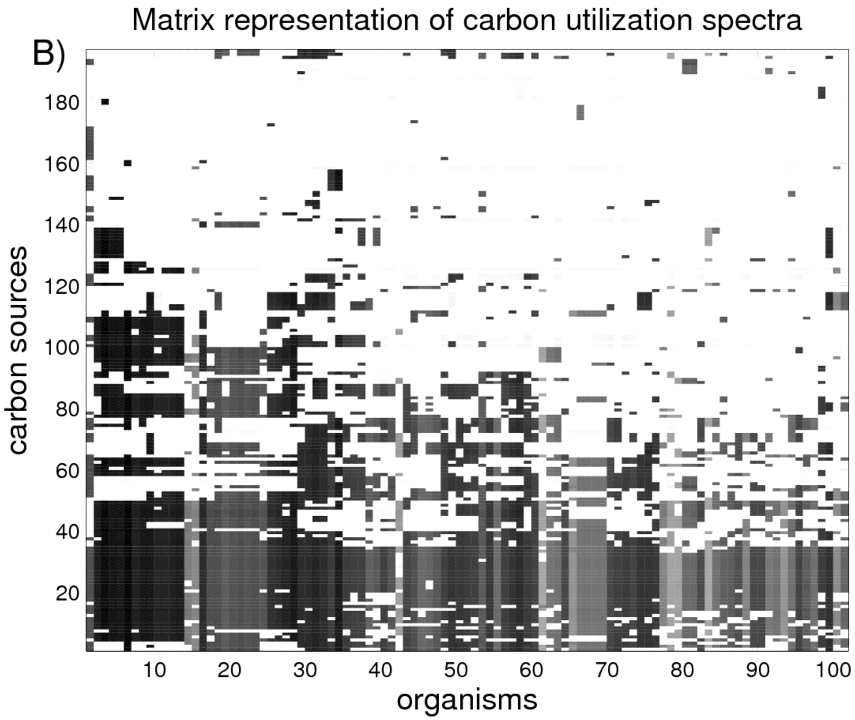
Single organisms

Investigate biosynthetic capacities of organisms on various carbon sources:

- 447 organism specific networks (KEGG)
- 200 carbon sources

⇒ **CARBON UTILIZATION SPECTRA**

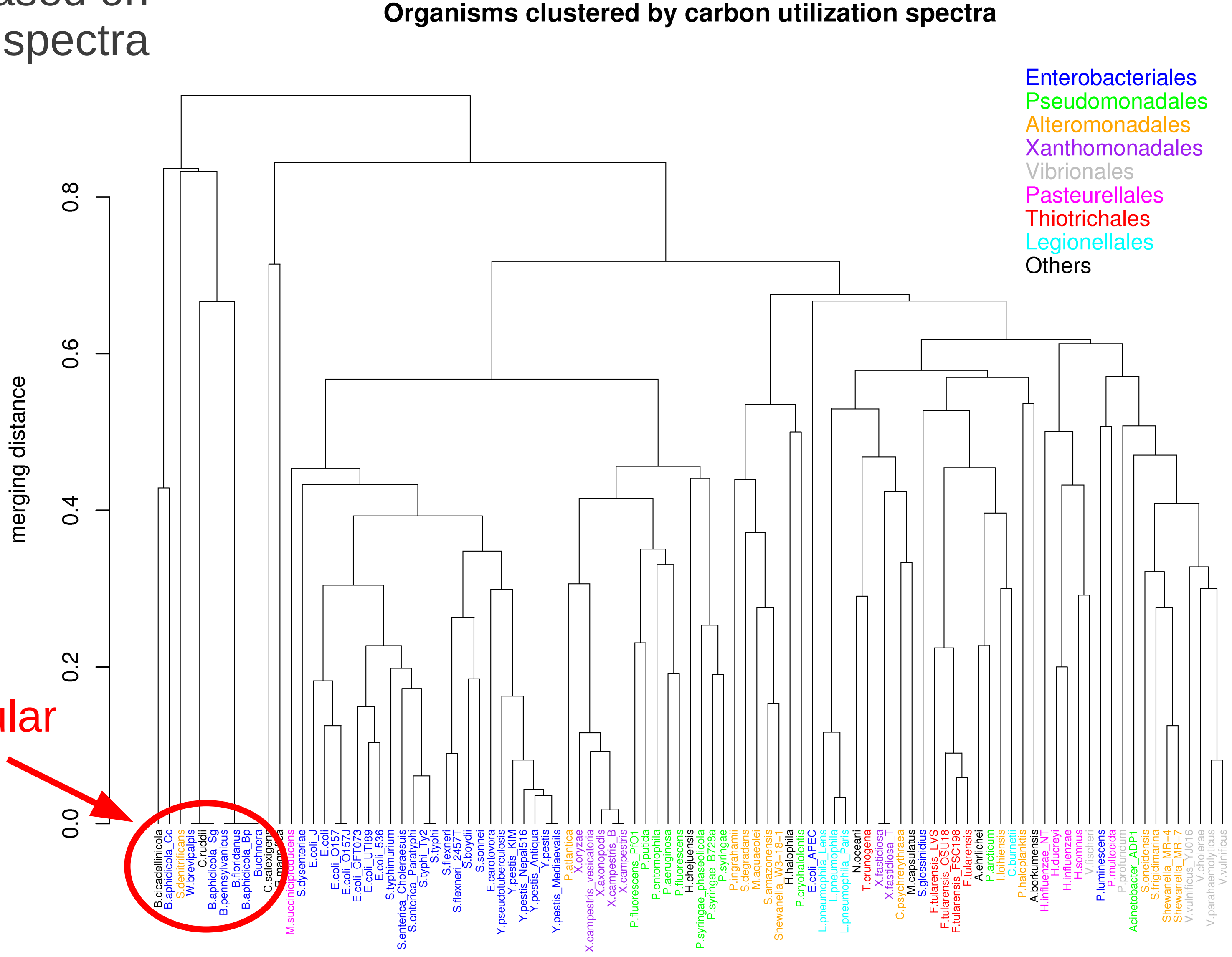

selected carbon utilization spectra


Matrix representation of carbon utilization spectra

It is in principle possible to distinguish between generalists and specialists

Single organisms

A phenetic tree based on carbon utilization spectra



Organisms clustered by carbon utilization spectra

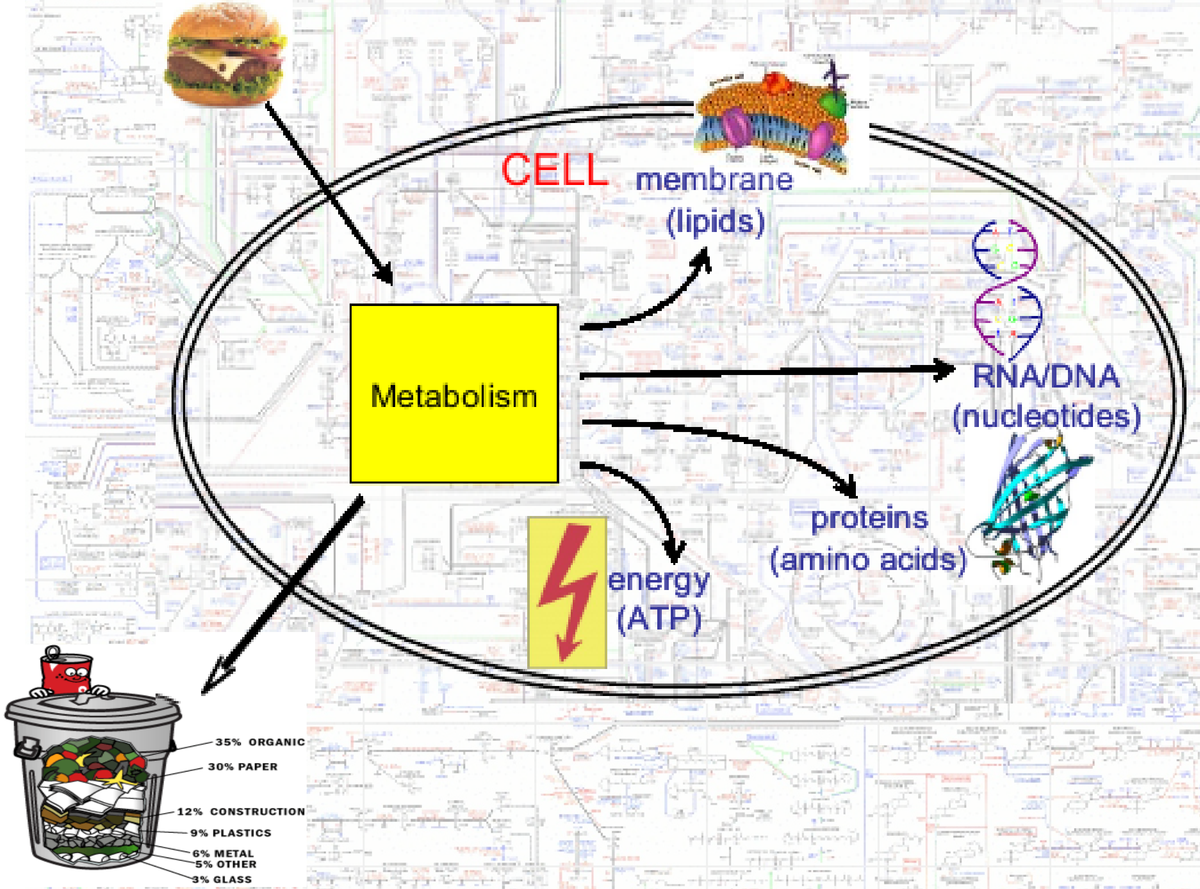
many intracellular parasites

⇨ Classification of organisms by 'lifestyle'?

(Ebenhöh and Handorf, EURASIP, 2009)

Combining expert knowledge with mathematics

Expert knowledge



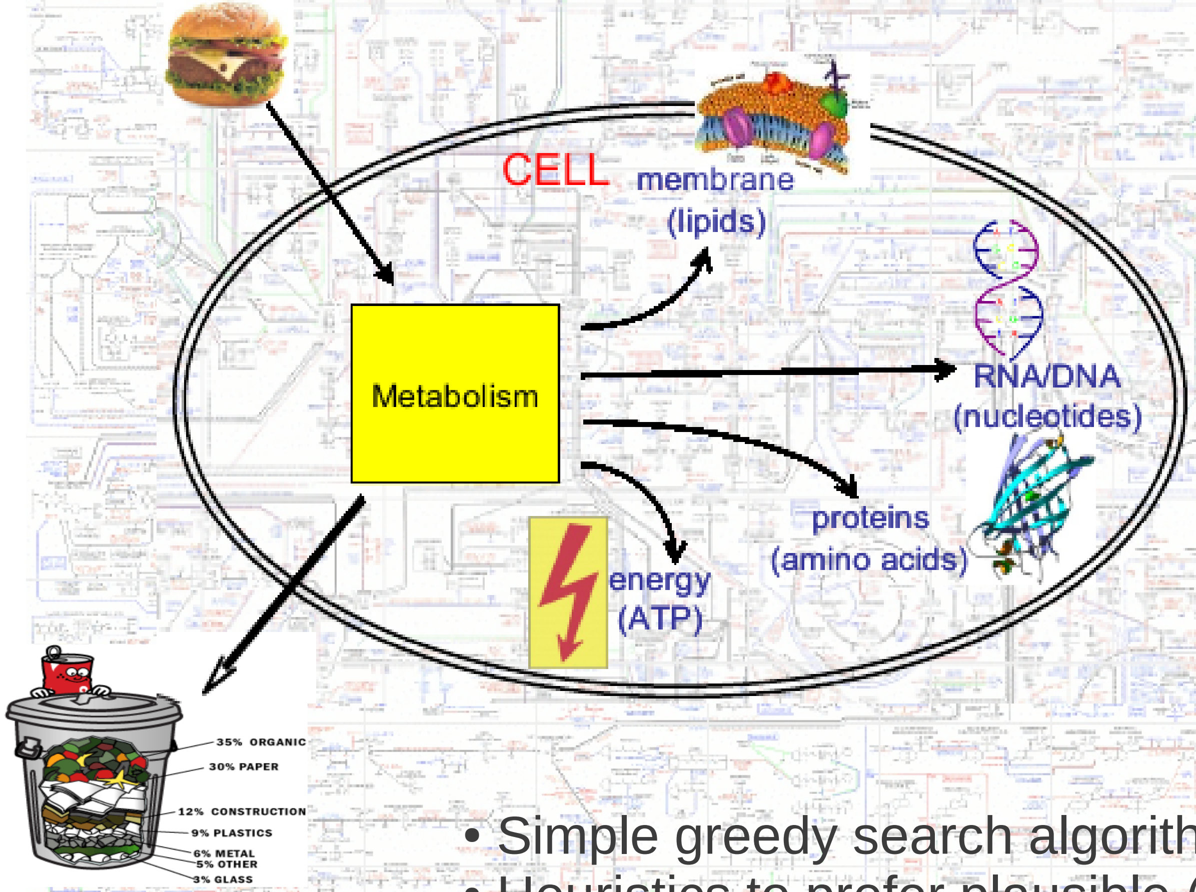
Mathematics

Every network must be able to produce precursors:

- amino acids
- nucleotides
- lipids
- energy
- etc...

Combining expert knowledge with mathematics

Expert knowledge



CELL

membrane
(lipids)

Metabolism

RNA/DNA
(nucleotides)

proteins
(amino acids)

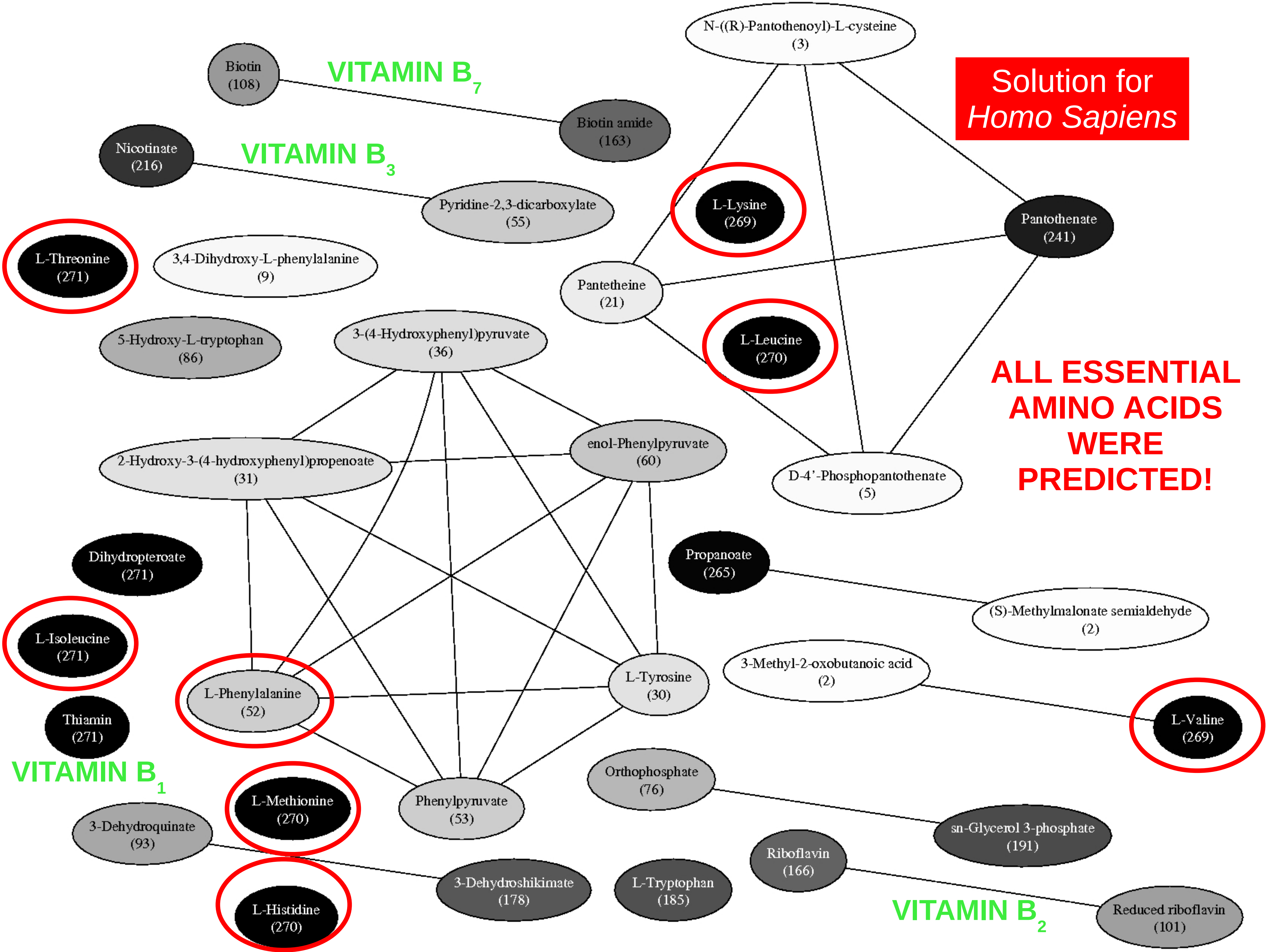
energy
(ATP)

Mathematics

Every network must be able to produce precursors:

- amino acids
- nucleotides
- lipids
- energy
- etc...

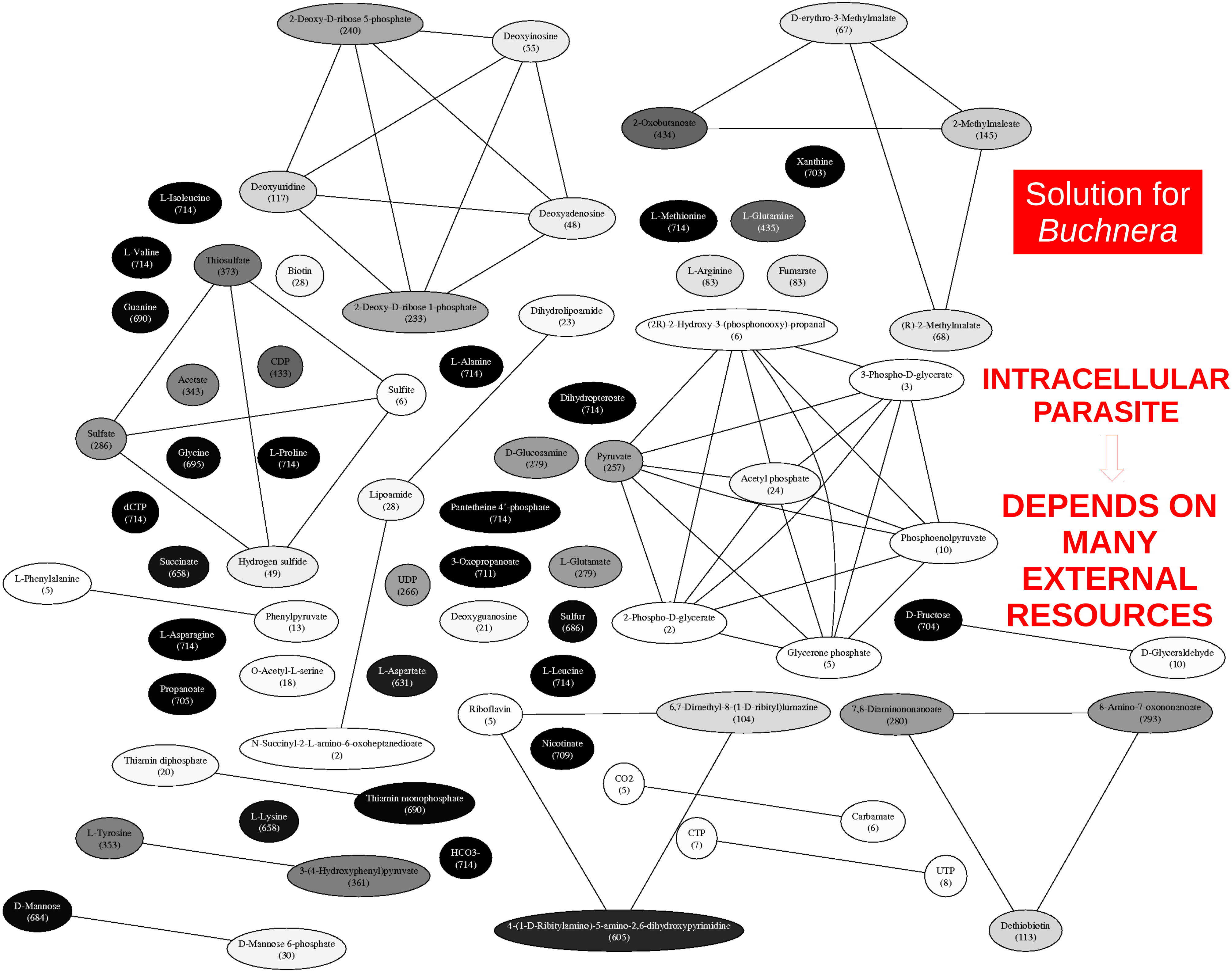
- Simple greedy search algorithm
- Heuristics to prefer plausible (small, transportable) nutrients



Solution for
Homo Sapiens

ALL ESSENTIAL AMINO ACIDS WERE PREDICTED!

Solution for *Buchnera*



INTRACELLULAR PARASITE

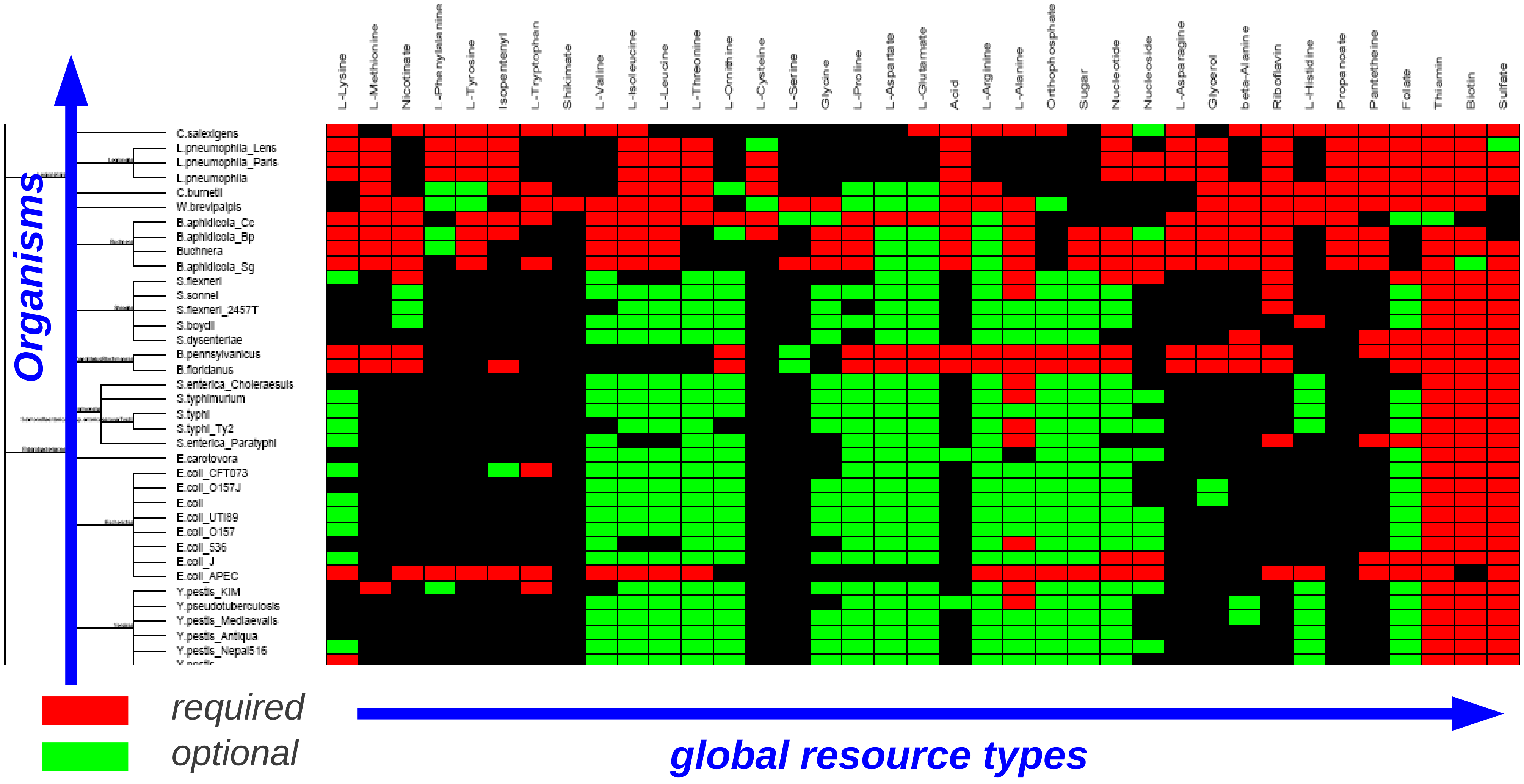
⇩

DEPENDS ON MANY EXTERNAL RESOURCES

Global resource types

The comparison of the results for 400 organisms allows to define

36 global resource types



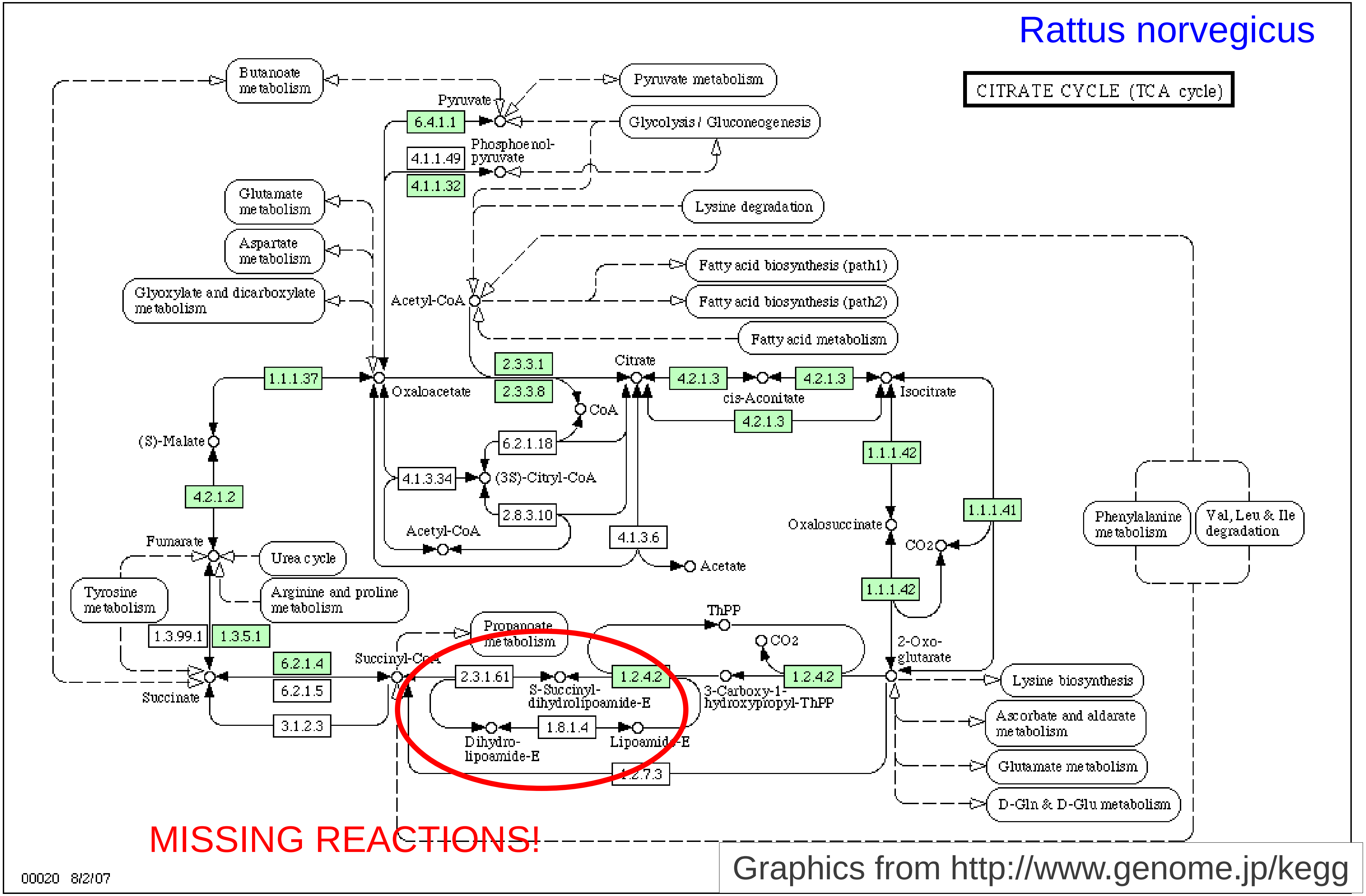
🟥 *required*

🟩 *optional*

⬛ *not required*

Handorf, Christian, Ebenhöh & Kahn, JTB, 2008

Closing gaps in metabolism

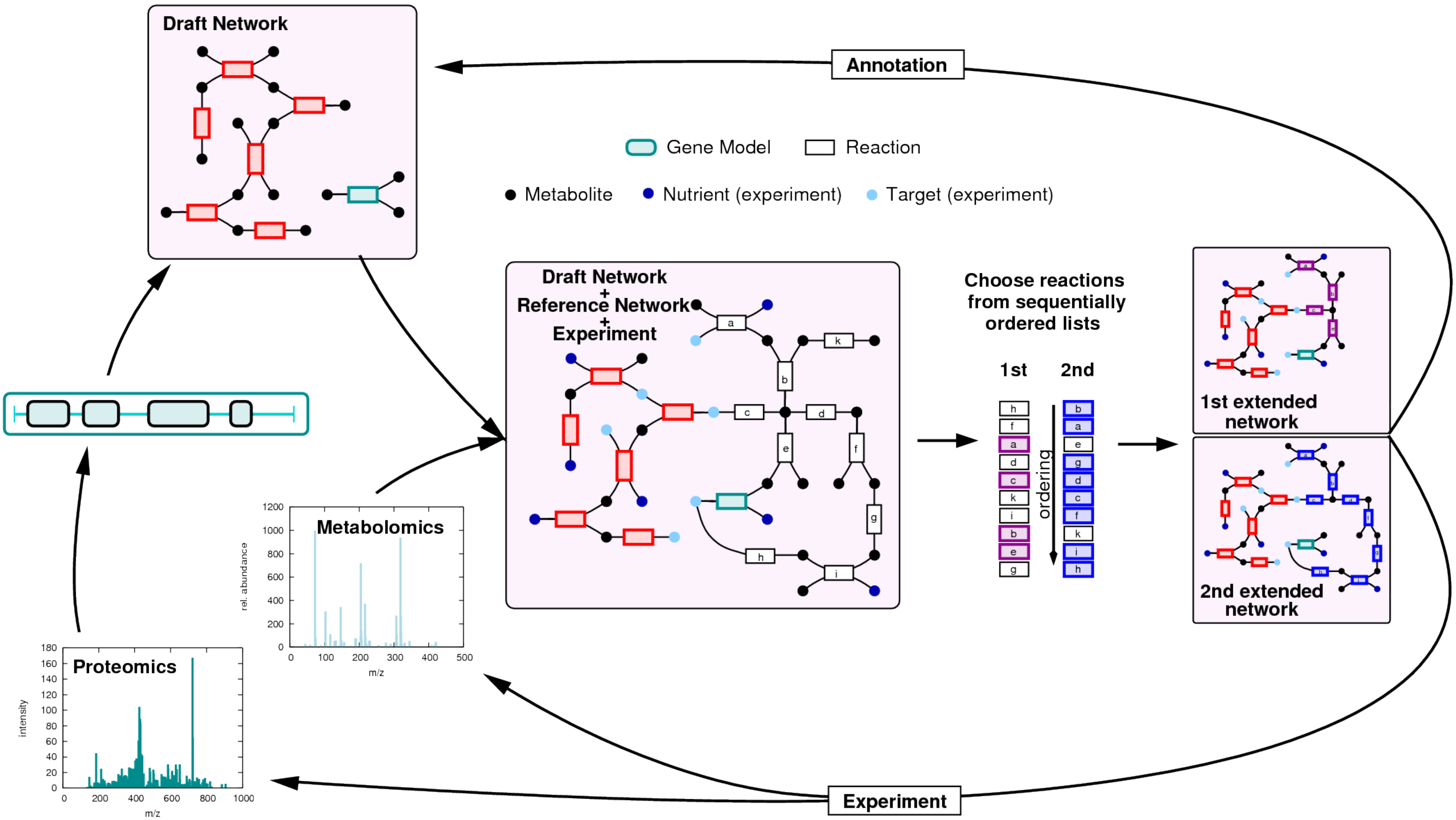


Rattus norvegicus

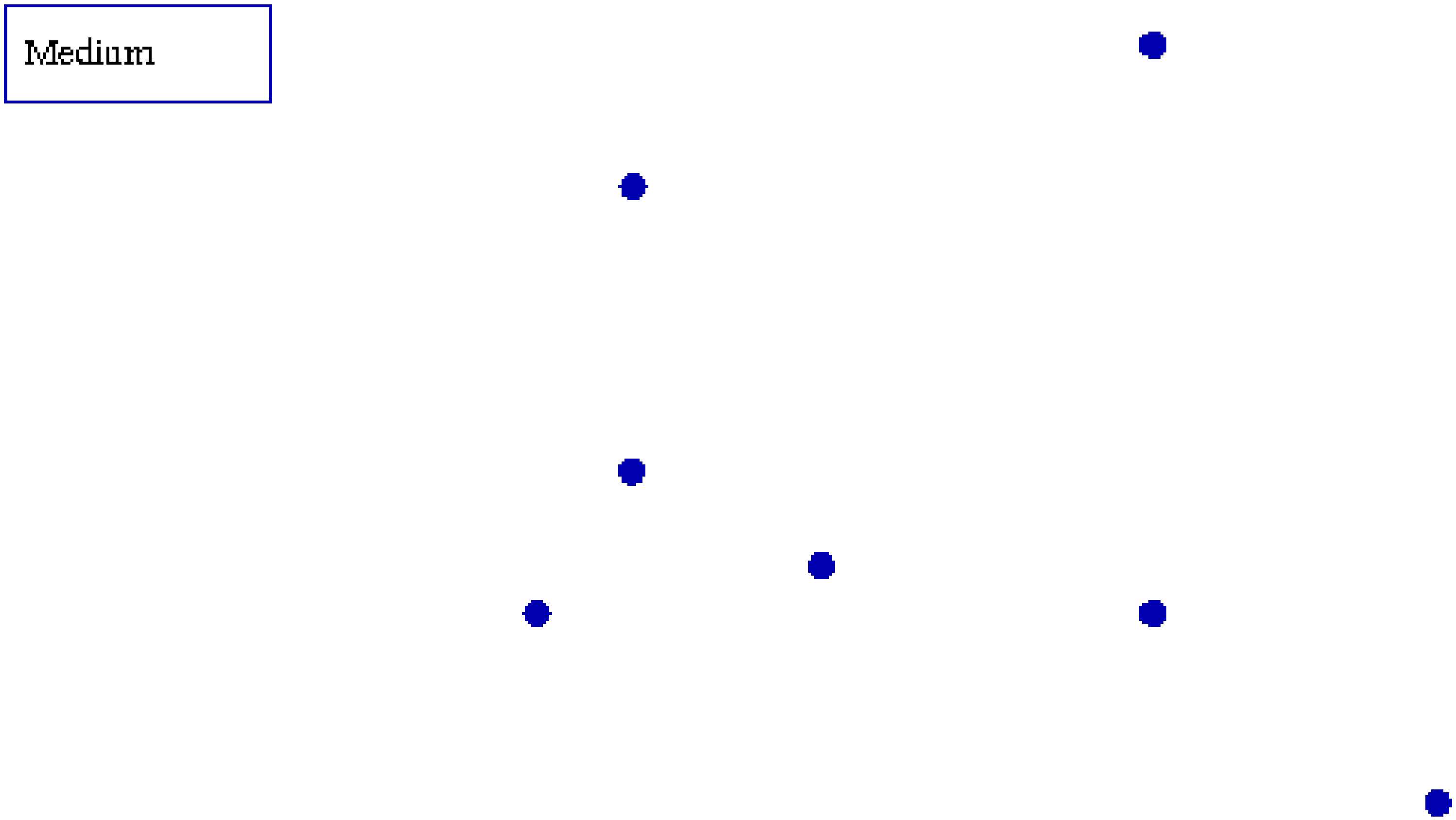
MISSING REACTIONS!

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

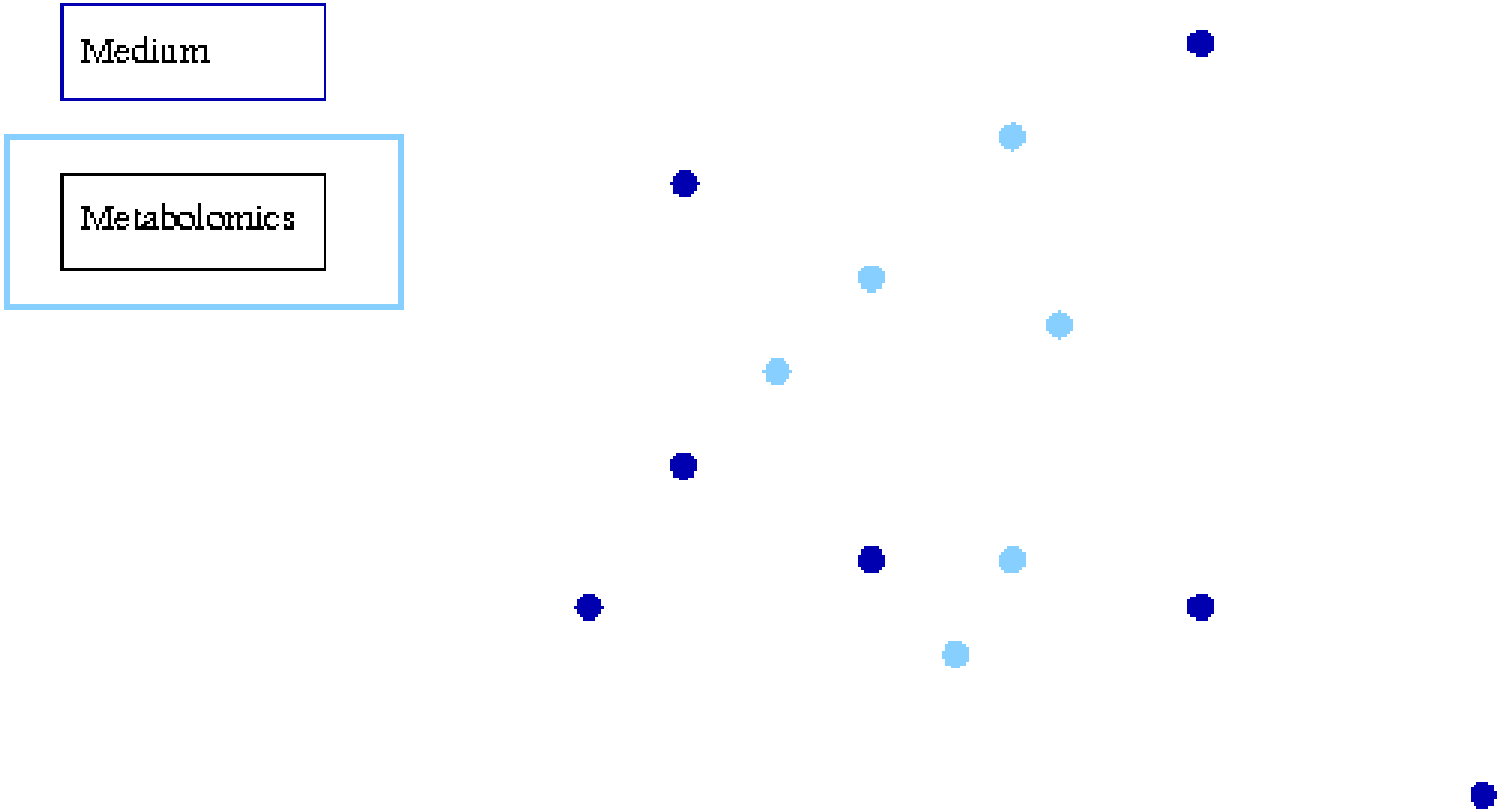
Closing gaps in metabolism



Filling the gaps



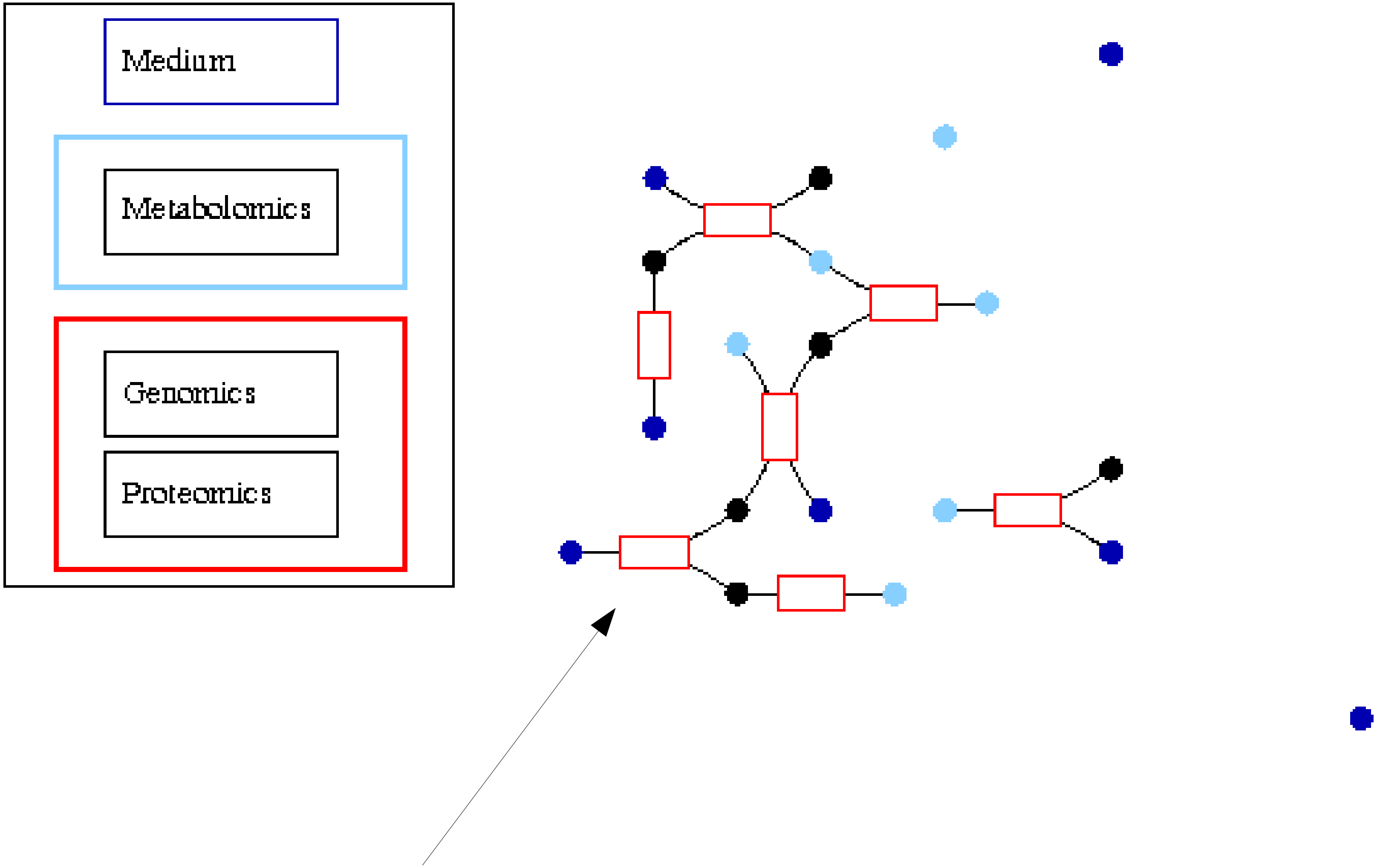
Filling the gaps



Medium

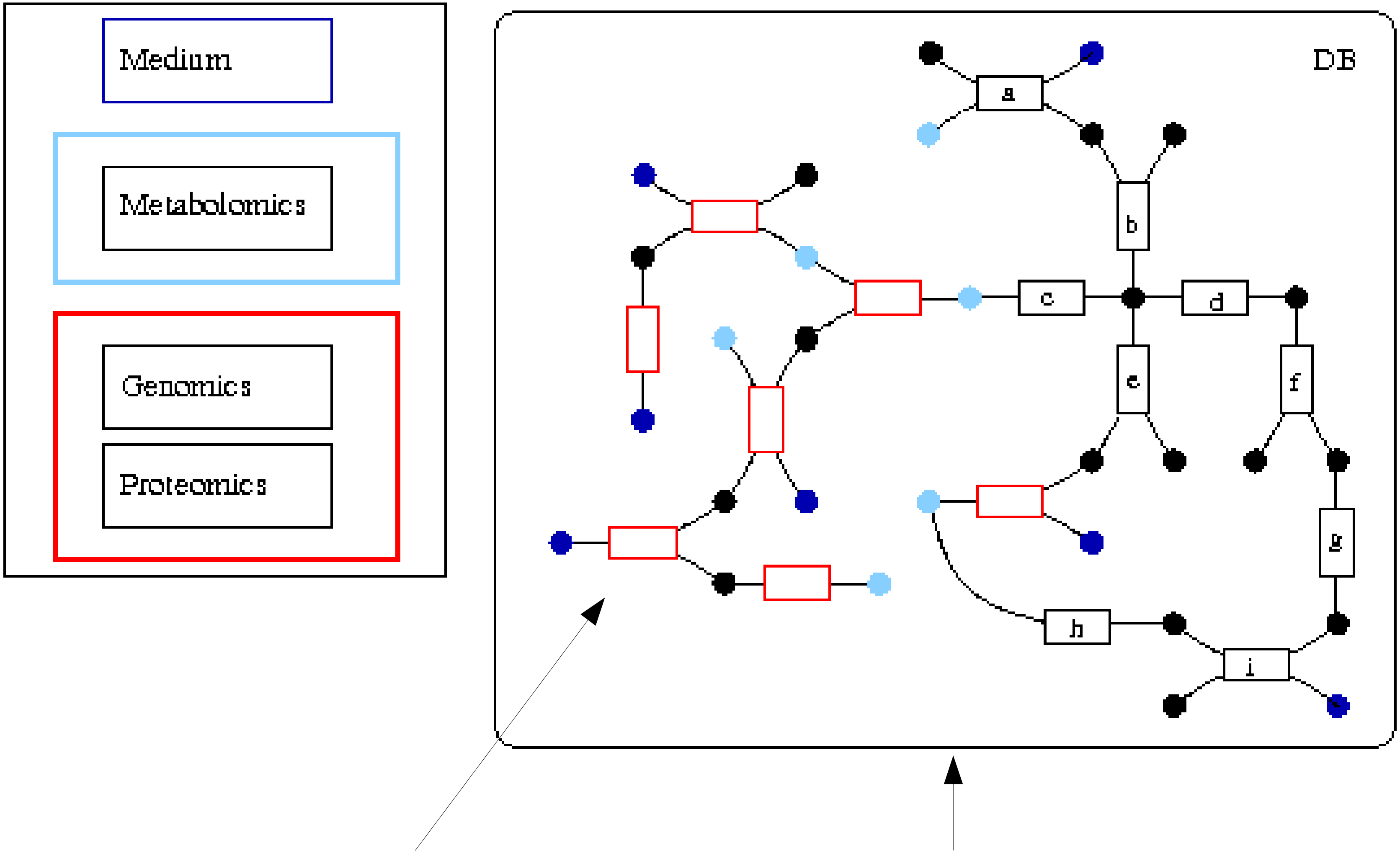
Metabolomics

Filling the gaps



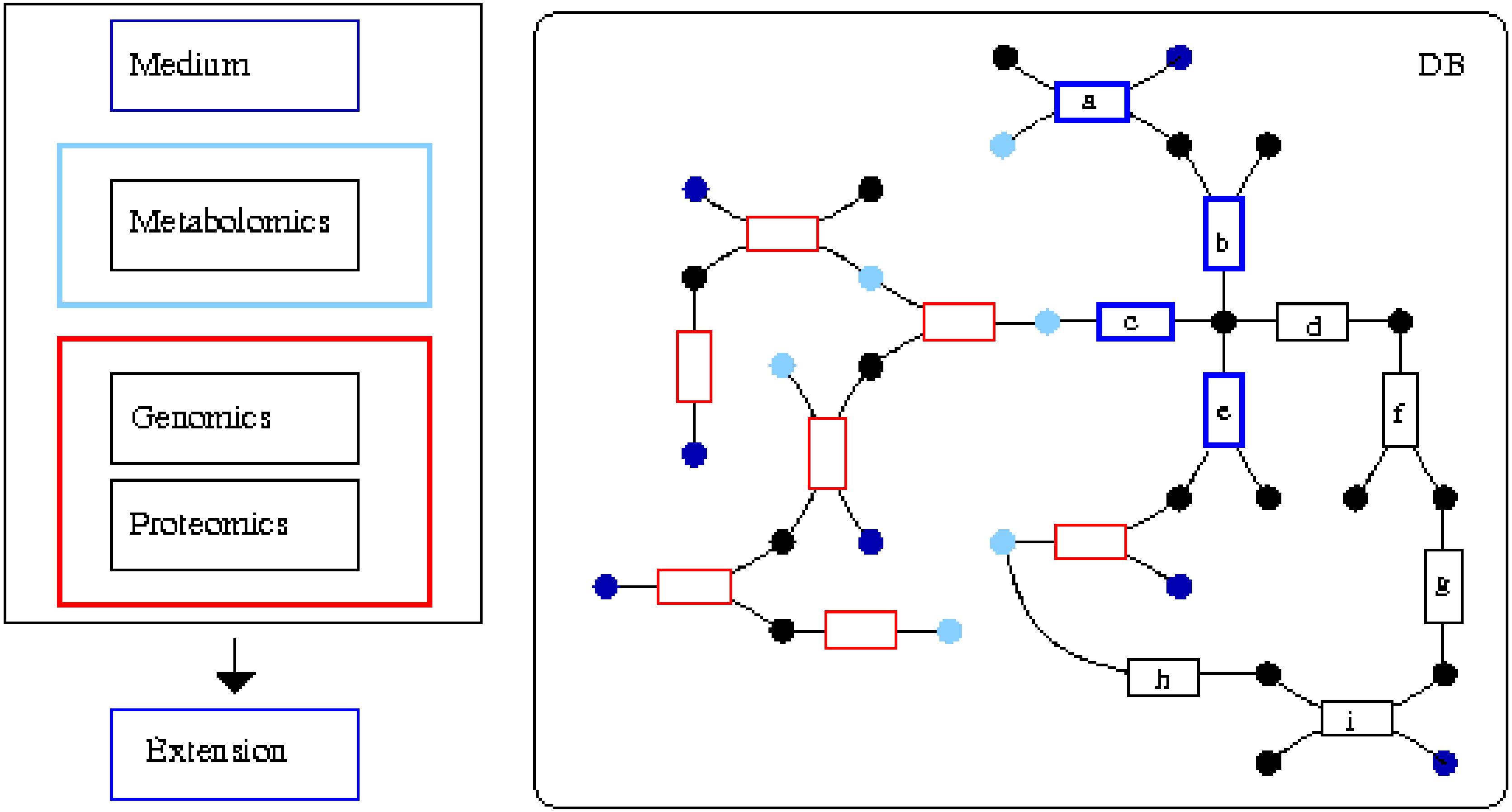
Draft network

Filling the gaps



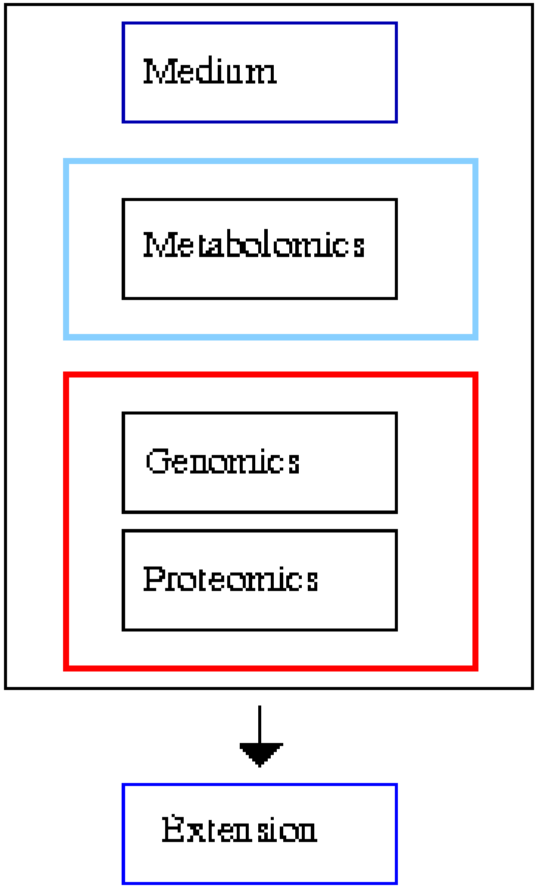
Draft network embedded in larger network (from database)

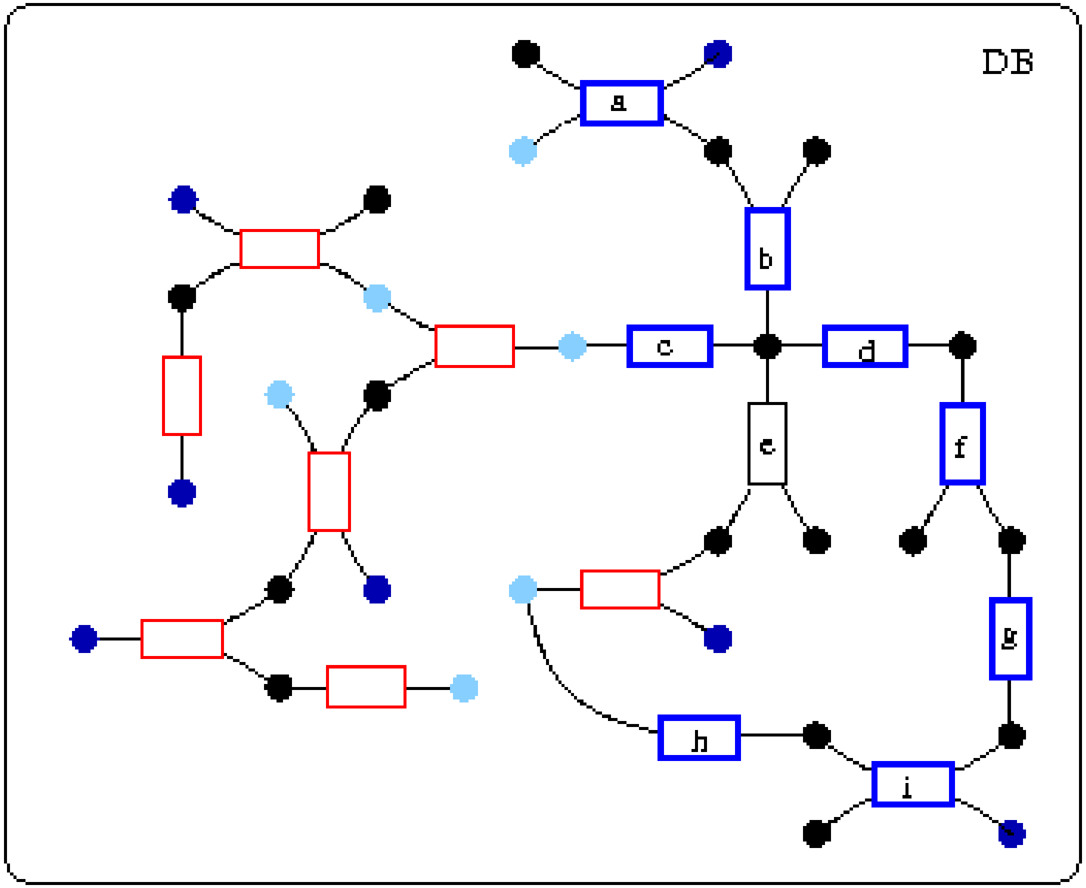
Filling the gaps



Solution 1: minimal extension with 4 reactions

Filling the gaps

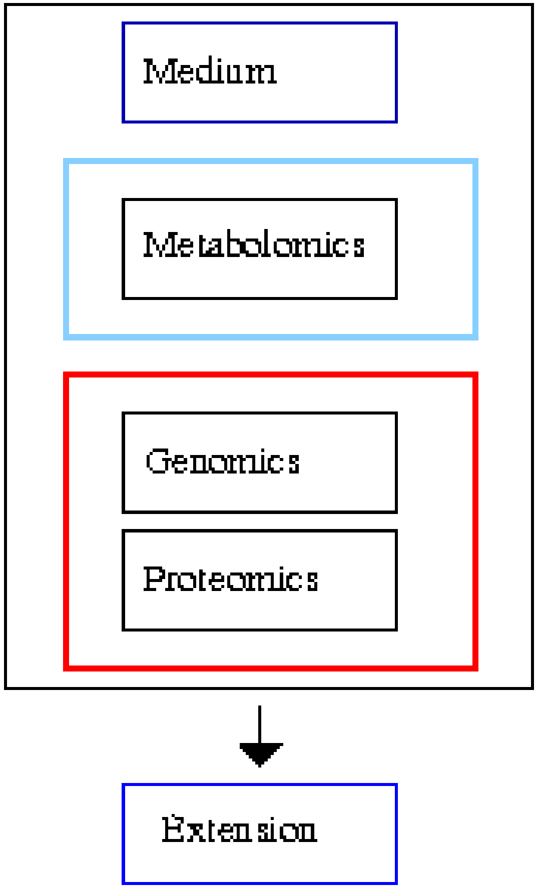


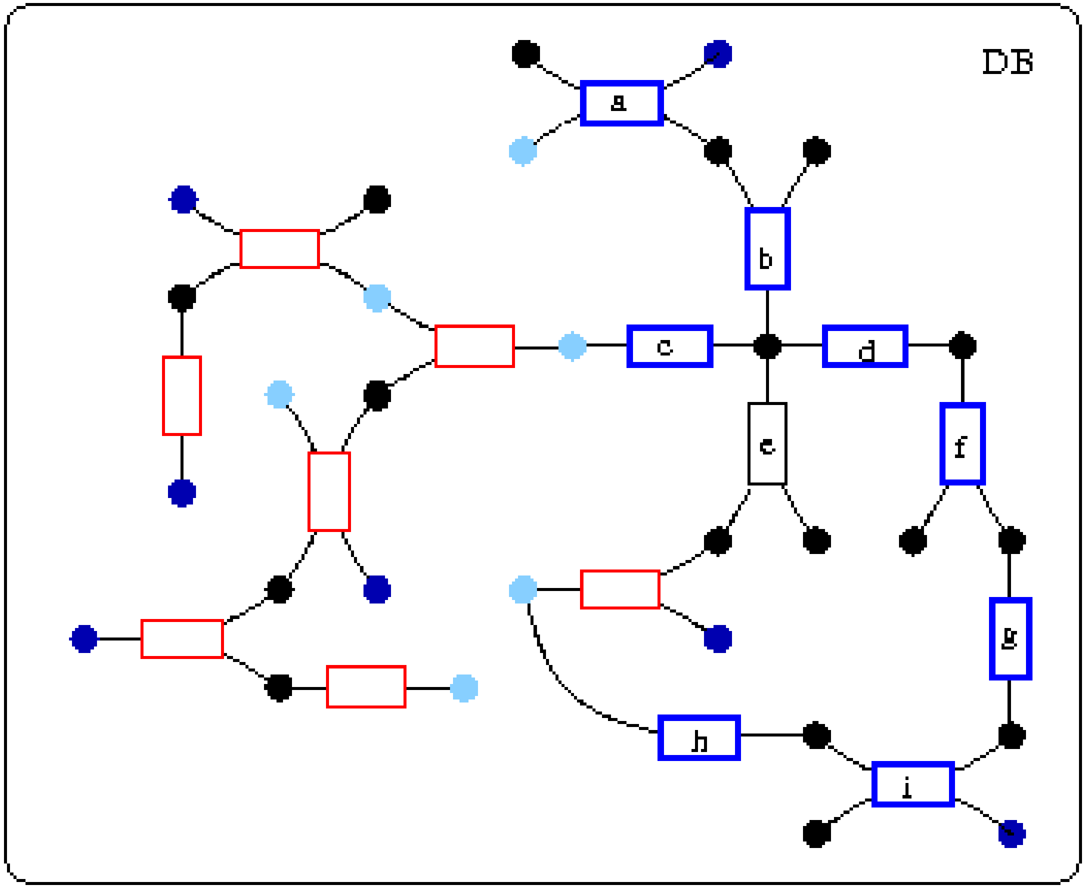


Solution 1: minimal extension with 4 reactions

Solution 2: minimal extension with 8 reactions

Filling the gaps



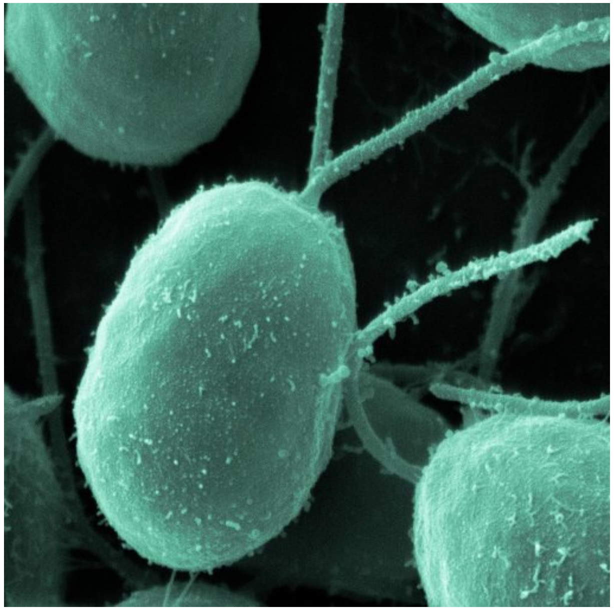


- greedy algorithm (traversing all reactions)
- depends on the order of the reaction lists

Solution 1: minimal extension with 4 reactions

Solution 2: minimal extension with 8 reactions

Chlamydomonas reinhardtii



Model organism of the *Go*FORSYS research consortium
(***photosynthesis and growth*** - http://www.goforsys.de)

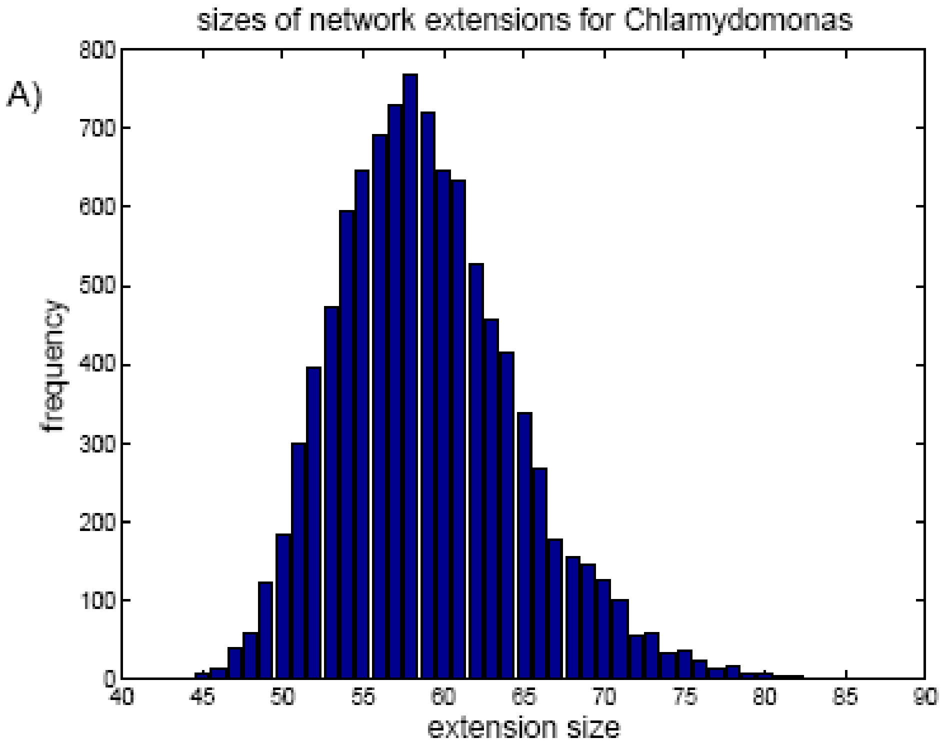
- 15143 genes (JGI)
- 6864 functional annotated genes in ChlamyCyc
- 1258 biochemical reactions (Patrick May)
- 159 measured metabolites (Stefan Kempa)
- 22 not producible by draft network

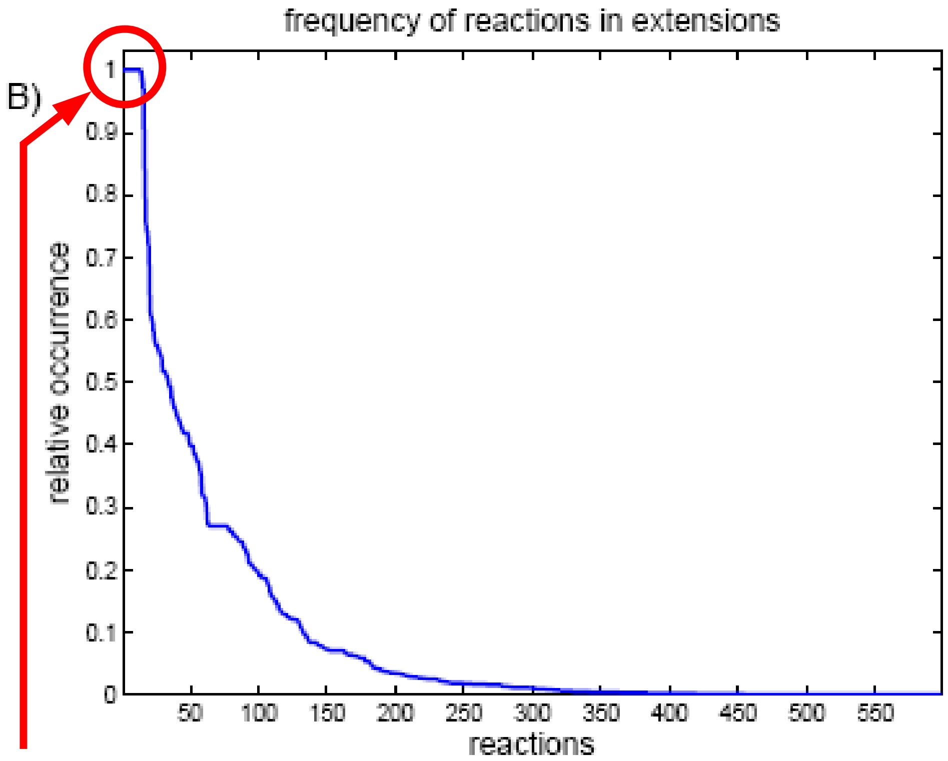
embed draft network in reference network (>6000 reactions from MetaCyc)

→ 598 distinct reactions in 10000 calculated minimal extensions

May et al., *Genetics* (2008) Chlamydomonas special issue

Chlamydomonas reinhardtii – some statistics





Minimum: 44 reactions

Maximum: 88 reactions

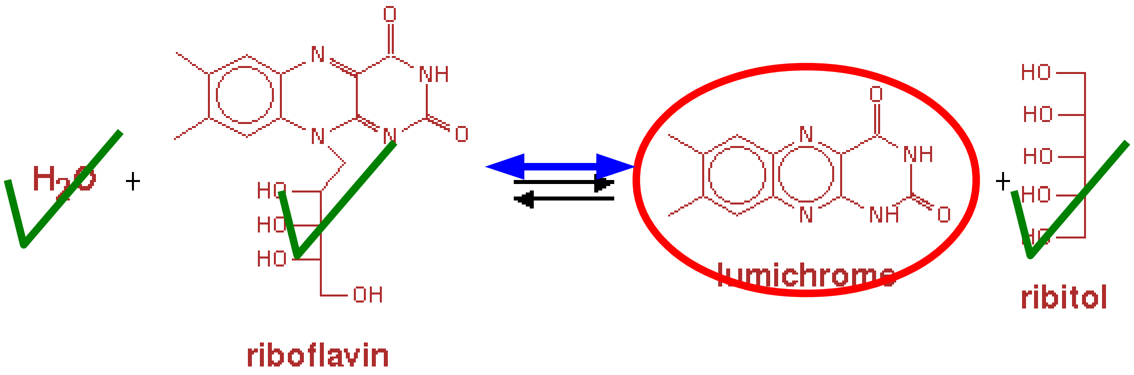
15 ubiquitious reactions!
(found in every extension)

Some specific examples

lumichrome

- secreted by Chlamydomonas
- disrupts quorum-sensing pathways in bacteria
- putative role in pathogen defense

(Rajamani et al., 2008)



○ experimentally observed metabolite

✓ metabolites/reactions annotated in Chlamy

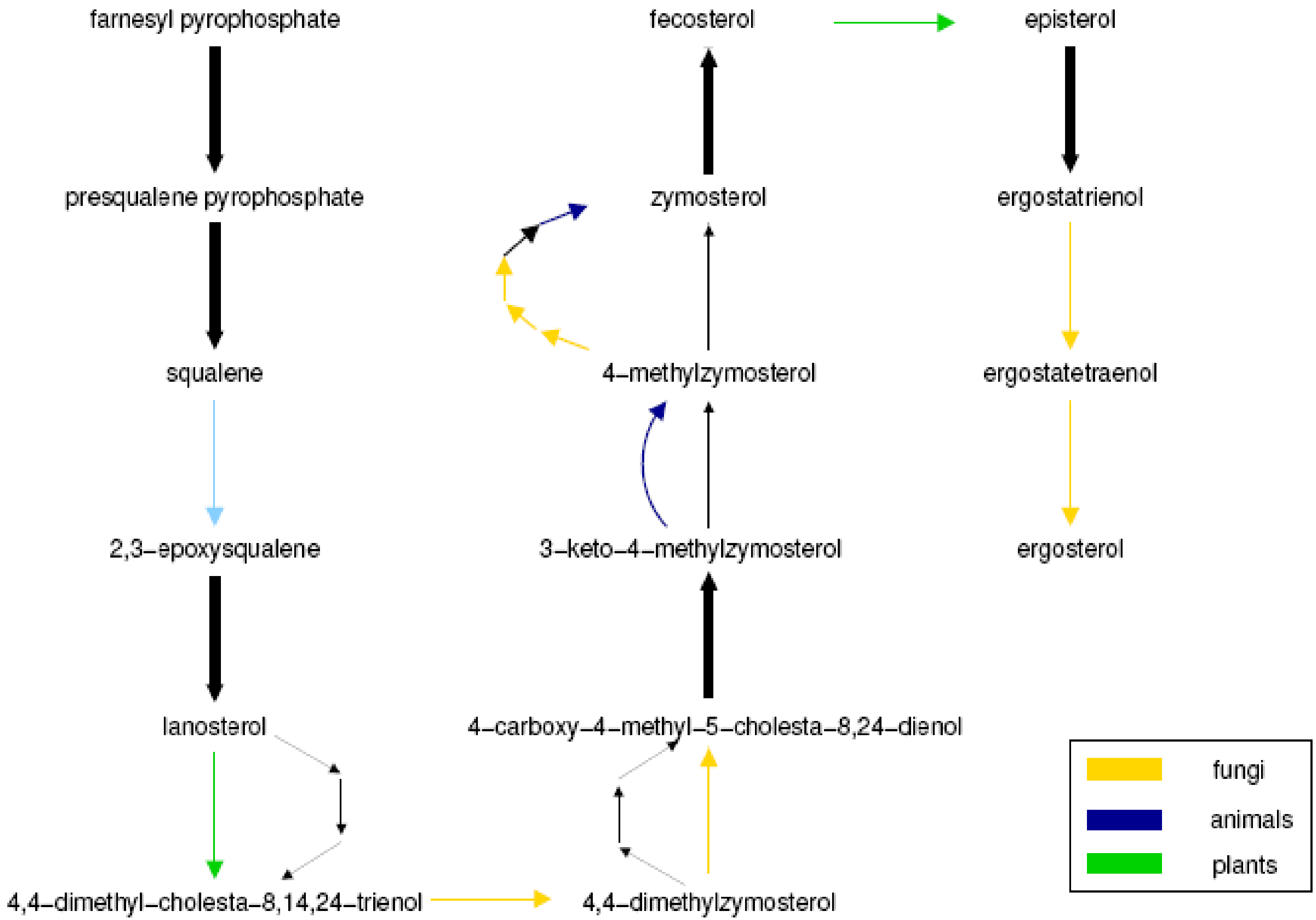
→ predicted reactions

graphics from www.metacyc.org

ergosterol

- component in ciliary membranes

Ergosterol synthesis

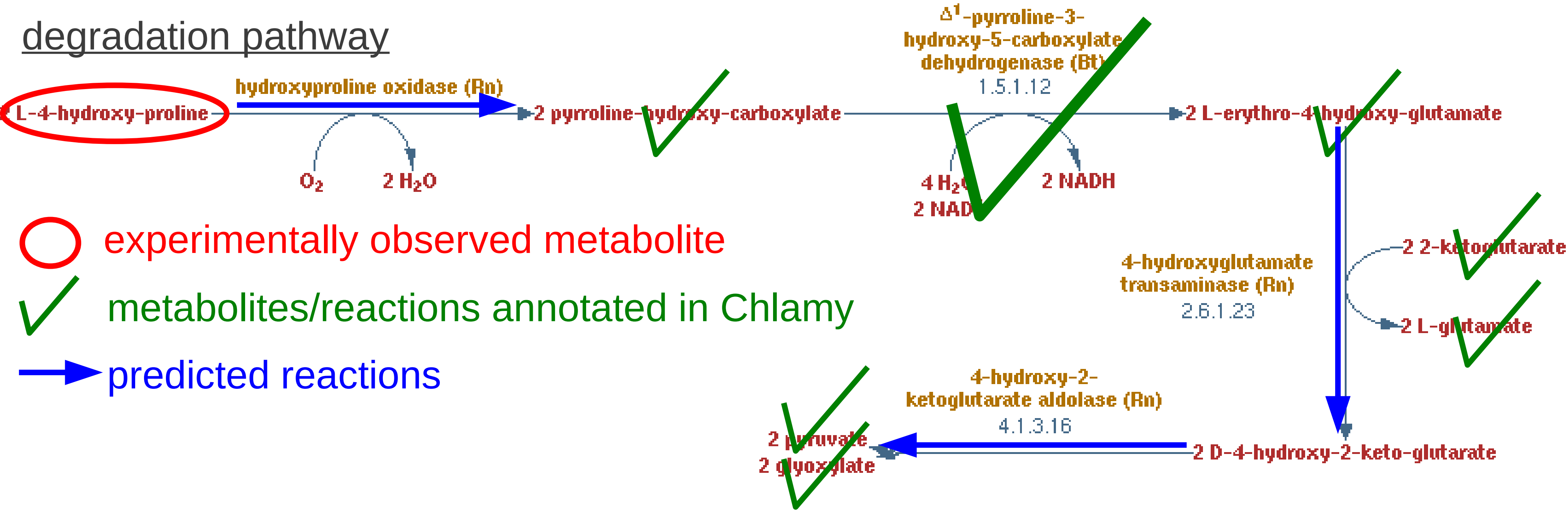


Completion of a pathway

L-4-hydroxyproline

- in animals: important structural component of collagen
- in plants: found in some glycoproteins and cell wall proteins

degradation pathway



○ experimentally observed metabolite

✓ metabolites/reactions annotated in Chlamy

→ predicted reactions

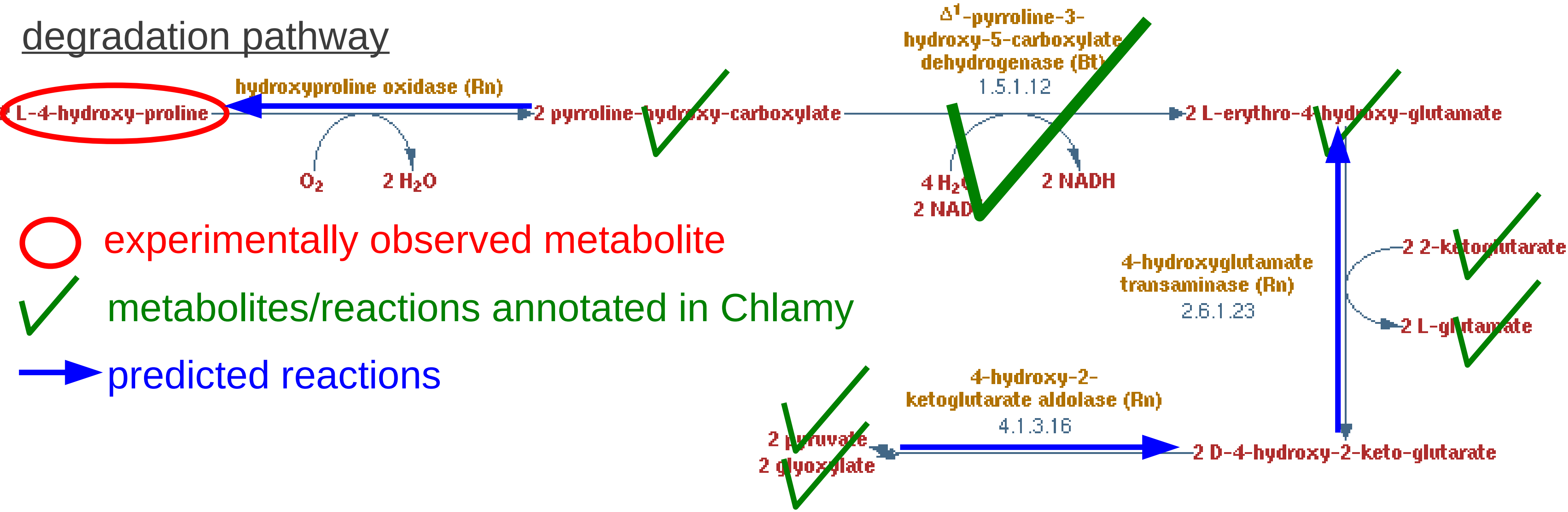
graphics from www.metacyc.org

Completion of a pathway

L-4-hydroxyproline

- in animals: important structural component of collagen
- in plants: found in some glycoproteins and cell wall proteins

degradation pathway



○ experimentally observed metabolite

✓ metabolites/reactions annotated in Chlamy

→ predicted reactions

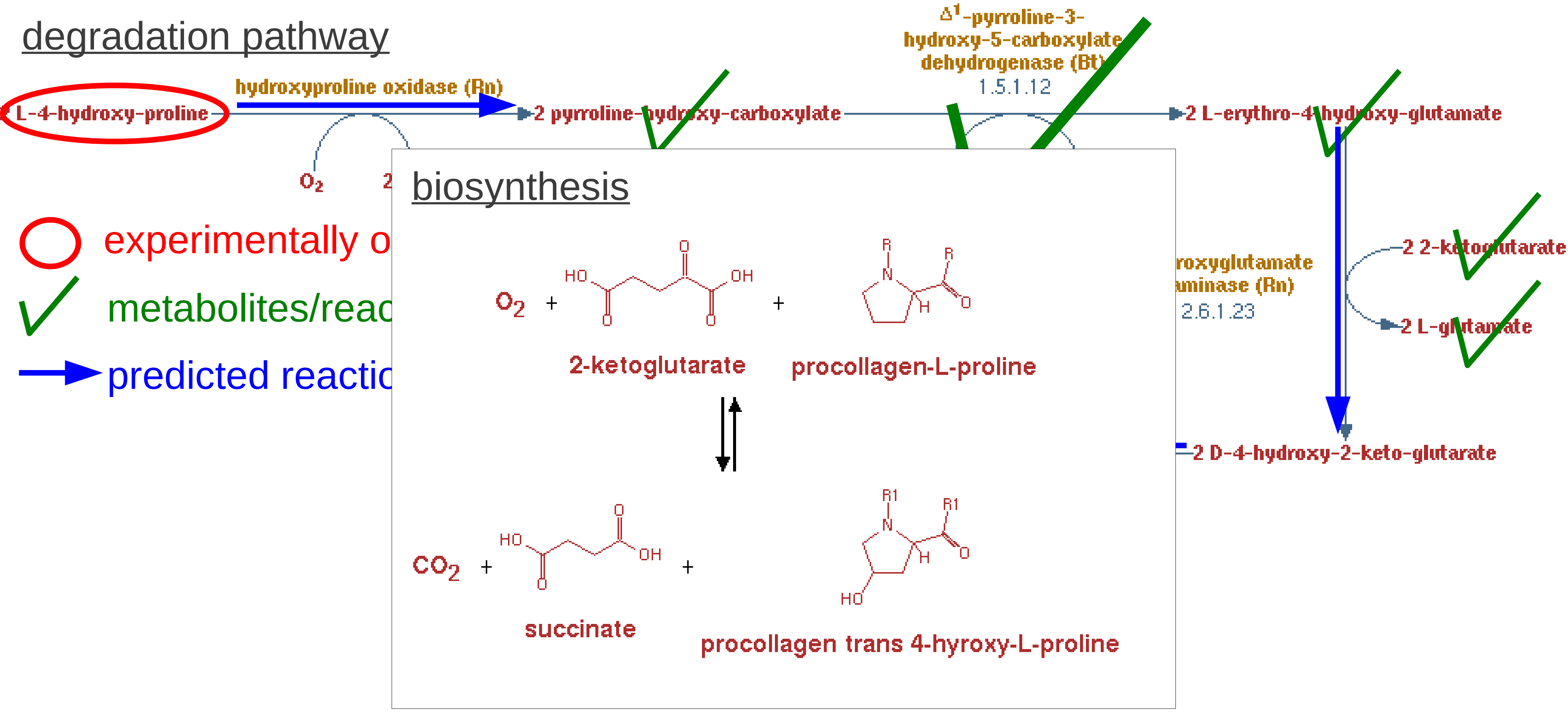
graphics from www.metacyc.org

Completion of a pathway

L-4-hydroxyproline

- in animals: important structural component of collagen
- in plants: found in some glycoproteins and cell wall proteins

degradation pathway



○ experimentally o

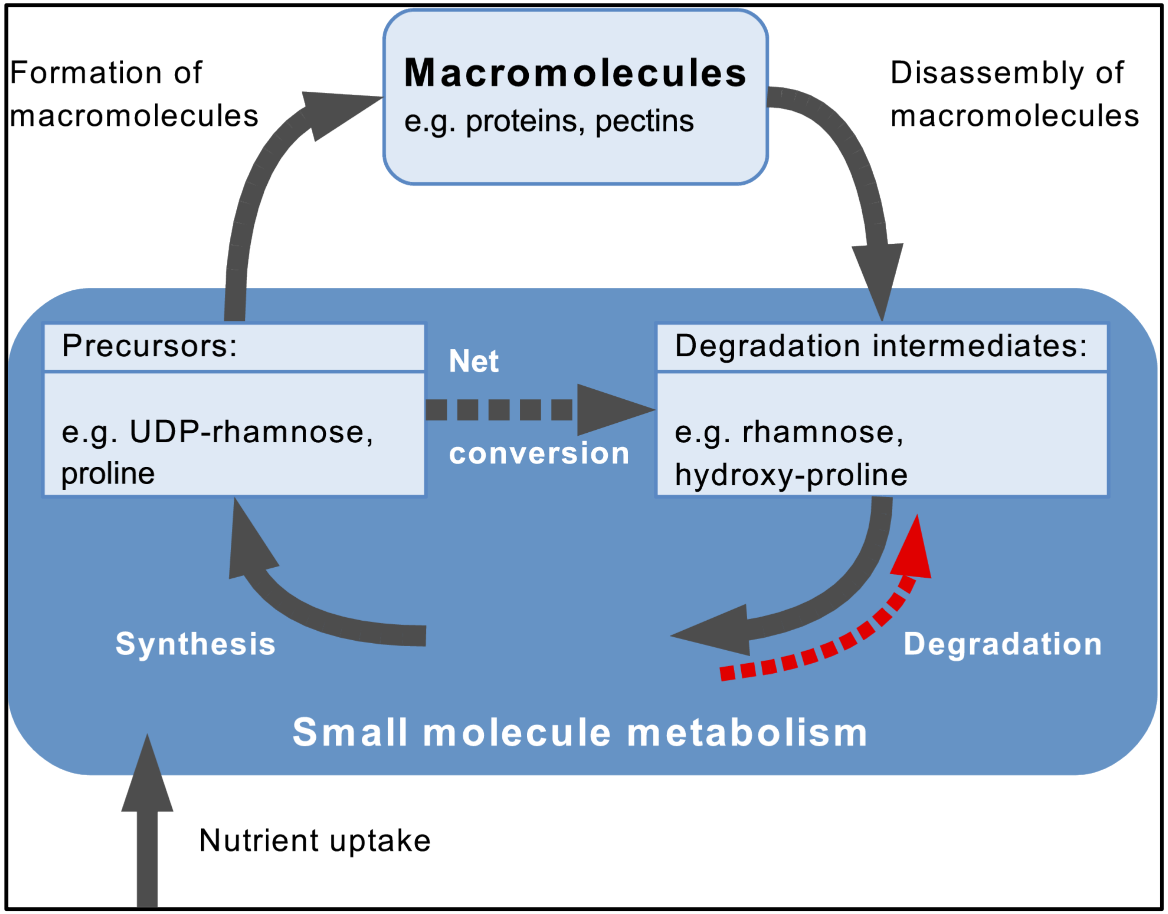
✓ metabolites/reac

→ predicted reactio

biosynthesis

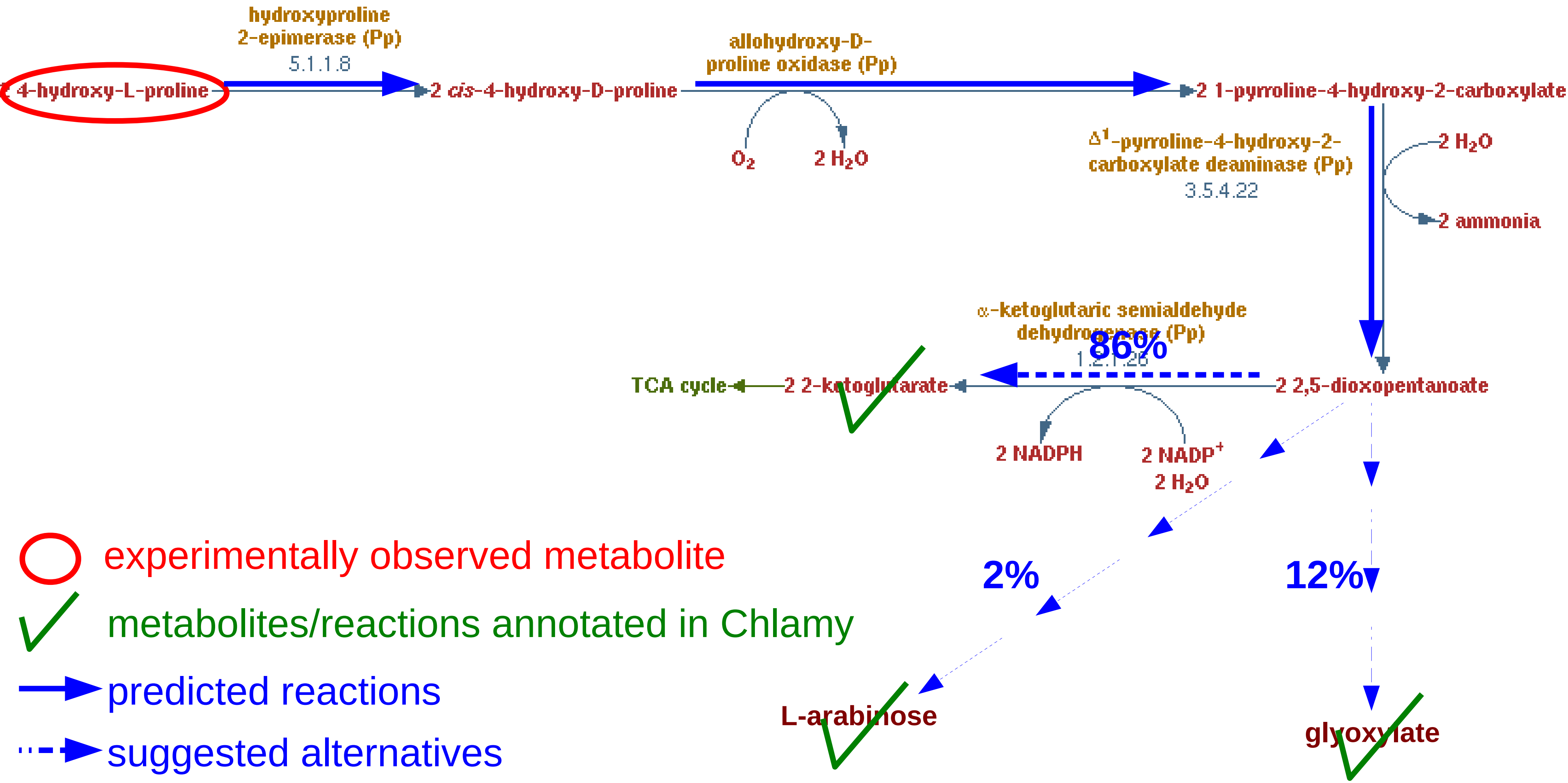
graphics from www.metacyc.org

Where are the limits of metabolism?



Prediction of alternative routes

4-hydroxy-L-proline



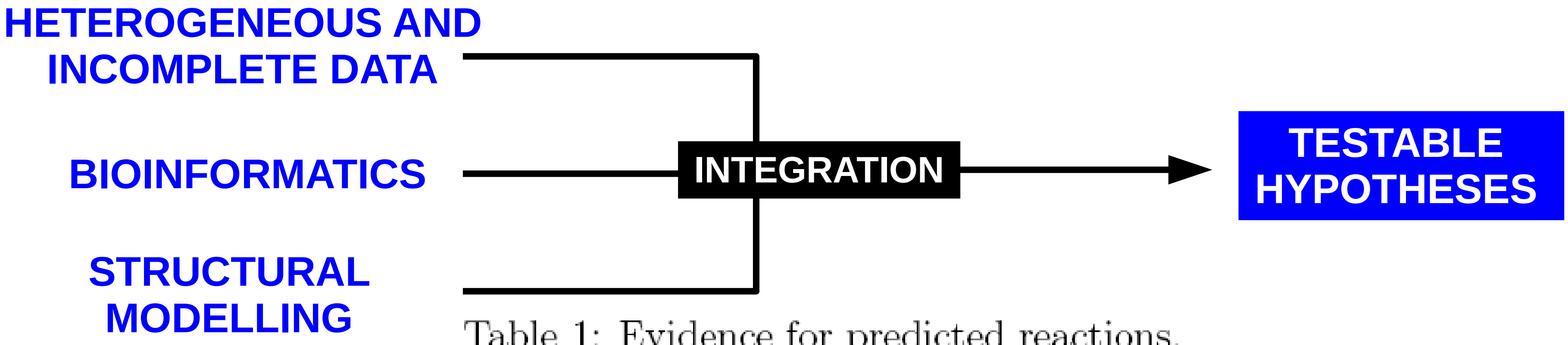
○ experimentally observed metabolite

✓ metabolites/reactions annotated in Chlamy

→ predicted reactions

⇢ suggested alternatives

graphics from www.metacyc.org



HETEROGENEOUS AND INCOMPLETE DATA

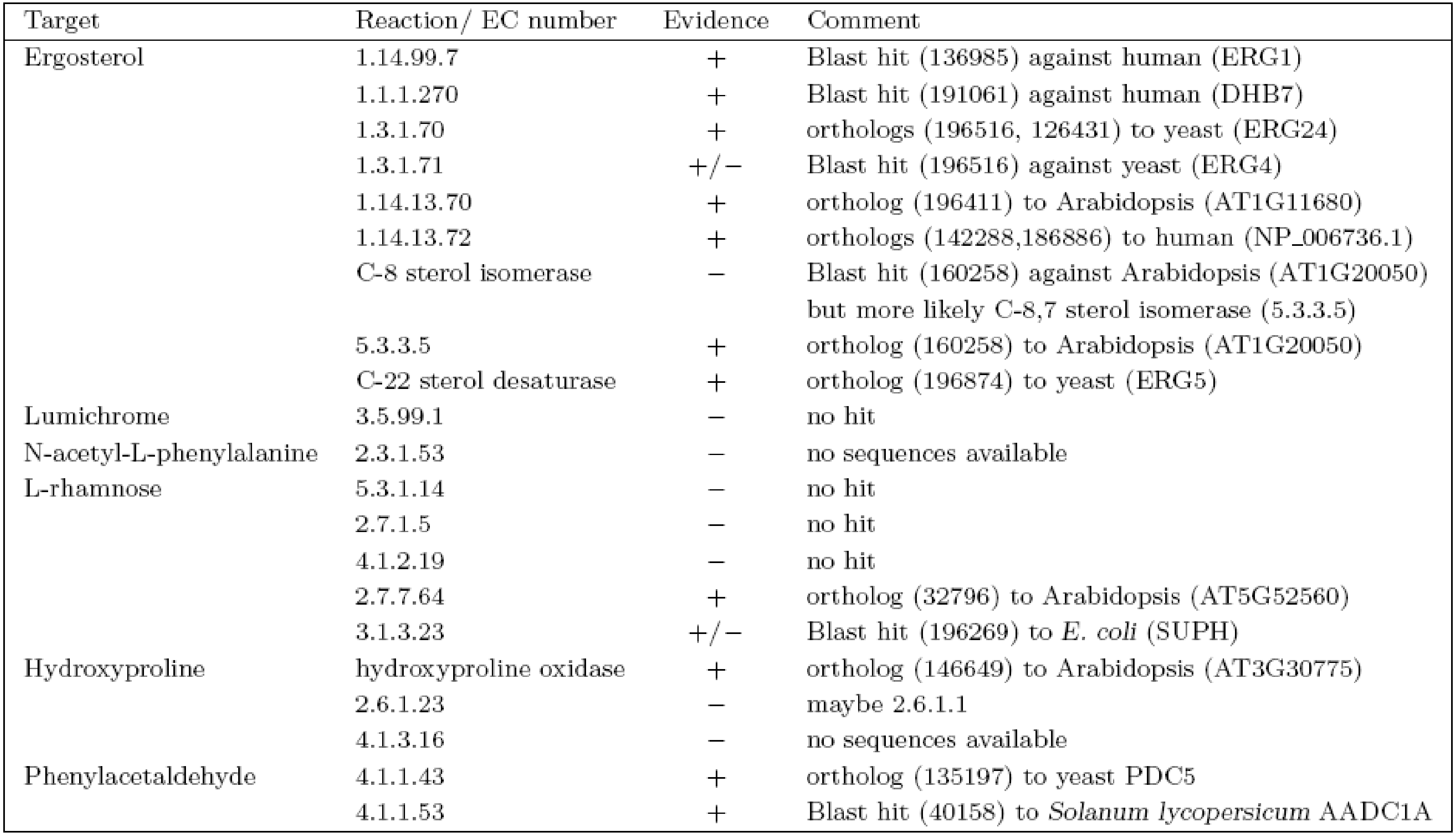
BIOINFORMATICS

STRUCTURAL MODELLING

INTEGRATION

TESTABLE HYPOTHESES

Table 1: Evidence for predicted reactions.

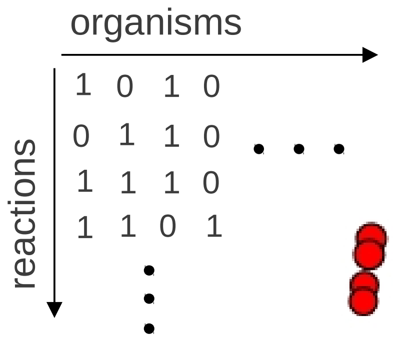


Target	Reaction/ EC number	Evidence	Comment
Ergosterol	1.14.99.7	+	Blast hit (136985) against human (ERG1)
	1.1.1.270	+	Blast hit (191061) against human (DHB7)
	1.3.1.70	+	orthologs (196516, 126431) to yeast (ERG24)
	1.3.1.71	+/−	Blast hit (196516) against yeast (ERG4)
	1.14.13.70	+	ortholog (196411) to Arabidopsis (AT1G11680)
	1.14.13.72	+	orthologs (142288,186886) to human (NP_006736.1)
	C-8 sterol isomerase	−	Blast hit (160258) against Arabidopsis (AT1G20050) but more likely C-8,7 sterol isomerase (5.3.3.5)
	5.3.3.5	+	ortholog (160258) to Arabidopsis (AT1G20050)
	C-22 sterol desaturase	+	ortholog (196874) to yeast (ERG5)
Lumichrome	3.5.99.1	−	no hit
N-acetyl-L-phenylalanine	2.3.1.53	−	no sequences available
L-rhamnose	5.3.1.14	−	no hit
	2.7.1.5	−	no hit
	4.1.2.19	−	no hit
	2.7.7.64	+	ortholog (32796) to Arabidopsis (AT5G52560)
	3.1.3.23	+/−	Blast hit (196269) to *E. coli* (SUPH)
Hydroxyproline	hydroxyproline oxidase	+	ortholog (146649) to Arabidopsis (AT3G30775)
	2.6.1.23	−	maybe 2.6.1.1
	4.1.3.16	−	no sequences available
Phenylacetaldehyde	4.1.1.43	+	ortholog (135197) to yeast PDC5
	4.1.1.53	+	Blast hit (40158) to *Solanum lycopersicum* AADC1A

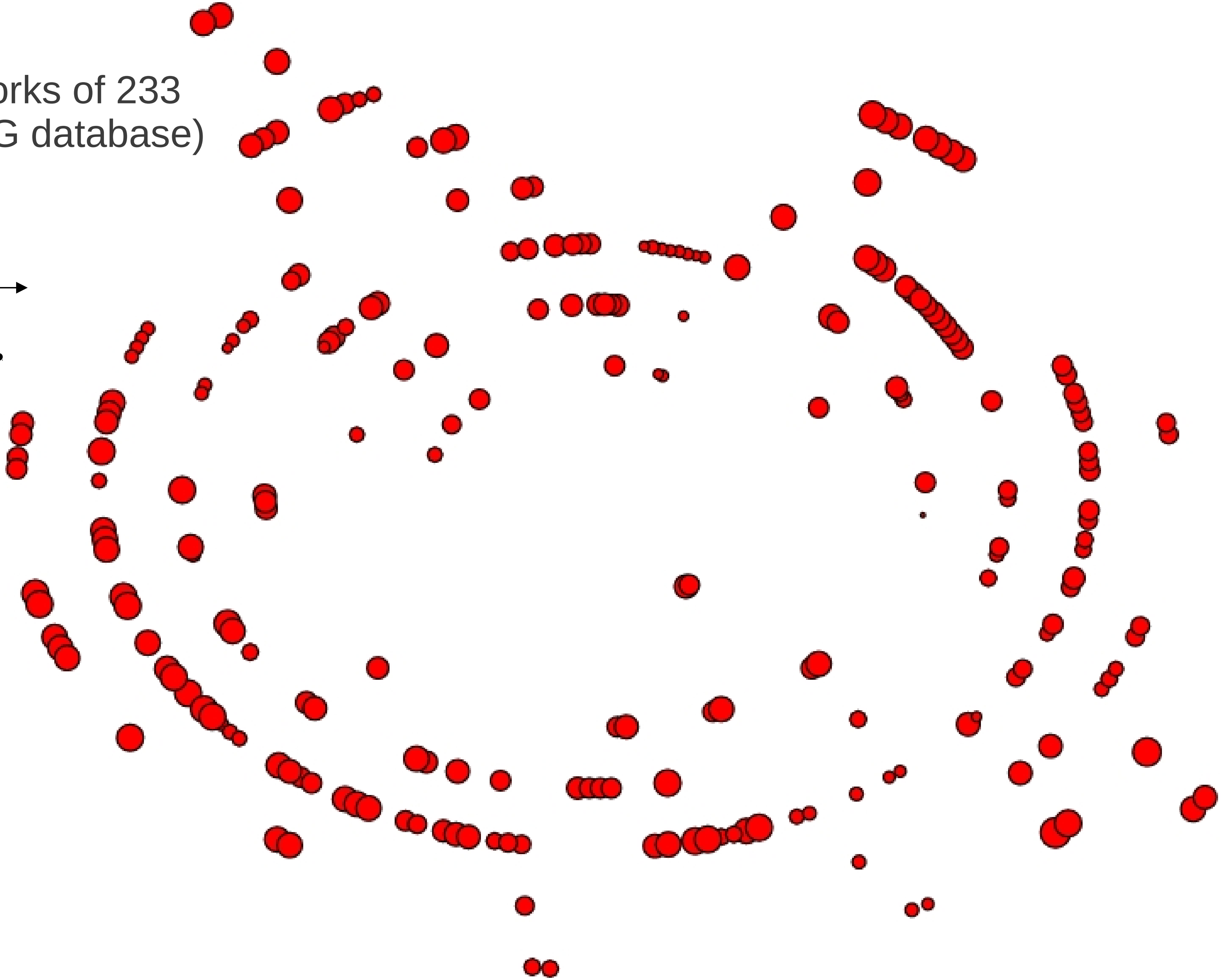
(Christian et al., Mol BioSystems, 2009)

Evolution of metabolic functions

Metabolic networks of 233 organisms (KEGG database)



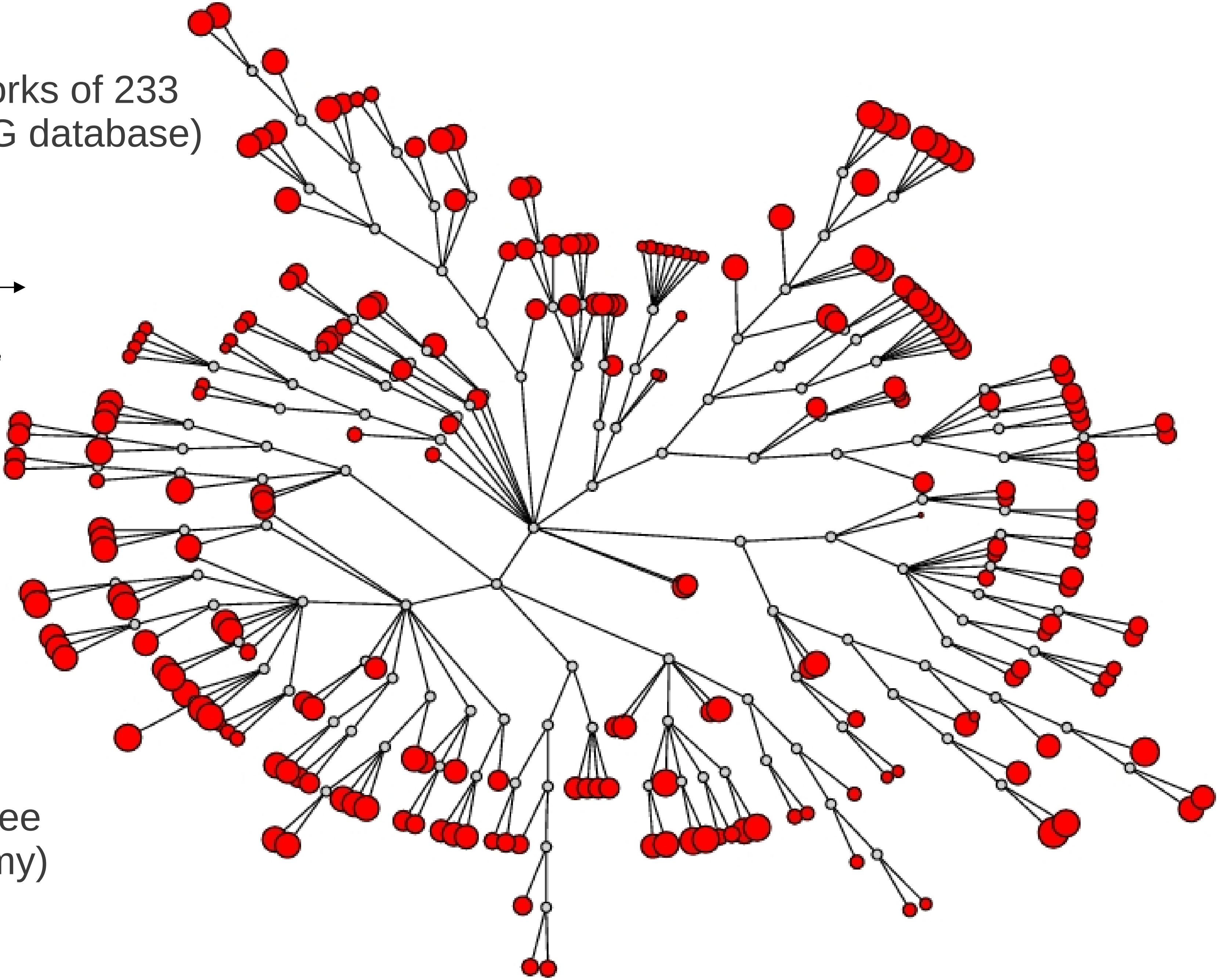
$$\begin{array}{c|cccccc} & \multicolumn{6}{c}{\text{organisms} \rightarrow} \\ \hline \text{reactions} \downarrow & 1 & 0 & 1 & 0 & & \\ & 0 & 1 & 1 & 0 & \cdots & \\ & 1 & 1 & 1 & 0 & & \\ & 1 & 1 & 0 & 1 & & \\ & & \vdots & & & & \end{array}$$



Evolution of metabolic functions

Metabolic networks of 233 organisms (KEGG database)

$$\begin{array}{cl} & \text{organisms} \longrightarrow \\ \text{reactions} \downarrow & \begin{matrix} 1 & 0 & 1 & 0 & \\ 0 & 1 & 1 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \\ 1 & 1 & 0 & 1 & \\ & \vdots & & & \end{matrix} \end{array}$$



Evolutionary tree (NCBI Taxonomy)

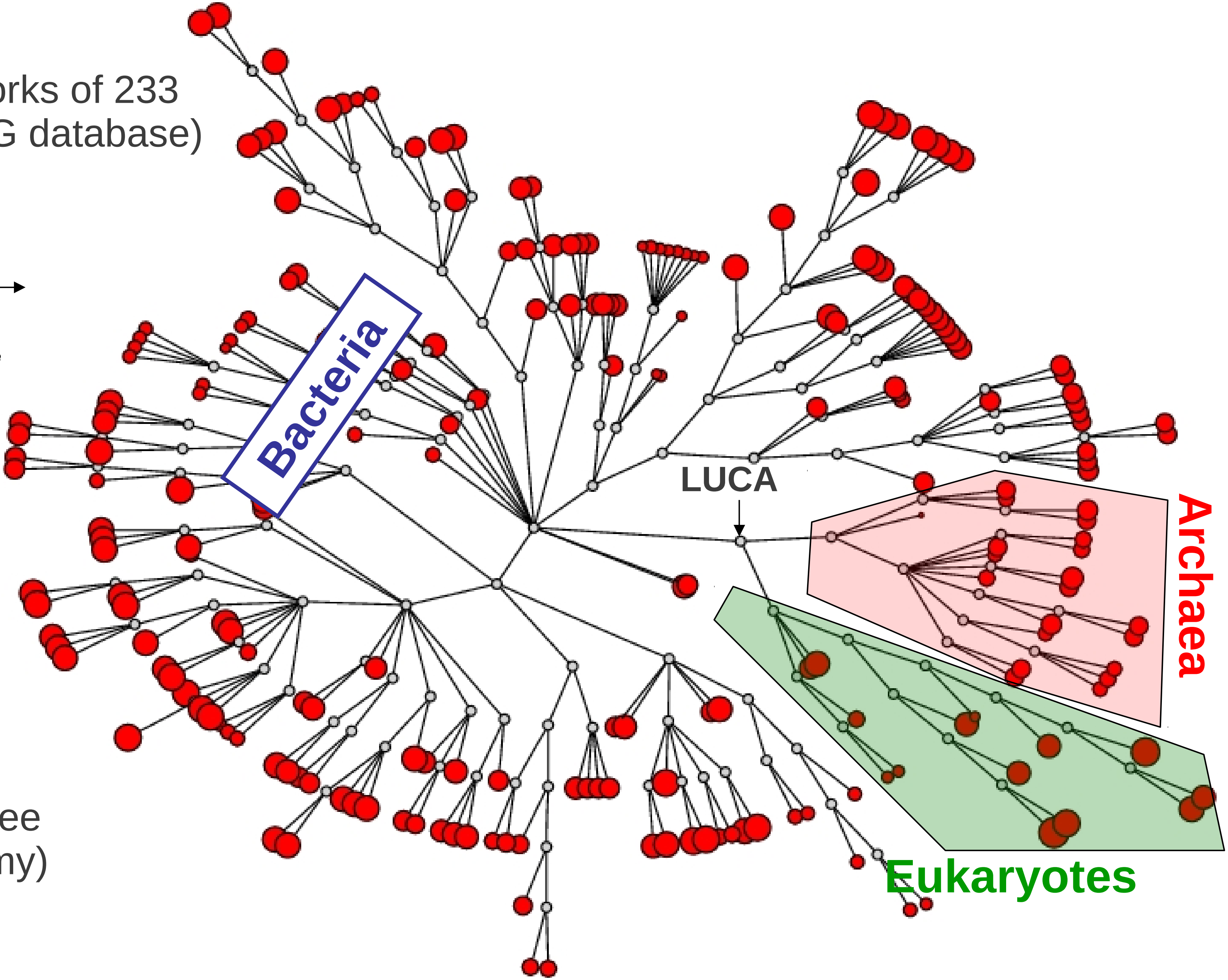
The tree of life

Metabolic networks of 233 organisms (KEGG database)

organisms

reactions

$$\begin{array}{ccccc} 1 & 0 & 1 & 0 & \\ 0 & 1 & 1 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \\ 1 & 1 & 0 & 1 & \\ & \vdots & & & \end{array}$$

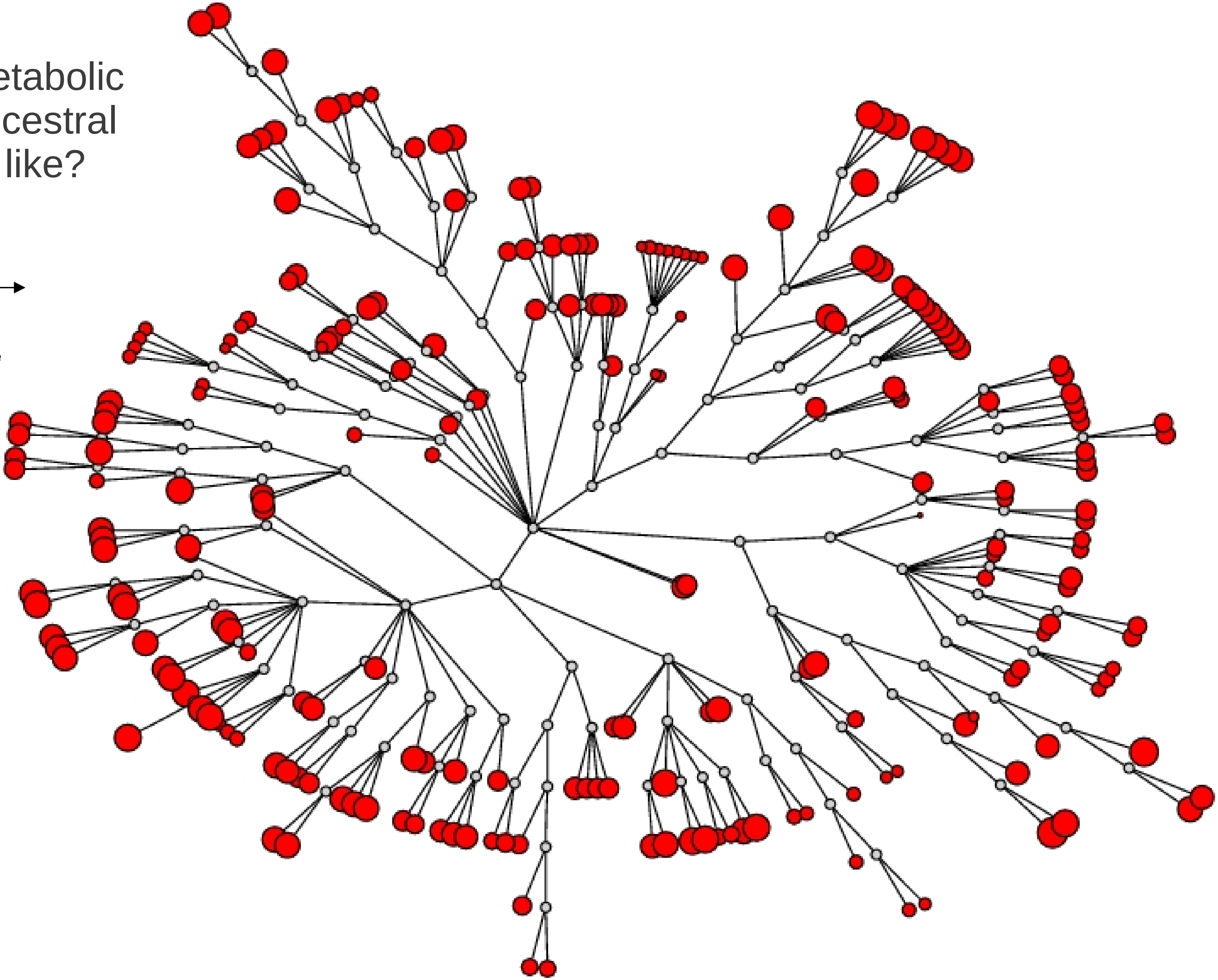


Evolutionary tree (NCBI Taxonomy)

Ancestral networks

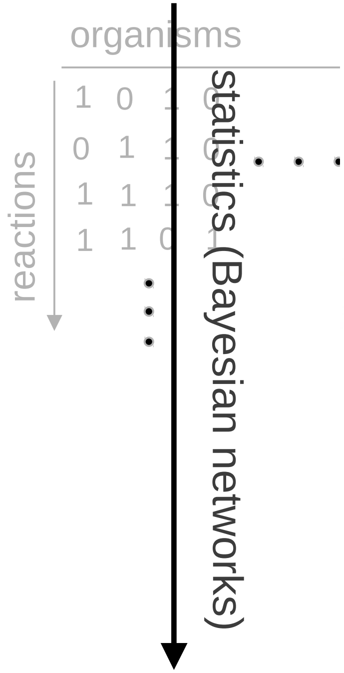
How did the metabolic networks of ancestral species look like?

$$\begin{array}{c|ccccc}
 & \multicolumn{5}{c}{\text{organisms} \longrightarrow} \\
\hline
\text{reactions} \downarrow & 1 & 0 & 1 & 0 & \\
 & 0 & 1 & 1 & 0 & \cdots \\
 & 1 & 1 & 1 & 0 & \\
 & 1 & 1 & 0 & 1 & \\
 & & \vdots & & &
\end{array}$$

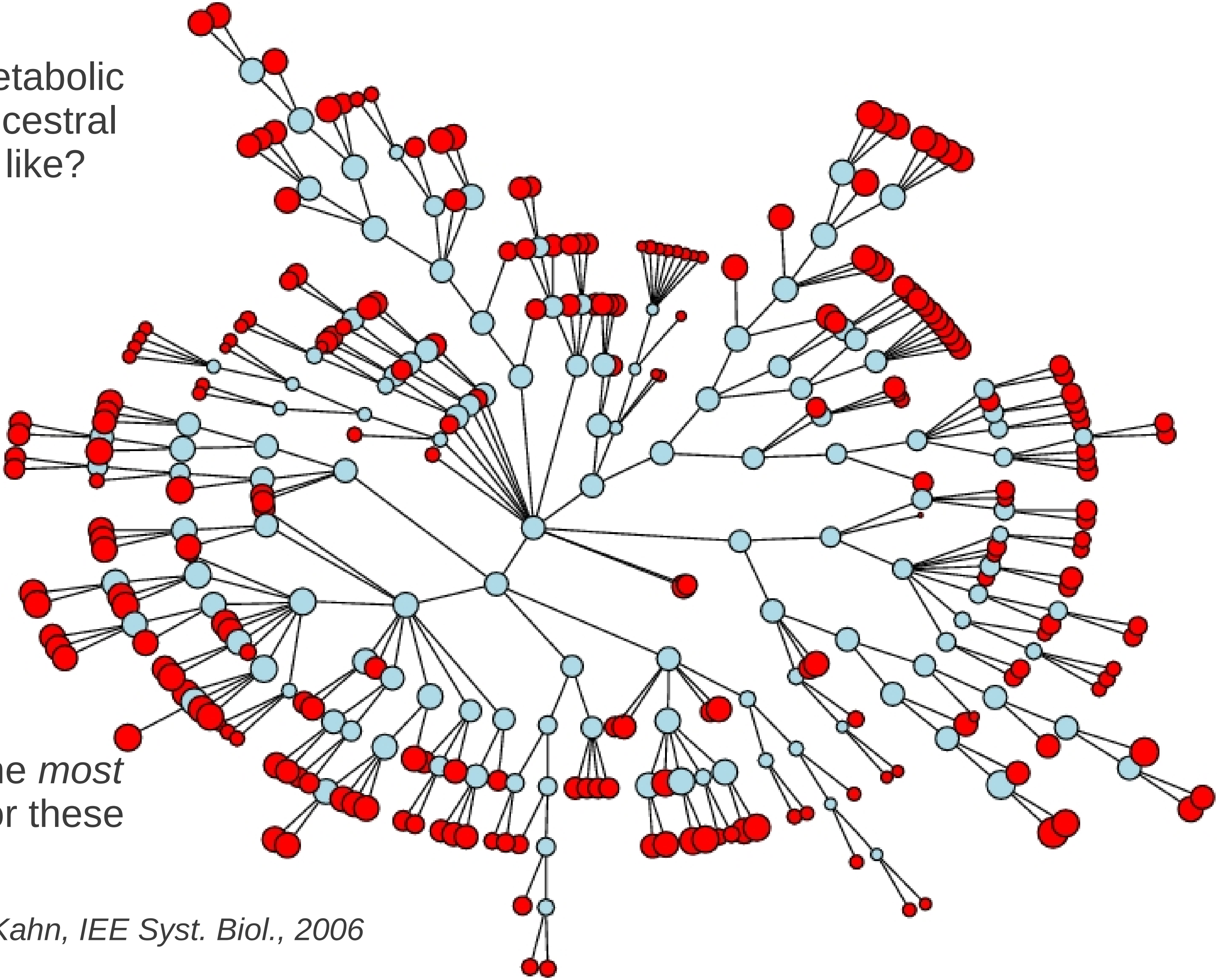


Reconstruction of ancestral networks

How did the metabolic networks of ancestral species look like?



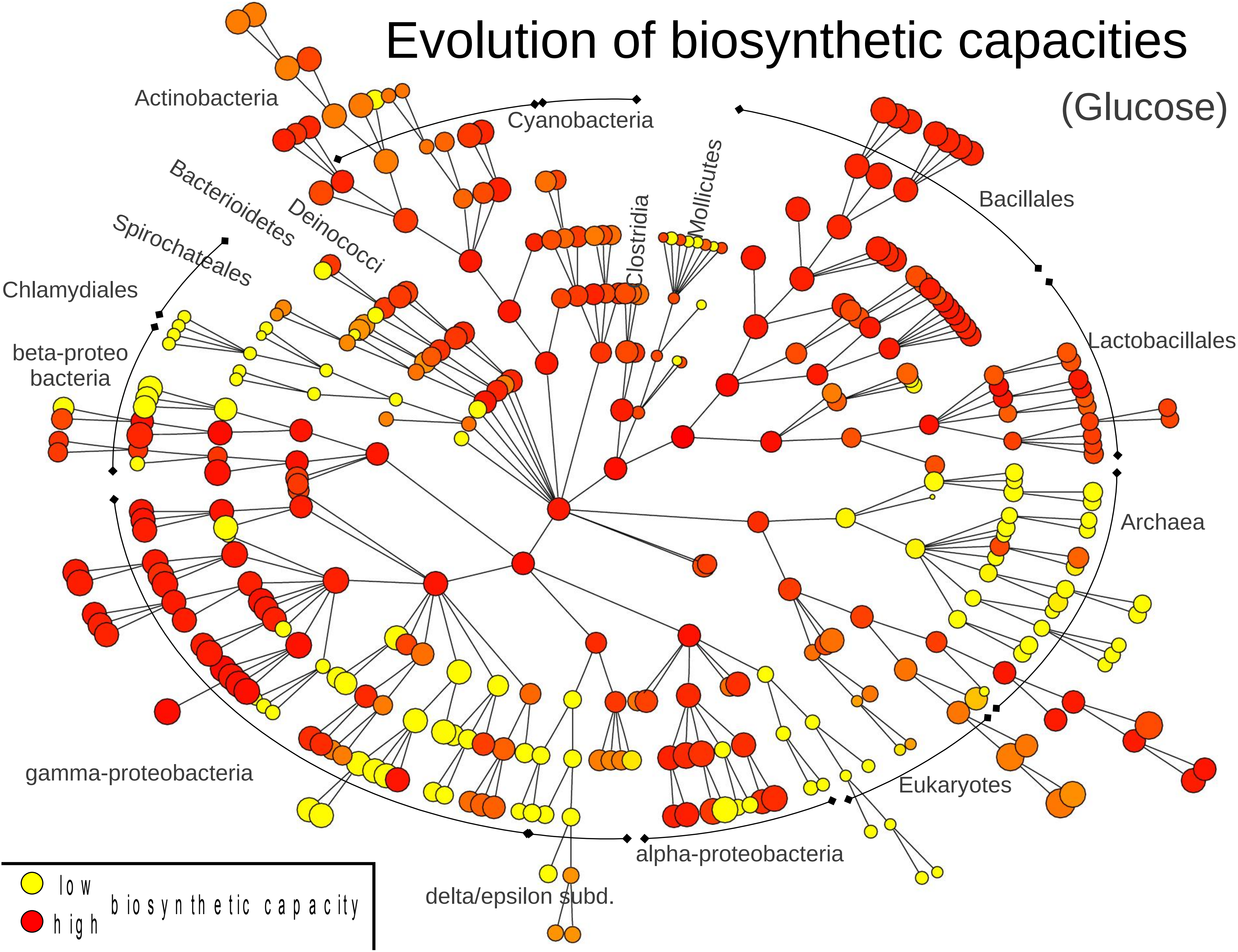
We calculated the *most likely scenario* for these networks



Ebenhöh, Handorf & Kahn, IEE Syst. Biol., 2006

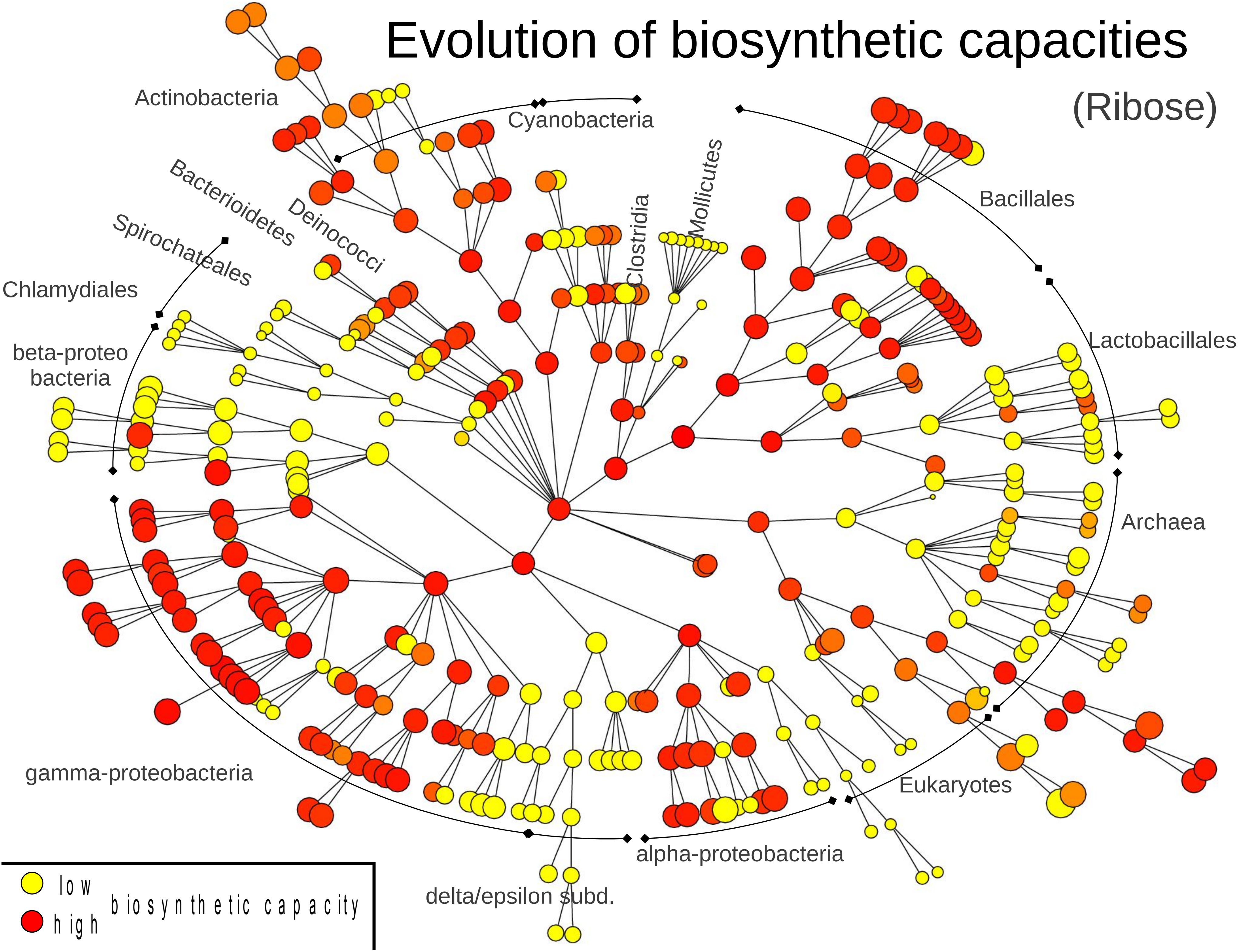
Evolution of biosynthetic capacities

(Glucose)



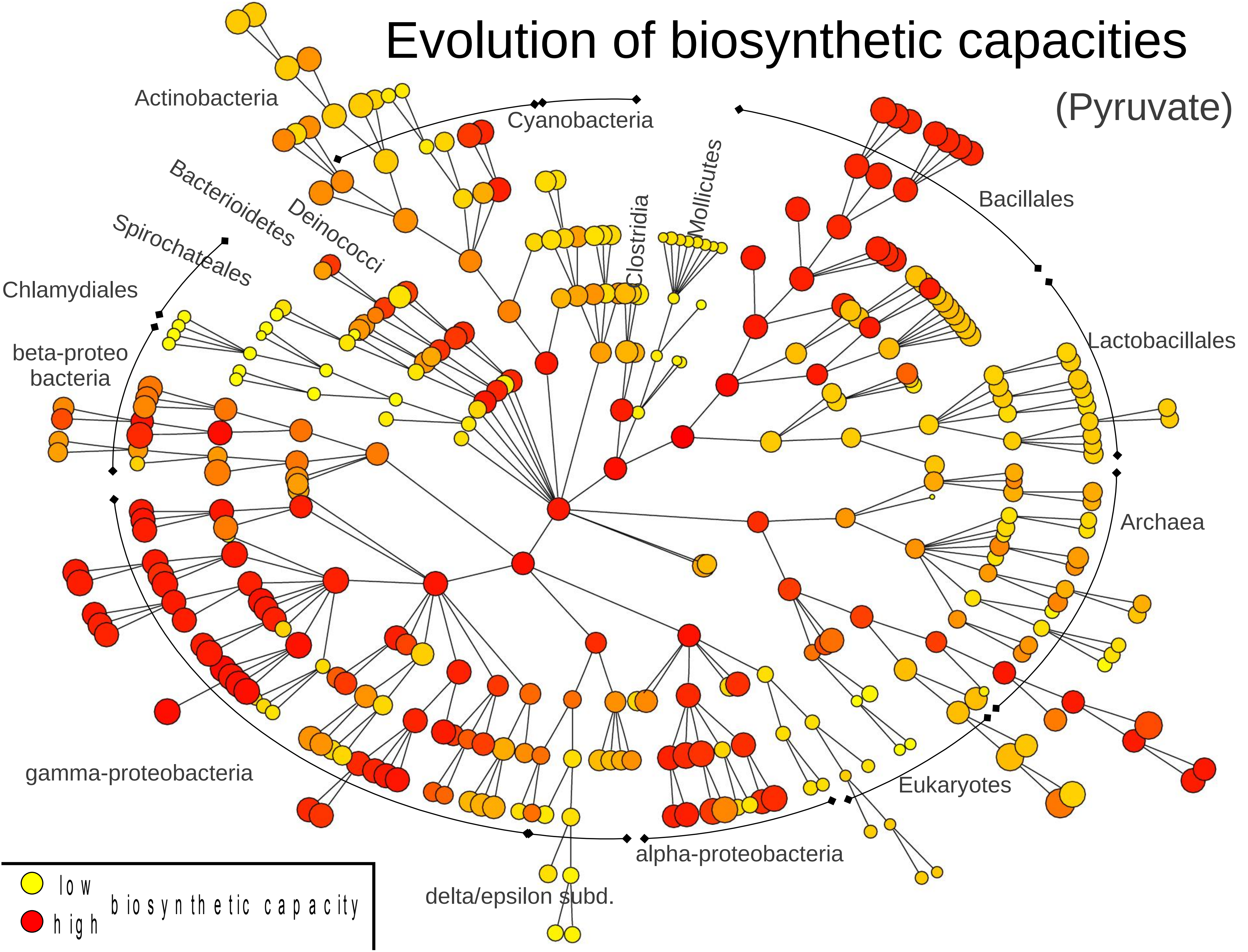
Evolution of biosynthetic capacities

(Ribose)



Evolution of biosynthetic capacities

(Pyruvate)

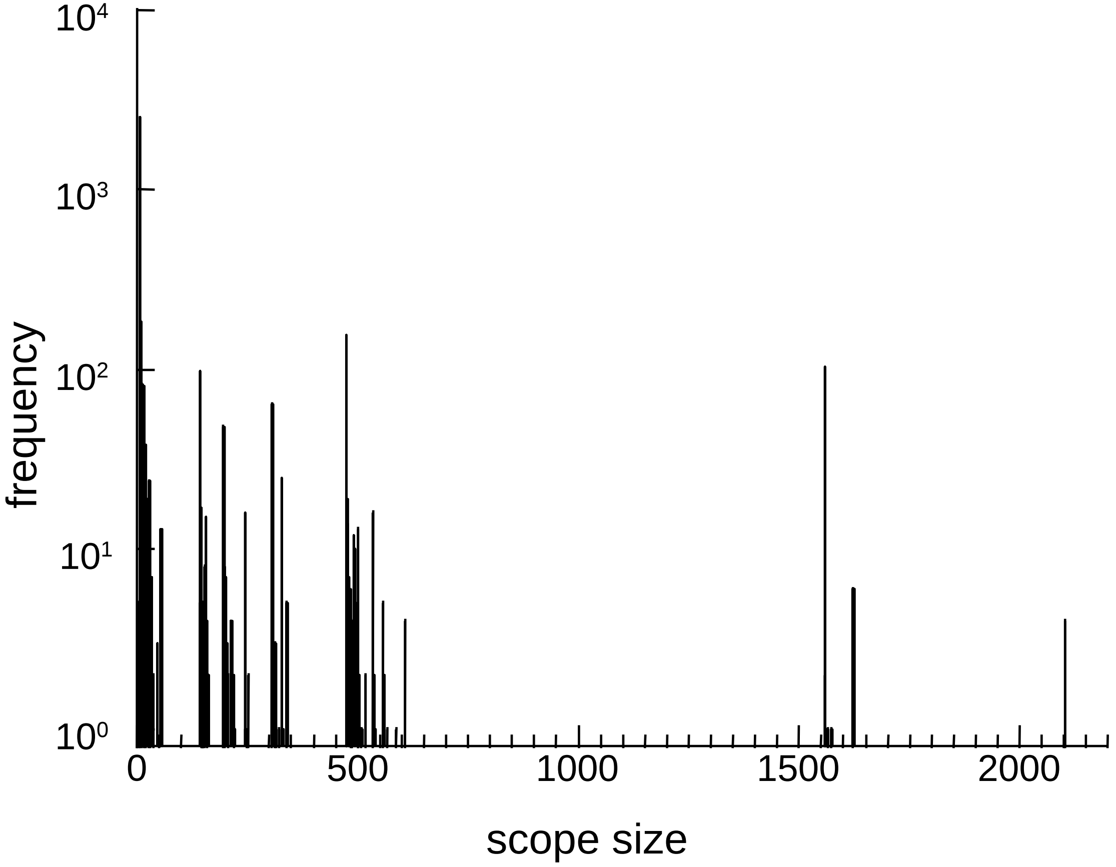


Analysis of the “global” network

The network containing all known biochemical reactions (biosphere network)

A scope of a ***single metabolite*** characterizes its ***biosynthetic potential***

Characterize all metabolites!



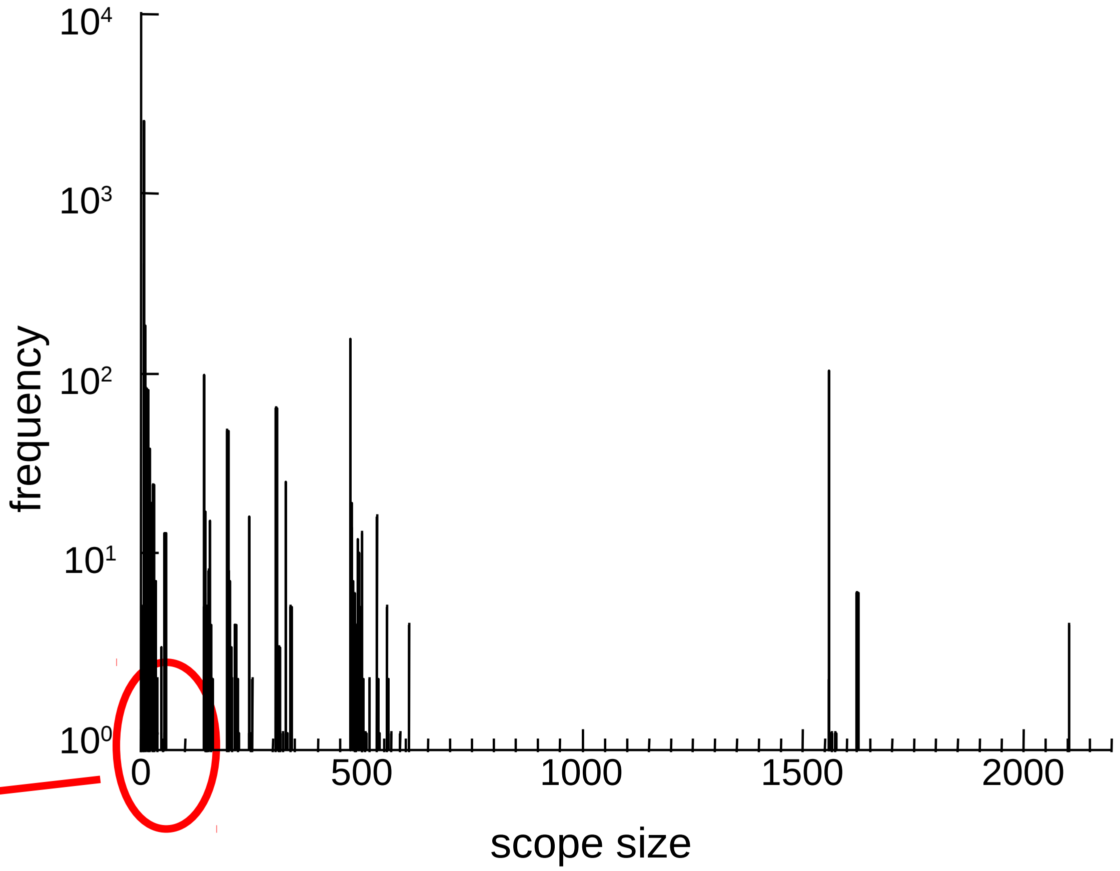
(Handorf, Ebenhöh and Heinrich, J. Mol. Evol., 2005)

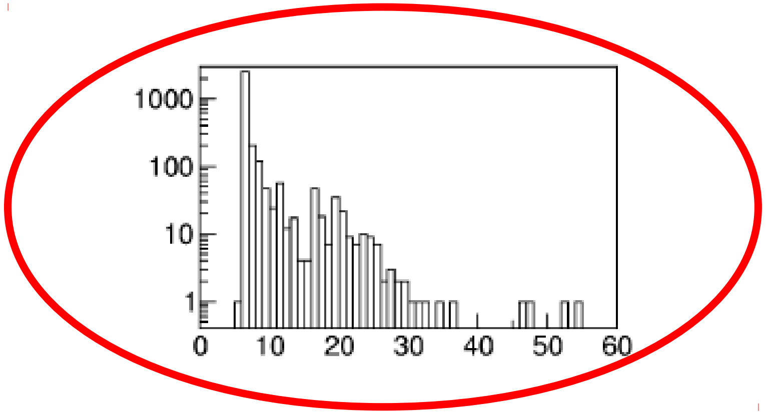
Analysis of the “global” network

A scope of a ***single metabolite*** characterizes its ***biosynthetic potential***

Characterize all metabolites!

MOST METABOLITES CARRY A LOW POTENTIAL





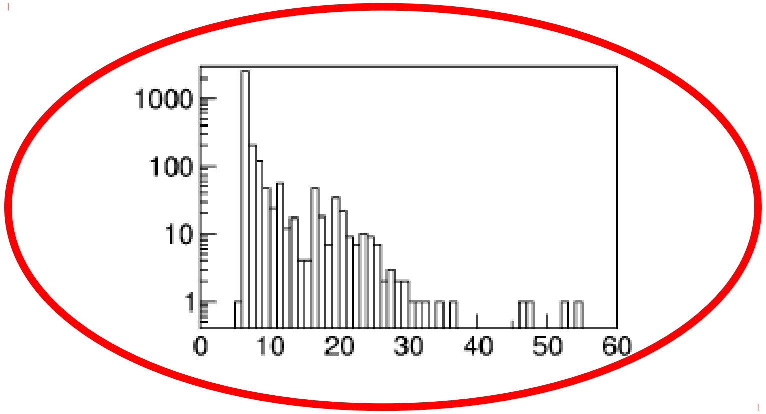
(Handorf, Ebenhöh and Heinrich, J. Mol. Evol., 2005)

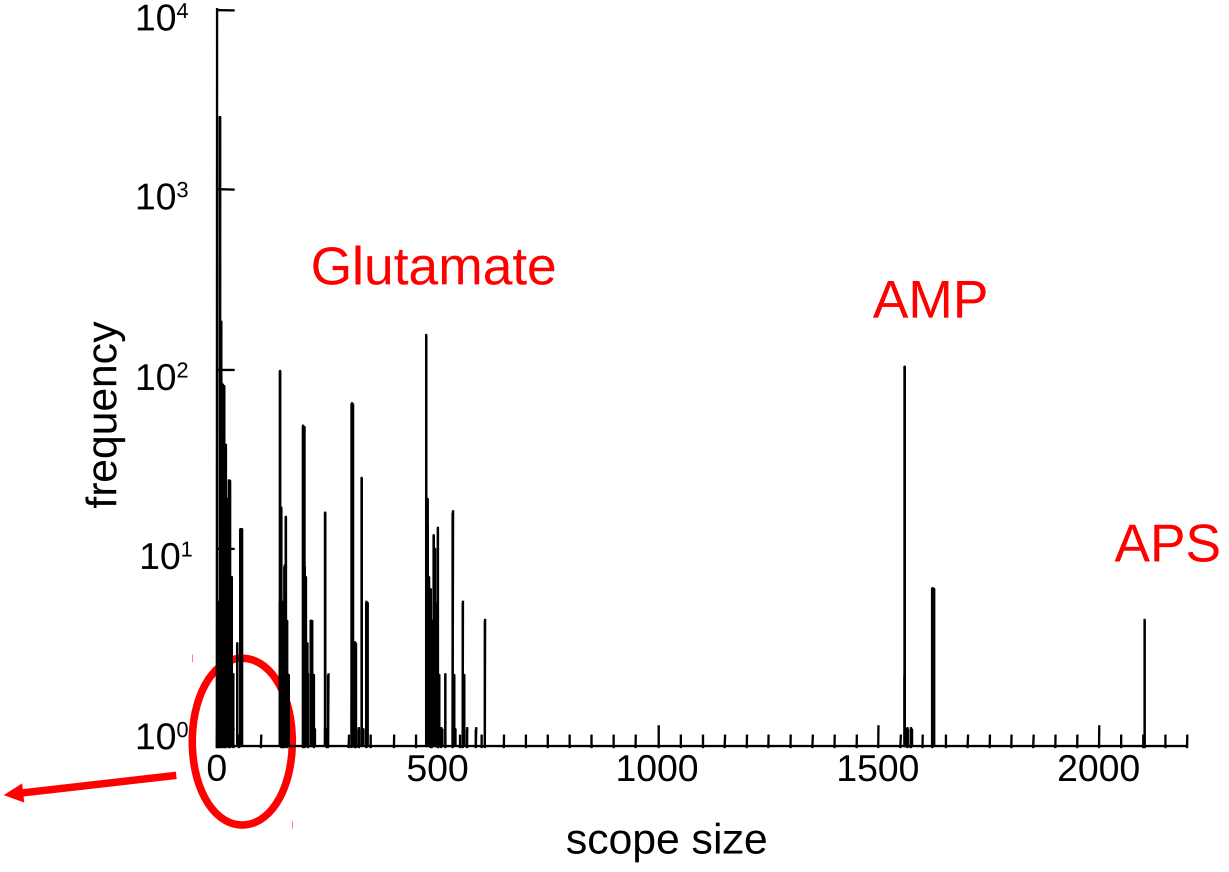
Analysis of the “global” network

A scope of a ***single metabolite***
characterizes its
biosynthetic potential

Characterize all metabolites!

MOST METABOLITES
CARRY A LOW POTENTIAL





(Handorf, Ebenhöh and Heinrich, J. Mol. Evol., 2005)

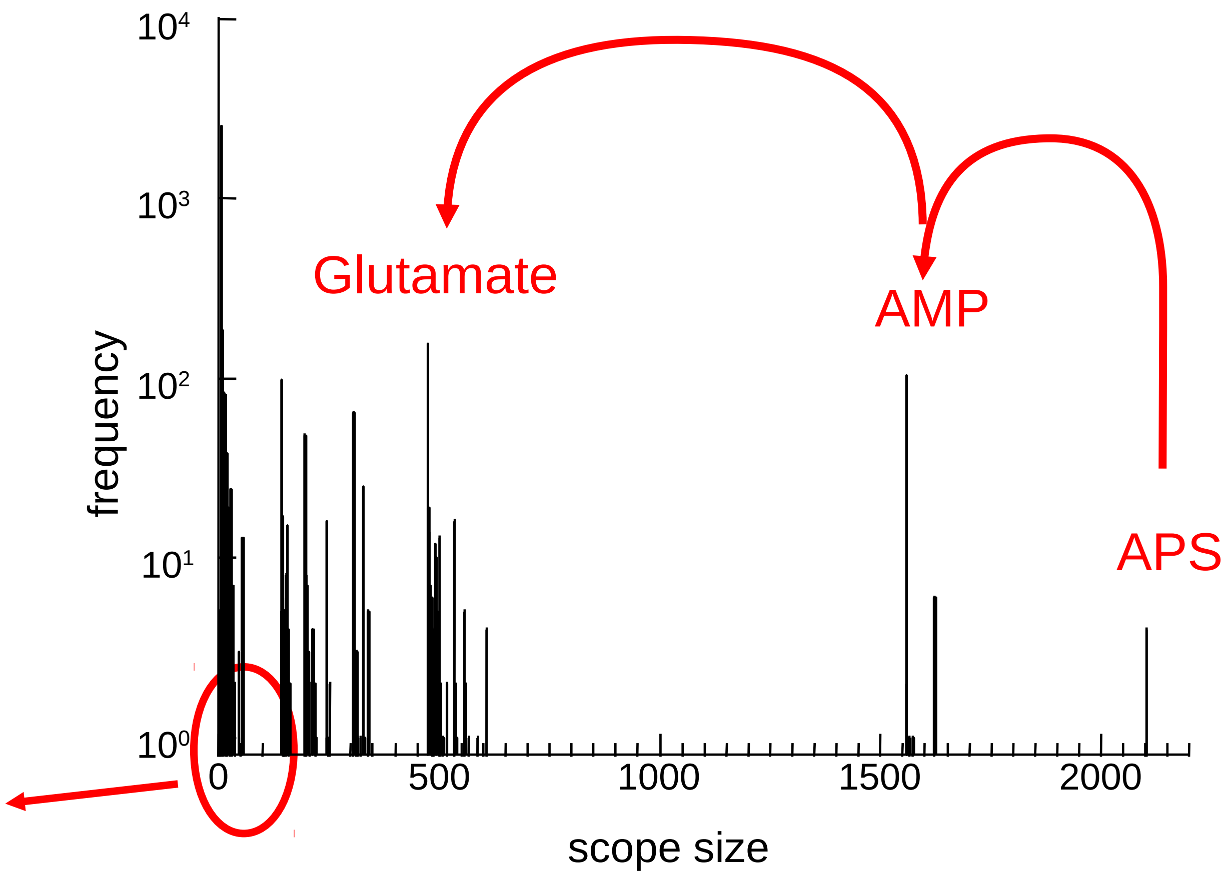
Analysis of the “global” network

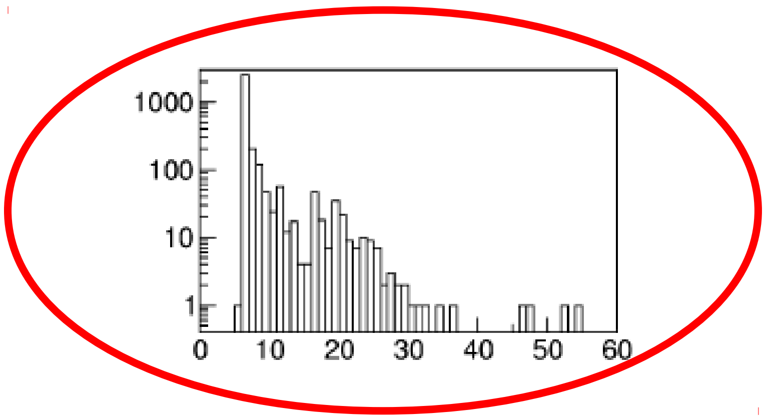
A scope of a ***single metabolite*** characterizes its ***biosynthetic potential***

SOME SCOPES ARE INCLUDED IN OTHERS

Characterize all metabolites!

MOST METABOLITES CARRY A LOW POTENTIAL





(Handorf, Ebenhöh and Heinrich, J. Mol. Evol., 2005)

Global organisation of metabolism

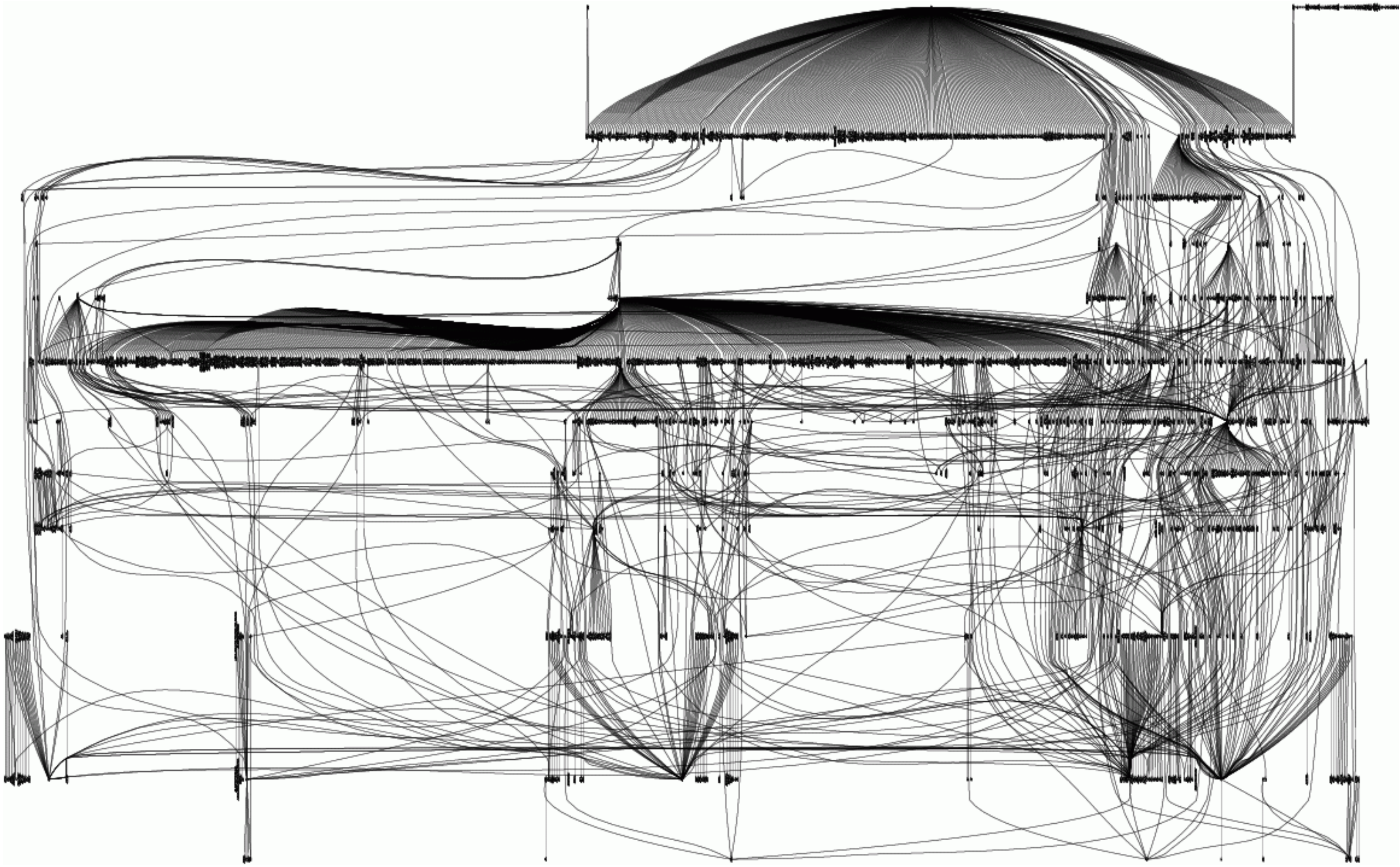
Reference network from KEGG

↓

calculate all single scopes

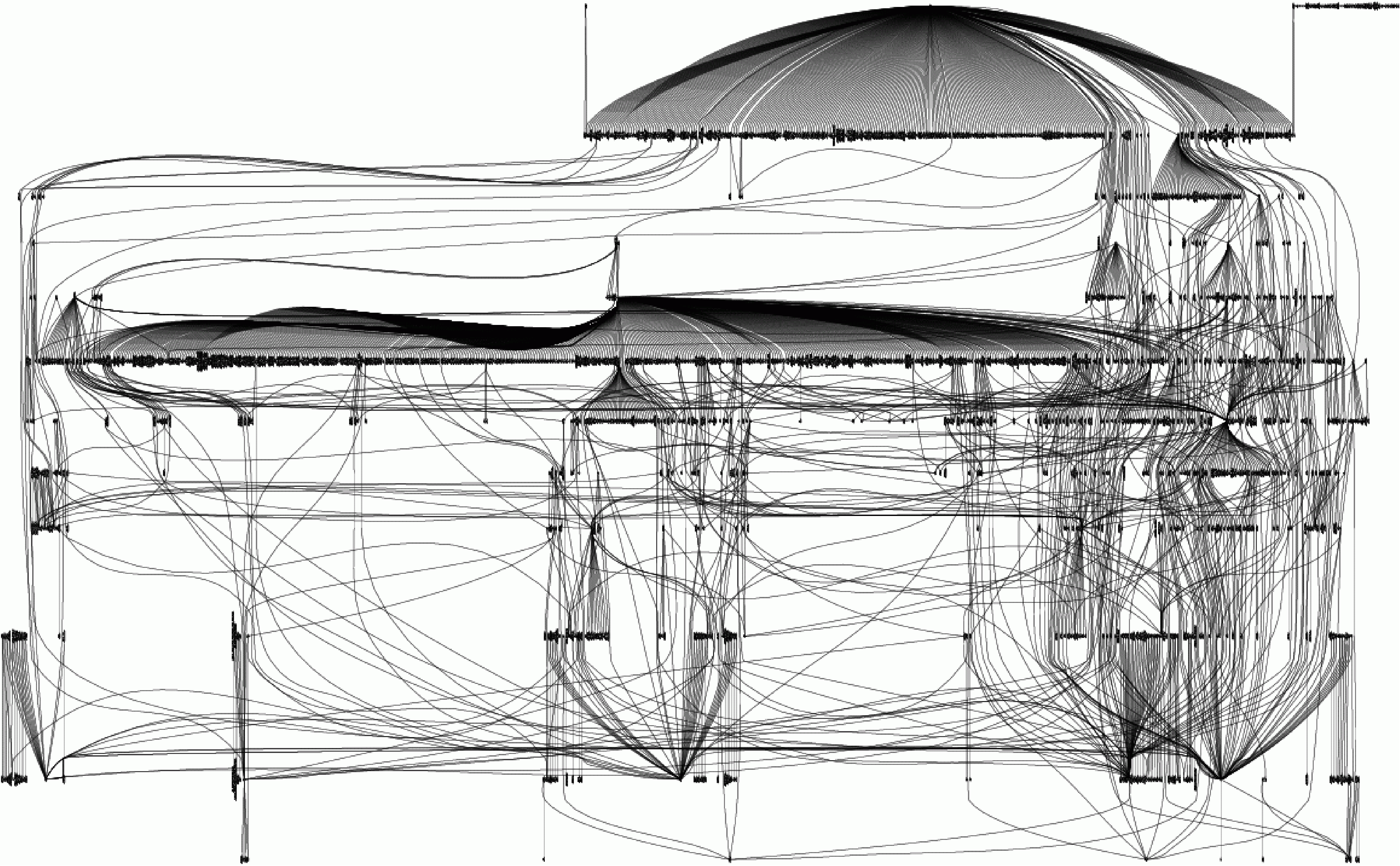
↓

determine all inclusion relations

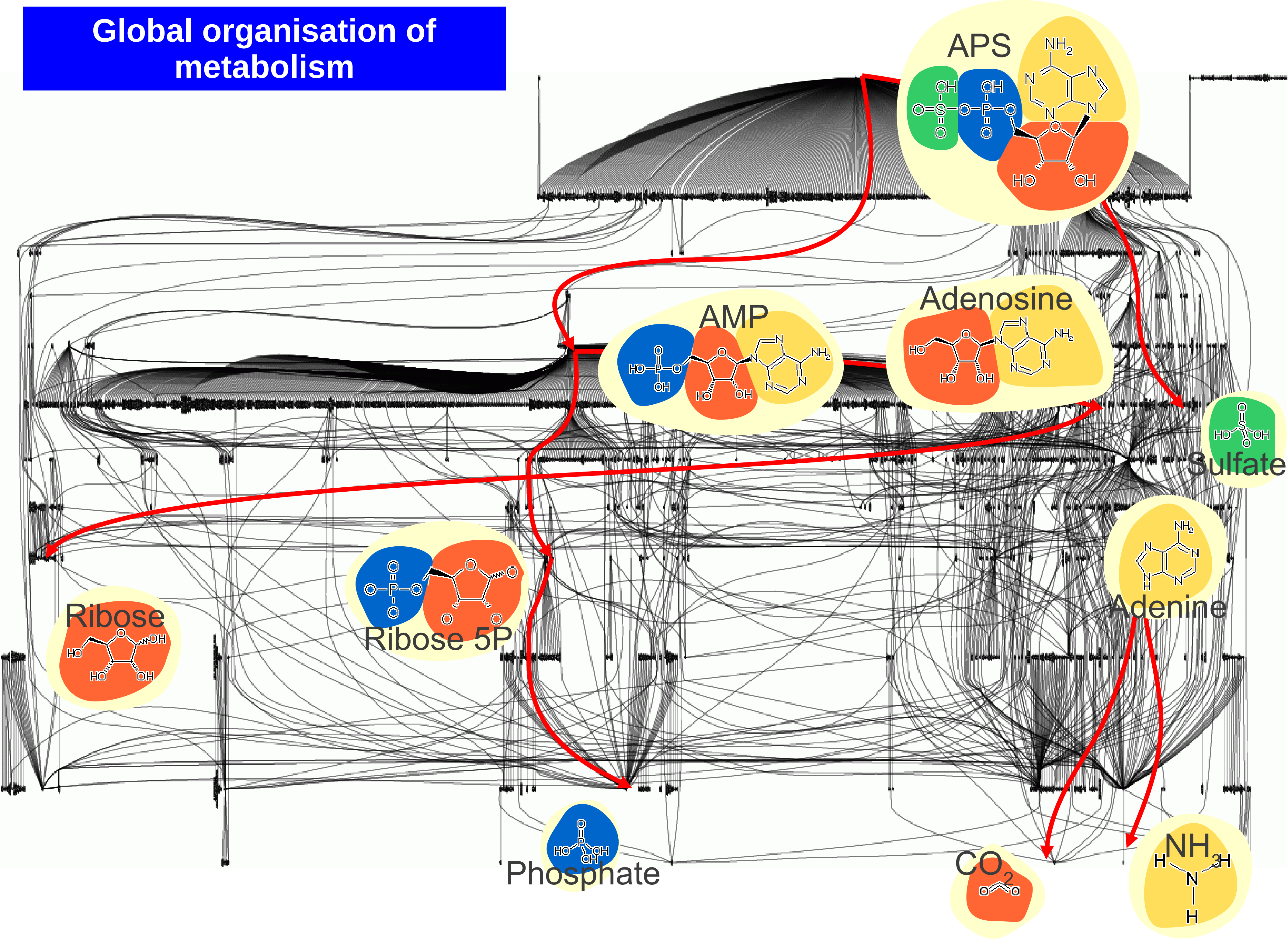


(Handorf, Ebenhöh, Kahn and Heinrich, IEE Syst Biol, 2006)

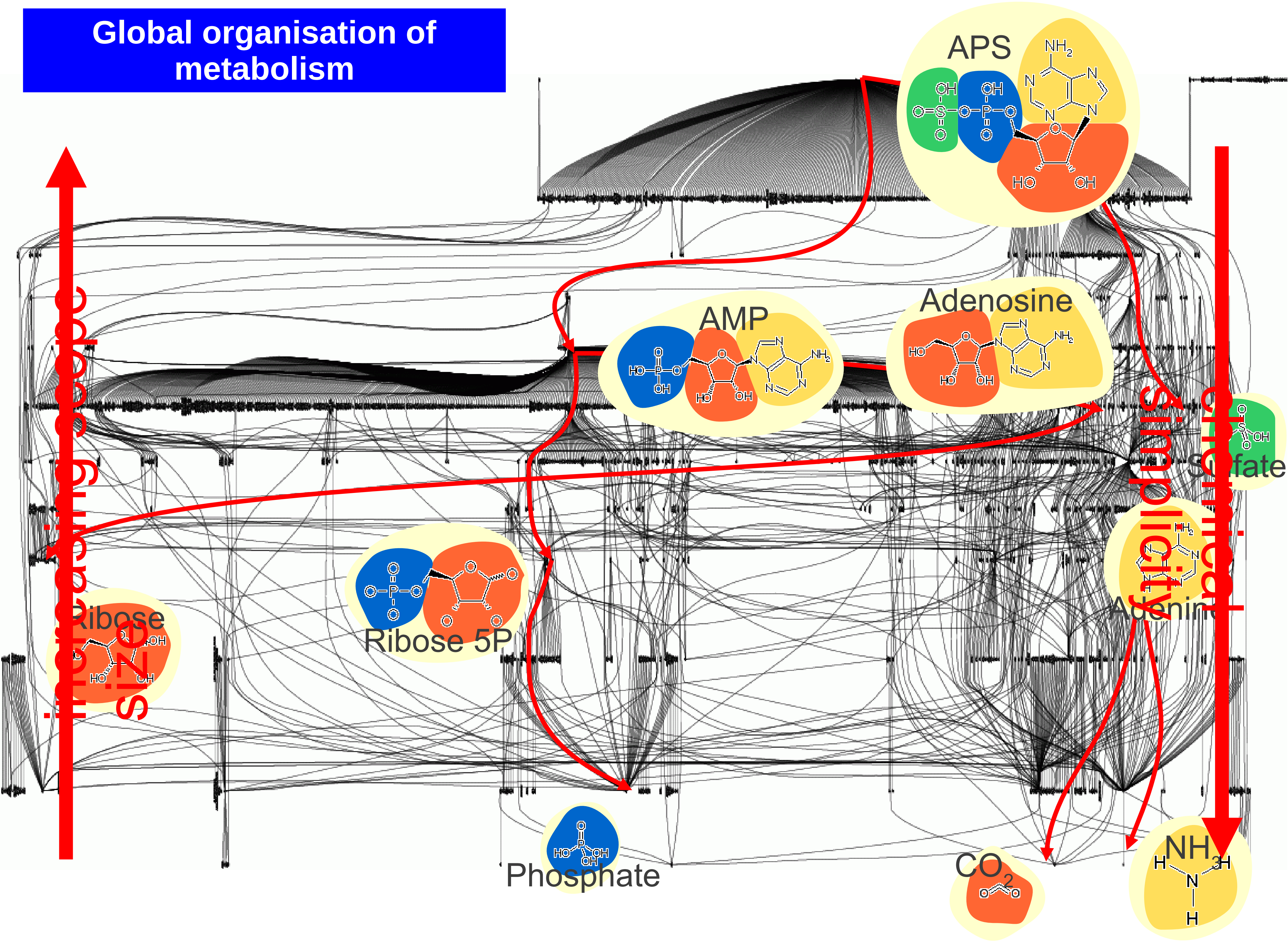
Global organisation of metabolism



Global organisation of metabolism



Global organisation of metabolism



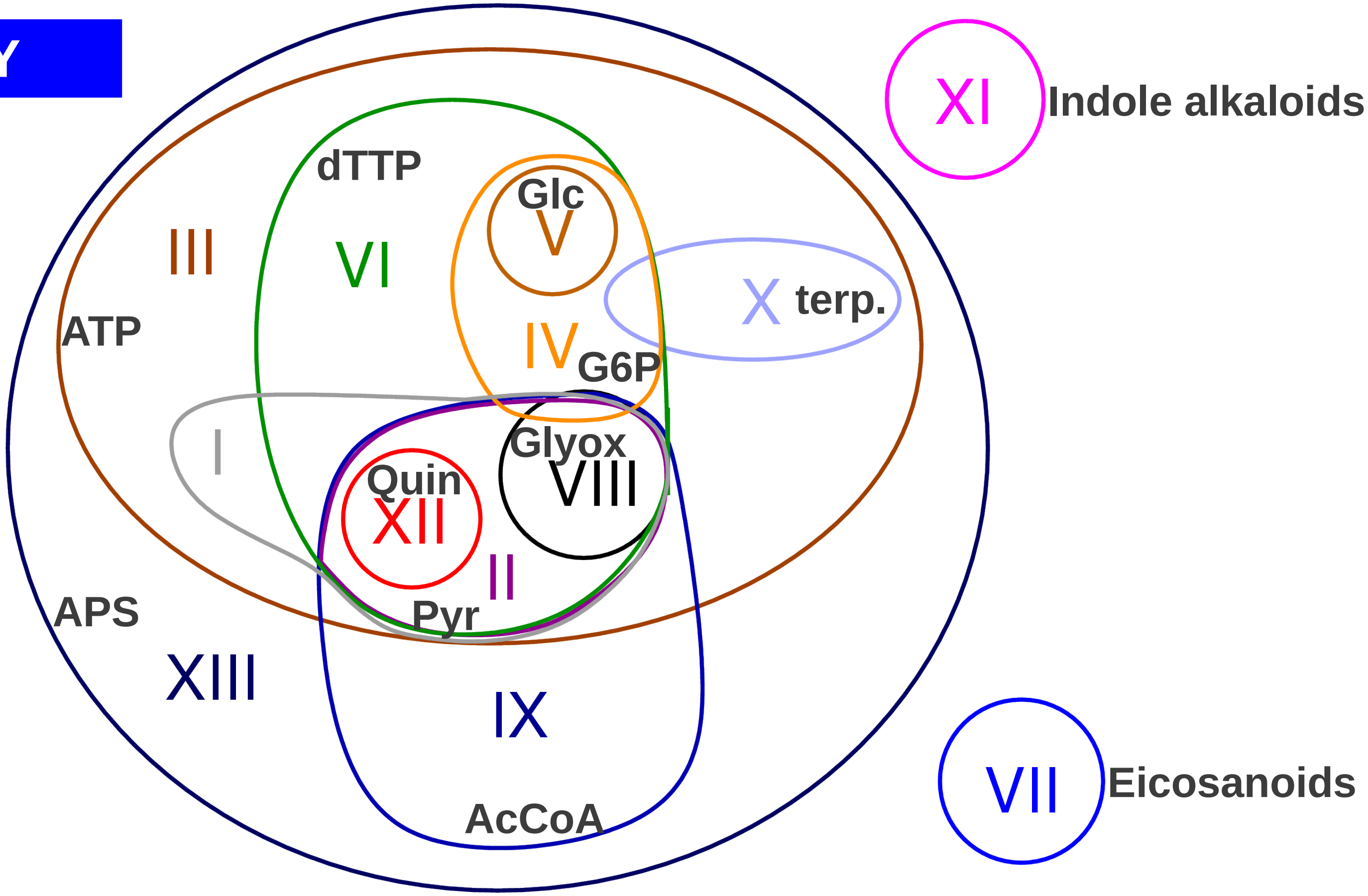
Similarity of biosynthetic potentials

Many metabolites carry ***similar*** biosynthetic potentials

⇨ ***"consensus scope"***
(typical biosynthetic potential)

Groups with similar potentials can be identified by ***clustering analysis***

HIERARCHY



(Matthäus, Salazar and Ebenhöh, PLoS Comp Biol, 2008)

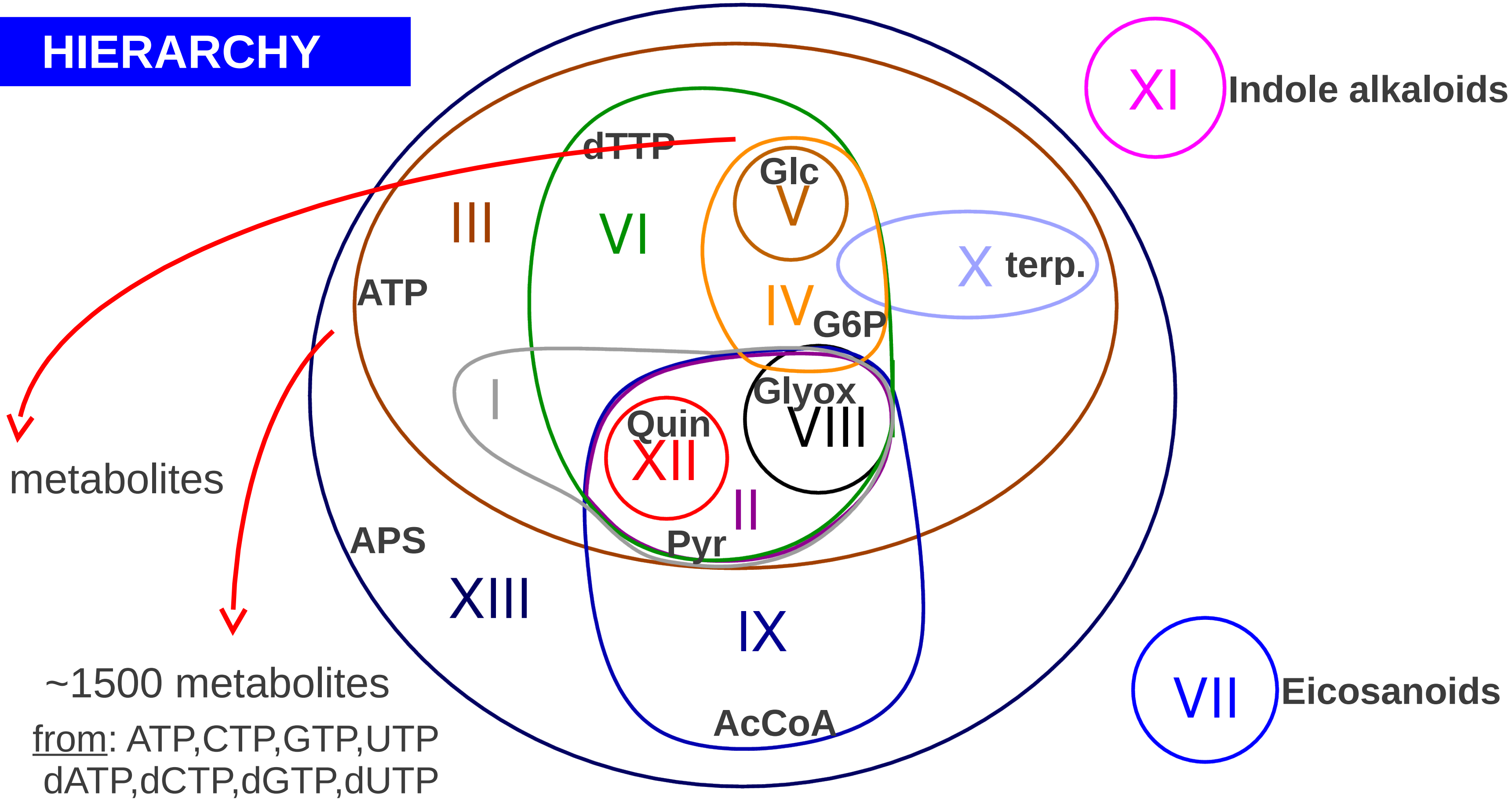
Similarity of biosynthetic potentials

Many metabolites carry ***similar*** biosynthetic potentials

⇒ ***“consensus scope“***
(typical biosynthetic potential)

Groups with similar potentials can be identified by ***clustering analysis***

HIERARCHY

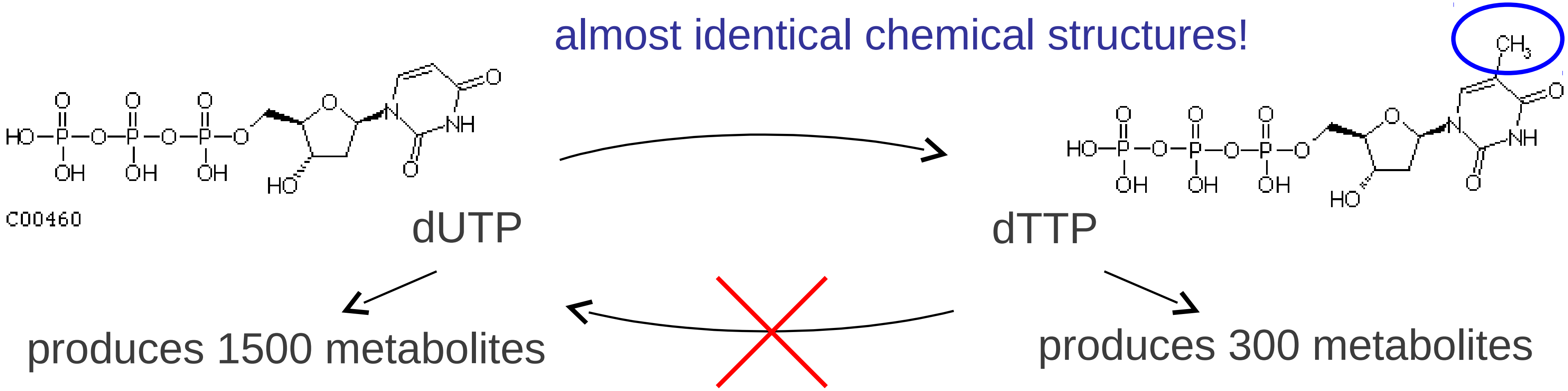


~300 metabolites

~1500 metabolites
from: ATP,CTP,GTP,UTP
dATP,dCTP,dGTP,dUTP

(Matthäus, Salazar and Ebenhöh, PLoS Comp Biol, 2008)

Separation of biosynthetic potentials



almost identical chemical structures!

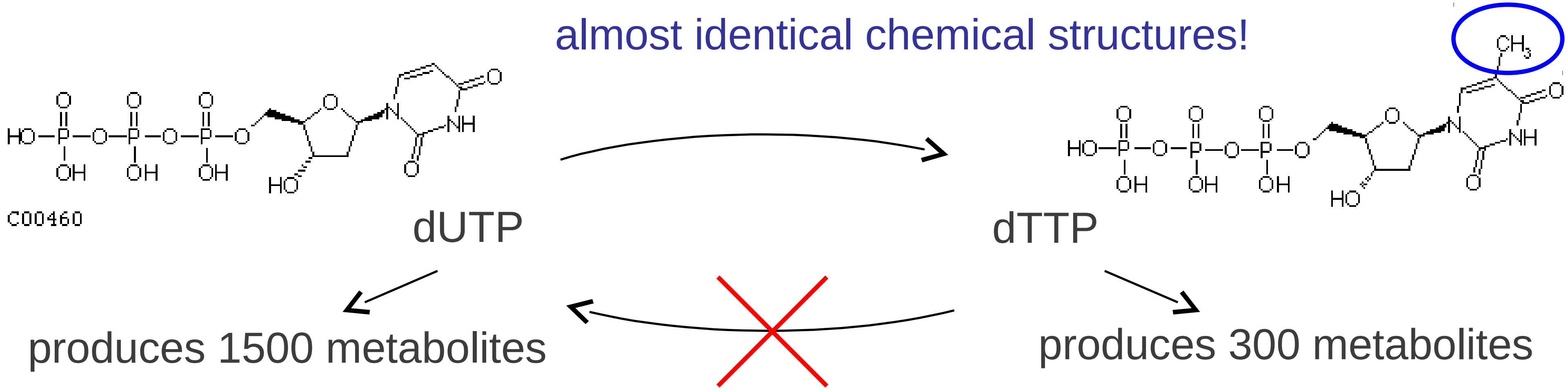
dUTP

dTTP

produces 1500 metabolites

produces 300 metabolites

Separation of biosynthetic potentials



almost identical chemical structures!

dUTP

dTTP

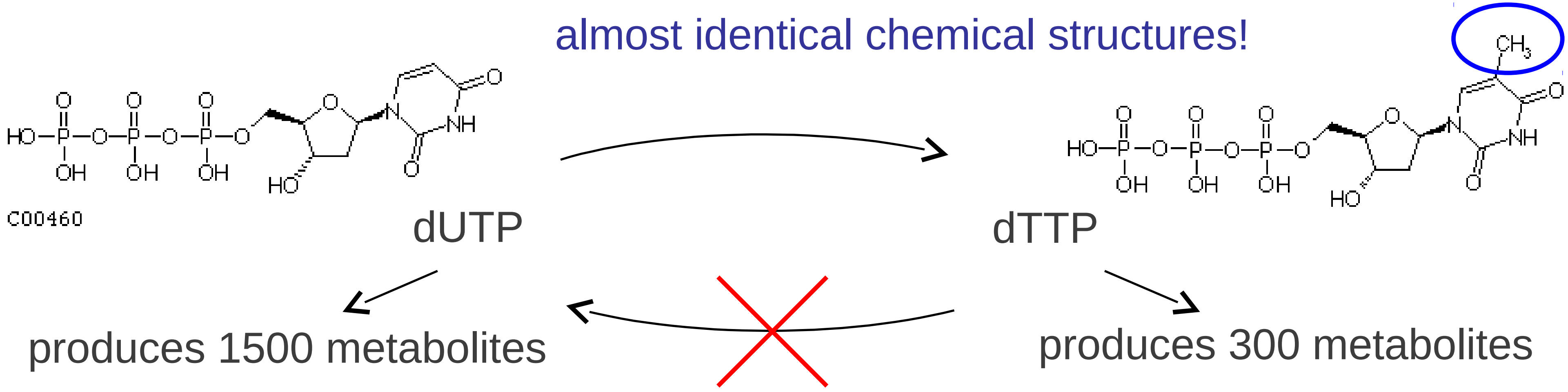
produces 1500 metabolites

produces 300 metabolites

in agreement with experiments (growth of *Physarum Polycephalum* with ^{14}C nucleosides)

Fink & Nygaard (1978), Eur. J. Biochem

Separation of biosynthetic potentials



almost identical chemical structures!

dUTP

dTTP

produces 1500 metabolites

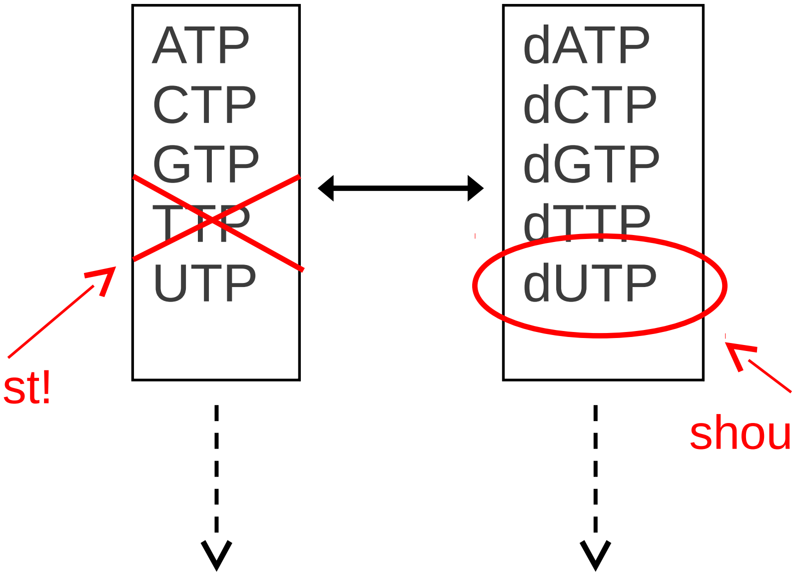
produces 300 metabolites

in agreement with experiments (growth of *Physarum Polycephalum* with ^{14}C nucleosides)

Fink & Nygaard (1978), Eur. J. Biochem

The chemical complexity alone does not determine the biosynthetic potential!

<u>Explanation</u>



ATP
CTP
GTP
~~TTP~~
UTP

dATP
dCTP
dGTP
dTTP
dUTP

does not exist!

should be low!

RNA

DNA

Organisms and their environment

No organism lives in complete isolation

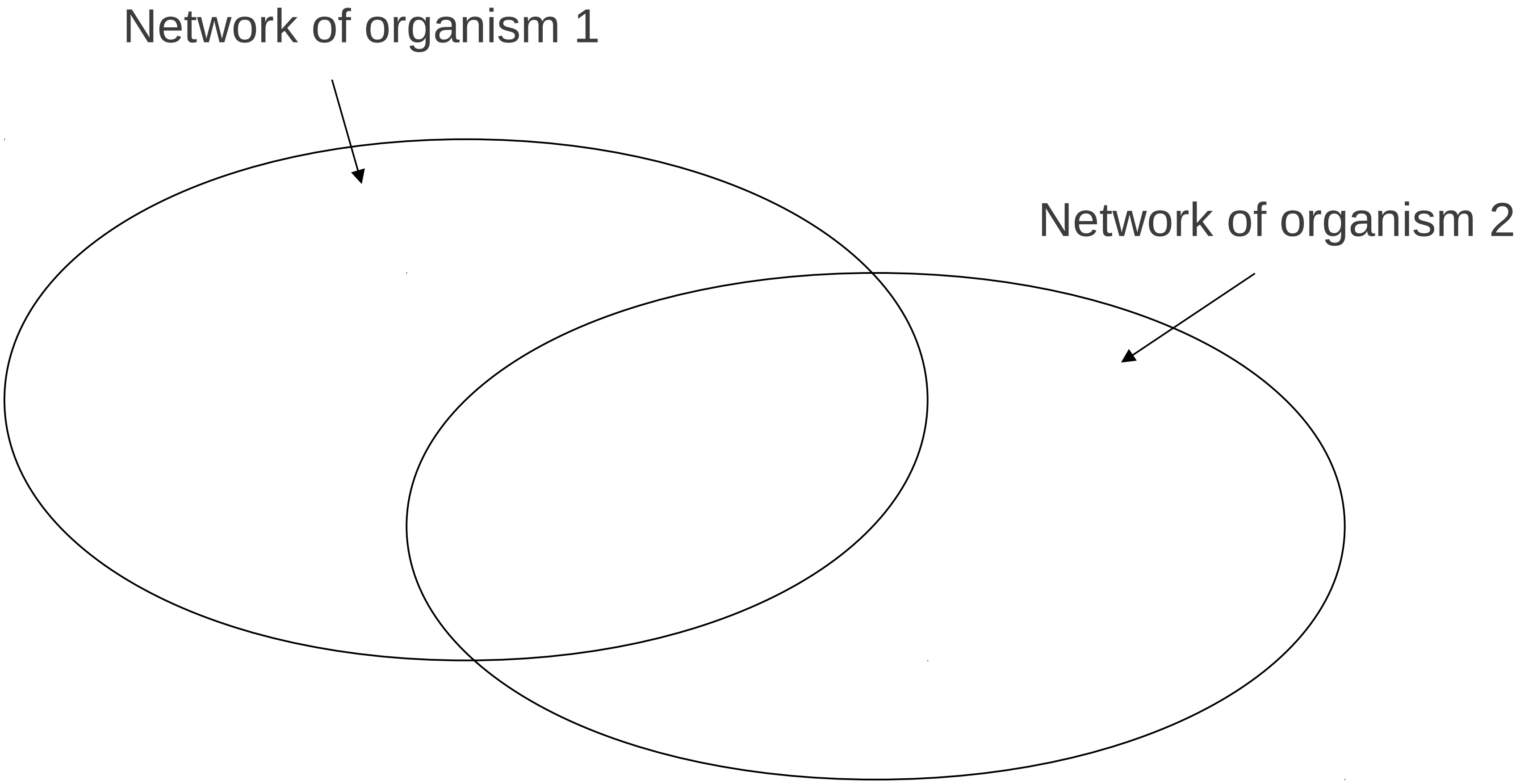
Organisms shape the environment (e.g. by excreted products)

Organisms are themselves part of the environment of others (ecosystem)

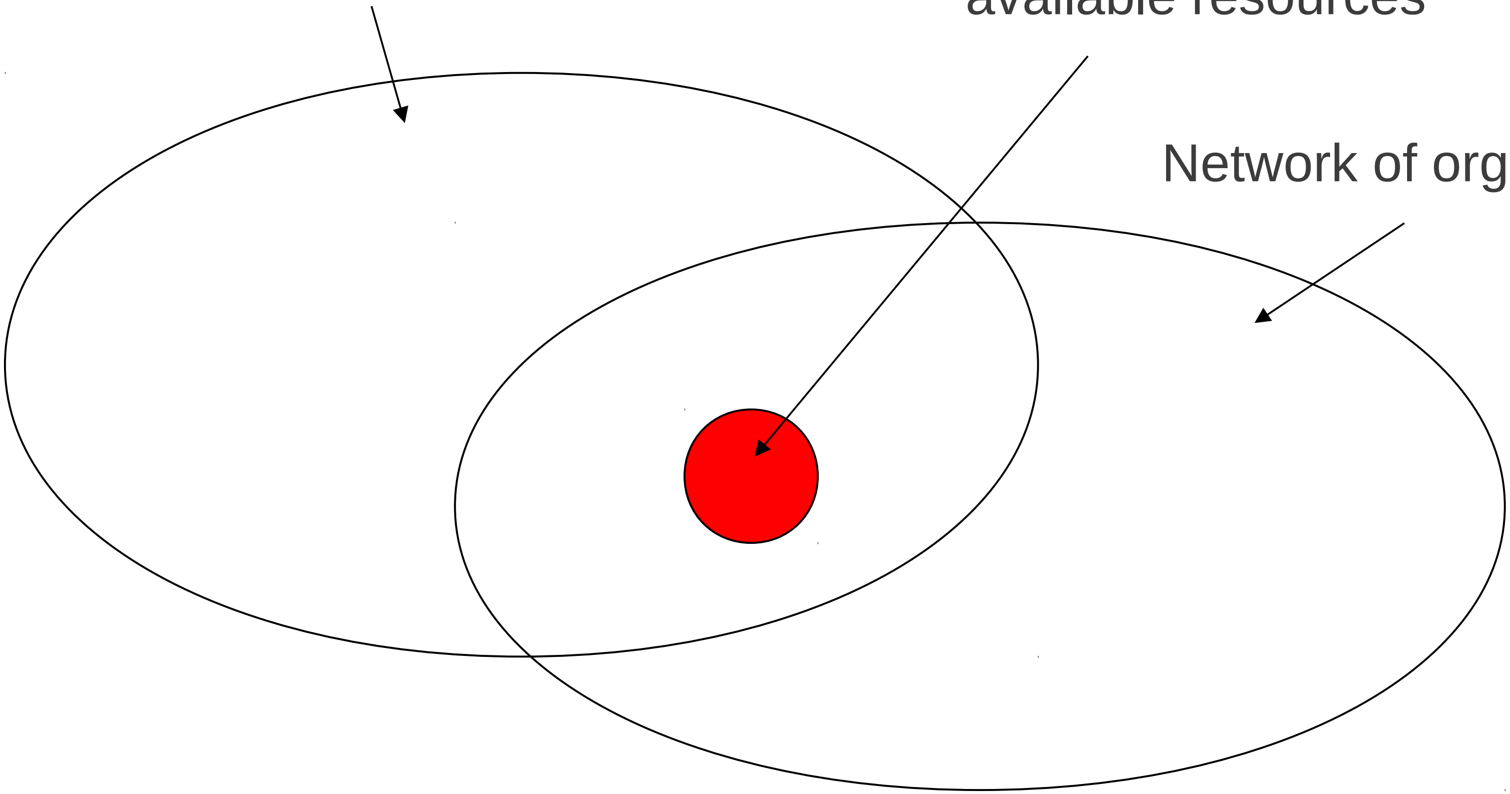
<u>Interaction on the level of metabolic networks</u>

- Biodegradation
  involves many microorganisms, requires the special metabolic capabilities
- Symbiosis
  e.g. plants (fabacaea) and Rhizobia (nitrogen fixing bacteria)
- Parasitism
  e.g. Wolbachia live inside insect cells

Metabolic synergy



Metabolic synergy

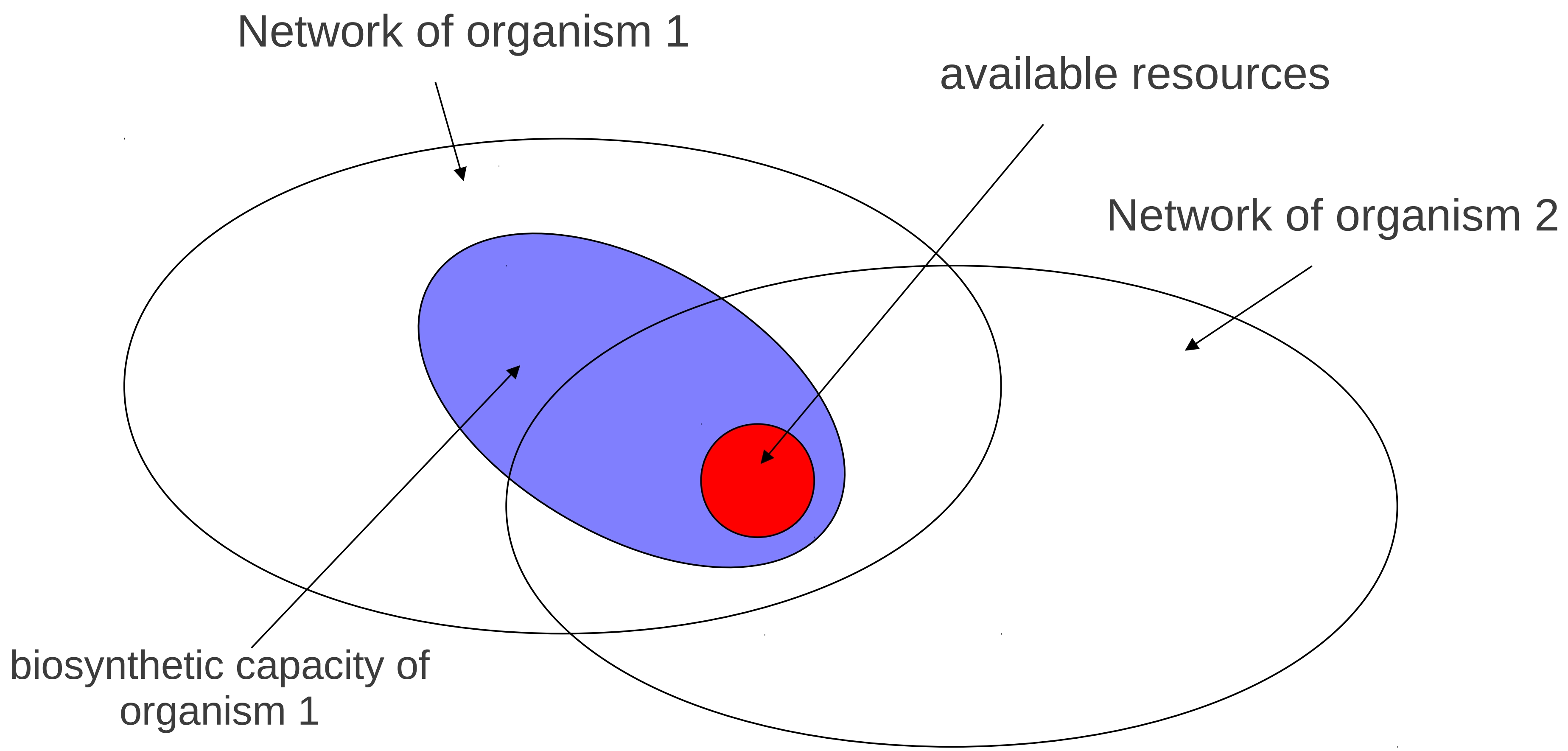


Network of organism 1

available resources

Network of organism 2

Metabolic synergy



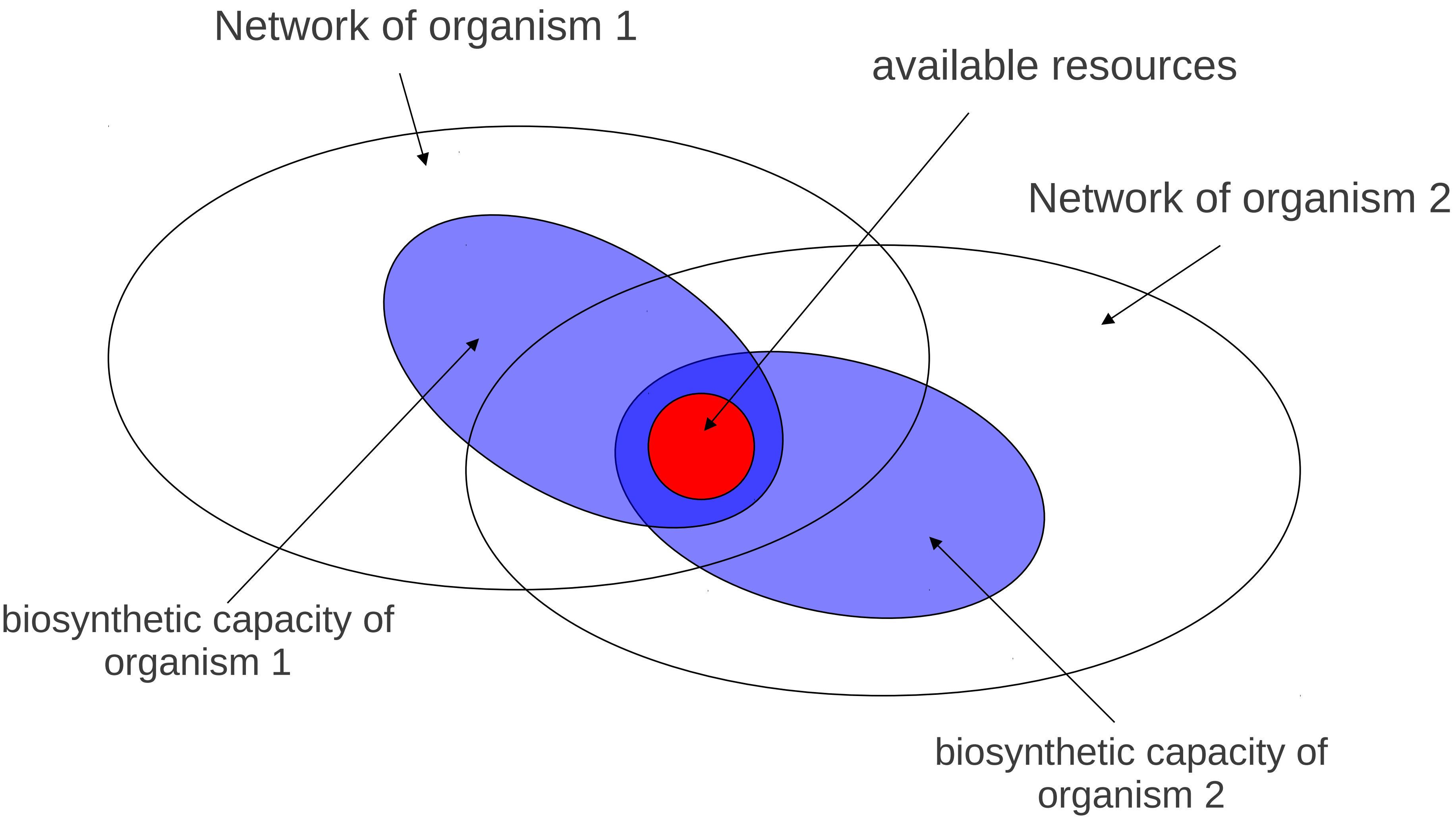
Network of organism 1

available resources

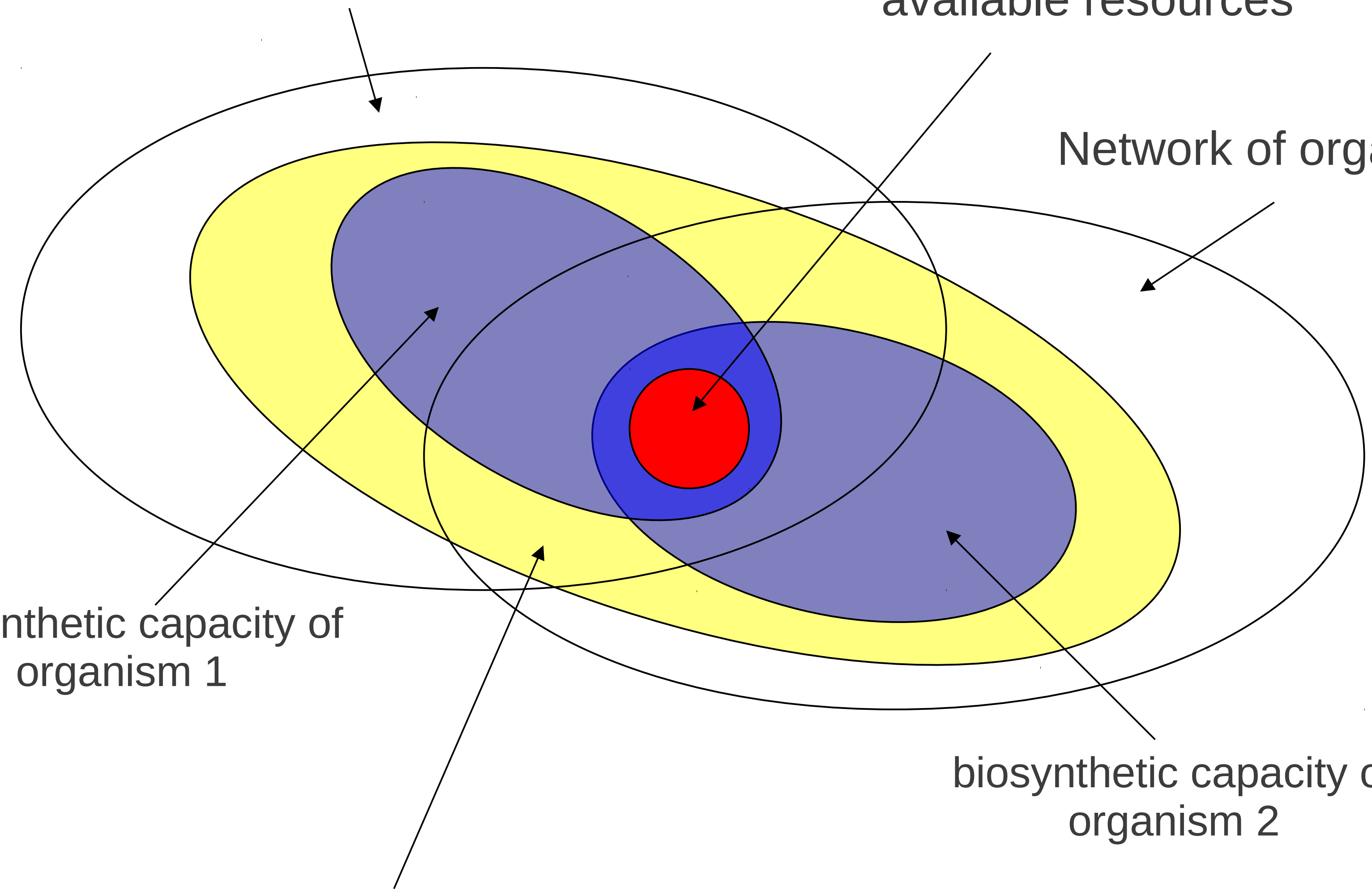
Network of organism 2

biosynthetic capacity of
organism 1

Metabolic synergy



Metabolic synergy



Network of organism 1

available resources

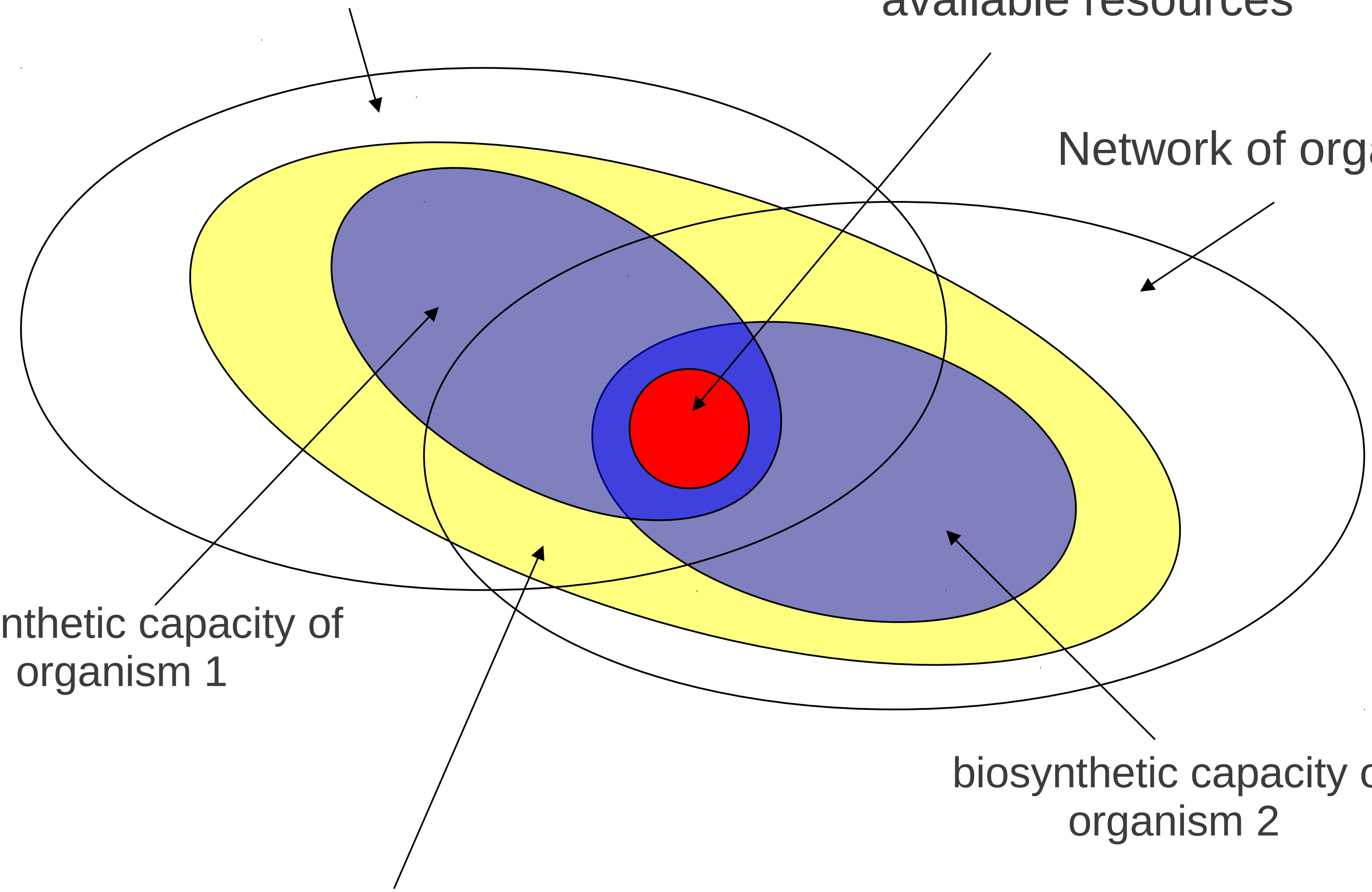
Network of organism 2

biosynthetic capacity of organism 1

biosynthetic capacity of organism 2

Biosynthetic capacity for the joined network is larger than the sum of the single capacities!

Metabolic synergy



Network of organism 1

available resources

Network of organism 2

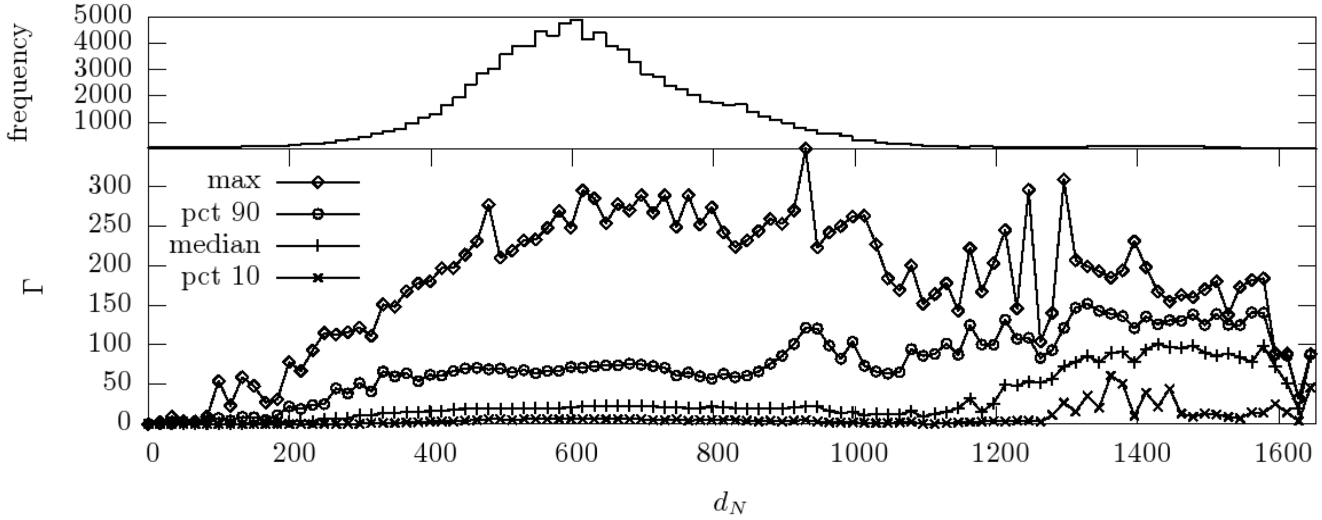
biosynthetic capacity of organism 1

biosynthetic capacity of organism 2

Biosynthetic capacity for the joined network is larger than the sum of the single capacities! → Synergy

Synergy vs. network dissimilarity

Statistics...
Which pairs are best suited to yield synergetic effects?



(Christian, Handorf and Ebenhöh, 2007)

A simplistic view with lots of space for improvement:

- Transport processes
- Quantification of negative effects (FBA)
- ...

⇒ Investigate specific examples

Acknowledgements

Financial support

Bundesministerium
für Bildung
und Forschung

SULSA
Scottish Universities Life Sciences Alliance

SEVENTH FRAMEWORK
PROGRAMME

BBSRC
bioscience for the future

NATURAL
ENVIRONMENT
RESEARCH COUNCIL

EXCELLENCE
IN SCIENCE

Cooperation and discussion partners

Aberdeen	Nils Christian Thomas Pfau

cooperation partners:
Patrick May (U Luxemburg)
Stefan Kempa (MDC Buch)
Stefanie Wienkoop (U Wien)
Daniel Kahn (U Lyon)
Franziska Matthäus (U Heidelberg)
Carlos Salazar (BioQuant Heidelberg)
Thomas Handorf (Humboldt U Berlin)
Kai Kruse (U Potsdam)