In silico Evolution of Early Metabolism

Alexander Ullrich

Bioinformatics
University of Leipzig

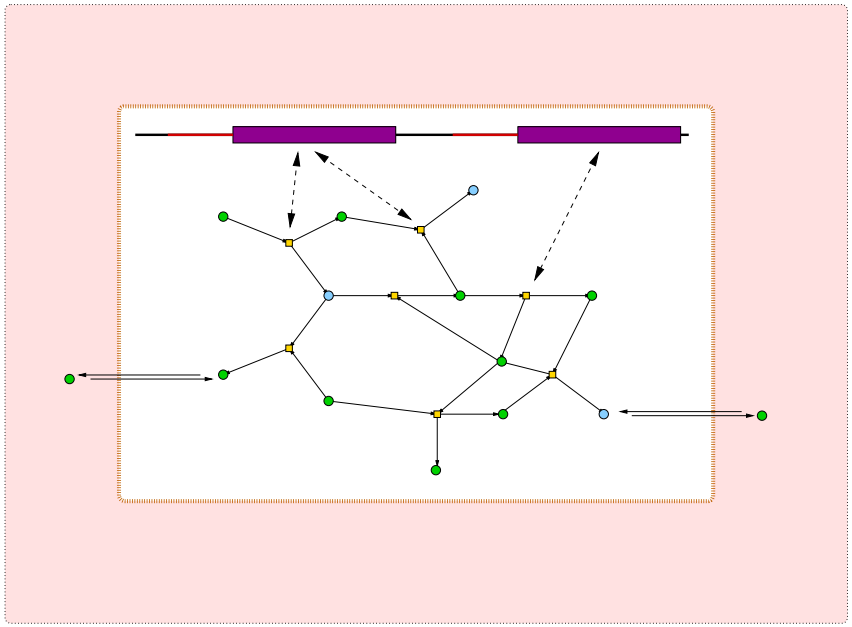
Jena, November 25

Motivation

- Understand evolutionary mechanisms of biological systems
- Study the early development of metabolism
  - not observable by conventional approaches
- Analyse different hypotheses for pathway evolution
  - finding scenarios for observations in present data
- Investigate the emergence of systemic properties
- Answers beyond analyzing real-world data

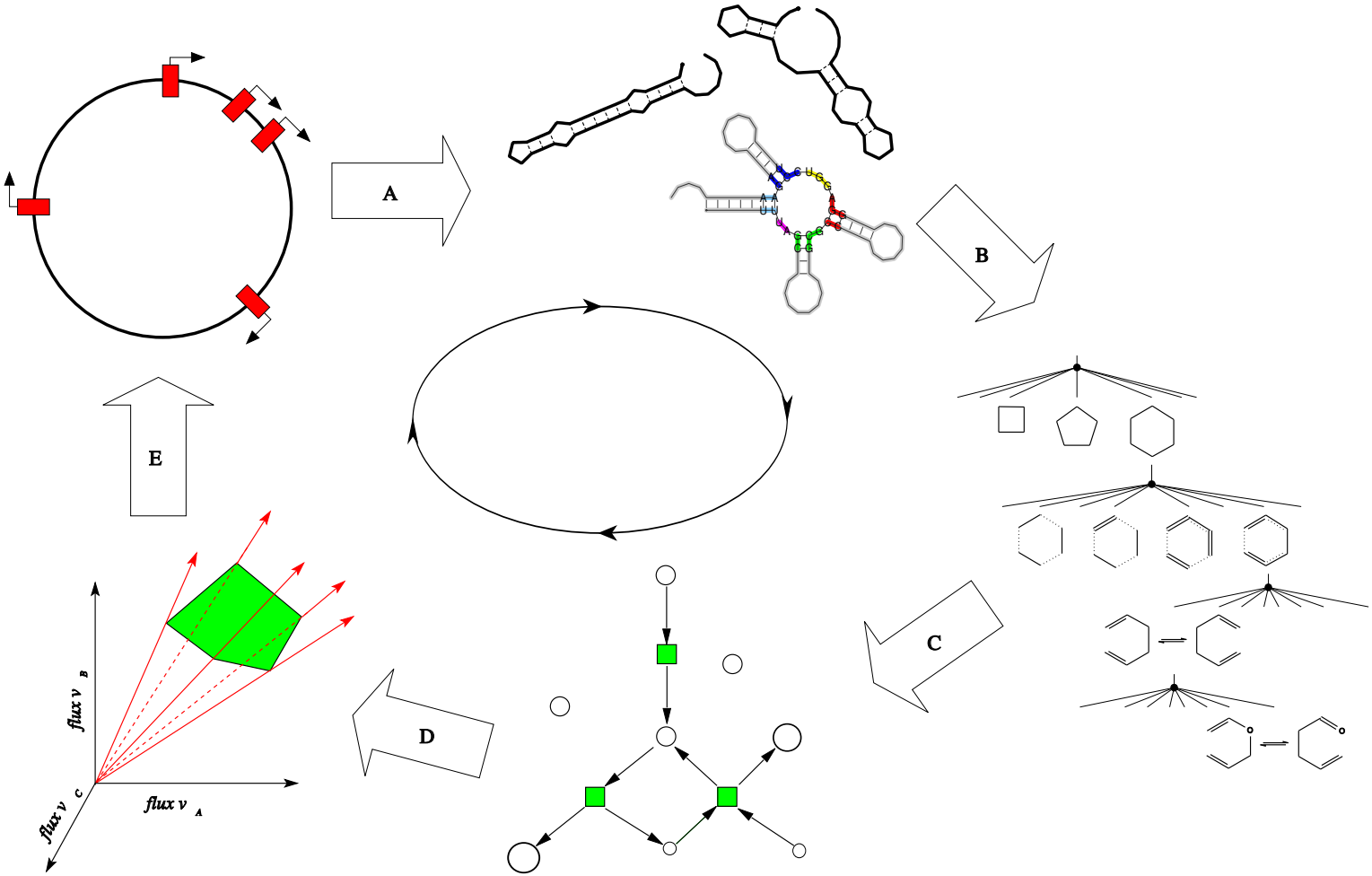
$\longrightarrow$ a multi-scale computational model of early metabolism

Simulation



- Protocellular entity
- Bag of ribozymes
- Algebraic chemistry model
- Exchange of molecules with the environment

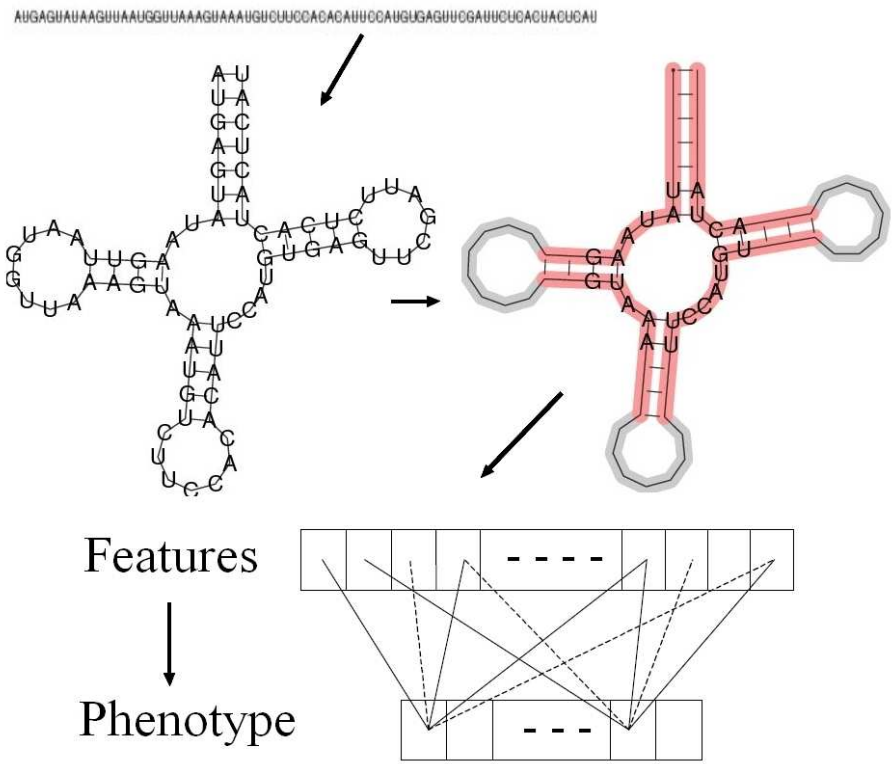
Simulation - Overview



Mapping from Gene to Enzyme - How?

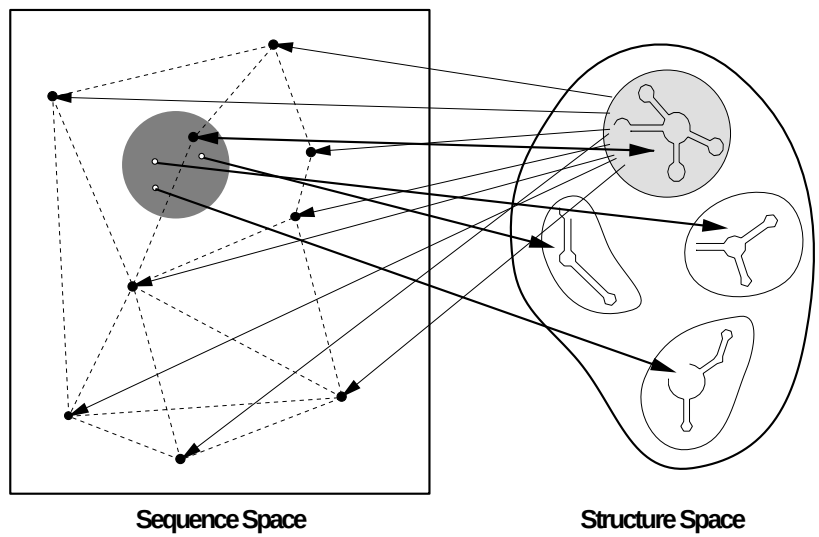
RNA-Sequence

- RNA-Structure
- reduced Structure
- Features
- ITS ID
- ITS
- Graph-rewriting Rule



RNA sequence-to-structure map

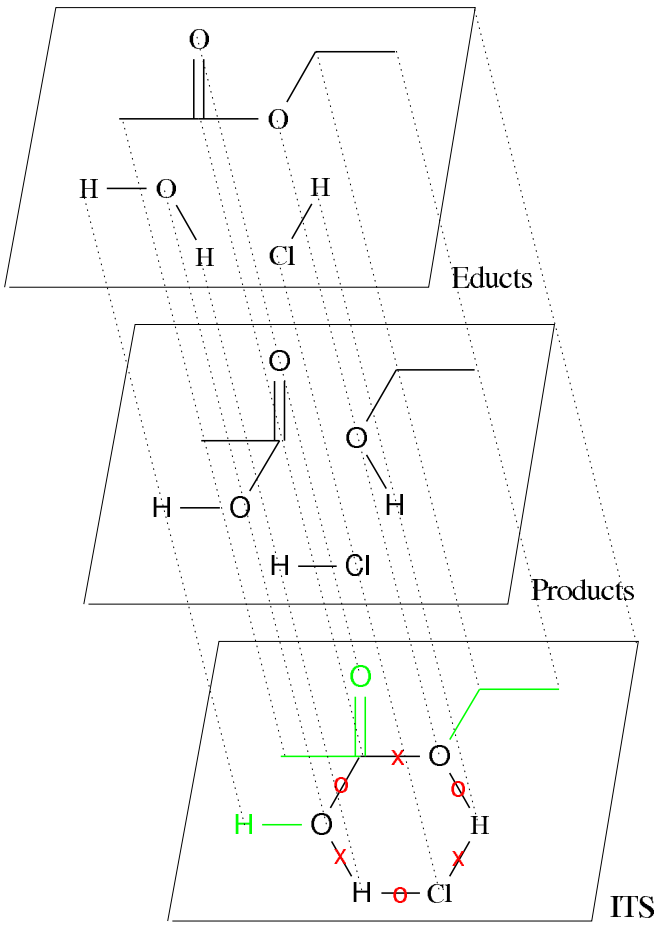
- Redundancy: Many more sequences than structures.
- Sensitivity: Small changes in the sequences may lead to large changes in the structure.
- Neutrality: A substantial fraction of mutations does not alter the structure.

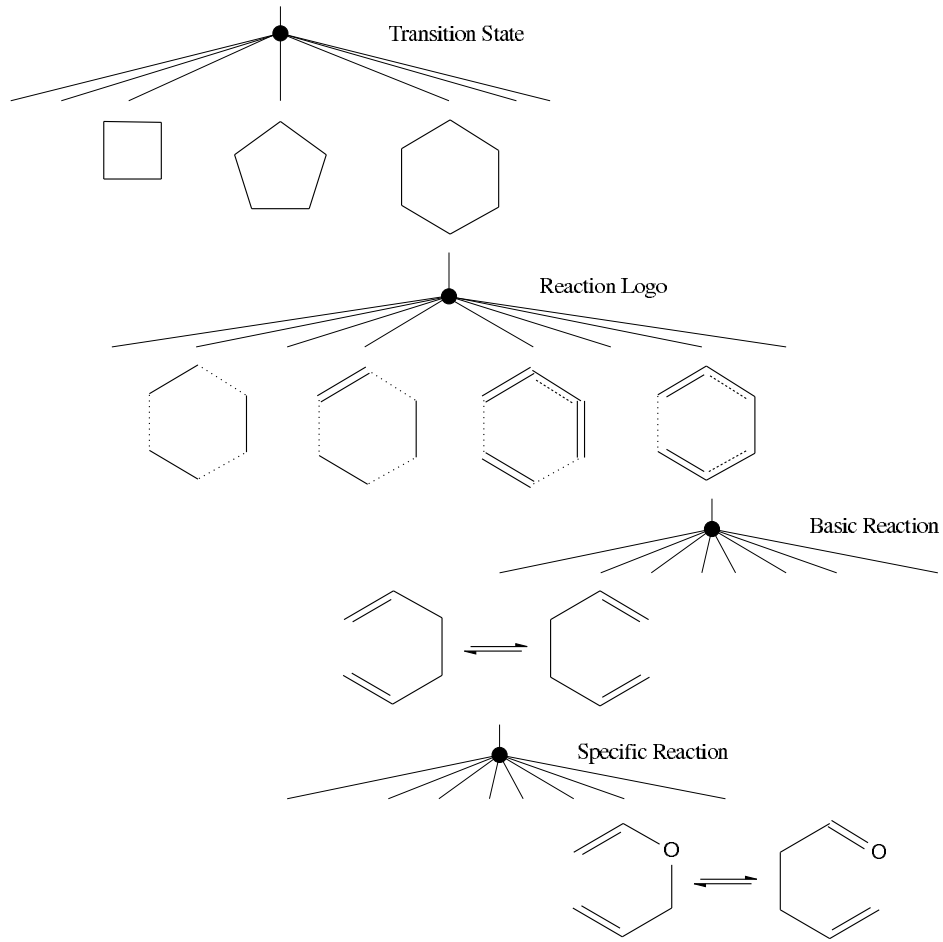


Walter Fontana & Peter Schuster, J. Theor. Biol. 194:491-515 (1998)

Reaction Classification

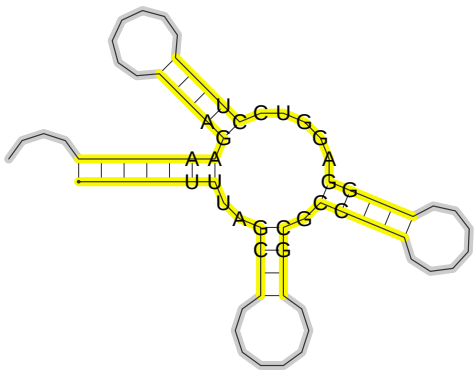
$$CH_3CO_2Et + HCl + H_2O \rightleftharpoons CH_3CO_2H + EtOH + HCl$$

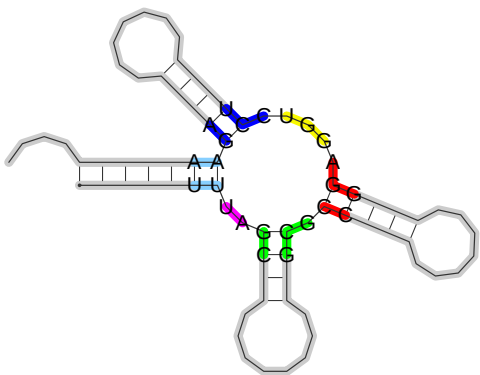


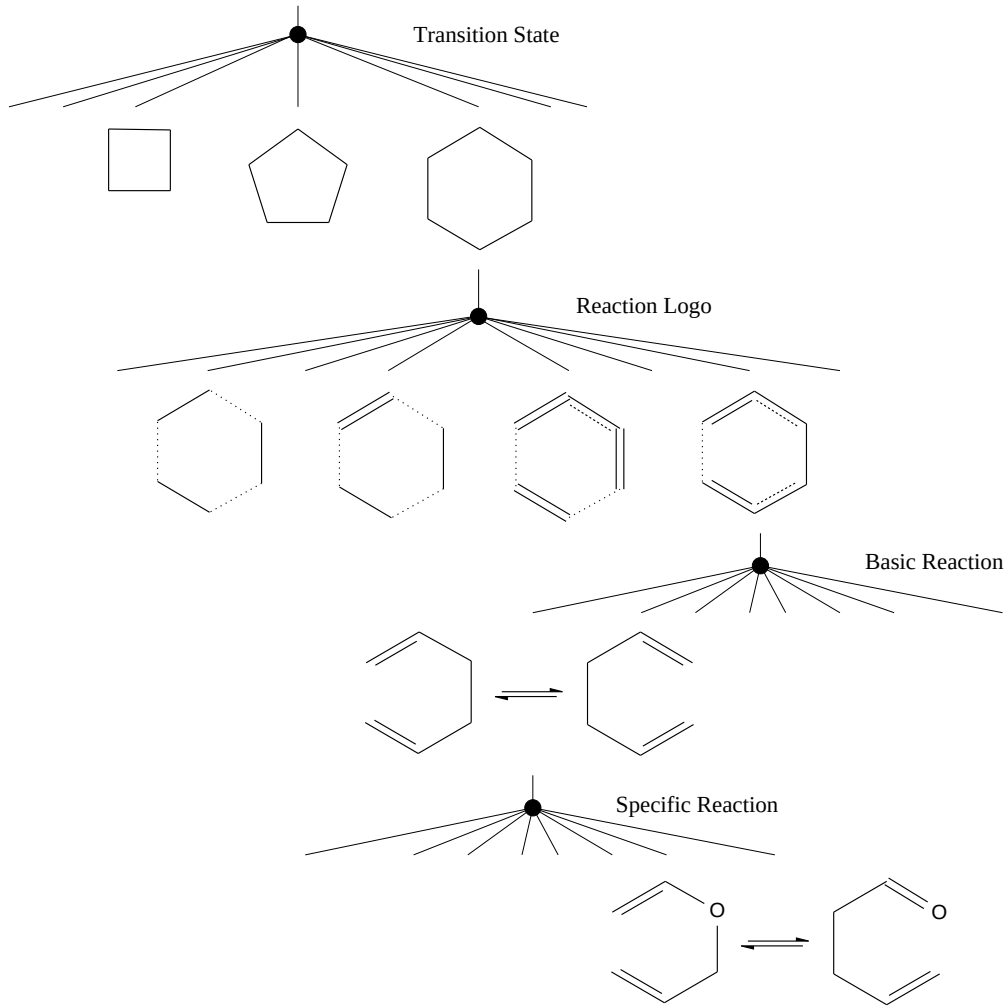


Fujita, Hendrickson, ...

From structure to function

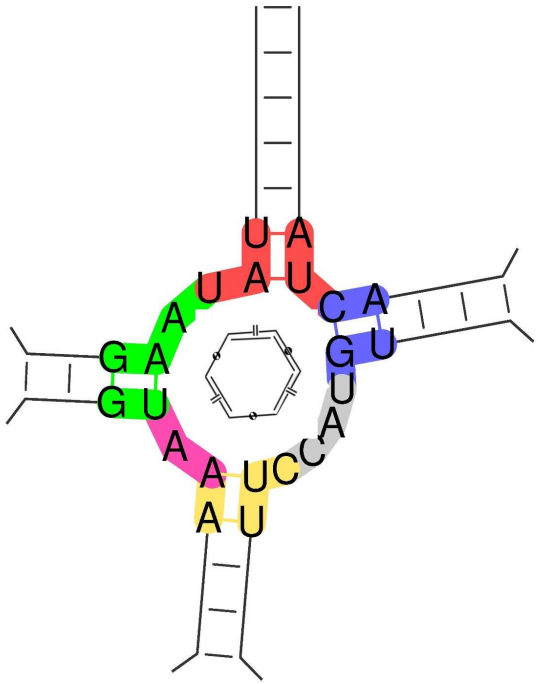




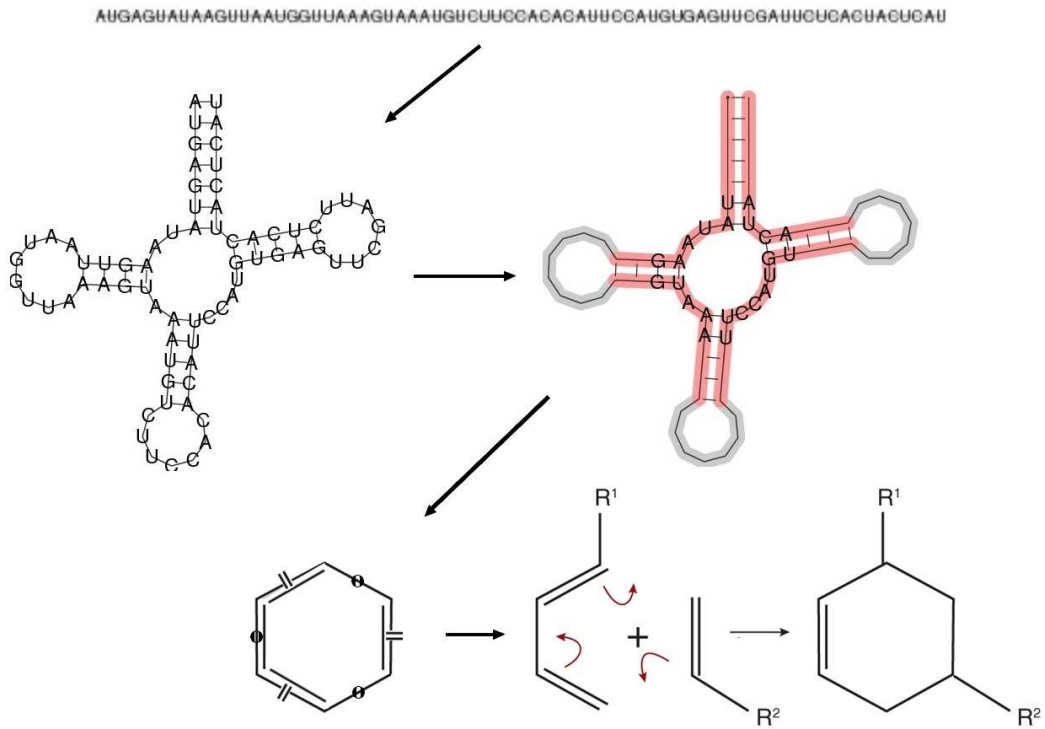


Neutrality is higher than in the RNA sequence-to-structure map.

Mapping from Gene to Enzyme - Example

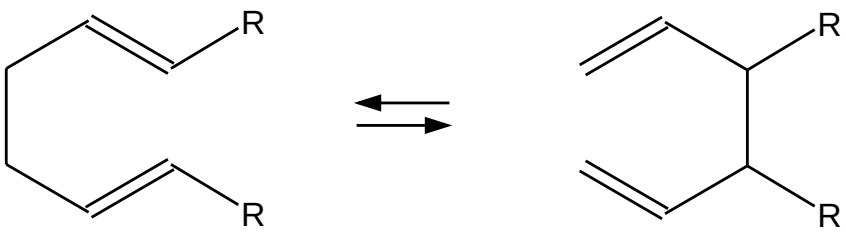


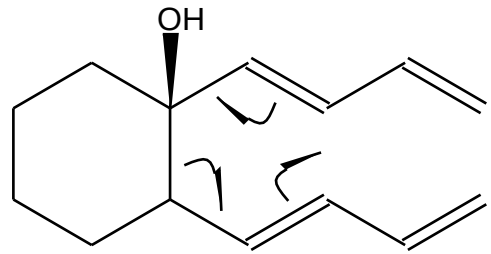
Section	Loop	C-G pair	Neighbor > 5 bp	**Bond**	Valence	Seq. (loop)	**Sequence**
1 (red)	yes	0	yes (+1)	**1** ”–”	2	4	**4 = C**
2 (blue)	yes	1	no	**1** ”–”	3	1	**3 = N**
3 (gray)	no	-	yes (+1)	**1** ”–”	3	4	**4 = C**
4 (yellow)	yes	0	no	**0** ” ”	2	4	**4 = C**
5 (pink)	no	-	yes (+1)	**1** ”–”	2	2	**2 = O**
6 (green)	yes	0	no	**0** ” ”	2	3	**3 = N**



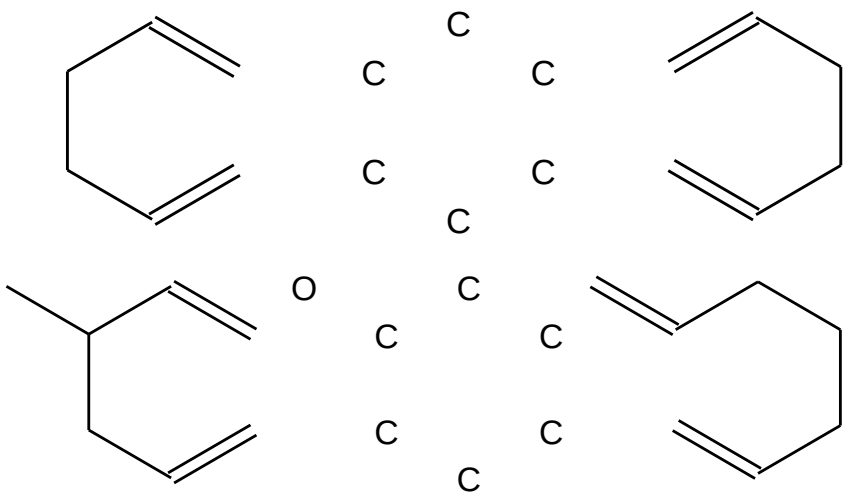
Chemical Reaction as Graph Rewrite Rule

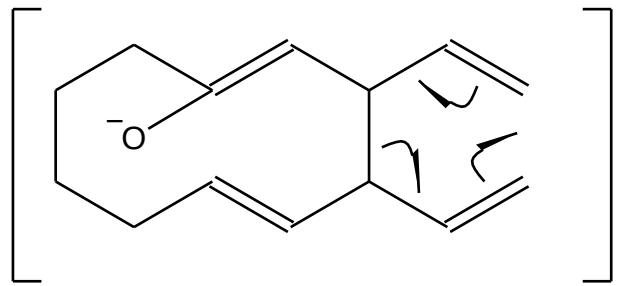
Cope Rearrangement

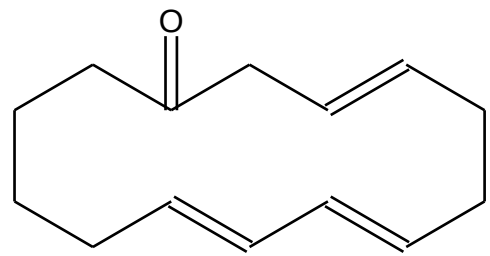




Rewrite Rules

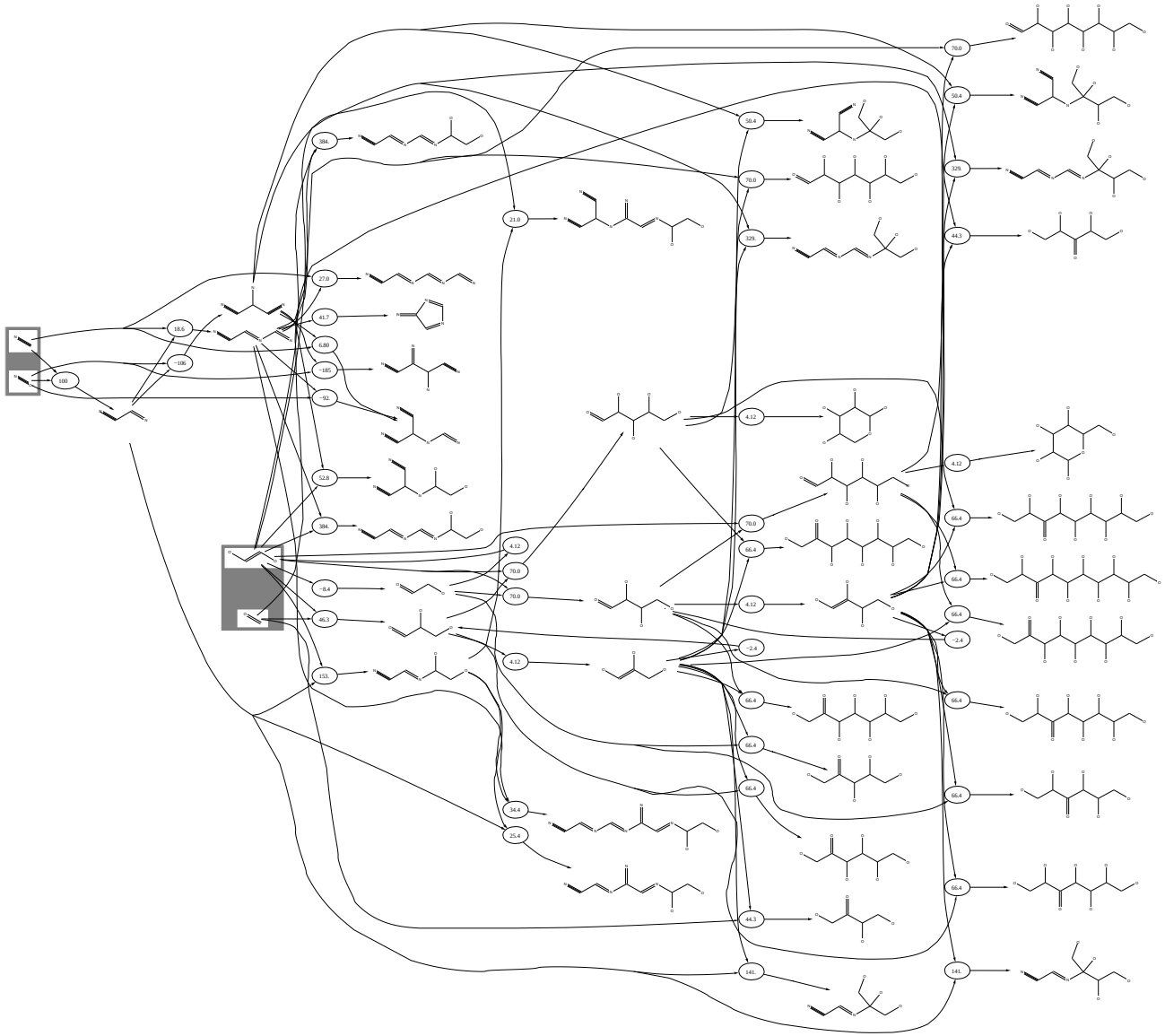






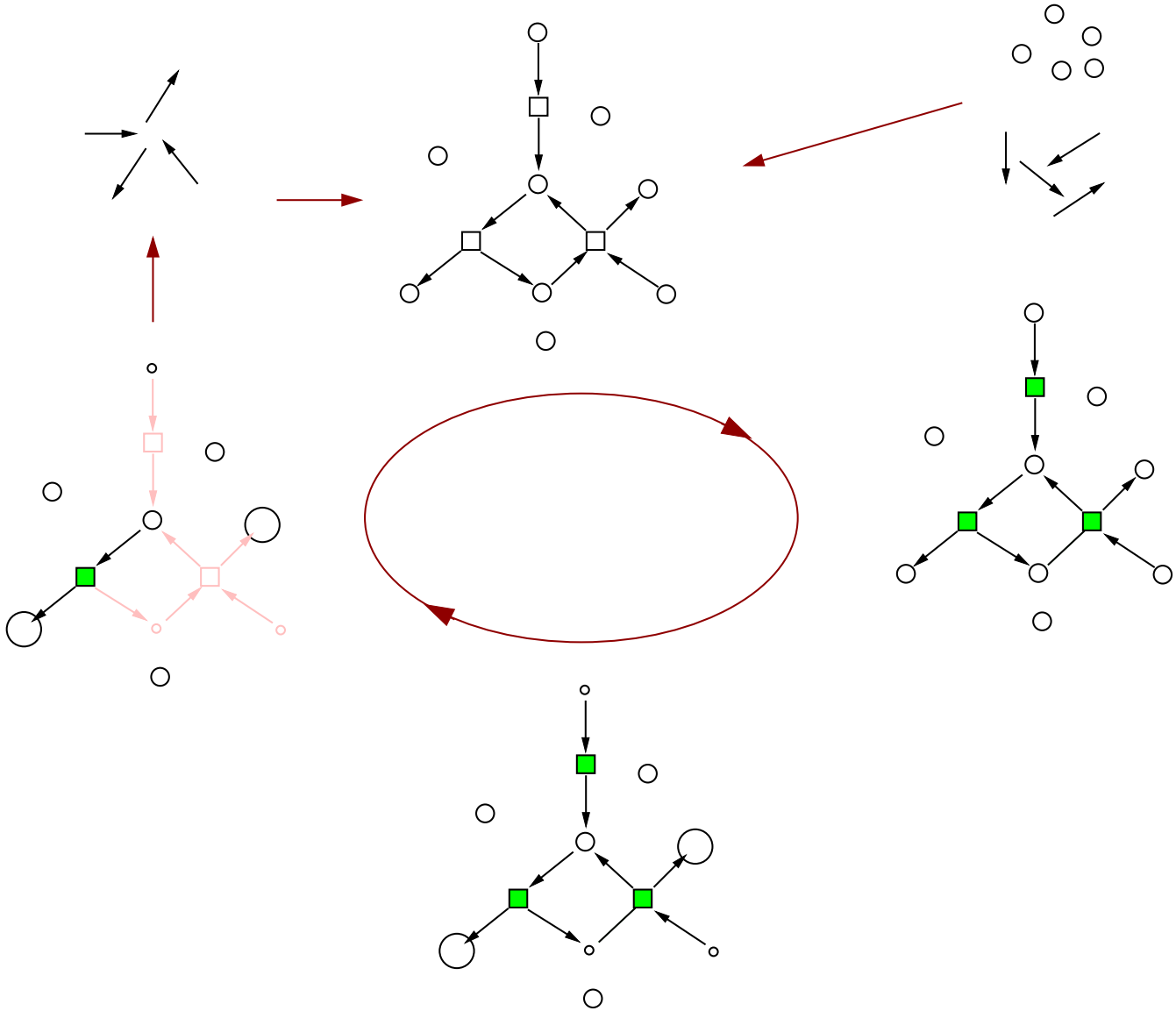
Graph-grammars are a context sensitive language!

Simulation - Growth



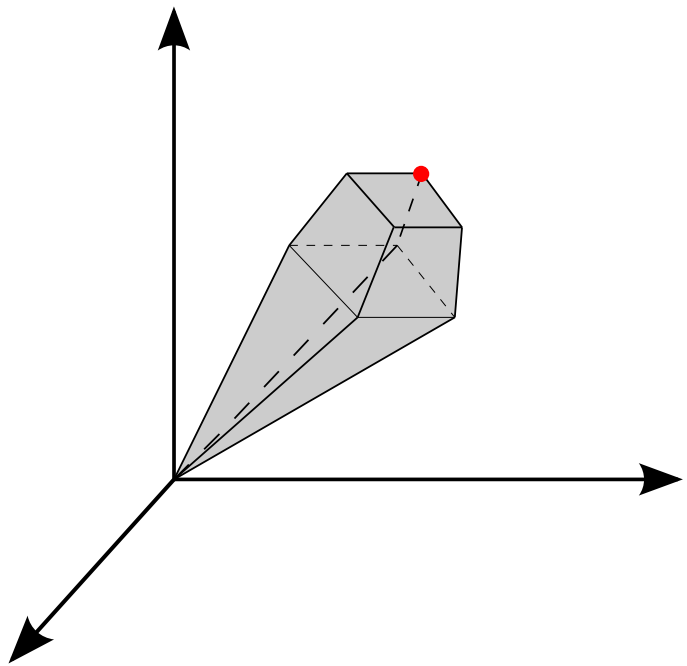
cyanide, formaldehyde glycol; aldolcondensation, tautomerization

Simulation - Stochastic Network Generator



Faulon, J-L, (2001) J Chem Inf Comput Sci 41:894-908

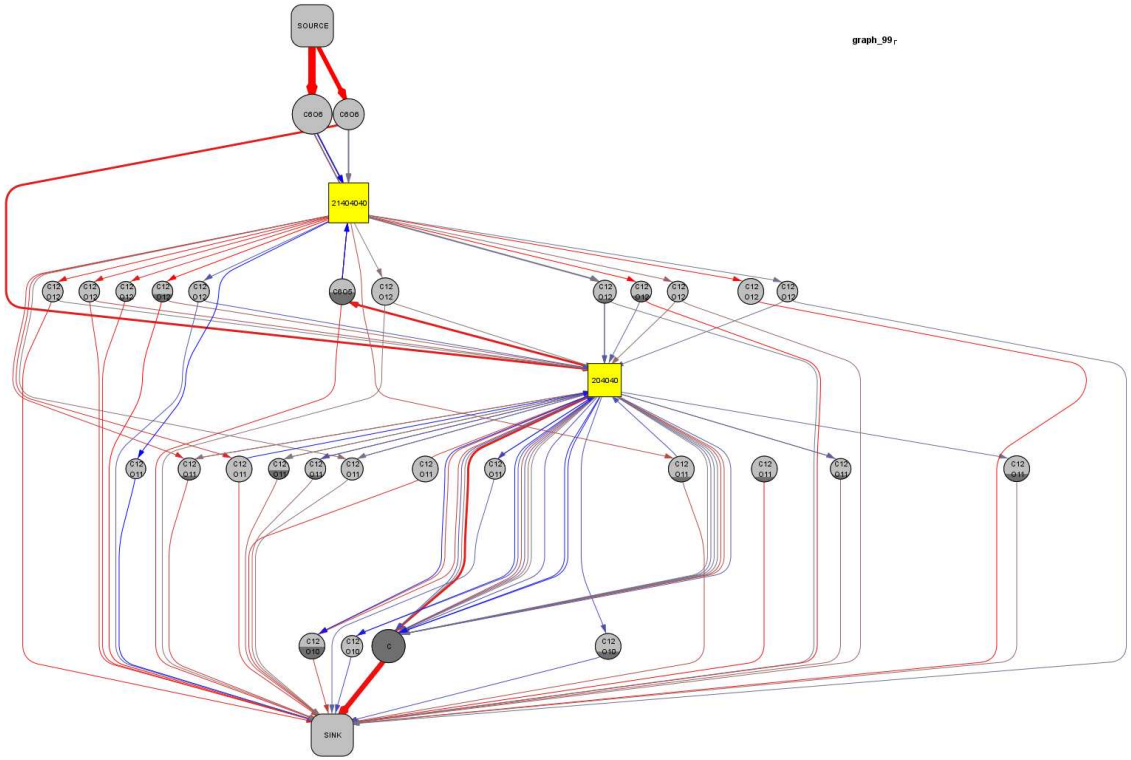
Simulation - Fitness



- Selection based on produced biomass
- maximizing biomass formation by linear optimization

Analysis

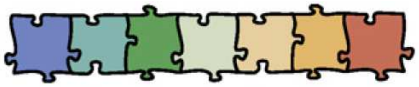
Visualization



- bidirectional, bipartite graph
- nodes: metabolites, enzymes/reactions
- edges: participation in the same reaction
- dot layout: flow of mass downwards in the graph (if possible)

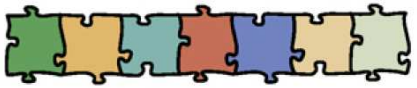
Visualization

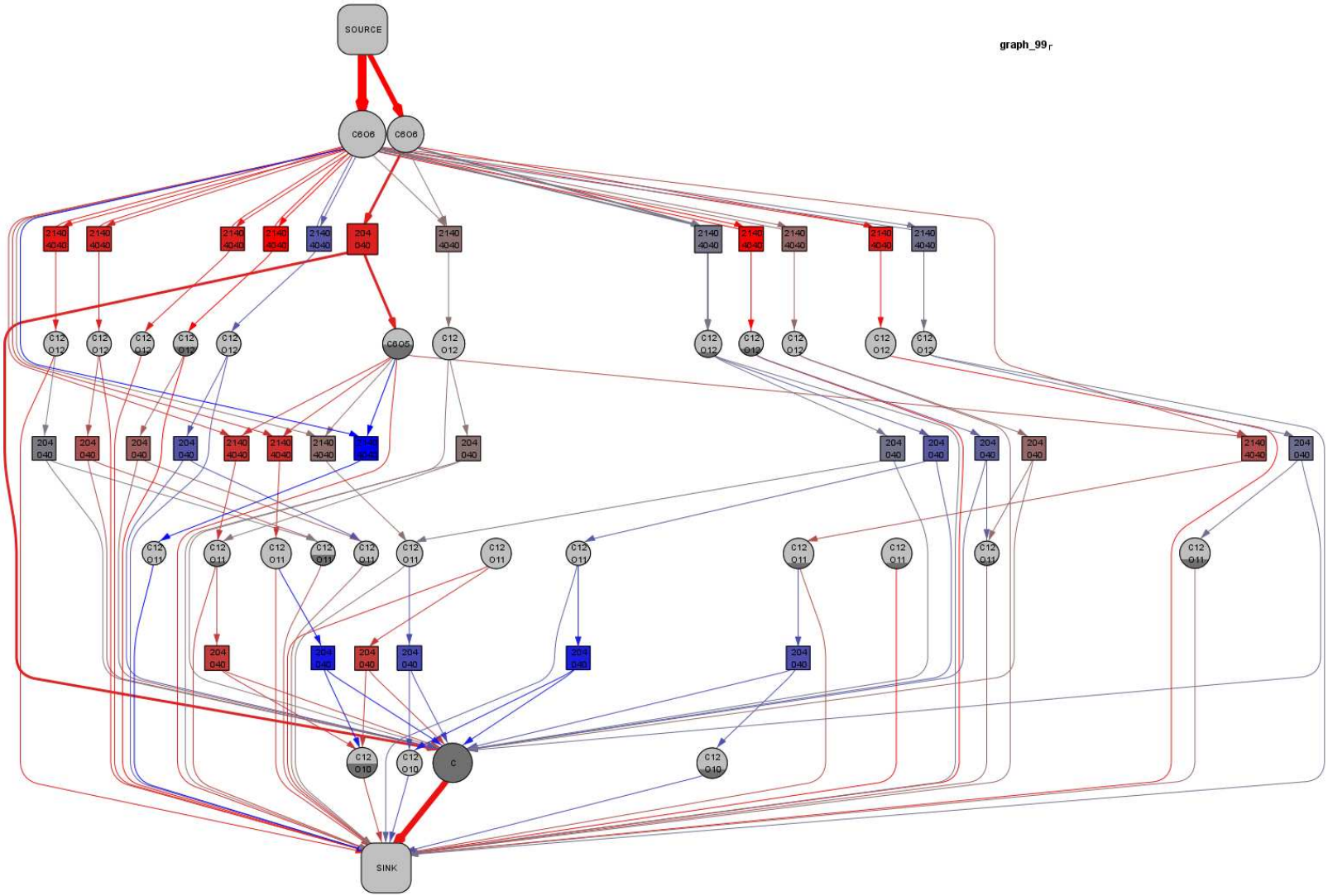
Retrograde Evolution



Forward Evolution

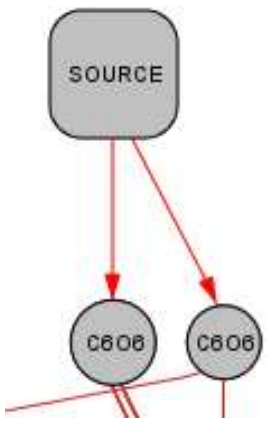
Patchwork Evolution



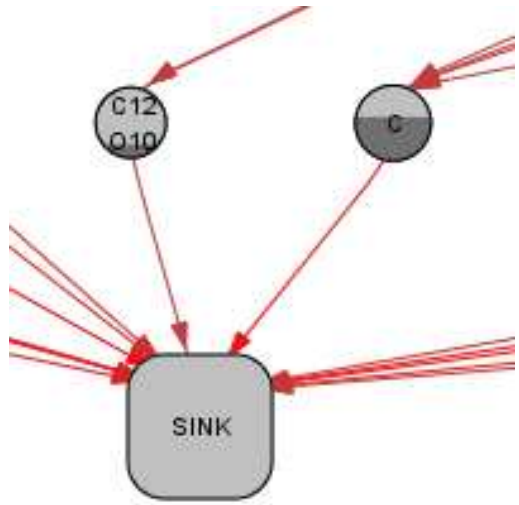


Visualization

Flow

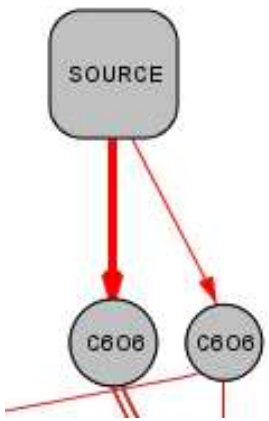


Concentration

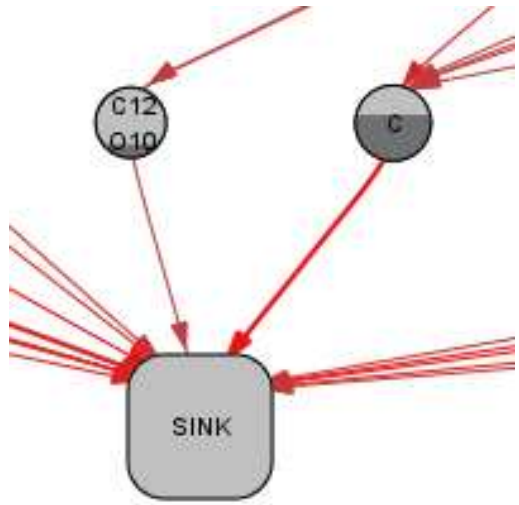


Visualization

Flow

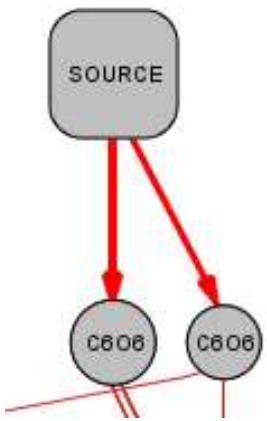


Concentration

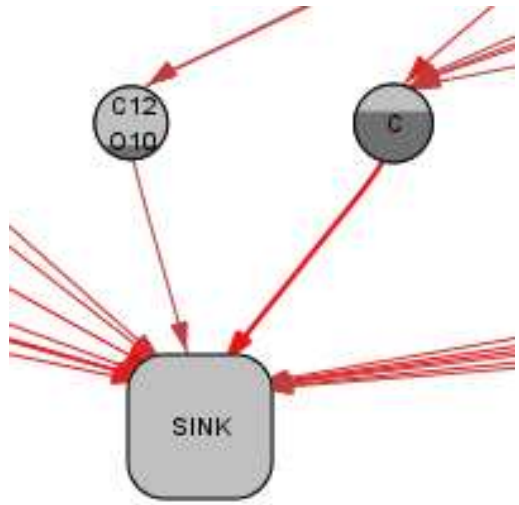


Visualization

Flow

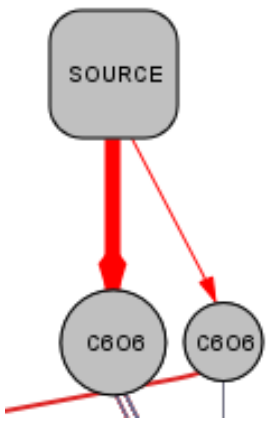


Concentration

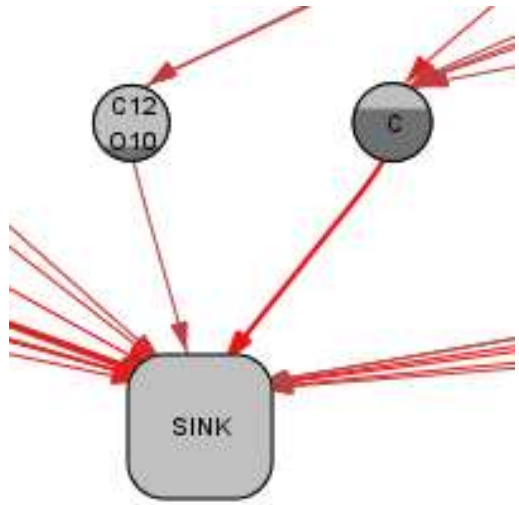


Visualization

Flow

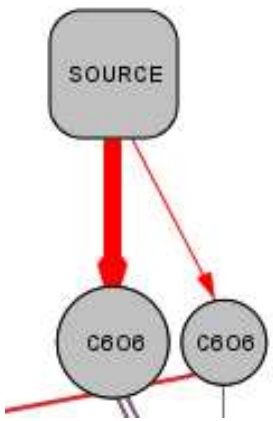


Concentration

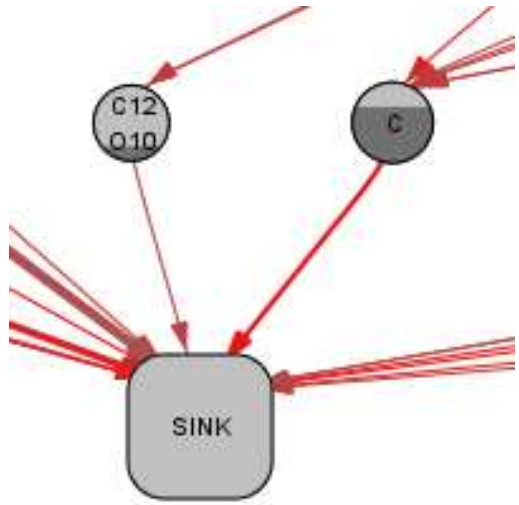


Visualization

Flow

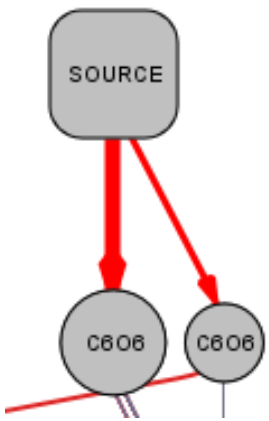


Concentration

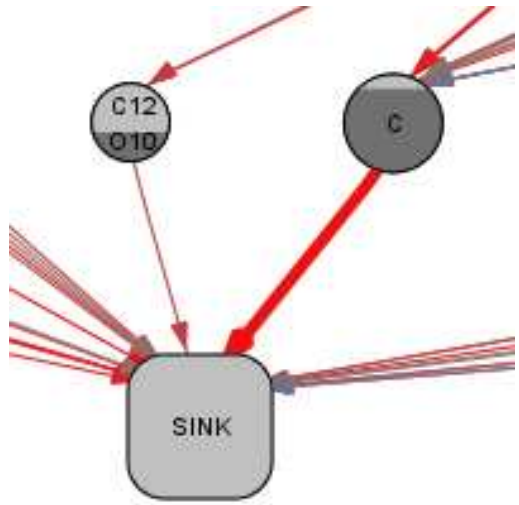


Visualization

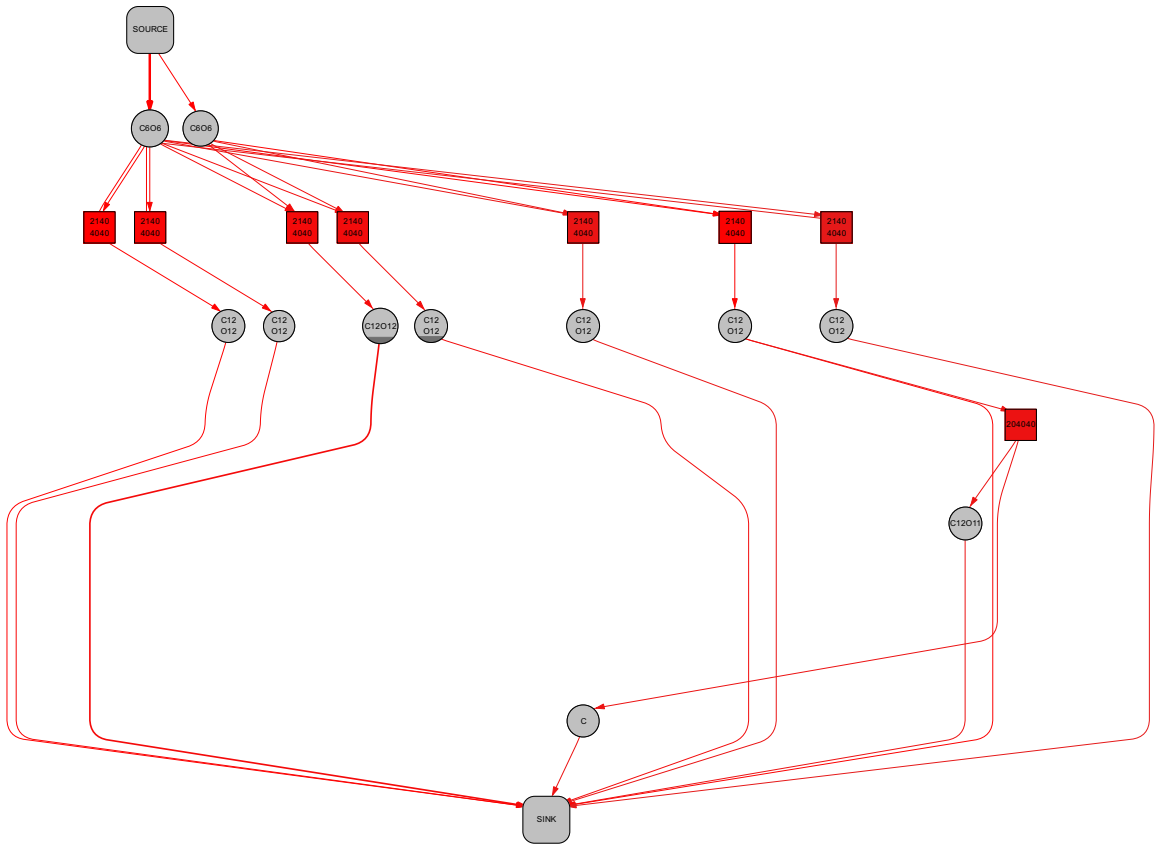
Flow



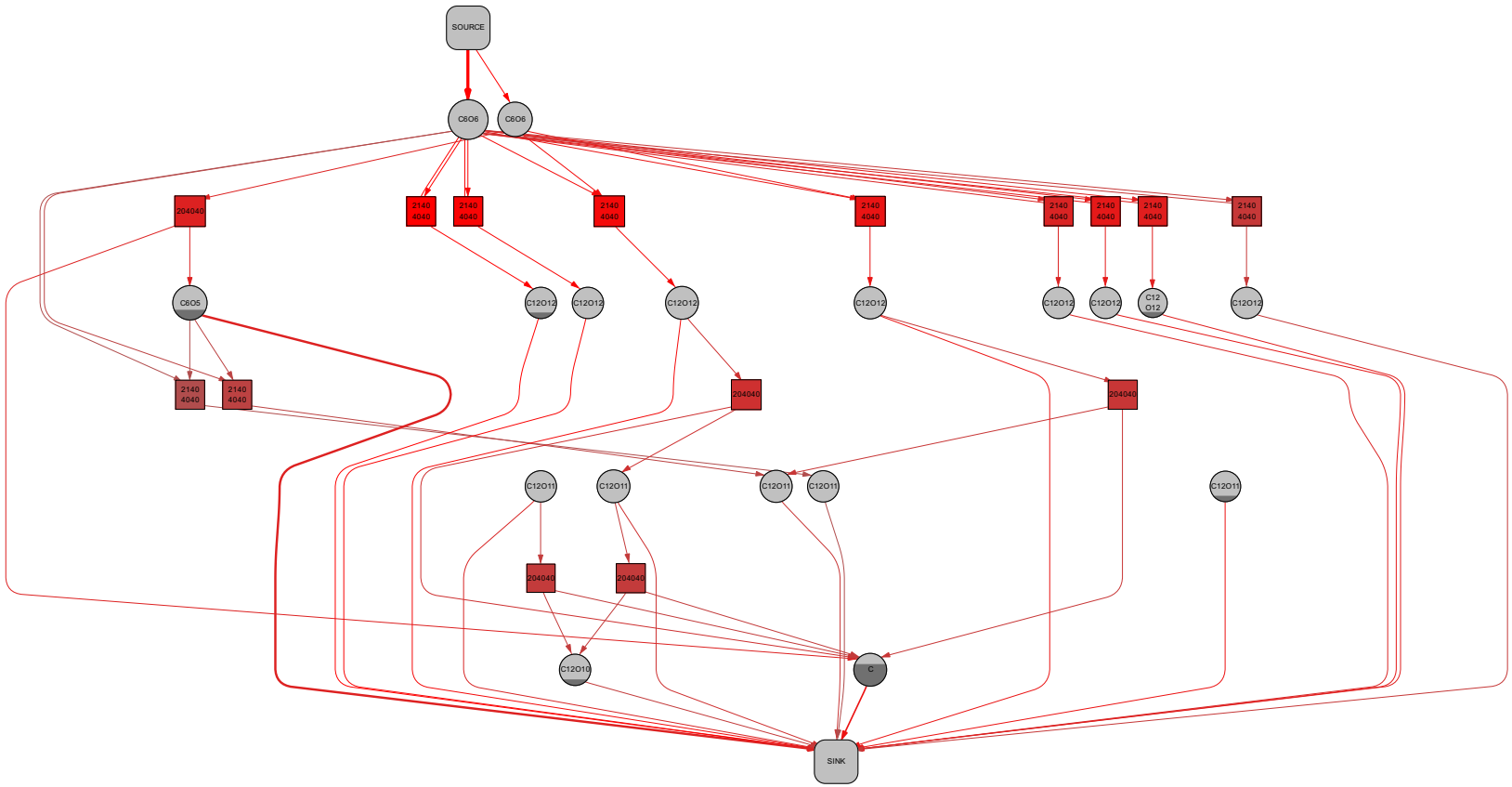
Concentration



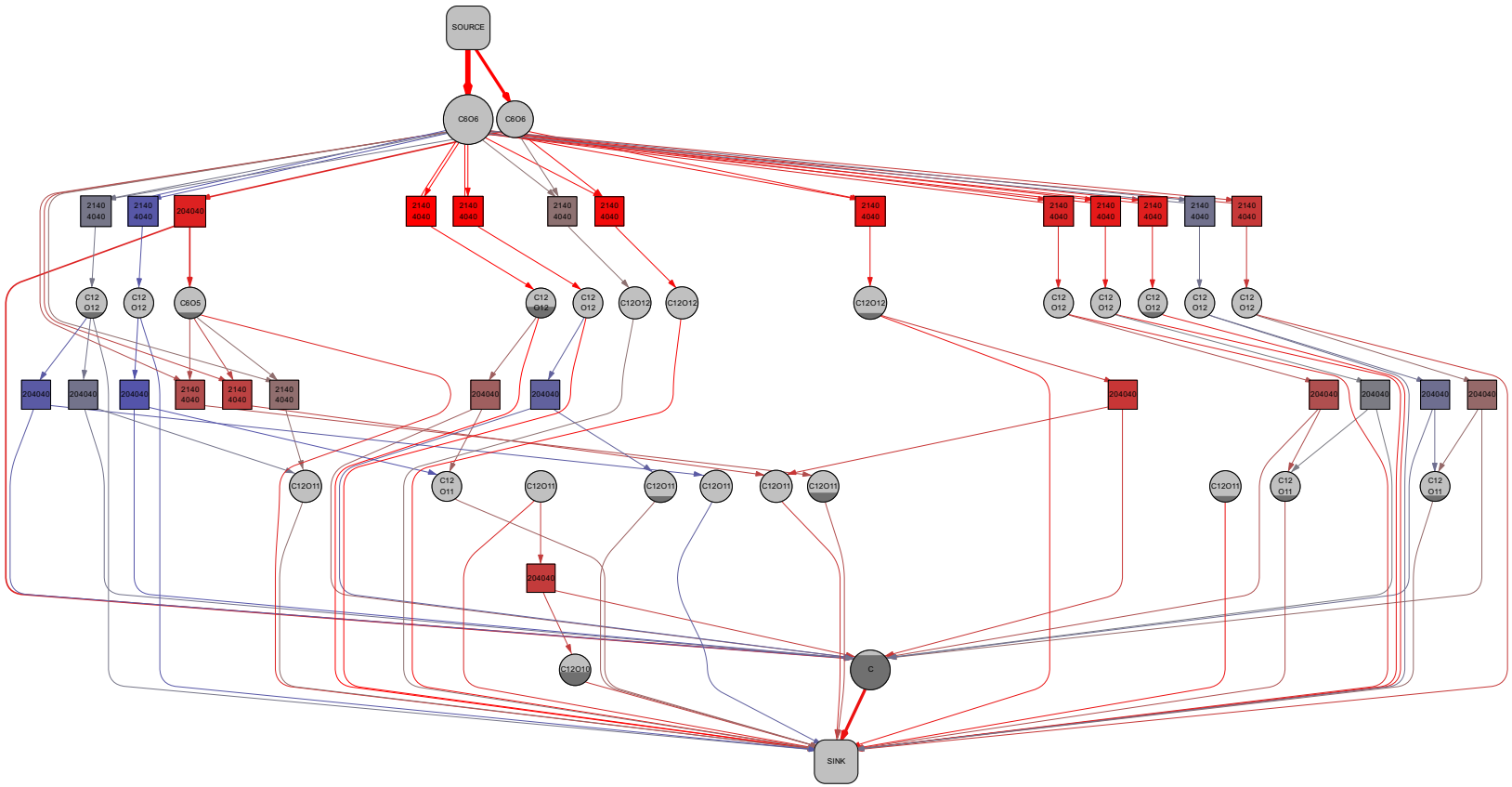
generation 10



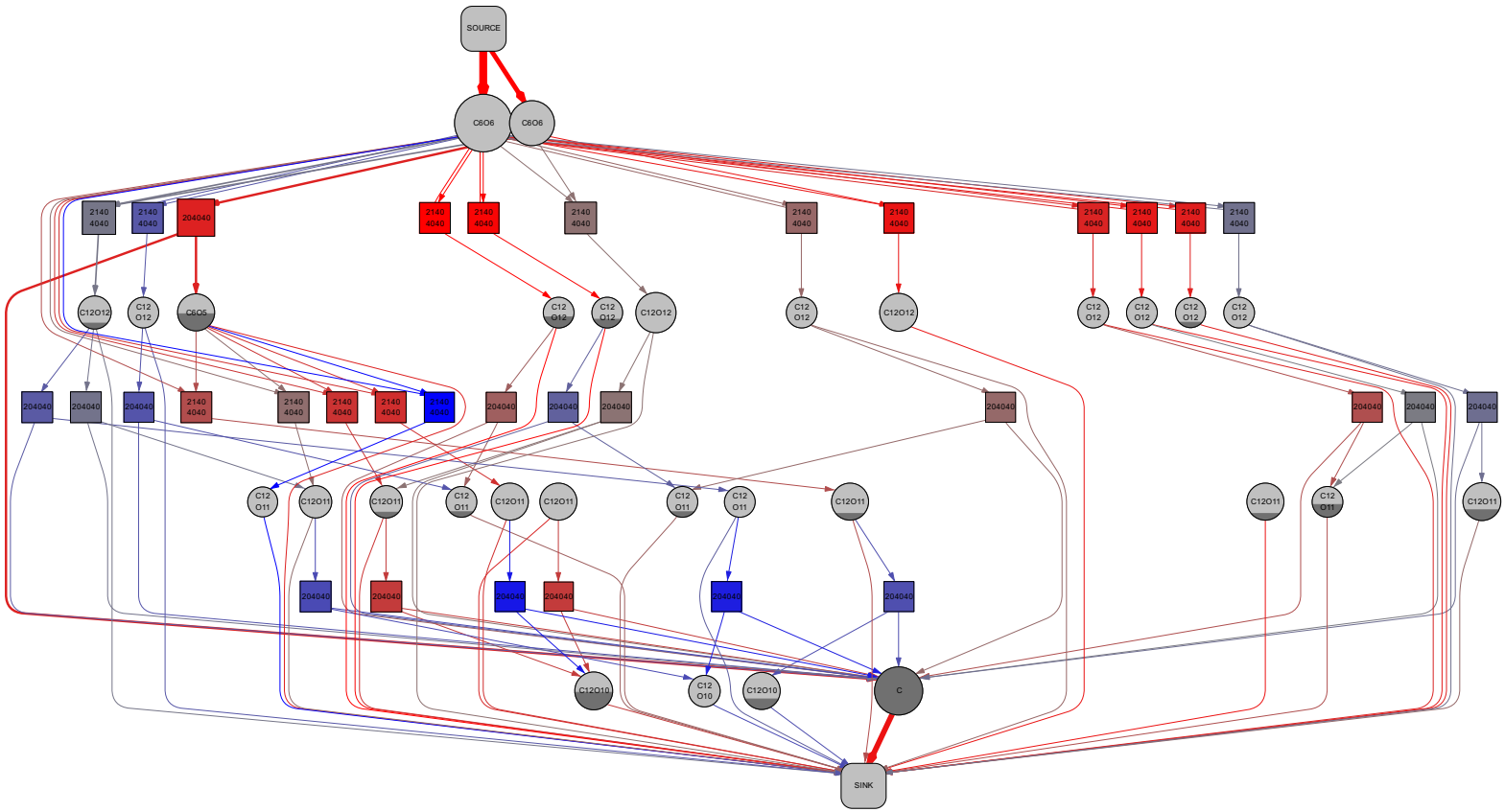
generation 30



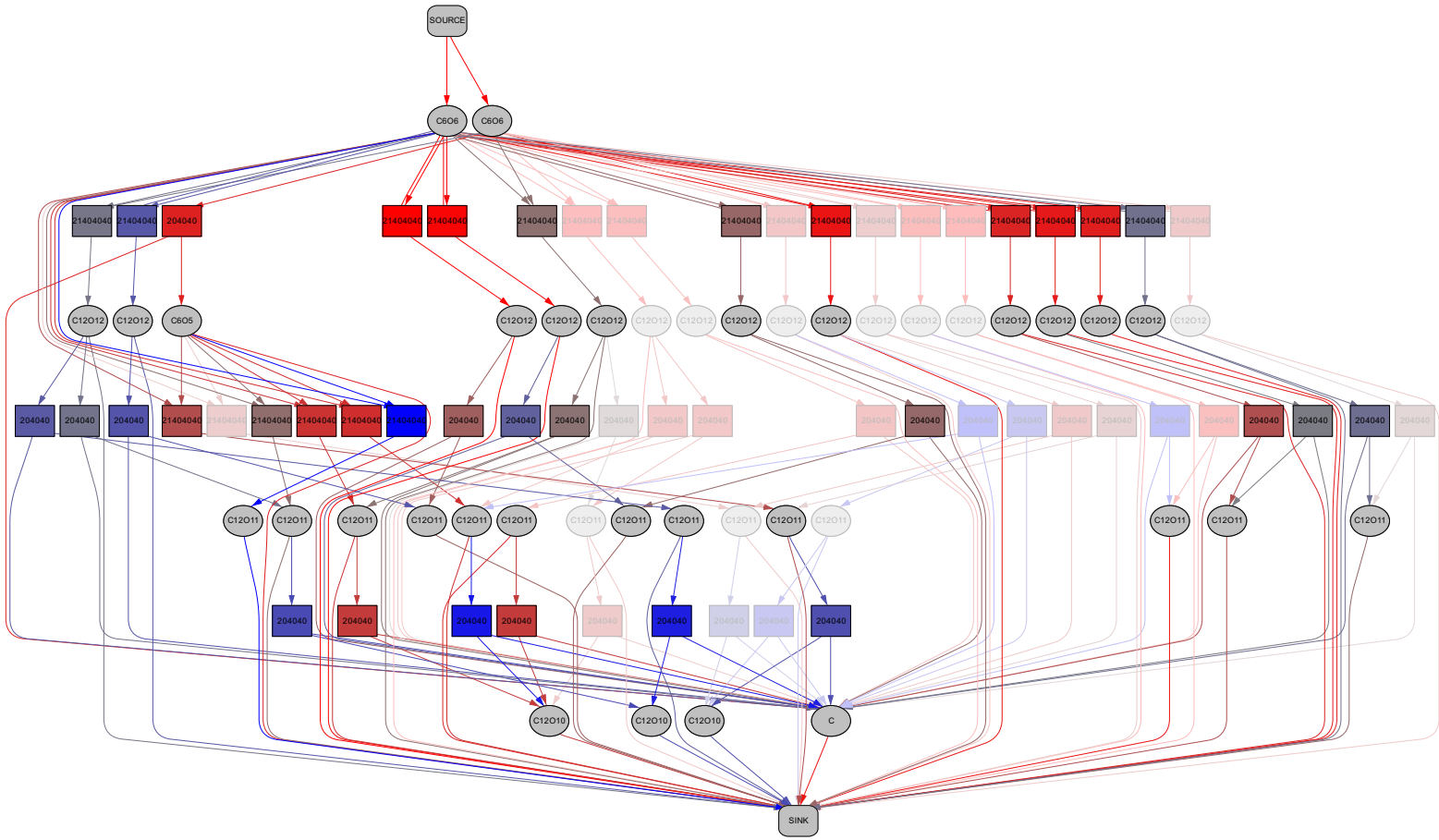
generation 66



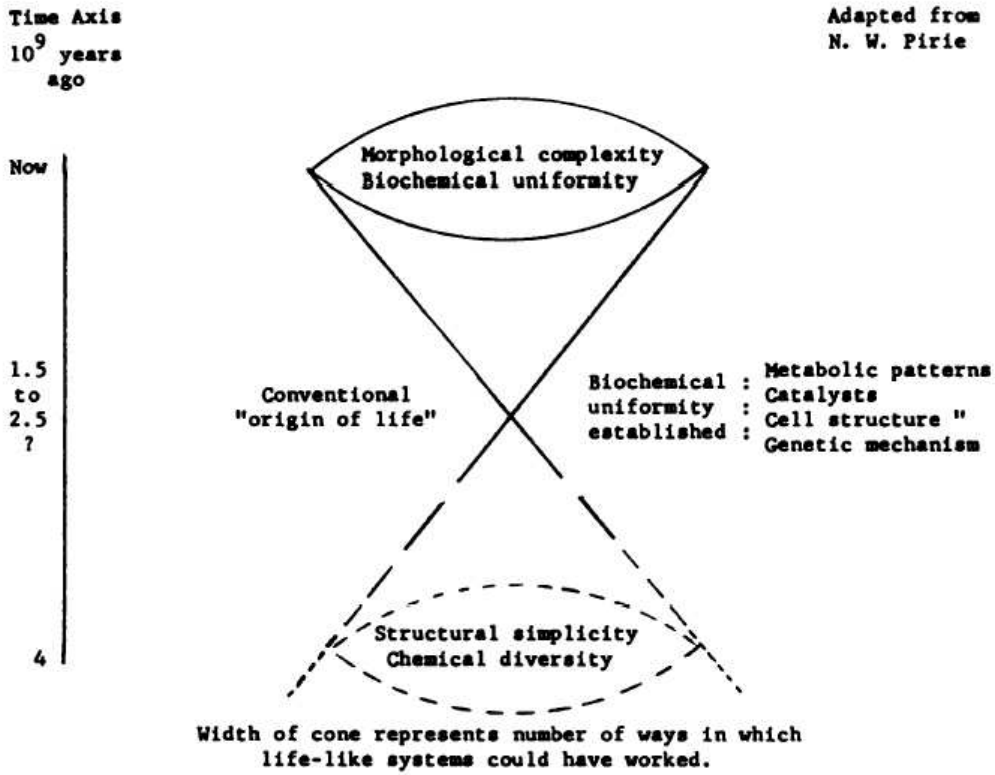
generation 100



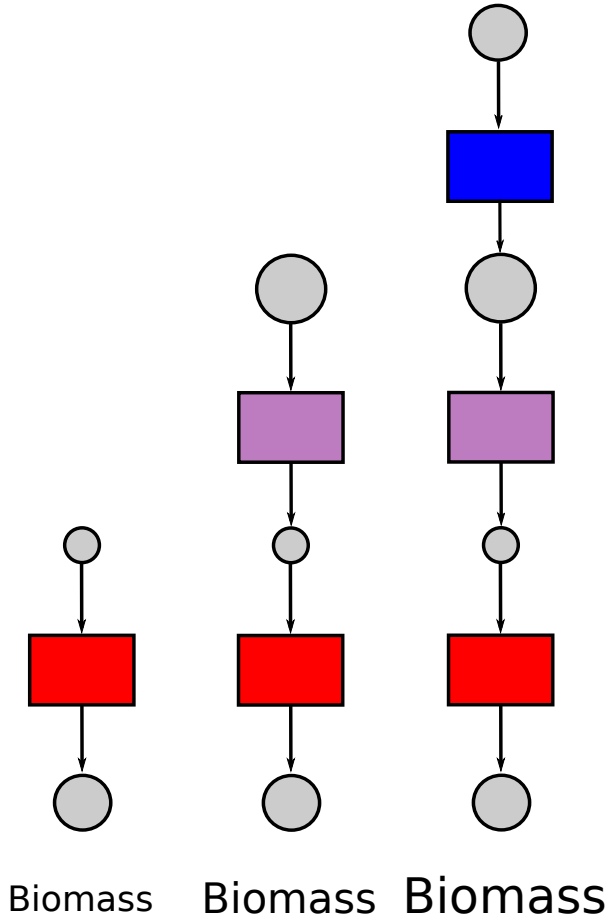
Union graph


SOURCE
C6O6
C6O6
21404040
204040
C12O12
C6O5
C12O11
C12O10
C
SINK

Evolution of early Metabolism

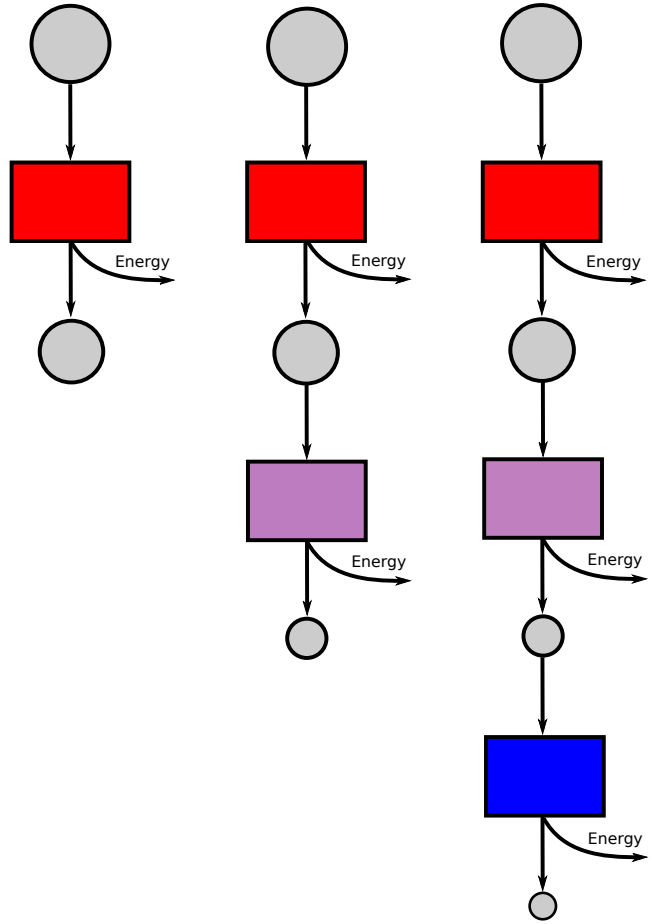


Retrograde Evolution



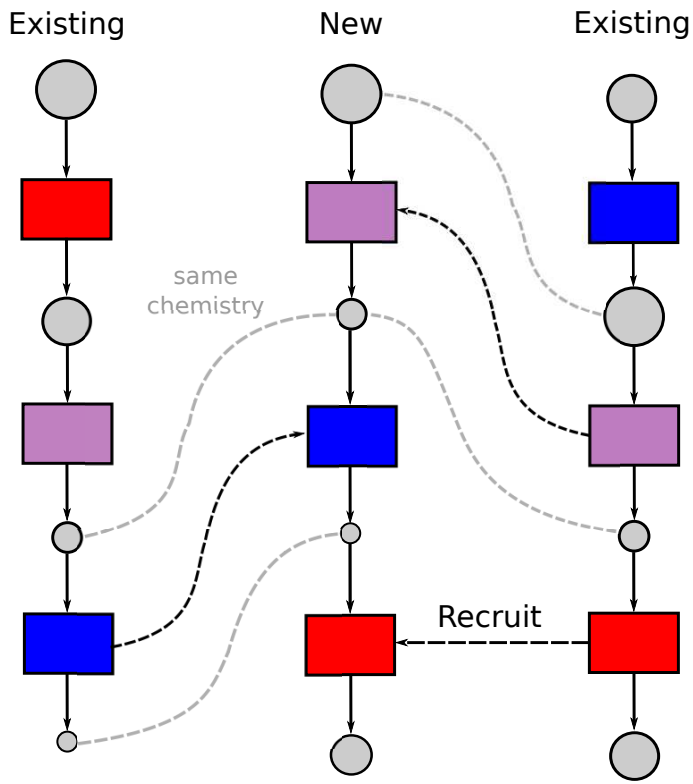
- End-product can be derived from more and more distant metabolites
- Example: glycolytic pathway, histidine biosynthesis

Forward Evolution



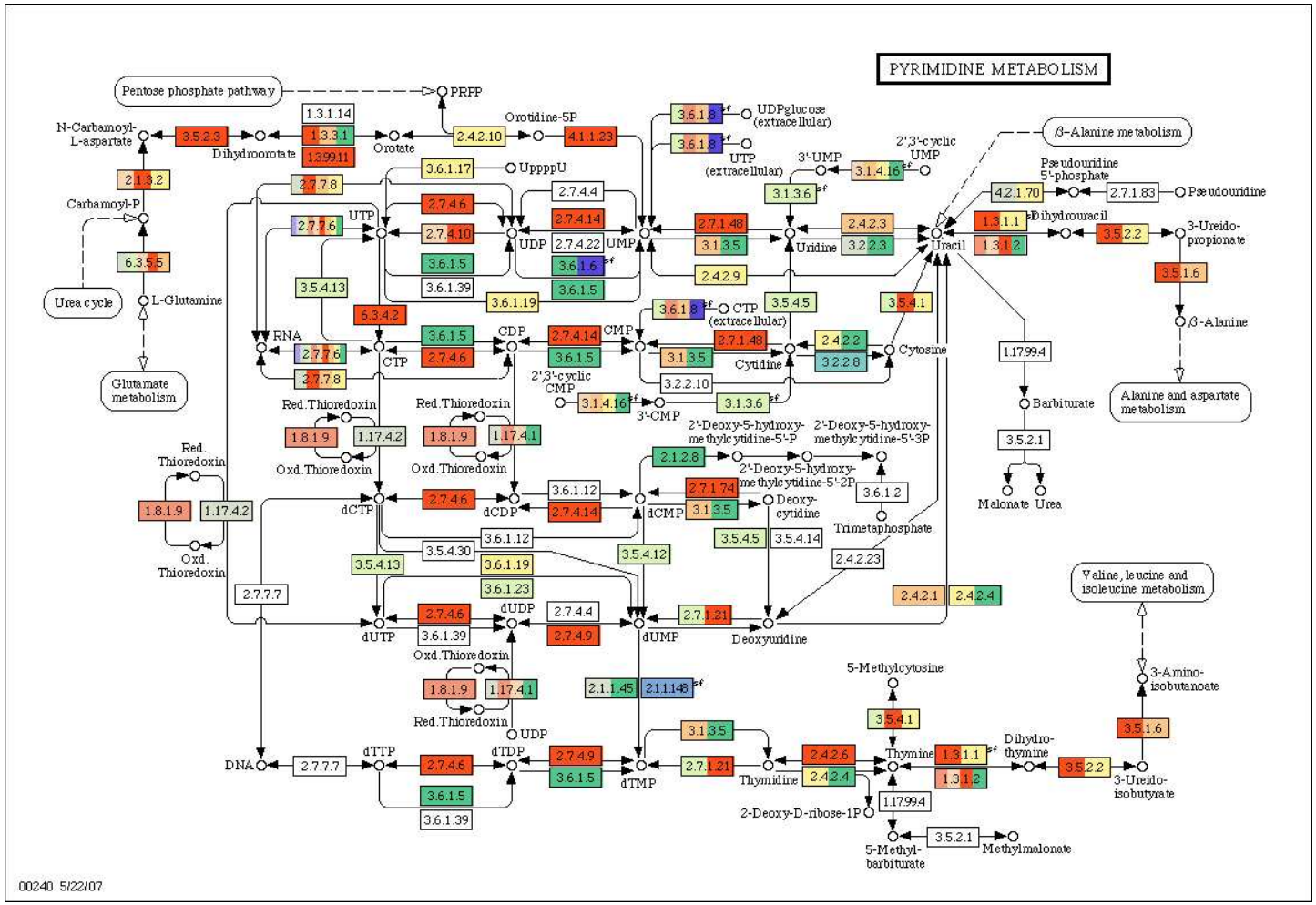
- more efficient extraction through deeper break-down of metabolites
- Example: isoprene lipid pathway

Patchwork Evolution



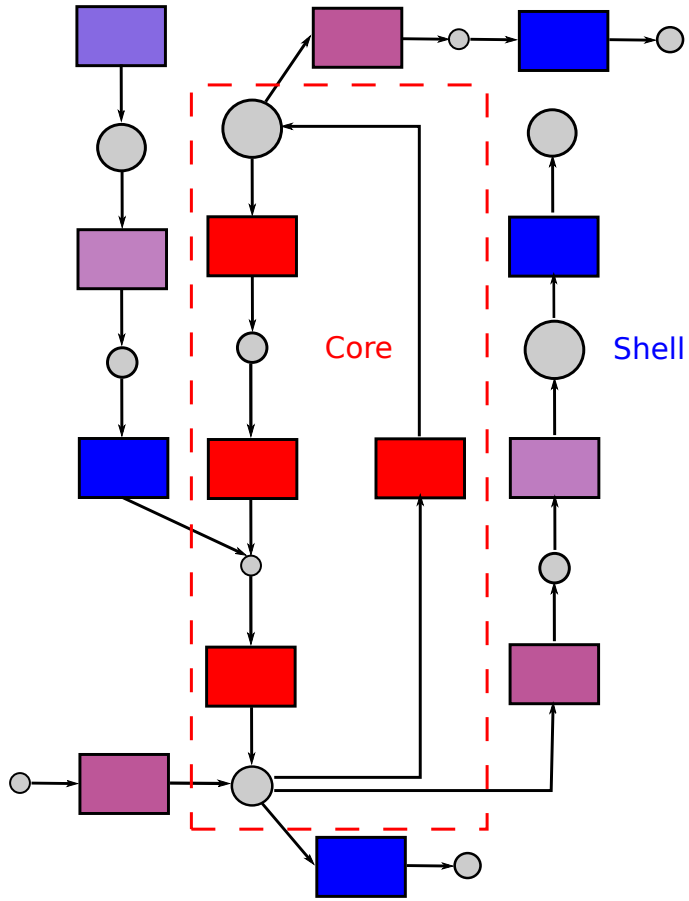
- Enzyme Recruitment from other Pathways
- Example: TIM β/α-barrel fold architecture in modern metabolism

Patchwork Evolution



- Pyrimidine metabolism (from MANET)

Shell Hypothesis

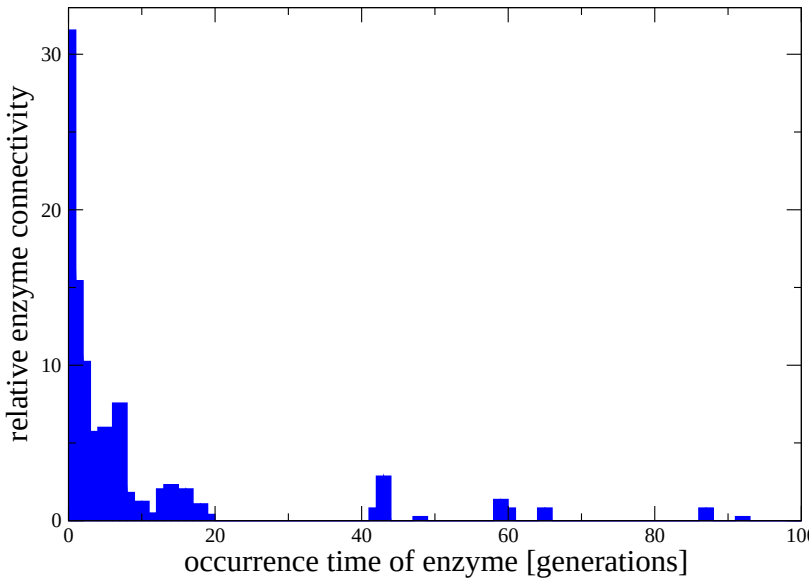


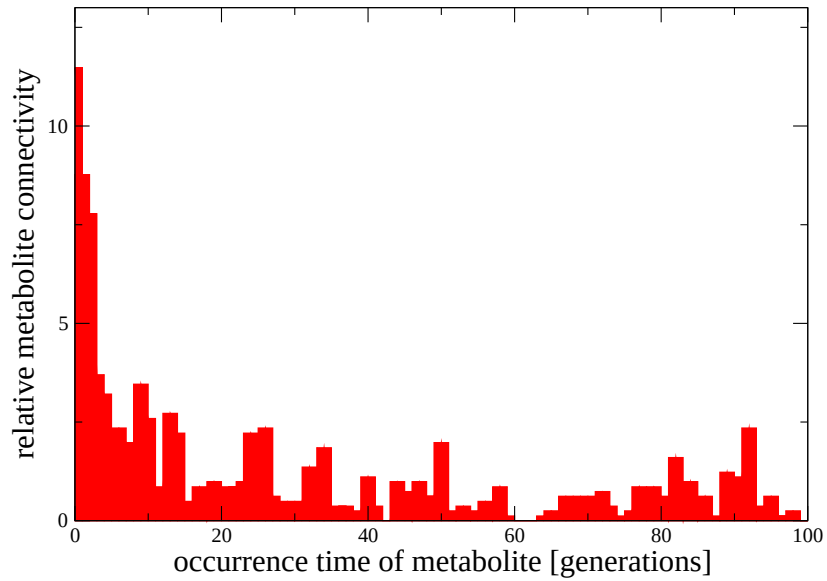
- A core from which pathways can be recruited
- Example:auto-catalytic core of the reductive citric acid cycle

Results

- Quantitive Analysis
  - Connectivity vs Age (Time of Occurence)
  - Evolution of Pathways (Direction)
- Study on Example
  - Evolution of Pathways (Life-time of enzymes, molecules)
  - Geneaology (History of Genes)

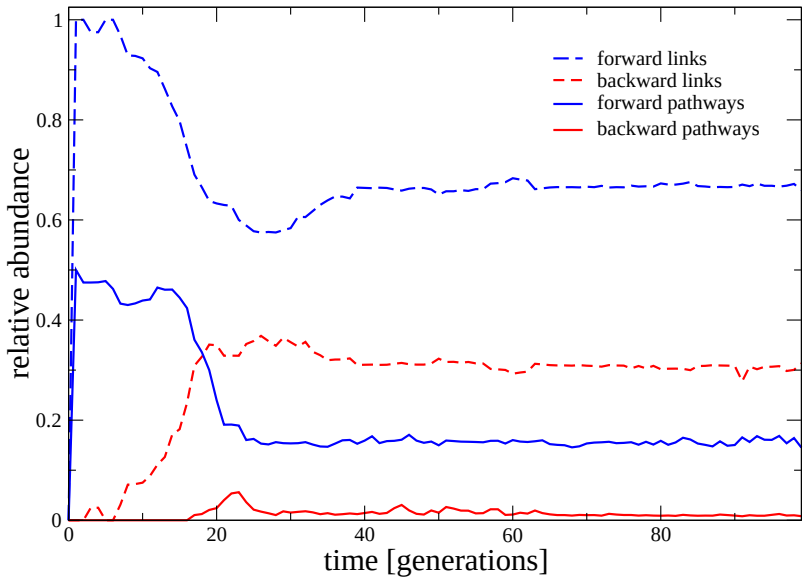
Results - QA

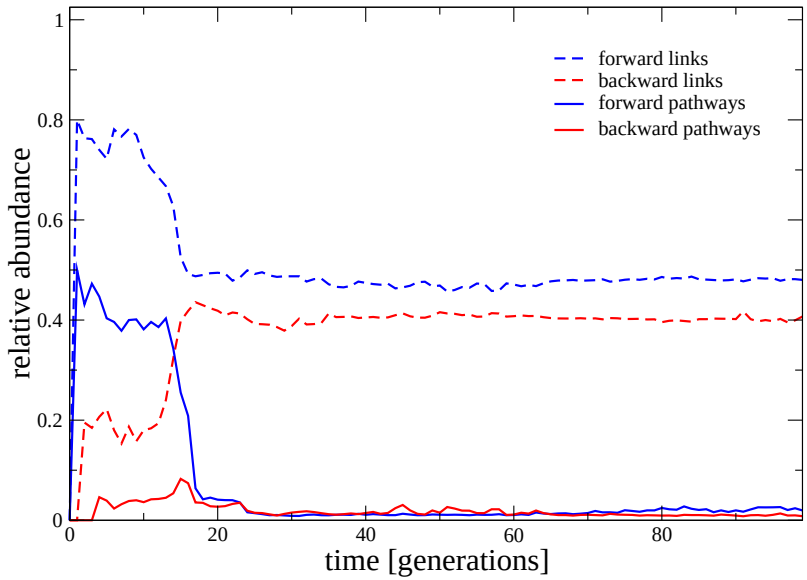




- Highly connected metabolites (hubs) originate from early generations.
- Enzymes from later stages have higher specificty.

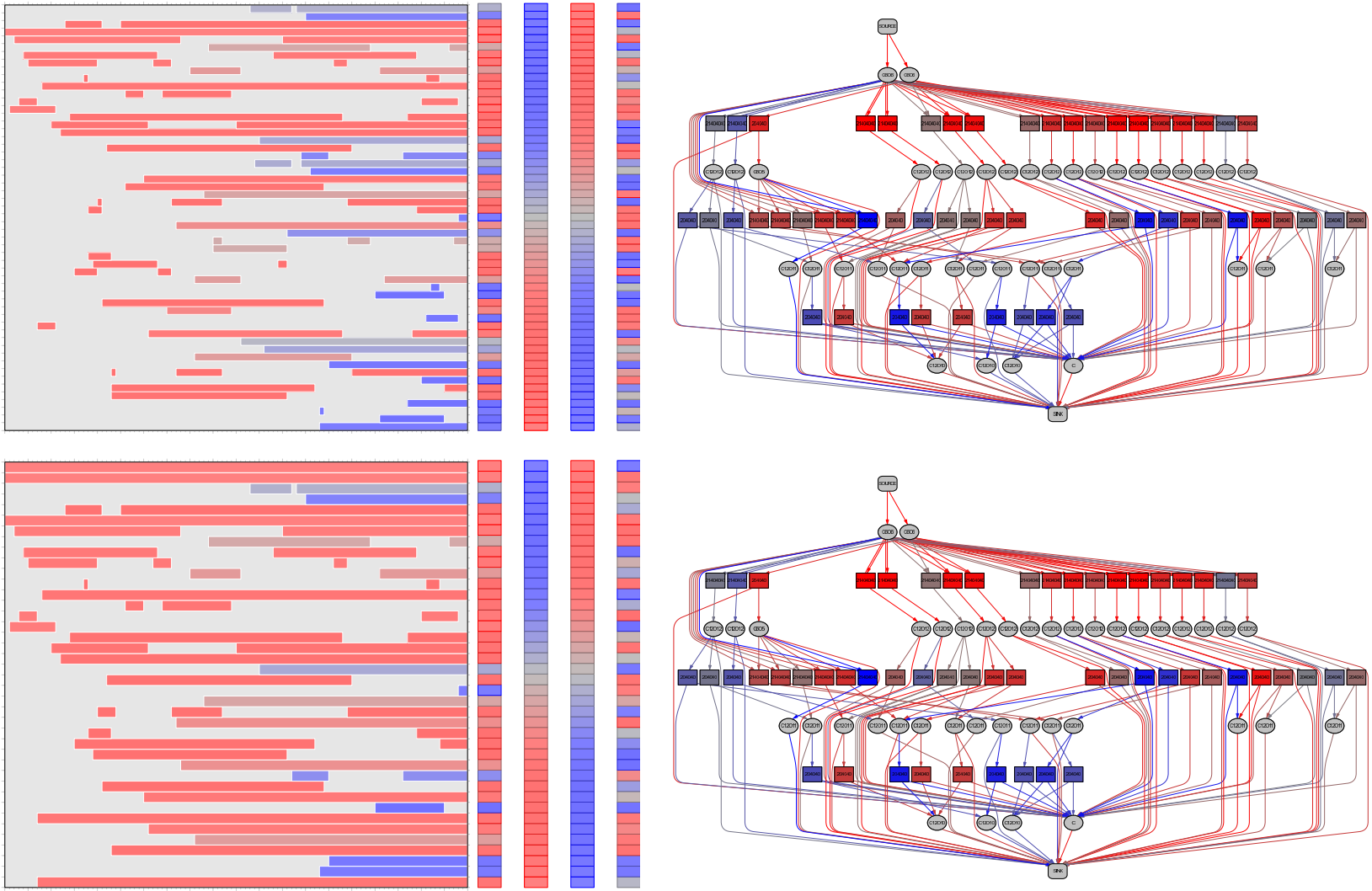
Results - QA





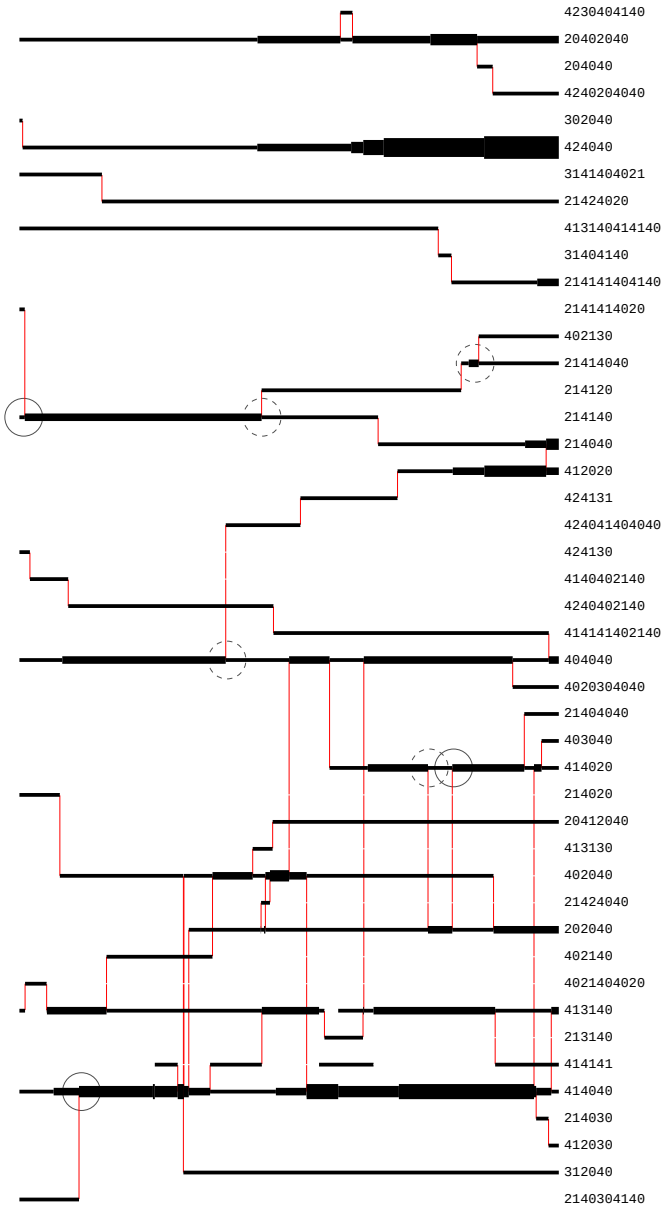
- First generations are dominated by forward evolution.
- When a certain network size is reached, enzyme recruitment takes over.
- A core of pathways from early generations is kept.

Results - small Example



- Enzyme pattern similar to forward simulation pattern.

Results - small Example

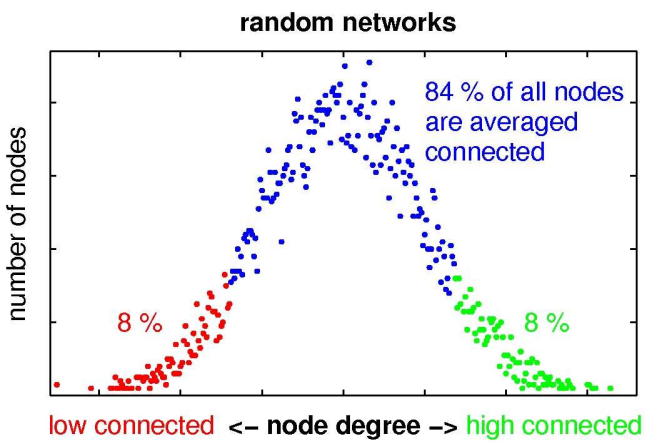


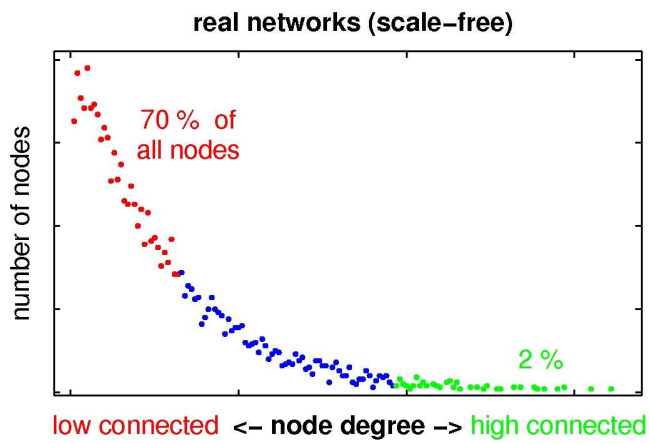
- Genealogy of catalytic functions and gene dosage over 2000 generations.
- Convergent as well as divergent events.

Emergence and Evolution of systematic Properties

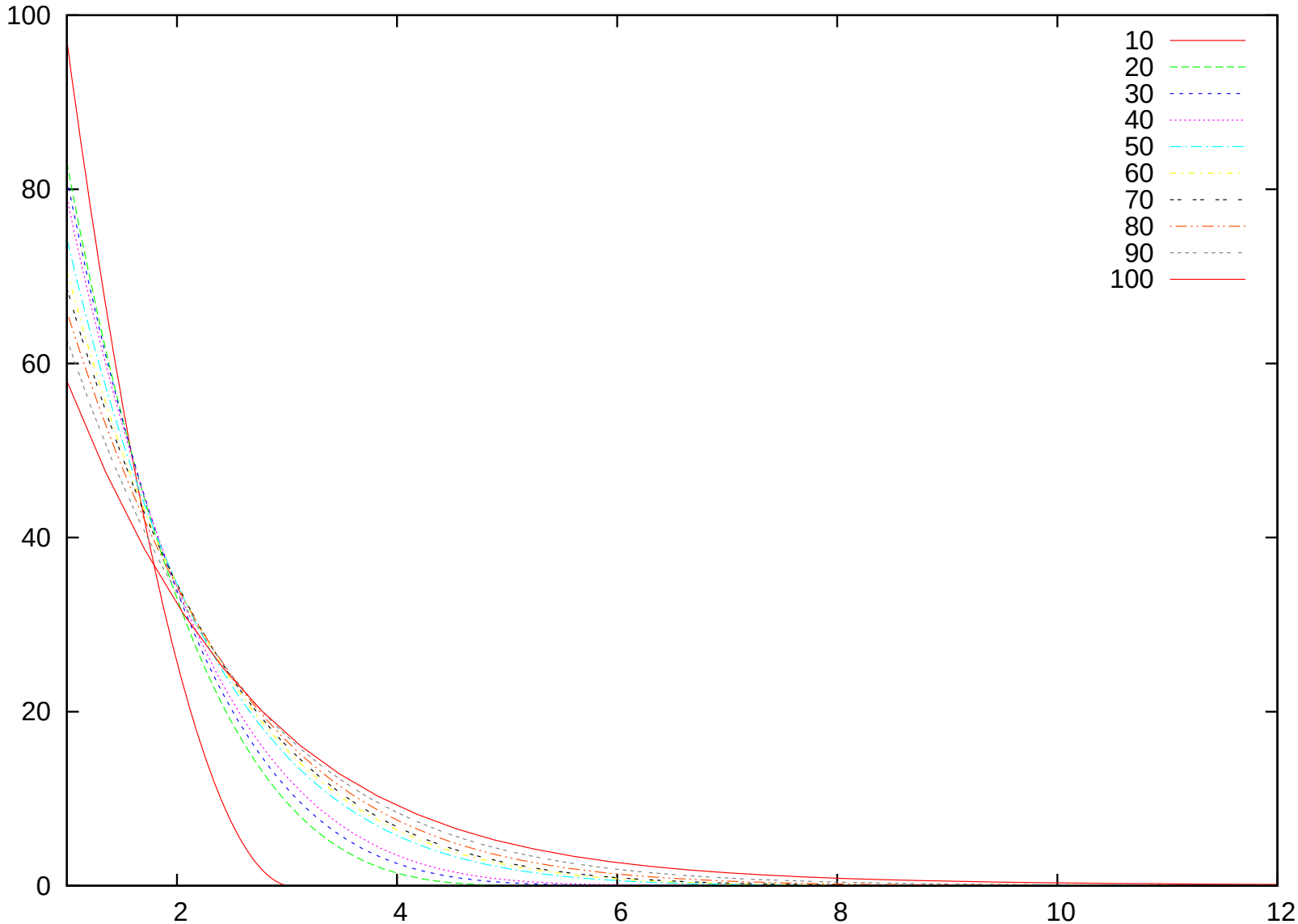
General network analysis

- Connectivity Distribution
  - small vs big
  - early vs evolved
- Centrality, Entropy, ...
  - simulated vs real world





Connectivity Distribution



Metabolic network analysis

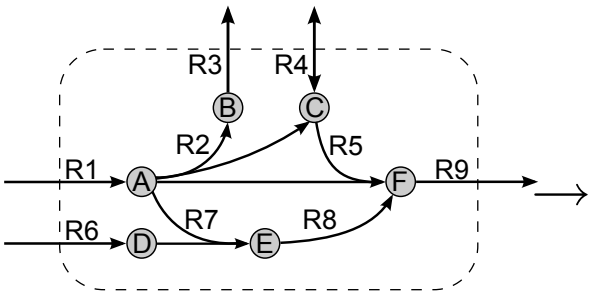
We have sets of edges forming meaningful complex entities

$\downarrow$

pathways

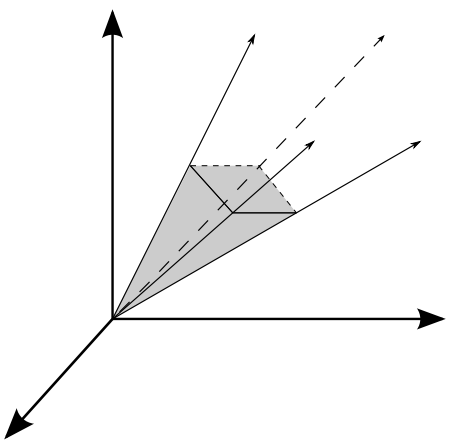
- number of pathways $\longrightarrow$ flexibilty
- change in case of single/multiple knockouts $\longrightarrow$ robustness
- number of acceptable knockouts $\longrightarrow$ robustness

Metabolic Pathway Analysis



$$
\begin{array}{c}
 \\
A \\
B \\
C \\
D \\
E \\
F
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 & R8 & R9 \end{array} \\
\left(\begin{array}{ccccccccc}
1 & -1 & 0 & 0 & -1 & 0 & -1 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & -1
\end{array} \right)
\end{array}
$$

$\downarrow$

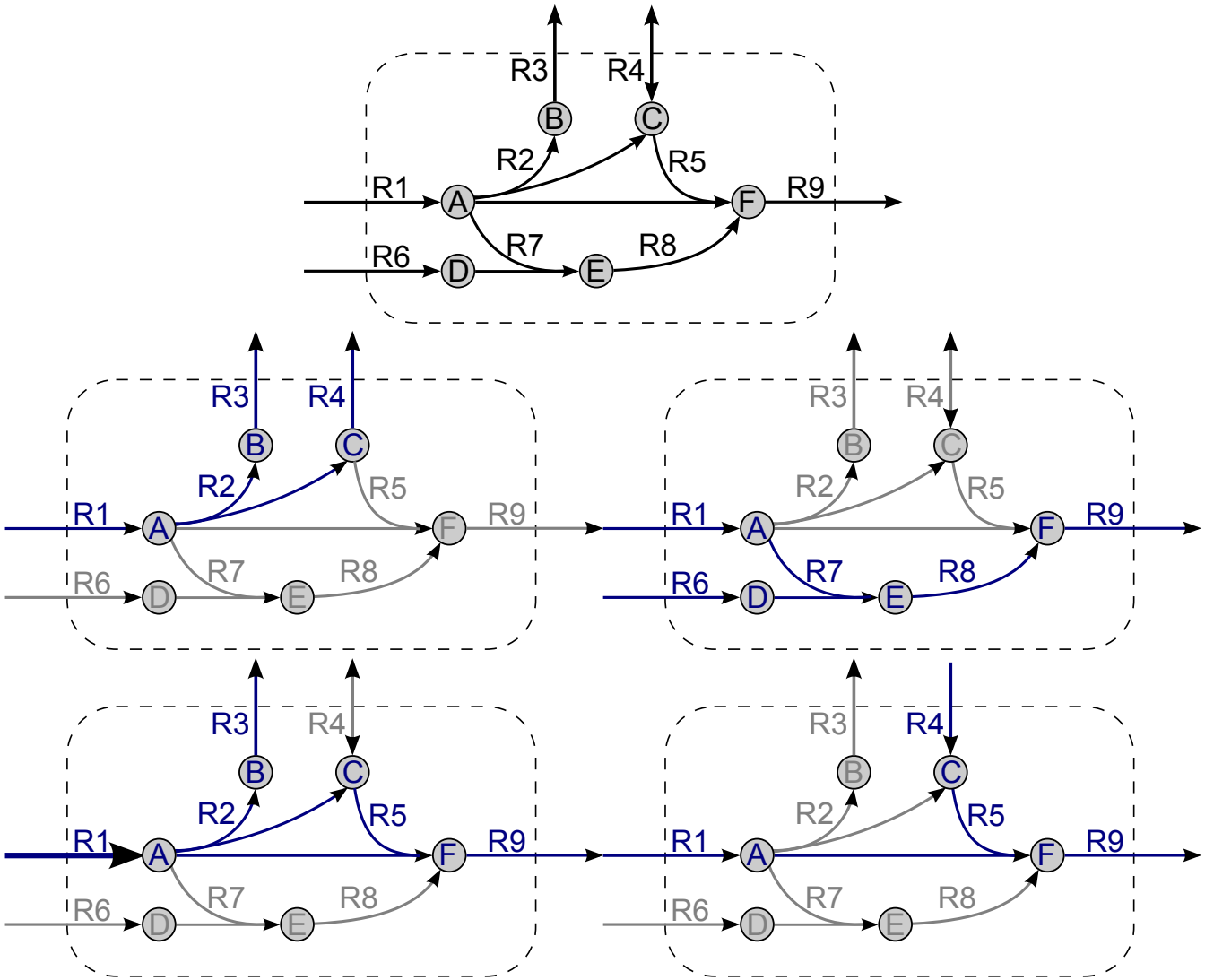


$\leftarrow$

$$S * v_i = 0$$

$$
\begin{array}{c}
 \\
R1 \\
R2 \\
R3 \\
R4 \\
R5 \\
R6 \\
R7 \\
R8 \\
R9
\end{array}
\begin{array}{c}
\begin{array}{cccccc} v_1 & v_2 & \cdots & v_{n-1} & v_n \end{array} \\
\left(\begin{array}{cccccc}
2 & -1 & \cdots & 0 & 0 \\
1 & -1 & \cdots & -1 & -1 \\
1 & -1 & \cdots & -1 & -1 \\
0 & 1 & \cdots & 1 & 0 \\
1 & 0 & \cdots & 0 & -1 \\
0 & 0 & \cdots & 1 & 2 \\
0 & 0 & \cdots & 1 & 2 \\
0 & 0 & \cdots & 1 & 2 \\
1 & 0 & \cdots & 1 & 1
\end{array} \right)
\end{array}
$$

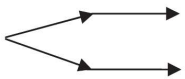
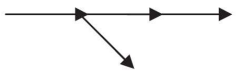
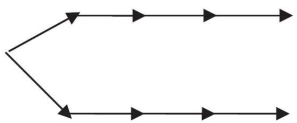
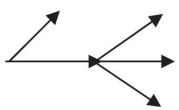
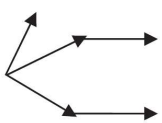
Metabolic Pathway Analysis



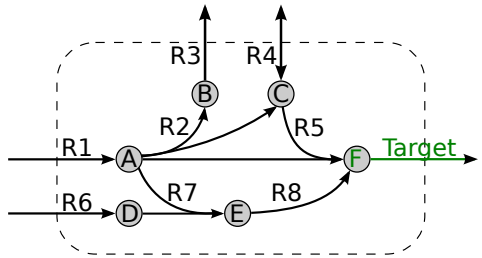
Knockout effects

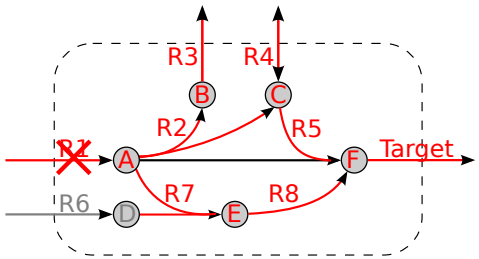
single $R_1 = \dfrac{\sum_{i=1}^{r} z^i}{r * z}$ depletion $R_2 = \dfrac{\sum_{i=1}^{n} R_1^i}{n}$

multiple $R_3(k) = \dfrac{\sum_{i=1}^{s(k)} z^i}{s(k) * z}$ overall $R_3(\leq K) = \sum_{k=1}^{K} R_3(k) p_k$

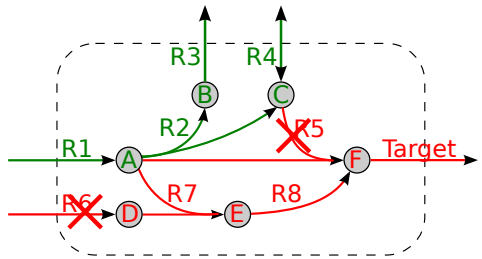
Example system	Number of reactions	Number of elementary modes	$R_1(1)$	$R_1(2)$	$R_1(3)$	$R_1(\leqslant 3)$
1 	4	2	$1/2 = 0.5$	$1/6 \approx 0.167$	0	0.414
2 	4	2	$1/2 = 0.5$	$1/4 = 0.25$	$1/8 = 0.125$	0.436
3 	4	2	$3/8 = 0.375$	$1/12 \approx 0.083$	0	0.302
4 	4	2	$1/4 = 0.25$	0	0	0.189
5 	8	2	$7/16 \approx 0.438$	$3/8 = 0.375$	$5/16 \approx 0.313$	0.418
6 	8	2	$1/2 = 0.5$	$3/14 \approx 0.214$	$1/14 \approx 0.071$	0.416
7 	5	4	$13/20 = 0.65$	$3/8 = 0.375$	$7/40 = 0.175$	0.573
8 	5	3	$2/3 \approx 0.667$	$2/5 = 0.4$	$1/5 = 0.2$	0.592

Minimal Knockout sets

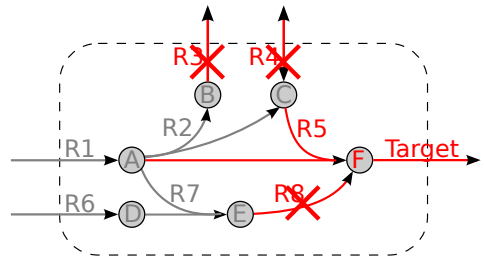




$\{R1\}$

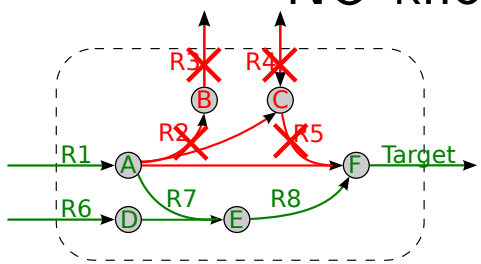


$\{R5, R6\}$

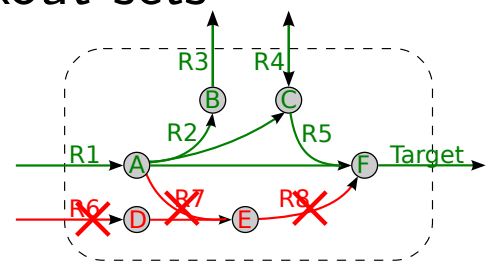


$\{R3, R4, R8\}$

NO knockout sets



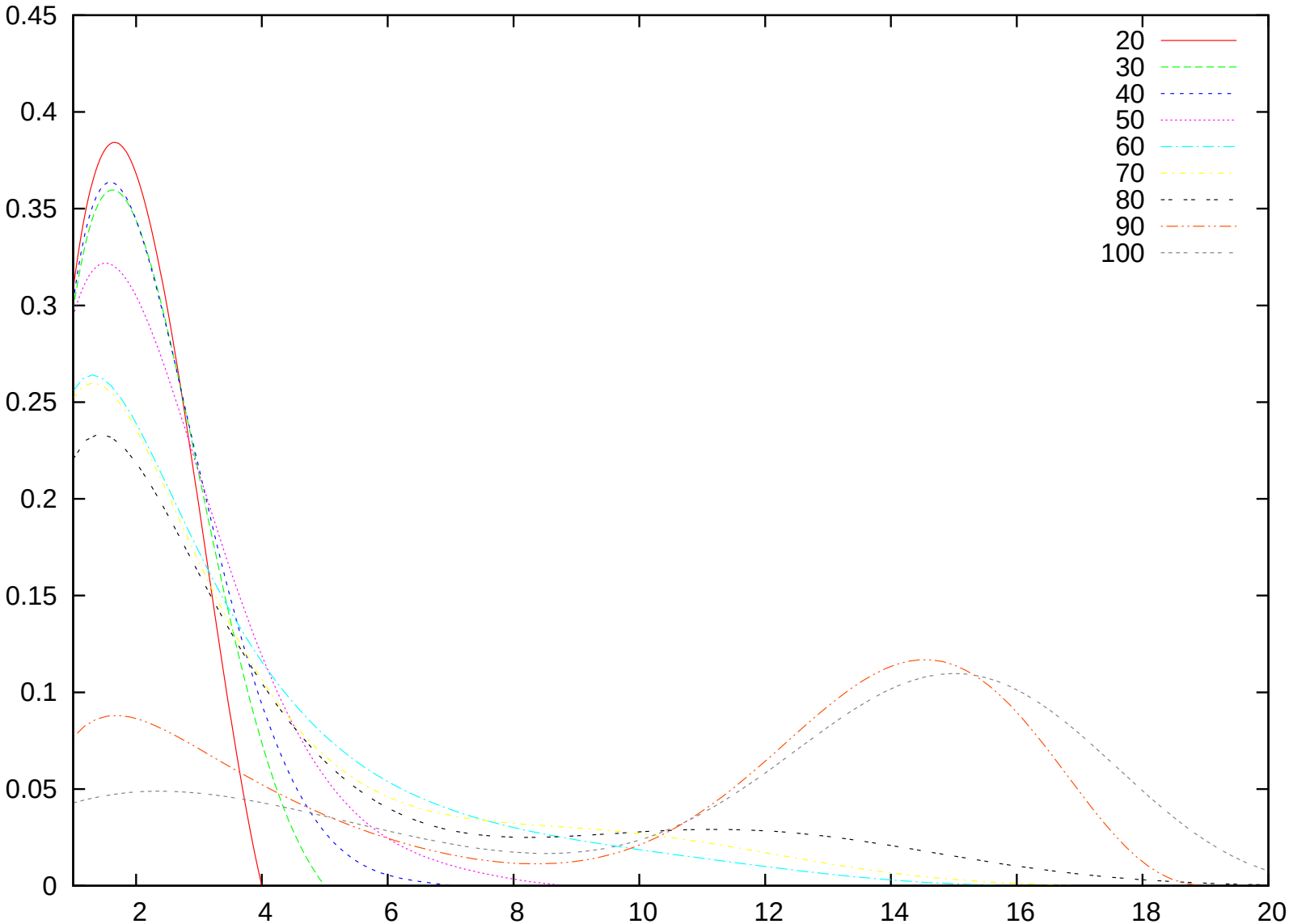
$\{R2, R3, R4, R5\}$



$\{R6, R7, R8\}$

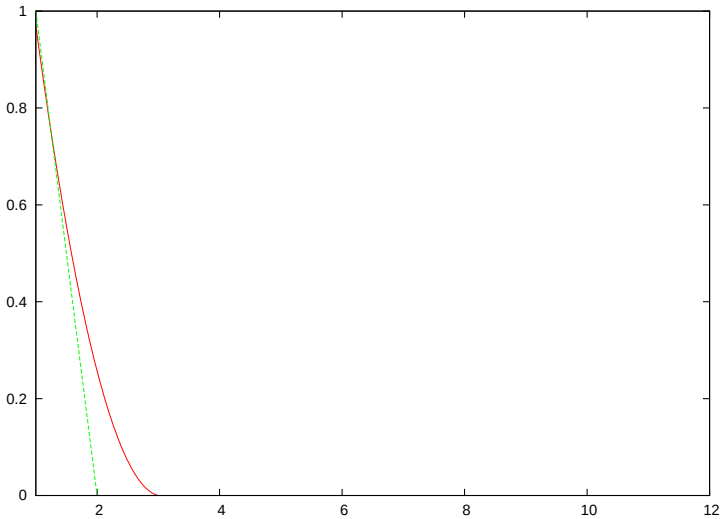
Knockout set size distribution $\longrightarrow$ Robustness

Knockout set size distribution

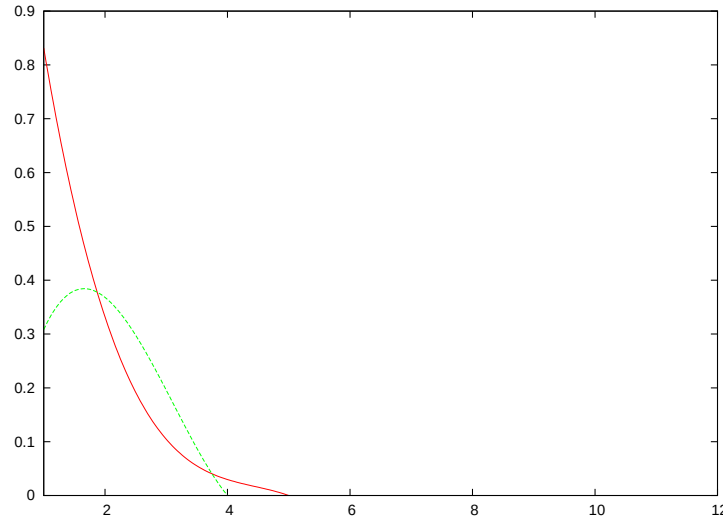


Robustness

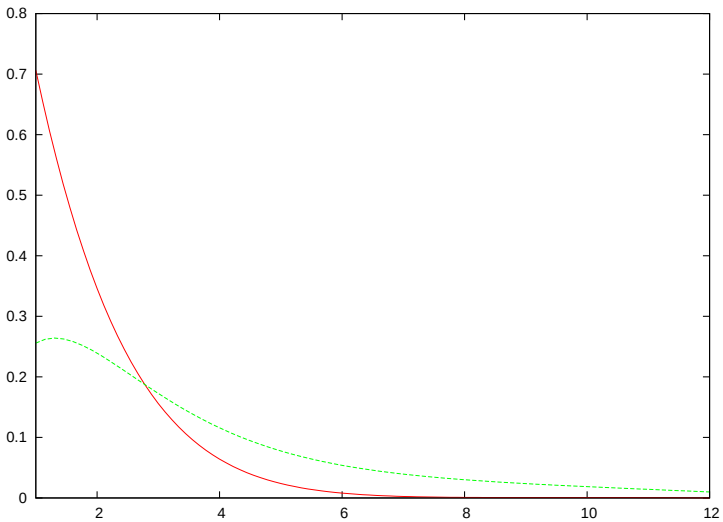
Robustness = 0.51



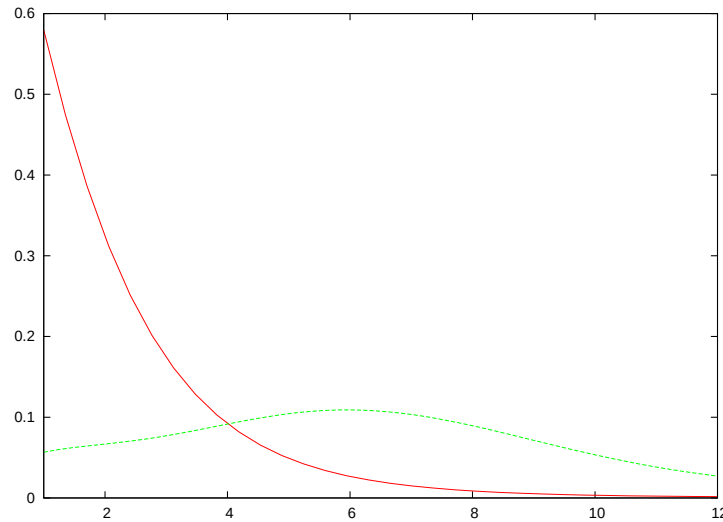
Robustness = 0.67



Robustness = 0.75

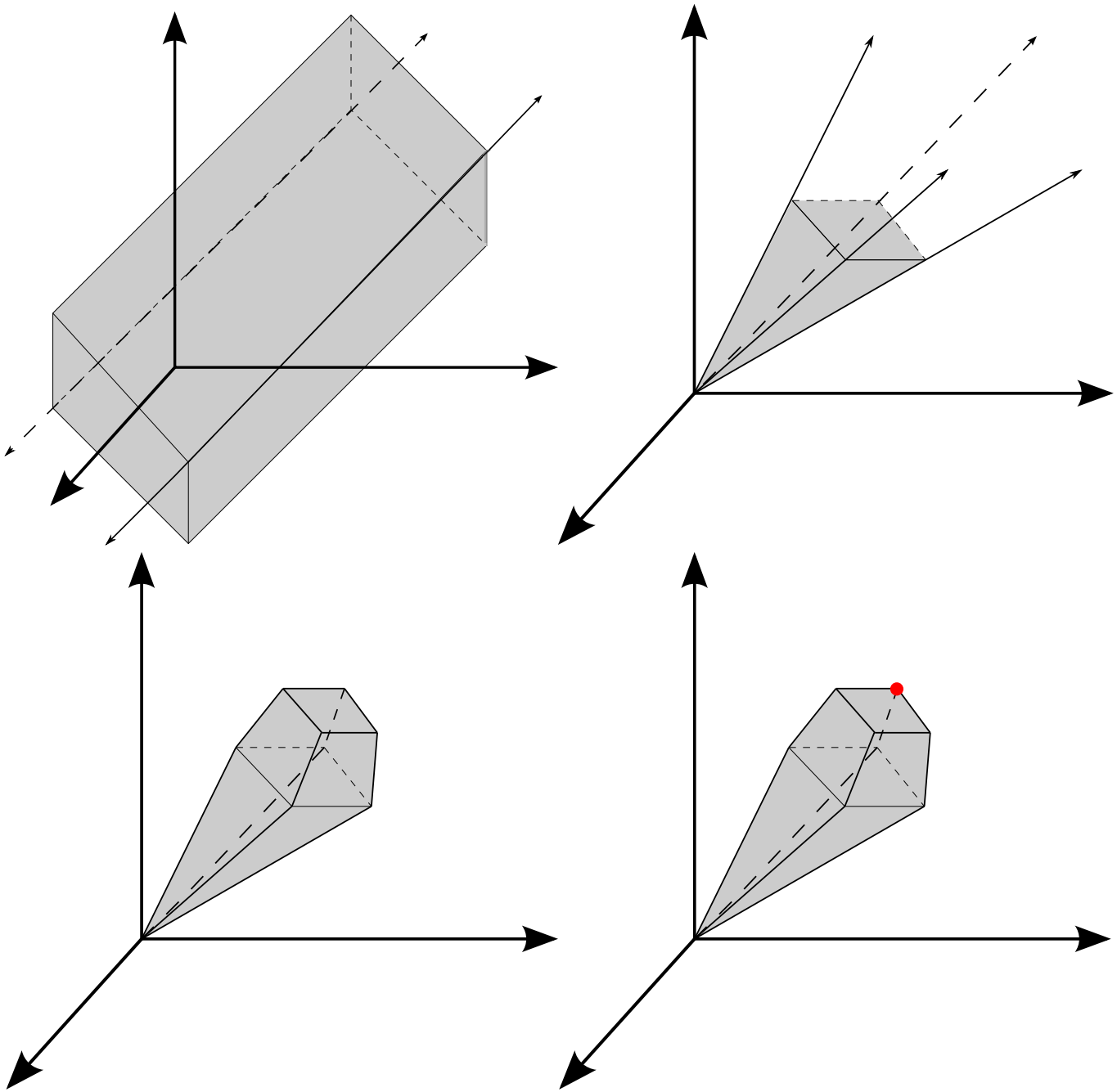


Robustness = 0.81



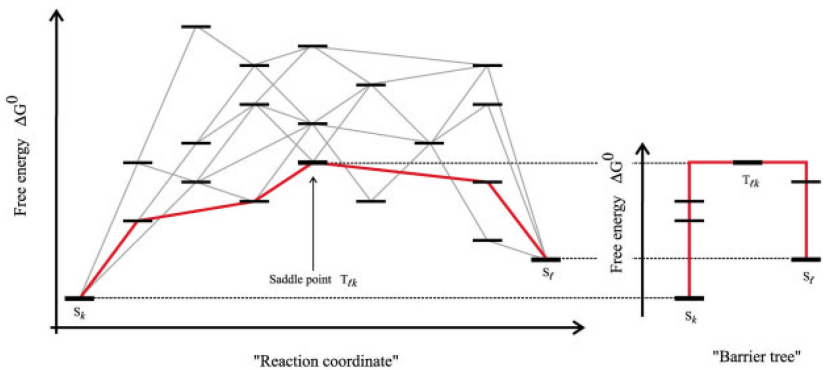
Work in Progress

FBA



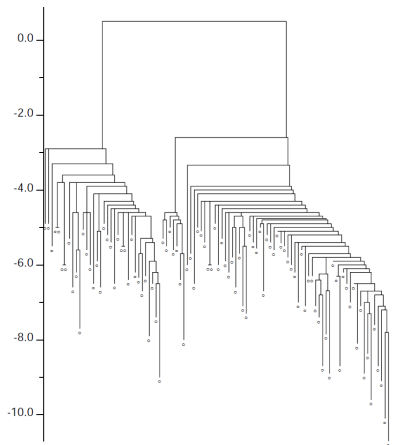
Flux barrier analysis

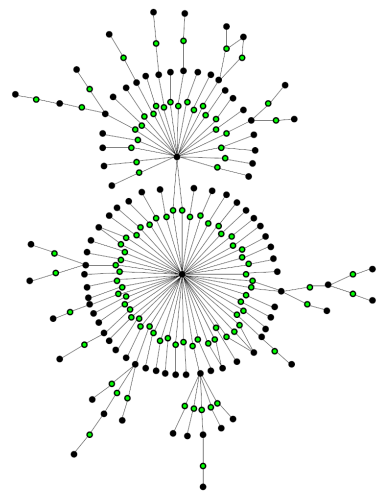
- linear optimization: EMs modeled as system of linear equations
- constraints: limits on reactions, exclusion of combinations of EMs
- barrier tree



Reaction barrier analysis

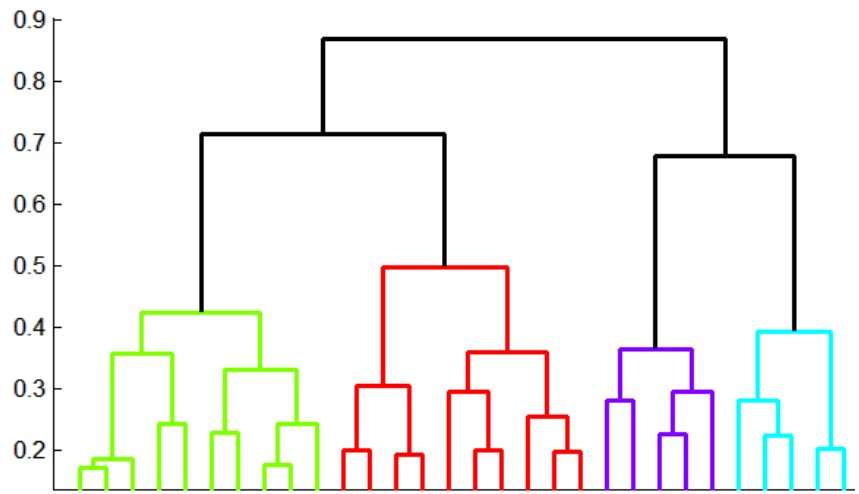
- linear optimization: stoichiometrix matrix
- constraints: limits on reactions, exclusion of combinations of reactions
- barrier tree





Flux similarity

- Compute pairwise similarity of elementary modes
- similarity between metabolites (in+out / all) through topological indices
- similarity between enzymes/reactions by comparing transition state structure



Conclusion

- Summary
  - Computational model of early metabolism
  - Insights in evolution of complex system
  - Combining different pathway evolution hypotheses
  - Explaining hypotheses through scenarios
  - Vizualisation + Network analysis
  - Emergence and evolution of network properties
- Outlook
  - Investigating further porperties (modularity with organizations)
  - Metabolic neutral network

Acknowledgements

Christoph Flamm

Peter Stadler
Konstantin Klemm
Martin Mann
Markus Rohrschneider