Evolution of Metabolism in a Graph-based Toy-Universe

Alexander Ullrich*[1], Christoph Flamm[1]
[1] University of Vienna, Institute for Theoretical Chemistry
* aullrich@tbi.univie.ac.at

WWTF Wiener Wissenschafts-, Forschungs- und Technologiefonds

Introduction

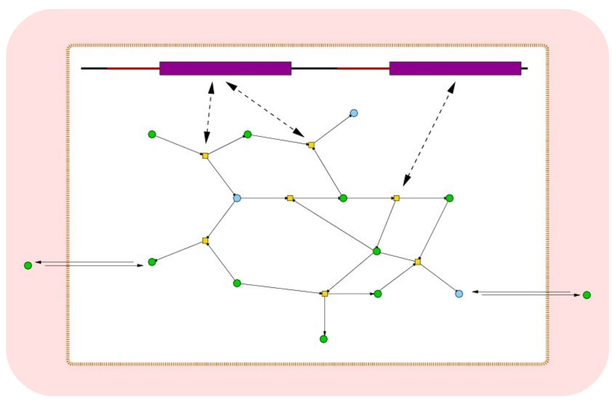
The origin and evolution of metabolism is an interesting field of research with many unsolved questions. Simulation approaches have proven to be helpful in explaining properties observed in real world metabolic reaction networks[1].
We propose here a more complex and intuitive graph-based model combined with an artificial chemistry. Instead of differential equations, enzymes are represented as graph rewriting rules and reaction rates are derived from energy calculations of the involved metabolite graphs.
The generated networks were shown to possess the typical properties of real world metabolisms. Using our metabolic pathway analysis tools we yield hypotheses about the evolution of catalytic molecules and its effect on the emergence of system level properties.

Model

All entities of metabolism are modeled as graphs. Processes are realized through applications on the graph structure.
The graphs for metabolites have the atoms of the chemical molecule as vertices and chemical bonds are represented by edges. The vertex labels indicate the atom-type (H,O,N,C), edge labels stand for the bond types (−, =, #).
Since enzymes are modeled as graph rewriting rules (consisting of a left rule and a right rule), enzyme graphs are superimpositions of two graphs (substrate and product). Consequently, two labels for each edge are used.
The connectivities between metabolites and enzymes and therefor the metabolic network are modeled in a bipartite labeled graph.

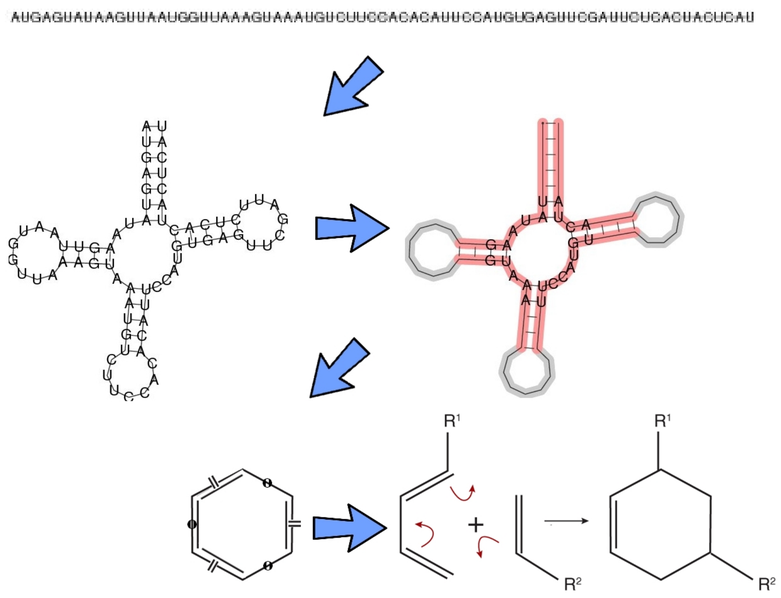
Simulation

Initially, the environment (set of metabolites) and the chemistry (set of chemical reactions) are set and a population of cells containing a genome and a (empty) metabolism are added to the simulation. The genes are expressed by mapping their RNA-sequence to an enzyme function. Enzymes look for reactants, produce new metabolites and thus build up the metabolic reaction network.



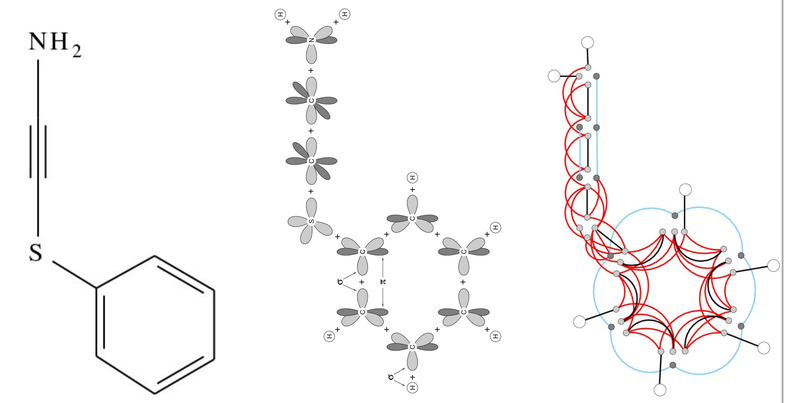
In each generation, individuals are selected based on their metabolic yield and then create new individuals with mutated genomes.

Mapping: Gene to Enzyme



The RNA-sequence is folded into its minimum free energy structure. Information is extracted only from the longest loop and its neighborhood (red marked parts and remaining sequence). Aptamers also mostly have their catalytic site in a loop region. The extracted information is used to map to a reaction index, encoding for a transition state structure[2], which in turn defines a graph rewriting rule. The application of the rewrite rule realizes the enzyme function.

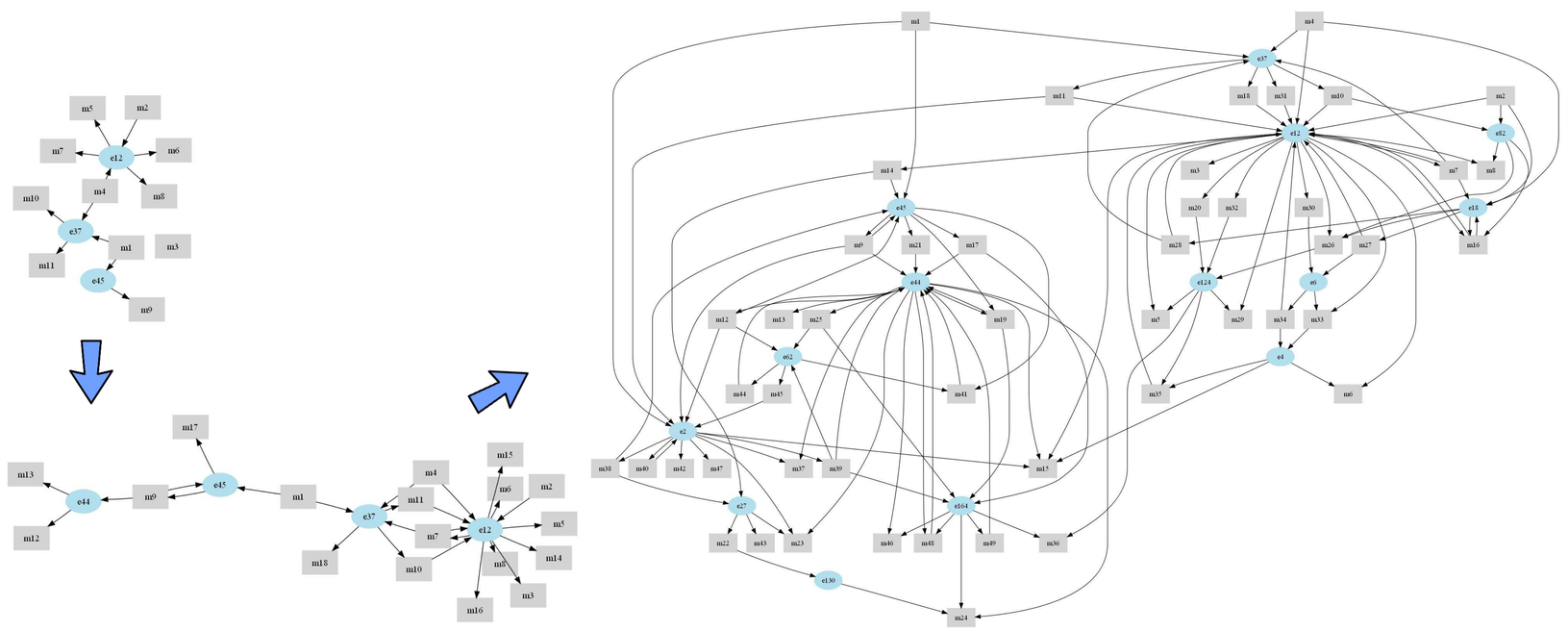
Toy-Universe



Chemical Graph | Hybrid Orbitals | Orbital Graph

For the energy calculation, molecules are represented by their orbital graphs which can be gained from the chemical structure (left graph) through the VSEPR rules. Vertices are atom orbitals (middle graph) and edges indicate overlaps of interacting orbitals (right graph). The energy calculation is an extreme simplification of the extended Hückel theory[3]. The energy is used to exclude unrealistic product molecules and for the calculation of the reaction rates. In the future we will also be able to calculate solvation energies, allowing us to model a more complex cell behavior.

Results

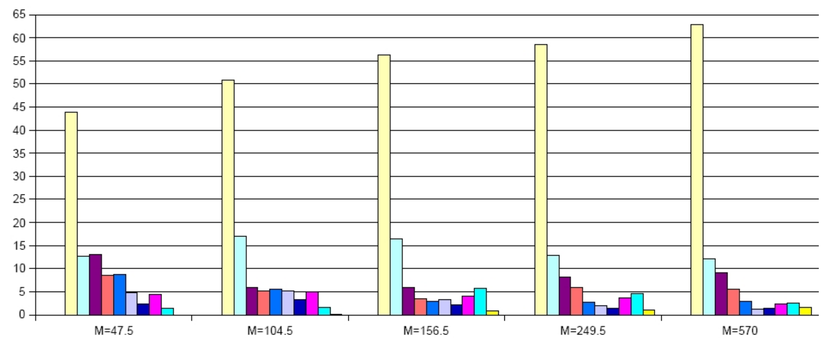


Connectivity / *Metabolites*	1	2	3	4	5	6	7	8 - 12	>12
avg(m)=47.5	43.91	12.61	13.03	8.61	8.82	4.83	2.31	4.41	1.47
avg(m)=104.5	50.89	17.1	6.02	5.2	5.61	5.2	3.28	5.06	1.64
avg(m)=156.5	56.21	16.36	5.85	3.56	2.92	3.29	2.1	4.02	5.67
avg(m)=249.5	58.59	12.89	8.13	6.01	2.69	2	1.37	3.67	4.64
avg(m)=570	62.8	12.1	9.12	5.49	2.89	1.2	1.46	2.34	2.63

Connectivity of metabolites in networks of different sizes. Frequency in %

Enzyme	e37	e45	e12	e44	e18	e27	e6	e82	e4	e62	e124	e130	e2
Generation	1	1	1	2	3	3	7	10	14	42	44	51	53
Connectivity	4	5	12	9	4	2	2	2	2	2	3	1	6

Specificity of enzymes in the example network



Several simulations with equal starting conditions were run for 100 generations. They yielded to metabolic networks of different sizes. We observed the connectivity of the metabolites in the networks. The table and histogram above show the distribution of the connectivities for groups of networks of different sizes. It can be stated that the connectivities in our networks follow the power law as is expected in real world metabolisms[1]. We further investigated the simulations one by one, for one example its network graphs from the first two and the last generation are displayed above. By looking at the connectivity of the enzymes and the generation in which they occurred in the sample simulation (see lower table), we make the observation that highly connected enzymes(broad specificity) mostly occur in the first generations, whereas later only rather specific enzymes are able. The same observation can be made for most simulations.
For more extensive studies we implemented a metabolic pathway analysis tool that produces the set of essential pathways and a fast and memory-efficient module for minimum knockout-set computation. Studying the length-distribution of these pathways and knockout-sets, we can gain deeper insights about the emergence of the robustness in our networks.

References

[1] T. Pfeiffer, O.S. Soyer, S. Bonhoeffer. The Evolution of Connectivity in Metabolic Networks In *PLoS Biology 2005*
[2] S. Fujita. Description of organic reactions based on imaginary transition structures. 1. Introduction of new concepts In *J Chem Inf Comput Sci 1986*
[3] G. Benkö and C. Flamm. A Graph-Based Toy Model of Chemistry In *J Chem Inf Comput Sci 2003*

Funding

This work was supported by WWTF - Wiener Wissenschafts-, Forschungs-, und Technologiefonds.