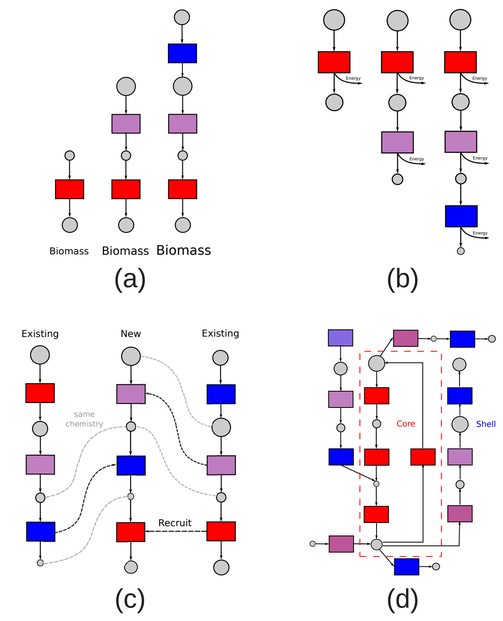
In Silico Evolution of early Metabolism.

Alexander Ullrich, Christoph Flamm,
Markus Rohrschneider, Peter F. Stadler

Abstract

We developed a simulation tool for the evolution of early metabolism, allowing us to speculate on the formation of metabolic pathways from catalyzed chemical reactions and network properties. Our model consists of a protocellular entity with a simple RNA-based genetic system and an evolving metabolism of catalytically active ribozymes that manipulate a rich underlying chemistry. Ensuring an almost open-ended and fairly realistic simulation is crucial for understanding the first steps in metabolic evolution. We show here how our simulation tool can be helpful in arguing for or against hypotheses on the evolution of metabolic pathways. We demonstrate that seemingly mutually exclusive hypotheses may well be compatible when we take into account that different processes dominate different phases in the evolution of a metabolic system. Our results suggest that forward evolution shapes metabolic network in the very early steps of evolution. In later and more complex stages, enzyme recruitment supersedes forward evolution, keeping a core set of pathways from the early phase.

Scenarios of Evolution



Hypotheses about the formation and evolution of metabolic pathways [1]. (a) Backward evolution, (b) Forward evolution, (c) Patchwork model, (d) Shell hypothesis. Colored squares represent enzymes, gray circles are metabolites. Color encoding for enzymes stand for their age, red being older and blue being younger enzymes.

Backward Evolution

Assumes that an organism is able to make use of certain molecules from the environment. Individuals that can produce these beneficial molecules by themselves gain an advantage in selection in the case of depletion of the "food source". Therefore, new chemical reactions are added that produce beneficial molecules from precursors that are abundant in the environment or that are produced in turn by the organism's metabolism. As a consequence, one should observe more ancient enzymes downstream in present-day metabolic pathways. Towards the entry point of the pathway, younger and younger going enzymes should be found.

Forward Evolution

Forward evolution could be seen as an extension or counterpart of the backward evolution hypothesis, reversing the direction of pathway evolution. Granick, and later Cordon, argue for a pathway evolution in forward direction, requiring that the intermediates are already beneficial to the organism. This is in particular plausible for catabolic pathways, where the organism can extract more energy by breaking food molecules downs to simpler and simpler end products. Older enzymes are then expected to be upstream in the pathway, with younger enzymes appearing further downstream.

Patchwork Model

The patchwork model explains the formation of pathways by recruiting enzymes from existing pathways. The recruited enzymes may change their reaction chemistry and metabolic function in the new pathways and specialize later trough evolution. This introduction of new catalytic activities lead to a selective advantage. Looking at the constitution of a pathway formed by enzyme recruitment, we should observe a mosaic-like picture of older and younger enzymes mixed throughout the pathway.

Shell Hypothesis

The shell hypothesis was proposed by Morowitz. It argues for the case of the reductive citric acid cycle that in the beginning an auto-catalytic core is formed from which new catalytic activities and pathways could be recruited and fed. Thus a metabolic shell would form around this core. Enzymes in the core would likely be less prone to mutational changes because they are essential for the organism. Thus, one should still be able to observe a core of ancient enzymes.

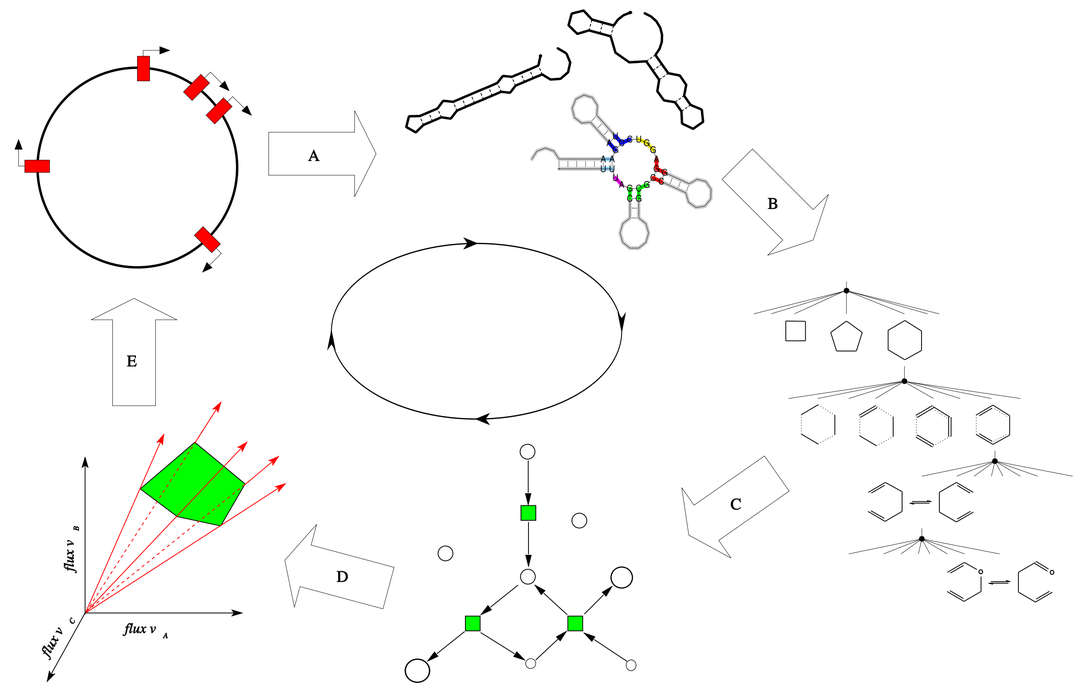
Computational Model

The computational model ([2]) is composed of a genetic and a metabolic subsystem. The genetic subsystem is implemented as a cyclic RNA genome. A special sequence motif indicates the start of genes which are of constant length. The RNA sequence corresponding to the "coding sequence" of a gene is folded into the (secondary) structure using the `Vienna RNA Package`(Step A).

During chemical reactions, bond formation/breaking is confined to a small subset of atoms of the reacting molecules. A cyclic graph abstraction, called the imaginary transition state (ITS) [3], can be used to capture the changes in the reactive center. Furthermore, over 90% of all known organic reactions can be classified by their ITS and organized in a hierarchical structure [3]. Sequence and structure features of the folded RNA gene products are mapped into the classification tree of organic reactions for functional assignment of the catalytic set (Step B). Thus we have implemented an evolvable sequence-to-function map, allowing the metabolic organization to escape from the confines of the chemical space set by the initial conditions of the simulation.

The metabolic subsystem is built upon a graph-based artificial chemistry endowed with a built-in thermodynamics. To generate the metabolic reaction network, induced by the catalytic set on the set of metabolites, a rule-based stochastic simulation is performed. Reaction rates are calculated "on the fly" from the chemical graphs of the reactants.

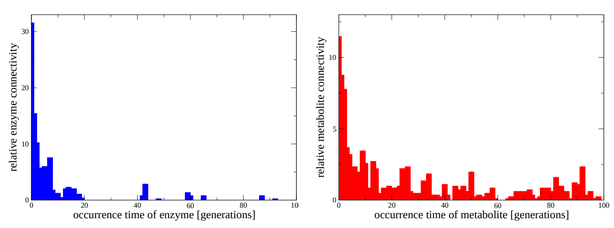
To identify the elementary flux modes, i.e., extreme pathways, of the resulting reaction network, a metabolic flux analysis is performed. (Step D). The fitness of an organism is computed as the maximum of the yield function (e.g. biomass production) over all extreme pathways. Finally, genetic variation operators are applied to the genome (Step E).



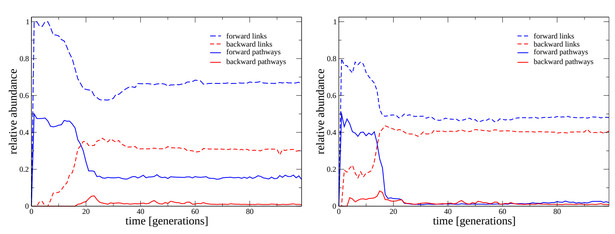
Scheme of the simulation system. (A) Decoding of (RNA-)genes to catalytic molecules; (B) Assignment of catalytic functions to "ribozymes", through mapping from structural and sequential information of the RNA molecule to a reaction logo in the hierarchy [3]; (C) Construction and stochastic simulation of the metabolic network; (D) Metabolic Flux analysis and fitness evaluation; (E) Application of genetic variation operators.

Results

For this study, we performed 100 runs with a population size of 100 cells running for 100 generations and performing 100 network expansion (stochastic simulation) steps per generation, the input molecules were cyclobutadiene, ethenol, phthalic anhydride, methylbutadiene, and cyclohexa-1,3-diene.

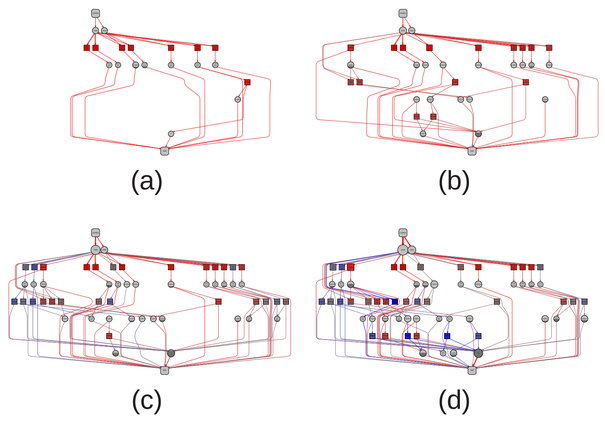


Average relative connectivity of (left) enzymes and (right) metabolites introduced in the same generation, for 100 generations. The height of the bars shows the fraction of the overall connections that are accounted by enzymes/metabolites from a particular generation. All values are averages over 100 simulation runs. Input molecules are not considered in the statistic, they account for nearly 50 percent of metabolite connectivity.

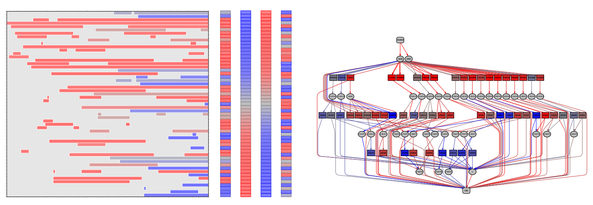


Evolutionary history of simulated metabolic networks. For the first 100 generations, we show the number of links and pathways that conform to the forward and backward evolution scenarios, respectively. Links are pairs of (left) consecutive reactions or (right) consecutive metabolites along a pathway. A pathway is identified as "forward-evolved" if at least one of its links is forward and none backward. In the first generations, the network consists predominantly of forward (reaction) links and pathways. After about 20 generations, the relative abundance of forward pathways decreases drastically but quickly reaches a persistent plateau value.

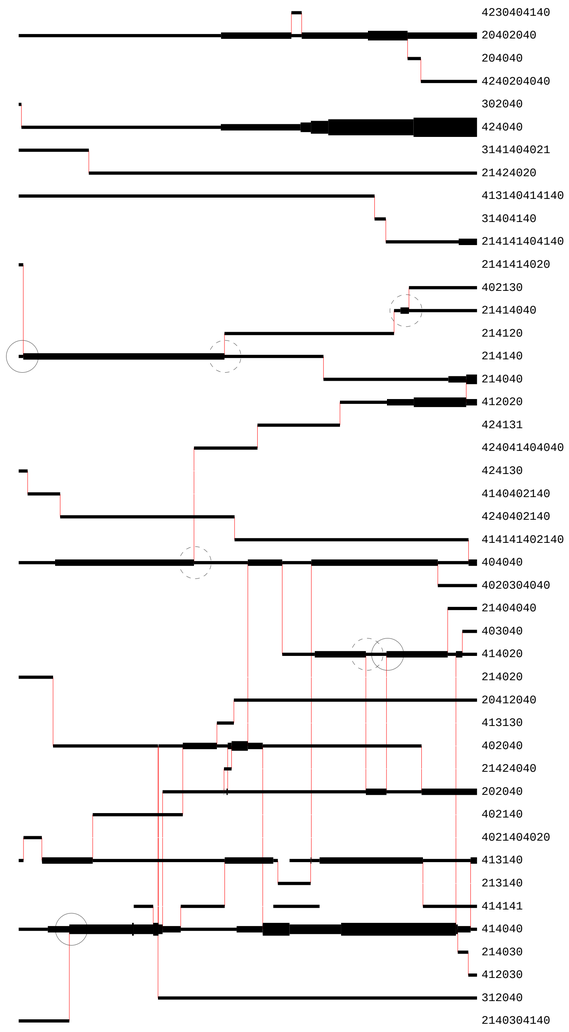
We use data from a simple simulation run, starting with only two input molecules and developing only few enzymes, for the visualization of an evolutionary time series and the reaction lifetime overview. Another, more complex, setting is used in a simulation run in which we investigate the evolutionary history of the involved genes/enzymes, depicted in the catalytic function genealogy for all generations.



A series of simulated metabolic networks after (a) 10, (b) 30, (c) 66, and (d) 100 generations. Colored squares represent chemical reactions, gray circles represent metabolites. Metabolites involved in a reaction are connected to it in the network graph. The coloring for the reactions encode their age, where red stands for older (occurrence in early generation) and blue for newer (later generation) reactions.



Life-time diagram for metabolic reactions. (left) Life-time of reactions, (right) union network graph over all 100 generations. The reactions in the life-time diagram are positioned corresponding to their position in the union network graph. The rows have colored entries if the corresponding reaction was present at a certain generation (columns 1-100). Older reactions are red, newer blue. The colored bars show the age distribution of reactions in the network in the same order as in the lifetime overview. The first bar represents our results, following the pattern for backward evolution, forward evolution and the patchwork model.



Genealogy of catalytic functions and gene dosage over 2000 generations. Each row represents an observed catalytic function. Black horizontal lines indicate time intervals in which genes coding for that catalytic function were present in the genome (0-200: from left to right). The thickness of the black lines indicates the number genes with a given function. Thin vertical red lines indicate points where the accumulation of mutations caused a transitions between catalytic functions. If the number of genes copies in a function class increases without a transition from another gene, then the increase is due to a gene duplication. A new gene can be created in the genome through the fortuitous formation of a TATA-box. Conversely, a gene can vanish if its TATA-box is destroyed by mutation. On the left of the chart a numerical encoding of the graph transformations performed by the "enzyme" is plotted.

Conclusion

We have introduced a simulation tool that models the early evolution of metabolism in a quite realistic setting and provides many tools for the detailed investigation of metabolic evolution. Using both simple example and a series of more complex simulation runs, the evolution of the components on the small scale (metabolites, enzymes) as well as on systems (pathways, networks) was investigated. The simulations allow to discriminate between different scenarios for the evolution of metabolic pathways. Based on the observations from this study, we argue that the different evolutionary hypotheses can be reconciled, in that they act in different phases of evolution, i.e. in different scenarios we might observe another strategy at work. Here, we suggest that forward evolution dominates in the earliest steps and is then superseded by a phase of enzyme recruitment, however, leaving behind a trace in form of a core set of forward evolved pathways.

Albeit our simulation environment is still a drastic simplification of chemistry, it is realistic enough to investigate the evolution of early metabolism. Computer simulations like this one are likely to provide new insights about the general evolutionary mechanisms governing biological systems in particular in regimes that are not readily observable. Our approach of a realistic, yet computationally feasible, model appears to be a promising step in this direction.

We find that enzymes from earlier stages are significantly higher connected than those from later stages. Further, we observe forward evolution in the first period succeeded by patchwork evolution, however keeping a core set of pathways from the early steps.

In an early and simple environment forward evolution is the main evolutionary force in our simulations. In longer simulations we observe all kinds of evolutionary events, such as divergence, convergence, multiplication and deletion.

References

[1] G. Caetano-Anolles et. al. The origin and evolution of modern metabolism. Inter J Biochem and Cell Biol (2009).
[2] Flamm et. al. Evolution of metabolic networks: A computational framework. J. Syst. Chem. (2010)
[3] J.B. Hendrickson. Comprehensive system for classification and nomenclature of organic reactions. J Chem Inf ComputSci (1997)