Computational Studies on the Evolution of Metabolism

Alexander Ullrich

PhD Thesis

Abstract

Living organisms throughout evolution have developed desired properties, such as the ability of maintaining functionality despite changes in the environment or their inner structure, the formation of functional modules, from metabolic pathways to organs, and most essentially the capacity to adapt and evolve in a process called natural selection. It can be observed in the metabolic networks of modern organisms that many key pathways such as the citric acid cycle, glycolysis, or the biosynthesis of most amino acids are common to all of them.

Understanding the evolutionary mechanisms behind this development of complex biological systems is an intriguing and important task of current research in biology as well as artificial life. Several competing hypotheses for the formation of metabolic pathways and the mechanisms that shape metabolic networks have been discussed in the literature, each of which finds support from comparative analysis of extant genomes. However, while being powerful tools for the investigation of metabolic evolution, these traditional methods do not allow to look back in evolution far enough to the time when metabolism had to emerge and evolve to the form we can observe today. To this end, simulation studies have been introduced to discover the principles of metabolic evolution and the sources for the emergence of metabolism properties. These approaches differ considerably in the realism and explicitness of the underlying models. A difficult trade-off between realism and computational feasibility has to be made and further modeling decisions on many scales have to be taken into account, requiring the combination of knowledge from different fields such as chemistry, physics, biology and last but not least also computer science.

In this thesis, a novel computational model for the in silico evolution of early metabolism is introduced. It comprises all the components on different scales to resemble a situation of evolving metabolic protocells in an RNA-world. Therefore, the model contains a minimal RNA-based genetics and an evolving metabolism of catalytic ribozymes that manipulate a rich underlying chemistry. To allow the metabolic organization to escape from the confines of the chemical space set by the initial conditions of the simulation and in general an openended evolution, an evolvable sequence-to-function map is used. At the heart of the metabolic subsystem is a graph-based artificial chemistry equipped with a built-in thermodynamics. The generation of the metabolic reaction network is realized as a rule-based stochastic simulation. The necessary reaction rates are calculated from the chemical graphs of the reactants on the fly. The selection procedure among the population of protocells is based on the optimal metabolic yield of the protocells, which is computed using flux balance analysis.

The introduced computational model allows for profound investigations of the evolution of early metabolism and the underlying evolutionary mechanisms. One application in this thesis is the study of the formation of metabolic pathways. Therefore, four established hypotheses, namely the backwards evolution, forward evolution, patchwork evolution and the shell hypothesis, are discussed within the realms of this in silico evolution study. The metabolic pathways of the

networks, evolved in various simulation runs, are determined and analyzed in terms of their evolutionary direction. The simulation results suggest that the seemingly mutually exclusive hypotheses may well be compatible when considering that different processes dominate different phases in the evolution of a metabolic system. Further, it is found that forward evolution shapes the 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. Backward evolution can only be observed under conditions of steady environmental change. Additionally, evolutionary history of enzymes and metabolites were studied on the network level as well as for single instances, showing a great variety of evolutionary mechanisms at work.

The second major focus of the in silico evolutionary study is the emergence of complex system properties, such as robustness and modularity. To this end several techniques to analyze the metabolic systems were used. The measures for complex properties stem from the fields of graph theory, steady state analysis and neutral network theory. Some are used in general network analysis and others were developed specifically for the purpose introduced in this work. To discover potential sources for the emergence of system properties, three different evolutionary scenarios were tested and compared. The first two scenarios are the same as for the first part of the investigation, one scenario of evolution under static conditions and one incorporating a steady change in the set of "food" molecules. A third scenario was added that also simulates a static evolution but with an increased mutation rate and regular events of horizontal gene transfer between protocells of the population. The comparison of all three scenarios with real world metabolic networks shows a significant similarity in structure and properties. Among the three scenarios, the two static evolutions yield the most robust metabolic networks, however, the networks evolved under environmental change exhibit their own strategy to a robustness more suited to their conditions. As expected from theory, horizontal gene transfer and changes in the environment seem to produce higher degrees of modularity in metabolism. Both scenarios develop rather different kinds of modularity, while horizontal gene transfer provides for more isolated modules, the modules of the second scenario are far more interconnected.