Robustness, Flexibility and Modularity in Complex Biological Systems.

Alexander Ullrich

Abstract

In the course of evolution, biological organisms have developed certain desirable properties such as the robustness against metabolic fluctuations or mutational errors, as well as the ability to switch flexibly between functional modules. Knowledge about the emergence of these properties is beneficial both for understanding the underlying evolutionary mechanisms as well as for developing principles for the construction of artificial systems. We investigate the formation of complex biological systems and the emergence of their properties using a multilevel computational model simulating the evolution of an early metabolism. Different evolutionary scenarios are simulated and compared regarding several measures for robustness, flexibility and modularity.

Complex Properties

Robustness is the ability of a system to adapt to changes in the environment or in itself. Therefore it is distinguished between robustness against genetic changes such as mutation, enzyme knockout and robustness against epigenetic changes or noise like fluctuations in the metabolite concentrations. Complex biological systems and scale-free architectures in general are known to be particularly robust. The power-law distribution of the node degrees is one explanation for the robustness against knockouts. Modularity in a system means that it is composed of several subsystems (modules) exhibiting distinct functions and containing further subsystems or elementary components. Most biological modules are modular and metabolic network products are found to be organized in an hierarchy of modules. The organization and preservation of modularity is not perfectly known, but changing environment or goals and horizontal gene transfer are suggested sources. A high average clustering coefficient and a power-law scaling of the clustering coefficient against the node degree are indicators for modular systems.

Connectivity Distribution

The connectivity distribution of metabolic networks follows a power law (-2). For comparison the connectivity distribution of the KEGG database we find that -1 is close to two. For the networks evolved under static conditions (also those with HGT) we find a similar distribution. The networks from simulations with changing environment we also find a scale-free distribution but with closer to one.

Clustering Coefficient

In modular networks the scaling of the clustering coefficient against the node degree depends on a power law. For the networks from the simulations with changing environment the simulations deviate in different parts from this scaling. In the static simulation the hub metabolites have a lower clustering coefficient compared to the simulations with changing environment. This could indicate that modules in these networks are less connected among each other.

Pathway Analysis

Metabolic pathway analysis is the calculation and analysis of the pathway distribution of a steady-state metabolic network to gain insights about its structure, functionality and properties. The calculation starts with the formation of the stoichiometric matrix presentation of the network and delivers the extreme pathways, spanning the entire steady state flux space, as the final result. Bounding the solution space through flow restrictions allows the computation of the optimal yield using a linear optimization process, this calculation is called flux balance analysis (FBA).

Computational Model

The computational model is composed of a metabolic network and a stochastic subsystem. The genetic system is implemented as a cyclic RNA genome. 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 target (Step B). 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 graph of the reactants. To identify the elementary flux modules, 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).

Summary

Three simulation scenarios were performed and analyzed to find potential causes for the emergence of complex properties such as robustness, flexibility and modularity. The first scenario assumes a static behavior, with steady environment and few mutations. The second scenario, in contrast, goes through a permanent change of environment. The third scenario works under a steady environment but an increased mutation rate and horizontal gene transfer (HGT) between individuals. The second measure indicates an increased modularity in the systems evolved in a static environment. Modularity seems only abundant in the scenarios with changing environment and horizontal gene transfer.

Ongoing Work

We follow two approaches to find signs of modularity by analysing the set of elementary pathways. In the first one we look at all possible combinations of these pathways and compute their optimal outcome with FBA. This information can be used like an energy landscape, from which we can in turn determine an energy barrier tree. An energy barrier between two states (combinations) tells us the metabolic cost of the cell to switch from one state to the other.

Chemical Organizations

A chemical organization is a closed and self-maintaining set of compounds, here metabolites and enzymes. It represents a steady-state of the cell in which it can produce all of the used components by itself. The organizations of a system are related to each other and can be represented in a hierarchy. The constitution of this hierarchy and the organizations on each level can give insights about the robustness and modularity of the system.

Connectivity Distribution

The connectivity distribution of metabolic networks follows a power law (-2). For comparison the connectivity distribution of the KEGG database we find that -1 is close to two. For the networks evolved under static conditions (also those with HGT) we find a similar distribution. The networks from simulations with changing environment we also find a scale-free distribution but with closer to one.

Clustering Coefficient

In modular networks the scaling of the clustering coefficient against the node degree depends on a power law. For the networks from the simulations with changing environment the simulations deviate in different parts from this scaling. In the static simulation the hub metabolites have a lower clustering coefficient compared to the simulations with changing environment. This could indicate that modules in these networks are less connected among each other.

Pathway Analysis

Metabolic pathway analysis is the calculation and analysis of the pathway distribution of a steady-state metabolic network to gain insights about its structure, functionality and properties. The calculation starts with the formation of the stoichiometric matrix presentation of the network and delivers the extreme pathways, spanning the entire steady state flux space, as the final result. Bounding the solution space through flow restrictions allows the computation of the optimal yield using a linear optimization process, this process is called flux balance analysis (FBA).

Computational Model

The computational model is composed of a metabolic network and a stochastic subsystem. The genetic system is implemented as a cyclic RNA genome. 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 target (Step B). 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 graph of the reactants. To identify the elementary flux modules, 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).

Summary

Three simulation scenarios were performed and analyzed to find potential causes for the emergence of complex properties such as robustness, flexibility and modularity. The first scenario assumes a static behavior, with steady environment and few mutations. The second scenario, in contrast, goes through a permanent change of environment. The third scenario works under a steady environment but an increased mutation rate and horizontal gene transfer (HGT) between individuals. The second measure indicates an increased modularity in the systems evolved in a static environment. Modularity seems only abundant in the scenarios with changing environment and horizontal gene transfer.

Ongoing Work

We follow two approaches to find signs of modularity by analysing the set of elementary pathways. In the first one we look at all possible combinations of these pathways and compute their optimal outcome with FBA. This information can be used like an energy landscape, from which we can in turn determine an energy barrier tree. An energy barrier between two states (combinations) tells us the metabolic cost of the cell to switch from one state to the other.

Chemical Organizations

A chemical organization is a closed and self-maintaining set of compounds, here metabolites and enzymes. It represents a steady-state of the cell in which it can produce all of the used components by itself. The organizations of a system are related to each other and can be represented in a hierarchy. The constitution of this hierarchy and the organizations on each level can give insights about the robustness and modularity of the system.

References