Robustness, Flexibility and Modularity in Complex Biological Systems.

Alexander Ullrich

IZBI UNIVERSITÄT LEIPZIG

Abstract

In the course of evolution, biological organisms have developed certain desirable properties such as the robustness against metabolite 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. We use well known graph measures as well as specifically developed flux- and steady-state-based measures based on Minimal knockout sets and chemical organizations.

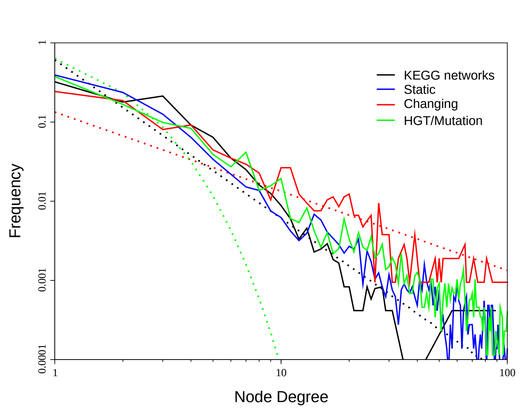
Complex Properties

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 systems are modular and metabolic networks are found to be organized in an hierarchy of modules [1]. The origin 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

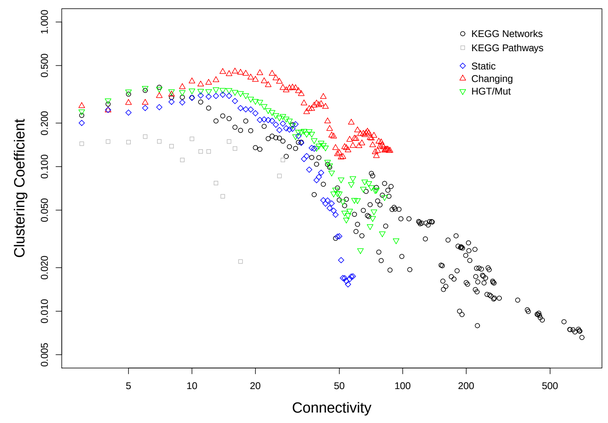
The connectivity distribution of metabolic networks follows a power law ($k^{-\gamma}$). For the networks from the KEGG database we find that γ is close to two. For the networks evolved under static conditions (also those with HGT) we find a similar distribution. The networks from simulations with changing environment we also find a scale-free distribution but with γ closer to one.



Connectivity distribution of metabolites from KEGG networks (black) and three simulation scenarios, static conditions (blue), changing conditions (red), HGT (green). For comparison the dotted lines show, exponential distribution (green) and power law distributions $\gamma = 1$ (red), $\gamma = 2$ (black).

Clustering Coefficient

In modular networks the scaling of the clustering coefficient against the node degree follows a power law with γ close to one. The KEGG networks show such a scaling, whereas the simulated networks deviate in different parts from that scaling. In the static simulation the hub-metabolites have a low clustering coefficient compared to the simulations with changing environment. This could indicate that modules in these networks are less connected among each other.

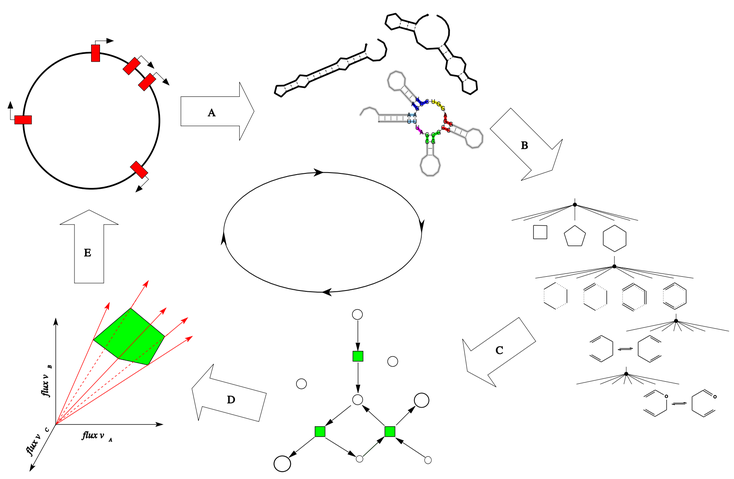


Clustering Coefficient of metabolites against the node degree. Investigated networks are network maps (black circles) and pathways (grey boxes) from the KEGG database, as well as networks from three different simulations, static behavior (blue diamonds), changing environment (red triangle) and HGT (green triangle).

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 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 shown measures indicate an higher mutational robustness of systems evolved in a static environment. Modularity seems only abundant in the scenarios with changing environment and horizontal gene transfer.

Computational Model



Scheme of the simulation system. (A) Decoding of (RNA-)genes to catalytic molecules; (B) Assignment of catalytic functions to "ribozymes"; (C) Construction and stochastic simulation of the metabolic network; (D) Metabolic Flux analysis and fitness evaluation; (E) Application of genetic variation operators.

The computational model [2] is composed of a genetic and a metabolic subsystem. The genetic subsystem 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).

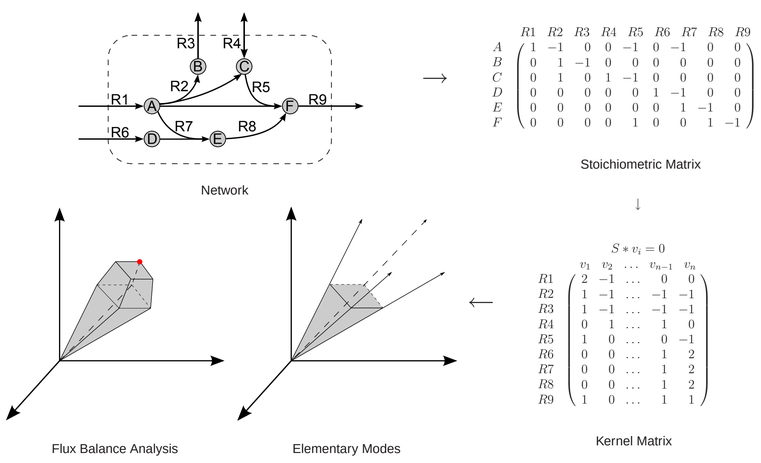
During chemical reactions, bond formation/breaking is confined to a small subset of atoms of the reacting molecules. A cyclic graph abstraction, called the imaginary transition state (ITS) [3], can be used to capture the changes in the reactive center. Furthermore, over 90% of all known organic reactions 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 set (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 graphs of the reactants.

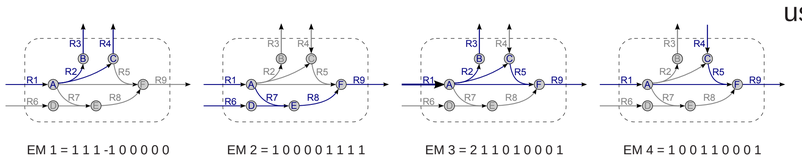
To identify the elementary flux modes, i.e., extreme pathways, of the resulting reaction network, a metabolic flux analysis is performed. (Step D). The fitness of an organism is computed as the maximum of the yield function (e.g. biomass production) over all extreme pathways. Finally, genetic variation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

Pathway Analysis

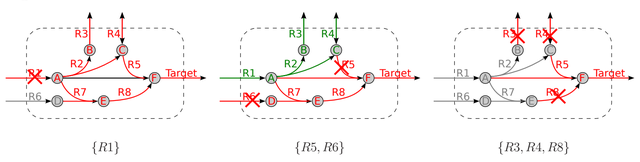
Metabolic pathway analysis is the calculation and analysis of the pathway distribution of a steady-state metabolic network to gain insights about its structure, functionality and properties. The calculation starts with the formation of the stoichiometric matrix presentation of the network and delivers 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, this process is called flux balance analysis (FBA).



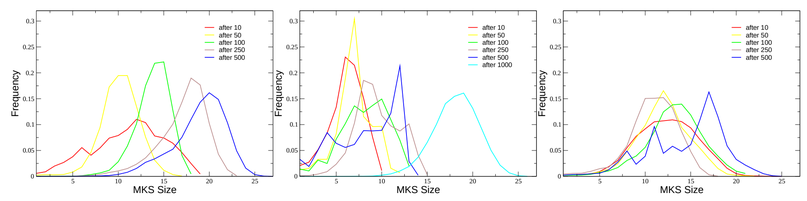
Extreme Pathways are the set of essential pathways through which all other possible pathways of the metabolic network can be generated, they are also minimal in the sense that they do not consist of smaller pathways.



Minimal knockout sets are sets of reactions that need to be removed in order to disable the function of a certain target reaction, this means that there may not be any extreme pathway containing this target reaction.



The size distribution of the minimal knockout sets are a good indication for the robustness of a network. If most sets are small then a few knockouts can block a vital function of the system. However, large sets imply a lower likelihood for blocked function since there would have to be many knockouts at the same time.

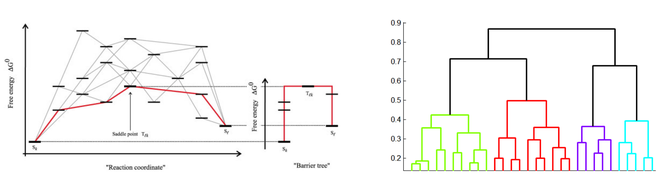


Minimal knockout set (MKS) size distribution for the three simulation scenarios, over several time steps (after 10, 50, 100, 250, 500 and 1000 generations). Static behavior (left), Changing Environment (middle) and increased mutation rate with horizontal gene transfer (right).

One can see that in all scenarios the robustness increases in the course of evolution. In the static scenarios, the increase is faster and stronger than in the scenario with changing environment. This underlines the results from the connectivity distribution that the robustness towards knockouts/mutations is decreased by additional perturbation via environmental change. The abundance of several peaks in the last time step of the HGT scenario might indicate different modules.

Ongoing Work

We follow two approaches two 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 Flux balance analysis. 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.

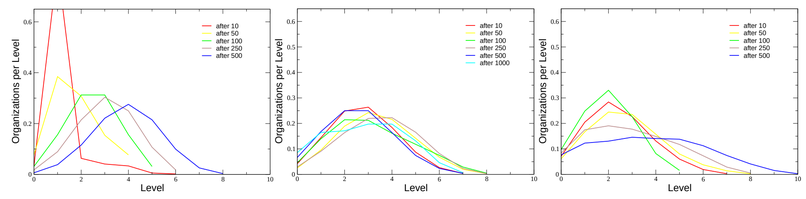


Left: Schema for a barrier tree, with an example for n energy barrier. States (black bars) represent combinations of active elementary modes. Right: Schema for a Hierarchical Clustering, leaves correspond to elementary modes.

If there are cheap connections between all states, the system is highly flexible. However, if we find high barriers between sets of combinations, the system is highly modular. The second approach is to define a similarity measure among pathways and then cluster all elementary pathways. Different similarity measures could indicate different types of modules.

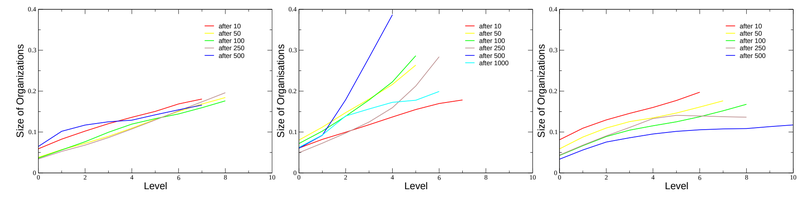
Chemical Organizations

A chemical organization is a closed and self-maintaining set of components, 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.



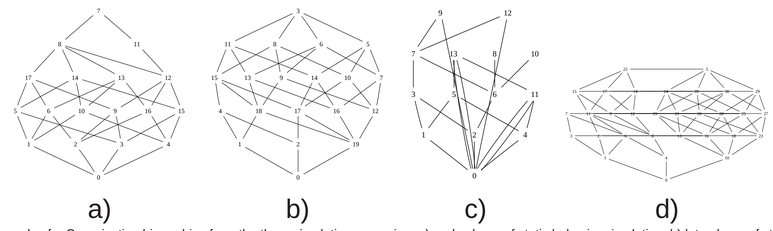
The distribution of Organizations per hierarchy level in relative frequencies for the three simulation scenarios and over several time steps (after 10, 50, 100, 250, 500 and 1000 generations). Static behavior (left), Changing Environment (middle) and increased mutation rate with horizontal gene transfer (right).

More organizations in the higher levels mean more alternative states which implies higher robustness. In the static simulations the trend goes to a more robust situation. The simulation with changing environment has more organizations close to the smallest organization making it more evolvable/flexible.



The distribution of Organization Sizes (relative size to the full organization) on an hierarchy level for the three simulation scenarios and over several time steps (after 10, 50, 100, 250, 500 and 1000 generations). Static behavior (left), Changing Environment (middle) and increased mutation rate with horizontal gene transfer (right).

A flatter slope means smaller differences between organizations of different levels, thus making the transition smoother. The high size differences in later stages of the changing environment scenario could also speak for abundance of modules.



Examples for Organization hierarchies from the three simulation scenarios. a) early phase of static behavior simulation, b) late phase of static behavior simulation, c) changing environment, d) increased mutation rate with horizontal gene transfer.

References

[1] E. Ravasz et.al. Hierarchical Organization of Modularity in Metabolic Networks. Science, Vol. 297-5586, pp. 1551-1555.
[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)