Using the RNA sequence-to-structure map for functional evolution of ribozyme catalyzed artificial metabolisms

Alexander Ullrich, Christoph Flamm and Lukas Endler

University of Vienna
aullrich@tbi.univie.ac.at

We introduce a novel genotype-phenotype mapping based on the relation between RNA sequence and its secondary structure for the use in evolutionary studies. The inspiration for this particular mapping emerged from the modeling of RNA enzymes within a simulation framework for the evolution of metabolic reaction networks. In our simulation we allow individuals, containing a genome and a metabolism, to evolve. The genome contains a number of RNA genes which then give rise to RNA enzymes acting on metabolites and thus shaping the metabolic network. Individuals are selected based on measures of this network and new individuals with mutated genomes are created. The use of our mapping allows not only for a more realistic study of the evolution of the entire system, but also enables us to observe the behavior of our enzymes itself and therefore possibly gain some insights about the evolution of catalytic molecules in general.

Enzymes typically have an active site where only few amino acids or bases determine its catalytic function and the remaining structure has mostly stabilization function. Accordingly, we extract structural and sequence information only from a restricted part of the fold. We decided to focus on the longest loop of the folded RNA. The idea for mapping the extracted information to a specific chemical reaction was encouraged by the fact that many enzymes catalyze a reaction by stabilizing its transition state. Recent work on hairpin ribozymes and other catalytic RNA support that as a common strategy for RNA enzymes. Given the definition of Fujita's imaginary transition structures (ITS), we developed a unique index for all possible pericyclic chemical reactions, describing the constitution of the reaction's transition state. Every RNA molecule is assigned such an reaction ID based on the information from its fold. The length of the longest loop specifies the number of involved atoms and the sequence within the loop determines the atom types. The bond types are derived from structural characteristics of the loop, such as the length and position of contained stems. Thus, a mapping from RNA sequence (genotype) to a chemical reaction (phenotype) is produced.

For many years it is known that neutral mutations have a considerable influence on the evolution in molecular systems. The folding of RNA sequences to secondary structures with its many-to-one property represents a mapping entailing considerable redundancy. Various extensive studies concerning RNA folding in the context of neutral theory yielded insights about properties of the structure space and the mapping itself. We intend to get a better understanding of some of these properties and especially of the evolution of RNA-molecules as well as their effect on the evolution of the entire molecular system.

Besides using the mapping in several simulation runs which yielded realistic metabolic networks and connectivities, we performed several statistical tests commonly used in neutral theory, such as the number of visited phenotypes and the average discovery rate during a random neutral walk. We compared it with results of approaches using cellular automatons, random boolean networks and other mappings based on RNA folding. It exceeds all non-RNA mappings in extent and connectivity of the underlying neutral network. Further, it has a significantly higher evolvability and innovation rate than the rest. Especially interesting is the highly innovative starting phase in RNA-based mappings.