A Sequence-to-Function Map for Ribozyme-catalyzed Metabolisms

Alexander Ullrich[1] and Christoph Flamm[2]

[1] University of Leipzig
Chair for Bioinformatics
Härtelstr. 16, 04275 Leipzig, Germany
[2] University of Vienna
Institute for Theoretical Chemistry
Währingerstr. 17, 1090 Vienna, Austria

Abstract. We introduce a novel genotype-phenotype mapping based on the relation between RNA sequence and its secondary structure for the use in evolutionary studies. 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. We investigate the constitution of the neutral network and compare our mapping with other artificial approaches using cellular automatons, random boolean networks and others also based on RNA folding. We yield the highest extent, connectivity and evolvability of the underlying neutral network. Further, we successfully apply the mapping in an existing model for the evolution of a ribozyme-catalyzed metabolism.

Key words: genotype-phenotype map, RNA secondary structure, neutral networks, evolution, robustness, evolvability

1 Introduction

For many years it is known that neutral mutations have a considerable influence on the evolution in molecular systems [14]. The RNA sequence-to-secondary map with its many-to-one property represents a system entailing considerable neutrality. It has been used successfully to shed light on the mechanisms of evolution [12, 13] and explain basic properties of biological systems, such as resolving the interplay between robustness and evolvability [19].

Several alternative scenarios have been proposed to describe the emergence and evolution of metabolic pathways (for a review see [4]) including spontanious evolution of enzymes with novel functions, the "backwards evolution scenario" suggesting that pathways evolve backward from key metabolites and the "enzyme recruitment scenario" or "patchwork model" which assumes that highly specialized an efficient enzymes evolved from a few inefficient ancestral enzymes
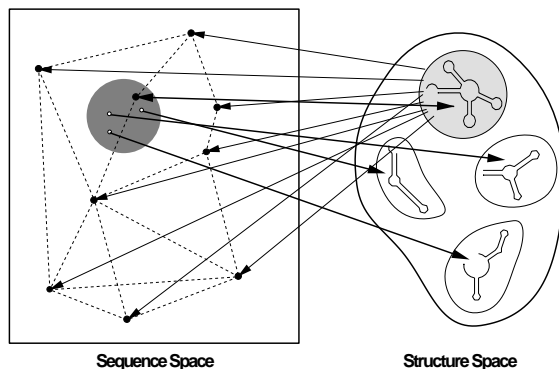
Fig. 1. The RNA sequence-to-structure map: There are many more sequences than structures which brings redundancy into the map. Sequences which fold into the same secondary structure form extended neutral networks in sequence space. The strong interweavement of the neutral networks implies that the sequences in a small volume around an arbitrary sequence realize all possible secondary structures.

with broad specificities. While these scenarios explain certain aspects of the early history of life and the evolution of metabolic pathways, the general mechanisms of the transition from an uncatalysed to a catalysed reaction network and the evolution of novel catalytic functions remains poorly understood.

We have proposed a multi-level computational model to study this transition in particular and the evolution of catalyzed reaction-networks in general [18]. The model combines a graph-based algebraic chemistry [1] with a RNA based structure-to-function-map within a protocellular entity. The structure-to-function-map projects sequence and structure features of the ribozyme into a hierarchical classification scheme for chemical reactions based on Fujita's "imaginary transition structure concept"[8]. The structure-to-function-map itself is evolvable and allows the emergence of completely new catalytic functions. Furthermore it forms the link between the genotype (i.e. the RNA sequence of the ribozymes) and the phenotype (i.e. the (partially) catalyzed reaction network).

2   The RNA Sequence-to-Structure Map

The folding of RNA sequences into their respective secondary structures [11] is a relatively well understood biophysically grounded process. Efficient algorithms [9] have been developed which allow the computation of nearly any desired thermodynamic or kinetic property of RNA sequences at the level of secondary structure. The relation between RNA sequences and their secondary structures can be viewed as a minimal model of a genotype-phenotype map (see Figure 1), where the RNA sequence carries the "heritable" information for the formation of the phenotype, which in turn is the RNA secondary structure upon which selection acts in evolutionary experiments [2, 3].
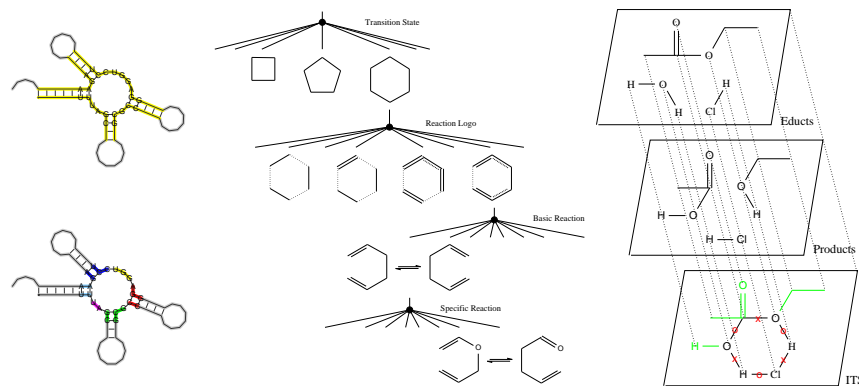
Fig. 2. The structure-to-function map: (lhs) The colored regions of the ribozyme fold determine the catalytic function i.e. which leaf in ITS-tree is picked; (middle) Along the levels of the ITS-tree the amount of chemical detail increases; (rhs) Acidic hydrolysis of ethyl acetat. The ITS of the reaction is gained from the superposition of educts and products by removing those atoms and bonds which do not directly participate in the reaction (shown in green).

The statistical architecture of the RNA sequence-to-structure map and it's implications for the evolutionary dynamics [6, 7] has been extensively studied over the past decade. In particular, the map possesses a high degree of neutrality, i.e. sequences which fold into the same secondary structure are organized into extended mutationally connected networks reaching through sequence space. A travel along such a "neutral network" leaves the structure unchanged while the sequence randomizes. The existence of neutral networks in sequence space has been demonstrated in a recent experiment [16]. Due to the fact, that the neutral networks are strongly interwoven, the sequence-to-structure map shows another interesting property called "shape space covering" [17]. Meaning that within a relatively small volume of sequence space around an arbitrary sequence any possible secondary structure is realized. Both features of the RNA sequence-to-structure map account for directionality and the partially punctuated nature of evolutionary change.

3   The Mapping

In the following, we elaborate the aspects of our novel genotype-phenotype mapping. First, the mapping is realized in two steps. While both steps have a biological motivation, they are also designed regarding statistical considerations. We will refer to the two steps as the sequence-to-structure and the structure-to-function map, respectively.

The sequence-to-structure map is equivalent to the folding of the RNA-sequence to its minimum free energy structure. We use the RNA sequence-to-structure map on the one hand to model ribozymes for our artificial metabolism (Sec-

tion 4) and on the other hand to add the properties described in Section 2 to the mapping. The second and more interesting part, the structure-to-function map, can vary in some aspects for different applications, since the resulting phenotype spaces may differ in size and structure as well as the constitution of the phenotype itself. Here we focus on the mapping for our artificial metabolism model (Section 4). Therefore, we map from the RNA secondary structure to a chemical reaction. The reactions are represented by their transition state structure. The classification of those structures is hierarchical (see Figure 2). Accordingly, we extract features for every level, to determine the resulting reaction. Thereby, we regard structural and sequential information only of the longest loop in the structure and adjacent stems. The length of the loop specifies the size of the transition state, i.e. the number of involved atoms. Further, length and starting base-pairs of the adjacent stems define the constitution of the basic reaction, i.e. the abundance and position of single, double or triple bonds. Finally, the sequence inside the loop region decides over the specific reaction, i.e. the position and type of the involved atoms. The decision to focus on the loop region is based on the observation that most catalytic RNA molecules have their active center in a hairpin loop and also generally enzymes have an active site that covers only a small part of the entire structure. Furthermore, some studies suggest that transition state stabilizing is an important catalytic strategy for ribozymes [15], matching our representation of chemical reactions as transition state structures.

4   Ribozyme-catalyzed Artificial Metabolism

In order to investigate the evolution of biological networks and the emergence of their properties, we have developed a sophisticated model of a ribozyme-catalyzed metabolism. The graph-based model is supported by an artificial chemistry, to allow for realistic kinetic behavior of the system. The basic elements of the model are represented as graphs. In case of the metabolites this complies with the intuitive presentation of molecules in chemistry. Similarly, metabolic reactions can be represented as a set of graphs or by one superimposition graph, combining the relevant parts of all involved molecules. The process of the reaction itself is performed by a graph-rewriting mechanism, through which new molecules can be generated. The conjunction of metabolites and their participation in metabolic reactions is captured in the metabolic network. We use several techniques to analyze the properties of the produced networks and compare them with analyzes of real-world networks [18]. Besides the metabolism, our model also comprises a genome that is divided into several genes by TATA-boxes. The genes are RNA-sequences of fixed length, intended to code for the catalytic elements, ribozymes, of the model system. The idea for the genotype-phenotype mapping that we introduce in this paper originates from the task to find a realistic and efficient mapping from these RNA-sequences (genotype) to the ribozymes (phenotype), or better, to the chemical reactions they implement.
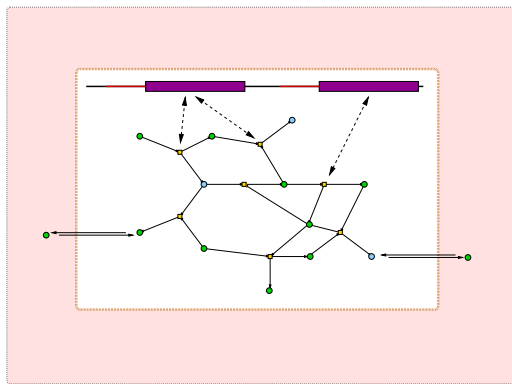
Fig. 3. Overview of the artificial metabolism model. Each gene of the genome expresses for a ribozyme. While these catalysts are trapped within the protocellular entities, metabolites under a certain size can pass through the membrane, allowing exchange with the environment.

5   Results

We will now compare our genotype-phenotype mapping with other existing approaches. Mappings based on cellular automaton (CA) and random boolean network (RBN) have been proven to exhibit desirable properties [5] and, thus, will serve here as reference. In previous evolutionary models and for the statistical analysis in this paper, we developed other mappings based on the RNA sequence-to-structure map but varying structure-to-function maps. Here we will mention only two of them. For the first mapping, we assign one target structure for each point in the phenotype space, the structure then maps to the phenotype of the closest target structure. The second mapping extracts structural features, similar to our mapping, but for the entire structure.

5.1   Random Neutral Walk

One way of gaining statistical results for the comparison of different genotype-phenotype mappings is through a random neutral walk on the genotype space. It starts from an arbitrary point of the mapping's genotype space. The first step is to choose randomly one neutral mutation out of the set of all possible one-point mutations. Applying the mutation, a new genotype is created that differs from its predecessor in only one position and maps to the same point in phenotype space. This procedure is repeated until the length of the walk reaches the pre-defined limit or no neutral mutation can be applied.

In each step, we keep track of several statistics that give insight into the structure of the neutral network. For this aim, we observe the one-point mutation neighborhoods of all genotypes of the random neutral walk. The fraction of neutral mutations within these neighborhoods will be referred to as the neutrality
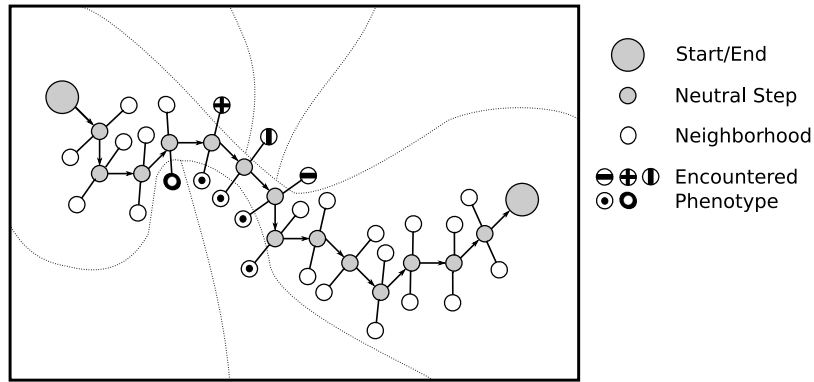
Fig. 4. Random neutral walk from an arbitrary point in genotype space to random neutral neighbors until a certain length is reached. In each step, encountered phenotypes in the one-point mutation neighborhood are protocolled.

of the underlying neutral network. In terms of biological systems this can also be regarded as the robustness that is gained through the mapping, since neutrality protects from harmful mutations. Further, we search for new phenotypes that are encountered during the course of the walk, i.e. the number of different phenotypes that were found in the one-point neighborhoods. This measurement indicates the rate with which the mapping discovers new phenotypes and other neutral components. The faster a system can access different points in the phenotype space, the more evolvable is it. Thus, the discovery rate can be equated with the notion of evolvability. Other measures that we regard as important here, are the extent and the connectivity of the neutral networks. Former, is given by the maximal distance between genotypes of the neutral walk. Latter, is defined by the fraction of possible phenotypes that ca be reached from an arbitrary starting configuration.

5.2   Comparison

For the following study, we performed 1000 random neutral walks of length 100 for each mapping. The size of the genotype spaces vary for the different mappings. For the RNA-based mappings the size is 4^{100}, for the random boolean network mapping it is 2^{144} and for the cellular automaton mapping 2^{76}. The phenotype space is $2^8 = 256$ for all mappings.

Figure 5 shows the results of the performed random walks. It can be seen that our mapping (RNA loop) reaches the most phenotypes ($\approx$ 200 of 256), following the other RNA-based mappings ($\approx$ 175 and 150), the random boolean network (145) and the cellular automaton (100). We can also see that the difference in the first steps is even more drastically, indicating a faster discovery rate of our mapping compared to all others. This can be explained by the connectivity, whereas CA and RBN have about 14 and 21 neighboring phenotypes,
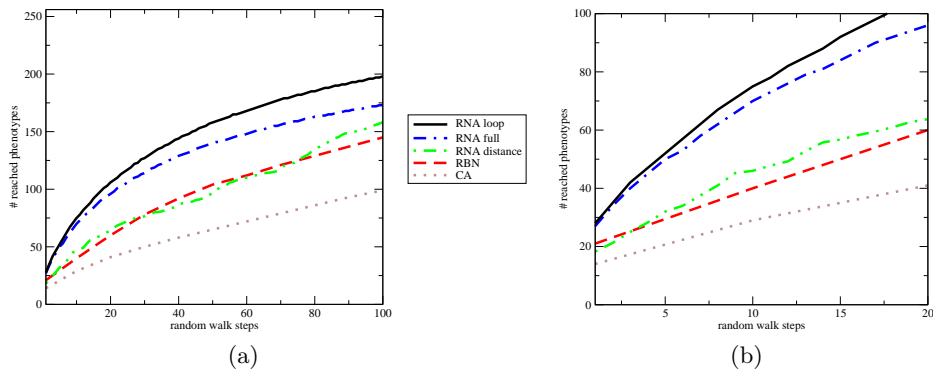
Fig. 5. Encountered phenotypes in a random neutral walk of length 100 (left) and for the first 20 steps (right). RNA loop = our mapping; RNA full = RNA-based mapping considering the entire structure; RNA distance = RNA-based mapping using target structures; RBN = random boolean network mapping; CA = cellular automaton mapping.

respectively, our mappings have about 27 distinct phenotypes in their one-point neighborhood. Furthermore, our mapping travels further in the genotype space, allowing a steady discovery of new phenotypes. However, in terms of neutrality the random boolean network mapping performs best, $\approx 58\%$ of its mutations are neutral, for CA it is $\approx 44\%$ and the RNA-based mappings are in between with around 50%.

5.3 Discussion

We have proposed here a genotype-phenotype mapping that is suitable for evolutionary studies, as shown for the given example and the comparative study with other mappings. Due to its advantageous properties, it is conceivable to use our mapping in studies where so far rather simple artificial genomes have been used and the RNA sequence-to-structure map can be integrated in a meaningful way, such as the evolution of regulatory networks [10]. First of all, our mapping allows open-ended evolution of the simulation model, since it continuously produces new phenotypes and does not get trapped in local optima. Further, it provides sufficient robustness to evolve a stable system and at the same time achieves high evolvabilty, allowing the system to react fast to perturbations or changes in the environment. We have also shown here, that a good genotype-phenotype mapping needs more than neutrality, the connectivity and extent of the underlying neutral network is at least as important.
Observing the mapping within the proposed evolutionary study of ribozyme-catalyzed metabolism, we hope not only to get insights about the emergence of pathways but also about the properties of the mapping itself, thus, the RNA sequence-to-structure map.

Acknowledgments

References

1. Benkö, G., Flamm, C., Stadler, P.F.: A Graph-Based Toy Model of Chemistry. J. Chem. Inf. Comput. Sci. 43, 1085-1093 (2003)
2. Bieberich, C.K.: Darwinian Selection of Self-Replicating RNA Molecules. In Evolutionary Biology, Plenum Publishing Corporation, volume 16, 1–52 (1983)
3. Bieberich, C.K., Gardiner, W.C.: Molecular Evolution of RNA *in vitro*. Biophys. Chem. 66, 179–192 (1997)
4. Caetano-Anollés, G., Yafremava, L.S., Gee, H., Caetano-Anollés, D., Kim, H.S., Mittenthal, J.E.: The origin and evolution of modern metabolism. Int. J. Biochem. & Cell Biol 41, 285–297 (2009)
5. Ebner, M., Shackleton, M., Shipman, R.: How neutral networks influence evolvability. Complex. 7, 19–33 (2001)
6. Fontana, W., Schuster, P.: Continuity in evolution: on the nature of transitions. Science 280, 1451-1455 (1998)
7. Fontana, W., Schuster, P.: Shaping Space: the Possible and the Attainable in RNA Genotype-phenotype Mapping. J. Theor. Biol. 194, 491-515 (1998)
8. Fujita, S.: Description of Organic Reactions Based on Imaginary Transition Structures. J. Chem. Inf. Comput. Sci. 26, 205–212 (1986)
9. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., Hofacker, I.L.: The Vienna RNA websuit. Nucl. Acids Res. 36, W70-W74 (2008)
10. Hallinan, J., Wiles, J.: Evolving genetic regulatory networks using an artificial genome. Proc. 2nd Conf. on Asia-Pacific Bioinformatics Vol. 29, 291-296 (2004)
11. Hofacker, I.L., Stadler, P.F.: RNA secondary structures. In Thomas Lengauer, editor, Bioinformatics: From Genomes to Therapies vol. 1, 439-489 (2007)
12. Huynen, M.A.: Exploring phenotype space through neutral evolution. J. Mol. Evol. 43, 165-169 (1996)
13. Huynen, M.A., Stadler, P.F., Fontana, W.: Smoothness within ruggedness: the role of neutrality in adaptation. Proc. Natl. Acad. Sci. USA 93, 397-401 (1996)
14. Kimura, M.: Evolutionary rate at the molecular level. Nature 217, 624-626 (1968)
15. Rupert, P.B., Massey, A.P., Sigurdsson, S.T., Ferre-D'Amare, A.R.: Transition State Stabilization by a Catalytic RNA. Science 298, 1421-1424 (2002)
16. Schultes, E.A., Bartel, D.P.: One sequence two ribozymes: Implications for the emergence of new ribozyme folds. Science 289, 448-452 (2000)
17. Schuster, P., Fontana, W., Stadler, P.F., Hofacker, I.L. From sequence to shape and back. Proc. Roy. Soc. (London) B 255, 279-284 (1994)
18. Ullrich, A., Flamm, C.: Functional Evolution of Ribozyme-Catalysed Metabolisms in a Graph-Based Toy-Universe. LNCS 5307, 0028 (2008)
19. A. Wagner, Robustness and evolvability: a paradox resolved Proceedings. Biological Sciences / The Royal Society 2758, 91–100 (2008)