A minimal and self-consistent *in silico* cell model based on macromolecular interactions

Christoph Flamm*[a,b], Lukas Endler[a], Stefan Müller[d], Stefanie Widder[a], and Peter Schuster[a,c]

[a] *Theoretical Biochemistry Group, Institut für Theoretische Chemie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*

[b] *Lehrstuhl für Bioinformatik, Institut für Informatik, Universität Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*

[c] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

[d] *Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria*

Summary: A self-consistent minimal cell model with a physically motivated schema for molecular interaction is introduced and described. The genetic and metabolic reaction network of the cell is modeled by multidimensional nonlinear ODEs, which are derived from biochemical kinetics. The strategy behind this modeling approach is to keep the model sufficiently simple in order to be able to perform studies on evolutionary optimization in populations of cells. At the same time the model should be complex enough to handle the basic features of genetic control of metabolism and coupling to environmental factors. Thereby the model system will provide insight into the mechanisms leading to important biological phenomena like homeostasis, (circadian) rhythms, robustness, and adaptation to a changing environment. One example of modeling a molecular regulation mechanism, cooperative binding of transcription factors, is discussed in detail.

Key Words: Biochemical ODE solver, cooperativity, genetic regulatory network, metabolic network, SBML

*Corresponding author, E-mail: xtof@tbi.univie.ac.at

Introduction

One of the great unsolved problems of biology is the intricate procedure that transforms heritable genetic information into the observable physical, chemical and biological characteristics of an organism subsumed in the phenotype. Unfolding a genotype in order to yield the corresponding phenotype is commonly a complex process that involves a great variety of molecular players. In case of multicellular organisms this process involves not only the formation of new cells by division but also cell differentiation and development, the next higher hierarchical level. Understanding the unfolding of a genotype, often subsumed under the notion of a "genotype-phenotype map", is further complicated by the fact, that the "interpretation" of the genotype by the molecular machinery of the cell is strongly influenced by environmental and epigenetic factors. At the present state of knowledge, a theoretical model which adequately describes the genotype-phenotype map at molecular level for a whole organism seems to be a hopeless task.

The simplest evolving entity, one can think of, comprises heritable information in a nucleotide sequence and a phenotype derived from it in one and the same object. The information stored in the sequence materializes as a self-replicating molecular species in the form of an RNA molecule. The structure of the RNA molecule is established by interactions of individual nucleotides joined by the polymer chain and determines the function of the molecule. In that sense the sequence carries the "heritable" information for the formation of the phenotype, which in turn is the molecular structure upon which selection acts in the evolution experiment [1, 2]. The folding process of RNA at the secondary structure level is accessible also to mathematical analysis and computer simulation, and can be viewed therefore as an abstract model of a genotype-phenotype map (The 'RNA model' is described in [3, 4]).

Intensive studies of the RNA sequence to structure map during the last decade revealed how the properties of this map influences the dynamics of evolutionary optimization. While being tremendously successful in elucidating the mechanisms governing molecular evolution [5–7], many concepts of biological genotype-phenotype maps, like signal transduction or developmental processes, have no concrete analogue [8] within the RNA model. The restriction of the phenotype to the structure of a single molecule makes it impossible to discuss aspects of organization since 'division of labor' is one of the key issues in biology that determines the properties of genotype-phenotype maps. In particular, the most striking deficiency is the absence of any form of control and regulation in the RNA model. It is indeed the regulatory network that builds the link between the genotype and the visual features of the phenotype [9, 10].

In this contribution we describe a model, which comes a step closer to biological genotype-phenotype maps while remaining sufficiently simple to allow for large scale evolutionary studies on the system.

The basic idea is to build a deterministic hierarchical mapping which in a sense *encrypts* a dynamical system representing the phenotype within a string (being the genotype). This approach separates the genotype, upon which the genetic variation operators act, from the phenotype which is under selection pressure. The dynamical system itself is a minimal version of a **gen**e regulatory and met**abolic** network[1] represented by a system of ordinary differential equations (ODEs).

Related Work

Our model follows the spirit of related work in the area of artificial regulatory network (ARN) models. Kauffman [11] used random boolean networks (RBN) to model gene regulatory networks. RBNs show a broad range of dynamical behavior from cyclic and multiple attractors to chaos. Most of this interesting dynamical features however vanish if the updating rule for the temporal evolution of the network's state is changed from a synchronous to a biologically more realistic, asynchronous one. Moreover, random boolean networks show only a limited ability to structurally represent genes and genomes.

Reil [12] introduced the concept of an artificial genome to overcome this structural weakness of RBNs. The artificial genome is essentially a biologically inspired representation of genes and their interactions. The model allows to manipulate the topology of the gene regulatory network at the level of the genome (implemented as string of digits) by a set of genetic variation operators which closely resemble their natural counterparts. This permits to study questions regarding the evolution of the ARN and it's quantized boolean dynamics from the point of view of the changing genome.

Delleart and Beer [13], Eggenberg [14] and later Bongard [15] embedded an ARN analogously to Reil's approach, into a hand-coded morphogenetic system to evolve "multi-cellular" objects capable of performing some predefined tasks. Bongard showed that within this framework commonly termed artificial embryogeny (AE) (see [16] for a recent review), gene reuse and modularity in terms of regulatory circuits can arise.

[1]Since gene regulation and metabolic control is intimately coupled in cellular dynamics we suggest to use the term *genabolic network* for the functional combination of genetics and metabolism.

Banzhaf [17] refrained from the boolean paradigm and expressed the dynamics of his ARN model in ordinary differential equations (ODEs). Many dynamical phenomena of natural gene regulatory networks i.e. point attractors, damped oscillations and heterochronic control [18] are reproduced by the ARN model. By introducing an arbitrary "virtual" binding site for a desired output function, networks could be evolved where the activation pattern of the virtual binding site follows a predefined mathematical function [19]. The model we propose differs from prior work with respect to the following points:

First, the competition of molecular species that bind to regulatory regions of genes is modeled explicitly in mass-action-governed elementary reactions. The reason for this decision lies in the facts that (i) competition for a common resource is obviously one of the core reactions in gene regulatory networks which warrants an accurate mechanistic description (ii) mechanistic details can have unexpected consequences [20] in terms of dynamic phenomena, especially if coarse-grained approximations i.e. Michaelis-Menten type kinetics or concentration weighted mean values are used.

Second, the genome and the gene products are modeled entirely in RNA molecules. At the level of RNA secondary structure efficient, well-established algorithms exist to compute nearly any desired molecular property. In particular the statistical properties of the sequence to structure map and its implications for the evolutionary process have been profoundly enlightened. Therefore operations on RNA molecules, as used in the presented model, possess a certain degree of physical realism which is lacking if binary or real-valued vectors are used.

Third, molecular interactions, another key feature of gene regulatory networks, are modeled within the framework of RNA secondary structures. This move provides us with a physically meaningful temperature dependent energy function which is not given for Hamming-distance based approaches for bit strings.

Finally, the model is equipped with a minimal version of a metabolism and a simple membrane similar to [21].

Model Description

A basic requirement for a model to be suitable for studying evolution is to be self-contained in the sense that it does not require input of parameters on the fly. This has been achieved in the RNA model by defining rules that provide the frame for the computation of the required parameters. The development of our model pursues the same strategy. In particular, the decoding step is done in such a way, that all the relevant parameters needed to compute the time evolution of the dynamical system are calculated from within the model. This enables the individual system to freely explore genotype space[2] by increasing its complexity without imposing limitations from the exterior. In molecular terms the genotype is thought to be a DNA or RNA molecule, which is transcribed in pieces to yield RNA that in turn is the source of various other molecules in the system. Two different scenarios are conceivable: (i) RNA is translated to yield protein molecules or processed to yield regulatory RNA molecules of the si-RNA type [22], or (ii) all molecular species of the gene regulatory and metabolic network are entirely represented as RNA molecules. There are no proteins involved in the second model and all regulatory and housekeeping functions are executed by ribozymes. Model (ii) is introduced here. It is based on the empirical evidence that natural occurring RNAs can fulfill a wide scope of different functions.

The Genotype-to-Phenotype Map

The genotype is represented by an RNA string of appropriate length. The genes which may overlap along the genome are structured as follows: The starting point of a gene is marked by a short sequence pattern which is reminiscent to the TATA-box of eukaryotic genes. Upstream of this sequence pattern two regulatory sites are located and in downstream direction a fixed length coding region follows (see figure 1).

Gene products fall into two major classes, (i) transcription factors (TF) and (ii) structural RNAs. While the former constitute the gene regulatory network, the latter fulfill metabolic tasks. The function of a given RNA molecule is determined by means of an energy-based comparison with classes of target structures: The sequence is folded into all target structures yielding a series of free energy values, and the lowest free energy value determines the function of the molecule. Since the probabilities of the predefined target structures obtained by folding random sequences is easily computable, the distribution of functions in the ensemble of gene products can be influenced by choosing target structures of different probabilities. This approach is supported by the fact that the function of naturally occurring RNAs is commonly determined by the structure and not by the sequence. A point mutation in a gene may or may not alter the function of the transcript depending on the degree of neutrality of the structure formed by the gene product. This procedure ensures a unique mapping (with exception of rare event that two folds

[2]Depending on the question to be addressed by the simulation the genotype space can be restricted to all (4^n) polynucleotide sequences of constant chain lengths n or the length may variable with insertion and deletion operators acting on genotypes.

of the sequence yield identical minimum free energies) and a tunable degree of redundancy. The problems arising from direct encoding of the phenotype by the genotype [23] are avoided thereby. Both, unique assignment and high degree of redundancy increase the accessibility of phenotypes leading to an increased evolvability of the population [24]. The latter effect is well known from simulation of RNA optimization where GC-only sequences are much harder to optimize by evolution than AUGC sequences because they have a lower degree of neutrality [25] (for a discussion on the parallels of the evolutionary search on neutral networks between RNA and genetic programming we refer the reader to [26]).

The Gene Regulation

Transcription factors (TFs) are grouped further into two types, activators and repressors. These molecules bind to one of the two binding sites in the upstream regulatory region (URS) of the gene. The activity of the gene is regulated by the fraction of repressors or activators bound to the URS. For modeling the transcriptional activity, we apply the common three state *regulated recruitment mechanism* observed with bacterial genes (See [27], pp.13-42): (i) Free genes are transcribed at a low basal rate, (ii) genes with activators bound to the regulatory region are transcribed at high rate, and (iii) genes with bound repressors are silenced. The transcription rate itself also depends on the concentration of active nucleotides, the analytic expressed of the dependence was adopted from the mechanism of RNA replication by the replicase of the phage $Q\beta$[28].

In order to regulate transcription the binding affinities of the whole ensemble of transcription factors (TFs) have to be transformed into a gene activity signal that falls into one of the three transcription states: basal, active, and silenced. Since the structure of the regulatory network itself should be a target of evolution, a model based on molecular interactions is required that decides upon two questions: (i) Which transcription factor (TF) binds to the distinct URS of the gene, and (ii) to what extend is the URS bound by the different TFs. Heteroduplex formation or RNA-RNA hybridization [29, 30] is used to quantify the binding strength of a given TF to a given regulatory site. The free binding energies computed in this way can be used directly to calculate dissociation constants, and under the assumption that association of the TF/URS complex is limited by diffusion, rate constants for complex formation and dissociation. Since computation of binding constants is straightforward, the topology of the genabolic network is readily recalculated, if some of the genes in the network changed sequence upon action of the genetic variation operators on the genome.

The definition of the model parameters provides an opportunity to design more complex regulatory mechanisms. As an example we present the consideration of cooperativity in TF binding. There are several possibilities to introduce cooperativity: (i) The effector molecule itself is a dimer, tetramer or even higher aggregate and oligomer formation is a cooperative process or (ii) alternatively two molecules bind to the binding sites within the URS and a stabilizing interaction between the two ligands leads to cooperative binding as sketched in figure 2. Whereas the first phenomenon leading to cooperativity is well known in biophysics, the direct cooperative interaction between two bound molecules has been studied much less frequently. For this reason we performed a large scale investigation with random RNA sequences. Two molecules were bound to the two URS binding sites and they are cofolded under the constraint that their structures bound to the URS sites are preserved. A negative free energy leads to an additional stabilization of the TF/URS complexes and results in cooperative binding (table 1). The results in the table present a 'proof of concept' of our model for cooperative interactions. A reasonable fraction of random sequences with sufficient GC content are suitable as regulators with cooperative interactions since they form stable aggregates. The fraction of stable complexes with only AU base pairs is very low. Pure AU sequences are characterized by relatively high free energies of cooperative binding but too low energies of interaction with the URS. In pure GC sequences we find reasonably strong binding to the URS and cooperativity of the same order of magnitude. Mixed sequences may be preferable since they have acceptable binding strength with still dominating cooperativity. The energetic differences between heterodimers and homodimers (self-cooperativity in table 1) are negligibly small.

The Metabolism

The metabolic RNAs are ribozymes which are capable of catalyzing either a chemical reaction that activates a mediator molecule or a reaction that transforms membrane precursor molecules into membrane building molecules (see figure 3). The active mediator, in turn, transfers energy to monomeric building blocks and converts it into active species, which can be directly used for transcription (or transcription and translation in case we use a model with mRNAs and proteins). Eventually we end up with biopolymers, RNAs (or RNAs and proteins), which may enhance their own production by producing more ribozyme in the manner of an auto-catalytic cycle. Catalysts are assumed to require specific predefined structural elements. The catalytic efficiency of a molecule in the catalysis of a metabolic reaction is derived from the activating energy $E_a = \varepsilon_n - \varepsilon_0$ that is required for the transition from the minimum free energy structure into the (lowest) suboptimal state, S_n, that carries the required element and constitutes the active form of the ribozyme (figure4). This concept for the evaluation of catalytic efficiency can also be interpreted as a distance measure between structures on an energy scale. It is preferred here over simpler distance measures like the Hamming distance based string comparison methods, because it retains the useful and realistic statistical properties of the RNA sequence to secondary structure map as exemplified by the evolutionary fitness landscape [5, 31]. Moreover the activation concept allows for an optimization of the catalytic efficiencies of ribozymes through accumulation of mutations that reduce E_a by stabilizing the suboptimal structure S_n relative to the minimum free energy conformation S_0. Ideally, if their minimum free energy structure S_0 is identical to the target structure S_n, $\exp(E_a/RT) = 1$, and the ribozyme catalyzes mediator activation at maximum velocity.

After the determination of the parameters by the respective mapping, the genabolic network is translated into a set of ordinary differential equations. In order to describe the system in a general and easily accessible format, it is implemented in the widely used Systems Biology Markup Language (SBML) [32]. From the concentration time course fitness values can be deduced to drive an evolutionary optimization procedure (see Figure 5). The integrator front end currently used is the SBML-ODE Solver, a versatile integrator for continuous ODE systems [33]. Due to the use of SBML a variety of integrators and analysis software can easily be adopted and a flexible handling is facilitated.

Results

Mutation studies without selection have already been performed and they show that a sufficiently high fraction of mutations gives rise to viable regulatory networks. To test the capabilities of the model we designed the following experiment. We ask the question: Is a cell with an initial random genome capable of adapting its cell volume to a predefined target volume during an mutational adaptive walk? Figure 6 shows the dynamical behavior of the final cell of the adaptive walk. The balance between regulatory and metabolic dynamics indeed adjusts the cell's volume exactly to the target volume. Figure 7 shows the respective gene regulatory and metabolic reaction network.

There is also a substantial fraction of neutral mutations, which was found to be a *conditio sine qua non* for efficient evolutionary optimization [5, 31, 34]. In order to study evolutionary phenomena, we have to consider genome replication and variation (mutation and/or recombination) too. The secondary structure computations can be readily extended to DNA-DNA and DNA-

RNA interactions since the same folding routines can be used with other sets of empirical parameters [35, 36]. The evolutionary evaluation of different genabolic networks is base on their fitness that results from a sophisticated interplay of replication rate and metabolic efficiency.

Conclusions and Outlook

With the presented model at hand several pending problems can be approached. The encryption of all relevant system information within a string genome allows the description and evolution of genabolic networks in an entirely independent fashion. No external sources of additive information are necessary, the system is self-determined and closed as far as rules and system-sustaining model functions are concerned. In contrast to prior exclusively RNA based auto-catalytic systems, the genotype and the phenotype in the presented model constitute separate objects. This allows an unhindered evolvability of the minimal cell on the way from a random dynamical network to an adapted functional system.

Due to the regulation mechanisms implemented, in particular the direct interaction of the transcription factors in a cooperative manner, the model allows for studies of the evolution of a great variety of regulatory networks. Experiments regarding an optimization of certain qualities or functions, for example high adaptability, high robustness, insensitivity to environmental stress, are conceivable.

The natural selection criterion is a short time of self-reproduction for the individual system. This reproduction efficiency will be a function of the cell's replication machinery and the cellular metabolism. Our model encodes all these features in the genome and the system itself unfolds the phenotype. One way is the use of a growing cell membrane as indicator for cell growth and replication. This approach includes an explicitly modeled membrane as described, for example, in the chemoton systems [37]. At a certain size and concentration of the cell components the individual could ready for cell division, giving an additional input to the fitness measure for selection.

Further extensions of the system are easily implemented due to the modular SBML format. Several aspects are of special interest: The extension of the system to a second class of biomolecules and for this goal lattice proteins with monomer sequences encoded in the transcribed structural molecules would be an excellent candidate. The introduction of membrane bound transporter molecules which are responsible for the exchange of high and low energy compounds fueling the minimal cell would allow a direct interaction with the environment. A further evolutionary step, the adaptation of the transporter to operate on a larger variety of molecules is likely to enable cell-cell communication. SBML on the other hand supports also the idea of a compartmentalization of this model. This would enable a spatial separation of different metabolic processes in a manner that is closer to physiology.

Acknowledgment

This work was supported in part by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) project number MA05.

References

[1] C. K. Biebricher. Darwinian selection of self-replicating RNA molecules. In M. K. Hecht, B. Wallace, and G. T. Prance, editors, *Evolutionary Biology, Vol. 16*, pages 1–52. Plenum Publishing Corporation, 1983.

[2] C. K. Biebricher and W. C. Gardiner. Molecular evolution of RNA *in vitro*. *Biophys. Chem.*, 66:179–192, 1997.

[3] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From Sequences to Shapes and back: A case study in RNA secondary Structure. *Proc. Roy. Soc. (London) B*, 255:279–284, 1994.

[4] P. Schuster. A testable genotype-phenotype map: Modeling evolution of RNA molecules. In M. Lässig and A. Valleriani, editors, *Biological Evolution and Statistical Physics*, pages 56–83. Springer-Verlag, Berlin, 2002.

[5] W. Fontana and P. Schuster. Continuity in Evolution: On the Nature of Transitions. *Science*, 280:145–165, 1998.

[6] P. Schuster. Evolution *in silico* and *in vitro*: The RNA model. *Biol. Chem.*, 382:1301–1314, 2001.

[7] P. Schuster. Molecular insight into the evolution of phenotypes. In J. P. Crutchfield and P. Schuster, editors, *Evolutionary Dynamics – Exploring the Interplay of Accident, Selection, Neutrality, and Function*, pages 163–215. Oxford University Press, New York, 2003.

[8] W. Fontana. Modelling 'Evo-Devo' with RNA. *BioEssays*, 24:1164–1177, 2002.

[9] E. H. Davidson. *Genomic Regulatory Networks*. Academic Press, 2001.

[10] W. Banzhaf and J. Miller. *The Challenge of Complexity*, chapter 1. Kluwer Academic Publishers, 2004.

[11] S. A. Kauffman. *The Origin of Order*. Oxford University Press, 1993.

[12] T. Reil. Dynamics of gene expression in an artificial genome-implications for biological and artificial ontogeny. *Advances in Artificial Life–Proceedings of the 5th European Conference on Artificial Life (ECAL '99)*, 1674:457–466, 1999.

[13] F. Dellaert and R. D. Beer. A Developmental Model for the Evolution of Complete Autonomous Agents. In P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack, and S. W. Wilson, editors, *From Animals to Animates 4: Proceedings of the 4th Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts USA, September 9-13 1996. MIT Press.

[14] P. Eggenberg. Evolving morphologies of simulated 3d organisms based on differential gene expression. In P. Husbands and I. Harvey, editors, *Proceedings of the 4th European Conference on Artificial Life (ECAL '97)*, London, UK Europe, September 9-13 1997. MIT Press.

[15] J. Bongard. Evolving modular genetic regulatory networks. In *Proceedings of the IEEE 2002 Congress on Evolutionary Computation (CEC '02)*, volume 2, pages 1872–1877, Honolulu, HI USA, May 12-17 2002. IEEE Press.

[16] K. O. Stanley and R. Miikkulainen. A taxonomy for artificial embryogeny. *Artificial Life*, 9(2):93–130, 2003.

[17] W. Banzhaf. On the dynamics of an artificial regulatory network. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, and J. Ziegler, editors, *Advances in Artificial Life–Proceedings of the 7th European Conference on Artificial Life (ECAL '03)*, pages 217–227, Dortmund, Germany Europe, September 14-17 2003. Springer.

[18] W. Banzhaf. *Artificial Regulatory Networks and Genetic Programming*, chapter 4. Kluwer Academic Publishers, 2003.

[19] D. P. Kuo, A. Leier, and W. Banzhaf. Evolving Dynamics in an Artificial Regulatory Network Model. In X. Yao, E. Burke, J. Lozano, J. Smith, J. Merelo-Guervs, J. Bullinaria, J. Rowe, P. Tino, A. Kabán, and H.-P. Schwefel, editors, *Proc. of the Paralell Problem Solving from Nature Conference*, volume LNCS 3242, pages 571–580. Springer, 2004.

[20] B. N. Kholodenko and H. V. Westerhoff. The macroworld versus the microworld of biochemical regulation and control. *TIBS*, 20(2):52–54, 1995.

[21] P. J. Kennedy and T. R. Osborn. A model of gene expression and regulation in an artificial cellular organism. *Complex Systems*, 13:33–59, 2001.

[22] M. T. McManus and P. A. Sharp. Gene silencing in mammals by small interfering RNAs. *Nature Rev. Genetics*, 3:737–747, 2002.

[23] P. Bentley and S. Kumar. Three Ways to Grow Designs: A Comparison of Embryogenies for an Evolutionary Design Problem. In W. Banzhaf, editor, *Genetic and Evolutionary Computation Conference (GECCO '99)*, pages 35–43, Orlando, Florida USA, July 14-17 1999.

[24] M. Ebner, M. Shackleton, and R. Shipman. How Neutral Networks Influence Evolvability. *Complexity*, 7(2):19–33, 2001.

[25] P. Schuster and P. F. Stadler. Modelling conformational flexibility and evolution of structure – RNA as an example. In U. Blastolla, M. Porto, H. E. Roman, and M. Vendruscolo, editors, *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, Biological and Medical Physics, Biomedical Engineering, chapter 1, pages 1–34. Springer-Verlag, Berlin, 2006.

[26] W. Banzhaf and A. Leier. *Evolution on Neutral Networks in Genetic Programming*, chapter 14. Kluwer Academic Publishers, 2006.

[27] M. Ptashne and A. Gann. *Genes & Signals*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2002.

[28] C. K. Biebricher, M. Eigen, and W. C. Gardiner. Kinetics of RNA replication. *Biochemistry*, 22(10):2544–2559, 1983.

[29] R. A. Dimitrov and M. Zuker. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.*, 87:215–226, 2004.

[30] S. H. Bernhart, H. Tafer, U. Mückstein, C. Flamm, P. F. Stadler, and I. L. Hofacker. Partition Function and Base Pair Probabilities of RNA Heterodimers. *Algo. Mol. Biol.*, 2006.

[31] W. Fontana and P. Schuster. Shaping space. The possible and the attainable in RNA genotype-phenotype mapping. *J. Theor. Biol.*, 194:491–515, 1998.

[32] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[33] R. Machné, A. Finney, S. Müller, J. Lu, S. Widder, and C. Flamm. The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics*, 22(11):1406–1407, 2006.

[34] M. A. Huynen, P. F. Stadler, and W. Fontana. Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93:397–401, 1996.

[35] J. SantaLucia, Jr., H. L. Allawi, and P. Seneviratne. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, 35:3555–3562, 1996.

[36] J. SantaLucia, Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95:1460–1465, 1998.

[37] T. Gantí. Organization of chemical reactions into dividing and metabolizing units: The Chemotons. *Biosystems*, 7(1):15–21, 1975.

Table 1: Binding energies of random RNA pairs of length 25 to two URS RNAs of length 5. Different base compositions in the short RNA sequences from pure AU to pure GC were studied. The free energies for cooperative interaction (ΔG_{coop}) are calculated for the conformations of lowest free binding energies (ΔG_{bind}) as indicated in figure 2 and given with the standard deviation.

Sequence	Sample	Sample size:	Hetero-cooperativity			Self-cooperativity		
% AU	# URS	Sequence pairs	Stable (%)	$-\Delta G_{\mathsf{coop}}$ [kcal·mol^{-1}]	$-\Delta G_{\mathsf{bind}}$ [kcal·mol^{-1}]	Stable (%)	$-\Delta G_{\mathsf{coop}}$ [kcal·mol^{-1}]	$-\Delta G_{\mathsf{bind}}$ [kcal·mol^{-1}]
100	11	1000000	0.1 ± 0.1	3.13 ± 2.26	0.32 ± 0.22	0.2 ± 0.3	4.00 ± 2.77	0.32 ± 0.21
90	56	824301	1.0 ± 1.2	3.02 ± 2.21	0.58 ± 0.55	0.9 ± 1.1	3.71 ± 2.65	0.72 ± 0.68
80	223	364473	8.8 ± 9.2	2.90 ± 2.15	0.87 ± 0.76	7.5 ± 8.5	3.53 ± 2.57	0.92 ± 0.78
70	557	163760	18.4 ± 12.7	2.85 ± 2.12	1.14 ± 0.98	14.2 ± 10.1	3.35 ± 2.46	1.17 ± 0.99
60	1021	62027	31.6 ± 14.3	2.80 ± 2.10	1.45 ± 1.17	23.9 ± 11.7	3.24 ± 2.39	1.47 ± 1.18
50	1265	28598	42.7 ± 13.3	2.78 ± 2.08	1.74 ± 1.35	32.2 ± 11.8	3.20 ± 2.35	1.74 ± 1.32
40	1007	16860	53.8 ± 11.0	2.75 ± 2.07	2.09 ± 1.50	41.6 ± 10.9	3.17 ± 2.33	2.03 ± 1.45
30	574	14305	60.9 ± 9.6	2.75 ± 2.07	2.43 ± 1.66	48.4 ± 10.2	3.21 ± 2.34	2.33 ± 1.59
20	238	13105	65.7 ± 8.6	2.75 ± 2.06	2.69 ± 1.76	54.0 ± 9.6	3.27 ± 2.37	2.55 ± 1.67
10	35	12054	70.9 ± 7.0	2.76 ± 2.07	3.07 ± 1.91	58.9 ± 9.1	3.31 ± 2.39	2.85 ± 1.80
0	8	11860	73.3 ± 8.4	2.85 ± 2.11	3.51 ± 2.07	66.0 ± 4.1	3.67 ± 2.63	3.29 ± 1.98

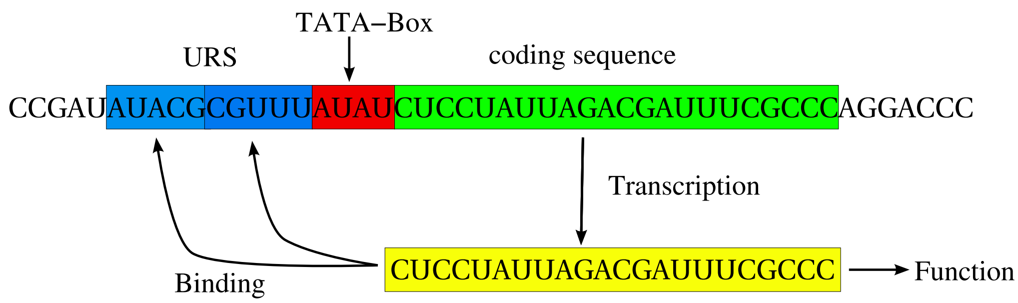

Figure 1: Gene control structure of a typical gene in the proposed model. Upstream of the coding sequence (green), lie the promotor region or TATA-box (red) and an upstream regulating sequence (blue), consisting of to transcription factor binding sites. The function of the gene transcript is determined by folding into secondary structures representing the different classes of functional RNAs.

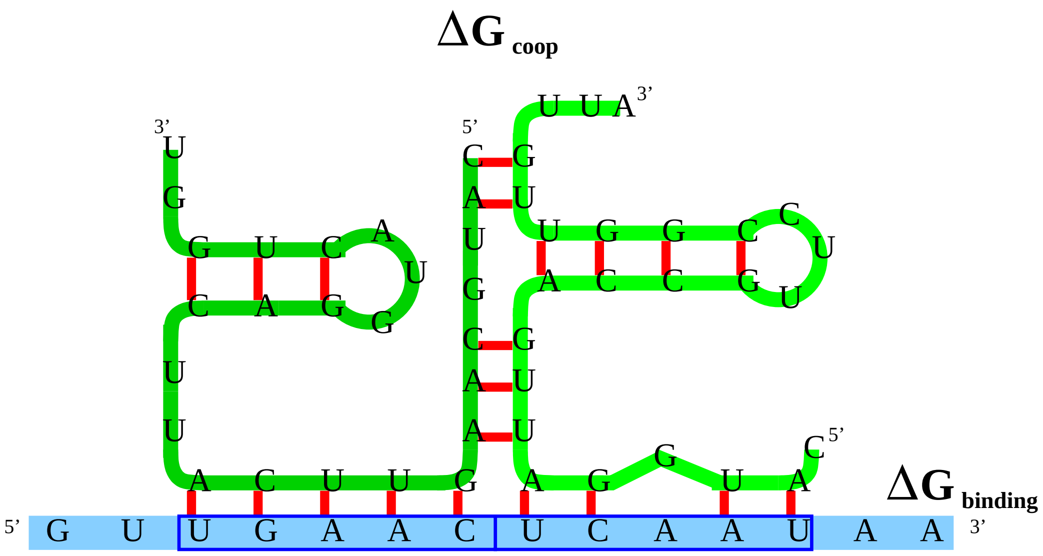

Figure 2: Cooperativity as modeled in the presented framework. Two TFs bind to the regulatory URS. The stabilizing free energy between the two functional structures enters additively into the regulatory mechanism.

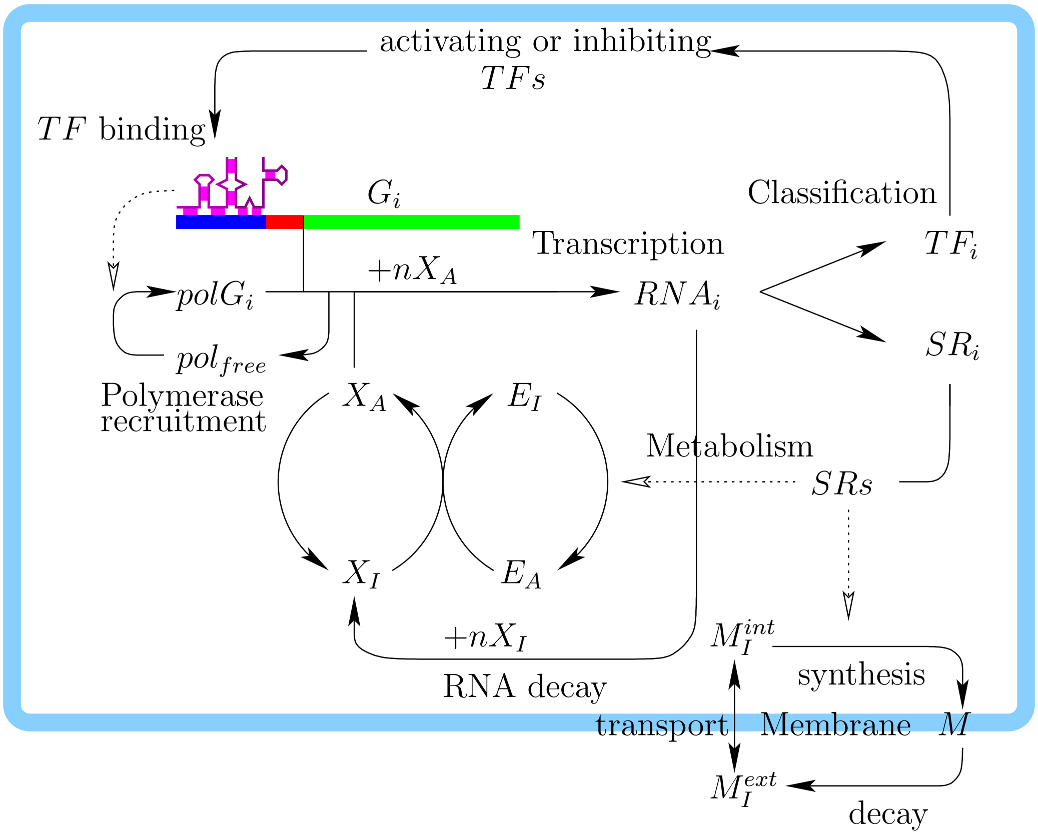

Figure 3: A sketch of the chemical reaction network of the presented model. The RNA polymerase is assumed to be available in a fixed amount and recruited to the genes (green) promoter region (red) at a rate determined by the TFs bound to the upstream regulating sequence (blue). The RNA transcription rate depends on the concentration of activated RNA building blocks (X_A) and consumes nX_A per RNA. The RNAs decay to inactivated components (X_I), which are reactivated via consumption of activated energy rich metabolites (E_A). The gene products are categorized into structural (SR) and gene regulatory (TF) RNAs. The SRs catalyze the activation of metabolites (E_I) and the incorporation of membrane building blocks (M_I^{int}) into the membrane (M). The internal pool of M_I^{int} is coupled to the exterior pool (M_I^{ext}) via diffusion through the membrane. All parameters for transcription factor binding to regulatory regions and the catalytic efficiencies of structural proteins are obtained by a mapping process (see text for details) and are therefore targets of evolution.

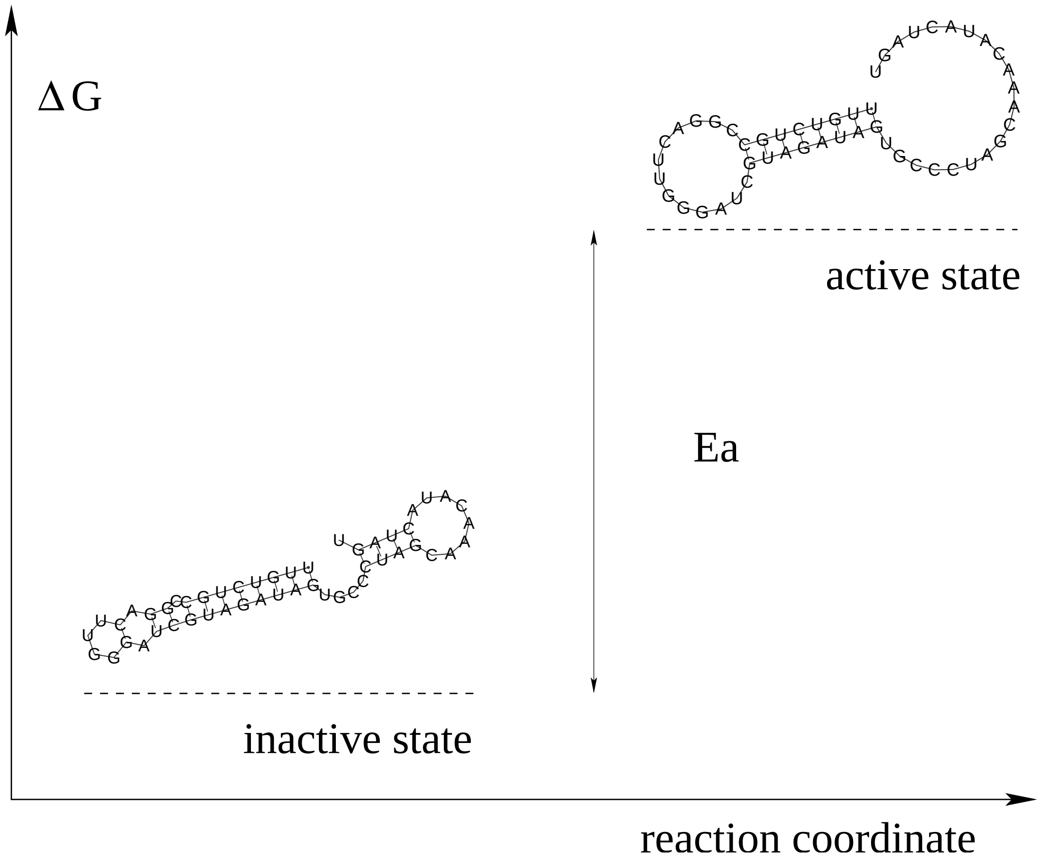

Figure 4: Activation of the ribozyme. The active structure catalyzes the metabolic reactions. The free energy needed to form the the secondary structure of the active ribozyme from the minimal free energy secondary structure of the RNA is used to derive the catalytic efficiency of the ribozyme.

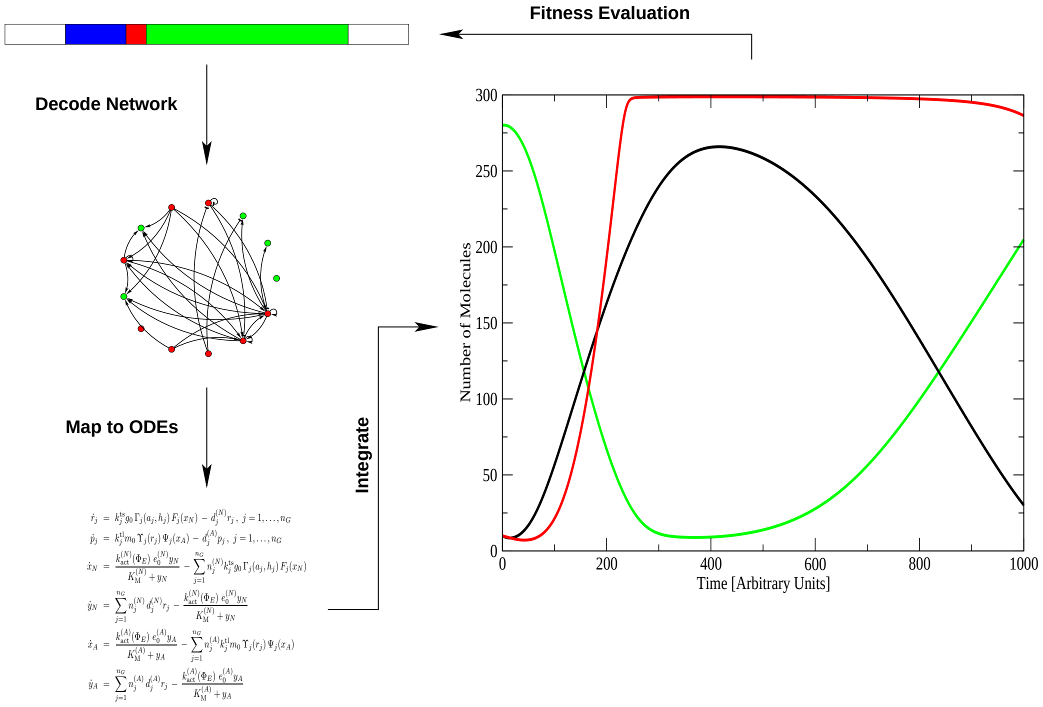

Figure 5: Schematic representation of an evolutionary cycle. The topology of the genabolic network together with the reaction parameters is *decoded* from the genotype. This information is translated into an ODE system which after numeric integration provides the concentration time coarse of the individual chemical species (the phenotype) which in turn modulates via a fitness function the reproductive efficiency of the genome.

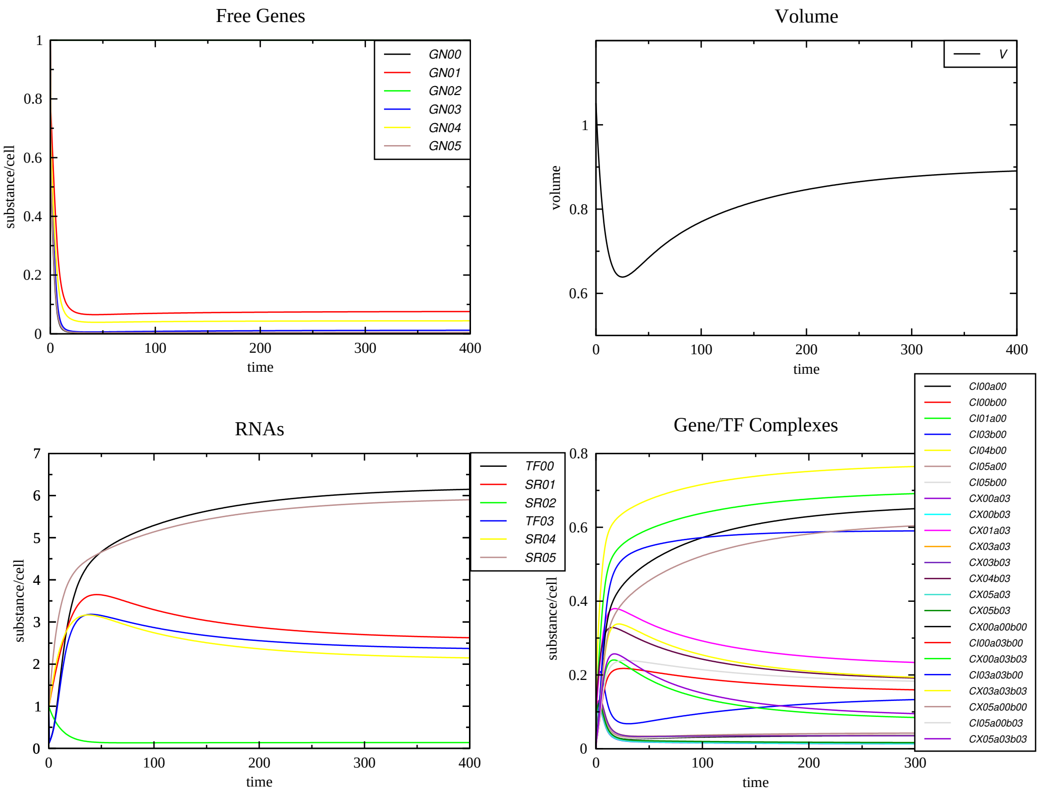

Figure 6: Integrated time course of a cell evolved via an adaptive walk targeting a cell volume of 1. A genome of length 100 and a gene length of 20 nucleotides were chosen. URS length was 5 per site and the promoter sequence motif was CC. Genes, transcription factors and ribozymes are labeled GN, TF and SR. The TF/gene complexes are labeled either CX or CI for activating or inhibiting complexes. The first index denotes the gene, letters a and b the sites the transcription factors bind to.

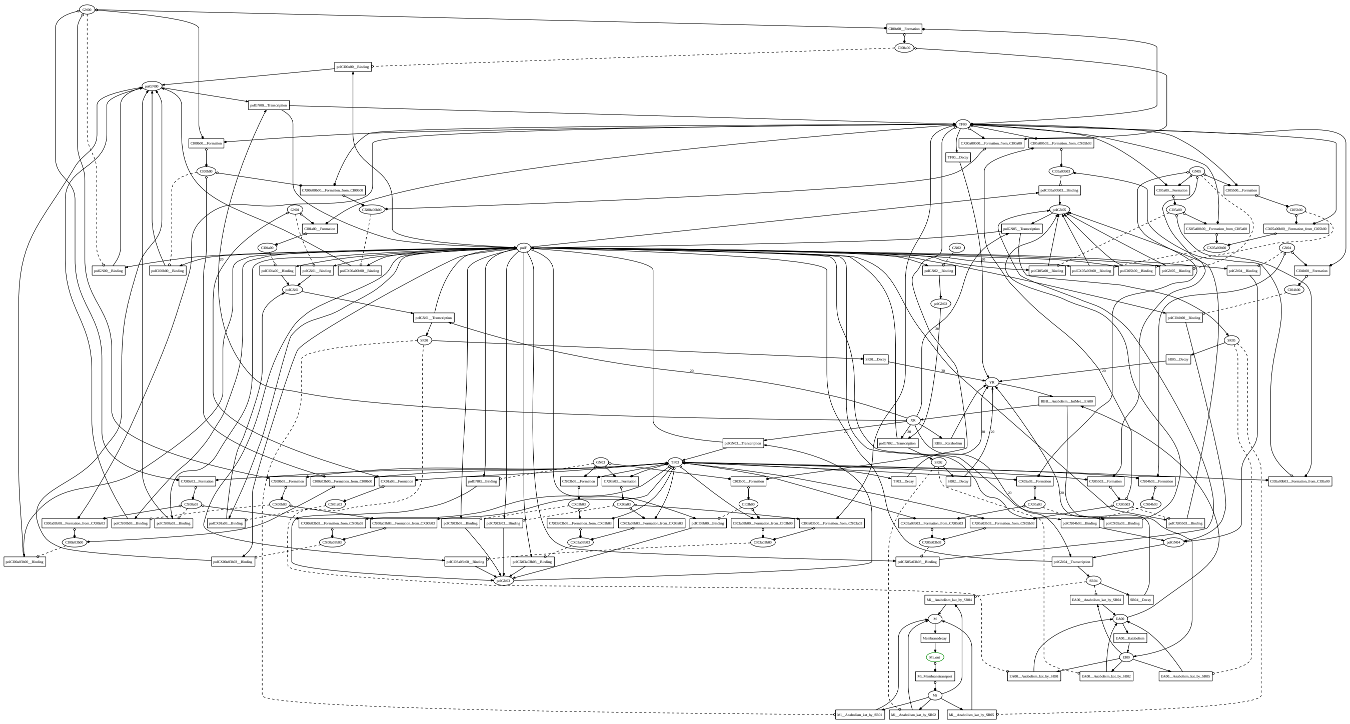

Figure 7: Reaction network of the cell in figure 6. The actual hyper graph is displayed in the König representation as bipartite graph. Rectangular nodes stand for reactions and circular nodes represent chemical species