Molecular Evolution of Catalysis

Christian V. Forst
IMB Jena
Beutenbergstr. 11
07745 Jena
Germany
E-mail: `chris @ imb-jena.de`

Abstract

In this paper we consider evolutionary dynamics of catalytically active species with a distinct genotype – phenotype relationship. Folding landscapes of RNA-molecules serve as a paradigm for this relationship with essential neutral properties. This landscape itself is partitioned by phenotypes (realized as secondary structures). To each genotype (represented as sequence) a structure is assigned in a unique way. The set of all sequences which map into a particular structure is modeled as random graph in sequence space (the so-called *neutral network*). A catalytic network is realized as a random digraph with maximal out-degree 2 and secondary structures as vertex set. Studying a population of catalytic RNA-molecules shows significantly different behavior compared to a deterministic description: hypercycles are able to co-exist and survive resp. a parasite with superior catalytic support. A "switching" between different dynamical organizations of the network can be observed, dynamical stability of a hypercyclic organization against errors and the existence of an error-threshold of catalysis can be reported.

1 Introduction

Todays biology heads towards a grand synthesis of combining knowledge from three different disciplines: molecular biology, developmental biology, and evolutionary biology. Initialized by the epochal discoveries made by Francis Crick and James Watson in the year 1953 molecular biology came into being. First steps in the direction of a *molecular evolutionary biology* were already taken in the late sixties by the pioneer work of Sol Spiegelman [22] who developed *serial transfer experiments* as a new method of molecular evolution in the test tube. Manfred Eigen [4] at about the same time formulated a kinetic theory of molecular evolution. Since then studying evolution of molecules in laboratory systems has become a research area of its own. Nowadays these experiments in the test tube indicate that evolution in the sense of Charles Darwin – by selection and variation – is not uniquely granted to cellular life only: optimization of replicating biopolymers by properties related to *fitness* is readily observed *in vitro* with *naked* RNA molecules in evolutionary experiments.

Darwinian dynamics is understood as a hill climbing process on a *fitness landscape* where fitness-values are assigned to genotypes (i.e. DNA or RNA sequences). Interpreted in terms of the theory of dynamical systems this dynamic is simple in the sense of following a gradient and finding a (local or global) fitness maximum. Even at this level of description controversy approaches have been discussed in history. Not only climbing to an optimum but genetic drift is seen as playing an important role for improving the *evolutionary search capacity* of a whole population [25, 26]. This *neutral theory* which often has been understood as antipole to Darwin's theory was proposed by Motoo Kimura in the late sixties [15, 14]. In the contrary he does not assume that selection plays *no* role but denies that any appreciable fraction of molecular change is caused by selective forces. He is saying in his theory that the *majority of changes* in evolution at the molecular level were the results of random drift of genotypes. Genetic drift and gradient walk are not the only feature of evolutionary systems. Interactions through catalysis, predator-prey or host-parasite behavior suppresses optimization of individual fitness. Formulating molecular evolution in terms of evolutionary dynamics has to take into consideration not only approaching a steady state but also complex dynamical phenomena like oscillations or chaotic behavior in space and time.

Evolutionary dynamics itself is a highly complex process. Therefore we omit additional difficulties in considering spatiotemporal patterns and introduce a comprehensive model which tries to account for most of the relevant features of molecular evolution. Peter Schuster proposed an interaction of three processes described in three different abstract metric spaces [21]:

- the *sequence space* of genotypes being DNA or RNA sequences,

- the *shape space* of phenotypes, and
- the *concentration space* of biochemical reaction kinetics.

The sequence space $\mathcal{Q}_\alpha^n$ is a metric space containing all sequences of chain-length n with alphabet-length α (i.e. the number of letters used in an alphabet $\mathcal{A}$). For DNA or RNA sequences α is 4, for peptides α would be 20 due to 4 distinct nucleotides common in DNA or RNA and 20 aminoacids in peptides and proteins resp. As a metric serves the Hamming metric [11]. The shape space covers all possible structures considered. The definition of structures itself is highly context dependent. Clearly two biopolymers with different sequences will form different spatial conformations at atomic resolution. This descriptions is, no question, adequate for comparison of active sites of enzymes (or ribozymes). In contrast, stating that all tRNAs have the same shape refers to a different understanding of structure on a coarse-grained level. Secondary structures may serve as one example which is suitable for our purpose of developing a mathematical model of molecular evolution. Similarities and dissimilarities of RNA structures can be expressed by means of mathematical measures with metric properties. Finally concentration space is the conventional space in which chemical reaction kinetics is described. Martin Feinberg formalized this concept in mathematical terms [6]. The conceptual cycle which is formed by a projection

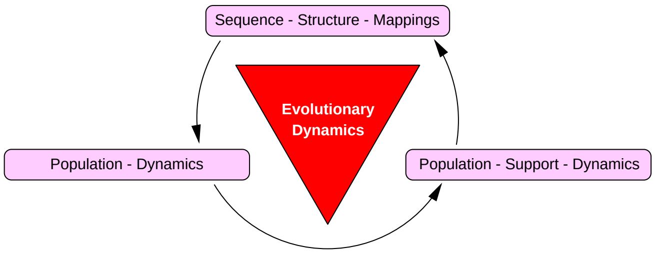

Figure 1: Evolutionary dynamics in sequence-, shape-, and concentration-space.

of evolutionary dynamics upon the three abstract spaces is sketched in Fig. 1.

Question raised in this papers deal with all three aspects of evolutionary dynamics. We give a mathematical description of a generic genotype – phenotype relation which maps RNA sequences to secondary structures. The topology of preimages in sequence-space with respect to this mapping is analyzed by graph theory. This preimage forms a so called *neutral network*. By subsequent mapping a complete folding landscape is constructed. A dynamical system is presented where the time evolution of a finite population of RNA molecules is described by stochastic point processes. By error-prone replication and unspecific elution the population moves in sequence space "guided" by evolutionary dynamics. We present a chemical reaction network which implies complex dynamical behavior and which is explored by this population. One critical parameter is the error-rate which determines the dynamics significantly. The occurrence of an *error-threshold of catalysis* can be reported.

2 Neutral Networks of RNA Structures

2.1 Secondary Structures and Compatible Sets

RNA secondary structures and the induced sequence – structure relationship are a suitable and generic description for genotype – phenotype mapping which is important in molecular evolutionary biology. RNA secondary structures represent a type of coarse-graining of biopolymer structures. They commonly are understood as list of Watson-Crick (**A=U** and **G≡C**) and Wobble (**G−U**) base-pairs which are compatible with unknotted and pseudoknot-free two-dimensional graphs (for a precise formal definition we refer to Waterman [24]).

Defining secondary structures independently of chemical or physical restrictions yields a general description based on contacts with respect to arbitrary alphabets $\mathcal{A}$ with arbitrary pairing rules Π. A *pairing rule* Π on $\mathcal{A}$ is given as a set of pairs of letters from the given alphabet (i.e. **AU**, **UA**, **GC**, **CG**, **GU**, **UG** for natural RNA-molecules with alphabet **A**, **U**, **G**, **C**). This concept can easily be extended to a general description of biopolymer structures via contact maps [16]. Similar to secondary structures a general *contact structure* c is determined by a *set of contacts* of c omitting the trivial contacts due to adjacent letters in the succession of the sequence.

A relevant concept in studying sequence – structure relation is how sequences has to be composed to fulfill necessary conditions for folding into a desired structure. In the following we define *compatibility* of a sequence to a given structure: A sequence x is said to be *compatible* to a structure s if all basepairs required by s can be provided by x_i and $x_j \in x$ with respect to the pairing rule Π for each base pair. $\mathbf{C}(s)$ is the set of all sequences which are compatible to structure s. The number of compatible sequences is readily computed for secondary structures (with n_u unpaired bases and n_p base pairs this evaluates to $4^{n_u} \cdot 6^{n_p}$).

2.2 Preimages and Complete Mappings

The relation between RNA sequences and secondary structures is understood as a (non-necessarily invertible) mapping f_n from sequence space $\mathcal{Q}_\alpha^n$ into shape space $\mathcal{S}_n$. The set of all sequences folding into a given structure is denoted as *neutral network* $\Gamma_n(s)$ with respect to s. $\mathcal{Q}_\alpha^n$ denotes the generalized hypercube of dimension n over an alphabet $\mathcal{A}$ of size α (i.e. the number of letters in $\mathcal{A}$ is α), and $s \in \mathcal{S}_n$ is a fixed secondary structure. Mathematically $\Gamma_n(s)$ refers to the induced subgraph of

$f_n^{-1}(s)$ in $\mathbf{C}(s)$ ($f_n^{-1}(s)$ indicates the *preimage* of a fixed structure s w.r.t the mapping f_n). A scetch of these embeddings is shown in Fig. 2.

Remark: The graph of compatible sequences $\mathcal{C}(s)$ to a fixed secondary structure s is

$$\mathcal{C}(s) = \mathcal{Q}_\alpha^{n_u} \times \mathcal{Q}_\beta^{n_p}$$

α is the number of different nucleotides, and β is the number of different *types* of base pairs that can be formed by α different nucleotides.

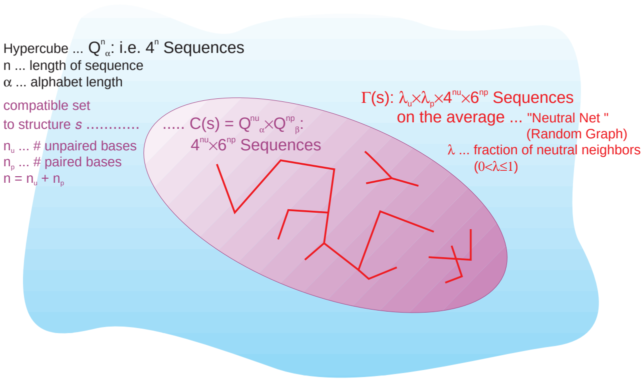

Figure 2: Sketch of a neutral network $\Gamma(s)$ (shown as solid line graph) embedded in the set of sequences compatible to structure s (i.e. $\mathbf{C}(s)$ – indicated as oval) which itself is embedded in sequence space $\mathcal{Q}_\alpha^n$ (realized as shaded background).

Sequences folding into the same structure are thus represented by a subgraph that is *randomly* induced on the set of compatible sequences corresponding to this structure [19]. Accordingly this model is of probabilistic nature and properties of random subgraphs are studied as functions of a single parameter – a probability measure over all possible induced subgraphs in a given sub-hypercube with a *choosing parameter* λ. This parameter represents the mean fraction of neighbors that are neutral with respect to the structure. – the *fraction of neutrality* with $0 \leq \lambda \leq 1$. Considering paired and unpaired regions we are dealing with two independent assignments for each corresponding choosing probability λ_p and λ_u resp. are used. Now vertices in both subcubes $\mathcal{Q}_\alpha^{n_u}$ and $\mathcal{Q}_\beta^{n_p}$ are chosen independently with these probabilities λ_u and λ_p resp. This is equivalent to choosing pairs of vertices in $\mathcal{C}(s)$ and yields exactly the desired neutral network $\Gamma_n(s)$ as probability space with its corresponding probability measure.

Two remarkable results are assertions about *density* and *connectivity* of subgraphs. In analogy to percolation theory these propererties are fulfilled almost sure if λ exceeds the threshold value

$$\lambda^* = 1 - \sqrt[1-\alpha]{\alpha}$$

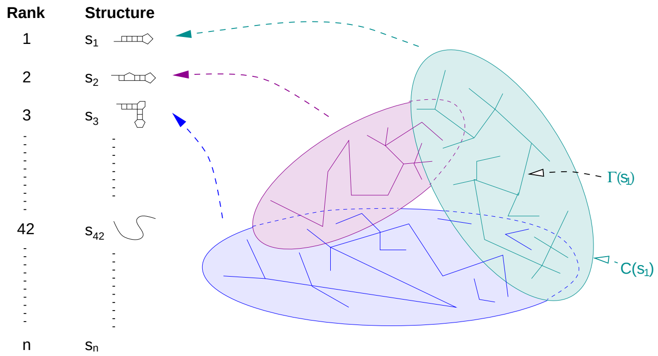

Figure 3: Complete sequence to structure mapping: neutral networks are constructed by subsequent (random) mapping of sequences to an ordered set of structures. Note that any pair of compatible sets have a non-empty intersection (for a proof see [19]).

Once we know how to construct a neutral network $\Gamma_n(s)$ we order the set of secondary structures $\mathcal{S}_n$ and define a complete mapping by iterating the construction process of the corresponding neutral network w.r.t. the ordering. The preimage for the structure with highest rank s_1 is assigned independently. For all other structures s_i, $i > 1$ the mapping depends on all previous assignments. A visualization of this process is shown in Fig. 3.

3 Dynamics of Catalytic Reaction Networks

3.1 Replication-Deletion Process

Let $\mathbf{P}$ be a (finite) population of size N defined as family of vertices $(P_i \,|\, i \in N)$ with $\{\, P_i \,|\, i \in N \,\} \subset \mathrm{v}(\mathcal{Q}_\alpha^n)$ and $N \geq 2$. The theory of point processes provides a powerful tool by identifying such a family $(v_i \,|\, i \in N)$ with an integer valued measure $\phi \colon \mathcal{Q}_\alpha^n \to \mathbb{R}$:

$$\mathbf{P} = (P_i \,|\, i \in N) \quad \longleftrightarrow \quad \phi \stackrel{def}{=\!=} \sum_{i=1}^{N} g_{P_i}, \tag{1}$$

$$\text{where } g_{P_i}(v) \stackrel{def}{=\!=} \begin{cases} 1 & \text{for } v \neq P_i \\ 0 & \text{otherwise .} \end{cases}$$

The set of sequences where ϕ is nonzero is referred to as *support* of ϕ. Thus the *restriction* of ϕ to a subgraph $Y \subset \mathcal{Q}_\alpha^n$ is tantamount to characterize subpopulations on $\mathrm{v}(Y)$.

The time evolution of ϕ is obtained by a mapping from $(P_i \,|\, i \in N)$ to the family $(P_i' \,|\, i \in N)$ as follows: we select an ordered pair (P_l, P_k) where $P_l, P_k \in \{ P_i \,|\, i \in N \}$. For this purpose let $\mathcal{F}_s$ be a "fitness" of s in a generalized sense which will be outlined in section 3.2. Accordingly the average "fitness" of ϕ reads

$$\mathcal{F}_\phi = \sum_s \phi(\mathrm{v}(\Gamma_n(s)))\mathcal{F}_s.$$

Now the first coordinate P_l of this ordered pair is chosen with probability $\mathcal{F}_s/\mathcal{F}_\phi$ among the elements of $\mathbf{P}$. The second coordinate of the above pair is selected with uniform probability on $(P_i \neq P_l \,|\, i \in N)$ i.e. $1/(N-1)$. We assume the times $\hat{T}$ between these mappings to be exponentially distributed (scaled by the mean fitness)

$$\boldsymbol{\mu}\{\hat{T} \leq t\} = e^{-\mathcal{F}_\phi t}.$$

We map $P_l = (x_1, ..., x_n)$ randomly into the vertex $P^* = (x'_1, ..., x'_n)$. This is done by mapping each coordinate x_i to a $x'_i \neq x_i$ with probability $\wp$ where all $x'_i \neq x_i$ are equally distributed and leave the coordinate fixed otherwise. This random mapping $v_l \mapsto v^*$ is called "replication". Finally, we delete the second coordinate of the pair (v_l, v_k), that is v_k and have a mapping $(P_l, P_k) \mapsto (P_l, P^*)$. Thereby we obtain a "new" family by substituting the P_k by the P^*. This process is referred to as *replication-deletion process.*

3.2 Exploring Catalytic Nets

The aim of this paper is to analyze the behavior of a population exploring a given (large) catalytic network. In terms of graph-theory we understand a *catalytic network* as digraph $\mathcal{N}$ with $\mathrm{v}(\mathcal{N}) \subseteq \mathcal{S}_n$ and an edge-set as follows: edges are valued by a function $g\colon \mathrm{e}(\mathcal{N}) \to \mathbb{R}_0^+$ which corresponds to a reaction-rate of catalysis. I.e. fixing an edge ϵ then we define $g(\epsilon)$ as reaction-rate for the catalysis of phenotype $s_i = t(\epsilon)$ by phenotype $s_k = o(\epsilon)$. We call $o(\epsilon)$ *catalyst* and $t(\epsilon)$ *template.* If $o(\epsilon) = t(\epsilon)$ then ϵ describes an *autocatalytic reaction.*

In our model we consider a maximal out-degree of $\mathcal{N} \leq 2$. In addition to catalysis there exists an uncatalyzed replication with rate $f\colon \mathrm{v}(\mathcal{N}) \to \mathbb{R}_0^+$

The generalized "fitness" introduced in section 3.1 corresponds directly with above defined reaction-rates f and g. Thus the "fitness" of an individual phenotype s is

$$\begin{aligned}\mathcal{F}_s \;=\;& \phi\{\mathrm{v}(\Gamma_n[s])\}f(s)\\ &+ \sum_{s=t(\epsilon)} \phi\{\mathrm{v}(\Gamma_n(s))\}\phi\{\mathrm{v}(\Gamma_n[o(\epsilon)])\}g(\epsilon). \quad (2)\end{aligned}$$

Translated to chemical reaction equation the dynamical behavior of the above system can be described as follows: We denote I_i, $i = 1, \ldots, n$ as reacting species with a fixed phenotype $s(I_i)$, W_{ij} indicates the probability of reproducing I_j by replicating I_i.:

$$\begin{aligned} I_i &\xrightarrow{f[s(I_i)]\cdot W_{ij}} I_i + I_j\\ I_i + I_k &\xrightarrow{g[s(I_i),s(I_k)]\cdot W_{ij}} I_i + I_j + I_k \qquad (3)\\ I_i &\xrightarrow{\Phi} \emptyset \end{aligned}$$

Thus the $f[s(I_i)]$'s refer to the level of phenotypes whereas the W_{ij}'s indicate the genotypes dynamic. The $g[s(I_i),\ s(I_k)]$'s denote the rate of the catalyzed reaction. By decoupling both levels from each other yields following schema (Fig. 4):

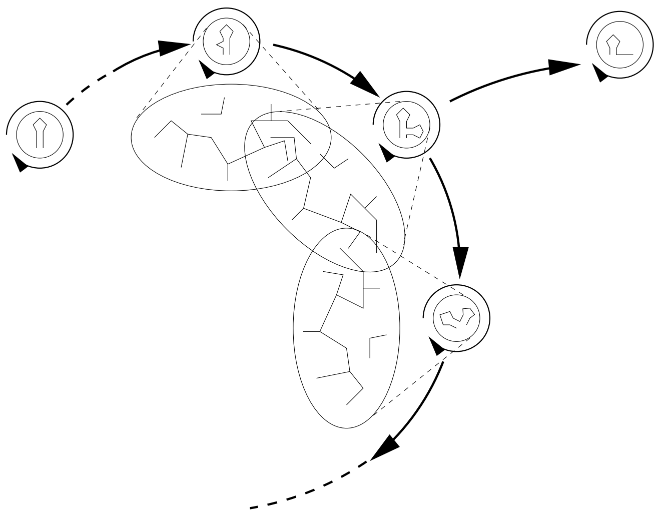

Figure 4: Hypercyclic reaction-network with a parasite including a genotype – phenotype mapping. Each phenotype — sketched by secondary structures — induce a corresponding compatible set and an embedded neutral net. The parasitic phenotype profits by a hypercyclic phenotype only without providing support.

Important for the dynamical characteristics of the system are the reaction-graphs of phenotypes and the error-prone replications on the level of genotypes. The topologies of the underlying neutral nets assure that there are couplings between each two of them. Parts of the population can switch from one net to the other and thereby cause a stabilizing effect for the hypercycle.

3.2.1 Hypercycles and Parasites

Considering stability of a hypercyclic organization one has to consider to what extend this stability can be maintained in drastic scenarios. One of these drastic scenarios is the competition of a hypercycle with a parasite (Fig. 4). The results known by stability analysis of this system in case of infinite populations let us expect the following [12]: depending on the ratio of the reaction rates of the competitors and of the initial concentrations either the hypercycle or the parasite will survive at the expense of the other.

Different approaches has been pursued by Boerlijst and Hogeweg [3] who studied hypercycles and parasites in an two-dimensional spatial system. They observed formation of spirals and a "cleaning"-capacity of the system against attacks of parasites.

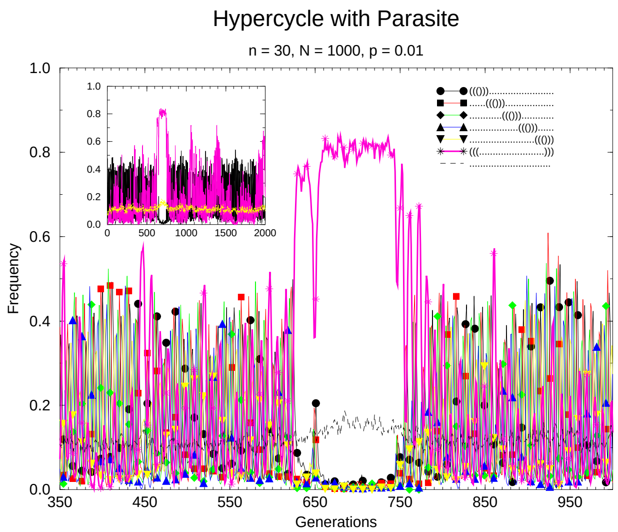

Figure 5: Fife-member Hypercycle with parasite. Starting with one member of the hypercycle the parasite becomes dominant at generation 600. The hypercycle is still able to coexist with the parasite and out-competes the latter at a later generation. Parameters are: population size = 1000, chain length = 30, catalyzed reactions are the same for hypercyclic members as for the parasite.

Similar effects can be noticed in our stochastic formulation of a homogeneous system with the above defined partly neutral genotype – phenotype mapping (equation 3): in contrast to homogeneous models with infinite population size the competition between parasite and hypercycle is not a 0-1 behavior. Both competitors can coexist for a long time and can over-populate but not extinguish each other. A characteristic time evolution is shown in Fig. 5.

3.2.2 *Voyaging large Catalytic Nets*

In order to study *formation*, structural *stabilization* against hostile environment, and *evolution* of catalytic active entities (i.e. hypercycles) we have to establish an artificial world of interacting catalytic networks. In contrast to approaches made by various people [1, 2, 7, 13, 18] where autocatalytic sets emerge automatically we setup a system with fixed but very large catalytic network which has to be explored by the population:

Using an ordered set of 1000 secondary structures as phenotypes, a catalytic network of about 420 random hypercycles of size 3 to 7, and about 18 additional random reaction paths is constructed. This yields a total of about 2000 possible reaction-paths. By an initial population of $N = 1000$ individual with phenotype of rank 1 the catalytic network is explored. For the corresponding reaction-rates we are dealing with two distinct scenarios:

- equal opportunity: function g is constant on $e(\mathcal{N})$ (Fig. 6.) therefore neutral evolution of catalytic units can be observed.
- optimization: g has a random distribution with distinct expectation- value and variance (Fig. 7). The population explores the catalytic network in search for better support.

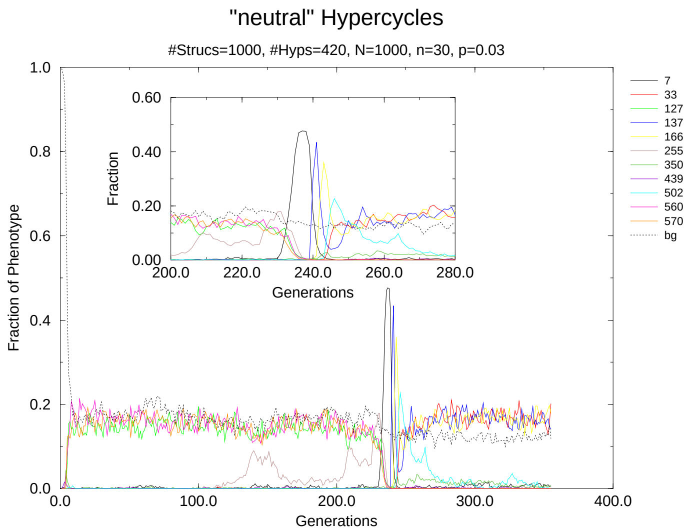

Figure 6: Neutral evolution of hypercycles: embedded in an artificial secondary-structure landscape of 1000 phenotypes 420 randomly chosen hypercycles of length three to seven are assigned. In addition 27 reaction paths between randomly chosen phenotypes are defined. Due to the constant function g no optimization is possible but "switching" from one hypercycle to another can be observed. In general only those phenotypes are displayed whichever reached the global, maximal concentration in the simulation at least once.

This still quite simple model gives deep insight how new catalytic active units can be formed, how they can reorganize by neutral drift and how optimization to more efficient units take place. Important for this purpose is the error-rate on genotype level which is contained in the stochastic mutation matrices W_{ij}. This leads us to the next section.

3.3 *Error-Threshold*

One essential question one may ask is, what is the influence of the error-rate on the dynamical behavior of the system. Well known is the quasi-species concept of Eigen *et al.*[4, 5] which has been described by an birth-death ansatz by Nowak & Schuster [17]. The critical error rate p_* of the genotypic error-threshold is determined by following expression:

$$(1 - p_*)^n = \frac{1}{\sigma} \tag{4}$$

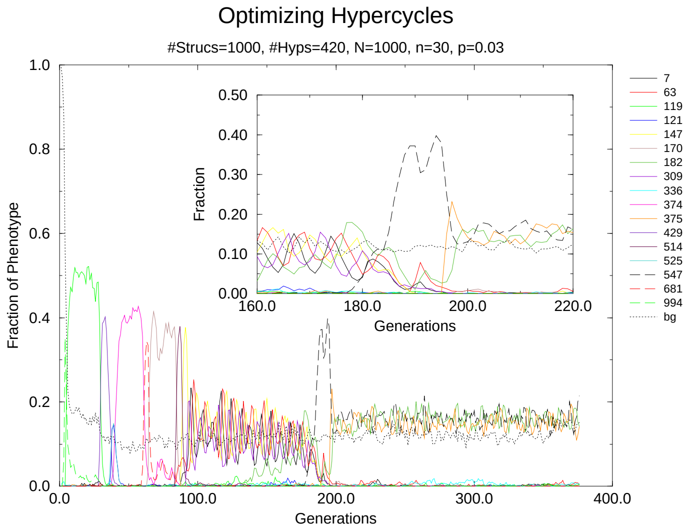

Figure 7: Voyaging large catalytic networks: similar to Fig. 6 420 randomly chosen hypercycles of length three to seven are assigned. In addition 18 reaction paths between randomly chosen phenotypes are defined. In this figure an optimization from one catalytic organization (four-member hypercycle) to a three-member hypercycle is shown. Analogue to Fig. 6 only those phenotypes are displayed whichever reached the global, maximal concentration in the simulation at least once.

where n is the chain-length and σ is the superiority of the master.

3.3.1 Phenotypic Error-Threshold

Extending the mean-field approach of Eigen *et al.*of an evolving population by a biologically motivated genotype – phenotype relationship yields the following system [8]: consider a single fixed secondary structure s and the preimage $f^{-1}(s)$ as subgraph $\Gamma(s)$ of the sequence-space. This induces in a natural way a *fitness landscape* f_Γ on the complete sequence space $\mathcal{Q}_\alpha^n$:

$$f(x) = \begin{Bmatrix} 1 & \text{if} & x \notin \Gamma \\ \sigma & \text{if} & x \in \Gamma \end{Bmatrix}. \qquad (5)$$

with $\sigma > 1$. We call f a *single shape* landscape in contrast to the single peak landscape of Eigen *et al.*. We used this landscape and applied a birth-death process of replicating RNA-molecules in a constant environment characterized in section 3.1. One significant result of this model is the existence of a so called *phenotypic error-threshold*. Beyond a critical error-rate p_* the information of a fit phenotype is lost for the population and is replaced by random drift. Fig. 8 shows a density plot for a single shape landscape with $\lambda_u = \lambda_p = 0.5$ and $N = 1000$.

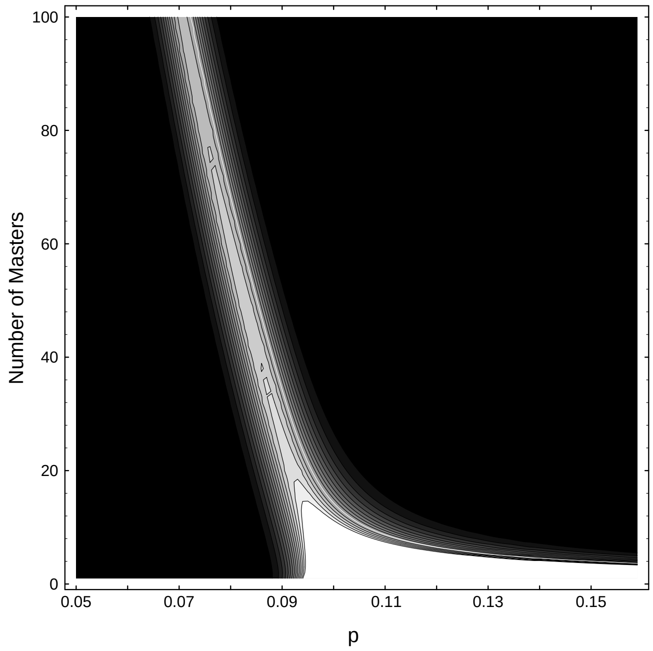

Figure 8: Density plot of the stationary distribution for a single-shape landscape with $\lambda_u = \lambda_p = 0.5$ and $N = 1000$. Light area denote high probability. Note that 0 is binding state beyond the error-threshold.

3.3.2 Error-Threshold of Catalysis

Extending the concept of the importance of a single phenotype for the evolving population to the maintenance of cooperative behavior of a set of phenotypes leads to a general understanding: Assume the importance of a hypercyclic organization. Then one may ask how stable is this catalytic network against mutational influences?

We are starting with a five-member hypercycle with population size $N = 1000$ and error-rate $p = 0$ (other parameters of the simulations are: chain-length $n = 30$, $n_u = n_p = 10$, catalytic support between members of hypercycle $h = 10$, all other reaction rates $r = 1$). By subsequently increasing p by 0.02 and "relaxation" of the system for 500 generations per new p-value we record the number of individuals which have the phenotype of the hypercycle (i.e. which are on the corresponding neutral networks). Fig. 9 shows the result of this simulation. One can notice that phases of organized hypercyclic dynamics alternates with aperiodicity where the hypercycle has to self-reorganize in sequence-space due to diffusion of its members (not shown). Beyond a critical error-rate $p_* = 0.11$ the information of the hypercycle is lost for the population. Comparing p_*-values between this system and the single-shape landscape with simple birth-death dynamics one observes a larger "buffering" capacity for the hypercyclic reaction-network. I.e. similar p_*-values are detected for $\lambda_u = \lambda_p = 0.5$ for the reaction-network and for $\lambda_u = \lambda_p = 0.8$ for the simple single-shape model.

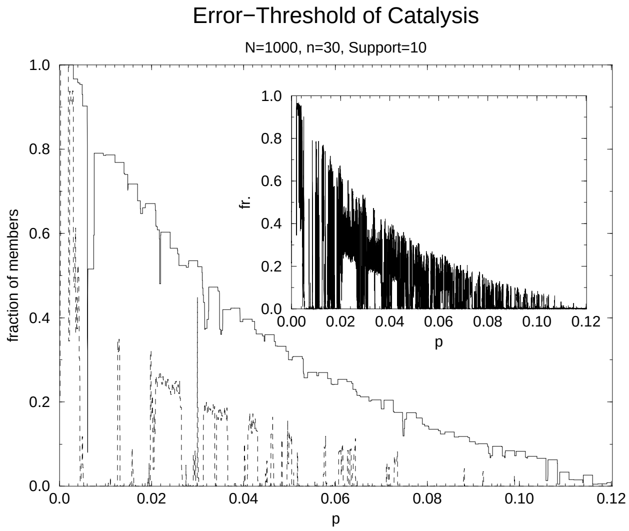

Figure 9: Error-Threshold of catalysis: the fraction of hypercyclic members depending on the error rate p is shown. The small plot refers to the member with maximal fraction per p-value. The large diagram show running averages of this data: the solid line denotes a running average over 500 samples of the maximum, the dashed line refers to a running average over 50 data-points of the minimum. Hypercyclic "activity" occur at extended maxima of the dashed line.

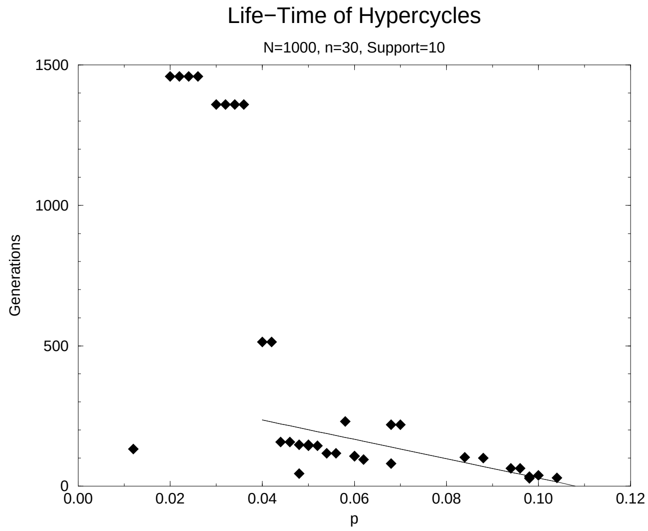

Figure 10: The life time of the periodic behavior of the hypercycle is shown depending on the error rate p. A "phase-transition" at $p \approx 0.04$ can be observed. The drawn solid line indicates a linear approximation of the data beyond this phase-transition. This line crosses the x-axes at $p = 0.108$ – the error-threshold.

There is a second effect of the error-rate on the stability of the periodic phases of the reaction-network. For higher p-values the observed lifetime of these phases become shorter. This can be understood by "mixing" properties of the mutation. Analogue studies of mutational influences on small deterministic autocatalytic reaction nets show a phase transition of the periodic and chaotic behavior of the system to a stable fixed point [20, 23]. Due to neutral properties of the underlying sequence – structure relationship the reaction-network is highly tolerant against these destabilizing effects. Although a significant change at $p_h = 0.04$ can be observed between long-living hypercyclic dynamics for $p < p_h$ and short lifetime for $p > p_h$ the existence of these dynamics itself persists up to high error-rates. Fig. 10 shows the life-time of the periodic behavior of the hypercycle depending on the error-rate p. The line refers to a linear regression for data-points at $p > p_h$ with $intercept = 374$ and $slope = -3449$. The linear regression function becomes 0 at $p = 0.108$ – the error threshold.

4 Conclusion

Evolutionary dynamics interpreted as an adaptive walk in an high-dimensional fitness landscape implies several assumptions which are crucial for the validity of the model: (i) constant environment and (ii) independent reproduction. In nature neither the first nor the second assumptions are realized even on a rudimentary scale. Constant environment can be interpreted that environmental changes shall occur on a slower time-scale than the adaption of the population. Coevolving species may influence the environment of a population. But this influence may take place on the same time-scale as adaption due to the same adaptive mechanism. Therefore landscape models of coevolution could easily fail and should be considered with special care. The second assumption of independent reproduction is dealt with similar care. Independent production implies independence of replication-rates from the concentration of the respective individual. Host-parasite systems, predator-prey scenarios and symbiotic interactions are examples of such dependencies and are impossible to be described by walks on complex landscapes. Experiments of directed molecular evolution performed in laboratories are almost always dealing with constant environments and the replication rate are commonly independent of the concentrations. In nature, however, neither the first nor the second assumptions are realized even on a rudimentary scale.

Recent experimental studies of quasi-viral amplification systems give a different picture [9, 10]. Utilizing even simple regulatory systems like quasi-viral amplification and understanding infectious cycles of Viridae depends crucially on insights of complex dynamical correlations. The model presented here is a first step in this direction: the neutral property of the underlying genotype – phenotype mapping both stabilizes the system and enables "save" optimization. Due to this property of the landscape catalytically organized entities (such as a hypercycle) become tolerant against parasites with even higher fitness. If the reaction rates of parasite and hypercycle are in the same order of magnitude an alternating revival and repression is observed. Hypercycles are capable for evolutionary optimization and reorganize accordingly (e.g. the periodicity can change with adaption). Not only the tolerance against competing systems but also tolerance against mutations has been observed. A catalytically organized entity maintains its organization up to high error rates. Before the dynamic behavior breaks down at the error-threshold of catalysis a phase transition occurs where short-living hypercycles follow hypercycles with long life time.

Evolutionary dynamics in its full capability of describing interacting evolving species provides a powerful tool for molecular evolutionary biology. Not only evolutionary experiments yield better models but also these models are capable for the design of optimal experimental strategies. This synthesis between theoretical model and experimental setup is the key of a comprehensive and complete description of molecular evolution.

Acknowledgments

I would like to thank Peter Schuster for fruitful discussions. Special thanks are directed to Christian Reidys for helpful hints especially by formulating graph-theoretical problems. Proof-reading of the manuscript by Ulrike Göbel and Janos Palinkas is gratefully acknowledged.

References

[1] Richard J. Bagley and J. Doyne Farmer. Spontaneous emergence of a metabolism. In Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, editors, *Artificial Life II*, volume X of *Santa Fe Institute Studies in the Science of Complexity*, chapter Origin/Self-Organization, pages 93–141. Addison Wesley, Redwood City, 1992.

[2] Richard J. Bagley, J. Doyne Farmer, and Walter Fontana. Evolution of a metabolism. In Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, editors, *Artificial Life II*, volume X of *Santa Fe Institute Studies in the Science of Complexity*, chapter Origin/Self-Organization, pages 141–158. Addison Wesley, Redwood City, 1992.

[3] M. C. Boerlijst and P. Hogeweg. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Physica D*, 48:17 – 28, 1991.

[4] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, 10:465–523, 1971.

[5] M. Eigen, J. McCaskill, and P. Schuster. The molecular Quasispecies. *Adv. Chem. Phys.*, 75:149 – 263, 1989.

[6] Martin Feinberg. Mathematical aspects of mass action kinetics. In L. Lapidus and N.R. Amundson, editors, *Chemical Reactor Theory, A Review*, pages 1–78, Englewood Cliffs, NJ, 1977. Prentice-Hall Inc.

[7] Walter Fontana and Leo W. Buss. “The Arrival of the Fittest”: Towards a theory of biological organization. *Bull. Math. Biol.*, 56:1–64, 1994.

[8] Christian V. Forst, Christian Reidys, and Jacqueline Weber. Evolutionary dynamics and optimization: Neutral Networks as model-landscape for RNA secondary-structure folding-landscapes. In F. Morán, A. Moreno, J.J. Merelo, and P. Chacón, editors, *Advances in Artificial Life*, volume 929 of *Lecture Notes in Artificial Intelligence*, Berlin, Heidelberg, New York, 1995. ECAL '95, Springer. Santa Fe Preprint 95-10-094.

[9] Michael Gebinoga. Hypercycles in biological systems. *J. Endocyt.*, 1995. submitted.

[10] Michael Gebinoga and Frank Oehlenschläger. Comparison of self-sustained sequence replication reaction systems. *Eur. J. Biochem.*, 235:256–261, 1996.

[11] R. W. Hamming. *Coding and Information Theory.* Prentice Hall, Englewood Cliffs, 1986.

[12] J. Hofbauer and K. Sigmund. *The Theory of Evolution and Dynamical Systems.* Cambridge University Press (Cambridge), 1988.

[13] Stuart A. Kauffman. Autocatalytic sets of proteins. *J. Theoret. Biol.*, 119(1):1–24, 1986.

[14] M. Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge Univ. Press, Cambridge, UK, 1983.

[15] Motoo Kimura. Evolutionary rate at the molecular level. *Nature*, 217:624 – 626, 1968.

[16] Stephan Kopp, Christian M. Reidys, and Peter Schuster. Exploration of artificial landscapes based on random graphs. In Frank Schweitzer, editor, *Self-Organization of Complex Structures: From Individual to Collective Behavior*, London, UK, 1996. Gordon and Breach.

[17] Martin Nowak and Peter Schuster. Error tresholds of replication in finite populations, mutation frequencies and the onset of Muller's ratchet. *Journal of theoretical Biology*, 137:375–395, 1989.

[18] Steen Rasmussen, Carsten Knudsen, and Rasmus Feldberg. Dynamics of programmable matter. In Christopher G. Langton, Charles Taylor, J. Doyne Farmer, and Steen Rasmussen, editors, *Artificial Life II*, volume X of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 211–254, Redwood City, 1991. Addison Wesley.

[19] Christian Reidys, Peter F. Stadler, and Peter Schuster. Generic properties of combinatory maps and neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 59(2):339–397, 1997.

[20] Wolfgang Schnabl, Peter F. Stadler, Christian Forst, and Peter Schuster. Full characterization of a strange attractor: Chaotic dynamics in low dimensional replicator systems. *Physica D*, 48:65 – 90, 1991.

[21] Peter Schuster. Artificial life and molecular evolutionary biology. In F. Morán, A. Moreno, J.J. Merelo, and P. Chacón, editors, *Advances in Artificial Life*, Lecture Notes in Artificial Intelligence, Berlin, Heidelberg, New York, 1995. ECAL '95, Springer.

[22] S. Spiegelman. An approach to the experimental analysis of precellular evolution. *Quart. Rev. Biophys.*, 17:213, 1971.

[23] Peter F. Stadler, Wolfgang Schnabl, Christian V. Forst, and Peter Schuster. Dynamic of small autocatalytic reaction networks – II. Replication, mutation and catalysis. *Bull. Math. Biol.*, 57(1):21–61, 1995.

[24] M. S. Waterman. Secondary structure of single - stranded nucleic acids. *Studies on foundations and combinatorics, Advances in mathematics supplementary studies, Academic Press N.Y.*, 1:167 – 212, 1978.

[25] Sewall Wright. Evolution in mendelian populations. *Genetics*, 16:97–159, 1931.

[26] Sewall Wright. The roles of mutation, inbreeding, crossbreeeding and selection in evolution. In D. F. Jones, editor, *int. Proceedings of the Sixth International Congress on Genetics*, volume 1, pages 356–366, 1932.