On Evolutionary Dynamics

Christian V. Forst[a1] **and Christian Reidys**[b]

[a]Institut für Molekulare Biotechnologie, Jena, Germany
Beutenbergstraße 11, p/o box 100 813, D-07708 Jena, Germany
Phone: **49 (3641) 65 6459 Fax: **49 (3641) 65 6450
E-Mail: chris@imb-jena.de

[b]Los Alamos National Laboratory, NM, USA
TSA-DO/SA, Los Alamos, NM 87545, USA
E-Mail: duck@santafe.edu

Abstract

Motivated by RNA secondary-structure folding-landscapes based on computer algorithms we construct a mathematical framework which provides a comprehensive and coherent description of evolutionary dynamics.

In this paper we study the evolutionary dynamics of a population of self-replicating RNA molecules with autocatalytic interactions (hypercycles and parasites). We use a stochastic replication dynamics in a mean field approximation (no spatial resolution), sitting on top of a genotype-phenotype mapping, including neutral networks (families of RNA genotypes with the same phenotype [31]). The neutral nets as well as the autocatalytic interactions are modeled by random graphs [29]. This approach enables us to develop a structure theory of landscapes on the one hand and to understand evolutionary dynamics by means of theory and simulation on the other hand.

The evolutionary dynamics select complex cooperative RNA structures. Hypercycles are found to be able to co-exist as well as survive parasites with superior catalytic support. These results are surprising when compared to a deterministic description of the same dynamics where neither co-existence nor survival in the presence of parasites are possible. Our findings are more in concert with the results obtained using a spatial description of the dynamics.

1. Introduction

Dynamics of self-replicating macromolecules has been modeled by two different approaches in literature:

Non-catalytically active entities: This concept of error-prone replication is based upon the selection-model of Fisher [11]. Individuals reproduce by their fitness, by accident mutations occur thus different individuals emerge by replication, and a constant organization maintains a selective force. Eigen *et al.*[8] extended these models by studying binary sequences as individuals. Applying mean-field-theory yields the well known *Quasispecies* model which has been analyzed by birth-death processes [23]. By extending this concept with fitness-landscapes based on RNA-secondary structure folding-landscapes [14, 13] one can study error-prone replication of RNA-molecules in a more realistic setup. These simulations can easily be correlated with biochemical experiments in the $Q\beta$-system of Biebricher and Eigen [4].
A recently developed mathematical model of a so-called *single-shape* landscape generalizes Eigen's concept. A *phenotypic error-threshold* can be characterized which describes the information-breakdown of replicating biopolymers [15, 28].

Catalytically active species: Self-replication of catalytically active species is usually modeled by higher order replicator equation. The famous classical approach of dynamical system is the *hypercycle* [9]. Motivation is a more natural model where various species do interact not only by competition for a given external resource, but also by predator-prey relations or symbiosis. The analogous to an ecosystem on a molecular level is an *autocatalytic network* consisting of self-replicating macromolecules which act as catalysts in replication reactions of other species or even catalyze their own replication [10, 22]. Although no *in vitro* system exhibiting this type of dynamic is know at present, the discovery of the catalytic potential of RNA, and especially the demonstration that RNA catalyzed template-induced synthesis of RNA is possible [6], suggests that an ecosystem of catalytically active RNAs which catalyze RNA replication might be feasible [26]. Recent experimental studies make evident that hypercyclic coupling in vitro is within reach [16]. In such a model, the fitness of a certain species will

[1]Author to whom correspondence should be addressed.

depend on the composition of the entire system, i.e. we have to replace the constant fitness value by a function of the population numbers of – in general – all species in the system.

Catalytic interaction between reacting species has been discussed recently in literature: Bagley *et al.* [2, 3] studied spontaneous emergence of metabolisms. Nuño *et al.* [1, 25, 24] suggested a catalytic network introduced through faulty self-replication into a mutant molecular species. Stadler *et al.* [34] investigated reaction-mutation networks and observed error-thresholds as first order phase transitions. Examples of low dimensional dynamical systems ($n = 2, 3$ and 4) are discussed and complete qualitative analysis is presented. Fontana *et al.* [12, 33] studied networks that are a generalization of replicator (or Lotka-Volterra[2]) equations. They model the dynamics of a population of object types whose binary interactions determine the specific type of interaction product. Similar studies about dynamics of programmable matter were performed by Rasmussen *et al.* [27].

Peter Schuster [30] proposed to incorporate a genotype phenotype mapping for the study of *evolutionary dynamics*. The scenario can be described by interactions of the following three processes (Fig. 1):

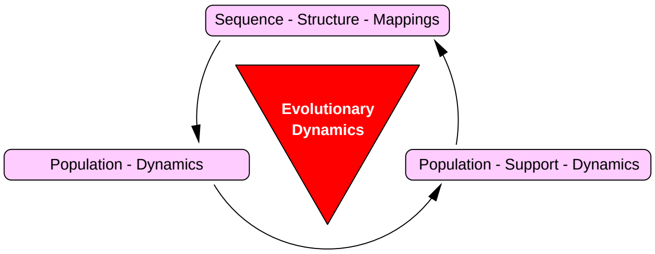

Figure 1: Evolutionary Dynamics

- *Genotype-Phenotype Mapping:* Modeling biological systems with selective forces and error-prone replication has to keep both genotypes and phenotypes in mind. In general mutations take place on the level of genotypes while selection acts on phenotypes. Thus it is the sequence to structure and further the structure to function mapping which provides the (kinetic) parameters for population dynamics.
- In *Population-Dynamics* the classic approach is to describe time-development of species or chemicals in a concentration-space by methods of chemical kinetics. Whenever a new entity emerges or a "weak" species dies out, the concentration-space adapts and changes it's dimensionality.
- *Population-Support-Dynamics* closes the cycle by establishing a relation to the genotype-phenotype-mappings. By exploring new regions of the sequence-space by migrating sets of genotypes it supplies genotype-phenotype-mapping with new input.

[2]According to Hofbauer [19] Replicator-equations with $n+1$ species are – apart from a transformation of the time scale – equivalent to a n-species Lotka-Volterra equation

2. Neutral Networks of RNA Secondary Structures

We begin by defining a RNA secondary structure [35]: A *secondary structure* is a vertex-labeled graph on n vertices with an adjacency matrix $A = (a_{i,k})_{1\le i,k\le n}$ fulfilling

(1) $a_{i,i+1} = 1$ for $1 \le i \le n-1$;
(2) For each i there is at most a single $k \ne i-1, i+1$ such that $a_{i,k} = 1$;
(3) If $a_{i,j} = a_{k,l} = 1$ and $i < k < j$ then $i < l < j$.

We call an edge (i,k), $|i-k| \ne 1$ a *bond* or *base pair* and write $[i,k] \in s$. A vertex i connected only to $i-1$ and $i+1$ shall be called *unpaired*. We shall denote the number of base pairs and the number of unpaired bases in a secondary structure s by $n_p(s)$ and $n_u(s)$ respectively. We denote the size of the alphabet by α and the number of distinct base pairs by β.

In this section we model the preimages of secondary structures –so called *neutral networks* – as random graphs. For this purpose we discuss properties of random induced subgraphs of generalized hypercubes.

2.1. Random induced Subgraphs of generalized Hypercubes

A generalized hypercube is a graph whose vertices are all n-tuples where the coordinates x_i are elements of the finite set $\mathcal{A}$ (of size α) and two vertices are adjacent if they differ in exactly one coordinate.
Each subset $X \subset \mathrm{v}[\mathcal{Q}_\alpha^n]$ induces the subgraph $\mathcal{Q}_\alpha^n[X]$ establishing a one-to-one correspondence between $X \subset \mathrm{v}[\mathcal{Q}_\alpha^n]$ and $\mathcal{Q}_\alpha^n[X]$. Let us denote the set of all induced subgraphs of $\mathcal{Q}_\alpha^n$ by $\mathcal{G}(\mathcal{Q}_\alpha^n)$. We suppose $0 \le \lambda \le 1$ to be given and set for $\Gamma_n \in \mathcal{G}(\mathcal{Q}_\alpha^n)$

$$\boldsymbol{\mu}_\lambda\{\Gamma_n\} \overset{\text{def}}{=\!=} \lambda^{|\,\mathrm{v}[\Gamma_n]\,|}\,(1-\lambda)^{\alpha^n - |\,\mathrm{v}[\Gamma_n]\,|}\,.$$

Obviously $\boldsymbol{\mu}_\lambda$ fulfills $\sum_{\Gamma_n} \boldsymbol{\mu}_\lambda\{\Gamma_n\} = 1$ and we thereby obtain the probability space

$$\Omega \stackrel{\text{def}}{=\!=} (\,\mathcal{G}(\mathcal{Q}_\alpha^n), \mathcal{P}(\mathcal{G}(\mathcal{Q}_\alpha^n)), \boldsymbol{\mu}_\lambda\,)\ .$$

For random induced subgraphs of generalized hypercubes there exists *one* threshold value λ^* for the density and the connectivity property. Here a subgraph $\Gamma_n < \mathcal{Q}_\alpha^n$ is called *dense* in $\mathcal{Q}_\alpha^n$ if and only if $\overline{\mathrm{v}[\Gamma_n]} = \mathrm{v}[\mathcal{Q}_\alpha^n]$. The main result on density reads

Theorem 1. *Let* $\lambda^* \stackrel{\text{def}}{=\!=} 1 - \sqrt[\alpha-1]{\alpha^{-1}}$ *then for* $\lambda > \lambda^*$ *holds*

$$\lim_{n\to\infty} \boldsymbol{\mu}_n\{\Gamma_n \textit{ is dense in } \mathcal{Q}_\alpha^n\,\} = 1$$

and for $\lambda < \lambda^*$ *we have*

$$\lim_{n\to\infty} \boldsymbol{\mu}_n\{\Gamma_n \textit{ is dense in } \mathcal{Q}_\alpha^n\,\} = 0\,.$$

For the connectivity property we have

Theorem 2. *Let* $(\mathcal{Q}_\alpha^n)$ *be a sequence of generalized hypercubes and* $\Gamma_n < \mathcal{Q}_\alpha^n$ *random induced subgraphs. Then*

$$\lim_{n\to\infty} \boldsymbol{\mu}_n\{\Gamma_n \text{ is connected}\} = \begin{cases} 1 & \text{for} \quad \lambda > 1 - \sqrt[\alpha-1]{\alpha^{-1}} \\ 0 & \text{for} \quad \lambda < 1 - \sqrt[\alpha-1]{\alpha^{-1}}\,. \end{cases} \tag{1}$$

A proof can be found in [29].

2.2. Compatibility and Construction of Neutral Nets

By definition of secondary structure its preimage is a subset of the *set of compatibles sequences.* A sequence x is said to be *compatible* to a secondary structure s if the nucleotides x_i and x_j at sequence positions i and j can pair whenever (i, j) is a base pair in s. Note that this condition does by no means imply that x_i and x_j will actually form a base pair in the structure $\varphi(x)$ obtained by some folding algorithm. The set of all sequences compatible with a secondary structure s will be denoted by $\mathbf{C}[s]$. There are two types of neighbors to sequence $x \in \mathbf{C}[s]$: each mutation in a position k which is unpaired in the secondary structure s leads again to a sequence compatible with s, while point mutations in the paired regions of s will in general produce sequences that are not compatible with s. This problem can be overcome by modifying the notion of neighborhood. If we allow the exchange base pairs instead of single nucleotides in the paired regions of s we always end up with sequences compatible with s. This definition of neighborhood allows us to view $x \in \mathbf{C}[s]$ as a graph. It can be shown [29] that this graph is the direct product of two generalized hypercubes

$$\mathbf{C}[s] = \mathcal{Q}_\alpha^{n_u} \times \mathcal{Q}_\beta^{n_p} \tag{2}$$

where n_u is the number of unpaired positions in s, α is the number of different nucleotides, i.e., $\alpha = 4$ in the case of natural RNAs, n_p is the number of base *pairs* in s, and β is the number of different *types* of base pairs that can be formed by the α different nucleotides; for natural RNAs we have $\beta = 6$. The sequence length is $n = n_u + 2n_p$.

We construct the *neutral network* w.r.t. s as a graph product of random induced subgraphs of $\mathcal{Q}_\alpha^{n_u}$ and $\mathcal{Q}_\beta^{n_p}$ respectively (corresponding to the unpaired and paired coordinates w.r.t. the secondary structure s) as illustrated in the figure below:

$$\begin{tikzcd} & \mathcal{Q}_\kappa^{n_u} \times \mathcal{Q}_\beta^{n_p} & \\ & \Gamma_{n_u} \times \Gamma_{n_p} \arrow[u, hook, "\iota"] & \\ \Gamma_{n_u} \arrow[ur, hook, "in"] & & \Gamma_{n_p} \arrow[ul, hook', "in"'] \end{tikzcd}$$

2.3. Complete Mappings

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. Thereby we obtain w.r.t. a given λ parameter a complete sequence to structure mapping. Formally speaking let $\mathbf{C}^* : \mathcal{S}_n \to \mathcal{P}(\mathrm{v}[\mathcal{Q}_\alpha^n])$ and $r : \mathcal{S}_n \to \mathbb{N}$ be two mappings such that $j \le i \implies r(s_j) \ge r(s_i)$.
A mapping $f : \mathcal{Q}_\alpha^n \to \mathcal{S}_n$ is called $\mathbf{C}^*$*-map* if and only if

$$(*) : \quad f(v) = s \quad \implies \quad v \in \mathbf{C}^*[s]$$

A mapping $f_r : \mathcal{Q}_\alpha^n \to \mathcal{S}_n$ is called $\mathbf{C}^*$*-random-map* if and only if f_r is given by

$$\begin{aligned} f_r^{-1}(s_0) &\stackrel{\text{def}}{=\!=} \Gamma_n[s_0] \\ f_r^{-1}(s_i) &\stackrel{\text{def}}{=\!=} \Gamma_n[s_i] \setminus \bigcup_{j<i} \left[\Gamma_n[s_i] \cap \Gamma_n[s_j]\right]\ . \end{aligned}$$

In particular any RNA folding map is a $\mathbf{C}^*$-map setting $\mathbf{C}^*[s] \stackrel{\text{def}}{=\!=} \mathbf{C}[s]$ since the neutral networks are constructed *a priori* in the graph of compatible sequences. We will assume that $\mathbf{C}^* = \mathbf{C}$ and then the recursion reads in the particular case of $\Gamma_n[s] = \mathbf{C}[s]$:

$$f^{-1}(s_{r_0}) \stackrel{\text{def}}{=\!=} \mathbf{C}[s_0]$$
$$f^{-1}(s_{r_i}) \stackrel{\text{def}}{=\!=} \mathbf{C}[s_i] \setminus \bigcup_{j<i} [\mathbf{C}[s_i] \cap \mathbf{C}[s_j]]$$

2.4. Coupling of Neutral Networks

In this section we will show that each two neutral networks being dense and connected come "close" in sequence space. To verify this we embed all secondary structures of fixed length n into the symmetric group S_n. Let S_n be the *symmetric group in n letters* and $\mathcal{S}$ be the set of all secondary structures. A *transposition* $\tau \in S_n$ is written as $\tau = (x_i, x_k)$. A base pair of a structure s is given by $[i, k]$. The set of all base pairs is denoted by $\Pi(s)$. Then

$$\imath : \mathcal{S} \to S_n; \qquad s \mapsto \imath(s) \stackrel{\text{def}}{=\!=} \prod_{[i,k]\in\Pi(s)} (i, k)\,.$$

Clearly the map $\imath$ is an *embedding* of the set $\mathcal{S}$ of secondary structure in the set S_n of permutations. It can be verified that $\imath(s)^2 = 1$, i.e., for any secondary structure the permutation assigned by $\imath$ is an *involution.* It is known from the theory [32] that any two involutions generate a *dihedral group*, D_m. Therefore $\imath$ in a natural way gives rise to the mapping

$$\jmath : \mathcal{S} \times \mathcal{S} \longrightarrow \{D_m < S_n\};$$
$$(s, s') \mapsto \jmath(s, s') \stackrel{\text{def}}{=\!=} \langle \imath(s), \imath(s')\rangle\,.$$

The key result of the above considerations is the following theorem

Theorem 3. **(Intersection–Theorem)** *Let s and s' be two arbitrary secondary structures. Then*

$$\mathbf{C}[s] \cap \mathbf{C}[s'] \neq \emptyset\,.$$

A proof can be found in [29].

That means for each pair of secondary structures we can always find sequences that are compatible to both structures. Hence any two connected and dense neutral networks come very close (Hamming distance $4 - (\lambda_1 + \lambda_2)$). The cyclic group $\langle \imath(s) \circ \imath(s')\rangle$ operates on the set of all 'positions' of the sequence $x = (x_1, \ldots, x_n)$. This operation induces a *cycle decomposition* of the generator $\imath(s) \circ \imath(s')$. In fact we can evaluate the coupling of the networks by the above cycle decomposition as follows: suppose a sequence on one network is given then how likely a sequences located on one network can move by point mutations to the other net. Using a simplified model the situation can be illustrated as shown in Fig. 2.

Figure 2: (Cf. [36]) By a simple replication-deletion process evolutionary dynamics of individuals are modeled. Two classes of individuals can occur (referring to two phenotypes or *bins*) and are transferred by an erroneous replication into each other. An unspecific dilution maintains a constant organization. In this figure the negative logarithm of the stationary distribution is plotted depending on the coupling rate between two bins (corresponding to two nets) and the possible states. For a strong coupling ≈ 1 a bistable situation occurs where either one or the other bin (i.e. neutral net) is completely occupied. For the decoupled case the most probable states of the distribution lie between those two extremas.

3. Hypercycles

3.1. Modeling Finite Populations in Sequence Space

Let us begin by mentioning a general concept for describing "populations" on graphs. For this purpose let N be a natural number such that $N \geq 2$ and let $\mathbf{V}$ be a (finite) family of vertices $(v_i \,|\, i \in \mathbb{N}_N)$ where $\{\, v_i \,|\, i \in \mathbb{N}_N \,\} \subset \mathrm{v}[\mathcal{Q}_\alpha^n]$. We shall call $\mathbf{V}$ a *population* in $\mathcal{Q}_\alpha^n$. The theory of point processes provides a powerful tool by identifying such a family $(v_i \,|\, i \in \mathbb{N}_N)$ with an integer valued measure:

$$\mathbf{V} = (v_i \,|\, i \in \mathbb{N}_N) \quad \longleftrightarrow \quad \phi \stackrel{\text{def}}{=\!=} \sum_{i=1}^{N} g_{v_i}, \tag{3}$$

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

A *replication-deletion process* is a mapping from $(v_i \,|\, i \in \mathbb{N}_N)$ to the family $(v'_i \,|\, i \in \mathbb{N}_N)$ as follows:
We select an ordered pair (v_l, v_k) where $v_l, v_k \in \{\, v_i \,|\, i \in \mathbb{N}_N \,\}$. For

$$\ell \stackrel{\text{def}}{=\!=} \mathrm{res}_{\mathrm{v}[\Gamma_n[s]]}\phi(\mathrm{v}[\Gamma_n[s]])$$

the first coordinate v_l is chosen with probability $\sigma\,\ell/[(N-\ell)+\sigma\ell]$ from $(v_i \,|\, i \in \mathbb{N}_N)$ where $\mathrm{res}_{\Gamma_n[s]}(v_i) \geq 1$ (the elements of $\mathbf{V}$ that are located on the neutral network) with uniform probability and from the remaining elements (with uniform probability) otherwise. The second coordinate of the above pair is selected with uniform probability on $(v_i \neq v_l \,|\, i \in \mathbb{N}_N)$ i.e. $1/(N-1)$. We assume the times $\hat{T}$ between these mappings to be exponentially distributed (scaled by the mean fitness)

$$\boldsymbol{\mu}\{\hat{T} \leq t\} = e^{-[(N-\ell)+\sigma\ell]\,t}\,.$$

We map $v_l = (x_1, ..., x_n)$ randomly into the vertex $v^* = (x'_1, ..., x'_n)$. This is done by mapping each coordinate x_i to a $x'_i \neq x_i$ with probability p 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 $(v_l, v_k) \mapsto (v_l, v^*)$. Thereby we obtain a "new" family by substituting the v_k by the v^*.

3.2. Hypercyclic Coupling

One important complex organization is the catalytic interaction of different phenotypes. A first and classical approach is to analyze cooperative behavior. The *hypercycle* [7, 9] serves as a paradigm for co-operation between different species. It is of particular relevance to study the stability of hypercycles against viruses and concurring hypercycles.

The dynamical behavior of the above system can be described by the following reaction equations. We denote I_i, $i = 1, \ldots, n$ as reacting species with a fixed phenotype $S(I_i)$, $f[S(I_i)]$ as fitness-function of phenotype $S(I_i)$, and W_{ij} indicates the probability of reproducing I_j by replicating I_i.:

$$\begin{aligned} I_i &\xrightarrow{f[S(I_i)]\cdot W_{ij}} I_i + I_j \\ I_i + I_k &\xrightarrow{g[S(I_i),S(I_k)]\cdot W_{ij}} I_i + I_j + I_k \\ I_i &\xrightarrow{\quad\Phi\quad} \emptyset \end{aligned} \tag{4}$$

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. 3):

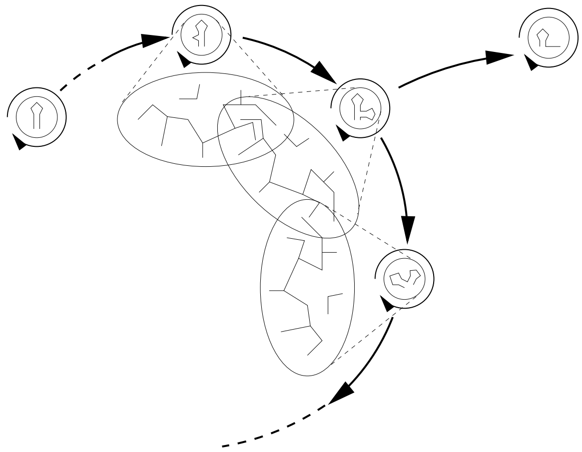

Figure 3: Population-Support-Dynamics of a hypercyclic organization with a parasite. Each phenotype — sketched by secondary structures — induce a corresponding compatible set and an embedded neutral net. The parasitic phenotype profits by a hypercyclic phenotype only without providing support.

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.

3.3. Hypercycles without Parasites

The stochastic process describing the hypercyclic coupled networks exhibits a more complex structure as shown in Fig. 4. We first investigated the time evolution of short hypercycles using the Gillespie algorithm [18]. The underlying stochastic process is closely related to the scheme we introduced in section 3.1 and can be seen as the stochastic analogue of the continuous flow reactor

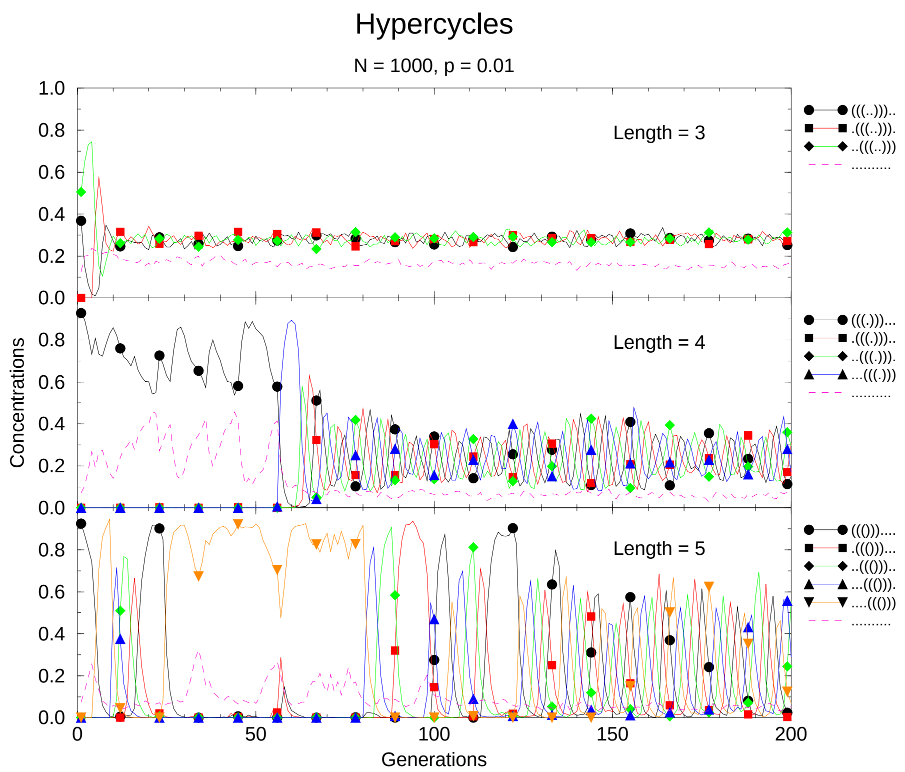

Figure 4: Time evolution of hypercycles of length $n = 3, 4, 5$. In case of infinite population size using theory of nonlinear differential equations there exist a stable inner fixed point for $n \leq 4$. For $n > 4$ this rest point becomes unstable and a stable limit-cycle emerges. Parameters for these simulations: population size = 1000, chain length = 10, reaction rates for catalyzed replication 10× faster than autocatalytic background replication.

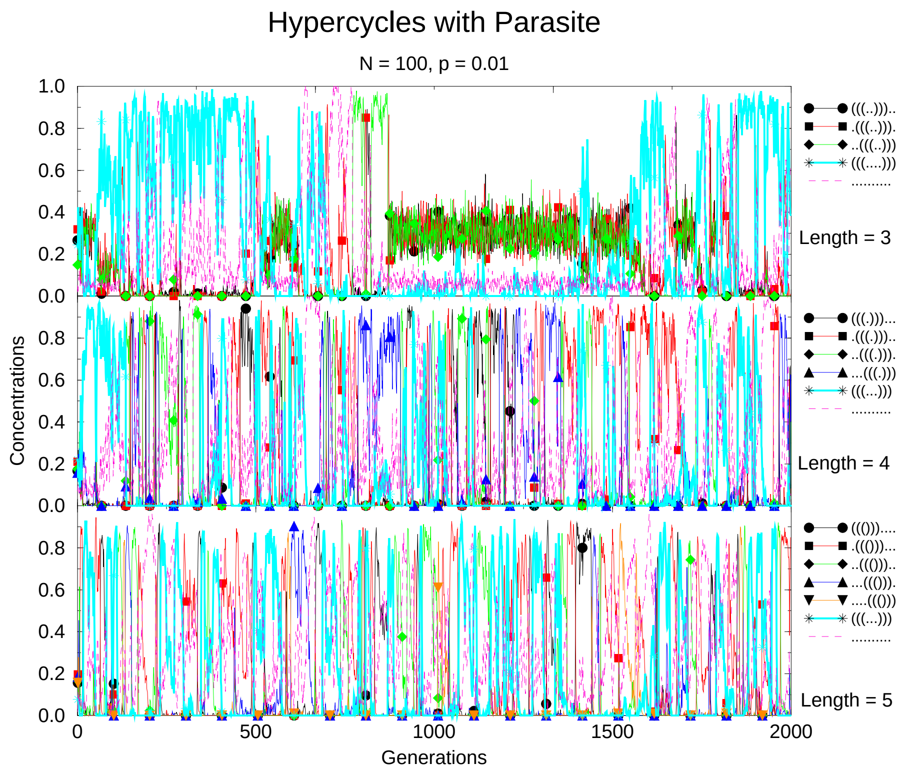

Figure 5: Time evolution of hypercycles with a parasite of length $n = 3, 4, 5$. Parameters for these simulations are the same as in Fig. 4 except for the population-size which is 100. The replication of the parasite is catalyzed by phenotype $\bullet$ with the same reaction rate as hypercyclic catalysis. Solid symbols denote members of the hypercycle, $*$ indicate the parasite, and the dashed line is the background.

usually modeled by differential equations. Phenomena like the transition between two neutral nets as monitored in section 2.4 show that it is adequate to model these systems fully stochastic. We were particularly interested to compare the time evolution of three- to five-phenotype hypercycles with the corresponding results from differential equation theory. The results are as follows:

For infinite populations we distinguish the following scenarios:

$n \leq 4$: There exists a stable inner fixed point.

$n > 4$: Via a supercritical Hopf-Bifurcation the inner rest point becomes instable and a stable limit-cycle emerges.

3.4. Hypercycles with Parasites

Interesting questions arise by introducing a competing parasite (Fig. 6). Recalling the results for hypercycles in case of infinite populations we would expect the following [21]: 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.

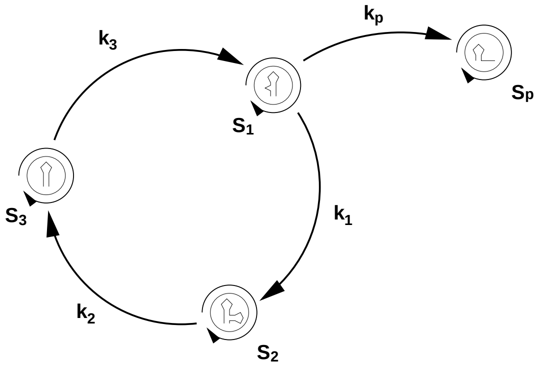

Figure 6: Hypercycle with three phenotypes and one parasite. S_i are the corresponding structures and k_i refer to the reaction rates of catalyzed replication $i = 1, 2, 3$, and p for the parasite.

Another approach has been studied by Boerlijst & Hogeweg [5] who investigated hypercycles with parasites using cellular automata and partial differential equations. On a two dimensional grid the authors observed spirals stabilizing the hypercycles versus the competing parasite. In the stochastic formulation we observe the following:

In contrast to homogeneous models with infinite population size the competing behavior between parasite and hypercycle is not a drastic *live and let die* scenario. Both competing systems can coexist for a long time. Due to the (transient) periodic behavior of hypercycles with four and more elements the parasitic concentration exhibits periodic variations which cannot be observed in deterministic, homogeneous systems (Fig. 5).

4. Discussion

The mathematical model of sequence to structure maps allows to study catalytic reaction networks on the level of RNA secondary structures in a realistic framework. Neutral nets with respect to secondary structures come close in sequence space and exhibit a topology that induces a diffusing populations. Transitions between these neutral nets of equal fitness have been observed. Further studies indicate that there is a strong evidence that these transition-phenomena also occur in more complex landscapes. By designing autocatalytic reaction networks as multi-layer dynamics we have two levels of dynamical systems:

- Level of Phenotypes
- Level of Genotypes

Both levels are correlated by a genotype-phenotype-mapping. Thus genotypic changes of the population will influence the characteristics of interactions between phenotypes in a certain extend. Due to the neutral properties of this mapping we will expect a “smooth” response and a “drifting” of catalytic networks in the space of genotypes.

In this paper we have studied hypercyclic coupling of autocatalytic phenotypes. The results obtained by computer simulations of the underlying stochastic process show a different behavior compared to the case of infinite population size. Using theory of nonlinear differential equations for $n \leq 3, 4$ there exists a stable inner fixed point and for $n = 5$ there exists a stable limit cycle [9]. The permanence of the limit cycle [20] is a nontrivial mathematical result and uses explicitly the compactness of the concentration simplex. The simulations of a corresponding stochastic process show substantial deviation from the predictions of the theory of differential equations. The computer experiments support that in the case of a hypercycle without parasites there is no convergence of random variables, whose states are the numbers of the population located on one particular neutral net, to a stationary distribution.

In case of hypercycles with parasites we observe that even when the parasite has a higher fitness the fraction of sequences located on neutral nets belonging to the hypercycle is significantly higher than a random one. There exists even a coexistence of parasite and hypercycle to a certain extent if the reaction rates of catalyzed reaction for both the hypercycle and the parasite are in the same magnitude.

Those preliminary results suggest a formulation of an analytical theory in terms of stochastic processes. As proposed by Peter Schuster [30] such a theory has to include the graph structure of the support —the neutral networks— and the sequence to structure mapping explicitly. In particular this theory could make nontrivial prediction on chemical reaction systems in biotechnology. Present day *in vitro* techniques are already capable to perform crucial parts of the work presented here. Recent experiments [17, 16] raise the hope that a system for powerful selection of catalysts are within reach of exploration.

Acknowledgments

Fruitful discussion with Walter Fontana, Michael Gebinoga and Peter Schuster are gratefully acknowledged. Special thanks to Jacqueline Weber who supplied unpublished material for Fig. 2. Part of the work has been done at the Santa Fe Institute, Santa Fe, NM in October 1995. Christian Forst thanks the institute for a great stay and the researchers for a stimulating atmosphere.

References

[1] M. A. Andrade, J. C. Nuño, F. Morán, F. Montero, and G. J. Mpitsos. Complex dynamics of a catalytic network having faulty replication into error-species. *Physica D*, 63:21–40, 1993.

[2] R. J. Bagley and J. D. Farmer. Spontaneous emergence of a metabolism. In C. G. Langton, C. Taylor, J. D. Farmer, and S. 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.

[3] R. J. Bagley, J. D. Farmer, and W. Fontana. Evolution of a metabolism. In C. G. Langton, C. Taylor, J. D. Farmer, and S. 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.

[4] C. K. Biebricher, M. Eigen, and W. C. Gardiner Jr. Kinetics of RNA replication: Competition and selection among self-replicating RNA species. *Biochemistry*, 24:6550–6560, 1985.

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

[6] J. A. Doudna and J. W. Szostak. RNA catalysed synthesis of complementary-strand RNA. *Nature*, 339:519–522, 1989.

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

[8] M. Eigen, J. McCaskill, and P. Schuster. The molecular Quasispecies. volume 75, pages 149 – 263. J. Wiley & Sons, 1989.

[9] M. Eigen and P. Schuster. *The Hypercycle: a principle of natural self-organization*. Springer, Berlin, 1979 (ZBP:23493).

[10] J. D. Farmer, N. H. Packard, and A. S. Perelson. The immune system, adaptation and machine learning. *Phsica*, 22 D:187–204, 1986.

[11] R. A. Fisher. *The genetical theory of natural selection*. Oxford: Clarendon Press., 1930.

[12] W. Fontana and L. W. Buss. "The Arrival of the Fittest": Towards a theory of biological organization. *Bull. Math. Biol.*, 1994. in press.

[13] W. Fontana, W. Schnabl, and P. Schuster. Physical aspects of evolutionary optimization and adaption. *Physical Review A*, 40(6):3301–3321, 1989.

[14] W. Fontana and P. Schuster. A computer model of evolutionary optimization. *Biophysical Chemistry*, 26:123–147, 1987.

[15] C. V. Forst, C. Reidys, and J. Weber. Evolutionary dynamics and optimization: Neutral Networks as model-landscape for RNA secondary-structure folding-landscapes. In F. Morán, A. Moreno, 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.

[16] M. Gebinoga. Hypercycles in biological systems. *J. Endocyt.*, 1995. submitted.

[17] M. Gebinoga and F. Oehlenschläger. Comparison of self-sustained sequence replication reaction systems. *Eur. J. Biochem.*, 1995. accepted.

[18] D. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Chem. Phys.*, 81:2340–2361, 1977.

[19] J. Hofbauer. On the occurance of limit cycles in the Lotka- Volterra equation. *Nonlinear Analysis*, 5:1003–1007, 1981.

[20] J. Hofbauer, J. Mallet-Paret, and H. Smith. Stable periodic solutions for the hypercycle system. *J. Dyn. Diff. Equ.*, 3(3):423, 1991.

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

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

[23] M. Nowak and P. 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.

[24] J. C. Nuño, M. A. Andrade, and F. Montero. Non-uniformities and superimposed competition in model of an autocatalytic network formed by error-prone selfreplicative species. *Bull. Math.Biol.*, 55:417–449, 1993.

[25] J. C. Nuño, M. A. Andrade, F. Morán, and F. Montero. A model of an autocatalytic network formed by error-rone self-replicative species. *Bull. Math. Biol.*, 55:385–415, 1993.

[26] S. Rasmussen. Toward a quantitative theory of the origin of life. In C. G. Langton, editor, *Artificial Life*, volume VI of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 79–104, Redwood City, 1989. Addison Wesley.

[27] S. Rasmussen, C. Knudsen, and R. Feldberg. Dynamics of programmable matter. In C. G. Langton, C. Taylor, J. D. Farmer, and S. 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.

[28] C. Reidys, C. V. Forst, and P. Schuster. Replication on neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 1995. almost submitted (since Oct. 1994).

[29] C. Reidys, P. F. Stadler, and P. Schuster. Generic properties of combinatory maps and neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 1995. submitted.

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

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

[32] J.-P. Serre. *Linear Representations of Finite Groups*. Springer, 1977.

[33] P. F. Stadler, W. Fontana, and J. H. Miller. Random catalytic reaction networks. *Physica D*, 63:378, 1993.

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

[35] 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.

[36] J. Weber. Transitions between two neutral networks. private communication, 1996.