Nature and Evolution of Early Replicons

PETER SCHUSTER[a,b,*] AND PETER F. STADLER[a,b]

[a]Institut für Theoretische Chemie und Strahlenchemie der Universität Wien,
Vienna, Austria

[b]Santa Fe Institute, Santa Fe, NM

[*]Mailing Address:
Institut für Theoretische Chemie und Strahlenchemie der Universität Wien
Währingerstraße 17, A-1090 Wien, Austria
Phone: +43 1 4277 527 42 Fax: +43 1 4277 527 93
E-Mail: pks@tbi.univie.ac.at

Abstract

RNA and protein molecules were found to be both templates for replication and specific catalysts for biochemical reactions. RNA molecules, although very difficult to obtain via plausible synthetic pathways under prebiotic conditions, are the only candidates for early replicons. Only they are obligatory templates for replication which can conserve mutations and propagate them to forthcoming generations. RNA based catalysts, called ribozymes, act with high efficiency and specificity on all classes of reactions involved in the interconversion of RNA molecules such as cleavage and template assisted ligation. The idea of an *RNA world* was conceived for a plausible prebiotic scenario of RNA molecules operating upon each other and constituting thereby a functional molecular organization. A theroretical account on molecular replication making precise the conditions under which one observes parabolic or exponential growth is presented. Exponential growth is observed in a protein assisted RNA world where plus-minus-($\pm$)-duplex formation is avoided by the action of an RNA replicase. Error propagation to forthcoming generations is analyzed in absence of selective neutral mutants as well as for predefined degrees of neutrality. A model of evolution is proposed that allows to deal explicitly with phenotypes.

1. Simple replicons and the origin of replication

A large number of successful experimental studies which tried to work out plausible chemical scenarios for the origin of early *replicons*, being molecules capable of replication, have been conducted in the past [60]. A sketch of such a possible sequence of events in prebiotic evolution is shown in figure 1. Most of the building blocks of present day biomolecules are available from different prebiotic sources, from extraterrestrial origins as well as from processes taking place in the primordial atmosphere or near hot vents in deep oceans. Condensation reactions and polymerization reactions formed non-instructed polymers, for example random oligopeptides of the protenoid type [33].

Template catalysis opens up the door to molecular copying and self-replication. Several small templates were designed by Julius Rebek and coworkers: These

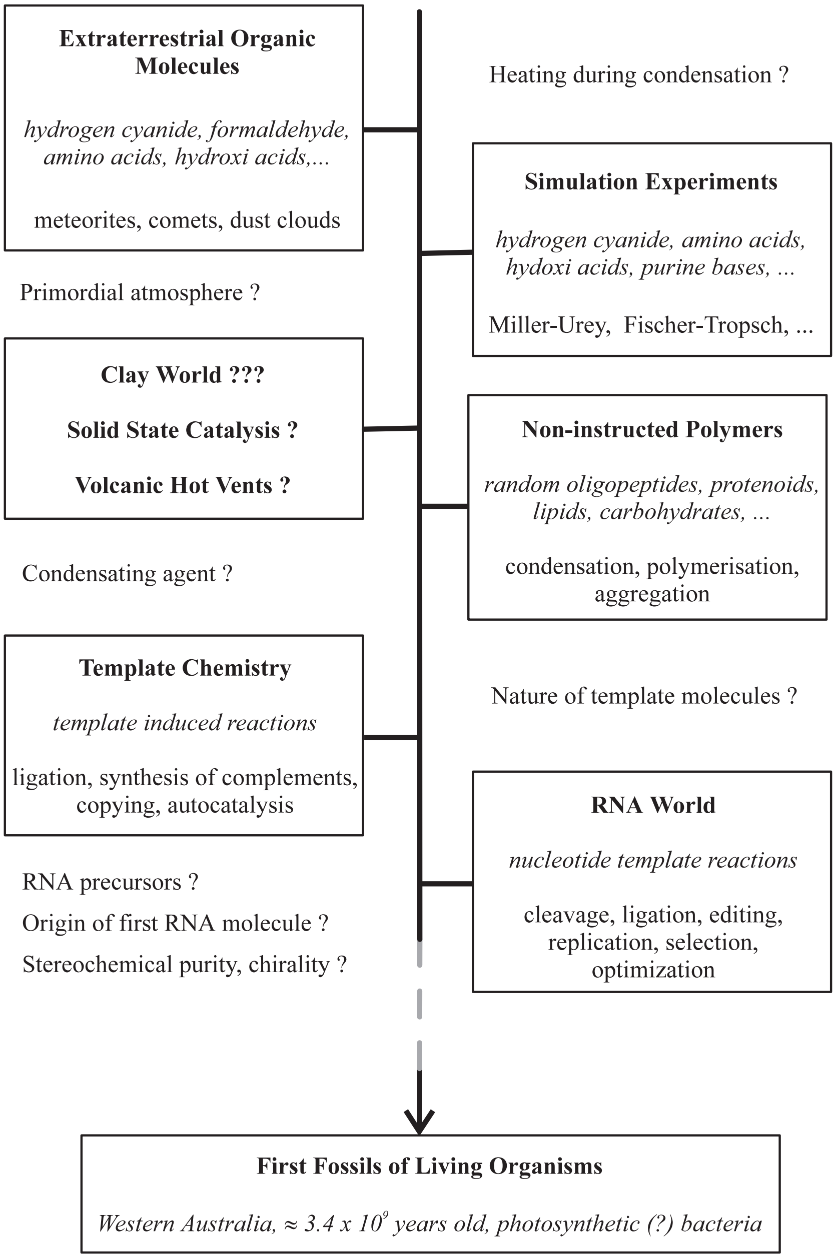


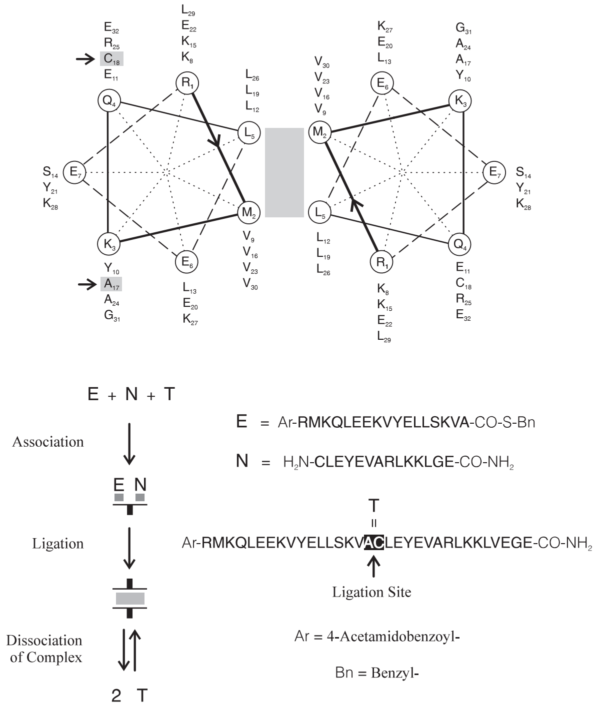
Figure 1: The RNA world. The concept of a precursor world preceeding present day genetics based on DNA, RNA and protein is based on the idea that RNA can act as both, storage of genetic information and specific catalyst for biochemical reactions. An RNA world in the first scenario on the route from prebiotic chemistry to present day organisms that allows for Darwinian selection and evolution. Problems and open questions are indicated by question marks. Little is known about further steps (not shown here explicitly) from early replicons to the first cells [22, 61].

molecules show indeed complementarity and undergo self-replication (see for example [68, 95]). Like nucleic acids they consist of a backbone whose role is to bring "molecular digits" in sterically appropriate positions, so that they can be read by their complements. Complementarity is also based on essentially the same principle as in nucleic acids: Specific patterns of hydrogen bonds allow to recognize complementary digits and to discriminate between "letters" of an alphabet. The hydrogen bonding pattern in these model replicons may be assisted by opposite electric charges carried by the complements. We shall encounter the same principle later in the discussion of Ghadiri's replicons based on stable coiled coils of oligopeptide α-helices [52]. Autocatalysis in small model systems is certainly interesting because it reveals some mechanistic details of molecular recognition. These systems are, however, are highly unlikely to be the basis of biologically significant replicons because they cannot be extended to large polymers in a simple and hence they are unsuitable for storing a sizeable amount of (sequence) information. Ligation of small pieces to larger units, on the other hand, is a source of combinatorial complexity providing sufficient capacity for information storage and, hence, evolution. Heteropolymer formation thus seems inevitable and we shall therefore focus on replicons which have this property: nucleic acids and proteins.

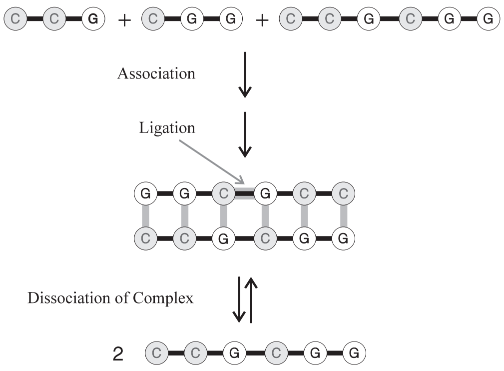
A first major transition leads from a world of simple chemical reaction networks to autocatalytic processes that are able to form self-organized systems which are capable of replication and mutation as required for Darwinian evolution. This transition can be seen as the interface between chemistry and biology since an early Darwinian scenario is tantamount to the onset of biological evolution. Two suggestions were made in this context: (i) autocatalysis arose in a network of reactions catalyzed by oligopeptides [49] and (ii) the first autocatalyst was a representative of a class of molecules with "obligatory" template function [17, 70]. The first suggestion works with molecules that are easily available under prebiotic conditions but lacks plausibility because the desired properties, conservation and propagation of mutants, are unlikely to occur with oligopeptides. The second concept suffers from opposite reasons: it is very hard to obtain the first nucleic acid like molecules but they would fulfill all functional requirements.

Until the eighties biochemists had an empirically well established but nevertheless prejudiced view on the natural and artificial functions of proteins and nucleic acids. Proteins were thought to be Nature's unbeatable universal catalysts, highly efficient as well as ultimately specific, and as in the case of immunoglobulins even tunable to recognize previously unseen molecules. After Watson and Crick's famous discovery of the double helix, DNA was considered as the molecule of inheritance, capable of encoding genetic information and sufficiently stable to allow for essential conservation of nucleotide sequences over many replication rounds. RNA's role in the molecular concert of Nature was reduced to the transfer of sequence information from DNA to protein, be it as mRNA or as tRNA. Ribosomal RNA and some rare RNA molecules did not fit well into this picture: Some sort of scaffolding functions were attributed to them such as holding supramolecular complexes together or bringing protein molecules into the correct spatial positions required for their functions.

This conventional picture was based on the idea of a complete "division of labor". Nucleic acids, DNA as well as RNA, were the templates, ready for replication and read-out of genetic information and but to do catalysis. Proteins were the catalysts and thus not capable of template function. In both cases these rather dogmatic views turned out to be wrong. Tom Cech and Sidney Altman discovered RNA molecules with catalytic functions [9, 10, 11, 37]. The name *ribozyme* was created for this new class of biocatalysts because they combine properties of ribonucleotides and enzymes (see section 2). Their examples were dealing with RNA cleavage reactions catalyzed by RNA: Without the help of a protein catalyst a non-coding region of an RNA transcript, a group I intron, cuts itself out during mRNA maturation. The second example concerns the enzymatic reaction of RNase P which catalyzes tRNA formation from the precursor poly-tRNA. For long time biochemists had known that this enzyme consists of a protein and an RNA moiety. It was tacitly assumed that the protein is the catalyst while the RNA component has only a backbone function. The converse, however, is true: The RNA acts as catalyst and the protein is merely a scaffol required for enhancing the efficiency.



A



B

Figure 2: Oligopeptide and oligonucleotide replicons. Part **A** shows an autocatalytic oligopeptide that makes use of the leucine zipper for template action. The upper part illustrates the stereochemistry of oligopeptide template-substrate interaction by means of the helix-wheel. The ligation site is indicated by arrows. The lower part shows the mechanism [52, 89]. Template-induced self-replication of oligonucleotides (part **B**; [98]) follows essentially the same reaction mechanism. The critical step is the dissociation of the dimer after bond formation which commonly prevents these systems from exponential growth and Dawinian behavior (see section 3).

The second prejudice was disproved only about two years ago by the demonstration that oligopeptides can act as templates for their own synthesis and thus show autocatalysis [52, 89, 53]. In this very elegant work, Reza Ghadiri and his coworkers have demonstrated that template action does not necessarily require hydrogen bond formation. Two smaller oligopeptides of chain lengths 17 (E) and 15 (N) are aligned on the template (T) by means of the hydrophobic interaction in a coiled coil of the leucine zipper type and the 32-mer is produced by spontaneous peptide bond formation between the activated carboxygroup and the free amino residue (figure 2). The hydrophobic cores of template and ligands consist of alternating valine and leucine residues and show a kind of knobs-into-holes type packing in the complex. The capability for template action of proteins is a consequence of the three-dimensional structure of the protein α-helix which allows the formation of coiled coils. It requires that the residues making the contacts between the helices fulfill the condition of space filling and thus stable packing. Modification of the oligopeptide sequences allows to alter the interaction in the complex and modifies thereby the specificity and efficiency of catalysis. A highly relevant feature of oligopeptide self-replication concerns easy formation of higher replication complexes: Coiled-coil formation is not restricted to two interacting helices, triple helices and higher complexes are known to be very stable too. Autocatalytic oligopeptide formation may thus involve not only a template and two substrates but, for example, a template and a catalyst that form a triple helix together with the substrates [89]. Only a very small fraction of all possible peptide sequences fold into three-dimensional structures that are suitable for leucine zipper formation and hence a given autocatalytic oligopeptide is very unlikely to retain the capability of template action on mutation. Peptides thus are *occasional* templates and replicons on peptid basis are rare.

In contrast to the volume filling principle of protein packing, specificity of catalytic RNAs is provided by base pairing and to a lesser extent by tertiary interactions. Both are the results of hydrogen bond specificity. Metal ions, in particular $\mathsf{Mg}^{2\oplus}$, are often involved in RNA structure formation and catalysis, too. Catalytic action of RNA on RNA is exercised in the cofolded complexes of ribozyme and substrate.

Since the formation of a ribozyme's catalytic center which operates on another RNA molecule requires sequence complementarity in parts of the substrate, ribozyme specificity is thus predominantly reflected by the sequence and not by the three-dimensional structure of the isolated substrate. Template action of nucleic acid molecules, being the basis for replication, results directly from the structure of the double helix. It requires an appropriate backbone provided by the antiparallel ribose-phosphate or 2'-deoxyribose-phosphate chains and a suitable geometry of the complementary purine-pyrimidine pairs. All RNA (and DNA) molecules, however, share these features which, accordingly, are independent of sequence. Every RNA molecule has a uniquely defined complement. Nucleic acid molecules, in contrast to proteins, are therefore *obligatory* templates. This impies that mutations are conserved and readily propagated into future generations.

Enzyme-free template-induced synthesis of longer RNA molecules from monomers, however, has not been successfully achieved so far (see e.g. [69]). A major problem, among others, is the dissociation of double stranded molecules at the temperature of efficient replication: If monomers bind with sufficiently high binding constants to the template in order to guarantee the desired accuracy of replication, then the new molecules are too sticky to dissociate after the synthesis has been completed. Autocatalytic template-induced synthesis of oligonucleotides from smaller oligonucleotide precursors was nevertheless successful: a hexanucleotide through ligation of two trideoxynucleotide precursors was carried out by Günter von Kiedrowski [98]. His system is the oligonucleotide analogue of the autocatalytic template-induced ligation of oligopeptides discussed above (figure 2). In contrast to the latter system the oligonucleotides do not form triple-helical complexes. Isothermal autocatalytic template-induced synthesis, however, cannot be used to prepare longer oligonucleotides because of the same duplex dissociation problem as mentioned for the template induced polymerization of monomers (see also section 3).

2. RNA catalysis and the RNA world

The natural ribozymes discovered early were all RNA cleaving molecules, the RNA moiety of RNase P [37], the class I introns [9] as well as the first small ribozyme

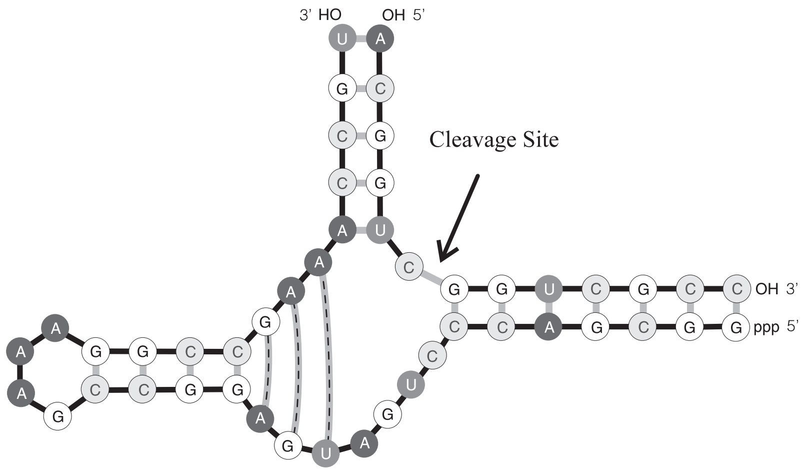


Figure 3: The hammerhead ribozyme. The substrate is a tridecanucleotide forming two double-helical stacks together with the ribozyme (n=34) in the cofolded complex [73]. Some tertiary interactions indicated by broken lines in the drawing determine the detailed structure of the hammerhead ribozyme complex and are important for the enzymatic reaction cleaving one of the two linkages between the two stacks. Substrate specificity of ribozyme catalysis is caused by the secondary structure in the cofolded complex between substrate and catalyst.

called "hammerhead" (figure 3) because of its characteristic secondary structure shape [96]. Three-dimensional structures are now available for three classes of RNA cleaving ribozymes [73, 87, 8, 29] and these data revealed the mechanism of RNA catalyzed cleavage reactions in full molecular detail. Additional catalytic RNA molecules were obtained through selection from random or partially random RNA libraries and subsequent evolutionary optimization (see section 6). RNA catalysis in non-natural ribozymes is not only restricted to RNA cleavage: Some ribozymes show ligase activity [4, 25] and many efforts were undertaken to prepare a ribozyme with full RNA replicase activity. The attempt that comes closest to the goal yielded a ribozyme that catalyzes RNA polymerization in short stretches [24]. RNA catalysis is not restricted to operate on RNA, nor do nucleic acid catalysts require the ribose backbone: Ribozymes were trained by evolutionary techniques to process DNA rather than their natural RNA substrate [5], and catalytically active DNA molecules were evolved as well [7, 12]. Polynucleotide kinase activity has been reported [56, 57] as well as self-alkylation of RNA on nitrogen [103].

Systematic studies also revealed examples of RNA catalysis on non-nucleic acid substrates. RNA catalyzes ester, amino acid, and peptidyl transferase reactions [55, 46, 105]. The latter examples are particularly interesting because they revealed close similarities between the RNA catalysis of peptide bond formation and ribosomal peptidyl-transfer [106]. A spectacular finding in this respect was that oligopeptide bond cleavage and formation is catalyzed by ribosomal RNA and not by protein: More than 90% of the protein fraction can be removed from ribosomes without loosing the catalytic effect on peptide bond formation [66, 36]. In addition, ribozymes were prepared that catalyze alkylation on sulfur atoms [100] and, finally, RNA molecules were designed which are catalysts for typical reactions of organic chemistry, for example an isomerization of biphenyl derivatives [74].

For two obvious reasons RNA was chosen as candidate for the leading molecule in a simple scenario at the interface between chemistry and biology: (i) RNA is thought to be capable of storing retrievable information because it is an obligatory template and (ii) it has catalytic properties. Although the catalytic properties of RNA are less universal than those of proteins, they are apparently sufficient for processing RNA. RNA molecules operating on RNA molecules form a self-organizing system that can develop a form of molecular organization with emerging properties and functions. This scenario has been termed the *RNA world* (see e.g. [35, 48] as well as the collective volume by Gesteland and Atkins [34]). The idea of an RNA world turned out to be fruitful in a different aspect too: It initiated the search for molecular templates and created an entirely new field which may be characterized as *template chemistry* [71]. Series of systematic studies were performed, for example, on the properties of nucleic acids with modified sugar moieties [27]. These studies revealed the special role of ribose and provided explanations why this molecule is basic to all life processes.

Chemists working on the origin of life see a number of difficulties for an RNA world being a plausible direct successor of the functionally unorganized prebiotic chemistry (see figure 1 and the reviews [70, 48, 71, 86]): (i) no convincing prebiotic synthesis has been demonstrated for all RNA building blocks, (ii) materials for successful RNA synthesis require a high degree of purity that can hardly be achieved under prebiotic conditions, (iii) RNA is a highly complex molecule whose stereochemically correct synthesis (3'-5' linkage) requires an elaborate chemical machinery, and (iv) enzyme-free template-induced synthesis of RNA molecules from monomers has not been achieved so far. In particular, the dissociation of duplexes into single strands and the optical asymmetry problem are of major concern. Template induced synthesis of RNA molecules requires pure optical antipodes. Enantiomeric monomers (containing L-ribose instead of the natural D-ribose) are "poisons" for the polycondensation reaction on the template since their incorporation causes termination of the polymerization process. Several suggestions postulating more "intermediate worlds" between chemistry and biology were made. Most of the intermediate information carriers were thought to be more primitive and easier to synthesize than RNA but nevertheless still having the capability of template action [86]. Glycerol, for example, was suggested as a substitute for ribose because it is structurally simpler and it lacks chirality. However, no successful attempts to use such less sophisticated backbone molecules together with the natural purine and pyrimidine bases for template reactions have been reported so far.

Starting from a world of replicating molecules, it took a series of many not yet well-understood steps [23] to arrive at the first organisms that formed the earliest identified fossils (Warrawoona, Western Australia, 3.4×10^9 years old, [78]) and possibly the even older kerogen found in the Isua formation (Greenland, 3.8×10^9 years old, [72, 77]), see (figure 1. It has been speculated that functionally correlated RNA molecules have developed a primitive translation machinery based on an early genetic code. After such a relation between RNA and proteins had been established the stage was set for concerted evolution of proteins and RNA. Proteins may induce vesicle formation into lipid-like materials and eventually lead to the formation of compartments. After a number of steps such an ensemble might have developed a primitive metabolism and thus led to the first protocells [23]. DNA being now the backup copy of genetic information is seen as a late comer in prebiotic evolution.

A successful experimental approach to self-reproduction of micelles and vesicles is highlighting one of the many steps enumerated above: prebiotic formation of vesicle structures [3]. The basic reaction leading to autocatalytic production of amphiphilic materials is the hydrolysis of ethyl caprilate. The combination of vesicle formation with RNA replication represents a particularly important step towards the construction of a kind of minimal synthetic cell [58]. Despite these elegant experimental studies and the attempts to build comprehensive models satisfactory answers to the problems of compartment formation and cell division are not at hand yet.

3. Parabolic and exponential growth

It is relatively easy to derive a kinetic rate equation displaying the elementary behavior of replicons if one assumes that catalysis proceeds through the complementary binding of reactant(s) to free template and that autocatalysis is limited by the tendency of the template to bind to itself as an inactive "product inhibited" dimer [99]. However, in order to achieve an understanding of what is likely to happen in systems where there is a diverse mixture of reactants and catalytic templates, it is desirable to develop a comprehensive kinetic description of as many individual steps in the reaction mechanism of template synthesis as is feasible and tractable from the mathematical point of view.

Szathmáry [93] over-simplified the resulting dynamics to a simple parabolic growth law $\dot{x}_k \propto x_k^p$, $0 < p < 1$ for the concentrations of the interacting template species. His model suffers from a conceptual and a technical problem: (i) Under no circumstances does one observe extinction of a species in any parabolic growth model, and (ii) the vector fields are not Lipschitz-continuous on the boundary of the concentration simplex, indicating that we cannot expect a physically reasonable behavior in this area.

In a recent paper [102] we have derived the kinetic equations of a system of coupled template-instructed ligation reactions of the form

$$A_i + B_j + C_{kl} \underset{\bar{a}_{ijkl}}{\overset{a_{ijkl}}{\rightleftarrows}} A_iB_jC_{kl} \overset{b_{ijkl}}{\longrightarrow} C_{ij}C_{kl} \underset{\bar{d}_{ijkl}}{\overset{d_{ijkl}}{\rightleftarrows}} C_{ij} + C_{kl} \tag{1}$$

Here $A_{.}$ and $B_{.}$ denote the two substrate molecules which are ligated on the template $C_{..}$, for example, the electrophilic, E, and the nucleophilic, N, oligopeptide in peptide template reactions or the two different trinucleotides, GGC and GCC, in the autocatalytic hexanucleotide formation (figure 2). This scheme thus encapsulates the experimental results on both peptide and nucleic acid replicons [52, 98].

The following assumptions are straightforward and allow for a detailed mathematical analysis:

(i) the concentrations of the intermediates are stationary in agreement with the "quasi-steady-state" approximation [88],

(ii) the total concentration c_0 of all replicating species is constant in the sense of *constant organization* [17],

(iii) the formation of hetero-duplices of the form $C_{ij}C_{kl}$, $ij \neq kl$ is neglected, and

(iv) only reaction complexes of the form $A_kB_lC_{kl}$ lead to ligation.

Assumptions (iii) and (iv) are closely related. They make immediate sense for hypothetical macromolecules for which the template instruction is direct instead of complementary. It has been shown, however, that the dynamics of complementary replicating polymers is very similar to direct replication dynamics if one considers the two complementary strands as "single species" by simply adding their concentrations [91].

Assumptions (iii) and (iv) suggest a simplified notation of the reaction scheme:

$$A_k + B_k + C_k \underset{\bar{a}_k}{\overset{a_k}{\rightleftarrows}} A_kB_kC_k \xrightarrow{b_k} C_kC_k \underset{\bar{d}_k}{\overset{d_k}{\rightleftarrows}} 2C_k \tag{2}$$

It can be shown that equ.(2) together with the assumptions (i) and (ii) leads to the following system of differential equations for the frequencies or relative total concentrations x_k, i.e., $\sum_k^M x_k = 1$ of the template molecules C_k in the system (Note that x_k accounts not only for the free template molecules but also for those bound in the complexes C_kC_k and $A_kB_kC_k$):

$$\dot{x}_k = x_k \left(\alpha_k \varphi(\beta_k x_k) - \sum_j^M \alpha_j x_j \varphi(\beta_j x_j) \right) \quad , \quad k = 1, \ldots, M \ , \tag{3}$$

where

$$\varphi(z) = \frac{1}{z}\,\left(\sqrt{z+1}-1\right)\,, \qquad \varphi(0) = \frac{1}{2}\,. \tag{3'}$$

and the effective kinetic constants α_k and β_k can be expressed in terms of the physical parameters a_k, $\bar{a}_k$, etc. It will turn out that survival of replicon species is determined by the constants α_k which we characterize therefore as Darwinian fitness parameters.

Equation (3) is a special form of a replicator equation with the non-linear response functions $f_k(x) := \alpha_k \varphi(\beta_k x_k)$. Its behavior depends strongly on the values of β_k: For large values of z we have $\varphi(z) \sim 1/\sqrt{z}$. Hence equ.(3) approaches Szathmáry's expression [93]

$$\dot{x}_k = h_l \sqrt{x_k} - x_k \sum_j^M h_j \sqrt{x_j}$$

with suitable constants h_k. This equation exhibits a very simple dynamics: the mean fitness $\Phi(x) = \sum_j^M h_j \sqrt{x_j}$ is a Ljapunov function, i.e., it increases along all trajectories, and the system approaches a globally stable equilibrium at which all species are present [102, 97]. Szathmáry's parabolic growth model thus does not lead to selection.

On the other hand, if z remains small, that is, if β_k is small, then $\varphi(\beta_k x_k)$ is almost constant $1/2$ (since the relative concentration x_k is of course a number between 0 and 1). Thus we obtain

$$\dot{x}_k = \frac{1}{2} x_k \left(\alpha_k - \sum_j^M \alpha_j x_j \right) \tag{4}$$

which is the "no-mutation" limit of Eigen's kinetic equation for replication [17]. (If condition (iv) above is relaxed, we in fact arrive at Eigen's model with a mutation term). Equ.(4) leads to survival of the fittest: The species with the largest value of α_k will eventually be the only survivor in the system. It is worth noting that the mean fitness also increases along all orbits of equ.(4) in agreement with the no-mutation case [85].

The constants β_k that determine whether the system shows Darwinian selection or unconditional coexistence is proportional to the total concentration c_0 of the templates. For small total concentration we obtain equ.(4), while for large concentrations, when the formation of the dimers C_kC_k becomes dominant, we enter the regime of parabolic growth.

Equ.(3) is a special case of a class of replicator equations studied in [42]. Restating the previously given result yields the following: All orbits or trajectories starting from physically meaningful points (these are points in the interior of the simplex S_M with $x_j > 0$ for all $j = 1, 2, \ldots, M$) converge to a unique equilibrium point $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M)$ with $\bar{x}_i \geq 0$, which is called the ω-limit of the orbits. This means that species may go extinct in the limit $t \to \infty$. If $\bar{\mathbf{x}}$ lies on the surface of S_M (which is tantamount to saying that at least one component $\bar{x}_j = 0$) then it is also the ω-limit for all orbits on this surface. If we label the replicon species according to decreasing values of the Darwinian fitness parameters, $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_M$, then there is an index $\ell \geq 1$ such that $\bar{\mathbf{x}}$ is of the form $\bar{x}_i > 0$ if $i \leq \ell$ and $\bar{x}_i = 0$ for $i > \ell$. In other words, ℓ replicon species survive and the $M - \ell$ least efficient replicators die out. This behaviour is in complete analogy to the reversible exponential competition case [84] where the Darwinian fitness parameters α_k are simply the rate constants a_k. If the smallest concentration dependent value $\beta_s(c_0) = \min\{\beta_j(c_0)\}$ is sufficiently large, we find $\ell = M$ and no replicon goes extinct ($\bar{\mathbf{x}}$ is an interior equilibrium point).

The condition for survival of species k is explicitly given by:

$$\alpha_k > 2\Phi(\bar{\mathbf{x}})\,.$$

It is interesting to note that the Darwinian fitness parameters α_k determine the order in which species go extinct whereas the concentration dependent values $\beta_k(c_0)$ collectively influence the flux term and hence set the "extinction threshold". In contrast to Szathmáry's model equation the extended replicon kinetics leads to both competitive selection and coexistence of replicons depending on total concentration and kinetic constants.

4. Molecular evolution experiments

In the first half of this century it was apparently out of question to do conclusive and interpretable experiments on evolving populations because of two severe problems: (i) Time scales of evolutionary processes are prohibitive for laboratory investigations and (ii) the numbers of possible genotypes are outragingly large and thus only a negligibly small fraction of all possible sequences can be realized and evaluated by selection. If generation times could be reduced to a minute or less, thousands of generations, numbers sufficient for the observation of optimization and adaptation, could be recorded in the laboratory. Experiments with RNA molecules in the test-tube fulfil indeed this time scale criterion for observability. With respect to the "combinatorial explosion" of the numbers of possible genotypes the situation is less clear. Population sizes of nucleic acid molecules of 10^{15} to 10^{16} individuals can be produced by random synthesis in conventional automata. These numbers cover roughly all sequences up to chain lengths of $n = 27$ nucleotides. These are only short RNA molecules but their length is already sufficient for specific binding to predefined target molecules, for example antibiotics [47]. In addition, sequence to structure to function mappings of RNA are highly redundant and thus only a small fraction of all sequences has to be searched in order to find solutions to given evolutionary optimization problems [30, 82].

The first successful attempts to study RNA evolution *in vitro* were carried out in the late sixties by Sol Spiegelman and his group [64, 90]. They created a "protein assisted RNA replication medium" by adding an RNA replicase isolated from *E. coli* cells infected by the RNA bacteriophage Qβ to a medium for replication that also contains the four ribonucleoside triphosphates (GTP, ATP, CTP, and UTP) in a suitable buffer solution. Qβ RNA and some of its smaller variants start instantaneously to replicate when transferred into this medium. Evolution experiments were carried out by means of the serial transfer technique: Materials consumed in RNA replication are replenished by transfer of small samples of the current solution into fresh stock medium. The transfers were made after equal time steps. In series of up to one hundred transfers the rate of RNA synthesis increased by orders of magnitude. The increase in the replication rate occurs in steps and not

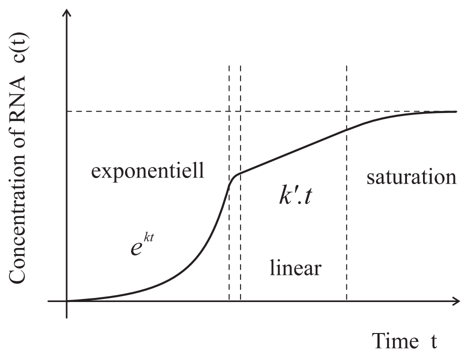


Figure 4: Replication kinetics of RNA with Qβ replicase. In essence, three different phases of growth are distinguished: (i) exponential growth under conditions with excess of replicase, (ii) linear growth when all enzyme molecules are loaded with RNA, and (iii) a saturation phase that is caused by product inhibition.

continuously as one might have expected. Analysis of the molecular weights of the replicating species showed a drastic reduction of the RNA chain lengths during the series of transfers: The initially applied Qβ RNA was 4220 nucleotides long and the finally isolated species contained little more than 200 bases. What happened during the serial transfer experiments was a kind of degradation due to suspended constraints on the RNA molecule. In addition to perform well in replication the viral RNA has to code for four different proteins in the host cell and needs also a proper structure in order to enable packing into the virion. In test-tube evolution these constraints are released and the only remaining requirement is recognition of the RNA by Qβ replicase and fast replication.

Evidence for a non-trivial evolutionary process came a few years later when the Spiegelman group published the results of another serial transfer experiment that gave evidence for adaptation of an RNA population to environmental change. The replication of an optimized RNA population was challenged by the addition of ethidium bromide to the replication medium [51]. This dye intercalates into DNA and RNA double helices and thus reduces replication rates. Further serial transfers in the presence of the intercalating substance led to an increase in the replication rate until an optimum was reached. A mutant was isolated from the optimized population which differed from the original variant by three point mutation. Extensive studies on the reaction kinetics of RNA replication in the Qβ

replication assay were performed by Christof Biebircher in Göttingen [6]. These studies revealed consistency of the kinetic data with many-step reaction mechanism. Depending on concentration the growth of template molecules allows to distinguish three phases of the replication process: (i) at low concentration all free template molecules are instantaneously bound by the replicase which is present in excess and therefore the template concentration grows exponentially, (ii) excess of template molecules leads to saturation of enzyme molecules, then the rate of RNA synthesis becomes constant and the concentration of the template grows linearly, and (iii) very high template concentrations impede dissociation of the complexes between template and replicase, and the template concentration approaches a constant in the sense of product inhibition. We neglect plus-minus complementarity in replication by assuming stationarity in relative concentrations of plus and minus strand [17] and consider the plus-minus ensemble as a single species. Then, RNA replication may be described by the over-all mechanism:

$$A + I_i + E \underset{\bar{k}_i}{\overset{k_i}{\rightleftarrows}} A + I_i \cdot E \xrightarrow{a_i} I_i \cdot E \cdot I_i \underset{\bar{k}_i'}{\overset{k_i'}{\rightleftarrows}} I_i \cdot E + I_i \ . \qquad (5)$$

Here E represents the replicase and A stands for the low molecular weight material consumed in the replication process. This simplified reaction scheme reproduces all three characteristic phases of the detailed mechanism (figure 4) and can be readily extended to replication and mutation.

Despite the apparent complexity of RNA replication kinetics the mechanism at the same time fulfils an even simpler over-all rate law provided the activated monomers, ATP, UTP, GTP, and CTP, as well as Qβ replicase are present in access. Then, the rate of increase for the concentration x_i of RNA species I_i follows the simple relation, $\dot{x}_i \propto x_i$, which in absence of constraints ($\Phi = 0$) leads to exponential growth. This growth law is identical to that found for asexually reproducing organisms and hence replication of molecules in the test-tube leads to the same principal phenomena that are found with evolution proper. RNA replication in the Qβ system requires specific recognition by the enzyme which implies sequence and structure restrictions. Accordingly only RNA sequences that fulfil these criteria can be replicated. In order to be able to amplify RNA free of such constraints many-step replication assays have been developed. The discovery of the DNA polymerase chain reaction (PCR) [65] was a milestone towards sequence independent amplification of DNA sequences. It has one limitation: double helix separation requires higher temperatures and conventional PCR works with a temperature program therefore. PCR is combined with reverse transcription and transcription by means of bacteriophage T7 RNA polymerase in order to yield a sequence independent amplification procedure for RNA. This assay contains two possible amplification steps: PCR and transcription. Another frequently used assay makes use of the isothermal self-sustained sequence replication reaction of RNA (3SR) [28]. In this system the RNA-DNA hybrid obtained through reverse transcription is converted into single stranded DNA by RNAse-digestion of the RNA strand, instead of melting the double strand. DNA double strand synthesis and transcription complete the cycle. Here, transcription by T7 polymerase represents the amplification step. Artificially enhanced error rates needed for the creation of sequence diversity in population can be achieved readily with PCR. Reverse transcription and transcription are also susceptible to increase of mutation rates. These two and other new techniques for RNA amplification provided universal and efficient tools for the study of molecular evolution under laboratory conditions and made the usage of viral replicases with their undesirable sequence specificities obsolete.

5. Error propagation and quasispecies

Evolution of molecules based on replication and mutation exposed to selection at constant population size has been formulated and analyzed in terms of chemical reaction kinetics [17, 20, 19]. Error-free replication and mutation are parallel chemical reactions,

$$A + I_i \stackrel{a_i Q_{ij}}{\longrightarrow} I_j \, + \, I_i \ , \tag{6}$$

and form a network which in principle allows to form every RNA genotype as a mutant of any other genotype. The materials required for or consumed by RNA synthesis, again denoted by A, are replenished by continuous flow in a reactor resemblimg a chemostat for bacterial cultures (figure 5). The object of interest is now the distribution of genotypes in the population and its time dependence. We present here a short account of the most relevant features of such replication-mutation assays, in particular the existence of thresholds in error propagation.

Selection in populations is described by ordinary differential equations. It has been shown for systems of type (6) that the outcome of selection is independent of the selection constraint applied. In particular, the flow reactor and constant organization yield essentially the same results [84, 39] and thus we used the latter simpler condition without loosing generality. Variables are again the frequencies of individual genotypes, x_i measuring that of genotype or RNA sequence I_i. The frequencies are nomalized, $\sum_{i=1}^{M} x_i = 1$ (due to constant organization), the population size is denoted by N, and the number of different genotypes by M. The time dependence of the sequence distribution is described by the kinetic equation

$$\dot{x}_i \;=\; x_i\left(a_i Q_{ii} - \bar{E}(t)\right) + \sum_{j=1, j\neq i}^{M} a_j Q_{ji} x_j \;, \quad i \;=\; 1, \ldots, M \;. \tag{7}$$

The rate constants for replication of the molecular species are a_i. Once a reaction has been initiated it can lead to a correct copy, $\mathrm{I}_i \rightarrow \mathrm{I}_i$, or to a mutant, $\mathrm{I}_i \rightarrow \mathrm{I}_j$. The frequencies of the individual reaction channels are contained in the mutation matrix $Q \doteq \{Q_{ij};\, i,j = 1, \ldots, M\}$, in particular the fraction of error copies of genotype I_i falling into species I_j is given by Q_{ij} and thus we have $\sum Q_{ij} = 1$. The diagonal elements of Q are the replication accuracies, i.e., the fractions of correct replicas produced on the corresponding templates. The time dependent excess productivity which is compensated by the flow in the reactor is the mean value $\bar{E}(t) = \sum a_i x_i(t)$. The quantities determining then the outcome of selection are the products of replication rate constants and mutation frequencies subsumed in the value matrix: $W \doteq \{w_{ij} = a_i Q_{ij}\,;\, i,j = 1, \ldots M\}$, its diagonal elements, w_{ii}, were called the selective values of the individual genotypes.

The selective value of a genotype is tantamount to its fitness in the case of vanishing mutational backflow and hence the genotype with maximal selective value, I_m,

$$w_{mm} \;=\; \max\{w_{ii} \,|\, i = 1, \ldots, M\} \;, \tag{8}$$

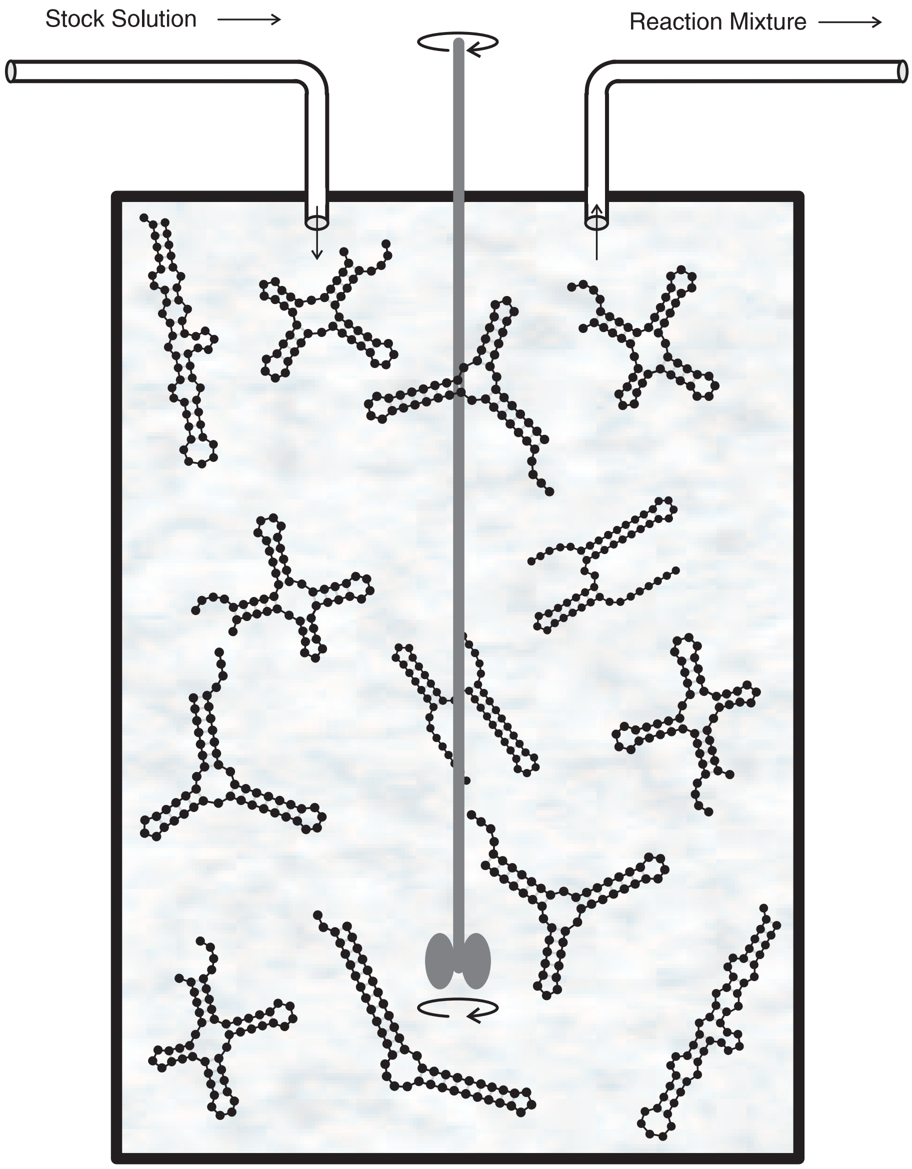


Figure 5: A flow reactor for the evolution of RNA molecules. A stock solution containing all materials for RNA replication including an RNA polymerase flows continuously into a well stirred tank reactor and an equal volume containing a fraction of the reaction mixture leaves the reactor. The population in the reactor fluctuates around a mean value, $N \pm \sqrt{N}$. RNA molecules replicate and mutate in the reactor, and the fastest replicators are selected. The RNA flow reactor has been used also as an appropriate setup for computer simulations [45, 31, 32]. There, other criteria than fast replication can be used for selection. For example, fitness functions are defined that measure the distance to a predefined target structure and fitness increases during the approach towards the target [45, 32].

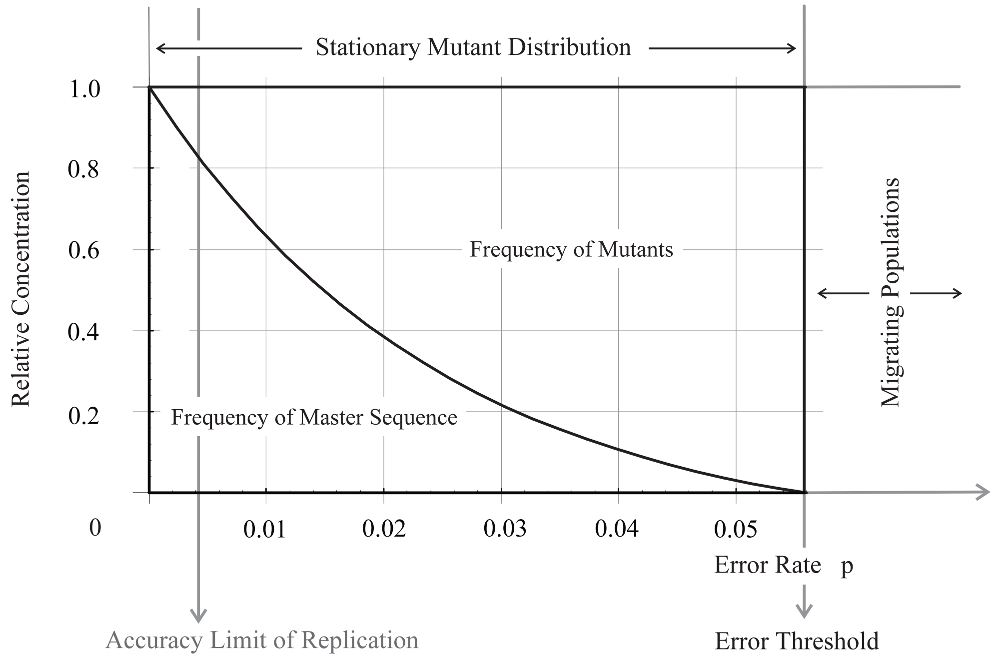


Figure 6: The genotypic error threshold. The fraction of mutants in stationary populations increases with the error rate p. Stable stationary mutant distributions called quasispecies require sufficient accuracy of replication: the single digit accuracy has to exceed a minimal value known as error threshold, $1-p=q>q_{\min}$. Above threshold populations migrate through sequence space in random walk like manner [45]. There is also a lower limit to replication accuracy which is given by the maximum accuracy of the replication machinery.

dominates a population after it has reached the selection equilibrium and hence it is called the *master* sequence. The notion *quasispecies* was introduced for the stationary genotype distribution in order to point at its role as the genetic reservoir of the population.

A simple expression for the stationary frequency can be found, if the master sequence is derived from the single peak model landscape that assigns a higher replication rate to the master and identical values to all others, for example $a_m = \sigma_m \cdot a$ and $a_i = a$ for all $i \neq m$ [92, 94, 1]. The (dimensionless) factor σ_m is called the superiority of the master sequence. The assumption of a single peak landscape is tantamount to lumping all mutants together into a mutant cloud with average fitness. The probability to be in the cloud is simply $x_c = \sum_{j=1,j\neq m}^{M} x_j = 1 - x_m$

and the replication-mutation problem boils down to an exercise in a single variable, x_m, the frequency of the master. The single peak model can be interpreted as a kind of mean field approximation since the mutant cloud is characterizable by "mean-except-the-master" properties, for example by the mean-except-the-master replication rate constant $\bar{a} = \sum_{j \neq m} a_j x_j / (1 - x_m)$. The superiority then reads: $\sigma_m = a_m / \bar{a}$. Neglecting mutational backflow we can readily compute the stationary frequency of the master sequence:

$$\bar{x}_m = \frac{a_m Q_{mm} - \bar{a}}{a_m - \bar{a}} = \frac{\sigma_m Q_{mm} - 1}{\sigma_m - 1} . \tag{9}$$

In this expression the master sequence vanishes at some finite replication accuracy, $Q_{mm} \big|_{\bar{x}_m = 0} = Q_{\min} = \sigma_m^{-1}$. Non-zero frequency of the master thus requires $Q_{mm} > Q_{\min}$. We introduce the uniform error rate model, which assumes that the mutation rate is p per site and replication event independently of the nature of the nucleotide to be copied and the position in the sequence [20]. Then, the single digit accuracy $q = 1 - p$ is the mean fraction of correctly incorporated nucleotides and the elements of the mutation matrix for a polynucleotide of chain length n are of the form:

$$Q_{ij} = q^n \left(\frac{1-q}{q} \right)^{d_{ij}} ,$$

with d_{ij} being the Hamming distance between two sequences I_i and I_j. The critical condition occurs at the minimum accuracy:

$$q_{\min} = 1 - p_{\max} = \sqrt[n]{Q_{\min}} = \sigma_m^{-1/n} , \tag{10}$$

which was called the *error threshold*. Above threshold no stationary distribution of sequences is formed. Instead, the population drifts randomly through sequence space. This implies that all genotypes have only finite life times, inheritance breaks down and evolution becomes impossible.

Figure 6 shows the stationary frequency of the master sequence as a function of the error rate. Variations in the accuracy of *in vitro* replication can indeed be easily achieved because error rates can be tuned over many orders of magnitude [54, 59]. The range of replication accuracies which are suitable for evolution is limited by the maximal accuracy that can be achieved by the replication machinery and the minimum accuracy determined by the error threshold. Populations in constant environments have an advantage when they operate near the maximal accuracy because then they loose as few copies through mutation as possible. In highly variable environments the opposite is true: it pays to produce as many mutants as possible because then the chance is largest to cope successfully with change.

In order to be able to study stochastic features of population dynamics around the error threshold, the replication-mutation system was modeled by a multitype branching process [13]. Main result of this study is the derivation of an expression for the probability of survival to infinite time for the master sequence and its mutants. In the regime of sufficiently accurate replication the survival probability is non-zero and decreases with increasing error rate. At the critical accuracy $q_{\min}$ this probability becomes zero. This implies that all molecular species which are currently in the populations, master and mutants, will die out in finite times and new variants will appear. This scenario is tantamount to migration of the population through sequence space. The critical accuracy $q_{\min}$, commonly seen as an error threshold for replication, can as well be understood as the localization threshold of the population in sequence space [62]. Later investigations aimed directly at a derivation of the error threshold in finite populations [67, 2].

In order to check the relevance of the error threshold for the replication of RNA viruses the minimum accuracy of replication can be transformed into a maximum chain length $n_{\max}$ for a given error rate p. The condition for stationarity of the quasispecies then reads:

$$n \; < \; n_{\max} \; = \; -\frac{\ln \sigma}{\ln q} \; \approx \; \frac{\ln \sigma}{1-q} \; . \tag{10 a}$$

The populations of most RNA viruses were shown to live indeed near the above mentioned critical value of replication accuracy [14, 15]. In particular, the chain length n was found to be roughly the inverse mutation rate per site and replication [16]. According to previously mentioned expectations these viruses should live in very variable environments in agreement with the highly active defense mechanisms of the host cells.

6. Evolution of phenotypes

If several molecular species have the same maximal fitness we are dealing with a case of neutrality [50]. The superiority of the master sequence becomes $\sigma_m = 1$ in this case, and the localizaton threshold of the quasispecies converges to the limit of absolute replication accuracy, $q_{\min} = 1$. Accordingly, the deterministic model fails, and we have to modify the kinetic equations. Genotypes are ordered with respect to non-increasing selective values. The first k_1 different genotypes have maximal selective value: $w_1 = w_2 = \ldots = w_{k_1} = w_{\max} = \tilde{w}_1$ (where $\tilde{.}$ indicates properties of groups of neutral phenotypes). The second group of neutral genotypes has highest but one selective value: $w_{k_1+1} = w_{k_1+2} = \ldots = w_{k_1+k_2} = \tilde{w}_2 < \tilde{w}_1$, etc. Replication rate constants are assigned in the same way: $a_1 = a_2 = \ldots = a_{k_1} = \tilde{a}_1$, etc. In addition, we define new variables, y_j $(j = 1, \ldots, L)$, that lump together all genotypes folding into the same phenotype:

$$y_j \;=\; \sum_{i\,=\,k_{j-1}+1}^{k_j} x_i \;\text{ with }\; \sum_{j=1}^{L} y_j = \sum_{i=1}^{M} x_i = 1 \;, \tag{11}$$

Without loss of generality we denote the phenotype with maximal fitness, the *master phenoptype*, by "m". Since we are heading again for a kind of zeroth-order solution, we consider only the master phenotype and put $k_1 = k$. With $y_m = \sum_{i=1}^{k} x_i$ we obtain the following kinetic differential equation for the set of sequences forming the neutral network of the master phenotype:

$$\dot{y}_m \;=\; \sum_{i=1}^{k} \dot{x}_i \;=\; y_m\left(\tilde{a}_m Q_{kk} - \bar{E}\right) \;+\; \sum_{i=1}^{k}\sum_{j\neq i} a_j Q_{ji} x_j \;. \tag{12}$$

The mean excess productivity of the population is, of course, independent of the choice of variables:

$$\bar{E} \;=\; \sum_{j=1}^{L} \tilde{a}_j y_j \;=\; \sum_{i=1}^{M} a_i x_i \;.$$

In order to derive a suitable expression for a phenotypic error threshold we split the mutational backflow into two contributions, (i) mutational backflow on the neutral network and (ii) mutational backflow from genotypes not on the network:

$$\sum_{i=1}^{k}\sum_{j\neq i} a_j Q_{ji} x_j \;=\; \left\{ \tilde{a}_m \sum_{i=1}^{k}\;\sum_{j=1,j\neq i}^{k} Q_{ji} x_j \right\} + \left\{ \sum_{i=1}^{k}\;\sum_{j=k+1}^{M} a_j Q_{ji} x_j \right\} \;.$$

We approximate by assuming a constant fraction of selectively neutral neighbors of the master phenotype (λ_m) and equal mutation rates ($Q_{ji} = \bar{Q}_j; i,j = 1,\ldots,k; i \neq j$) on the master network and find:

$$\sum_{i=1}^{k} \sum_{j=1,j\neq i}^{k} Q_{ji} x_j \approx \frac{\lambda_m(1-Q_{mm})}{k-1} \sum_{i=1}^{k} \sum_{j=1,j\neq i}^{k} x_j =$$
$$= \frac{\lambda_m(1-Q_{mm})}{k-1} \sum_{j=1,j\neq i}^{k} \sum_{i=1}^{k} x_j = \lambda_m(1-Q_{mm})\, y_m \ .$$

Mutational backflow from other networks ($y_j, j \neq m$) need not be evaluated explicitly since it has also been neglected in the derivation of the genotypic error threshold. The kinetic equation for the master phenotype can now be rewritten:

$$\dot{y}_m = \left(\tilde{a}_m \tilde{Q}_{mm} - \bar{E}\right) y_m + \text{Mutational Backflow}$$

They are identical with those in the variables expressing genotype concentrations except the use of an effective replication accuracy of

$$\tilde{Q}_{mm} = Q_{mm} + \lambda_m(1-Q_{mm}) = q^n\left(\Phi(q)\lambda_m + 1\right), \text{ with } \Phi(q) = \left(\frac{1}{q^n - 1}\right) \ .$$

The last part of the equation has the advantage that the over-all accuracy can be factorized into contributions from classes of nucleotides corresponding to positions on the sequence with different degrees of neutrality, $\lambda^{(k)}$:

$$\tilde{Q}_{mm} = q^n \prod_k \left(\Phi_k(q)\lambda_m^{(k)} + 1\right), \text{ with } \Phi_k(q) = \left(\frac{1}{q^{n_k} - 1}\right) \ .$$

The numbers of nucleotides in class k is denoted by n_k; clearly we have $\sum_k n_k = n$. Recently, it has been shown that a four class approximation of the distribution of λ-values yields excellent results for tRNAs [75].

Neglecting mutational backflow from non-master phenotypes we finally find in complete analogy with the derivation of the genotypic error threshold

$$\tilde{Q}_{\min} = Q_{mm} + \lambda_m(1-Q_{mm}) = \sigma_m^{-1}$$

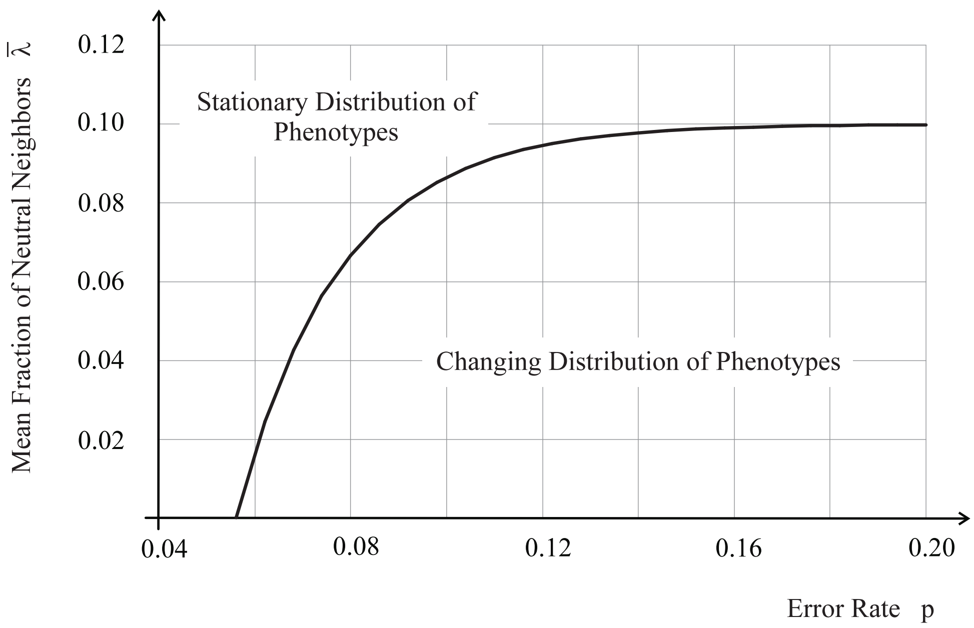


Figure 7: The phenotypic error threshold. The error threshold is shown as a function of the error rate p and the mean degree of neutrality $\bar{\lambda}$. The line separates the domains of stationary quasispecies and migrating populations. More replication errors can be tolerated at higher degrees of neutrality.

where σ_m is the superiority of the "master phenotype". Introducing the uniform error rate model we obtain by neglecting mutational backflow for the stationary frequency of master phenotypes:

$$\bar{y}_m(p) = \frac{\tilde{Q}_{mm}(p)\,\sigma_m - 1}{\sigma_m - 1} = \frac{(1-p)^n\,\sigma_m\,(1-\lambda_m) + \sigma_m\lambda_m - 1}{\sigma_m - 1} .$$

Eventually we find for the phenotypic error threshold by applying the "zeroth-order approximation" ($\bar{y}_m = 0$):

$$q_{\min} = (1 - p_{\max}) = \left(\frac{1-\lambda_m\sigma_m}{(1-\lambda_m)\,\sigma_m}\right)^{1/n} .$$

The function $q = q_{\min}(n, \lambda_m, \sigma_m)$ is illustrated in figure 7. The limits are easily visualized: (i) the phenotypic error threshold converges to the genotypic value $q_{\min} = \sigma_m^{-1/n}$ in the limit $\lambda_m \to 0$ and (ii) the minimal replication accuracy $q_{\min}$ approaches zero in the limit $\lambda_m \to \sigma_m^{-1}$. The second case implies that single digit accuracy plays no role in case the degree of neutrality is larger than the reciprocal value of the superiority.

Recapitulating the results on stationary distributions of phenotypes derived in this section we state that selective neutrality allows to tolerate more replication errors than in the non-neutral case. We are dealing with a distribution of changing genotypes corresponding to a population which drifts randomly [45] on the neutral network of the fittest or master phenotype. In this drift the master phenotype is conserved as long as the replication accuracy is above a critical minimal value, $q_{\min}$. In case the accuracy falls also below this critical value the population drifts through sequence space and through shape space and no more stationarity, neither with genotypes nor with phenotypes, is observed. It is particularly interesting to note that there is a degree of neutrality related to the superiority of the master phenotype ($\lambda = \sigma^{-1}$) above which the error rate does not matter. In other words, the master phenotype will never be lost when the degree of neutrality exceeds a limit being the inverse superiority.

So far phenotypes were only considered in terms of parameters contained in the kinetic equations. Mutation acts on genotypes whereas selection is dealing with phenotypes since fitness is a property of the phenotype. The relations between genotypes and phenotypes are thus an intrinsic part of evolution and no theory can be complete without considering them. A comprehensive theory of evolution which is explicitly dealing with phenotypes has been introduced a few years ago [79, 81, 80]. The model is shown in figure 8. The complex process of evolution is partitioned into three simpler phenomena: (i) population genetics, (ii) migration of populations, and (iii) genotype-phenotype mapping. Conventional population genetics is extended by two more aspects: population support dynamics describing the migration of populations through sequence space and genotype-phenotype mapping providing the source of the parameters for populations genetics. In general, phenotypes and their formation from genotypes are so complex that they cannot be handled appropriately. In test-tube evolution of RNA, however, the phenotypes are molecular structures. Then, genotype and phenotype are two features of the same molecule. In this simplest known case the relations between

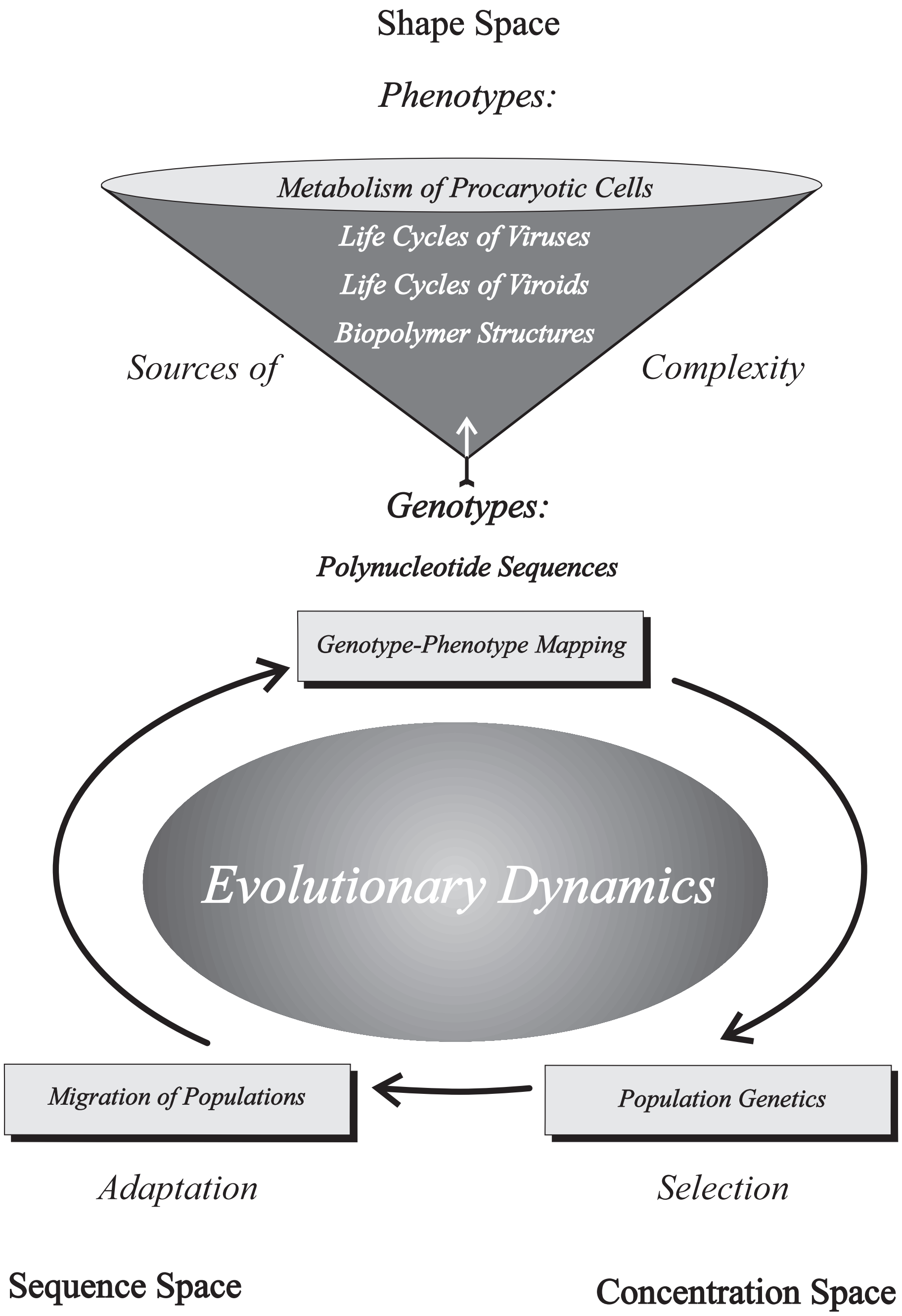

Shape Space
Phenotypes:
Metabolism of Procaryotic Cells
Life Cycles of Viruses
Life Cycles of Viroids
Biopolymer Structures
Sources of
Complexity
Genotypes:
Polynucleotide Sequences
Genotype-Phenotype Mapping
Evolutionary Dynamics
Migration of Populations
Population Genetics
Adaptation
Selection
Sequence Space
Concentration Space

Figure 8: A comprehensive model of molecular evolution. The highly complex process of biological evolution is partitioned into three simpler dynamical phenomena: (i) population genetics, (ii) migration of populations, and (iii) genotype-phenotype mapping. Population genetics describes how optimal genotypes with optimal genes are chosen from a given reservoir by natural (or artificial) selection. The basis of population genetics is replication, mutation and recombination modeled by differential equations as derived from chemical reaction kinetics. In essence, population genetics is concerned with selection and other evolutionary phenomena occurring on short time-scales. Population support dynamics describes how the genetic reservoirs change when populations migrate in the huge space of all possible genotypes. Issues are the internal structure of populations and the mechanisms by which the regions of high fitness are found in sequence or genotype space. Support dynamics is dealing with the long-time phenomena of evolution, for example, with optimization and adaptation to changes in the environment. Genotype-phenotype mapping represents a core problem of evolutionary thinking since the dichotomy between genotypes and phenotypes is the basis of Darwin's principle of variation and selection: all genetically relevant variation takes place on the genotypes whereas the phenotypes are subjected to selection. Variations and their results are uncorrelated in the sense that a mutation yielding a fitter phenotype does not occur more frequently because of the increase in fitness. The problem is the enormous complexity of the unfolding of genotypes that involves sophisticated processes from the formation of biopolymer structures to cellular metabolism and higher up to the almost open ended increase in complexity with the development of multicellular organisms.

genotypes and phenotypes are reduced to the mapping of RNA sequences onto structures. Folding RNA sequences into structures is an essential part of the RNA optimization process and can be considered explicitly provided a coarse-grained version of structure, the secondary structure, is used. The model is self-contained in the sense that it is based on the rules of RNA secondary structure formation, the kinetics of replication and mutation as well as the structure of sequence space, and it needs no further inputs. The three processes shown in figure 8 are indeed connected by a cyclic mutual dependence in which each process is driven by the previous one in the cycle and provides the input for the next one: (i) Folding sequences into structures yields the input for population genetics. (ii) Population genetics describes the arrival of new genotypes through mutation and the dying of old ones through selection, and determines thereby how and where the population migrates. (iii) Migration of the population in sequence space finally defines the new genotypes that are to be mapped into phenotypes and thus completes the cycle. The model of evolutionary dynamics has been applied to interpret the experimental data on molecular evolution and it was implemented for computer simulations of neutral evolution and RNA optimization in the flow reactor [32]. The computer simulations allow to follow the optimization process in full detail on the molecular level. Individual runs are monitored as time series of structures which eventually lead to the optimized molecule. The simulations helped to clarify the role of neutral variants in evolution. Recording of evolution experiments [26] as well as computer simulations [45, 44, 32] have shown first that optimization does not occur continuously. Instead, stepwise increases of fitness are observed. The periods of increase are interrupted by long phases of almost constant fitness. Inspection of populations during the quasi-static phases revealed that constancy is restricted to the level of phenotypes or their properties, respectively. The genotypes are changing all the time and the apparent stasis is a result of selective neutrality or, in other words, populations drift randomly through sequence space but stay on neutral networks.

Selective neutrality plays an active role in optimization. On a rugged landscape without neutrality populations are regularly caught in evolutionary traps: whenever a population reaches a local optimum in sequence space, i.e. a point that has no neighbors with higher fitness values, optimization comes to an end. If we are dealing with a sufficiently high degree of neutrality, however, the landscape consists of extended neutral networks for all common phenotypes [76]. Almost all points having no further advantageous neighbors belong to one of the extended neutral network. When a population reaches such a point at the end of an adaptive phase, it starts drifting randomly on the network until it comes to an area that contains also points of higher fitness. There, the next adaptive period starts and the population continues the hill-climbing process. The role of neutral variants is to enable populations to leave local fitness optima and to proceed towards areas of higher fitness in sequence space. Optimization on realistic landscapes is a process on two time scales: Fast adaptive phases with substantial increase in fitness are interrupted by periods of random drift during which fitness is essentially constant. The combination of adaptation and drift allows to escape from evolutionary traps and, depending on the degree of neutrality, eventually leads to the global optimum of the landscape.

7. RNA perspectives

Molecular evolution experiments with RNA molecules and the accompanying theoretical descriptions made three important contributions to evolutionary biology:

(i) The role of replicative units in the evolutionary process has been clarified, the conditions for the occurrence of error thresholds have been layed down, and the role of neutrality has been elucidated.

(ii) The Darwinian principle of (natural) selection has shown to be no privilege of cellular life since it is valid also in serial transfer experiments, flow-reactors, and other laboratory assays such as SELEX.

(iii) Evolution in molecular systems is faster than organismic evolution by many orders of magnitude and thus allows to observe optimization and adaptation on easily accessible time-scales, i.e. within days or weeks.

The third issue made selection and adaptation subjects of laboratory investigations. In all these systems the coupling between different replicons is weak: in the simplest case there is merely competition for common resources, for example the raw materials for replication. With more realistic chemical reaction mechanisms a sometimes substantial fraction of the replicons is unavailable as long as templates are contained in complexes. None of these systems, however, comes close to the strong interactions and interdependencies characteristic for ecosystems.

In contrast to the weakly coupled networks of replicons considered in this contribution, *hypercycles* [17, 21] involve specific catalysis beyond mere template instruction (see figure 9). In the simplest case, where we consider catalyzed replication reactions explicitly, the reaction equations are of the form:

$$(A) + I_k + I_l \to 2I_k + I_l \ . \tag{13}$$

Here a copy of I_k is produced using another macromolecular species I_l as a specific catalyst for the replication reaction. A more realistic version of (13) that might be experimentally feasible is

$$\begin{aligned} A_i + B_j + C_{kl} + C_{rs} \underset{\bar{a}_{ijkl}}{\overset{a_{ijkl}}{\rightleftarrows}} A_iB_jC_{kl} + C_{rs} \underset{\bar{e}_{ijklrs}}{\overset{e_{ijklrs}}{\rightleftarrows}} A_iB_jC_{kl}C_{rs} \overset{b_{ijklrs}}{\longrightarrow} \\ C_{ij}C_{kl}C_{rs} \overset{f_{ijklrs}}{\longrightarrow} C_{ij} + C_{kl}C_{rs} \underset{\bar{d}_{ijkl}}{\overset{d_{ijkl}}{\rightleftarrows}} C_{ij} + C_{kl} + C_{rs} \end{aligned}$$

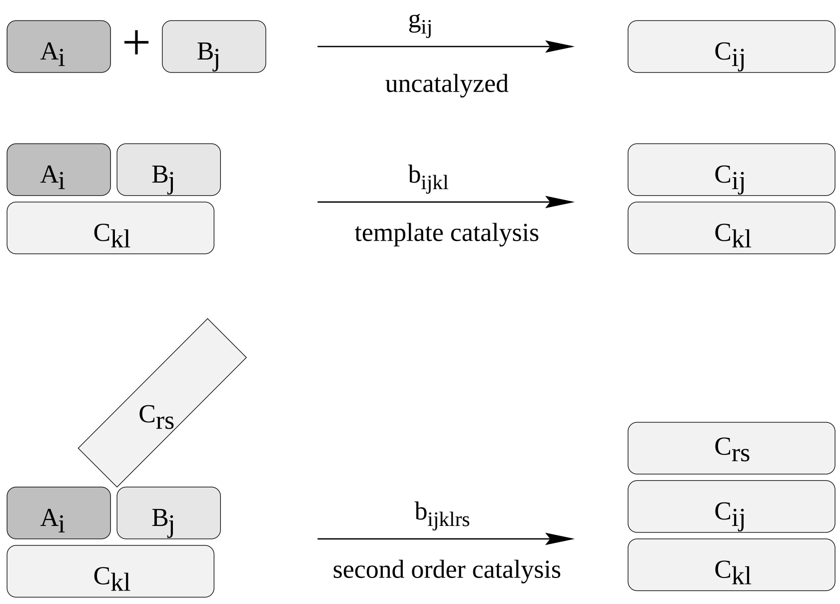


Figure 9: Modes of template formation. In complex systems of mixed templates and depending on the underlying mechanism of template synthesis, different modes of dynamic behavior are possible. Uncatalyzed synthesis generally corresponds to linear growth. Template-instructed synthesis gives parabolic or exponential growth. The coupling of systems involving second order autocatalysis can also give rise to hyperbolic growth, as has been predicted for hypercycles [21].

Here the template C_{rs} plays the role of a ligase for the template-directed replication step.

The kinetic differential equation

$$\dot{x}_k = x_k \left(\sum_l a_{kl} x_l - \Phi(x) \right)$$

corresponding to the mechanism (13) has been termed *second order replicator equation* [83]. These systems can display enormous diversity of dynamic behavior [43] depending on the structure of the matrix (a_{kl}) of coupling constants which describes the catalytic activity of one species (I_l) on the replication of another one (I_k). Second order replicator equations are mathemically equivalent to Lotka-Volterra equations used in mathematical ecology [41]. Indeed, recent research in the group of John McCaskill in Jena [104, 63] is dealing with *molecular ecologies* of strongly interacting replicons.

The work with RNA replicons has had a pioneering character. Both the experimental approach to evolution in the laboratory and the development of a theory of evolution are much simpler for RNA than in case of proteins or viruses. On the other hand, genotype and phenotype are more closely linked in RNA than in any other system. The next logical step in theory [21, 38] and experiment [18] consists of the development of a coupled RNA-protein system that makes use of both replication and translation. This achieves the effective decoupling of genotype and phenotype that is characteristic for all living organisms: RNA is the genotype, protein the phenotype and thus, genotype and phenotype are no longer housed in the same molecule. The development of a theory of evolution in the "RNA-protein world" requires little more than an understanding of the sequence-structure relations in proteins. There, a huge body of theoretical and empirical knowledge is already available and the daily growing sequence and structure databanks provide a substantial amount of not yet exploited information.

Virus life cycles represent the next logical step in increasing complexity of genotype-phenotype interactions. RNA viruses are the simplest candidates and indeed the development of a phage in a bacterial cell has already been modeled in a pioneering paper by Charles Weissmann [101]. Complete viral RNA-genomes are now accessible to computational investigations searching for functional substructures [40] and we can expect progress in understanding viral phenotypes in the not to distant future.

Acknowledgments

The work reported here was supported financially by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Projects No. 11065-CHE, 12591-INF, and 13093-GEN, by the European Commission, Project No. PL970189, and by the Santa Fe Institute.

References

[1] D. Alves and J. F. Fontanari. Population genetics approach to the quasispecies model. *Phys. Rev. E*, 54:4048–4053, 1996.

[2] D. Alves and J. F. Fontanari. Error threshold in finite populations. *Phys. Rev. E*, 57:7008–7013, 1998.

[3] P. A. Bachmann, P. L. Luisi, and J. Lang. Autocatalytic self-replicating micelles as models for prebiotic structures. *Nature*, 357:57–59, 1992.

[4] D. P. Bartel and J. W. Szostak. Isolation of new ribozymes from a large pool of random sequences. *Science*, 261:1411–1418, 1993.

[5] A. A. Beaudry and G. F. Joyce. Directed evolution of an RNA enzyme. *Science*, 257:635–641, 1992.

[6] C. K. Biebricher and M. Eigen. Kinetics of RNA replication by Qβ replicase. In E. Domingo, J. J. Holland, and P. Ahlquist, editors, *RNA Genetics. Vol.I: RNA Directed Virus Replication*, pages 1–21. CRC Press, Boca Raton, FL, 1988.

[7] R. R. Breaker and G. F. Joyce. Emergence of a replicating species from an *in vitro* RNA evolution reaction. *Proc. Natl. Acad. Sci. USA*, 91:6093–6097, 1994.

[8] J. H. Cate, A. R. Gooding, E. Podell, K.Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, and J. A. Doudna. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science*, 273:1678–1685, 1996.

[9] T. R. Cech. RNA splicing: Three themes with variations. *Cell*, 34:713–716, 1983.

[10] T. R. Cech. RNA as an enzyme. *Sci. Am.*, 255(5):76–84, 1986.

[11] T. R. Cech. Self-splicing of group I introns. *Ann. Rev. Biochem.*, 59:543–568, 1990.

[12] B. Cuenoud and J. W. Szostak. A DNA metalloenzyme with DNA ligase activity. *Nature*, 375:611–614, 1995.

[13] L. Demetrius, P. Schuster, and K. Sigmund. Polynucleotide evolution and branching processes. *Bull.Math.Biol.*, 47:239–262, 1985.

[14] E. Domingo. Biological significance of viral quasispecies. *Viral Hepatitis Rev*, 2:247–261, 1996.

[15] E. Domingo and J. J. Holland. RNA virus mutations and fitness for survival. *Ann.Rev.Microbiol.*, 51:151–178, 1997.

[16] J. W. Drake. Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci. USA*, 90:4171–4175, 1993.

[17] M. Eigen. Selforganization of matter and the evolution of macromolecules. *Naturwiss.*, 58:465–523, 1971.

[18] M. Eigen, C. K. Biebricher, M. Gebinoga, and W. C. Gardiner jr. The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. *Biochemistry*, 30:11005–11018, 1991.

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

[20] M. Eigen and P. Schuster. The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften*, 64:541–565, 1977.

[21] M. Eigen and P. Schuster. *The Hypercycle – A Principle of Natural Self-Organization*. Springer-Verlag, Berlin, 1979. Translated into Russian, Bulgarian and Chinese.

[22] M. Eigen and P. Schuster. Stages of emerging life - Five principles of early organization. *J. Mol. Evol.*, 19:47–61, 1982.

[23] M. Eigen and P. Schuster. Stages of emerging life - Five principles of early organization. *J.Mol.Evol.*, 19:47–61, 1982.

[24] E. H. Ekland and D. P. Bartel. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature*, 382:373–376, 1996.

[25] E. H. Ekland, J. W. Szostak, and D. P. Bartel. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science*, 269:364–370, 1995.

[26] S. F. Elena, V. S. Cooper, and R. E. Lenski. Punctuated evolution caused by selection of rare beneficial mutants. *Science*, 272:1802–1804, 1996.

[27] A. Eschenmoser. Hexose nucleic acids. *Pure and Applied Chemistry*, 65:1179–1188, 1993.

[28] E. Fahy, D. Y. Kwoh, and T. R. Gingeras. Self-sustained sequence replication (3SR): An isothermal transcription-based amplification system alternative to PCR. *PCR Methods Appl.*, 1:25–33, 1991.

[29] A. R. Ferré-D'Amaré, K. Zhou, and J. A. Doudna. Crystal structure of a hepatitis delta virus ribozyme. *Nature*, 395:567–574, 1998.

[30] W. Fontana, D. A. M. Konings, P. F. Stadler, and P. Schuster. Statistics of RNA secondary structures. *Biopolymers*, 33:1389–1404, 1993.

[31] W. Fontana and P. Schuster. A computer model of evolutionary optimization. *Biophys.Chem.*, 26:123–147, 1987.

[32] W. Fontana and P. Schuster. Continuity in evolution. On the nature of transitions. *Science*, 280:1451–1455, 1998.

[33] S. W. Fox and H. Dose. *Molecular Evolution and the Origin of Life*. Academic Press, New York, 1977.

[34] R. F. Gesteland and J. F. Atkins, editors. *The RNA World*. Cold Spring Harbor Laboratory Press, Plainview, NY, 1993.

[35] W. Gilbert. The RNA world. *Nature*, 319:618, 1986.

[36] R. Green and H. F. Noller. Ribosomes and translation. *Ann. Rev. Biochem.*, 66:679–716, 1997.

[37] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, and S. Altman. The RNA moiety of Ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35:849–857, 1983.

[38] R. Happel, R. Hecht, and P. F. Stadler. Autocatalytic networks with translation. *Bull. Math. Biol.*, 58:877–905, 1996.

[39] R. Happel and P. F. Stadler. Autocatalytic replication in a cstr and constant organization. *J. Math. Biol.*, page in press, 1998. SFI preprint 95-07-062.

[40] I. L. Hofacker, M. Fekete, C. Flamm, M. A. Huynen, S. Rauscher, P. E. Stolorz, and P. F. Stadler. Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucl. Acids Res.*, 26:3825–3836, 1998.

[41] J. Hofbauer. On the occurrence of limit cycles in the Volterra-Lotka differential equation. *Nonlin. Anal.*, 5:1003–1007, 1981.

[42] J. Hofbauer, P. Schuster, and K. Sigmund. Competition and cooperation in catalytic selfreplication. *J. Math. Biol*, 11:155–168, 1981.

[43] J. Hofbauer and K. Sigmund. *Dynamical Systems and the Theory of Evolution*. Cambridge Univ. Press, Cambridge, UK, 1998.

[44] M. A. Huynen. Exploring phenotype space through neutral evolution. *J. Mol. Evol.*, 43:165–169, 1996.

[45] 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.

[46] A. Jenne and M. Famulok. A novel ribozyme with ester transferase activity. *Chemistry & Biology*, 5:23–34, 1998.

[47] L. Jiang, A. K. Suri, R. Fiala, and D. J. Patel. Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex. *Chemistry & Biology*, 4:35–50, 1997.

[48] G. F. Joyce. The rise and fall of the RNA world. *The New Biologist*, 3:399–407, 1991.

[49] S. A. Kauffman. *The origins of Order. Self-Organization and Selection in Evolution*. Oxford University Press, New York, 1993.

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

[51] F. R. Kramer, D. R. Mills, P. E. Cole, T. Nishihara, and S. Spiegelman. Evolution *in vitro*: Sequence and phenotype of a mutant RNA resitant to ethidium bromide. *J. Mol. Biol.*, 89:719–736, 1974.

[52] D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin, and M. R. Ghadiri. A self-replicating peptide. *Nature*, 382:525–528, 1996.

[53] D. H. Lee, K. Severin, Y. Yokobayashi, and M. R. Ghadiri. Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature*, 390:591–594, 1997.

[54] D. W. Leung, E. Chen, and D. V. Goeddel. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*, 1:11–15, 1989.

[55] P. A. Lohse and J. W. Szostak. Ribozyme-catalyzed amino-acid transfer reactions. *Nature*, 381:442–444, 1996.

[56] J. R. Lorsch and J. W. Szostak. *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature*, 371:31–36, 1994.

[57] J. R. Lorsch and J. W. Szostak. Kinetic and thermodynamic charakterization of the reaction catalyzed by a polynucleotide kinase ribozyme. *Biochemistry*, 33:15315–15327, 1995.

[58] P. L. Luisi, P. Walde, and T. Oberholzer. Enzymatic RNA synthesis in self-reproducing vesicles: An approach to the construction of a minimal synthetic cell. *Ber. Bunsenges. Phys. Chem.*, 98:1160–1165, 1994.

[59] M. A. Martinez, J. P. Vartanian, and S. Wain-Hobson. Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc. Natl. Acad. Sci. USA*, 91:11787–11791, 1994.

[60] S. F. Mason. *Chemical evolution. Origin of the elements, molecules, and living systems.* Clarendon Press, Oxford (UK), 1991.

[61] J. Maynard Smith and E. Szathmáry. *The Major Transitions in Evolution.* W. H. Freeman, Oxford, UK, 1995.

[62] J. McCaskill. A localization threshold for macromolecular quasispecies from continuously distributed replication rates. *J. Chem. Phys.*, 80:5194–5202, 1984.

[63] J. S. McCaskill. Spatially resolved *in vitro* molecular ecology. *Biophys Chem*, 66:145–158, 1997.

[64] D. R. Mills, R. L. Peterson, and S. Spiegelman. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. USA*, 58:217–224, 1967.

[65] K. B. Mullis. The unusual origin of the polymerase chain reaction. *Sci. Am.*, 262(4):36–43, 1990.

[66] H. F. Noller, V. Hoffarth, and L. Zimniak. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science*, 256:1416–1419, 1992.

[67] M. Nowak and P. Schuster. Error thresholds of replication in finite populations. Mutation frequencies and the onset of Muller's ratchet. *J.Theor.Biol.*, 137:375–395, 1989.

[68] J. S. Nowick, Q. Feng, T. Ballester, and J. Rebek, Jr. Kinetic studies and modeling of a self-replicating system. *J. Am. Chem. Soc.*, 113:8831–8839, 1991.

[69] L. E. Orgel. RNA catalysis and the origin of life. *J. Theor. Biol.*, 123:127–149, 1986.

[70] L. E. Orgel. Evolution of the genetic apparatus. A review. *Cold Spring Harbor Symposia on Quantitative Biology*, 52:9–16, 1987.

[71] L. E. Orgel. Molecular replication. *Nature*, 358:203–209, 1992.

[72] H. D. Pflug and H. Jaeschke-Boyer. Combined structural and chemical analysis of 3.800-Myr-old microfossils. *Nature*, 280:483–486, 1979.

[73] H. Pley, K. Flaherty, and D. McKay. Three-dimensional structures of a hammerhead ribozyme. *Nature*, 372:68–74, 1994.

[74] J. R. Prudent, T. Uno, and P. G. Schultz. Expanding the scope of RNA catalysis. *Science*, 264:1924–1927, 1994.

[75] C. Reidys, C. Forst, and P. Schuster. Replication and mutation on neutral networks. *Bull. Math. Biol.*, 1998. Submitted. Also published as: Preprint No. 98-04-036, Santa Fe Institute, Santa Fe, NM 1998.

[76] C. M. Reidys, P. F. Stadler, and P. Schuster. Generic properties of combinatory maps: Neural networks of RNA secondary structures. *Bull. Math. Biol.*, 59:339–397, 1997.

[77] M. Schidlowski. A 3.800-million-year isotope record of life from carbon in sedimentary rocks. *Nature*, 333:313–318, 1988.

[78] J. W. Schopf. Microfossils of the early archean apex chert: New evidence of the antiquity of life. *Science*, 260:640–646, 1993.

[79] P. Schuster. Artificial life and molecular evolutionary biology. In F. Morán, A. Moreno, J. J. Merelo, and P. Chacón, editors, *Advances in Artificial Life. Proceedings of Third European Conference on Artificial Life, Ganada, 1995*, volume 929 of *Lecture Notes in Artificial Intelligence.*, pages 3–19. Springer-Verlag, Berlin, 1995.

[80] P. Schuster. Genotypes with phenotypes: Adventures in an RNA toy world. *Biophys. Chem.*, 66:75–110, 1997.

[81] P. Schuster. Landscapes and molecular evolution. *Physica D*, 107:351–365, 1997.

[82] 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.R.Soc.Lond. B*, 255:279–284, 1994.

[83] P. Schuster and K. Sigmund. Replicator dynamics. *J.Theor.Biol.*, 100:533–538, 1983.

[84] P. Schuster and K. Sigmund. Dynamics of evolutionary optimization. *Ber. Bunsenges. Phys. Chem.*, 89:668–682, 1985.

[85] P. Schuster and J. Swetina. Stationary mutant distribution and evolutionary optimization. *Bull.Math.Biol.*, 50:635–660, 1988.

[86] A. W. Schwartz. Speculation on the RNA precursor problem. *J. Theor. Biol.*, 187:523–527, 1997.

[87] W. G. Scott, J. T. Finch, and A. Klug. The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell*, 81:991–1002, 1995.

[88] L. A. Segel and M. Slemrod. The quasi-steady state assumption: A case study in perturbation. *SIAM Rev.*, 31:446–477, 1989.

[89] K. Severin, D. H. Lee, J. R. Granja, J. A. Martinez, and M. R. Ghadiri. Peptide self-replication via template directed ligation. *Chemistry*, 3:1017–1024, 1997.

[90] S. Spiegelman. An approach to the experimental analysis of precellular evolution. *Quart. Rev. Biophys.*, 4:213–253, 1971.

[91] P. F. Stadler. Complementary replication. *Math. Biosc.*, 107:83–109, 1991.

[92] J. Swetina and P. Schuster. Self-replication with errors - A model for polynucleotide replication. *Biophys.Chem.*, 16:329–345, 1982.

[93] E. Szathmáry and I. Gladkih. Sub-exponential growth and coexistence of non-enzymatically replicating templates. *J. Theor. Biol.*, 138:55–58, 1989.

[94] P. Tarazona. Error-tresholds for molecular quasi-species as phase transitions: From simple landscapes to spinglass models. *Phys. Rev. A[15]*, 45:6038–6050, 1992.

[95] T. Tjivikua, P. Ballester, and J. Rebek Jr. A self-replicating system. *J. Am. Chem. Soc.*, 112:1249–1250, 1990.

[96] O. C. Uhlenbeck. A small catalytic oligoribonucleotide. *Nature*, 328:596–600, 1987.

[97] S. Varga and E. Szathmáry. An extremum principle for parabolic competition. *Bull. Math. Biol.*, 59:1145–1154, 1997.

[98] G. von Kiedrowski. A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed. Engl.*, 25:932–935, 1986.

[99] G. von Kiedrowski. Minimal replicator theory I: Parabolic versus exponential growth. In *Bioorganic Chemistry Frontiers, Volume 3*, pages 115–146, Berlin, Heidelberg, 1993. Springer-Verlag.

[100] M. Wecker, D. Smith, and L. Gold. *In vitro* selection of a novel catalytic RNA: Characterization of a sulfur alkylation reaction and interaction with a small peptide. *RNA*, 2:982–994, 1996.

[101] C. Weissmann. The making of a phage. *FEBS Letters (Suppl.)*, 40:S10–S12, 1974.

[102] P. R. Wills, S. A. Kauffman, B. M. Stadler, and P. F. Stadler. Selection dynamics in autocatalytic systems: Templates replicating through binary

ligation. *Bull. Math. Biol.*, 1998. in press, Santa Fe Institute Preprint 97-07-065.

[103] C. Wilson and J. W. Szostak. *In Vitro* evolution of a self-alkylating ribozyme. *Nature*, 374:777–782, 1995.

[104] B. Wlotzka and J. S. McCaskill. A molecular predator and its prey: Coupled isothermal amplification of nucleic acids. *Chemistry & Biology*, 4:25–33, 1997.

[105] B. Zhang and T. R. Cech. Peptide bond formation by *in vitro* selected ribozymes. *Nature*, 390:96–100, 1997.

[106] B. Zhang and T. R. Cech. Peptidyl-transferase ribozymes: *Trans* reactions, structural characterization and ribosomal RNA-like features. *Chemistry & Biology*, 5:539–553, 1998.