ELSEVIER

Biophysical Chemistry 66 (1997) 179–192

Biophysical Chemistry

Molecular Evolution of RNA in vitro

Christof K. Biebricher

Max-Planck-Institute for Biophysical Chemistry, D-37070 Göttingen, Germany

William C. Gardiner

Department of Chemistry and Biochemistry, University of Texas, Austin, TX 78712, USA

Received 8 April 1997; accepted 17 April 1997

Abstract

Experimental studies of RNA evolution in vitro are reviewed in the context of Eigen's 1971 theory and its subsequent extensions. Current research activity and future prospects for using automated molecular biology techniques for in vitro evolution experiments are surveyed. © 1997 Published by Elsevier Science B.V.

Keywords: in vitro evolution; aptamer; selection; RNA replication; ribozymes; genetic engineering; combinatorial chemistry

1. Introduction

In 1971, a century after Charles Darwin and Gregor Mendel laid the groundwork for our understanding of natural evolution and heredity—and millenia after beneficial uses of competition and selection had been reduced to everyday practice by our ancestral experimenters with crops and livestock—Manfred Eigen, inspired by experiments done with bacteriophage RNA in Urbana in the laboratory of Sol Spiegelman, presented a theory of in vitro evolution. It was targeted not only at physicochemical interpretation of data on RNA self-replication under laboratory conditions, but also at understanding how molecules could have evolved under conditions that we assume existed on the Earth before self-sustaining cellular life forms were present, in what later became known as "the RNA world".

Just as the laws of variation, competition, and selection introduced by Darwin and Wallace came to pervade our understanding of all of biology, Eigen's theory provides a comprehensive framework for describing molecular evolution throughout the ever-expanding spectrum of molecular biological situations in which we discover that it occurs. These now range from laboratory evolution experiments under controlled conditions to evolution of antibody specificity in the mammalian immune system and defeat of that specificity by opposing evolution of the coat proteins of viral invaders.

The explosive development of molecular biology techniques since 1971 has provided powerful tools—automated sequencing and synthesizing of biopolymers in particular—that enable one to do molecular evolution experiments today that were only Gedankenexperiments then. Indeed, we are clearly now just crossing the threshold of in vitro studies of molecular evolution. Our goal in this paper is thus to describe the current research scene as well as to outline the course of molecular evolution research since the early contributions of Spiegelman and Eigen.

PII S0301-4622(97) 00059-8

2. Early extracellular evolution experiments

Bacteriophage infection was recognized as a source of insight into basic life processes around the middle of the 20th Century, when refinements in microbiological techniques allowed phages and infected bacteria to be visualized and advances in genetic and physicochemical methods allowed identification and characterization of the molecular species involved. Among the phages that allowed fast progress in the study of their molecular organization were phages of E. coli found abundantly in sewage[1] that eventually became characterized as plus-strand RNA viruses. They proved to have a single RNA as their genome and to require only one enzyme, the RNA replicase, to amplify the viral RNA genome. In contrast to the RNA replicases of many RNA coliphages, the replicase of phage Qβ was found to be sufficiently stable to be isolated [2,3] and to retain its activity in frozen storage. Because of this stability, it has provided the cornerstone for over three decades of research on RNA replication itself and on the consequent molecular evolution of the RNA being replicated.

Once Spiegelman and his coworkers had established conditions under which Qβ replicase would replicate its RNA in vitro, they proceeded to study the replication process in detail and recognized that they had discovered "An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule", the title of their classic 1967 paper [4]. The essence of these and later experiments is that the replicase itself is present as a constant environmental factor, not subject to evolution, while the autocatalytically amplified RNA generates selection pressure among the "mutant" RNA strands formed by imperfect replication and does evolve.

The first selection pressure studied was that provided by the high discrimination of the replicase in accepting RNA templates. (Discrimination is a required feature of the replicase; else viral RNA would be unable to compete against the high concentration of host-cell RNA it encounters at the start of an infection cycle.) The design of the experiment targeted selection for fecundity, the fittest RNA being the one that replicated most rapidly for given experimental conditions. At the start a small amount of wild-type Qβ RNA was added to an incubation mixture containing replicase and nucleoside triphosphates at concentrations similar to those prevailing in host cells. After waiting long enough for large amplification of the RNA to occur, an aliquot was transferred to fresh incubation mixture and the process repeated. After 75 such transfers the RNA was characterized: The replication rate had accelerated to 2.6 times that of wild-type Qβ, while the chain length had diminished to only 15% of the wild-type value.

In a subsequent study [5] serial transfer experiments were done for a variety of conditions and selection pressures. Adaptation to the different conditions took place, "revealing an unexpected wealth of phenotypic differences which a replicating nucleic acid can exhibit". Quantititive interpretation of these early evolution experiments is difficult, however [6]; it is a pity that they have been never repeated under better defined conditions.

The difficulties attending evolution experiments beginning with the viral RNA were apparent, and so Spiegelman and his coworkers searched for and found well-defined shorter ones better suited for quantitative study [7–9]. With them they were able to follow the selection of a mutant adapted to replicate in the presence of the replication inhibitor ethidium bromide [10] and to show that three consecutive point mutations were responsible for it.

For physical understanding of the Spiegelman et al. experiments a new theoretical framework was needed, for the classical neo-Darwinistic theory developed by Haldane, Muller, Fisher and others [11] assumes reproduction of organisms, not of molecules. This new theoretical basis was provided by Eigen [12], who based his analysis on the fundamental nature of the physico-chemical processes underlying a generalized replication-mutation system. He showed that it was the autocatalytic amplification inherent in all molecular evolution experiments that enabled a general dynamical description of in vitro molecular evolution, independent of the specific system in which it occurs.

The general description must be formulated in a specific way for the conditions of any molecular evolution experiment under study. Quantitative descriptions of evolution based on Eigen's formalism were carried out later for short RNAs using Spiegelman's experimental system combined with quantitative radiographic methods for monitoring RNA production rates [13–16]. As described in detail in the following section, it was found that reliable selection values can indeed be derived systematically from physical-chemical parameters and that trends in evolution can be accurately predicted from them.

While molecular as well as organismic evolution is mainly concerned with adaptation of existing species and emergence of new ones from them, it also deals with the origin of life. Eigen's theory addresses this subject as well.

A large literature has developed around the theme of prebiotic chemistry. While much of it reports solid chemical findings [17], inferences drawn about the prebiotic roles that such chemistry may have played are quite speculative, and relatively little of it relates to precursors of nucleic acids. An important part that does was contributed by Orgel and coworkers [18–20]. Even though a practical chemical replication system did not emerge, their work did confirm that evolution in vitro could allow the evolutionary emergence of molecules with novel properties without deriving them from biological prototypes. In the first experiment, reported in 1973 by Biebricher and Orgel [21], a replicating RNA was selected from a random mixture of polynucleotides. Their approach was later reintroduced after the advent of synthetic methods that allow one to synthesize partially randomized sequences reproducibly. Automated synthesizers now permit one to condense appropriate monomeric building blocks, e.g., nucleotides, amino acids or unnatural monomers, in predetermined sequences routinely. This technology offers a pathway to "reproduce" and mutagenize "genotypes", i.e., sequence information, synthetically. Appropriate iterative strategies of selection and amplification have been devised [22]. Quantitative theoretical interpretation of the results, however, has not been undertaken so far. These essentially in vitro techniques have also invaded research on in vivo selection systems to such an extent that distinctions between them are often no longer possible.

3. Quantitative analysis of RNA evolution in vitro

Spiegelman's evolution experiments did not just demonstrate that Darwinian evolution does take place in vitro. The RNA replication system he introduced also has the two great advantages of allowing experiments to be done under controlled conditions and having only a few biochemical reaction steps. Using Spiegelman's system it is possible to describe molecular evolution quantitatively and to derive selection values from measured rate constants.

RNA replication requires two catalysts working in concert, the template RNA and the replicase [13]. In the Spiegelman system, however, the enzyme acts purely as an environmental factor—its concentration does not change during the course of an experiment and it is subject neither to mutation nor to selection.

The mechanism of single-strand RNA replication has accordingly been investigated in detail using short-chained RNA species replicated by $Q\beta$ RNA replicase [13–16, 23–29]. The basic replication mechanism is as follows (Fig. 1). Replicase binds to the single-stranded template, synthesizes and then releases the complementary replica strand, and finally releases the template RNA in a slow step to recycle both of them. The relative slowness of the last step is clearly a basic requirement for a viable phage replicase; were the template-replicase complex not tightly bound, production of incomplete replica strands would soon lead to extinction. Provided that the complementary strand is also accepted as a template for the replicase, autocatalytic amplification of a template takes place in a double cycle.

At the start of an experiment, enzyme is generally supplied in large excess over RNA, and the plus and minus strand concentrations of self-replicating RNA species are amplified exponentially. Once the growing RNA concentration reaches the enzyme concentration, the enzyme becomes saturated with template and the RNA con-

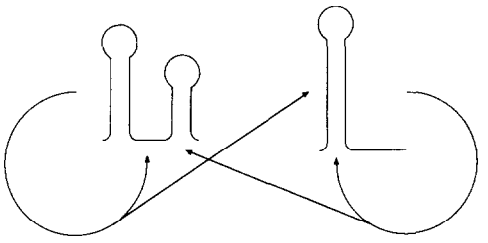

Fig. 1. Schematic RNA replication mechanism. Both plus and minus RNA are single stranded and both must be replicated by the replicase. Not included in the diagram is formation of double-stranded RNA, which eventually competes with replication and limits the concentration growth.

centration increases linearly with time. At this stage free plus and minus strands also begin to combine to double strands that are inactive as templates. Eventually the RNA concentration reaches a steady state where RNA synthesis is balanced by loss through double-strand formation.

The laws of selection can be readily investigated by using two different template RNAs that can be separated by physical methods and are neither interchangeable by mutation nor tending to form hybrid heteroduplices [15,28,29]. Selection between them depends strongly upon the growth conditions: When the RNA concentration is low and enzyme and nucleotides are in large excess, each RNA species grows exponentially without interference by the other one. The selection kinetics is inferred from Eigen's theory to be simple for this case—the selection coefficient is a constant, namely the exponential growth constant of the RNA species. The selection process was indeed found to behave as predicted [12,33,10].

When the enzyme is saturated with template the selection criterion is completely different. Often a species with a lower replication rate is selected [6,15,28,29]. In simple terms this reflects the fact that when a critical resource is limiting the growth, competition for that resource determines the outcome of the selection. For the Spiegelman system the selection kinetics under enzyme saturation conditions depends on the rate of binding the enzyme and the concentration of free enzyme. Since the latter depends on the concentration of the RNA species, selection values in the linear growth phase are not constant [15,16]. When the RNA concentration is high, selection for minimizing the template concentration also comes into play. It can even happen that two species not forming heteroduplices may form a stable two-component ecosystem where each species is present at a steady concentration determined by its rates of enzyme binding and double strand formation.

Computer modelling of the replication kinetics by numerical integration of the rate equations using the measured rate coefficients for the most important steps and reasonable estimates for the rate coefficients of the other steps has provided useful insight. Comparisons of the concentration profiles obtained by computer simulation with the rather complicated experimental profiles eventually showed excellent agreement. Such modelling has been applied successfully to RNA replication by $Q\beta$ replicase [13–15] as well as to the mechanism of HIV reverse transcriptase [30]. The simulated selection behavior of several competing RNA species [15,16] also matched the complicated experimental results, allowing the calculation of selection values. Depending on the experimental conditions, positive, negative and zero selection values are observed. The importance of this is that only in a few limiting cases is it possible to find compact equations as analytical solutions of the rate equations [16].

Carrying the understanding achieved for the replication process forward, modelling was even shown to describe the essential features of the entire infection cycle of plus-strand RNA phages [31] and the infection dynamics [32].

The fidelity of RNA replication is much lower than the fidelity of DNA replication. For intracellular synthesis of RNA by transcription from DNA, the accuracy obtainable by nucleotide discrimination suffices—no form of proofreading or postsynthetic repair has evolved because mRNA is a short-lived transient intermediate anyway and production of a few incorrect tRNA or ribosomal RNA sequences was found in evolution to be bet-

ter tolerated than suppressed; occasional mistakes are neither toxic nor transmitted to offspring. For RNA viruses this is another matter, but viruses with RNA genomes are nevertheless quite frequent and thus obviously evolutionary successful. Indeed, the relatively high mutation rate compared to DNA replication proves to confer an advantage for mutational escape from defenses evolved by targeted hosts. On the other hand, no cellular organism that uses RNA as its genetic information has ever been discovered. The explanation, as expressed by the "stability criterion" of Eigen's theory [33]: Genome sizes of RNA viruses are limited by the replication fidelity of their replication apparatus to the range of 10,000 nucleotides. This is one essential reason why viruses have to exploit the genetic information provided by the host cell for most of their needs.

It was predicted by Eigen that nucleic acid populations are never homogeneous in sequence but rather a mixture of different species centered around a master sequence [33,34; Brakmann and Runzheimer, this issue]. The validity of this prediction was rapidly verified experimentally: All RNA virus populations studied show high sequence heterogeneities [34–37].

Eventually, the genomic distribution in an RNA population reaches a steady state, the quasispecies distribution, where each mutant type represents a constant fraction of the total population, its mutant frequency. As Eigen showed, the mutant frequency is by no means a simple function of the mutation frequency from wild type to mutant, but rather a complicated one of its synthesis rate by erroneous synthesis from all templates present (mutational gain) as well its success in competing with the other members of the population (selective value).

The validity of this prediction was confirmed by analysis of mutation spectra of a short (86 nucleotides) RNA known as MNV-11 replicating with $Q\beta$ replicase [38]. Even though the mutation frequency of the replicase has been deduced to lie between 10^{-3} and 10^{-4} per nucleotide incorporated [39], a broad mutant distribution was found: The consensus sequence never made up more than 40% of the total population and was accompanied by many mutants. Most mutants had several base exchanges, insertions and/or deletions; up to 9 of the total 86 nucleotides were changed. The mutants found had replication rates comparable to that of the wild type and were thus enriched in the population by selection forces. When the growth conditions were changed, the mutant distribution center shifted. The published consensus sequence of MNV-11 did not have the highest growth rate of all mutants—it was instead the one best adapted to the combined selection forces governing the growth phases the replicating RNA went through, i.e., it had found an optimal compromise between the rates of overall replication, enzyme binding and template loss through double strand formation [38].

The quasispecies nature of RNA populations has high biological significance: RNA viruses are able to adapt rapidly to changes in the environment, e.g., the defense of a host organism by an immune response. Apparently some of the "new" virus species that appear to arise suddenly do so because they were already present as rare mutants in a quasispecies distribution of another species that could grow with an altered host range and were thus able, indeed compelled, to grow out as parasites of the different host to become a new wild type [40].

When selection pressure is applied in in vitro evolution experiments, the quasispecies distribution adapts rapidly to the new conditions. The adaptation scenario is different from what had been assumed before Eigen's theory was introduced: Upon a change of experimental conditions, selection of the best adapted mutants already present in the population occurs first; once these mutants have accumulated, further mutations occur until eventually a new quasispecies has formed. For controlled evolution experiments, it is advantageous to keep the evolving population in the exponential growth phase. Because better-adapted mutants are formed unpredictably, causing changes in the growth rates during the evolution experiments, it is necessary to record the replication speed in real time and to trigger the next serial transfer automatically before saturation of the replicase. The design of such experiments and the details of selection are described elsewhere in this issue [Strunk and Eder-

hof; Schober; Koltermann and Kettling].

RNA replication by RNA replicases has two features that are still incompletely understood [41]. The first is the strict selectivity, i.e., the fact that only a few RNA species can be replicated, in contrast to the near universality of DNA replication even in the most primitive procaryotes. The second is the mechanism of strand separation, i.e., the fact that a single-stranded RNA template is used and a complementary single-stranded replica is produced. Sequence comparisons among a great number of RNA species replicated by Qβ replicase [42] did not reveal sequence homologies except for the invariant ends. Only one structural element required for replication has been identified—for both plus and minus strands, "leader" stem structures were found at the 5′ termini, while the 3′ termini were unpaired. Parallel structures in complementary strands rather than antiparallel ones require the occurrence of wobble pairs and other imperfections in stem regions. This suggests that the RNA structures of template and replica participate directly in the catalysis itself, in particular in the separation of strands during replication [41,42].

To confirm that the leader structures are required for replication, artificial RNA sequences were synthesized by transcription from synthetic oligodeoxynucleotides with T7 RNA polymerase and assayed for their ability to be replicated by Qβ replicase. A synthetic short RNA species known to be replicated was amplified to form a stable quasispecies whose consensus sequence was conserved for hundreds of replication rounds. A synthetic mutant of this sequence that stabilized the leader in one strand but favored a 3′ terminal stem in the other one led to complete loss of template activity. In another experiment, completely new RNA sequences with the described structural requirements were designed and synthesized. It was found that their template activity was too low to be measured directly. After longer incubation with replicase, however, replicating RNA with a consensus sequence closely related to the synthesized RNA species was formed [43].

This observation can be explained in terms of the sequence dependence of competitive advantage, the so-called "fitness landscape"—the designed sequences were originally located in a low mountainous region and consequently could be optimized during amplification by Qβ replicase to evolve to a nearby fitness peak. The structural features postulated to be required for replication were not only conserved but even improved in the outgrowing mutants. These results corroborate the postulated structural requirement, but they also indicate that still other features are required to achieve measurable replication efficiency [43]. Polypyrimidine clusters in the template are favorable for replication [44, 45] and RNA species optimized for replication have been shown to contain structural motifs similar to the replicase binding site of the phage Qβ genome [46].

The active participation of the RNA template in the replication mechanism suggests a third fundamental question: Is it possible that an RNA replicated by one viral replicase can instruct other enzymes to replicate it? For example, DNA-dependent RNA polymerases have been shown to accept RNA as template when offered a suitable one, resulting in autocatalytic amplification of the RNA template if both complementary strands are accepted. Several RNA species replicating with T7 RNA polymerase have been characterized and sequenced [47,48]. While the primary sequences of the different replicating RNA species showed few sequence homologies, their secondary structures did prove to have a common building principle: The structures were hairpins, the two half sequences of which were partly palindromic. The replication mechanism was found to be analogous to that observed in replication by phage-instructed RNA-dependent RNA replicases. The signal to carry out replication does not reside solely in the RNA part of the replicating complex, however: RNA species replicated by Qβ replicase are not accepted by T7 RNA polymerase and vice versa. Limited cross-reactivity of templates for the RNA polymerases of the phages T7 and T3 has been observed [48].

In summary, quantitative rate and sequence studies on in vitro replication and evolution of RNA have confirmed the predictions of Eigen's theory of molecular evolution in a variety of ways. We anticipate that the questions which remain about the structural aspects of the replication

mechanism will eventually be resolved, bringing the kinetics of RNA replication to the same level of understanding that physical organic chemists have achieved for reactions involving molecules that do not evolve as a consequence of their reactions.

4. Creation of stable biological information from nowhere

The experiments discussed so far were concerned with Darwinian adaptation to the environment, i.e., optimization of a pre-existing biological function. What about generating self-replicating molecules without providing any templates at all? Most biologists would be reluctant to consider such a scenario possible: The medieval belief in vermin emerging de novo from organic waste was clearly disproved by experiments, culminating in Pasteur's experiments that led to the "Omne vivum de vivo" dogma. However, most biologists would also admit that evolution is able not only to adapt but also to create. It is doing so at the molecular level that is the sticking point.

The essential difficulty arises from the extremely low probability of creating a well defined new molecular function by pure chance. If 50 nucleotides are required to build a specified RNA, then if created by a random synthesis there would be only one copy of it, on average, in $4^{50} = 10^{30}$ strands, or 580 tons of RNA. Fortunately, as we saw in the previous section, only a certain structure and a small number of invariant bases is required to provide a function, e.g., specific binding of a certain protein. The large number of total blanks in the lottery is compensated in part by the large number of minor wins; sequences with low fitness values, once any are created by chance, can be improved by instructed evolution.

A feasible procedure for creating new information is to extract it by amplification starting from a huge library containing randomly assembled sequences. The first example of such a process to be reported was selection of an RNA amplified by E. coli RNA polymerase out of a random RNA mixture [21]. The random nucleotide copolymer starting material was synthesized by using as catalyst the nonspecific enzyme polynucleotide phosphorylase. This success was possible essentially because self-replicability is a particularly sensitive function to select for, as the great majority of the strands could not be amplified at all and disappeared from the scene after a few serial transfers. New RNA species replicated by $Q\beta$ replicase have also been made by this method [44,45]. Artificial selection from a bank of random nucleotide sequences [49] will be discussed further in the next section.

Natural selection of a replicating RNA molecule has also been achieved without providing any RNA template molecules at all to start with.

Prolonged incubation of an RNA replicase at concentrations substantially higher than those present in phage infections, in the presence of high nucleotide triphosphate concentrations but in absence of any detectable RNA or oligonucleotides, does produce self-replicating RNA [50–52]. The enormous amplification factors implied by going from a single strand to macroscopic appearance makes the interpretation of the underlying process difficult, and so the mechanism of the process remains controversial [52–55]. It could be ruled out that the outgrowing RNA strands were present in the incubating solution as an impurity. (Partial instruction by non-replicating material cannot be excluded, but such effects were below the level of detection.) Evidence has been presented in support of a mechanism in which slow condensation of precursors to a random mixture of short sequences is followed by selection and optimization of strands that are able to replicate effectively [52]. In each individual aliquot of incubation mixture different sequences were selected [51,42] even when the "template-free" incubation mixture was subdivided after an amplification period. In the presence of non-replicatable RNA, replicating species can be formed by recombination events [56], probably by a copy-choice mechanism [57].

It has been found that DNA-dependent RNA polymerases can also create, in template-free synthesis, RNA that can be amplified specifically by that RNA polymerase, e.g., the RNA polymerase from E. coli [Wettich and Biebricher, unpublished] and bacteriophage T7 [48]. It seems likely that many other enzymes will be found to show this phenomenon.

5. Molecular Selection of RNA with a Function

The evolving systems described so far concern RNA dependent on and adapted to strongly selective non-evolving amplification systems. Because natural selection in these systems is particularly simple, they provide excellent paradigms for studying molecular evolution itself. For evolving new functions they are less suitable, however, because too many constraints for selecting the function are imposed, particularly if artificial selection is desired.

Two strategies for evolutionary development of new or altered molecular functions are available, rational and "irrational" design.

Rational design is a well-tested method that may be feasible for well-understood systems. It is frequently used to modify protein properties [58]. Mutations are site-directed, not random, introduced say to modify a substrate binding pocket or the site of a catalytic function. It is evolutionary in the sense that it relies on evaluation and selection of successful mutants, and generally to be preferred for modifying existing functions.

If one has no well-characterized starting point, an irrational design approach has to be taken. While much discussed, this approach is hardly new: Government agencies and pharmaceutical companies have long been randomly screening biogenic and anthropogenic chemicals for drug activity. Once a molecule with biological activity is identified by chance, further searching concentrates on derivatives of it or on molecules with related structures.

In principle, irrational design does not depend on prior knowledge of how the function is achieved. In reality, however, random shotgun approaches have such low expectation of success that mixed approaches are much preferred: Whatever constraint information is available is used to restrict the shotgunning to a narrow field and keep expenses within bounds. (It is of interest that irrational design methods have been used successfully even for such practical ends as engineering complex machines [59]; genetic and evolutionary computing algorithms have been developed for them to reduce the number of trials, improve efficiency and avoid getting trapped in local optima [60].)

In the design of drugs or specific catalysts the so-called combinatorial chemistry is a recent development [61–63]. Like biopolymers, these substances are polymers having a defined sequence of monomeric blocks with suitable chemical side groups; also as with biopolymers, the combinatorial possibilities increase rapidly with chain length and systematic searching of longer chains is not feasible.

Whatever success one may anticipate for a shotgun molecular design approach on theoretical grounds, an effective screening method to identify hits is needed. The ability of RNA self-replication to amplify sequences as well as to evolve them would make replicating RNA a promising basis for selecting a function [64], despite the limited reactivity of its functional groups. The development of reliable general amplification systems for RNA would be a crucial step in this development. While RNA replication by phage replicase is a comparatively simple reaction, it is found to be just too discriminatory for particular sequences to be generally applicable. At the other extreme, while DNA replication is essentially sequence-independent, the DNA replication apparatus of even simple prokaryotes [65] is far too complicated to be suitable for an in vitro amplification system.

Several successful in vitro amplification systems have been devised over the past decade, all requiring several steps to recycle the template (Table 1). The first and still most popular one is the polymerase chain reaction (PCR): A DNA double strand is melted into single strands by heating, oligonucleotide primers corresponding to the 5′ termini of both complementary strands are annealed at low temperature and completed to full strands by a DNA polymerase [66]. Iterating these reaction steps doubles (almost) the DNA concentration in each cycle, i.e., the concentrations increase exponentially (Fig. 2). The PCR method gained technical dominance by using a thermostable DNA polymerase that does not need to be renewed at each replication step [67]; the temperature cycling needed has been automated in commercial units. Several related DNA amplification methods have been invented.

Enzyme	Template	Product	Primer-dependent?	Sequence-specific?	Role in vivo	Source
DNA polymerase	ss-DNA cs	ds-DNA	Yes	– –	DNA repair and replication	all organisms
RNA polymerase	ds-DNA rc	ss-RNA	No	+ –	transcription	all organisms
Reverse transcriptase	ss-RNA ss-DNA cs	RNA:DNA ds-DNA	Yes Yes	– –	virus replication	retroviruses
RNA replicase	ss-RNA rc	ss-RNA	No	+ +	virus replication	levivirus-infected bacteria
Chemical	ss-RNA cs	ds-RNA	Partial	–	none	artificial

Table 1: Properties of DNA und RNA polymerases.
Notes: cs–template consumed; rc–template recyled; ss–single-stranded; ds–double-stranded. The template-instructed chemical polynucleotide synthesis included as the final row refers to the cited papers of Orgel et al. [18–20]; in those experiments the primer requirements and sequence specificities were found to vary widely with the nature of the template. The level of sequence specificity indicated for enzyme-catalyzed polymerization ranges from essentially none for natural DNA polymerases to very strict for RNA replicases: the + – notation for RNA polymerase indicates the highly sequence-specific initiation followed by unspecific elongation, the + + notation for RNA replicase indicates sequence specificity for both initiation and elongation, and the – – notation for DNA polymerase indicates lack of sequence specificity for both. Not included are nucleic acid synthesizers, for which the template exists only conceptually rather than as a complementary nucleic acid sequence, and polymerases such as poly(A) polymerase or polynucleotide phosphorylase that do not copy a template.
The basic requirement for designing an amplification system can be traced in the table: The combined steps must recycle template and lead to a net increase in nucleic acid concentration.

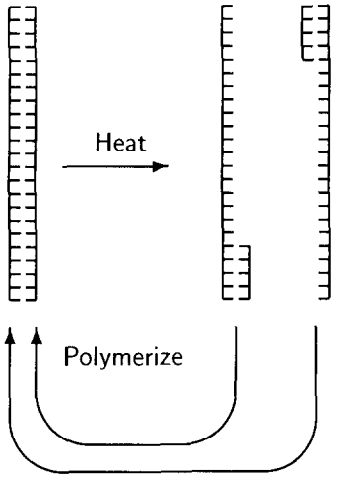

Fig. 2. Amplification mechanism of the PCR method. In each cycle of melting and oligomer-initiated replica elongation the DNA concentration approximately doubles. The times for melting and polymerizing are adjusted so that the reactions go to completion. Because of the synchronization by the temperature cycle, possible differences in the rates of hybridizing the primers and elongation by DNA polymerase do not matter: the kinetics of different species in the solution are identical, as long as the primers match.

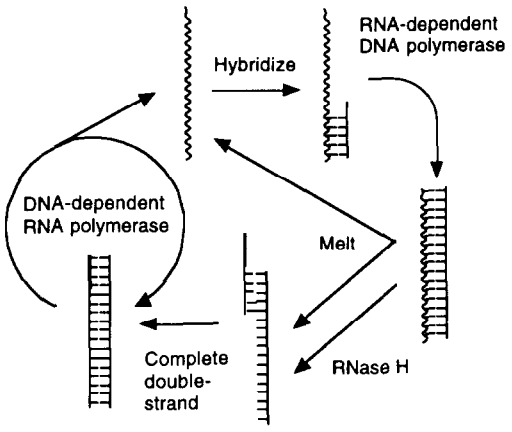

Fig. 3. Amplification mechanism of the 3SR method. Retrotranscription of the RNA (wavy lines) produces an RNA:DNA hybrid. The cDNA strand is set free either by heating or by RNase H digestion and completed by a DNA polymerase to a double strand with a primer containing a promoter sequence at its 5′-terminus. From the resulting double-strand an RNA polymerase can synthesize multiple copies of the RNA single strand.

Another in vitro amplification sytem provides cross-catalytic RNA amplification by a network of transcription and retrotranscription called self-sustained sequence replication (3SR; Fig. 3). An RNA template is transcribed into cDNA by a reverse transcriptase. The DNA strand is released either by heating [68] or by digesting the RNA template with RNase H [69]. The resulting single-stranded DNA is completed to a double-stranded cDNA by a DNA polymerase using a primer containing a promoter sequence followed by the 5′ terminal sequence of the RNA template. The resulting double-stranded DNA is used to produce some 100 strands of RNA by RNA polymerase. While this experimental procedure is more complicated than direct RNA replication, there are few constraints on the RNA sequence. The temperature cycle method of RNA amplification has been mostly replaced by its isothermal alternative using RNase H. Because reverse transcriptases have DNA polymerase and RNase H activity as well as retrotransciption activity, sometimes two enzymes suffice for the entire amplification procedure [70]. Automation is not required, although for large amplification factors a serial transfer machine is useful. An advantage of the 3SR method is that it produces single-stranded RNA of only one polarity; thus interferences by the antisense strand produced by the other amplification methods are avoided. The success of the 3SR amplification method in fulfilling the fundamental requirements of an invitro exolution system (Fig. 4) is not surprising.

If an enrichment method is not inherent to the experiment, the clones obtained in a shotgun approach have to be screened for activity. Such screening is generally quite slow and must be automated to obtain a sufficiently high throughput. Eigen and coworkers have described a 960-channel fluorimeter to carry this out [71; Strunk and Ederhof, this issue].

In most cases studied so far, the function under study has involved—among other interactions—binding of RNA to a substrate molecule. This binding can be in an affinity chromatography column onto which the substrate has been immobi-

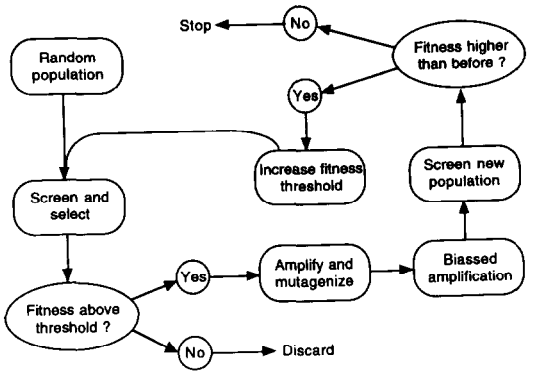

Fig. 4. General flow diagram for in vitro molecular evolution experiments.

lized (SELEX) [22,72]. Classical test-tube experiments involving cycles of amplification, mutagenesis and selection have led to optimization of RNA sequences that bind to specific targets. A surprisingly broad collection of aptamers, i.e., RNA able to bind highly selectively to different substances, has been produced by this technique [73–78].

Efficient catalysis was long considered the sole domain of proteins. Evolutionary approaches to devise proteins with altered catalytic properties by site-directed mutagenesis were highly successful [79] and new promoter elements found by an evolutionary approach have been reported [80]. Success has also been achieved in creating proteins that catalyse non-biological reactions by selecting for binding to an analog of the presumed transition state of a reaction, followed by screening for catalytic enhancement of the reaction rate over its background value. The selection system used was based on generation of antibodies against molecules with the presumed transition state structure [81–83]. Unfortunately, a catalytic antibody approach is inherently able to evolve only the substrate binding aspect of the catalytic process. Because catalysis also requires chemical events like electron or proton transfer, which can only arise by chance in the antibody selection process, i.e., can not be evolved, catalytic antibodies offer only limited enhancement of background reaction rates.

Unfortunately, a simple selection protocol for proteins has not been found so far, essentially because an efficient in vitro translation system has not yet been devised. The in vivo part has been considerably simplified by the technique of phage display [84–87], whereby functional proteins are inserted into a phage coat protein. Phages with recombinant phage coat proteins are screened and advantageous mutants selected by binding the phage to an immobilized hapten and amplifiying phages that bind in a suitable host. Sequencing and further mutations are readily possible, because the genetic information for the selected protein is included in the phage genome. Phage display has been done so far mostly using the DNA bacteriophage M13, but RNA phages have a much higher potential for adaptation [Skamel, Besendahl and Biebricher, in preparation].

The exclusive role of proteins as biological catalysts was disrupted by the discovery of ribozymes, catalysts composed solely of RNA. Natural ribozymes [88–90] have so far been found to participate only in phosphoester transfer and phosphoester hydrolysis reactions. However, it has been recognized that the roles of RNA that participates in gene expression, e.g., ribosomal RNA, include catalytic functions [91–93], as predicted already in 1968 by Crick [91]. The putative ribozyme in 23S rRNA is thought to accelerate the peptidyl transfer reaction [93].

The facile amplification of RNA has stimulated an active search for RNA that catalyses novel biological or nonbiological reactions. Despite the absence of RNA monomers with reactive side groups, ribozymes with an astonishing repertoire of reactions have been evolved. Selection from randomized or partially randomized sequences, followed by iterative cycles of amplification and selection, has facilitated the design of artificial ribozymes with catalytic functions of RNA ligase [95-97], aminoacyl-RNA synthetase [98,99], polynucleotide kinase [100], RNA polymerase [101] and alkylation [102]. Many of these novel ribozymes were created entirely de novo, but evolutionary optimization of existing ribozymes or

extending their catalytic repertoire to similar reactions by coupling the wanted activity with a selective advantage has also been highly successful, as reported by Joyce and coworkers [103–105]. It has been demonstrated that even DNA single strands can catalyze specific reactions [106]. Binding of RNA to an analog of a Diels-Alder transition state analog has been reported [107], although catalytic rate enhancement could not be detected. Since RNA lacks groups with effective acid-base or redox reactivities, doping RNA with noncanonical bases in a postreplication modification reaction—as occurs naturally in tRNA modification—or incorporation of nucleotide analogs [108] offer further possibilities to extend the repertoire of ribozymic functions. It is reasonable to anticipate that artificial coenzymes will soon be incorporated into ribozymes. Ribozymic self-incorporation of a coenzyme has already been realized [108].

We expect that research on functional nucleic acids will produce many discoveries even more spectacular than the ones that have emerged in the still very brief history of this field. We also expect that the principles of evolution explored within the framework of Eigen's theory using experimental systems where the only nucleic acid function was self-replication will play vital roles in molecular-level understanding of these discoveries.

Acknowledgments

The authors wish to express their gratitude to Manfred Eigen for his inspired and inspiring contributions to the research projects we have undertaken together with him over the years. C.K.B acknowledges support by the European Commission and by the Deutsche Forschungsgemeinschaft. W.C.G. expresses his thanks for financial support by the Fulbright, Guggenheim, Humboldt, Thyssen and Robert A. Welch Foundations that enabled him to contribute to parts of the research described in this review.

References

1. T. Loeb and N.D. Zinder, Proc. Nat. Acad. Sci. USA 47 (1961) 282.
2. I. Haruna, K. Nozu, Y. Ohtaka and S. Spiegelman, Proc. Nat. Acad. Sci. USA 50 (1963) 905.
3. S. Spiegelman, I. Haruna, I.B. Holland, G. Beaudreau and D.R. Mills, Proc. Nat. Acad. Sci. USA 54 (1965) 919.
4. D.R. Mills, Peterson, R.L. and S. Spiegelman, Proc. Nat. Acad. Sci. USA 58 (1967) 217.
5. R. Levisohn and S. Spiegelman, Proc. Nat. Acad. Sci. USA 63 (1969) 807.
6. C.K. Biebricher, in: Evolutionary Biology, ed. M.K. Hecht, B. Wallace and G.T. Prance, Vol. 16 (Plenum Press, New York, 1983), p. 1.
7. D.L. Kacian, D.R. Mills and S. Spiegelman, Proc. Nat. Acad. Sci. USA 69 (1972) 3038.
8. D.R. Mills, F.R. Kramer and S. Spiegelman, Science 180 (1973) 916.
9. D.R. Mills, F.R. Kramer, C. Dobkin, T. Nishihara and S. Spiegelman, Proc. Nat. Acad. Sci. USA 72 (1975) 4252.
10. F.R. Kramer, D.R. Mills, P.E. Cole, T. Nishihara and S. Spiegelman, J. Mol. Biol. 89 (1974) 719.
11. J. Maynard-Smith, Evolutionary Genetics, Oxford University Press, Oxford, England, 1989.
12. M. Eigen, Naturwissenschaften 58 (1971) 465.
13. C.K. Biebricher, M. Eigen and W.C. Gardiner, Biochemistry 22 (1983) 2544.
14. C.K. Biebricher, M. Eigen and W.C. Gardiner, Biochemistry 23 (1984) 3186.
15. C.K. Biebricher, M. Eigen and W.C. Gardiner, Biochemistry 24 (1985) 6550.
16. C.K. Biebricher, M. Eigen and W.C. Gardiner, in: Biologically inspired Physics, ed. L. Peliti, NATO ASI Series B, Vol. 263 (Plenum Press, New York, 1991), p. 317.
17. S.L. Miller and L.E. Orgel, The Origins of Life on Earth, Prentice-Hall, Englewood-Cliff, N.J., 1974.
18. L.E. Orgel and R. Lohrmann, Acc. Chem. Res. 7 (1974) 368.
19. T. Inoue and L.E. Orgel, J. Mol. Biol. 162 (1982) 201.
20. L.E. Orgel, Cold Spring Harb. Symp. Quant. Biol. 52 (1987) 9.
21. C.K. Biebricher and L.E. Orgel, Proc. Nat. Acad. Sci. USA 70 (1973) 934.
22. C. Tuerk and L. Gold, Science 249 (1990) 505.
23. D.R. Mills, T. Nishihara, C. Dobkin, F.R. Kramer, P.E. Cole and S. Spiegelman, in: Nucleic Acid–Protein Recognition, ed. H.J. Vogel, (Academic Press, New York, 1977), p. 533.

24. D.R. Mills, C. Dobkin and F.R. Kramer, Cell 15 (1978) 541.
25. C. Dobkin, D.R. Mills, F.R. Kramer and S. Spiegelman, Biochemistry 18 (1979) 2038.
26. C.K. Biebricher, M. Eigen and R. Luce, J. Mol. Biol. 148 (1981) 391.
27. C.K. Biebricher, S. Diekmann and R. Luce, J. Mol. Biol. 154 (1982) 629.
28. C.K. Biebricher, Chemica scripta 26B (1986) 51.
29. C.K. Biebricher, Cold Spring Harb. Symp. Quant. Biol. 52 (1987) 299.
30. M.P. Pop and C.K. Biebricher, Biochemistry 35 (1995) 5054.
31. M. Eigen, C.K. Biebricher, M. Gebinoga and W.C. Gardiner, Biochemistry 30 (1991) 11005.
32. C.K. Biebricher, M. Eigen, W.C. Gardiner, Y. Husimi, H.-C. Keweloh and A. Obst, in: Complex Chemical Reaction Systems: Mathematical Modelling and Simulation, ed. J. Warnatz and W. Jäger, (Springer-Verlag, Berlin, 1987), p. 17.
33. M. Eigen and P. Schuster, Naturwissenschaften 64 (1977) 541.
34. M. Eigen and C.K. Biebricher, in: RNA Genetics, ed. E. Domingo, P. Ahlquist and J.J. Holland, Vol. III: Variability of RNA Genomes (CRC Press, Boca Raton, Fl, 1987), p. 211.
35. E. Domingo, D. Sabo, T. Taniguchi and C. Weissmann, Cell 13 (1978) 735.
36. E. Domingo and J.J. Holland, in: RNA Genetics, ed. E. Domingo, P. Ahlquist and J.J. Holland, Vol. III: Variability of RNA Genomes (CRC Press, Boca Raton, Fl, 1988), p. 3.
37. E. Domingo, J. Holland, C. Biebricher and M. Eigen, in: Molecular Basis of Viral Evolution, ed. A. Gibbs and C.H. Calisher, (Cambridge University Press, Cambridge, 1995), p. 181.
38. N. Rohde, H. Daum and C.K. Biebricher, J. Mol. Biol. 249 (1995) 754.
39. J.W. Drake, Proc. Nat. Acad. Sci. USA 90 (1993) 4171.
40. M. Eigen, Sci. Am. 369 (1993) N1 42.
41. C.K. Biebricher, Ber. Bunsenges. 98 (1994) 1122.
42. C.K. Biebricher and R. Luce, Biochemistry 32 (1993) 4848.
43. H. Zamora, R. Luce and C.K. Biebricher, Biochemistry 34 (1995) 1261.
44. D. Brown and L. Gold, Biochemistry 34 (1995) 14765.
45. D. Brown and L. Gold, Biochemistry 34 (1995) 14775.
46. J. Beekwilder, The secondary structure of the RNA genome of bacteriophage $Q\beta$, PhD thesis, University of Leiden, 1996.
47. M.M. Konarska and P.A. Sharp, Cell 63 (1990) 609.
48. C.K. Biebricher and R. Luce, EMBO J. 15 (1996) 3458.
49. S.A. Kauffman, J. theor. Biol. 157 (1992) 1.
50. M. Sumper and R. Luce, Proc. Nat. Acad. Sci. USA 72 (1975) 162.
51. C.K. Biebricher, M. Eigen and R. Luce, J. Mol. Biol. 148 (1981) 369.
52. C.K. Biebricher, M. Eigen and R. Luce, Nature (London) 321 (1986) 89.
53. Hill, D. and Blumenthal, T., Nature 301 (1983) 350.
54. C.K. Biebricher, M. Eigen and J.S. McCaskill, J. Mol. Biol. 231 (1993) 175.
55. J.S. McCaskill and G.J. Bauer, Proc. Nat. Acad. Sci. U.S.A. 90 (1993) 4191.
56. A.V. Munishkin, L.A. Voronin and A.B Chetverin, Nature 333 (1988) 473.
57. C.K. Biebricher and R. Luce, EMBO J. 11 (1992) 5129.
58. A.R. Fersht, J.-P. Shi, A.J. Wilkinson, D.M. Blow, P. Carter, M.M.Y. Wayre and G.P. Winter, Angew. Chem. Int. Ed. Engl. 23 (1984) 467.
59. I. Rechenberg, Evolutionsstrategie, Problemata Formann-Holzboog, Stuttgart-Bad Cannstatt, 1973.
60. S. Forrest, Science 261 (1993) 872.
61. R.M.J. Liscamp, Angew. Chem. Int. Ed. Engl. 33 (1994) 633.
62. S. Brenner and R.A. Lerner, Proc. Nat. Acad. Sci. USA 89 (1992) 5381.
63. G. Lowe, Pure Appl. Chem. 56 (1984) 967.
64. A. Kornberg, DNA replication, Freeman, San Francisco, Calif, 1980.
65. K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich, Cold Spring Harbor Symp. Quant. Biol. 51 (1986) 263.
66. R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis and H.A. Erlich, Science 239 (1988) 487.
67. D.Y. Kwoh, G.R. Davis, K.M. Whitfield, H. Chapelle, L. DiMichele and T.R. Gingeras, Proc. Nat. Acad. Sci. USA 86 (1989) 1173.
68. J.C. Guatelli, K.M. Whitfield, D.Y. Kwoh, K.E. Barringer, D.D. Richman and T.R. Gingeras, Proc. Nat. Acad. Sci. USA 87 (1990)

1874.
69. M. Gebinoga and F. Oehlenschläger, Eur. J. Biochem. 235 (1996) 256.
70. A. Schober, N.G. Walter, U. Tangen, G. Strunk, T. Ederhof, J. Dapprich and M. Eigen, BioTechniques 18 (1995) 652.
71. A.D. Ellington and J.W. Szostak, Nature 346 (1990) 818.
72. P. Niewlandt, D. Decker and L. Gold, Biochemistry 34 (1995) 5651.
73. M. Wecker, D. Smith and L. Gold, RNA 2 (1996) 982.
74. C. Lauhon and J. Szostak, FASEB J. 7 (1993) A1087.
75. M.G. Wallis, U. von Ahsen, R. Schroeder and M. Famulok, Chem. & Biol. 2 (195) 543.
76. L. Jiang, A.K. Suri, R. Fiala and D.J. Patel, Chem. & Biol. 4 (1996) 35.
77. S.M. Lato, A.R. Boles and A.D. Ellington, Chem. & Biol. 2 (1995) 291.
78. K.M. Munir, D.C. French and L.A. Loeb, Proc. Nat. Acad. Sci. USA 90 (1993) 4012.
79. M.S.Z. Horwitz and L.A. Loeb, J. Biol. Chem. 263 (1988) 14724.
80. R.A. Lerner and A. Tramontaur, Sci. Am. 258 (1988) N3 42.
81. P.G. Schultz, Angew. Chem. Int Ed. 28 (1989) 1283.
82. P.G. Schultz and R.A. Lerner, Acc. Chem. Res. 26 (1993) 391.
83. S.F. Parmley and G.P. Smith, Gene 73 (1988) 305.
84. J.K. Scott and G.P. Smith, Science 249 (1990) 386.
85. G.P. Smith and J.K. Scott, Meth. Enzymol. 217 (1993) 228.
86. A.D. Griffiths, Annales Biol. Clin. 51 (1993) 554.
87. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, and S. Altman, Cell 35 (1983) 849.
88. A.J. Zaug, and T.R. Cech, Science 231 (1986) 470.
89. A.C. Forster, and R.H. Symons, Cell 49 (1987) 211.
90. A.E. Dahlberg, Cell 57 (1989) 525.
91. M. O'Connor, C.A. Brunelli, M.A. Firpo, S.T. Gregory, K.R. Lieberman, J.S. Lodmell, H. Moine, D.I. van Ryk and A.E. Dahlberg, Biochem. Cell Biol. 73 (1995) 859.
92. H.F. Noller, V. Hoffarth and L. Zimnick, Science 256 (1992) 1416.
93. F.H.C. Crick, J. Mol. Biol. 38 (1968) 367.
94. B. Cuenoud and J.W. Szostak, Nature 375 (1995) 611.
95. E.H. Ekland, J.W. Szostak and D.P. Bartel, Science 269 (1995) 364.
96. K.B. Chapman and J.W. Szostak, Chem. & Biol. 2 (1995) 325.
97. M. Illanga-Sekare, G. Sanchez, T. Nicklcs and M. Yarus, Science 267 (1995) 643.
98. P.A. Lohse and J.W. Szostak, Nature 381 (1996) 442.
99. J.R. Lorsch and J.W. Szostak, FASEB J. 9 (1995) A1422.
100. J.W. Doudna and J.A.Szostak, Nature 339 (1989) 519.
101. C. Wilson and J.W. Szostak, Nature 374 (1995) 777.
102. A.A. Beaudry and G.F. Joyce, Science 257 (1992) 635.
103. X.C. Dai, A. Demesmaeker and G.F. Joyce, Science 267 (1995) 237.
104. R.R. Breaker and G.F. Joyce, Chem. & Biol. 2 (1995) 655.
105. J.A. Piccirilli, T. Krauch, S.E. Moroney and S.A. Benner, Nature 343 (1990) 33.
106. K.N. Morris, T.M. Tarasow, C.M. Julin, S.L. Simons, D. Hilvert and L. Gold, Proc. Nat. Acad. Sci. USA 91 (1994) 13028.
107. R.R. Breaker and G.F. Joyce, J. Mol. Evol. 40 (1995) 551.