Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Mutation, Competition and Selection as Measured with Small RNA Molecules

By Christof K. Biebricher

INTRODUCTION

Darwin's theory of natural selection is one of the greatest milestones in science. It provides answers to deep questions that are otherwise unanswerable. As Dobzhansky put it: "Nothing makes sense in Biology except in light of evolution" (Dobzhansky et al., 1977).

Yet even a century after Darwin we still cannot understand organismic evolution in detail, let alone make quantitative predictions about its course. To be sure, neo-darwinistic theory does provide insights into evolution processes in quantitative terms, but they essentially comprise only quantitative descriptions of reproduction, in particular in mendelian populations, not of evolution itself. For the fundamental processes operating in evolution – mutation and selection – its parameters have to be adjusted to fit, more or less, the evolutionary outcome. They cannot be derived from measurable properties of the organisms themselves or of their genes.

In contrast to our lack of quantitative descriptions of evolution, its molecular basis is very well understood: the information needed for morphogenesis and function is encoded into the genotype as the sequence of nucleotides in each organism's genome (a few exceptions notwithstanding). The key step in reproduction is copying that nucleotide sequence. While replication error rates leading to accepted mutations vary over the genome, and also depend to some degree upon environmental conditions, these errors are more or less random. No teleology directing mutations to an advantageous result has ever been observed.

Information informs only if it is understood; thus genetic information has to be decoded by cellular machinery in order to operate a life-sustaining program of biochemical reactions. The program depends on the environment and leads to the properties of the organism that we observe as its phenotype. Evolutionary success depends on two components of the phenotype: those that determine survival in the prevailing environment and those that establish the rate of producing viable offspring. The combination of these two determines the population trend that we call fitness.

Environments are typically complex and variable. Species interact by competition for resources, predation or symbiosis; individuals of the same species influence one another socially, and the individuals themselves may be composed of large numbers of specialized cells, all containing the same genetic information, that have to cooperate with one another for the organism they compose to survive and reproduce. While the way that gene expression produces the metabolic apparatus is now felt to be well understood, we know much less about how cells acquire information about the environment and how this information triggers appropriate genetic responses. Biologists still debate the identity of the target of selection: is it
an ecosystem, a species, a variant, a subpopulation, an individual, a gene or merely a "replicator unit" (Dawkins, 1982)? There is no ultimate answer to this question: selection takes place at all of these levels. Which of the selection levels dominates depends on the environment.

Evolution is a dynamical self-organization process in which causal correlations between the performance of the process as a whole and its component subprocesses are not identifiable (Biebricher et al., 1995). We can correlate fitness to the function of a single gene only if this gene happens to be absolutely required for survival under prevailing conditions. Further, translation of a genotype into a phenotype is far too complicated a process to be evaluated. Fitness values have to be determined a posteriori, i.e. so as to describe the observed changes in the composition of the population under study. Prediction of an evolutionary outcome from fitness values obtained from process-independent parameters is generally impossible. Is the darwinian concept of evolution then merely a tautomerism describing the survival of the survivors, as some have criticised? No, it is not. The studies with RNA viruses described in this volume witness that in many cases the molecular basis of fitness can be clearly identified.

Nevertheless, even the simplest RNA viruses are too complicated to allow quantitative descriptions of their mutation, competition and selection, in particular because their complex interactions with host cells are inevitably involved in the evolutionary process. The 1961 discovery of RNA bacteriophages by Loeb and Zinder, more than 80 years after the discovery of plant RNA viruses, was instrumental in accelerating progress in understanding molecular processes in the infection cycles. Ten years after their discovery they were already by far the best understood viruses. In particular, because of their essential nature as parasitic messenger RNA, they became invaluable experimental tools in studying the expression of genetic information.

### The Experimental System

Most RNA bacteriophages belong to the plus strand virus family leviviridae. Except for a few members of the reoviridae family, no other RNA virus families have been found to infect prokaryotes. Leviviridae are particularly simple, in all respects, and their genome sizes are the smallest among autonomously infecting viruses. Their properties, gene map and infection cycle are described in Figures 4.1 and 4.2.

Shortly after their discovery, several research groups succeeded in detecting a novel RNA-dependent RNA polymerase in levivirus-infected cells (August et al., 1963; Haruna et al., 1963; Weissmann et al., 1963). After being found to be highly specific in amplifying viral RNA it became known as a replicase. The replicase of the coliphage Qβ was found to be particularly stable and thus the most suitable one for in-vitro studies. Purification of the Qβ replicase (Kamen, 1970; Kondo et al., 1970) revealed four subunits, one coded by the R gene of the phage, the others provided by the host. Together with an additional host factor (Franze de Fernandez et al., 1968) they perform all steps necessary to amplify viral RNA.

The experimental procedure for replication experiments is simple. An RNA template is incubated with replicase purified from Qβ infected cells, appropriate amounts of the four nucleoside triphosphate precursors and an appropriate buffer. When using viral RNA as template, the progeny RNA synthesized in vitro was found to be infectious (Spiegelman et al., 1965). However, when Spiegelman and collaborators tried to dilute out parental RNA by serially diluting aliquots of growing RNA into fresh test tubes containing replicase and precursors, further production of infectious RNA stopped after the fifth serial transfers, while incorporation of nucleoside triphosphates into RNA continued, even at steadily increasing rates. Spiegelman and collaborators recognized that they had performed an "extra-cellular Darwinian
4. MUTATION, COMPETITION AND SELECTION AS MEASURED WITH SMALL RNA MOLECULES

FIGURE 4.2 Infection cycle of phage Q8. ●, ribosome; □, replicase; ←, 5' end of RNA; →, 3' end of RNA; △, ribosome binding site available; ▲, ribosome binding site unavailable. Reproduced from Biebricher and Eigen, 1987, in: RNA Genetics, vol. I: RNA-directed Virus Replication (eds Domingo, E., Ahlquist, P. and Holland, J.J.), CRC Press, Boca Raton, FL, with permission.

At the beginning of the infection cycle, a phage particle adsorbs to an F-pilus of the host and injects its RNA into the host cell. Ribosomes bind to the binding site of the C-gene, which is the only available site. During translation of the C-gene the RNA structure is modified in such a way that the ribosome site of the R-gene also becomes available (Weissmann, 1974). Both pC and pR proteins are synthesized, the latter combining with EFTu-Ts and ribosomal S1 protein to form mature replicase. The replicase binds to a site (rb) directly at the ribosome site of the C-gene. Therefore, replicase and ribosome compete for the RNA. If sufficient replicase has been synthesized it binds to the RNA, but cannot start replication before completion of translation and clearance from the ribosome. The replicase produces minus strands that bind neither ribosomes nor coat protein but are excellent templates for the production of plus strands. With increasing amounts of coat protein, binding of coat protein occurs at a site (cb) next to the ribosome binding site of the R-gene. Binding of coat protein dimer shuts off R-gene translation and is the nucleus for the capsid formation. A2-gene translation is only available when a nascent plus strand is formed in minus strand replication (Beekwilder et al., 1996).

The rate of RNA synthesis is hardly measurable in the first 15 min, because replicase must first be synthesized, then it explosively increases at a hyperbolic (faster than exponential) rate until 20 min after infection (Eigen et al., 1991). After all the replicase is bound (nearly exclusively to minus strand), linear growth of RNA at maximum rate is observed. Mature phages accumulate in the interior of the host cell and are liberated at an eclipse time of about 50 min. Liberation occurs by host cell lysis triggered by A2 protein (Karnik and Billeter, 1983). A2 is also responsible for binding of the phage particles to the F-pilus.
experiment with a self-duplicating nucleic acid molecule”, the title of a paper that became seminal for *in-vitro* evolution studies (Mills *et al.*, 1967). After 74 serial transfers, the RNA was analyzed and found to have eliminated 83% of its chain length in the course of increasing its replication rate. Experiments under different replication conditions, e.g. reduced levels of one nucleoside triphosphate or in the presence of inhibitors (Levisohn and Spiegelman, 1969; Saffill *et al.*, 1970) were performed; the RNA was shown to adapt to these conditions, “revealing an unexpected wealth of phenotypic differences which a replicating nucleic acid can exhibit”.

The experiment had an immediate impact. While most scientists were enthusiastic about the new possibilities, some scoffed that the experiment represented “a search for the best carcass”. Indeed, the evolution experiments did start with an infectious RNA able to perform many different roles – to serve as messenger, to replicate and to be packed in a protein coat – and end with a “variant RNA” that had lost most of the information and was only able to replicate. Evolution thus produced degeneration, as have many other experimental evolution trials.

Furthermore, the reaction and its products were declared to be “unphysiological”. Indeed, they have to be, for otherwise the consequences for phage reproduction would be disastrous. This criticism applies also to the other evolution studies reported below; the templates and the reactions are all unphysiological. This does not detract from their value, however, in studying evolution. The experiments I describe abstract the essential features from the enormously complex net of physiological interactions at work in physiological phage infection cycles *in vivo*. While replicase and RNA synthesis precursors must be synthesized during infection cycles, they are environmental factors in the *in-vitro* experiments. The *in-vitro* experiments provide precisely controllable and reproducible conditions indispensable for quantitative studies.

Some of the key results of the classical *in-vitro* RNA evolution studies were difficult to understand. The RNA species selected in many different such evolution experiments were indistinguishable from one another in their physical properties. One would have expected that RNA evolution would proceed by rare events leading to different and irreproducible products. Furthermore, in the early work there was the problem that the enzyme used was a partially purified crude extract not entirely free from RNA. Spiegelman and coworkers recognized these difficulties and searched for short-chained RNA molecules where genotypic changes could be tracked by sequence analysis (Mills *et al.*, 1973).

**THE MECHANISM OF RNA REPLICATION**

Qβ replicase has evolved features absolutely necessary for virus infection.

1. It is highly specific in accepting its own RNA as template. It is vital for phage reproduction that only the viral RNA be amplified, while a huge excess of host RNA is ignored. Therefore, the RNA itself cannot be considered to be merely a substrate of the replicase: it shares the catalytic role. Its template activity, i.e. its efficiency in instructing the replicase to replicate it, is a phenotypic expression of the RNA species that is crucial for its evolutionary success.

2. Only single-stranded RNA is accepted as template. The replica formed is complementary and antiparallel to the template, suggesting Watson–Crick base-pairing at the replication site. A few other templates, e.g. poly(C) and C-rich nucleotide copolymers that are accepted by Qβ replicase, result in a perfect double strand (Hori *et al.*, 1967; Mitsunari and Hori, 1973). Double strands are not replicated (Biebricher *et al.*, 1982) and thus synthesis using these templates stops after transcribing the RNA; neither template nor enzyme is recycled. Autocatalytic amplification of a template takes place only when template and replica strands separate during replication and are released individually in single-stranded form (Dobkin *et al.*, 1979; Biebricher, 1983). An RNA species thus always consists of both complementary sequences.

3. The replicase is highly processive, i.e. it usually does not dissociate from the template before a round is completed, because there is no way to complete a released incomplete replica strand in a subsequent reaction. As a consequence, release of the template after a replica-
tion round is the rate-limiting step in the whole cycle.

A large collection of artificial short-chained RNA species that are efficiently amplified by Qβ replicase have been selected, in most cases by the so-called template-free synthesis procedure described below under Creating biological information from scratch. They are replicated by the mechanism shown in Figure 4.3.

At the start of an amplification experiment, enzyme is typically present in large excess over RNA template. Newly synthesized replica as well as released template strands quickly bind to enzyme molecules, and exponential growth of the RNA concentration results. Once the enzyme is saturated with template, the RNA concentration increases linearly with time and the main products are free plus and minus strands. The overall replication rate in the linear growth phase is lower than in the exponential growth phase, because recycling of the enzyme becomes rate-limiting. Free complementary strands react to form double strands that are inactive as templates. Eventually, a steady state is reached where the concentration of single strands does not change, because the synthesis of new strands is balanced by loss of double-strand formation. At the final steady state, only the concentration of double-strand increases. (The double strands do bind to replicase, but they rapidly dissociate again. As the concentration of double strands increases, this slows down the synthesis rate, by lowering the concentration of available free enzyme.)

The essential chemical steps of the replication were identified in a series of experiments (Biebricher et al., 1981, 1983, 1984, 1985), and rate coefficients for some of them were
measured; for others, reasonable estimates could be introduced to enable kinetic modeling. A kinetic model set up this way is an oversimplification in the sense that the identified steps are not elementary chemical reactions. Binding of protein to RNA, for example, is not a simple bimolecular event. In reality, a cascade of chemical steps distorting bond angles, establishing short-range van der Waals interactions, hydrogen bonds and pushing out water molecules is involved for both macromolecular components. Fortunately, incorporating this level of detail in the model was not necessary to rationalize the experimental RNA concentration profiles. On the contrary, the kinetic model was able to describe the complicated experimental RNA growth profiles precisely, and is thus adequate for drawing conclusions about the roles of different parts of the replication process in determining the fitness of mutant RNA species. The quantitative determination of RNA replication is described in Box 4.1 and Figures 4.4 and 4.5.

**SELECTION OF RNA SPECIES**

When two or more RNA species are present in the starting template population, they compete with one another. If the sequences and the physical properties of the species differ sufficiently, the outcome of the selection process can be followed.

---

**Box 4.1 Quantitative measurements of the replication rate**

A simplified replication mechanism is shown in Figure 4.4. It is an adequate description for the replication time course when the replicase concentration exceeds 150 nmol/l and the concentrations of the triphosphates are higher than 0.300 mmol/l each. From the last equation in Figure 4.4, we can calculate the relative population increase \( \frac{d[I]}{d[t]} \) of type \( i \) due to its fecundity \( A_i = \frac{K_i[E][I]}{[I_0]} \), which may be formulated in a more biological language as follows: the relative rate of population increase is equal to the proportion of the population in the reproducing age times its birth rate.

Experimentally, overall replication rates are determined by radioactive nucleotide incorporation measurements. The obtained profiles are shown in Figure 4.5. Two growth phases are clearly distinguished: an exponential one where enzyme is in excess and a linear one where enzyme is saturated with template. The overall growth rate in the linear growth phase is determined from the slope of the linear part. The overall growth rate in the exponential growth phase is calculated from the time displacement \( \Delta t \) of the profile caused by dilution by the factor \( F_{di} \), according to \( \kappa = \Delta t^{-1} \ln F_{di} \). In the exponential growth phase, the intermediates and the total population show after an equilibration period coherent growth (Biebricher et al., 1983):

\[
\frac{d[I]}{d[t]} = \frac{d[E][I]}{d[t]} = \frac{d[E][I]}{d[t]} = \frac{d[I]}{d[t]} = \frac{d[I]}{d[t]} = \kappa,
\]

where \( \kappa \) is the overall (exponential) replication rate. It is easily determined from the displacement \( \Delta t \) on the time axis after dilution of the template by a factor \( F_{di} = \kappa = \Delta t^{-1} \ln F_{di} \). In the linear growth phase, virtually all of the enzyme is bound to template, and a steady state is established where the intermediate concentrations do not change and the flux through each step is equal to the total flux \( V \):

\[
V = k_a[E][I] = k_e[E] = k_{ae}[IE] = p[E],
\]

where \( p[E] = p[E]_0 \), where \( [E]_0 \) is the total enzyme concentration, free and bound. The relative fecundities \( A_i \) are constant in the exponential growth phase \( A_i = \kappa \), but decrease in the linear growth phase with increasing \( [I] \) (\( A_i = \rho[E]/[I] \)).

The population change of type \( i \) is of course also dependent on its mortality rate. Under the conditions of the described RNA replication experiments, the mortality is caused by the loss of template molecules by double-strand formation. Other contributions to mortality-like decomposition can be neglected. The loss rate can be described with the equation \( -d[I]/dt = \frac{1}{2}d[I]/dt = k_a[I] \). Hence, the relative mortality of type \( i \) is \( D_i = k_{ae}[I]/[I] \). In the exponential growth phase, the concentrations of free strands are very small and the mortality is negligible.

The net population growth is the balance between fecundity and mortality: \( E = A - D_i \). In the exponential growth phase, we obtain \( E_i = \kappa \). In the linear growth phase, \( A_i \) decreases and \( D_i \) increases until steady state is reached where \( A_i = D_i \) and \( E_i = 0 \). The steady state concentrations can be calculated when the rates values have been determined (Biebricher et al., 1984, 1991).
4. MUTATION, COMPETITION AND SELECTION AS MEASURED WITH SMALL RNA MOLECULES

\[
\begin{align*}
\frac{d[E]}{dt} &= k_A[E][I] - k_E[E] \\
\frac{d[E]}{dt} &= k_E[E] - k_D[E] \\
\frac{d[E]}{dt} &= k_A[E][I] - k_D[E] = -\frac{d[E]}{dt} \\
\frac{d[I]}{dt} &= k_E[E] + k_D[E] - k_A[E][I] \\
\frac{d[I]}{dt} &= k_D[E]
\end{align*}
\]

**FIGURE 4.4** Simplified mechanism of RNA replication. Shown are the steps involving binding and releasing of complex components. All steps involved in replica synthesis are combined into a single step.

The simplest case is exponential growth of small populations, which prevails when all resources required for amplification are present in excess. Under these conditions, each species grows as it would in the absence of the others. Each species grows independently with its own characteristic growth rate, but as this goes on the composition of the population changes: the population gradually becomes enriched in species with higher growth rates and relatively depleted in species that grow more slowly. A quantitative description of such experiments can be calculated (Box 4.2). The fitness of each species under exponential growth conditions is characterized by its fecundity, i.e. its replication rate, alone.

This strong selection makes working with an RNA replicase rather difficult. Assume that an RNA species with a replication rate one-tenth that of an optimized species has to be amplified by a factor of 10. While that happens, a single strand of the optimized species is amplified by a factor of 10^9, i.e. to macroscopic appearance! This illustrates that the techniques of amplification with replicase are technically not as easy as they might seem to be; severe precautions have to be made to avoid contamination of an RNA population with optimized species (Biebricher et al., 1993). Synchronized amplification techniques like the polymerase chain reaction are much easier to handle. Purification of an RNA species by physico-chemical methods, e.g. by electrophoresis, must always be followed by a cloning procedure, because otherwise only the fastest species will be found. If we know that the separation method has reduced impurities to a level of say 1/1000, then it suffices to start an amplification experiment with fewer than 1000 template strands to get a pure species as product.

The outcomes of selection experiments carried out under linear growth conditions are at first glance surprising. Amplification of a mixed population by a factor of 10 can result in a change in the population composition by many orders of magnitude, and often a species with a lower replication rate is selected. As observed in organismic evolution, species with low fecundity can be quite successful if they are able to outcompete their competitors for limiting resources. In the linear growth phase of RNA replication, the limiting resource is the replicase itself, and the species that

**FIGURE 4.5** Incorporation profiles. MNV-11 was added at the indicated concentrations to a replication mix containing 200 nM replicase. The incorporation of labeled AMP was measured and calculated as number of strands synthesized per enzyme molecule. Two growth phases are clearly distinguished: an exponential one where enzyme is in excess and a linear one where enzyme is saturated with template. The overall growth rate in the linear growth phase is determined from the slope of the linear part. The overall growth rate in the exponential growth phase is calculated from the time displacement \( \Delta t \) of the profile caused by dilution by the factor \( F_{\text{rep}} \) according to \( \kappa = \Delta t \ln F_{\text{rep}} \). Experimental values under optimal growth conditions of MNV-11 were \( \kappa = 1.92 \times 10^{-7} \text{s}^{-1} \) and \( \rho = 6.37 \times 10^{-4} \text{s}^{-1} \) (Biebricher et al., 1985).
is fastest in binding to newly liberated replicase molecules will be selected whatever its replication rate may be. The quantitative description is somewhat more complicated. The kinetic model described previously is a good help. To describe competition in the linear growth phase, it is sufficient to set up the rate equations such that two species share the resources and the calculated profile again precisely matches the experimental results. Instead of fitness, we work better with selection (rate) values, the relative change of the relative population in time. The definitions are listed in Table 4.1. The selection values, which vary with time and the concentration of the competitors in the linear growth phase, can be precisely determined from the computed concentration profiles (Figure 4.6).

In the late linear growth phase, the loss terms caused by double-strand formation must also be taken into account in the model. For the case that the nucleotide sequences of two competing species are rather different, formation of heteroduplex strands can be neglected. Species with low concentrations of free single strands are favored by low loss rates through double strand formation and so the population can eventually reach a steady state, where its relative composition no longer changes: a stable ecosystem has been formed.

Even under the controlled external conditions of in-vitro evolution experiments, selection patterns can thus be quite complex, basically because the growing RNA species change their own environment. A typical example would be starting with two RNA species (MNV-11 and MDV-1), for which a computer simulation is shown in Figure 4.6. Initially both species are present in small equimolar amounts and exponential growth begins for both. When the enzyme is saturated, MNV-11 has conquered, because of its higher replication rate, most of the enzyme, and shortly afterwards it reaches the steady state of double-strand formation and its selection value vanishes. However, MDV-1 continues to grow, and eventually it displaces MNV-11 from the enzyme because of its higher enzyme binding rate. Eventually an ecosystem is formed where both species coexist; their selection values have both vanished.

**MUTATION IN REPLICATING RNA**

Any alteration of a genotype is called a mutation. Mutation may occur by chemical modification of a base, such as deamination of a cytidylate to a uridylate, but most mutations are produced by an erroneous replication, i.e. the
TABLE 4.1 Symbols and parameters used for the quantitative studies of in vitro evolution.

### Concentrations

| Symbol | Definition |
|--------|------------|
| [I]    | Concentration of free single-stranded RNA of type i [mol/l] |
| [EI]   | Concentration of active replication complex [mol/l] |
| [IE]   | Concentration of inactive replication complex [mol/l] |
| [E]    | Concentration of free enzyme [mol/l] |
| [E_i]  | Total concentration of template strands of type i complexed to enzyme [mol/l] |
| [E]    | Total concentration of enzyme, bound or free [mol/l]** |
| [I]    | Total concentration of template strands of type i [mol/l]† |
| Σ[I]   | Total concentration of RNA [mol/l]* |
| [II]   | Concentration of double strands (homoduplex) of type i [mol/l]† |
| [III]  | Concentration of double strands between plus strand of type i and minus strand of type j (heteroduplex) [mol/l] |

### Rate constants

| Symbol | Definition |
|--------|------------|
| k_s   | Association rate constant for binding of replicase to RNA of type i [1/mol s⁻¹]* |
| k_e   | Rate constant for synthesizing and releasing a replica from a replication complex of type i [s⁻¹] |
| k_d   | Dissociation rate of inactive replication complex [s⁻¹]* |
| k_m   | Rate constant for double strand formation between plus strand of type i and minus strand of type j [1/mol s⁻¹]* |
| κ     | Overall replication rate constant of type i in the exponential growth phase [s⁻¹]* |
| ρ     | Experimentally measured relative rate of RNA synthesis per template strand of type i [s⁻¹]* |

### Parameter definitions

| Symbol | Definition |
|--------|------------|
| x_i   | Mutant frequency; Fraction of type i in the total population † |
| A_i   | Relative fecundity of type i [s⁻¹] |
| D_i   | Relative mortality of type i [s⁻¹] |
| E_i   | Relative net excess production rate of type i; \( E_i = A_i - D_i \) [s⁻¹]* |
| E̅   | Relative net excess production rate for all types [s⁻¹]* |
| Q_i   | Probability of producing type i per reproduction process of type j |
| Q_a   | Probability of producing a correct copy per reproduction process of type i [s⁻¹] |
| W_a   | Intrinsic selection rate value, \( W_a = Q_a A_i - D_i \) [s⁻¹] |
| μ     | Mutational gain rate (synthesis by miscopying other templates) [s⁻¹] |
| ζ     | Selection rate value [s⁻¹] |
| ξ     | Evolution rate value; relative rate for relative increase of type i, \( \xi = \zeta + \mu \) [s⁻¹] |

* parameter can be readily measured
† parameter set at beginning of experiment
‡ parameter can be readily measured when types can be easily distinguished
FIGURE 4.6  Competition among RNA species. Calculated growth profiles of two species, 1 (solid symbols) with standard values (MNV-11) and 2 (open symbols) having $k_e = 4 \times k_d$ and $k_o = 1/4 k_o$. Starting conditions were $[E] = 200$ nmol/l and $[I] = [II] = 1$ pmol/l. In the exponential growth phase (0–8 min), the smaller $k_o$ value is detrimental and species 1 grows more rapidly (see the semilogarithmic plot at upper left). It saturates the enzymes and enters the steady state, where its net growth (lower left) stops. Species 2 continues to grow exponentially at a lower rate (due to the smaller amount of free enzyme); it conquers most of the enzyme because of its higher binding rate (note the diminishing concentration of free enzyme). After 60 min the final steady state is reached where each species occupies a constant part of the enzyme. Calculations were done by numerical integration of the rate equations, using a much more detailed mechanism than shown in Figure 4.4. Approximate analytical solutions of the rate equations of the simplified mechanism can be found for certain cases. Reproduced from Biebricher et al., 1985, in Biochemistry, 24, 6550–6560, with permission.
progeny genotype differs at one or more positions from the parental genotype. Luria and Delbrück showed in a classic experiment in 1943 that the mutation event in bacteria is stochastic and that the mutated type may spread by error propagation. Different mutant types compete one with the other and selection occurs. Mutation can be studied quantitatively if selection is excluded by restricting amplification to a single replication round. Mutation rates can be measured, defined as the probability of incorporating a non-cognate base per incorporation event.

It is generally believed that DNA replaced RNA as genetic material during evolution because of its superior replication fidelity and chemical stability. Several energy-consuming error correction systems were invented to accomplish this fidelity. The systematic error caused by the tautomerization reaction of the pyrimidines is mainly corrected directly after phosphodiester formation by proof-reading. A mismatch – particularly the one caused by a base that is returning from the wrong tautomeric structure to its favored one – is removed during the replication process itself. Extensive postreplicative repair systems detect and remove imperfections in newly formed DNA double helix. Neither fidelity-enhancing method is implemented in RNA synthesis. Because cellular RNA is normally produced in many copies per cell and is degraded after some time anyway, occasional errors do not cause permanent harm. Remarkably, no repair mechanism has yet been found among viruses with RNA genomes, where mutations can be lethal. Indeed, their mutation rates are so high that only a small fraction of the copies are identical to their parents. The high price for this – that most offspring of RNA viruses are defective – is apparently offset by the higher potential to adapt to changing environments, perhaps caused by host defenses, that is ultimately provided by error-prone replication.

Leiviruses were the first examples where the high mutant diversity of RNA viruses was detected. Watanabe collected a large number of leiviruses from all continents (Yonesaki et al., 1982). The leiviridae could be grouped into four classes. Their organization was nearly identical, yet the RNA fingerprints of species belonging to different classes did not indicate any sequence relationships. Today, while several such species have been sequenced, alignment of the sequences of species from different classes is difficult, because the information of the archetype founding the phylus has been almost totally diffused by mutations.

Within each virus species, however, RNA fingerprints were remarkably stable and well-defined, indicating a clearly defined wild-type sequence (Billeter et al., 1969; Fiers et al., 1976). The apparent simplicity of this result, however, was shattered when it was shown that clones derived from single phage plaques of a virus population showed differences in their fingerprint patterns (Domingo et al., 1978). It came as a shock to realize that viral populations are predominantly composed of an array of mutants, in which only a small fraction is what one would call a wild-type genome based on the dominant occupation of each nucleotide position in the sequence. Passaging the phage with a series of lysates restored the wild-type sequence. This result left only one explanation: the wild type sequence is nothing more than the average of all of the sequences present in the viral population.

Eigen and Schuster (1977) predicted this result with straightforward theoretical considerations. Error propagation causes a spread of mutations in the population, leading eventually to a stable population, the “quasispecies”, where each mutant type maintains a constant share of the population, its mutant frequency, which depends on its production by mutation and its selective value. A high mutant frequency does not necessarily correlate with a particularly high mutation rate (“hot spot”); nearly-neutral multi-error mutants may have substantial mutant frequencies even though their rates of production by mutation are quite small. The theoretical background is covered in detail by Schuster and Stadler in Chapter 1, and Domingo et al., discuss the evolution of virus populations in vivo in Chapter 7.

From in-vitro studies (Batschelet et al., 1976) and in-vivo data (Drake, 1993) it was possible to estimate average RNA mutation rates per incorporated nucleotide; the two studies give values between $10^{-3}$ and $10^{-4}$. On average, therefore, each
phage RNA replica contains about one mutation. On the other hand, for the much shorter RNA sequences used in \textit{in-vitro} experiments, the vast majorities of the copies should be correct.

A natural mutant spectrum of the replicating RNA species MNV-11 was investigated by Rohde \textit{et al.} (1995). Each of their experiments began with a homogeneous RNA population created by cloning. A large number of serial transfers, under constant growth conditions, were then made to allow establishment of an equilibrium population. The same procedure was then repeated, starting with the same RNA clone, for different growth conditions, e.g. higher ionic strength or a different growth phase. In order to determine the sequence and other properties of the mutants in each equilibrated population, representative collections of the mutants had to be cloned. This cloning could not be accomplished by amplifying single RNA strands with Qβ replicase because of the high error rate and intrinsic bias introduced by the replicase. Lethal or seriously disadvantaged mutants, for example, would not show up at all, because they would undergo evolutionary optimization as the clones were amplified to levels where sequencing is possible. Cloning RNA first into DNA, and then amplifying the DNA, does not have these drawbacks, however, and was thus adopted for analyzing the equilibrium populations. The cloning procedure was designed in such a way that the same RNA sequence that provided each clone could be reconstructed from the DNA clone by DNA-directed RNA synthesis (Biebricher and Luce, 1993). It was shown that the RNA populations obtained by transcription were quite homogeneous. To be sure, the fidelity of transcription is no better than that of replication by viral replicase, but because transcription uses only the DNA and never the RNA copy as a template, error propagation is avoided. The sequences of some of the mutants found by Rohde \textit{et al.} are shown in Figure 4.7.

The mutant spectra were found to be quite broad. When the linear growth phase was investigated, for example, "wild-type" RNA composed less than 40\% of the quasispecies population. Mutations were not distributed randomly. At some positions mutations were frequent, while some regions were conserved, indicating parts of the RNA that are required for replication to occur. Single-error mutants were rare, and multi-error mutants appeared with up to 10\% of the positions altered. Base transitions, transversions, deletions and insertions were observed, in one case even duplication of a 7-base segment.

It is clear that the mutant spectra observed by Rohde \textit{et al.} are not simply correlated to mutation rates. Base transitions were not found more frequently than transversions, and multi-error mutants were strongly over-represented in comparison with what one would expect on the assumption that mutation rates governed the mutant spectra. Mutations themselves are essentially independent events, and if they generated the observed mutant spectra one would find a high frequency of one-error mutants, a much smaller frequency of two-error mutants, and multi-error mutants would be extremely rare. One has to conclude that the mutant spectra were governed instead by selection values. When this is true, frequently found mutants are expected to be neutral or nearly so. This was found to be true: the mutant replication rates were measured and found to be close to that of the wild type. The rate measurements also showed that the "wild type" found most frequently in equilibrated populations from the linear growth phase was not the mutant with the highest overall replication rate, but rather the best compromise among the rates of replication, replicase binding and double-strand formation.

The main reason for the high incidence of multi-error mutants must be that structural elements within the RNA are crucial for maintaining replication efficiency (Zamora \textit{et al.}, 1995) and disturbance of such a structural element can be compensated by other mutations to restore replication efficiency.

Darwinian evolution of replicating RNA species offers more than an opportunity to do qualitative evolution experiments \textit{in vitro}. It is also possible to predict evolutionary outcomes by deriving quantitative selection values from the physicochemical parameters of the competing RNA species, as outlined in Box 4.3. These parameters can readily be measured for individual RNA species. In addition, interactions among mutants such as formation of heteroduplex strands between single strands of different mutants must be taken into account. Since it has
### FIGURE 4.7 Mutant spectrum of MNV-11

**A.** Mutants within a population are indicated by number, different MNV-11 populations by letters: m, linear growth for 5 h; n, p, growth in the linear phase for 2 h followed by separation of plus (p) and minus (m) strands; 11, growth in the exponential growth phase (100 replication rounds); e, exponential growth in the presence of 1 µM ethidium bromide; x, growth in the linear growth phase for 8 h in the presence of 50 mM (NH₄)₂SO₄. HD is the Hamming distance, i.e., the number of base exchanges.

Reproduced from Rohde et al., 1995, in *J. Mol. Biol.*, 249, 754-762, with permission.

**B.** A more detailed analysis of a mutant spectrum of a quasispecies grown in the exponential growth phase. The last number in each column is the number of clones found in the population. Reproduced from Rohde et al., 1995, in *J. Mol. Biol.*, 249, 754-762, with permission.

---

**TABLE A**

| HD | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|----|----|----|----|----|----|----|----|----|
|    | GGGUCAUCGCUUUCGGCUUUAAGGCCCUUUUUCCUCGCUAGCUAGCUACGAGGGCACCCCCGCCGUGCAGGAGGGGUGCCA | HD |
| m2 |     |     |     |     |     |     |     |     |
| m3 |     |     |     |     |     |     |     |     |
| m6 |     |     |     |     |     |     |     |     |
| m13 |     |     |     |     |     |     |     |     |
| m14 |     |     |     |     |     |     |     |     |
| m25 |     |     |     |     |     |     |     |     |
| p12 |     |     |     |     |     |     |     |     |
| p20 |     |     |     |     |     |     |     |     |
| p21 |     |     |     |     |     |     |     |     |
| p22 |     |     |     |     |     |     |     |     |
| n8  |     |     |     |     |     |     |     |     |
| n14 |     |     |     |     |     |     |     |     |
| n19 |     |     |     |     |     |     |     |     |
| n34 |     |     |     |     |     |     |     |     |
| 11  |     |     |     |     |     |     |     |     |
| 14  |     |     |     |     |     |     |     |     |
| 19  |     |     |     |     |     |     |     |     |
| 111 |     |     |     |     |     |     |     |     |
| e80 |     |     |     |     |     |     |     |     |
| e84 |     |     |     |     |     |     |     |     |
| e85 |     |     |     |     |     |     |     |     |
| x3  |     |     |     |     |     |     |     |     |
| x4  |     |     |     |     |     |     |     |     |
| x6  |     |     |     |     |     |     |     |     |
| x9  |     |     |     |     |     |     |     |     |
| x14 |     |     |     |     |     |     |     |     |
| x16 |     |     |     |     |     |     |     |     |
| x23 |     |     |     |     |     |     |     |     |
| x46 |     |     |     |     |     |     |     |     |

*: duplication of GCGAGGU

---

**TABLE B**

| n  | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|----|----|----|----|----|----|----|----|----|
|    | GGGUCAUCGCUUUCGGCUUUAAGGCCCUUUUUCCUCGCUAGCUAGCUACGAGGGCACCCCCGCCGUGCAGGAGGGGUGCCA | HD |
| 124 |     |     |     |     |     |     |     |     |
| 128 |     |     |     |     |     |     |     |     |
| 130 |     |     |     |     |     |     |     |     |
| 150 |     |     |     |     |     |     |     |     |
| 117 |     |     |     |     |     |     |     |     |
| 19  |     |     |     |     |     |     |     |     |
| 111 |     |     |     |     |     |     |     |     |
| 113 |     |     |     |     |     |     |     |     |
| 153 |     |     |     |     |     |     |     |     |
| 168 |     |     |     |     |     |     |     |     |
| 1103 |     |     |     |     |     |     |     |     |
| 1112|     |     |     |     |     |     |     |     |
| wt |     |     |     |     |     |     |     |     |

*GGGUCAUCGCUUUCGGCUUUAAGGCCCUUUUUCCUCGCUAGCUAGCUACGAGGGCACCCCCGCCGUGCAGGAGGGGUGCCA*
been found that the rate constants for homoduplex and heteroduplex formation are essentially the same (Biebricher, unpublished measurements), these interactions can be quantified. Only when double-strand formation with other well-populated mutants is affected does a selective advantage result. Nevertheless, even in this system with its minimal number of biochemical reactions, calculating selection values is a challenging exercise. The reason for this is that the experiments can be modeled using constant selective values only for the conditions of infinite dilution of the competing populations and unlimited resources. Under normal laboratory conditions, however, even in the constant environment of the test tube, the rates of production by mutation (“mutational gain”) and the selection values change continuously as the population changes in composition and concentration (Eigen and Biebricher, 1987; Biebricher et al., 1991). Once again, computer simulations by numerical integration of the rate equations are of great help for getting insight, even though it is of course not possible, or reasonable, to try to account in computer simulations for all of the mutation possibilities that exist in the experiments.

Of particular interest is evolution in the pure exponential growth phase, because the selection values are then indeed constant and equal to the overall replication rate coefficient, which can be readily measured. The mutant distribution of such a quasispecies is shown in Figure 4.7B (Rohde et al., 1995). Among the 35 clones that were sequenced, the wild-type sequence was not found, because its replication rate is not the maximal one. There are fewer constraints on species evolving in the exponential phase, simply because competition and loss are excluded. A consequence of this is that the master sequence in the exponential growth phase is degenerate, the typical result being that several different mutants are nearly equally populated. These were not found in the mutant distributions in the linear growth phase, because their rate of binding replicase was reduced.

Adaptation to minor changes of growth conditions was found to be quite rapid. This is because the route of adaptation is different from what one might naively assume: when growth conditions change, there is no delay until appropriate mutations occur. Selection of the best adapted mutant already in the quasispecies is much faster than generation of new mutants. Its frequency rapidly increases and with it the (absolute) rate of producing mutants from it. The “center of gravity” within the existing quasispecies floats quickly through sequence space.
to a new position. Floating continues until a new evolutionary stable mutant spectrum emerges. What is thus observed is what has been described in organismic evolution as “punctuated equilibrium” (Gould and Eldredge, 1977).

The chance that a specific mutant is present depends strongly on the population size. When the population is small, more steps are required to reach a new equilibrium and adaptation takes longer. Furthermore, the route must then traverse a long staircase, on which each intermediate must have a selective advantage per se or it cannot be a part of the climb. We saw in analyzing the quasi-species, however, that this is seldom the case. Multi-error mutants were advantageous because the adverse effects of one mutation were compensated by subsequent ones. In a large population this is no problem, because some downward steps on the fitness landscape can be tolerated. The likelihood of generating a multi-error mutant depends on the number of steps necessary, the number of possible routes to reach it and the depth of the valleys that have to be crossed. Very deep canyons (i.e. where one of the intermediates represents a lethal mutation) must be crossed with a single jump, i.e. the two-error mutant be formed in one replication round.

Several adaptation experiments of short-chained RNA species have been reported. The first quantitative one was on replication of the species MDV-1 in the presence of a low concentration of ethidium bromide (Kramer et al., 1974), which resulted in selection of a three-error mutant. Adaptation was achieved slowly, because each transfer began with a population of only 10⁶ RNA strands. The first mutant was already present in the quasispecies population and the next mutations occurred in the seventh and 12th transfers, respectively. A disadvantage of the serial transfer technique is that small aliquots are used for inoculation of succeeding transfers. The probability of finding a newly formed mutant in an aliquot may be quite small, depending on its size and the time when the mutant emerged. Furthermore, each step in these experiments involved amplification in both the exponential and the linear growth phase. It was therefore not possible to calculate selection values.

Eigen and collaborators (Strunk and Ederhof, 1997) developed a machine that avoids these disadvantages. It always remains in the exponential growth phase, because the RNA concentration is measured in real time, triggering a serial transfer before the enzyme is saturated. The 1:10 aliquot used for the next transfer insures that mutant populations do not drop to low values. Using this machine a variant MNV-11 resistant to RNase A was selected after a rather large sequence change, including a deletion (Strunk and Ederhof, 1997). The evolution route taken in this process has not yet been reported.

Site-directed mutagenesis experiments with leviviridae genomes have shown that almost any mutation of the genome affects the fitness of the virus. Studies of the revertants and pseudo-revertants revealed an intricate influence of the RNA structure on replication, translation and regulation of the virus (Arora et al., 1996; Klovins et al., 1997; Poot et al., 1997). These experiments brought many insights into the subtle control of the biochemical processes involving RNA and illustrated that the fitness of a viral type is a highly complex function that makes quantitative predictions extremely difficult.

RECOMBINATION AMONG RNA MOLECULES

In organisms with DNA genomes, the high replication fidelity makes large mutational jumps impossible as evolutionary routes. An alternative route is taken: DNA recombination. DNA from other organisms is occasionally inserted and sections of the native genome are occasionally deleted, duplicated, inverted or transposed to remote positions. Normal cells contain many enzymes involved in catalyzing DNA recombination, underscoring the importance of this process.

RNA recombination is far less frequent. In initial experiments using double infection of host with leviviruses containing defects in different cistrons no defect-free recombinants could be isolated. Recent experiments have shown that recombination does occur, but only at a very low rate (Palasingam and Shaklee, 1992). RNA
recombination has been observed with many different viruses (King et al., 1982; Lai et al., 1985; Lai, 1992), but molecular devices that would catalyze it have not been identified. It seems that RNA recombination results from errors in the replication process itself. Several models have been proposed, the simplest and most plausible being "copy choice", i.e. a jump from one template to another (or to the same template, but on a different position) during replication (Lai, 1992).

An RNA species replicated by Qβ replicase has been isolated that is obviously a recombinant between part of the replicase gene of Qβ and host cell tRNA (Munishkin et al., 1988). RNA recombination in vitro is a very rare event, but has also been reported (Biebricher and Luce, 1992). Even a very rare event, however, can quickly become evident in an evolution experiment if an advantageous mutant is created. Thus MNV-11 grown to equilibrium builds up a stable mutant spectrum (see above), but under conditions of high ionic strength and growth in the late linear growth phase a new RNA species with a higher chain length (135) than MNV-11 (86) eventually emerges and is rapidly selected. Repetition of the experiment under identical conditions showed that the eventual result is reproducible, indicating an instructed process, while the time lapse to emergence of the new species is not (Biebricher et al., 1982).

RNA recombination events are more frequent at higher ionic strength. In the numerous cases we observed (Biebricher and Luce, 1992; Zamora et al., 1995; Biebricher and Luce, unpublished work), a short repetitive sequence was usually found, indicating a copy choice mechanism. Chetverin et al. (1997) described examples where this is not the case. Since only the sequence changes that are genetically fixed can be observed, a clear decision between different models is not possible.

CREATING BIOLOGICAL INFORMATION FROM SCRATCH

So far I have described experiments that showed darwinian adaptation to the environment, i.e. optimization of a pre-existing biological function. Evolution, however, is able not only to adapt but also to create. Is it possible to generate a self-replicating RNA without offering a template? In the last years, many RNA species with novel functions have indeed been selected starting from completely random RNA sequences (Tuerk and Gold, 1990). In other words, new biological function has been formed without any ancestry at all. In these experiments human ingenuity (to set up the experiments in the first place), random chance and darwinian evolution are the driving forces that create information from nowhere. However, even a century after Darwin, doubts continue to be expressed that such information can be formed without human interference. The main conceptual difficulty that such doubters experience derives from the vanishingly low probability of creating a predefined sequence by chance. To find a specific sequence of chain length 50 would require a population of $4^{50} = 10^{30}$ strands. Fortunately, there is not just a single winner in the sequence lottery: the large number of total blanks partially is compensated by the large number of minor wins. Sequences with low fitness values, once any are created by chance, are optimized by adaptive evolution on a much quicker time scale.

Two basic strategies were found to create replicable RNA without any template being present in the starting mixture. (1) In the first it was intentionally extracted from a huge library of randomly assembled sequences (Biebricher and Orgel, 1973; Brown and Gold, 1995). (2) The second was an unexpected finding: incubation of an RNA replicase at high concentration in the presence of high nucleotide triphosphate concentrations produced replicable RNA after long incubation times despite the absence of detectable RNA in the starting material (Sumper and Luce, 1975). Different RNA species are selected in each experiment of this kind (Biebricher et al., 1981; Biebricher, 1987). Evidence has been presented to show that in absence of template the replicase condenses nucleotides, at a rate five orders of magnitude less than that of template-instructed synthesis, to produce a random mixture of sequences (Biebricher et al., 1986). Once any accepted template is produced, no matter how inefficient it
may be, it is amplified and optimized. Indeed, replicability is a particularly sensitive function to select for, because the overwhelming majority of unaccepted RNA is ignored by the replicase.

Impurity RNA would have some genetic origin; however, the emerging species show no base homology to the genomes of the virus or of the host; moreover, they cannot be detected in infected or non-infected cells (Avota et al., 1998). In vitro, template-free synthesis was even found to be suppressed by addition of non-replicable RNA or DNA. Aggregation of enzyme molecules increased the efficiency of template-free synthesis. Modification of the non-replicable RNA by instructed terminal elongation has been observed (Biebricher and Luce, 1992). In vivo, only weakly replicable RNA species were derived from host RNA, in particular from 16S ribosomal RNA (Avota et al., 1998).

Sequence analysis and quantitative characterization of the properties of the early products of template-free synthesis were quite instructive (Biebricher and Luce, 1993). As mentioned earlier, the first feature noticed was that these short-chain RNA strands differ in chain length and sequence in each experiment. The low probability of assembling long replicable RNA strands favors small early products. Experimentally, strands with 25–40 nucleotides dominated. Their replication rates were low compared to those of optimized RNA. During subsequent serial transfers, early RNA products underwent rapid evolutionary optimization. During the optimization the molecular weight increased, in nearly all cases by recombination-like events such as duplications or insertion of sections of the complementary sequence. The optimization rate depended on experimental conditions. At high ionic strength optimization was fast. Otherwise it was so slow that a short inefficient template could be amplified at high amplification factors (> 10^9) during many serial transfers without changes of the average sequence.

The large number of short-chain replicable RNA species found offered a possibility to investigate the minimum sequence requirement for replication. Sequence comparison of the replicable species, however, did not reveal anything like a consensus sequence at all. Except for the invariant ends – pppGG[G] at the 5' termini, CCA at the 3' termini (a terminal A is attached without template instruction) – no homologies could be found. However, when the secondary structures were calculated, it appeared that the structures of all replicable RNA had a stem at the 5' termini, while the 3' termini were unpaired. The alternative folding, with 5' and 3' paired with each other, was energetically disfavored (Biebricher and Luce, 1993). The constraints for the more stable structure are more severe than it might seem at first. If base-pairing only involved the canonical base pairs, then a stem at the 5' strand of one strand would correspond to a 3' stem for the complementary sequence. Only non-canonical base pairs and outlooped bases at strategic positions makes the conserved replicable structure possible. Site-directed mutation replacing these positions with canonical base pairs destroyed the template activity of the RNA entirely (Zamora et al., 1995).

What is the reason for this structure? It is not known yet, but there are arguments in favor of stems at the 5' termini (Biebricher, 1994). As a replica is formed, the structure is transiently double-stranded. With progressing elongation the replica leaves the template. Rapid stem formation reduces the danger that replica and template reform a double strand. There are additional features common to many, but not all, replicable RNA species. A pyrimidine cluster in the interior of the sequence seems to be favorable for enzyme-RNA binding (Brown and Gold, 1993). However, binding strength to replicase is only poorly correlated with template activity. Some RNA sequences, notably 16S rRNA, bind quite well to Qβ replicase but have no template activity, while binding of some early products of template-free synthesis is only weak even though their replication rates are substantial.

The structural features described here appear to be necessary for RNA replication. To test whether they are sufficient for replication, the author's group designed and synthesized RNA strands with sequences predicted to give these structural features (Zamora et al., 1995). Their template activities, however, were found to be barely measurable. Upon prolonged incubation
with replicase, replicable RNA did grow out. The selected RNA species differed from each other, but were all clearly mutants of their respective initial templates. In some cases two or three base exchanges sufficed to make the species replicable, while in other cases recombination events were also involved. In all cases the above-described structural features were not only conserved but even enhanced. We conclude that there are unidentified additional requirements for adequate template activity. Clearly it is unlikely to strike a fitness peak when designing templates on the basis of the structural features we have been able to identify. During amplification, drifting of the designed sequences to nearby fitness peaks is thus inevitable.

CONCLUSIONS

Many quantitative insights into the nature of evolution have been gained from studying the model system provided by Qβ replicase. Are the often surprising results obtained from these experiments only a misleading caprice of nature, with no relevance to evolution in general? There are many reasons to think that this is not the case. The principal one is that studies using other enzymes, including some that are not viral, lead to similar results. If, as is generally believed, the origin of viruses is formation of intracellular parasites that eventually develop an apparatus for horizontal gene transfer among different hosts, then viral genes must derive from cellular ones. RNA replication with Qβ replicase requires catalytic participation of RNA; if template RNA is able to instruct Qβ replicase to replicate it, why should it be unlikely that cellular RNA can find ways to instruct a cellular DNA-dependent RNA polymerase to carry out RNA replication? Indeed, it has been shown that RNA templates exist that are accepted by the DNA-dependent RNA polymerases of the bacteriophages T7 and T3 (Konarska and Sharp, 1989, 1990; Biebricher and Luce, 1996) and E. coli (Biebricher and Luce, 1993; Wettich, 1995). Since for these enzymes no physiological RNA templates exist, replicable RNA species were selected by the described methods from random nucleotide libraries (Biebricher and Orgel, 1973) or obtained by template-free synthesis (Biebricher and Luce, 1996; Wettich, 1995). The templates are specific for their cognate enzyme and not accepted by other RNA polymerases. (One exception that has been found is an RNA species replicated by T7 RNA polymerase as well as by T3 RNA polymerase.) The features like exponential and linear growth and strand separation during replication were nearly identical to what has been observed over the years with Qβ replicase.

Recently, newly developed RNA amplification methods with lower sequence specificity than Qβ replicase have been shown to be superior for artificial selection of functional RNA by evolutive biotechnology (Guatelli et al., 1990; Breaker and Joyce, 1994). For quantitative studies of natural selection under controlled conditions, however, amplification of RNA by Qβ replicase is still unsurpassed.

ACKNOWLEDGEMENTS

I am indebted to Dr W.C. Gardiner for critical reading of the manuscript. Support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

REFERENCES

Arora, R., Priano, C., Jacobson, A.B. and Mills, D.R. (1996) Cis-acting elements within an RNA coliphage genome – fold as you please, but fold you must. J. Mol. Biol., 258, 433–446.
August, J.T., Cooper, S., Shapiro, L. and Zinder, N.D. (1963) RNA phage-induced RNA polymerase. Cold Spring Harbor Symp. Quant. Biol., 28, 95–97.
Avota, E., Berzins, V., Grens, E., Vishnevsky, Y., Luce, R. and Biebricher, C.K. (1998) The natural 6S RNA found in Qβ-infected cells is derived from host and phage RNA. J. Mol. Biol., 276, 7–17.
(1976) The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene*, 1, 27–32.

Beekwilder, J., Nieuwenhuizen, R., Poot, R. and van Duin, J. (1996) Secondary structure model for the first 3 domains of Qβ RNA - control of A-protein synthesis. *J. Mol. Biol.*, 256, 8–19.

Biebricher, C.K. (1983) Darwinian selection of self-replicating RNA. *Evol. Biol.*, 16, 1–52.

Biebricher, C.K. (1987) Replication and evolution of short-chained RNA species replicated by Qβ replicase. *Cold Spring Harbor Symp. Quant. Biol.*, 52, 299–306.

Biebricher, C.K. (1994) The role of RNA structure in RNA replication. *Ber. Bunsenges.*, 98, 1122–1126.

Biebricher, C.K. and Eigen, M. (1987) Kinetics of RNA replication by Qβ replicase In: *RNA Genetics, vol. I: RNA-directed Virus Replication* (eds Domingo, E., Ahlquist, P. and Holland, J.J.), pp. 1–21. CRC Press, Boca Raton, FL.

Biebricher, C.K. and Luce, R. (1992) In vitro recombination and terminal elongation of RNA by Qβ replicase. *EMBO J.*, 11, 5129–5135.

Biebricher, C.K. and Luce, R. (1993) Sequence analysis of RNA species synthesized by Qβ replicase without template. *Biochemistry*, 32, 4848–4854.

Biebricher, C.K. and Luce, R. (1996) Template-free synthesis of RNA species replicating with T7 RNA polymerase. *EMBO J.*, 15, 3458–3465.

Biebricher, C.K. and Orgel, L.E. (1973) An RNA that multiplies indefinitely with DNA-dependent RNA polymerase: selection from a random copolymer. *Proc. Natl Acad. Sci. USA*, 70, 934–938.

Biebricher, C.K., Eigen, M. and Luce, R. (1981) Product analysis of RNA generated *de novo* by Qβ replicase. *J. Mol. Biol.*, 148, 369–390.

Biebricher, C.K., Diekmann, S. and Luce, R. (1982) Structural analysis of self-replicating RNA synthesized by Qβ replicase. *J. Mol. Biol.*, 154, 629–648.

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

Biebricher, C.K., Eigen, M. and Gardiner, W. C. (1984) Kinetics of RNA replication: plus–minus asymmetry and double-strand formation. *Biochemistry*, 23, 3186–3194.

Biebricher, C.K., Eigen, M. and Gardiner, W.C. (1985) Kinetics of RNA replication: competition and selection among self-replicating RNA species. *Biochemistry*, 24, 6550–6560.

Biebricher, C.K., Eigen, M. and Luce, R. (1986) Template-free RNA synthesis by Qβ replicase. *Nature*, 321, 89–91.

Billeter, M.A., Dahlberg, J. E., Goodman, H.M., Hindley, J. and Weissmann, C. (1969) Sequence of the first 175 nucleotides from the 5' terminus of Qβ RNA synthesized *in vitro*. *Nature*, 224, 1083–1086.

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

Brown, D. and Gold, L. (1995) Selection and characterization of RNAs replicated by Qβ replicase. *Biochemistry*, 34, 14775–14782.

Chetverin, A.B., Chetverina, H.V., Demidenko, A.A. and Ugarov, V.L. (1997) Non-homologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. *Cell*, 88, 503–513.

Dawkins, R. (1982) *The Extended Phenotype*. Freeman, San Francisco, CA.

Dobkin, C., Mills, D.R., Kramer, F.R. and Spiegelman, S. (1979) RNA replication: required intermediates and the dissociation of template, product, and Qβ replicase. *Biochemistry*, 18, 2038–2044.
Dobzhansky, T., Ayala, F.J., Stebbins, G.L. and Valentine, J.W. (1977) *Evolution*. Freeman, San Francisco, CA.

Domingo, E., Sabo, D., Taniguchi, T. and Weissmann, C. (1978) Nucleotide sequence heterogeneity of an RNA phage population. *Cell, 13*, 735–744.

Drake, J.W. (1993) Rates of spontaneous mutation among RNA viruses. *Proc. Natl Acad. Sci. USA, 90*, 4171–4175.

Eigen, M. and Biebricher, C.K. (1987) Sequence space and quasispecies distribution. In: *RNA Genetics, vol. III: Variability of RNA Genomes* (eds Domingo, E., Ahlquist, P. and Holland, J.J.), pp. 211–245. CRC Press, Boca Raton, FL.

Eigen, M., Biebricher, C.K., Gebinoga, M. and Gardiner, W.C. (1991) The hypercycle – a principle of natural self-organization. Part A: Emergence of the hypercycle. *Naturwissenschaften, 64*, 541–565.

Eigen, M., Biebricher, C.K., Eoyang, L. and August, J.T. (1968) Factor fraction required for the synthesis of bacteriophage Qβ RNA. *Nature, 219*, 588–590.

Gould, S.J. and Eldredge, N. (1977) Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology, 3*, 115–151.

Guatelli, J.C., Whitfield, K.M., Kwoh, D.Y., Barringer, K.E., Richman, D.D. and Gingeras, T.R. (1990) Isothermal in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. *Proc. Natl Acad. Sci. USA, 87*, 1874–1878.

Haruna, I., Nozu, K., Ohtaka, Y. and Spiegelman, S. (1963) An RNA “replicase” induced by and selective for a viral RNA: isolation and properties. *Proc. Natl Acad. Sci. USA, 50*, 905–911.

Hori, K.L., Eoyang, L., Banerjee, A.K. and August, J.T. (1967) Template activity of synthetic ribopolymers in the Qβ RNA polymerase reaction. *Proc. Natl Acad. Sci. USA, 57*, 1790–1797.

Kamen, R. (1970) Characterization of the subunits of Qβ replicase. *Nature, 228*, 527–533.

Karnik, S. and Billeter, M. (1983) The lysis function of RNA bacteriophage Qβ is mediated by the maturation (A2) protein. *EMBO J., 2*, 1521–1526.

King, A.M.Q., McCahon, D., Slade, W.R. and Newman, J.W.I. (1982) Recombination in RNA. *Cell, 29*, 921–928.

Klovins, J., Tsareva, N.A., de Smith, M.H., Berzins, V. and van Duin, J. (1997) Rapid evolution of translational control mechanisms in RNA genomes. *J. Mol. Biol., 265*, 372–384.

Konarska, M.M. and Sharp, P.A. (1989) Replication of RNA by the DNA-dependent RNA polymerase of phage T7. *Cell 57*, 423–431.
(1973) Complete nucleotide sequence of a replicating RNA molecule. *Science*, **180**, 916–927.

Mitsunari, Y. and Hori, K. (1973) Qβ replicase-associated, poly(C)-dependent poly(G) polymerase. *J. Biochem.*, **74**, 263–271.

Munishkin, A.V., Voronin, L.A. and Chetverin, A.B. (1988) An *in vivo* recombinant RNA capable of autocatalytic synthesis by Qβ replicase. *Nature*, **333**, 473–475.

Palasingam, K. and Shaklee, P.N. (1992) Reversion of Qβ RNA phage mutants by homologous RNA recombination. *J. Virol.*, **66**, 2435–2442.

Poot, R.A., Tsareva, N.V., Boni, I.V. and van Duin, J. (1997) RNA folding kinetics regulates translation of phage MS2 maturation gene. *Proc. Natl Acad. Sci. USA*, **94**, 10110–10115.

Rohde, N., Daum, H. and Biebricher, C.K. (1995) The mutant distribution of an RNA species replicated by Qβ replicase. *J. Mol. Biol.*, **249**, 754–762.

Saffhill, R., Schneider-Bernloehr, H., Orgel, L.E. and Spiegelman, S. (1970) *In vitro* selection of bacteriophage Qβ RNA variants resistant to ethidium bromide. *J. Mol. Biol.*, **51**, 531–539.

Spiegelman, S., Haruna, I., Holland, I.B., Beaudreau, G. and Mills, D.R. (1965) The synthesis of a self-propagating and infectious nucleic acid with a purified enzyme. *Proc. Natl Acad. Sci. USA*, **54**, 919–927.

Strunk, G. and Ederhof, T. (1997) Machines for automated evolution experiments *in vitro* based on the serial-transfer concept. *Biophys. Chem.*, **66**, 193–202.

Sumper, M. and Luce, R. (1975) Evidence for *de novo* production of self-replicating and environmentally adapted RNA structures by bacteriophage Qβ replicase. *Proc. Natl Acad. Sci. USA*, **72**, 162–166.

Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment. *Science*, **249**, 505–510.

Weissmann, C. (1974) The making of a phage. *FEBS Lett. (Suppl.)*, **40**, S10–S18.

Weissmann, C., Simon, L., Borst, P. and Ochoa, S. (1963) Induction of RNA synthetase in *E. coli* after infection by the RNA phage MS2. *Cold Spring Harbor Symp. Quant. Biol.*, **28**, 99–104.

Wettich, A. (1995) RNA replication mechanisms of DNA-dependent RNA polymerases. Master’s thesis, University of Göttingen.

Yonesaki, T., Furuse, K., Haruna, I. and Watanabe, I. (1982) Relationships among four groups of RNA coliphages based on the template specificity of GA replicase. *Virology*, **116**, 379–381.

Zamora, H., Luce, R. and Biebricher, C.K. (1995) Design of artificial short-chained RNA species that are replicated by Qβ replicase. *Biochemistry*, **34**, 1261–1266.