Biochemistry

 Volume 30, Number 46 November 19, 1991

Perspectives in Biochemistry

The Hypercycle. Coupling of RNA and Protein Biosynthesis in the Infection Cycle of an RNA Bacteriophage†

Manfred Eigen,* Christof K. Biebricher, and Michael Gebinoga
Max-Planck-Institut für Biophysikalische Chemie, D-3400 Göttingen, Federal Republic of Germany
William C. Gardiner
Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas 78712
Received June 19, 1991; Revised Manuscript Received September 11, 1991

Life is based on genetic information originating and changing through natural selection. Natural selection requires reproduction, and reproduction—as we know it today—means replication of nucleic acids. The present sophisticated replication machinery, however, was assembled and refined over a protracted evolution period of life on Earth. At the dawn of life, molecular genotypes and molecular phenotypes may have been the very same molecules, and selection based on the feedback loop of genotype replication marked the onset of Darwinian evolution. The target of selection was the fittest genotype quasispecies, i.e., the most efficiently reproducing mutant distribution centered around its wild type. Once the decoding and reproduction machineries were themselves coded by the genotype, however, their coevolution required a second feedback loop. This 2-fold feedback principle is what we termed a hypercycle (Eigen, 1971; Eigen & Schuster, 1977, 1978a,b). There are countless chemical–kinetic possibilities, with all sorts of positive and negative loops, to realize it. Can we find living examples showing how Nature actually achieves hypercyclic coupling?

Clear-cut cases are to be found in the infection cycles of RNA plus-strand viruses, which can be met in all kingdoms of life, in bacteria, plants, and animals. One of the viral genes instructs the synthesis of a replication factor that specifically recognizes viral RNA while ignoring a vast excess of host RNA. The virus-encoded replication factor provides the phenotypic feedback link of the viral hypercycle: Upon entering the host cell, the viral RNA, serving as messenger to the host translation apparatus, produces the replication factor that catalyzes amplification of the viral (and complementary minus strand) RNA in a cyclic reaction. The interplay of translation and replication cycles realizes a hypercycle.

The hypercycle concept was first introduced as a principle of self-organization relevant to early evolution, i.e., to the first expressions of genetic information in phenotypes, whether in the form of functional RNAs like present-day ribozymes or in the form of protein-like molecules translated by some means from the genetic messages. Its role in early evolution was analogous to its role in viral evolution today: Hypercyclic coupling provides functional evaluation and evolutionary improvement of the phenotype by specifically favoring the replication of its own genotype. The host's role in virus infections had to be played in precellular evolution by the (eventually compartmentalized) primordial reaction medium.

Viruses are presumably all of postbiotic origin, i.e., descendants of host-cell components. Studying their infection cycles, however, offers the best route we have to understanding how the hypercycle principle is actually realized in nature, which may help us understand how prebiotic hypercycles may have been realized. To attain such understanding it is necessary not only to describe viral hypercycles in vivo but also to reduce the chemical steps involved to an in vitro level where the tools of biochemistry, and in particular biochemical kinetics, can be applied for analysis. We were able to do this for a particular example, the infection cycle of the coliphage $Q\beta$, by means of chemical probing of in vivo infection profiles, in vitro characterization of the replication process, and analytical and computer simulation explorations of the reaction kinetics.

A reaction mechanism based upon in vitro results was found to account for the dynamics of the three experimentally observed phases of the infection cycle: (1) an initial phase where the invading RNA is used as mRNA for the production of viral proteins with a (practically) constant concentration of RNA

† This research was carried out under the auspices of the Max-Planck-Gesellschaft. Additional support from the Alexander von Humboldt, Fritz Thyssen, and Robert A. Welch Foundations is also acknowledged.

* To whom correspondence should be addressed.

0006-2960/91/0430-11005$02.50/0

and a linear increase of the viral protein concentration, (2) a short phase where RNA and viral protein grow hyperbolically, and (3) a late phase where RNA and viral protein grow linearly and the production rates of both are regulated at the maximum permissible levels.

Measurement of RNA and Protein Synthesis Rates

Rationale of the Experimental Approach. In the first decade after the discovery of the RNA coliphages, the replication intermediates (Ellis & Paranchych, 1963; Ammann et al., 1964; Fenwick et al., 1964; Franklin, 1966; Weissmann et al., 1967, 1968, 1973) and the products of phage-directed protein biosynthesis (Lodish & Zinder, 1966; Lodish, 1968, 1971, 1975; Nathans et al., 1969; Kozak & Nathans, 1972) were characterized and their intracellular synthesis was studied extensively in vivo. Concentration profiles of the viral RNA and the gene products in infected cells were measured (Lodish & Zinder, 1966; Gussin et al., 1966; Godson & Sinsheimer, 1967; Viñuela et al., 1967; Godson, 1968; Horiuchi & Matsuhashi, 1970; Jockusch et al., 1970; Cramer & Sinsheimer, 1971; Beremand & Blumenthal, 1979). However, it is virtually impossible to measure accurate synthesis rates in vivo, because the precursor concentrations can be neither measured nor labeled by appropriate markers. Addition of a defined precursor pool has been accomplished for in vitro studies of the replication (Weissmann et al., 1964; Haruna & Spiegelman, 1965a,b; August et al., 1965), phage assembly (Hung & Overby, 1969; Hung et al., 1969; Hofstetter et al., 1974) and protein synthesis processes (Engelhardt et al., 1967; Davies & Kaesberg, 1973; Happe & Jokusch, 1973). The efficiencies of the in vitro reactions under the conditions investigated were found to be much lower than the in vivo synthesis efficiencies except for the RNA replication process, which proceeded with efficiency comparable to the in vivo process. The cited in vitro studies disregarded the interference of the different processes with each other entirely.

An approach to measuring the synthesis rates using permeabilized bacteria (Moses & Richardson, 1970; Vosberg & Hoffmann-Berling, 1970) seemed appropriate. In such systems, lipids are extracted from the cytoplasmic membrane after selected incubation times by adding an organic solvent to the incubation medium. The semipermeabilized membrane allows rapid exchange of low molecular weight synthesis precursors with the surrounding medium, while the physical integrity of the bacterial cell and the locations and activities of the high molecular weight intracellular components are retained. The defined concentrations of the extraneously added low molecular weight synthesis precursors permit incorporation rates to be determined accurately. The infection process itself starts and continues under normal in vivo conditions; only at the analysis time is the process interrupted by extraction with the organic solvent, and so only the incorporation rate measurements themselves are measured quasi in vitro. The relative rate profiles obtained should thus reflect the profile of the unperturbed in vivo process, while the absolute values of the rates may well be perturbed by the experimental manipulations.

Parallel in vitro measurements with $Q\beta$ replicase established that the treatment with toluene under similar conditions does not affect replicase activity. Incorporation of nucleoside triphosphates into RNA as well as incorporation of amino acids into protein was observed to proceed linearly for about 10 min. However, incorporation rates then leveled off; furthermore, the infection cycles of permeabilized infected bacteria never finished normally with lysis of the bacterial host and liberation of a burst of phages. Obviously, some vital function of the cell in the synthesis process is seriously affected. The experimental details have been given elsewhere (Gebinoga, 1990).

RNA Synthesis Rate Measurement. In Leviviridae infections there is no direct interference with host synthesis of RNA and protein; the phage-directed processes simply outcompete the host ones (Ellis & Paranchych, 1963; Hudson & Paranchych, 1967; Lodish & Zinder, 1966). Viral RNA synthesis, however, is performed by a virus-coded replicase, while host RNA synthesis proceeds by transcription from cellular DNA. It was possible to suppress host RNA synthesis nearly completely with drugs without affecting viral RNA replicase activity (Fromageot & Zinder, 1968; Marino et al., 1968). The residual noise from host RNA synthesis was about 20 strands per cell.

The rate profile of nucleoside triphosphate incorporation is shown in Figure 1. Different phases can be distinguished: after a lag period of about 10 min a sudden explosive onset of RNA replication is observed. The rate of RNA synthesis reaches its maximum at about 15–20 min after infection and then gradually decreases. The yield of RNA strands calculated by integration of the measured production rates is about 10000 strands of $Q\beta$ RNA per host cell. In vivo, $(1–4) \times 10^4$ mature phages are produced per infected cell (Weissmann, 1974), representing about half of the final RNA concentration (Godson, 1968). Apparently the rate of RNA synthesis as measured in permeabilized cells is about half of the true in vivo rate. This factor is readily accounted for by the altered environment (absence of membranes and oxidative phosphorylation) and the drugs used for suppression of transcription.

Viral Protein Synthesis Rate Measurement. Viral and host protein biosynthesis share the same synthesis apparatus; therefore, specific suppression of host protein synthesis is not possible. Transcription inhibitors like rifampicin, however, should have an indirect inhibitory action on host translation by mRNA depletion. Differentiation between host and viral synthesis requires product analysis, e.g., by gel electrophoresis (Viñuela et al., 1967; Beremand & Blumenthal, 1979). The rate of amino acid incorporation per ribosome in permeable cells was found to be lower by about an order of magnitude than rates measured in vivo even though well above the usual in vitro rates. The reason for the reduced efficiency of protein synthesis is not known; possibly essential low molecular weight cofactors may have been lost and not supplied by the incorporation medium, or the treatment with organic solvent affects the activities of one or several of the many high molecular weight components of the translation apparatus. The source of the suppression is unimportant for present purposes, for these rates were only used for sampling the in vivo growth rate profiles.

The protein with the greatest influence on viral growth is the product of the replicase gene, which contributes one subunit to the viral RNA replicase, the other three being supplied by the host (Kamen, 1970; Kondo et al., 1970). It can be assumed that the concentrations of this subunit and of active replicase are nearly identical, because the host-coded subunits are found in high concentrations in the cell (Gordon & Weissbach, 1970). The accuracy of measurement of the replicase production rate in permeable cells did not suffice to give reliable data, because the synthesized quantities were too difficult to discriminate from host proteins. On the other hand, the most abundant gene product by far is that of the coat protein gene, and measuring it provides the additional advantage that no cellular synthesis products comigrate with it in electrophoresis. The sensitivity of the incorporation measurement is nonetheless lower than that of RNA synthesis because of the limited specific activities of the protein precursors. The amino acid

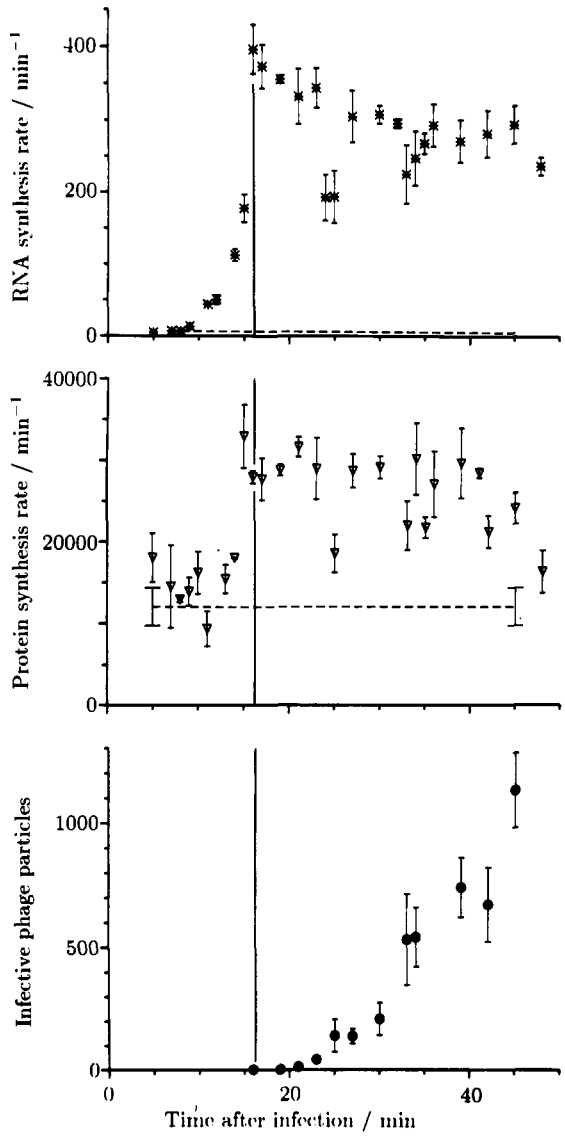

FIGURE 1: Experimental rate profiles of RNA and protein synthesis and concentration profile of intracellular infective phage particles. The rates were measured in Qβ-infected *E. coli* cells that were treated with toluene at the indicated times after infection. Procedure: *E. coli* Q13 cells were grown in rich medium in shaker flasks at 37 °C. At a cell density of 6×10^8/mL the culture was divided into two parts. One was infected with Qβ phage at a multiplicity of infection of 5 while the other was diluted with a corresponding volume of phage buffer (10 mM Tris-HCl buffer, pH 8.1, 10 mM NaCl, 1 mM $MgCl_2$, and 10 mg/L gelatin). The infected culture was further divided into aliquots. Eight minutes before the chosen harvesting times, rifampicin was added to give a concentration of 120 μM. At the harvesting times, cells were chilled with crushed frozen buffer and treated with toluene by the method of Moses and Richardson (1970). The cells were centrifuged, and the pellets resuspended in TCB buffer (60 mM HEPES-KOH buffer, pH 7.5, 200 mM potassium glutamate, 10 mM magnesium acetate, 10 mM NaCl, 0.5 mM dithiothreitol, 20% glycerol, and 120 μM rifampicin), incubated for 10 minutes at 0 °C with toluene (5% v/v) and washed three times with TCB buffer. The cells were resuspended at a concentration of 7×10^9 cells/mL in TCB buffer containing 35% glycerol and 2 mM dithiothreitol and shock-frozen in liquid nitrogen. Rate measurement: The frozen cells were diluted with an equal amount of prewarmed TCB buffer containing (as final concentrations) 2 mM ATP, 1 mM GTP, 1 mM UTP, 0.5 mM CTP, 2 mM phosphoenolpyruvate, 0.2 mM NAD, 20 μM leucin, 40 μM each of the amino acids, 25 μM each of dATP, dCTP, dGTP, and dTTP, 4 mM vanadyl ribosidyl complex, 0.5 TBq/mM [α-^{32}P]CTP, 7.5 TBq/mM [U-^{14}C]leucine, and 10 μg/mL phenol red. Samples were withdrawn after 1, 2, and 4 min of incubation at 37 °C and stopped by an excess of a solution of stop mix (5% trichloroacetic acid, 100 mg/L tRNA, 100 mg/L bovine serum albumin, and 200 mg/L leucine), filtered with nitrocellulose filters, washed, dried, and counted in a liquid scintillation spectrometer. Top: RNA synthesis rate, calculated as phage RNA strands equivalent synthesized per minute and cell. The RNA synthesis background measured in noninfected control cells is shown as a broken horizontal line. The values are averages of eight independent (i.e., different cell preparations) measurements and the error bars indicate standard deviations. The equivalence point (see Discussion of Results) is shown as a thin vertical line. In the late phase of infection several points had reproducibly substantially lower rates of RNA synthesis. The cause of these relatively large deviations, which also show unusually large scatter, is not known. Cooperative packing could cause sudden drops in RNA concentration, which, given the high saturation rate of RNA formation, could be balanced within minutes. Middle: Experimental rate profiles for coat protein synthesis, calculated as incorporated leucine residues per minute and cell. Predominantly coat protein is synthesized, which contains 12 leucine residues. The background measured in uninfected cells is shown as a broken horizontal line. The error bars indicate standard deviations. Bottom: Experimental profile of intracellular phage concentration, calculated as infectious phage particles per cell. On average, about 10% of the mature phage particles are infectious. The error bars indicate standard deviations. Cells were prepared and incubated for 0.5 min at 37 °C as described above. The cells were incubated in lysis buffer (10 mM HEPES-KOH buffer, pH 7.5, 1 mM EDTA, and 0.1 mg/mL lysozyme) for 15 min at 30 °C. DNase I (RNase-free grade) and $MgCl_2$ were added to give final concentrations of 100 units/mL and 10 mM, respectively, and incubation was continued for an additional 10 min. The suspension was centrifuged and the pellet reextracted with lysis buffer. The phage titers of the combined supernatants were determined by the top agar method of Adams (1959).

incorporation profile for coat protein is shown in Figure 1.

The similarity of the RNA synthesis and protein synthesis profiles is striking: the lag periods, the explosive onsets, and the late phases with approximately constant rates cover the same time ranges. The apparent similarity of the profiles is readily explained. During the lag period, RNA and protein synthesis levels are too low to detect. The sudden onset of detectable RNA synthesis rate reflects the rapid increase of RNA concentration produced by replication. Part of the RNA serves as replication template while the rest is bound to ribosomes for synthesis of (mostly coat) protein. The constant rates in the late infection period indicate saturation of the available replicase as well as of the ribosomes with free RNA strands. The concentrations of both ribosomes and replicase remain constant in the late infection phase. Unfortunately, absolute concentrations of replicase in infected cells have not been measured; estimates from the literature data (Viñuela et al., 1967; Horiuchi & Matsuhashi, 1970; Cramer & Sinsheimer, 1971; Beremand & Blumenthal, 1979) give values of approximately 1000 replicase molecules per cell.

Intracellular Phage Formation. The first intracellular phages are found after 20 min and accumulate further to a burst size of 600 infective particles, corresponding to about 10000 phage particles (Figure 1). The profiles found closely resemble profiles published before (Godson, 1968; Viñuela et al., 1967). Suppression of cellular RNA synthesis reduced the burst size by about half and abolished cell lysis entirely. Permeable cells provide no advantage in measuring the intracellular phage pool over conventional methods; RNA and viral proteins synthesized in permeabilized cells were not found to be incorporated in mature phages. Ten thousand phage particles corresponds to a coat protein pool of 2×10^6 copies per infected cell at the end of the infection cycle; most of the coat protein in the late infection phase is thus incorporated into mature phage particles. Coat protein (both free and assembled to capsids), on the other hand, continues to grow and reaches its peak concentration values at much later times.

Interpretation of the Profiles. The results presented here and the data in the literature show infection cycles with well-defined phases. In the first phase, the injected parental

phage RNA is found predominantly in the polyribosome fraction (Godson & Sinsheimer, 1967). Coat protein and replicase are synthesized, but the absolute amount of viral protein synthesis remains small due to the low copy number of RNA strands.

After about 10 min, a second and most remarkable phase sets in. Its duration is short (5–7 min); for this reason its important features are only resolved when a large number of measurements are made. In the previously reported concentration profiles (Godson, 1968; Viñuela et al., 1967), exponential growth of intracellular viral RNA and protein has been assumed but is not supported by the published profiles, because the semilogarithmic plot has only two or three measured points in the pertinent range. A semilogarithmic plot of the measured points of Figure 1, however, as well as a plot of doubling times against time after infection (not shown) clearly indicates that the growth rate increases more rapidly in this phase than would be expected for an exponential growth law. It will be shown in the following sections that this *hyperbolic* growth represents the first direct experimental evidence for hypercyclic coupling, as predicted before by one of the present authors (Eigen & Schuster, 1977).

The hyperbolic growth rapidly comes to its limits of synthesis: the ribosomes become saturated with viral messenger, and later the nucleoside triphosphate pools also become limiting. Slowing of protein synthesis results in rapid saturation of plus and minus strands with replicase, and at about 20 min after infection replicase production is fully shut off by translation repression. Growth retardation by changing reaction kinetics and eventually also by limited resources produces an "overshoot" at the point where the resources become rate-determining; the profile shown in Figure 1 clearly shows this feature. In the late phase of the infection cycle the production rates of viral RNA, coat protein, and mature phage are constant and eventually decline at the end of the infection cycle when the cell is stuffed with viral products.

Mechanistic Studies of the Infection Process

Modeling of the Infection Process. Chemical kinetics studies based upon numerical solutions of large sets of rate equations have contributed broadly to our understanding of complicated chemical processes (Warnatz & Jäger, 1988). In previous work (Biebricher et al., 1983, 1984, 1985, 1991) we have shown that a biochemical process like RNA replication can be modeled numerically to describe successfully the experimentally observed behavior. The in vitro kinetics of replication differs from in vivo kinetics by the absence of translation. Substrates and enzyme are added at defined concentrations and the reaction environment is well defined physically and chemically. Replication is triggered by addition of measured amounts of RNA strands that are accepted by the replicase as templates (Haruna & Spiegelman, 1965a).

The kinetics and mechanism of RNA-instructed RNA synthesis have been studied in detail (Mills et al., 1978; Dobkin et al., 1979; Biebricher et al., 1981a,b, 1983, 1984, 1985). The sequence of reactions comprises five phases (Figure 2, left), which may be replaced for analytical treatment by a three-step mechanism (Figure 2, right). Both analytical and numerical studies give quantitative descriptions of experimentally observed reaction profiles (Figure 3), as shown by the Fibonaccean mechanism in Appendix A. At low template/enzyme ratios, template RNA is amplified exponentially (Haruna & Spiegelman, 1965b). The exponential nature of the growth curve under conditions where enzyme is in excess of template is seen not only in the curve shapes but also—even more convincingly—by the parallel shift of growth curves caused

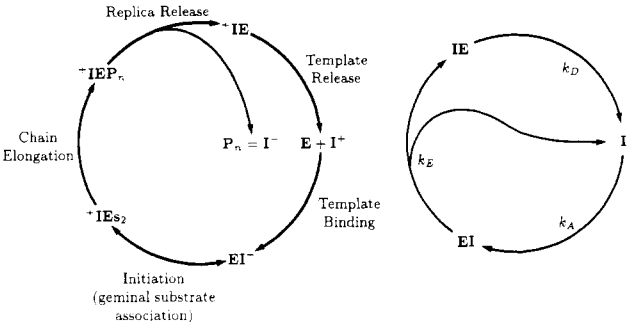

FIGURE 2: Simplified replication mechanism: left, five-step cycle; right, three-step cycle. I complexes to E with the rate $k_A[E]$ to form the active complex EI, which then synthesizes I with the rate k_E. The remaining inactive replication complex IE dissociates at the rate k_D into its components I and E. This mechanism is treated analytically in Appendix A.

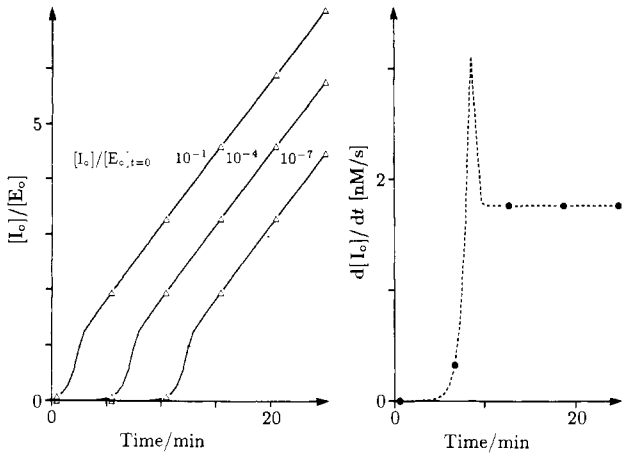

FIGURE 3: Computer-simulated incorporation profiles for in vitro RNA replication kinetics. Left: Nucleotide incorporation profiles as used for experimental measurements of overall replication rates in the exponential and the linear phases. Experiments are performed with various initial ratios of $[I_0]$ to $[E_0]$. Replication rates in the linear growth phase are taken from the slopes of the curves; replication rates in the exponential growth phase are by dividing the displacement of the curves on the time axis by the natural logarithm of the ratio of the starting template concentrations. Right: Simulated RNA synthesis rate profile for $[I_0]/[E_0]_{t=0} = 10^{-6}$. Exactly this behavior was found experimentally in in vitro studies (Biebricher et al., 1981b).

by serial dilutions of initial template (Figure 3; Sumper & Luce, 1975; Biebricher et al., 1981b; Biebricher, 1986). The transition from exponential to linear growth appears quite abruptly when the growing RNA concentration becomes equal to the (constant) replicase concentration (Biebricher et al., 1981b; Biebricher, 1986), like reaching the sharp end point of a titration curve (Winkler-Oswatitsch & Eigen, 1979), here of enzyme with templates (Figure 3 and Appendix A).

Life cycles of entire organisms, however, are so complex that a similar approach might appear at first glance scientifically uninteresting even if computationally possible. For the case at hand such a pessimistic attitude is unwarranted: the essential dynamic features of this infection cycle can be accounted for with a kinetic model having reasonable rate constants for all the steps involved.

An infection process has so many possible sources of interactions with host processes that neglecting most of them is imperative. For understanding the dynamics of the infection process, however, it is absolutely necessary to consider the interferences of the different viral processes with each other: The same viral RNA serves as message for protein synthesis, as template for RNA replication, and as infectious RNA to

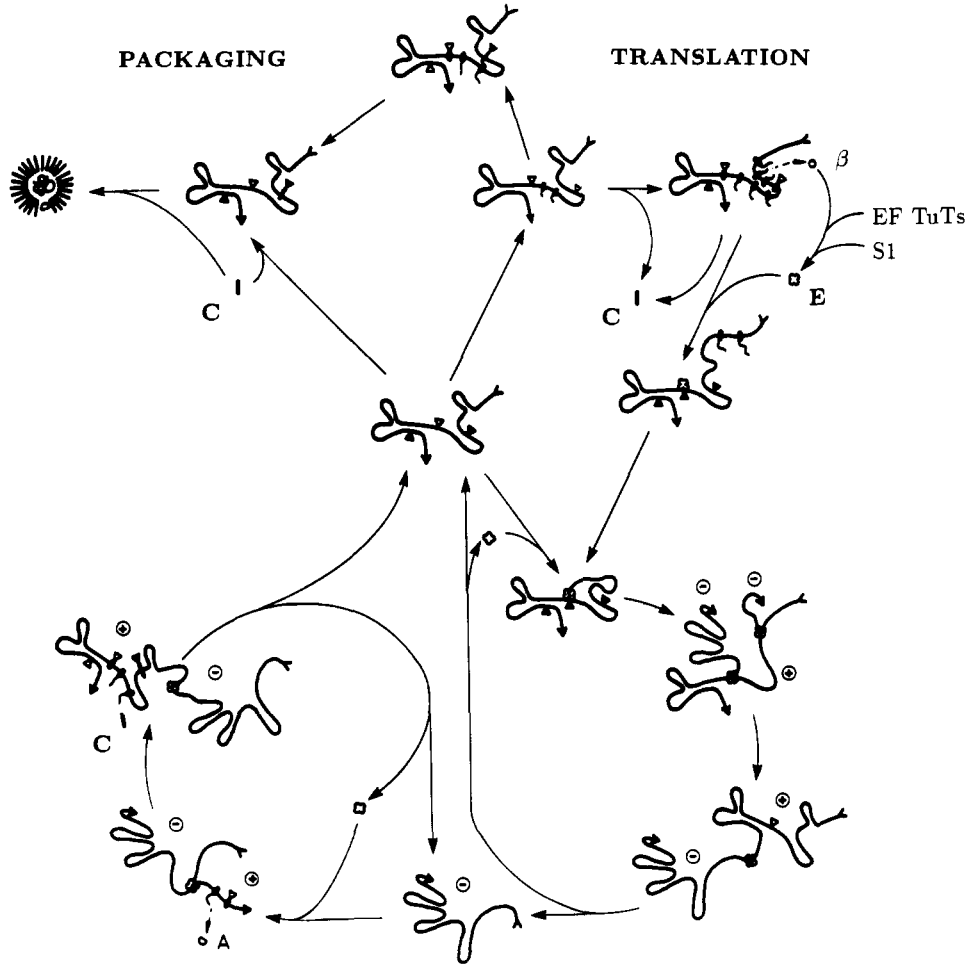

FIGURE 4: Cartoon of the infection cycle. The plus-strand RNA (center) can react with ribosomes to enter translation, with replicase to enter replication, or with coat protein to begin packaging. Translation: Initially, ribosomes can bind RNA only at the coat cistron Shine–Dalgarno site; the other two boxes are unavailable. Once coat protein synthesis has started, the Shine–Dalgarno box of the replicase cistron becomes available, resulting in a polysomal RNA producing coat protein and the viral-coded replicase subunit, which combines with the host proteins EF Tu·Ts and the ribosomal protein S1 to form active replicase. Replication: Uncomplexed RNA and RNA complexed with coat protein can bind to replicase directly to form an active replication complex able to start synthesis at the 3′-terminus of the template. RNA bound to ribosomes binds the replicase at a specific binding site interfering with the Shine–Dalgarno box of the coat protein. Synthesis at the 3′-terminus of the template has to wait until the ribosomes fall off the template after completing their job. RNA replication proceeds with a cross-catalytic mechanism similar to that described in Figure 2. Nascent plus strands synthesized on minus-strand templates can begin protein synthesis immediately. Packaging: Coat protein binds to RNA at a certain site overlapping with the Shine–Dalgarno sequence of the replicase cistron. The complex is unable to synthesize the β subunit of the replicase but can bind ribosomes to synthesize more coat protein or replicate. The RNA-coat protein can assemble with further coat protein (and with other capsid proteins) to form mature viruses. Open arrowhead: Shine–Dalgarno box available for ribosome binding. Solid arrowhead: Shine–Dalgarno box unavailable for ribosome binding.

be packed into mature virions. The different roles of the RNA are incompatible with each other and different processes can not take place simultaneously on the same RNA strand. Separate processors are responsible for each of the three different functions: (i) the ribosome and its ancillary translation apparatus, (ii) the replicase, aided by a host factor, and (iii) packaging into virions, which occurs spontaneously when sufficient amounts of RNA and protein become available. Switching from one function to another is effected by the processors themselves: an RNA molecule bound to a ribosome serves exclusively as message, while the message function of a replicase-bound RNA molecule is unavailable. The exclusive competition of the different processors provides a regulated series of operations at the RNA, defining indeed a "phage clock". The timing of the phage clock is essentially due to the fact that two of the processors are coded for by the viral RNA and are thus formed and increase in concentration during the infection cycle, while the translation processor is provided by the host and is present at constant concentration during the entire infection cycle. In an adequate model the timing of the phage clock must appear as the direct consequence of the cascade of chemical processes that it describes.

The numerous data collected over the past three decades were compiled and distilled into a simplified mechanism discussed in Appendix B and shown in Figures 4 and 5. This mechanism has been shown by computer simulations and by analytical treatment of the infection kinetics to describe and predict the measured concentration profiles of the viral components during the infection cycle.

Analytical Treatment of the Infection Cycle. The purpose of analytical treatment is to provide basic *understanding* of the experimental results, while numerical simulations aim at a quantitative *description.* Hence, for the sake of transparency, additional simplifications were introduced into the analytical treatment, the validity of which had to be, and was, then checked against numerical simulations.

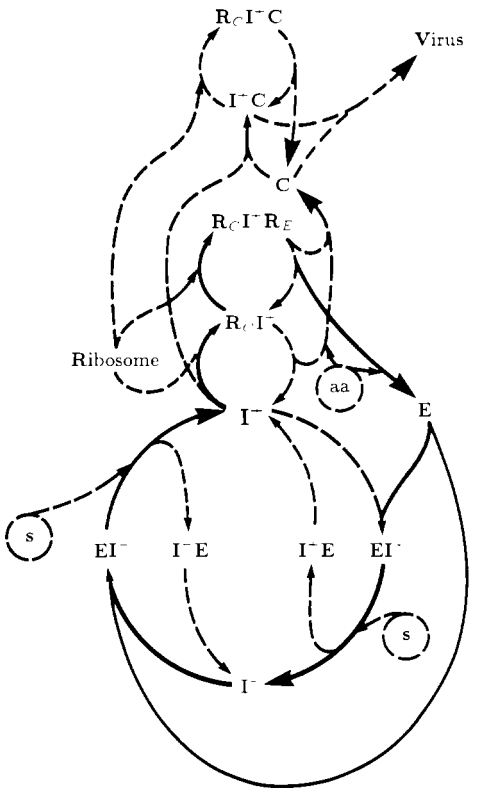

FIGURE 5: Mechanism of the infection process. The pathways belonging to the hypercyclic coupling are shown as solid curves; other reaction pathways are dashed. Large arrows denote synthesis of the pointed-to substances; small arrows denote association or dissociation reactions. Numerical simulation of this mechanism is described in Appendix D.

While in the simulations individual values of the rate and equilibrium constants of plus and minus strands were used throughout, we have treated the mechanism in Appendix C as if plus and minus strands were identical. Hence the different involvements of plus and minus strands with ribosomes and replicase are not taken into account. Furthermore, binding of several ribosomes is also disregarded and ribosome binding is assumed to lead only to replicase production. Binding of RNA (I) to ribosomes (R) and replicase (E) can then be treated as a straightforward competition process, assuming that all active templates are bound to either ribosome or replicase. Free template is accordingly neglected, which corresponds to reality except in the late phase of the infectious cycle. As was stressed before (Appendix A), binding at the template—here for both replicase and ribosome—is not equilibrated. The competition between replicase and ribosome is of a kinetic nature, governed by the rate constants of RNA association.

The procedure outlined in Appendix C starts with the rate equations (C1) for translation and replication and the conservation relations (C2) for all species involved in enzyme or ribosome binding to RNA. The kinetic competition between ribosome and replicase for binding to RNA is described by relation C3, which leads to the equation for the concentrations of the active complexes that determine the rates of translation and replication. It can be seen that eq C3, while physically based on a dynamic competition, has the mathematical form of a mass-action expression with a "competition constant" K_c.

For the initial phase of infection the root in eq C4 can be expanded and eqs C7 and C8 hold. Equations C7 and C8 describe—in agreement with the experimental profiles—hyperbolic growth. The initially delayed accumulation of template eventually speeds up to a burst. The essential difference between exponential and hyperbolic growth characteristics lies in the progressive reduction of doubling periods of the latter, which, if no inhibition were effective, would lead to a singularity. While in the in vitro RNA replication experiments the enzyme concentration remained constant, it grows during the infection cycle concomitantly with the RNA concentration.

However, RNA and replicase production rates are not simply proportional to one another. While the replicase concentration is increasing, the concentration of ribosomes remains constant. Therefore, the proportion of RNA engaged in replication to RNA engaged in protein biosynthesis is steadily increasing, and the pace of RNA growth acceleration thus exceeds that of protein biosynthesis. Eventually, an equivalence point is reached where the amount of RNA exceeds the combined concentration of ribosomes and replicase, and free RNA (or RNA bound to coat protein) appears. At this point a sharp transition from hyperbolic to linear growth behavior is observed and the equations derived in Appendix C do not apply to the late phase of infection. The equivalence point has many similarities to the point of transition from exponential to linear growth in the in vitro RNA replication (Appendix A). The persistance of hyperbolic growth up to the equivalence point, which is to be expected only for balanced competition of translation and replication (K_c values near 1), and the exactly simultaneous burst of both replicase and RNA concentrations are the characteristic features of the hypercycle mechanism.

Numerical Simulation of the Infection Cycle. In Appendix B some aspects of modeling the infection cycle (shown in Figure 5) are described and estimations or measurements of reaction rates discussed. The rate constant data used for numerical simulations are presented in Appendix D, and a set of typical simulation profiles is shown in Figure 6. Despite the simplifications that had to be introduced, all essential dynamic features found in the experimental studies were indeed seen in the computed profiles. The simulations also provide insights into what may be happening at the molecular level for parts of the infection cycle that are not accessible to experimental measurements.

Discussion of Results and Description of the Infection Cycle

The Early Phase of Infection. After invading the host cells, viral RNA binds to ribosomes and is involved almost exclusively in protein synthesis. Free coat and replicase begins to accumulate (Figure 6). While production of coat protein is more or less constant, production of replicase is soon braked by the onset of translation repression by coat protein. This first phase comprises the first 5–8 min of the infection cycle.

With increasing replicase concentration, RNA begins to replicate. At this time the hyperbolic growth phase begins with its mutual enhancement of replicase and RNA formation. Deviations from the analytical descriptions are found to be due to binding of coat protein to viral RNA and to the binding asymmetry to the coding and noncoding strands. Plus-strand RNA is bound to ribosomes, replicase, and coat protein, while minus-strand RNA is entirely saturated with replicase. However, during the hyperbolic phase, asymmetry of the synthesis rates of plus- and minus-strand RNA is offset by their strong synthesis coupling.

Complex formation between coat protein and plus-strand RNA is quite strong. Extensive in vitro measurements of the association rate and equilibrium constants have produced values for k_{IC} of about 10^9 M^{-1} s^{-1} (Uhlenbeck, 1986) and for K_{IC} in the range of 10^7 M^{-1} (Talbot et al., 1990) to 2×10^8 M^{-1} (Witherell & Uhlenbeck, 1989, 1990). Computer sim-

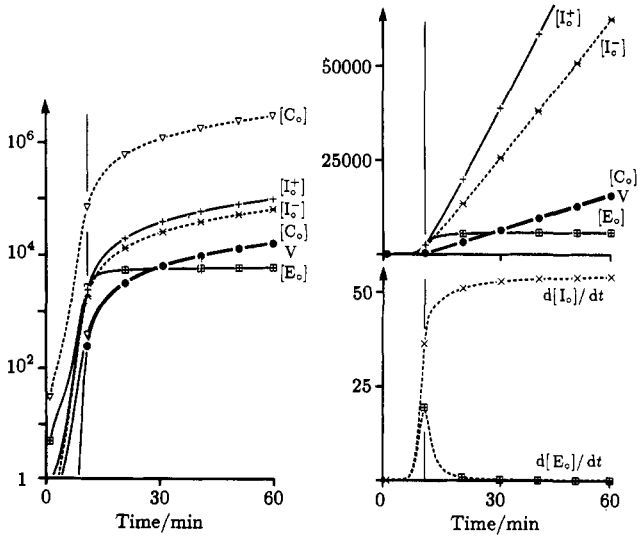

FIGURE 6: Computer simulation of the standard mechanism. The concentration profiles were obtained by numerical integration of the differential equations derived from Appendix D with the Gear algorithm (Hindmarsh, 1980). The standard rate constant values in Appendix D were used. The initial concentration of $[I^+]$ was one copy, of $[R_0]$ 5000 copies, of buffered substrate the equivalent of 2000 RNA strands; all other components were initially absent. One molecule per host cell corresponds to a concentration of about 1 nM. The rather low concentration of ribosomes was chosen to compensate for the polyribosomal state, which probably binds up to 10 ribosomes per phage RNA strand. Left: Semilogarithmic plot of concentration profiles. Total (complexed or free) concentrations of $[I_0^+]$ (+), $[I_0^-]$ (*), and $[E_0]$ (⊞), calculated as copies per cell, of $[C_0]$ (▽, full line calculated as phage equivalents per cell, broken line in copies per cell), and of mature phages per cell (●). The equivalence point is shown as a vertical line. Under these conditions nearly all coat protein present in the late infection phase is incorporated into mature viruses. Note the hyperbolic growth of the RNA and protein species between 6 and 12 min after infection and the rapid change to linear growth at the equivalence point. Top right: Linear plot of concentration files. Concentrations and symbols as in the semilogarithmic plot. Bottom right: Rate profiles for synthesis of replicase ($d[E_0]/dt$) and RNA ($d[I_0]/dt$), calculated as copies synthesized per cell and second. In this profile, substrate is buffered, i.e., always present in constant amount. If the substrate is not buffered but is consumed by replication and resupplied by biosynthesis at a constant rate, the RNA synthesis rate goes through a sharp maximum at the equivalence point and comes down in the late infection phase to a lower constant rate.

ulations show that using these in vitro values without additional assumptions could cause coat protein competing so well with ribosomes that accumulation of replicase to a concentration sufficient to trigger the hyperbolic growth phase would be prevented. (Note that one molecule per cell corresponds to about 10^{-9} M.) There are several plausible ways to treat this question and obtain simulated concentration profiles in satisfactory agreement with the experimentally measured profiles. The simplest one is to assume that when ribosomes, particularly as polyribosomes, translate the coat protein (and A1) cistron, the approach of the ribosomes to the coat protein binding site facilitates dissociation of the coat protein–RNA complex. The second one, which is possible only with class III phages where the ribosome and coat protein binding sites do not fully overlap, is that ribosomes may bind and await dissociation of the coat protein–RNA complex to start replicase synthesis. The braking of the replicase production contributes to a faster pace of RNA growth in comparison to replicase production.

The Equivalence Point. The hyperbolic phase, no matter how the rate constant values for the different steps are chosen, invariably ends at the equivalence point discussed above. Over a very short time span the metabolism of viral growth is totally changed as the RNA is suddenly saturated with replicase and ribosomes. While most of the plus strand is bound to ribosomes, the minus strands are saturated with replicase. The processes are discussed in detailed in the caption to Figure 6.

The Late Phase of Infection. In the late phase of infection the concentration of replicase was found to be constant; this phenomenon was attributed to translation repression. However, the simulations show that while translation repression certainly retards replicase production, it cannot shut it off entirely; another factor may cause the shut-off. Binding of ribosomes at the replicase ribosome recognition box requires previous binding of ribosomes at the coat protein recognition box. The accessibility of the replicase recognition box, which does not seem to be blocked by strong secondary structures (Skribkin et al., 1990), is probably due to the unfolding of the tertiary structure, and it seems plausible to attribute the unfolding of the RNA tertiary structure to the polyribosomal state. In the numerical simulations, cooperative binding of ribosomes can be easily introduced by a higher order dependence of the rate on the ribosome concentration, and under these conditions a complete arrest of replicase binding is observed as soon as the ribosomes have been titrated out by RNA. As discussed in Appendix B, it is unlikely that replication of the RNA is severely affected by the binding of coat protein, and coat protein biosynthesis is not inhibited at all. Thus RNA and coat protein are produced at constant rates. Their free forms cannot accumulate further in the cell, because at higher coat protein concentrations assembly to mature phages sets in. Indeed, in the late infection phase, the entire host cell metabolic apparatus is absorbed in making mature phages, forcing the turnoff of all other synthesis. The rate is high: the entire burst of about 10000–20000 phages is synthesized in the last 20 min of infection. This synthesis would use up a pool corresponding to at least 80 mM in nucleotide triphosphates and 160 mM in amino acids. This synthetic activity is about equal to what is required to double the host cell content at maximum speed, i.e., the synthetic capacity of the host cell is stressed to its limits. Furthermore, the required synthetic productivity demands economy, i.e., the production of material wasted at host lysis must be largely avoided. The simulations demonstrate that the proposed mechanism does indeed provide for this achievement, for practically all material synthesized after the equivalence point does show up in mature phages. The rather insoluble hydrophobic coat protein would precipitate in the cell at the concentrations approached in the infection were it not removed in the process of packaging plus-strand RNA into virions.

The economy in synthesis made it unnecessary in the simulations to limit growth by assuming a limited supply of nucleotides. Would the replicase production not arrest, the cell metabolic apparatus would be incapable of supplying the required nucleotide pool and synthesis would slow down due to the insufficient precursor supply. Indeed, simulations where replicase production was forced to continue in the late phase made it necessary to assume supply limitations in order to avoid unrealistic production rates.

The asymmetry in plus- to minus-strand production is high in the late phase of infection even when equal values are assumed for the rate constants in the plus and minus replication cycles. Due to the steady state of the synthesizing complexes there are no restrictions in asymmetry; the slow rise in the minus-strand concentration in the late infection cycle further stimulates plus-strand synthesis at the expense of minus-strand production. At lysis time the ratios of total plus strand to total minus strand may reach values of about 10,

comparable what has been measured experimentally. In the cell, minus-strand synthesis requires the presence of the α subunit of the replicase as well as a host factor, both of which are not necessary for plus-strand synthesis. Limitations of both factors may further contribute to the asymmetry found in vivo; indeed, replicase isolated from infected cells always contains substantial amounts of enzyme lacking the α subunit (Sumper & Luce, 1975).

In this mechanism, we deliberately disregarded the additional proteins A1 and A2. They do of course play a vital role in the infection, but it was apparently correct, for accounting for the observed kinetics, to assume that they do not interact directly with the synthesis apparatus.

Robustness of the Mechanism. How strongly do the computed infection profiles depend on the parameters used in the simulation? It is obvious that a valid mechanism has to be qualitatively robust against moderate changes in the rate constant values, because it can be expected that changes in growth conditions (e.g., nutrients in the growth medium) affect substrate concentrations and growth rate constants.

The mechanism presented proved in exhaustive trials to be rather insensitive to moderate changes in the rate constant values. It was not necessary to choose the production rate constant values carefully in order to get balanced growth. Too small or too high rate constant values just stretch or shrink the profiles along the time axis without altering their essential features.

There are some conditions, however, which cannot be violated without sacrificing profiles resembling the in vivo and in vitro observations. The main one is that the rate constants for RNA binding to replicase, ribosomes, and coat protein must not differ from one another by more than 2 orders of magnitude. Coat protein binding to RNA should be faster than binding of ribosomes or replicase, otherwise translation repression is not tight enough, and its relative rate of binding must not be too high, to avoid premature turnoff of replicase production and thus RNA synthesis; otherwise not even a single mature phage can be produced.

In the same way, RNA and protein synthesis must be balanced by about equal rates of replicase and protein binding. A moderate competitive advantage of replicase binding is also tolerable. Note (cf. Appendix C) that a hypercyclic mechanism does not necessarily result in hyperbolic growth. Kinetics based on enzyme–template binding, in order to be hyperbolic, requires simultaneous growth of both enzyme and template and balanced competition between translation (requiring binding of ribosomes) and replication (binding of the replicase translation product).

Outlook

The principal feature of the hypercycle is the feedback loop coupling genotype and phenotype. A genotype, to be conserved from generation to generation, has to provide phenotypic information for its (replicative) conservation expressed in the inherently autocatalytic, and thus entirely selfish, nature of the replication cycle. The replicator may express additional information by transcription or translation that is functionally useful for the larger ensemble of genotypes defining the composite genome of a living entity. Here the second feedback loop that integrates phenotypes with their genotypes comes into play.

The system studied in this paper is a prototype hypercycle. The viral genome is a 2-fold source of information. As a template it instructs its own reproduction; as a messenger it expresses information which is materialized through translation. Translation products—the replicase subunit and the coat protein—provide a specific feedback loop that regulates viral growth. The genotypic information, potentially general, is made specific for the viral replicator rather than for the many other potential replicators present in the host cell through the hypercyclic feedback loop.

Experimentally, the hypercycle expresses itself by the exposive growth profiles of RNA and protein: The coincidence of the bursts of RNA and protein synthesis indicates their synchronous and cooperative involvement, while the shrinkage of the doubling times demonstrates hyperbolic rather than exponential growth. At the same time the coat protein, by its inhibitory action, functions as a regulator.

The internal regulation due to the hypercyclic kinetics of amplification of virus protein and RNA inside the host cell is not reflected in the growth kinetics of the virus population in a cell culture. In the case of phage Qβ an infected host cell bursts every 40 min and releases about a thousand phage particles that are able to infect other cells, resulting in exponential kinetics (Biebricher et al., 1987) not at all reminiscent of the internal hypercyclic tuning of the replication mechanism.

In fact, the example of RNA plus-strand virus infection shows not only the typical features of hypercyclic mechanisms but also their limitations. The originally proposed hypercycle models using exclusively positive (i.e., enhancing) feedback loops had three shortcomings that are very apparent in the virus case: (1) The hyperbolic growth, as selectively decisive as it is, is per se too rambunctious. It requires regulation by negative feedback, i.e., inhibition, for otherwise it would exhaust all environmental resources. The virus has adapted one of its proteins, whose better known function is to provide a protective envelope, to throttle production of the self-enhancing replicase late in the infection cycle. The coat protein is a translational repressor. (2) A replicator alone is self-selective. In other words, a mutant that replicates more efficiently in a given environment or that is more stable will outgrow its competitors and be selected. In this case, the mutant genotype itself is the target of selection. A genotype, however, that encodes a function that is important for the system as a whole is not the immediate and exclusive target of its functional translation product. Hence, a mutant that has a translation product with advantageous functional efficiency usually will not be specifically selective for its own genotype; it will be just as advantageous for its precursor genotype, which encodes the less advantageous protein. On the other hand, there could be mutants that are functionally disadvantageous for the system but as targets selectively advantageous and are thus parasites. The "6S RNA" found in Qβ-infected cells (Banerjee et al., 1967) is probably a class of such parasites. (3) Since there are two types of mutations, those affecting the genotype as a target and those affecting the phenotype as functional entity, all sorts of positive and negative combinations are possible. A drastic effect is seen in phage infection: Only about 10% of the phages proliferated by a burst turn out to be infectious, 90% being "lethal" mutants. In previous theoretical work on hypercycles we suggested that compartmentation of the hypercycle is required to avoid takeover by parasites (Eigen et al., 1980), which indeed takes place in infection. In the host cell (where the hypercycle is active), all genotypes that are accepted as targets of replication are amplified regardless of whether they encode deleterious functions. Selection for intact functional properties occurs only after infecting a new cell, i.e., after recompartmentation of the virus genotype.

The viral system described in this paper is but the simplest case of a (one-member) hypercycle. Eukaryotic RNA viruses are more complicated (Perez Bercoff, 1987): the logistic

difficulties to regulate different cistrons of eukaryotic positive-strand viruses independently may have provided the selection pressure for the evolution of minus-strand viruses and/or multisegmented RNA genomes. Viruses with several RNA genome segments and different replicase genes (e.g., influenza viruses) might offer the possibility of a higher order hypercycle.

The hypercycle originally had been introduced as an organization principle that could overcome the size limitation on genotypes set by the error threshold. A genotype can maintain its information during iterative replications if its size remains below a threshold value—usually the maximum genome size is inversely proportional to the error rate (Eigen & Schuster, 1977). For single-stranded RNA viruses with well-adapted replicases, error rates are of the magnitude of (3×10^{-4})–10^{-5} per nucleotide and per replication round. Primordial systems must have had higher error rates, probably limiting the selectively stable information content to the size of one (small) gene. In order to build up larger information contents several genes have had to cooperate rather than to compete. Such a cooperation could have been organized by combining hypercyclic coupling with compartmentation; the latter alone would not provide a stable coexistence of the otherwise competitive replicators (Eigen et al., 1980).

The virus system we have studied suggests that hypercyclic organization functions optimally if it contains both enhancing and inhibitory feedback loops. Negative feedback by coat protein may well be rather common in RNA viruses; plant viruses, e.g., TMV, usually contain a binding region for coat protein which is known to start as nucleus for packaging. In nature it is critically timed so as to terminate but not to prevent the hyperbolic burst phase. Our simulations have shown that the presence of too-abundant coat protein in the very early phase of Qβ infection would prevent hyperbolic amplification entirely. This finding may suggest possibilities for new antiviral strategies.

In early evolution, negative feedback loops probably played roles that were just as important then as the roles they play in viral infections are today. One could imagine as a possible hypercyclic organization a system of genes located on RNA strands that have promoter and repressor regions. Such a system would need only one or a few replicase genes as positive feedback elements to produce an organization analogous to the known operon structure of prokaryotic DNA, now fixed inside the prokaryotic cell and entirely adapted to regulate and synchronize metabolism and cell division but which, in fact, may be the direct descendant of an evolutionary product of early hypercyclic organization.

Appendix A: Analytical Treatment of RNA Replication in Vitro: The Fibonaccian Mechanism

The reaction cycle in Figure 2 (left) can be condensed into the diagram shown in Figure 2 (right), where E is free replicase and I is free RNA, k_A describes the formation of the active enzyme–template complex EI, k_E comprises all steps of initiation, elongation, and replica release, and k_D is the rate constant of dissociation of the inactive template–enzyme complex IE, which is found to be the slowest step in the in vitro replication kinetics of short-chain self-replicating RNA species.

The rate equations are

$$\frac{d[EI]}{dt} = k_A[E][I] - k_E[EI]$$

$$\frac{d[IE]}{dt} = k_E[EI] - k_D[IE]$$

$$\frac{d[I]}{dt} = k_E[EI] + k_D[IE] - k_A[E][I] \qquad (A1)$$

where $[E_0] = [E] + [EI] + [IE]$ is a constant and $[I_0] = [I] + [EI] + [IE]$ is a variable. The first step (associated with k_A) is so fast for the conditions generally used in in vitro experiments that for $[E_0] > [I_0]$ free [I] will remain small compared to [EI] and [IE]. Thus a partial steady-state assumption

$$\frac{d[EI]}{dt}, \frac{d[IE]}{dt} \gg \frac{d[I]}{dt} \approx 0 \qquad (A2)$$

and

$$k_A[I][E] \approx k_E[EI] + k_D[IE] \qquad (A3)$$

reduces system A1 to

$$\frac{d[EI]}{dt} = k_D[IE] \quad \text{and} \quad \frac{d[IE]}{dt} = k_E[EI] - k_D[IE] \qquad (A4)$$

The eigenvalues can be obtained from the characteristic equations

$$\begin{pmatrix} -\lambda & k_D \\ k_E & -(\lambda + k_D) \end{pmatrix} = 0 \qquad (A5)$$

$$\lambda^2 + \lambda k_D - k_E k_D = 0 \Rightarrow \lambda^{\pm} = -\frac{k_D}{2}\left[1 \pm \left(1 + \frac{4k_E}{k_D}\right)^{1/2}\right] \qquad (A6)$$

The negative eigenvalue λ^- describes the internal relaxation of the coupled reaction system, while the positive one, λ^+, is the growth constant for the exponential amplification of I_0. The complete solution is hence an exponential growth function following an induction period.

Three limiting cases are of interest:

$$k_D \ll k_E \qquad \lambda^{\pm} = \pm(k_E k_D)^{1/2} \qquad (A7)$$

$$k_D = k_E \qquad \lambda^{\pm} = k_D\frac{\pm\sqrt{5} - 1}{2} \qquad (A8)$$

$$k_D \gg k_E \qquad \lambda^+ = k_E, \quad \lambda^- = -k_D \qquad (A9)$$

The first case is closest to what is observed experimentally, i.e., the absolute values of λ^+ and λ^- observed are of the same magnitude. The second case corresponds to the Fibonacci series, here in the form of continuous growth starting with an induction period resulting from superposition of the two exponentials with positive and negative eigenvalues [where $(\sqrt{5} - 1)/2$ is the golden ratio and $(\sqrt{5} + 1)/2$ its inverse]. Only the third case represents the simple autocatalytic form of the net reaction: $I_0 \xrightarrow{E} 2I_0$ with $[I] + [IE] \ll [EI] \approx [I_0]$. For very large ratios k_D/k_E, the system reduces to a single equation and the induction period is compressed into a negligibly small initial time interval:

$$\frac{d[I_0]}{dt} = k_E[I_0] \Rightarrow [I_0]_t = [I_0]_{t=0}e^{k_E t} \qquad (A10)$$

For $[E_0] < [I_0]$ eqs A2 and A3 are no longer fulfilled, because [EI] and [IE] become saturated due to $[E_0]$ = constant. Here $d[I]/dt$ $(= d[I_0]/dt)$ becomes finite while $d[IE]/dt$ and $d[EI]/dt$ vanish and system A1 reduces to

$$\frac{d[I_0]}{dt} = \frac{d[I]}{dt} = k_A[I][E] = k_E[EI] = k_D[IE] = \text{constant} \equiv c$$

$$[I_0]_t = [I_0]_{t=0} + ct \qquad (A11)$$

[I] and hence $[I_0]$ grows linearly with time, [EI] and [IE] remain (approximately) constant, and [E] decreases in inverse proportion to $[I_0]$. An apparent enzyme binding constant $K_{IE} \equiv [E_c]/[E][I]$ can be defined, where $[E_c] = [IE] + [EI]$ ($\approx [E_0]$ if $[I_0] > [E_0]$) comprises all RNA–enzyme complexes involved with initiation, elongation, and replica release. It should be stressed that K_{IE} is a steady-state constant, because thermodynamically equilibrated binding cannot occur when replication is proceeding; initiation and elongation are practically irreversible and much faster than dissociation of the active enzyme–template complex. The thermodynamic binding constant in the absence of replication has been determined to be 10^7 M^{-1} (Werner, 1991), but the steady-state constant K_{IE} is even larger ($\sim 10^9$ M^{-1}), causing a very sharp transition from exponential to linear growth when $[E_0] \approx [I_0]$. The sharpness is dependent upon the width of the period for transition from $[I] \ll [E]$ to $[I] \gg [E]$. The sharp transition point suggests an analogy to a titration end point (Winkler-Oswatitsch & Eigen, 1979). At the transition point the RNA synthesis rate parameter changes abruptly from $(k_E k_D)^{1/2}$ to k_D.

Appendix B: Biochemical Data for the Infection Mechanism

The Virus Infection Cycle. The RNA bacteriophages (Leviviridae), first discovered in 1961 (Loeb & Zinder), have genome sizes of only 3.5–4.3 kb, coding for only a very few genes, and are thus the most primitive organisms we know. Their nucleotide sequences (Fiers et al., 1976; Mekler, 1981; Inokuchi et al., 1986, 1988; Berzin et al., 1987), genome organizations (Jeppesen et al., 1970; Engelhardt et al., 1967; Weissmann et al., 1973), and the biochemistry of most synthesis steps are known. In 1974 Weissmann derived a schematic description of the infection cycle of these phages.

In the first step, viruses adsorb to the F-pili of the host (Crawford & Gesteland, 1964) and the viral RNA enters the cell. It uses the translation apparatus of the host to synthesize the viral subunit for the RNA replicase (Kamen, 1970; Kondo et al., 1970)—the other three subunits being the ribosomal protein S1 (Wahba et al., 1974) and the elongation factors (of protein biosynthesis) Tu and Ts (Blumenthal et al., 1972)—and the coat protein. The biosynthesis of viral proteins has been studied in detail (Kozak & Nathans, 1972; Lodish, 1975; Robertson, 1975; Steitz, 1975).

Viral RNA replication requires only the viral RNA replicase and a single host factor (Franze de Fernandez et al., 1968; Kajutani & Ishihama, 1991) for performing all steps of RNA replication.

Mature phages, produced by self-assembly of viral RNA and coat proteins in the late phase of infection, accumulate in the cell (Godson, 1968). The replication cycle ends with lysis of the host cell, setting free a burst of about 10^4 phages, only part of which are infectious. Lysis proteins are produced late in the infection process (Beremand & Blumenthal, 1979; Karnik & Billeter, 1983), but the effect of these proteins as well as that of other minor constituents of the capsid was neglected for the simulations.

RNA Replication. Of the different processes, the replication of viral Qβ RNA has been particularly well studied, both in vivo and in vitro (Weissmann et al., 1967, 1968, 1973). The essential mechanistic features of RNA replication have been described in Appendix A.

Viral plus strands are only accepted by replicase when a specific host factor is also present; under saturating concentrations of host factor about equal amounts of plus and minus strands are synthesized (Kamen, 1975). Different RNA phages use different host factors (Yonesaki & Aoyama, 1981), influencing the specificity of the replicases for their viral RNA (Miyake et al., 1971; Yonesaki et al., 1982). Under conditions where enzyme is in excess over RNA, more than one replicase molecule may be active on the same template, i.e., multiple replication forks can be observed (Fenwick et al., 1964; Franklin, 1966). This effect is probably adequately taken into account simply by assuming a higher rate constant value for chain elongation (300 nucleotides/s under optimal conditions).

Biosynthesis of Viral Proteins. The required viral proteins are made in vastly different amounts at different times of the infection process (Viñuela et al., 1967). Regulation occurs at the translation level by use of overlapping and thus interfering specific binding sites at the viral RNA genome.

Host ribosomes bind to recognition sites at the RNA Shine–Dalgarno boxes (Shine & Dalgarno, 1974). However, the RNA is highly structured and at first only allows access of the ribosome to the exposed Shine–Dalgarno box of the coat protein cistron (Lodish, 1968; Steitz, 1969; Hindley & Staples, 1969; Staples et al., 1971). During biosynthesis of coat protein the Shine–Dalgarno box of the replicase cistron becomes available for binding to ribosomes. This was included in the reaction mechanism through formation of the corresponding double translation complex. A high rate of protein production of the ribosome–RNA complex was assumed in order to take polyribosomal protein synthesis into account. Ribosome binding was assumed to be a second-order reaction. The available ribosome concentration, however, is an uncertain quantity. Host cells growing at maximum speed have about 2×10^4 ribosomes; initially many of them are bound to host mRNA, which continues to be produced after phage infection starts. If we assume (with obvious good reason) that viral RNA competes successfully with mRNA, we can calculate the total ribosome concentration using an appropriate binding rate constant.

Translation Repression. Translation and replication are separated by the mutually excluding binding sites for ribosome (at the coat cistron initiation site) and replicase (Vollenweider et al., 1976; Meyer et al., 1981). Once replicase is bound at this site, already bound ribosomes can finish biosynthesis and fall off, after which the 3′-terminus of the RNA is ready for initiation of replication (Weissmann, 1974), which proceeds in the 5′ to 3′ direction of the replica strand. RNA serving as template is not available for ribosome binding. The inactive template–replicase complex remaining after completion of the replica chain, however, is assumed to be able to bind ribosomes normally.

Coat protein serves as translation repressor for biosynthesis of the replicase (Weber et al., 1972; Peabody, 1990) by binding specifically to the translation operator, a hairpin also containing the Shine–Dalgarno box of the replicase cistron (Robertson et al., 1968; Bernardi & Spahr, 1972). The kinetics of binding of coat dimer to the translation operator has been extensively investigated (Uhlenbeck, 1986; Beckett & Uhlenbeck, 1988; Beckett et al., 1988; Witherell & Uhlenbeck, 1989; Witherell et al., 1990). Replicase production is thus turned off in the late phase, while biosynthesis of coat protein is not affected. Binding of coat protein to RNA neither inhibits initiation nor seriously affects elongation of replication, because of the rapid dissociation and reassociation of the coat protein–RNA complex.

Packaging. Binding of coat protein to plus-strand RNA also provides the nucleus for aggregating further coat molecules to form the virion shell (Beckett et al., 1988). Phage assembly takes place in vitro spontaneously, i.e., without additional host factors. Even though 179 further protein mol-

ecules bind, one step, likely the first, is probably rate determining; therefore, a second-order rate constant for the packaging reaction was assumed for the kinetic model. The time lag in appearance of mature intracellular phages indicates a cooperative effect of coat proteins in the packaging process.

Plus/Minus Asymmetry. Mature virus particles contain only plus-strand RNA. Furthermore, plus strands predominate in the late stage of infection. It has been thought that the host factor, required for synthesizing minus strands, might be responsible for this asymmetry; different rate constants for the production of plus and minus strands are also conceivable. However, the asymmetry can also be caused simply by the fact that replicase has to compete for plus strands with ribosomes and coat protein while minus strands are all free for production of plus strands. Indeed, this fact alone was found in the simulations to suffice for causing a substantial excess of plus strands (free and bound) over minus strands; therefore, additional ad hoc assumptions to increase the asymmetry were found to be unnecessary.

Appendix C: Analysis of Simplified Infection Kinetics

We are interested in the rate laws for concentration growth of total RNA and total replicase, namely

$$\frac{d[I_0]}{dt} = k_{repl}[EI]$$
$$\frac{d[E_0]}{dt} = k_{trans}[RI] \qquad (C1)$$

In these equations EI denotes replicase (E) bound to RNA, RI denotes ribosomes (R) bound to RNA, and the total concentrations of RNA, replicase, and ribosome are defined through the conservation equations:

$$[EI] + [RI] = [I_0]$$
$$[EI] + [E] = [E_0]$$
$$[RI] + [R] = [R_0] \qquad (C2)$$

Since ribosome is available from the onset at relatively high concentration and since binding of I to either R or E is strong, competition of R and E for I keeps the concentration of [I] negligibly small with respect to [RI], [R], [EI], and [E], and therefore the free RNA concentration [I] has been neglected in the conservation equation for $[I_0]$. The assumption of negligible free [I] holds up to an equivalence point where $[I_0] = [E_0] + [R_0]$. In order to determine a solution to eqs C1 and C2, one additional equation is required. We introduce the expression that the ratio of translation complex to replication complex concentrations is proportional to the ratio of free ribosome to free replicase concentrations, i.e., the relation

$$\frac{[RI]}{[EI]} = K_c\frac{[R]}{[E]} \qquad (C3)$$

Since reversible binding does not occur, the competition constant K_c cannot be a true equilibrium constant, because, as for the in vitro replication described in Appendix A, we are dealing with a nonequilibrium situation. Ribosome and enzyme bind at their respective recognition sites and immediately start to translocate along the chain, driven by the practically irreversible polymerization reactions of protein and RNA biosynthesis, until finally set free after completing the synthesis at their termination signals of the template RNA strand (at the next downstream stop codon in protein synthesis, at the 5′-terminus in RNA synthesis) to bind again competitively to I and reenter a new synthesis round. The constant K_c may be thought of as representing the ratio of the two association rate constants for binding of I to either R or E. Equations C2 and C3 can be combined to yield a quadratic equation for [EI] with the physically significant root

$$[EI] = \frac{\alpha}{2(1-K_c)}\left[1 - \left(1 - \frac{4[E_0][I_0](1-K_c)}{\alpha^2}\right)^{1/2}\right] \qquad (C4)$$

where

$$\alpha = K_c[R_0] + [E_0] + (1-K_c)[I_0] \qquad (C5)$$

Furthermore we have from eq C2

$$[RI] = [I_0] - [EI] \qquad (C6)$$

Equations C4 and C6 represent titration relations with the equivalence point $[I_0] = [R_0] + [E_0]$. They hold up only to this point where, due to saturation of R and E, free RNA suddenly rises to nonnegligible values. As in the case treated in Appendix A, the equivalence point is a sharp end point, because binding of RNA to both enzyme and ribosomes is practically irreversible.

As long as the K_c values are sufficiently close to one, the root in eq C4 may be expanded for all values of $[E_0]$ and $[I_0]$, yielding

$$[EI] = \frac{[I_0][E_0]}{K_c[R_0] + [E_0] + (1-K_c)[I_0]} \qquad (C7)$$

$$[RI] = [I_0]\left(1 - \frac{[E_0]}{K_c[R_0] + [E_0] + (1-K_c)[I_0]}\right) \qquad (C8)$$

For $K_c[R_0] \gg [E_0] \gg (1-K_c)[I_0]$, one obtains rate equations for $[I_0] = x$ and $[E_0] = y$ of the form

$$\frac{dx}{dt} = axy(1-cy) \quad \text{and} \quad \frac{dy}{dt} = bx(1-cy) \qquad (C9)$$

where a, b, and c are constants. As long as $cy \ll 1$, y will grow in proportion to x; thus, x (and y) will follow a hyperbolic growth law according to

$$\frac{dx}{dt} = k_2x^2 \Rightarrow x(t) = \frac{x_{t=0}}{1 - x_{t=0}k_2t} \qquad (C10)$$

The above approximation is also valid, irrespective of the value of K_c, if either $[I_0]$ or $[E_0]$ is small with respect to the other or to $K_c[R_0]$. For K_c values appreciably larger or smaller than 1, the solutions change from hyperbolic to exponential when approaching the equivalence point. Such a transitory type of growth behavior may be represented by rate equations of the form

$$\frac{dx}{dt} = k_px^p \quad \text{with} \quad 1 < p < 2 \qquad (C11)$$

having the hyperbolic solutions

$$x(t) = \frac{x_{t=0}}{[1 - (p-1)x_{t=0}^{p-1}k_pt]^{1/(p-1)}} \qquad (C12)$$

with the limiting case of the exponential growth law for $p = 1$:

$$x(t) = x_{t=0}\exp(k_1t) \qquad (C13)$$

The doubling period τ_{double}, constant for exponential growth, decreases steadily with time for hyperbolic growth:

$$\tau_{double}(t) = \tau_{double}(t=0)[1 - (p-1)x_{t=0}^{p-1}k_pt] \qquad (C14)$$

A comparison of the calculated values for τ_{double} with the experimentally determined values (Figure 1) reveals that at

the time of maximal growth rate the p value is between 2 and 1.5.

Appendix D: Data for the Simulation

Protein Synthesis. Binding of ribosomes to RNA had to be estimated due to the lack of experimental measurements (standard rate constant values are in s^{-1} and $M^{-1}\ s^{-1}$ units):

$$I^+ + R \rightarrow R_C I^+ \ (k_{IR} = 10^6)$$

$$R_C I^+ \rightarrow C + I^+ + R \ (1/T_c = 0.5)$$

$$R_C I^+ + R \rightarrow R_C I^+ R_E \text{ (cf. below)}$$

$$R_C I^+ R_E \rightarrow R_C I^+ + E + R \ (1/T_E = 0.2)$$

$$R_C I^+ R_E \rightarrow R_C I^+ R_E + C \ (1/T_c)$$

Replication. Equal rate constants for plus and minus cycles were chosen in the range of those measured for short-chained RNA species. Elongation of plus and minus strands requires substrate, which may become limiting in the late infection phase. k_{sub} is the rate of substrate formation (biosynthesis). Most simulations assumed buffered substrate concentrations, however:

$$I^+ + E \rightarrow EI^+ \ (k_{IPE} = 10^7)$$

$$I^- + E \rightarrow EI^- \ (k_{IME} = 10^7)$$

$$EI^+ + S \rightarrow I^+E + I^- \ (k_{TRP}[S] = 0.01)$$

$$EI^- + S \rightarrow I^-E + I^+ \ (k_{TRM}[S] = 0.01)$$

$$I^+E \rightarrow I^+ + E \ (1/T_{D+} = 0.02)$$

$$I^-E \rightarrow I^- + E \ (1/T_{D-} = 0.02)$$

$$\rightarrow S \ (k_{sub} = 0 \text{ and } d[S]/dt = 0 \text{ for substrate buffering})$$

Product Inhibition and Packaging. The complex of coat protein with plus strand serves as the condensation initiator for packaging (Witherell et al., 1990). Rates and equilibrium constants of coat protein binding to viral RNAs have been measured (Uhlenbeck, 1987; Witherell & Uhlenbeck, 1989):

$$I^+ + C \rightarrow I^+C \ (k_{IC} = 2 \times 10^9)$$

$$I^+C \rightarrow I^+ + C \ (k_{DIC} = 20)$$

$$I^+C + R \rightarrow R_C I^+C \ (k_{CIR} = 10^6)$$

$$R_C I^+C \rightarrow C + R + I^+C \ (1/T_{DCIR} = 0.5)$$

$$I^+C + (179)C \rightarrow V \text{ (cf. below)}$$

$$I^+C + E \rightarrow EI^+ + C \ (k_{IPE})$$

Inactive Replicates–Plus-Strand Complexes. The inactive complexes bind ribosomes and produce coat protein. Coat protein–plus-strand complexes bind to replicase; for simplicity, shedding of the coat protein is assumed to accompany the binding step:

$$I^+E + R \rightarrow R_C I^+E \ (k_{IR})$$

$$R_C I^+E \rightarrow R_C I^+ + E \ (1/T_{D+})$$

$$I^+C + R \rightarrow R_C I^+C \ (k_{IR})$$

$$R_C I^+C \rightarrow R_C I^+ + C \ (k_{DIC})$$

$$R_C I^+C \rightarrow I^+C + R + C \ (1/T_C)$$

Cooperative Processes. It is plausible to assume that the process of converting the complexes of RNA and coat protein into mature phage particles involves a cooperative association of coat protein molecules. There is also good reason to assume that polysomal occupation of RNA by ribosomes facilitates binding of ribosomes at the Shine–Dalgarno box for the replicase cistron. These two cooperative effects were taken into account in the mechanism by assuming second-order concentration dependences for coat protein and ribosomes, respectively, in these processes. In the stoichiometry, however, 179 coat protein and molecules and one ribosome, respectively, were counted as reacting:

$$I^+C + 2(179)C \rightarrow V \ (k_{pack} = 10^8)$$

$$R_C I^+ + 2(1)R \rightarrow R_C I^+ R_E \ (k_{CR} = 10^{12})$$

References

Adams, M. H. (1959) *Bacteriophages*, Interscience Publishers, New York.

Ammann, J., Delius, H., & Hofschneider, P. H. (1964) *J. Mol. Biol. 10*, 557–561.

August, J. T., Shapiro, L., & Eoyang, L. (1965) *J. Mol. Biol. 11*, 257–271.

Banerjee, A. K., Rensing, U., & August, J. R. (1967) *Proc. Natl. Acad. Sci. U.S.A. 47*, 986–993.

Beckett, D., & Uhlenbeck, O. C. (1988) *J. Mol. Biol. 204*, 927–938.

Beckett, D., Wu, H.-N., & Uhlenbeck, O. C. (1988) *J. Mol. Biol. 204*, 939–947.

Beremand, M. N., & Blumenthal, T. (1979) *Cell 18*, 257–266.

Bernardi, A., & Spahr, P.-F. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3033–3037.

Berzin, V., Avots, A., Jansone, I., Ginthere, L., & Tsimanis, A. (1987) *Nucleic Acids Res. 15*, 6741–6741.

Biebricher, C. K. (1986) *Chem. Scr. 26B*, 51–57.

Biebricher, C. K., Eigen, M., & Luce, R. (1981a) *J. Mol. Biol. 148*, 369–390.

Biebricher, C. K., Eigen, M., & Luce, R. (1981b) *J. Mol. Biol. 148*, 391–410.

Biebricher, C. K., Eigen, M., & Gardiner, W. C. (1983) *Biochemistry 22*, 2544–2559.

Biebricher, C. K., Eigen, M., & Gardiner, W. C. (1984) *Biochemistry 23*, 3186–3194.

Biebricher, C. K., Eigen, M., & Gardiner, W. C. (1985) *Biochemistry 24*, 6550–6560.

Biebricher, C. K., Eigen, M., Gardiner, W. C., Husimi, Y., Keweloh, H.-C., & Obst, A. (1987) In *Complex Chemical Reaction Systems, Mathematical Modeling and Simulation* (Warnatz, J., & Jäger, W., Eds.) pp 17–38, Springer-Verlag, Berlin.

Biebricher, C. K., Eigen, M., & Gardiner, W. C. (1991) In *Biologically Inspired Physics* (Peliti, L., Ed.) NATO series, Plenum Publishing Corp., New York (in press).

Blumenthal, T., Landers, T. A., & Weber, K. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1313–1317.

Cramer, J. H., & Sinsheimer, R. L. (1971) *J. Mol. Biol. 62*, 189–214.

Crawford, E. M., & Gesteland, R. F. (1964) *Virology 22*, 165–167.

Davies, J. W., & Kaesberg, P. (1973) *J. Virol. 12*, 1434–1441.

Dobkin, C., Mills, D. R., Kramer, F. R., & Spiegelman, S. (1979) *Biochemistry 18*, 2038–2044.

Eigen, M. (1971) *Naturwissenschaften 58*, 465–523.

Eigen, M., & Schuster, P. (1977) *Naturwissenschaften 64*, 541–565.

Eigen, M., & Schuster, P. (1978a) *Naturwissenschaften 65*, 7–41.

Eigen, M., & Schuster, P. (1978b) *Naturwissenschaften 65*, 341–369.

Eigen, M., Gardiner, W., & Schuster, P. (1980) *J. Theor. Biol. 85*, 407–411.
Ellis, D. B., & Paranchych, W. (1963) *J. Cell. Comp. Physiol. 62*, 207–213.
Engelhardt, D. L., Webster, R. E., & Zinder, N. D. (1967) *J. Mol. Biol. 29*, 45–58.
Fenwick, M. L., Erikson, R. L., & Franklin, R. M. (1964) *Science 23*, 527–530.
Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymakers, A., Vandenberghe, A., Volckaert, G., & Isebaert, M. (1976) *Nature 260*, 500–507.
Franklin, R. M. (1966) *Proc. Natl. Acad. Sci. U.S.A. 55*, 1504–1508.
Franze de Fernandez, M. T., Eoyang, L., & August, J. T. (1968) *Nature 219*, 588–590.
Fromageot, H. P. M., & Zinder, N. D. (1968) *Proc. Natl. Acad. Sci. U.S.A. 61*, 184–191.
Gebinoga, M. (1990) Kinetik der intrazellulären Entwicklung des Phagen Qβ in *Escherichia coli*, Ph.D. Thesis, Braunschweig University.
Godson, G. N. (1968) *J. Mol. Biol. 34*, 149–163.
Godson, G. N., & Sinsheimer, R. L. (1967) *J. Mol. Biol. 23*, 495–521.
Gordon, J., & Weissbach, H. (1970) *Biochemistry 9*, 4233–4236.
Gussin, G. N., Capecchi, M. R., Adams, J. M., Argetsinger, J. E., Tooze, J., Weber, K., & Watson, J. D. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 257–271.
Happe, M., & Jokusch, H. (1973) *Nature New Biol. 245*, 141–143.
Haruna, I., & Spiegelman, S. (1965a) *Proc. Natl. Acad. Sci. U.S.A. 54*, 579–587.
Haruna, I., & Spiegelman, S. (1965b) *Science 150*, 884–886.
Hindley, J., & Staples, D. H. (1969) *Nature 224*, 964–970.
Hindmarsh, A. C. (1980) *Assoc. Comput. Mach., Signum Newsl. 15*, 10–11.
Hofstetter, H., Monstein, H. J., & Weissmann, C. (1974) *Biochim. Biophys. Acta 374*, 238–247.
Horiuchi, K., & Matsuhashi, S. (1970) *Virology 42*, 49–60.
Hudson, J. B., & Paranchych, W. (1967) *J. Virol. 1*, 529–537.
Hung, P. P., & Overby, L. R. (1969) *Biochemistry 8*, 820–828.
Hung, P. P., Ling, C. M., & Overby, L. R. (1969) *Science 166*, 1638–1640.
Inokuchi, Y., Takahashi, R., Hirose, T., Inayama, S., Jacobson, A. B., & Hirashima, A. (1986) *J. Biochem. (Tokyo) 99*, 1169–1180.
Inokuchi, Y., Jacobson, A. B., Hirose, T., Inayama, S., & Hirashima, A. (1988) *Nucleic Acids Res. 16*, 6205–6221.
Jeppesen, P. G. N., Argetsinger-Steitz, J., Gesteland, R. F., & Spahr, P. F. (1970) *Nature 226*, 230–237.
Jockusch, H., Ball, L. A., & Kaesberg, P. (1970) *Virology 42*, 401–414.
Kajitani, M., & Ishihama, A. (1991) *Nucleic Acids Res. 19*, 1063–1066.
Kamen, R. (1970) *Nature 228*, 527–533.
Kamen, R. (1975) in *RNA Phages* (Zinder, N. D., Ed.) pp 203–234, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Karnik, S., & Billeter, M. (1983) *EMBO J. 2*, 1521–1526.
Kondo, M., Gallerani, R., & Weissmann, C. (1970) *Nature 228*, 525–527.
Kozak, M., & Nathans, D. (1972) *Bacteriol. Rev. 36*, 109–134.
Lodish, H. F. (1968) *Nature 220*, 345–350.
Lodish, H. F. (1971) *J. Mol. Biol. 56*, 627–632.
Lodish, H. F. (1975) in *RNA Phages* (Zinder, N. D., Ed.) pp 301–318, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Lodish, H. F., & Zinder, N. D. (1966) *Science 152*, 372–377.
Loeb, T., & Zinder, N. D. (1961) *Proc. Natl. Acad. Sci. U.S.A. 47*, 282–289.
Marino, P., Baldi, M. I., & Tocchini-Valentini, G. P. (1968) *Cold Spring Harbor Symp. Quant. Biol. 33*, 125–127.
Mekler, P. (1981) Determination of nucleotide sequences of the bacteriophage Qβ genome: Organization and evolution of an RNA virus, Dissertation, Zürich University.
Meyer, F., Weber, H., & Weissmann, C. (1981) *J. Mol. Biol. 153*, 631–660.
Mills, D. R., Dobkin, C., & Kramer, F. R. (1978) *Cell 15*, 541–550.
Miyake, T., Haruna, I., Shiba, T., Itoh, Y. H., Yamane, K., & Watanabe, I. (1971) *Proc. Natl. Acad. Sci. U.S.A. 68*, 2022–2024.
Moses, R. E., & Richardson, C. C. (1970) *Proc. Natl. Acad. Sci. U.S.A. 67*, 674–681.
Nathans, D., Oeschger, M. P., Polmar, S. K., & Eggen, K. (1969) *J. Mol. Biol. 39*, 279–292.
Peabody, D. S. (1990) *J. Biol. Chem. 265*, 5684–5689.
Perez Bercoff, R. P., Ed. (1987) *Molecular Basis of Viral Replication*, Plenum Press, New York.
Robertson, H. D. (1975) in *RNA Phages* (Zinder, N. D., Ed.) pp 113–145, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Robertson, H. D., Webster, R. E., & Zinder, N. D. (1968) *Nature 218*, 533–536.
Shine, J., & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1342–1346.
Skripkin, E. A., Adhin, M. R., de Smit, M. H., & van Duin, J. (1990) *J. Mol. Biol. 211*, 447–463.
Staples, D. H., Hindley, J., Billeter, M. A., & Weissmann, C. (1971) *Nature New Biol. 234*, 202–204.
Steitz, J. A. (1969) *Nature 224*, 957–964.
Steitz, J. A. (1975) in *RNA Phages* (Zinder, N. D., Ed.) pp 319–352, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sumper, M., & Luce, R. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 162–166.
Talbot, S. J., Goodman, S., Bates, S. R. E., Fishwick, C. W. G., & Stockley, P. G. (1990) *Nucleic Acids Res. 18*, 3521–3528.
Uhlenbeck, O. C. (1986) *Chem. Scr. 26B*, 97–101.
Viñuela, E., Algranati, I. D., & Ochoa, S. (1967) *Eur. J. Biochem. 1*, 3–11.
Vollenweider, H. J., Koller, T., Weber, H., & Weissmann, C. (1976) *J. Mol. Biol. 101*, 367–377.
Vosberg, H.-P., & Hoffmann-Berling, H. (1971) *J. Mol. Biol. 58*, 739–753.
Wahba, A. J., Miller, M. J., Niveleau, A., Landers, T. A., Carmichael, G. G., Weber, K., Kawley, D. A., & Slobin, L. I. (1974) *J. Biol. Chem. 249*, 3314–3316.
Warnatz, J., & Jäger, W. (1987) *Complex Chemical Reaction Systems. Mathematical Modelling and Simulation*, Proceedings of the Second Workshop, Heidelberg, Germany, August 11–15, 1986, Springer-Verlag, Berlin.
Weber, H., Billeter, M. A., Kahane, S., Weissmann, C., Hindley, J., & Porter, A. (1972) *Nature New Biol. 237*, 166–170.

Weissmann, C. (1974) *FEBS Lett. 40*, S10–S18.
Weissmann, C., Borst, P., Burdon, R. H., Billeter, M. A., & Ochoa, S. (1964) *Proc. Natl. Acad. Sci. U.S.A. 51*, 890–897.
Weissmann, C., Feix, G., Slor, H., & Pollet, R. (1967) *Proc. Natl. Acad. Sci. U.S.A. 57*, 1870–1877.
Weissmann, C., Feix, G., & Slor, H. (1968) *Cold Spring Harbor Symp. Quant. Biol. 33*, 83–100.
Weissmann, C., Billeter, M. A., Goodman, H. M., Hindley, J., & Weber, H. (1973) *Annu. Rev. Biochem. 42*, 303–329.
Werner, M. (1991) *Biochemistry 30*, 5832–5838.
Winkler-Oswatitsch, R., & Eigen, M. (1979) *Angew. Chem., Int. Ed. Engl. 18*, 20–49.
Witherell, G. W., & Uhlenbeck, O. C. (1989) *Biochemistry 28*, 71–76.
Witherell, G. W., Wu, H. N., & Uhlenbeck, O. C. (1990) *Biochemistry 29*, 11051–11057.
Yonesaki, T., & Aoyama, A. (1981) *J. Biochem. (Tokyo) 89*, 741–750.
Yonesaki, T., Furuse, K., Haruna, I., & Watanabe, I. (1982) *Virology 116*, 379–381.

Accelerated Publications

Direct Identification of the Active-Site Nucleophile in a DNA (Cytosine-5)-methyltransferase†

Lin Chen,‡ Andrew M. MacMillan,‡ Wilbur Chang,‡ Khosro Ezaz-Nikpay,‡ William S. Lane,§ and Gregory L. Verdine*,‡

Department of Chemistry and Microchemistry Facility, Harvard University, Cambridge, Massachusetts 02138

Received August 29, 1991; Revised Manuscript Received September 27, 1991

ABSTRACT: The overproduction, purification, and determination of the active-site catalytic nucleophile of the DNA (cytosine-5)-methyltransferase (DCMtase) enzyme M.*Hae*III are reported. Incubation of purified M.*Hae*III with an oligodeoxynucleotide specifically modified with the mechanism-based inhibitor 5-fluoro-2′-deoxycytidine [Osterman, D. G., et al. (1988) *Biochemistry 27*, 5204–5210], in the presence of the cofactor *S*-adenosyl-L-methionine (AdoMet), resulted in the formation of a covalent DNA–M.*Hae*III complex, which was purified to homogeneity. Characterization of the intact complex showed it to consist of one molecule of the FdC-containing duplex oligonucleotide, one molecule of M.*Hae*III, and one methyl group derived from AdoMet. Exhaustive proteolysis, reduction, and alkylation of the DNA–M.*Hae*III complex led to the isolation of two DNA-bound peptides—one each from treatment with Pronase or trypsin—which were subjected to peptide sequencing in order to identify the DNA attachment site. Both peptides were derived from the region of M.*Hae*III containing a Pro-Cys sequence that is conserved in all known DCMtases. At the position of this conserved Cys residue (Cys_{71}), in the sequence of each peptide, was found an unidentified amino acid residue; all other amino acid residues were in accord with the known sequence. It is thus concluded that Cys_{71} of M.*Hae*III forms a covalent bond to DNA during catalytic methyl transfer. This finding represents a direct experimental verification for the hypothesis that the conserved Cys residue of DCMtases is the catalytic nucleophile [Wu, J. C., & Santi, D. V. (1987) *J. Biol. Chem. 262*, 4778–4786]. Furthermore, the present studies provide ready access to large quantities of a homogeneous, covalent protein–DNA complex that is trapped at an intermediate stage in catalysis.

The enzymatic addition of methyl groups to DNA is an essential element of genomic function in organisms ranging from bacteria to mammals (Razin et al., 1984; Adams & Burdon, 1985). In prokaryotes, DNA methylation directs the mismatch repair and restriction–modification systems, which correct errors of replication and prevent transformation by non-self-DNA, respectively. In eukaryotes, DNA methylation serves an essential yet poorly understood role in cell differentiation and regulation of gene expression. These complex and intriguing biological effects overlay the fundamental mechanistic questions surrounding how these proteins transfer a methyl group from the cofactor *S*-adenosyl-L-methionine (AdoMet)[1] to duplex DNA. The class of DNA methyltransferase enzymes that direct methylation of the 5-position of C [DNA (cytosine-5)-methyltransferases (DCMtases); eq

† This work was supported by the donors of the Petroleum Research Fund, administered by the American Chemical Society (to G.L.V.), by NIH Grant GM 44853-02 (to G.L.V.), and by gifts from Hoffmann-La Roche, Pfizer, and Bristol-Myers Squibb. G.L.V. is a Presidential Young Investigator, a Searle Scholar, a Fellow of the Alfred P. Sloan Foundation, and a Lilly Grantee.

‡ Department of Chemistry.
§ Harvard Microchemistry Facility.

[1] Abbreviations: AdoMet, *S*-adenosyl-L-methionine; AdoHcy, *S*-adenosyl-L-homocysteine; BSA, bovine serum albumin; DCMtase, DNA (cytosine-5)-methyltransferase; ds, double stranded; DTT, dithiothreitol; ECPCR, expression-cassette polymerase chain reaction; FdC, 5-fluoro-2′-deoxycytidine; FdU, 5-fluoro-2′-deoxyuridine; IPTG, isopropyl β-D-thiogalactopyranoside; kDa, kilodaltons; M.*Hae*III, DNA (cytosine-5)-methyltransferase enzyme from *Haemophilus aegyptius*; M.*Hha*I, DCMtase from *Haemophilus haemolyticus*; oligo, oligo-2′-deoxynucleotide; PTH, phenylthiohydantoin; R.*Hae*III, restriction endonuclease from *H. aegyptius*; ss, single stranded; TMP-FdU, 4-*O*-(2,4,6-trimethylphenyl)-5-fluoro-2′-deoxyuridine.

0006-2960/91/0430-11018$02.50/0