THE MOLECULAR QUASI-SPECIES

MANFRED EIGEN AND JOHN M^{C}CASKILL

Max Planck Institut für Biophysikalische Chemie
D-3400 Göttingen, FRG

PETER SCHUSTER

Institut für Theoretische Chemie und Strahlenchemie der Universität Wien
A-1090 Wien Austria

CONTENTS

I. INTRODUCTION

1. Model and Reality

Physical chemists are well aware of the usefulness of models. An under-
standing of the fundamental properties of matter can hardly be gained from
watching reality, requiring instead the posing of if–then questions that can be
answered only by models. The nature of pressure or temperature of a gas as a
collective property of its individual atomic or molecular constituents became
obvious only through the billiard ball models of Clausius, Maxwell, and
Boltzmann, despite our later insights that true atoms or molecules have
quantized motion.

   In trying to understand the physical behavior of *viable* matter, we are
facing a quite similar situation today as did our forerunners the physical
chemists in the late-nineteenth century, when they started to look into the
nature of inanimate matter. What is the kind of physical problem that we are
really concerned with? Analyzing life at the molecular level, we realize the
fundamental dichotomy of a genotypic legislative and phenotypic executive
represented by particular forms of macromolecular organization. Asking how
such an order became established, we are referred to Darwin's principle of
natural selection. There is general consensus among biologists that this
principle not only has guided the evolution of species but must also have been
as instrumental in the evolution of the molecular forms of organization, up to
a state that might be called "the first living species." As we have learned more
about the details of this organization, it has seemed more difficult to provide
any answer to the question of the physical nature of life. Indeed, what stands
out from our increasingly detailed perspective is the tremendous complexity
of molecular organization in even the most primitive viable system. We are
facing questions such as: why did that particular organization proving viable
come about while myriads of alternative states of the system, structurally just
as stable, do not share the property of being alive and is there a guiding
principle that narrows down the number of alternatives so as to render the

appearance of life on our planet possible with meaningful physical expectation? Another, not less stringent, question then immediately comes up: Why is this molecular organization of life, which historically evolved, so perfect? Why, for instance, are enzymes optimal catalysts? The number of structurally stable alternatives that could have been assembled from the same monomeric material is so large that one could not accommodate it within the spatial and temporal limits of our universe. If, as a consequence, we must conclude that all these alternatives never could be tested, why is it that the particular choice of nature turns out to be optimal in its dynamical performance? Half a century of enzyme kinetics has clearly demonstrated this fact. The performance of enzymes can hardly be improved. Nature's solution for the enzymes represents the optimum of what can be achieved with this type of macromolecular organization.

In this chapter we shall try to give answers to these questions concerning an understanding of the physics basic to life. We shall use defined models in order to derive quantitative results, and we shall discuss the relevance of those models for realistic scenarios that can be simulated in laboratory experiments. We shall *not*, however, present hypotheses regarding the historical origin of life. The historical process of evolution depends not only on physical principles but also on historical boundary conditions. These render the historical process unique. Physics, as Eugene Wigner [1] once stated clearly, cannot deal with such unique events. Physics can deal only with regularities among events. The regularity "life," not the historical reality "life," will be the subject of this Chapter.

2. Darwinian Systems

The logic of life may be condensed into four statements:

1. Life came about through evolution.
2. Evolution is the result of variation and natural selection under conditions far from thermodynamic equilibrium.
3. a) Natural selection is a consequence of self-reproduction under conditions far from thermodynamic equilibrium.

   b) Variation is due to imprecise reproduction or other modifications involved in the processing.
4. Self-reproduction is based on structural complementarity of a particular class of molecules.

Systems adhering to this scheme may be called Darwinian because Charles Darwin was the first to reason along this line. Statements such as the preceding are too general to be of more than heuristic value. Let us therefore

apply this logic to a defined model and then analyze its consequences in more depth.

The central issue will be Darwin's principle of natural selection. In its original formulation as "survival of the fittest," it does not provide much new insight and it was promptly misunderstood. We really have to define the terms *survival* and *fittest* independently of this principle in order not to get caught in the tautological loop of "survival of the survivor."

Survival certainly is to be measured in terms of population numbers or, to use the language of the physical chemist, in terms of concentrations. A surviving type has to be present with a nonzero number of copies. Fittest, on the other hand, refers to a value parameter that characterizes the surviving type, for example, the one most efficient in producing offspring. For complex living beings with all their mutual interferences, it may be quite difficult to express it through numbers. For a self-reproductive molecular entity I_i, such as an RNA or DNA molecule possibly in combination with enzymic machinery, a selective value W_i can be precisely defined given that the reaction mechanism has been established. Then the quantitative correlation between W_i and the population number N_i is the essence of the selection problem. Survival of the fittest means that all types but the one with maximum selective value are bound to die out. In this form the selection principle is certainly only a zero-order approximation. In other words: there is no real correlation between population numbers and selective values except for the "fittest," which represents the entire system.

It is obvious that such a ruthless "all-or-none" decision could neither be a consequence of random production nor result from interactions as they are responsible for chemical equilibrium, which always settles on finite concentration ratios. It is indeed the peculiar mechanism of the reproduction process far from equilibrium that accounts for the fact of survival, and this mechanism is even active when the competitors are degenerate in their selective values, that is, if they are neutral competitors. In this limiting case, considered to be very important for the evolution of species, Darwin's principle indeed reduces to the mere tautology: survival of the survivor. Nevertheless, there are, even here, systematic quantitative regularities in the way that macroscopic populations of wild types rise and fall in a deterministic manner (as far as the process, not the particular copy choice, is concerned), which make it anything but a *trivial* correlation. This case of neutral selection has been called non-Darwinian. It should be emphasized, however, that Darwin was well aware of this possibility and described it verbally in a quite adequate way. The precise formulation of a theory of neutral selection, which then allows us to draw quantitative conclusions on the evolution of species is an achievement of the second half of this century. Kimura [2] has pioneered this new branch of population genetics.

We have called this all-or-none type of selection a *zero-order approxima-tion*. As such, and wherever the boundary conditions allow for selection rather than for coexistence, it may hold only for competitors that do not have close kinship relations in their genetic inheritance. In fact, even with strong differences in selective value, evolution would not have proceeded very far if it were based on such a correlation for natural selection. For instance, we could not explain why the enzymes are catalysts of optimal efficiency. Evolution would soon have come to a standstill on a minor hill in the value landscape, waiting for an advantageous mutation that would appear much too rarely considering the huge number of possible alternatives. It is common to rely on Darwin's principle as a deterministic tool for selection of the advantageous mutant, while at the same time assuming that this advantageous mutant could come about by nothing but pure chance. One could shock today's biologists by saying that there *must* exist also some guidance, assuring that the advantageous mutants will appear with much higher rates than the disadvantageous ones, and indeed it is considered a heresy to say that evolution can be guided other than by chance.

We shall show that in a Darwinian system evolution indeed is guided to the peaks in the value landscape through biased mutations. To be sure, there is no correlation between the (intrinsically stochastic) act of mutation and the fitness of the product; yet there is bias provided by the fitness-dependent population distribution of mutants — so to speak a mass action guidance of probabilities of mutations. To appreciate this, we have to look at the higher approximations of the selection principle. The target of selection is not a singular wild-type sequence. The fact that sequence analysis as carried out for complete viral genomes yields unique primary structures is no proof that the individual sequence as such is really present to a major extent. It only confirms that at each position the found symbol (nucleotide or amino acid residue) is the prevailing one. A typical RNA virus has 10^4 different one-error copies. If each of them would appear with equal frequency, a population consisting solely of mutants would yield the wild-type sequence at each position with an accuracy of 0.9999 despite the fact that this particular sequence as an individual may not be present at all. It has been shown by Weissmann [3] and co-workers through cloning and rapid amplification of single mutants that wild-type distributions consist predominantly of mutants, the wild type as an individual remaining below the limit of detectability. These mutants — even a diluted test tube fraction usually contains some 10^9–10^{12} such individuals — must be included if we ask *how* are population numbers related to selective values. Each mutant in the wild-type population, though inferior to the wild type, must be assigned a selective value too. Hence mutants are produced not only as error copies of the wild type but also through self-replication, and this will bias the distribution of mutants accord-

ing to their selective values. Value peaks will be clustered as mountains are in any landscape on earth. Evolutionary optimization then may be viewed as a hill-climbing process that proceeds to the mountain areas, eventually finding the highest peaks. How much this process is biased by the detailed structure of the mountain distribution came out as a true surprise from the rigorous treatment of the selection problem. In this chapter we shall start with the formulation and detailed analysis of a particular deterministic model, the quasi-species [4–7].

II. DETERMINISTIC APPROACH TO SELECTION

In conventional chemical kinetics, time changes of concentrations are described deterministically by differential equations. Strictly, this approach applies to infinite populations only. It is justified, nevertheless, for most chemical systems of finite population size since uncertainties are limited according to some $\sqrt{N}$ law, where N is the number of molecules involved. In a typical experiment in chemical kinetics N is in the range of 10^{18} or larger, and hence fluctuations are hardly detectable. Moreover, ordinary chemical reactions involve but a few molecular species, each of which is present in a very large number of copies. The converse situation is the rule in molecular evolution: the numbers of different polynucleotide sequences that may be interconverted through replication and mutation exceed by far the number of molecules present in any experiment or even the total number of molecules available on earth or in the entire universe. Hence the applicability of conventional kinetics to problems of evolution is a subtle question that has to be considered carefully wherever a deterministic approach is used. We postpone this discussion and study those aspects for which the description by differential equations can be well justified.

1. The Sequence Space

We consider a set of sequences of uniform length comprising ν monomeric subunits of which κ classes exist. If the model is applied to single-stranded RNA or double-stranded DNA sequences — as throughout this chapter — κ assumes the value 2 or 4, depending on whether only purines (R = G, A) and pyrimidines (Y = C, U, or T) or the four individual bases (G, A, C, U, or T, respectively) are specified. The total number of different sequences in such a set amounts to κ^ν. To give some examples: a particular tRNA ($\nu = 76$) represents one choice out of 10^{46}, a particular ribosomal 5S mRNA ($\nu = 120$) one out of 10^{72}, and a particular viral genome such as $Q\beta$ ($\nu = 4200$) one out of 10^{2529} alternative sequences of the given length. We specify now two sequences I_i and I_k in such a set. The Hamming distance $d(i, k)$ counts how

many positions in these sequences are occupied by different monomers (symbols: R and Y or G, A, C and U/T). For any given reference sequence I_m (which later will usually be the master sequence of the wild type) the Hamming distance $d = d(m, k) \in (1, 2, \ldots, v)$ partitions the mutants into classes, each comprising N_d different sequences [cf. eqn. (A1.1) in Appendix 1]. A correct ordering of mutants according to their mutual Hamming distances requires a v-dimensional space, in which each dimension consists of κ equivalent points (Figure 1). *Equivalent* here means that one can jump to any of the κ points within a given dimension according to a one-error mutation. (It does not necessarily mean that all such jumps occur with equal probabilities.) In nucleic acids, transitions, which are mutations that conserve the base class R or Y, respectively, indeed do occur more frequently than transversions, which are mutations that change those classes. Considering only transversions leads to the simpler limiting case of binary sequences ($\kappa = 2$).

An important feature of this high-dimensional sequence space, as is obvious from Figure 1, is the enormous increase in the number of shortest mutational routes between two given mutant sequences with increasing Hamming distance d. There are $d!$ such connecting 2^d states. Another related effect of high dimensionality is that many states become confined to a close neighborhood. In other words all distances among such an enormous number of states remain small so that a target state can be reached in relatively few steps provided a guiding gradient exists. Moreover, all states are connected through many alternative routes, which may pass through saddle points of higher order, that is, points where the selective value may increase in k and decrease in $v - k$ directions. In fact, each point is linked to v neighboring points, thereby yielding a multiply looped network. The value function, or landscape, of such a space looks quite different from what we are used to with the topographical landscape over the two-dimensional surface of earth. Such a landscape would appear very bizarre, typically showing drastic changes on short distances. This will be of importance if we look at the (selective) value and population landscapes over such a sequence space.

Now consider the sequences to be self-reproductive. Mutants will appear through copying errors. We introduce a fidelity q_i as the probability that a symbol i is copied correctly ($0 < q_i < 1$). Correspondingly, the probability of producing an error at that position is $1 - q_i$. The probability of copying a complete sequence I correctly then is $q_1 \cdot q_2 \cdots q_v = \bar{q}^v$, where $\bar{q}$ is the geometric mean taken over the fidelities for all positions of the sequence and $\bar{q}^v$ defines a quality of sequence copying and is identical with the Q_d value for $d = 0$, as listed in Appendix 1.

In order to characterize the average mutational behavior, we now assume uniform fidelities q for all positions and obtain for the probabilities of
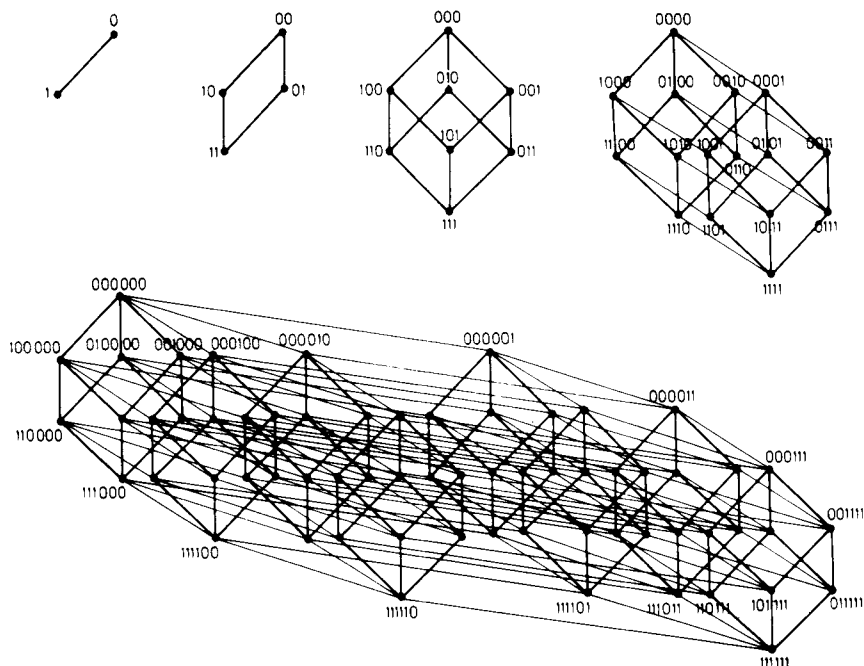
Figure 1. "Lexicographic" ordering of sequences through successive duplications of sequence space. As shown, for binary sequences the sequence space of dimension n can be constructed by duplication of sequence space of dimension $n - 1$. This iterative procedure is used in Appendix 2 to construct mutation matrix in such a way that eigenvalues and eigenvectors can be computed easily [8]. Each of 2^ν points specifies binary (R, Y) sequence. If, in addition, two alternative base classes ($R = G$ or A, $Y = C$ or U) are specified, then to each of points in binary sequence space another subspace of binary specification is added, yielding total of 4^ν points or dimension of hypercube of 2ν.

producing mutants with Hamming distances d the expressions in Appendix 1. (By this procedure we neglect, of course, inhomogeneities such as hot or cold spots.) Note that the probability of producing a particular mutant I_i from a reference copy I_k strongly decreases with increasing Hamming distance $d(i, k)$ [Eqn. (A1.5) in Appendix 1)] since q is usually close to 1.

The mutation frequencies between all pairs (I_i, I_k) of the κ^ν sequences of chain length ν are best given a matrix representation

$$Q = \{ Q_{ik}; i, k = 1, 2, \ldots, \kappa^\nu \}$$

Q being called the *mutation matrix*.

For the uniform error model it is suggestive to order the individual elements according to classes defined by their Hamming distances. Rumschitzki [8], however, has proposed a different procedure, which constructs the mutation matrix for chain length v from that for $v - 1$ recursively (cf. Appendix 2). In this way it is possible to calculate the eigenvalues and eigenvectors of Q for arbitrary chain lengths v, which in this particular model are all real and positive.

Physically, the elements Q_{ik} are associated with the production rates of the mutants (i.e., producing I_i by miscopying I_k). They do not resemble the frequencies by which these mutants appear in the population. First, these frequencies depend also strongly on the value topology of the mutant space, that is, on the individual efficiencies of reproduction. Second, in order to arrive at population numbers, we have to solve the kinetic equations.

2. The Kinetic Equations

The model is based on the following assumptions [9]:

(i) Sequences as defined in Section II.1 form and decompose steadily. Any individual type I_i of sequence is present with $n_i(t)$ copies per unit volume. This concentration may vary with time t, the rate being $\dot{n}_i(t) \equiv dn_i(t)/dt$.

(ii) Sequences form exclusively through either faithful or erroneous copying of sequences already present. The rates are first order in the concentrations of those sequences that act as templates.

(iii) The substrates of the formation reaction, that is, the energy-rich monomers from which new sequences are to be assembled, are assumed to be present in large excess and hence are buffered. Their constant concentration terms then can be included in the rate coefficients of the formation reactions. (This assumption may be relaxed later on if systems under constant-flow conditions are considered. This assumption turned out not to be decisive for the main behavior of the model.)

(iv) Sequences decompose according to a first-order rate law. The corresponding constant half-lives refer to a constant environment.

Conditions (ii)–(iv) allow the definition of two types of rate terms of a linear system: a diagonal term $W_{ii}n_i(t)$ and a nondiagonal term $W_{ik}n_k(t)$, referring to a value matrix $W = \{ W_{ik}; i, k = 1, 2, \ldots, \kappa^v \}$. The diagonal term comprises the effects of faithful replication and of decomposition, both being proportional to $n_i(t)$: $W_{ii} = A_iQ_{ii} - D_i$. Here, A_i describes autocatalytic amplification, that is, replication catalyzed by template I_i, of which only the fraction Q_{ii} (cf. Section III.1) leads to replica that are identical with the

template i. The decomposition term D_i counts the average number of decays of template i per unit time. The inverse of D_i is the average lifetime of i (cf. Figure 2).

The nondiagonal terms refer to the fraction of erroneous copying processes for which the conservation relation must hold

$$\sum_l A_l(1 - Q_{ll})n_l(t) = \sum_l \sum_k W_{lk} n_k(t). \tag{II.1}$$
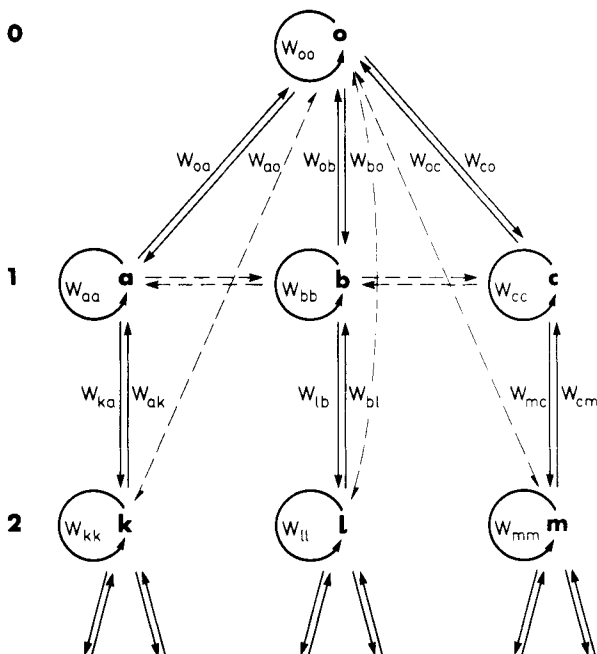
Figure 2. Reaction scheme of quasi-species. Index zero is assigned to the master copy, W_{00} being rate coefficient for correct (excess) production of master copy. Likewise W_{ii} refers to correct (excess) production of mutant copy i. Off-diagonal rate coefficients refer to mutation rates. W_{ij} represents the production of mutant i by miscopying template j. Rate coefficients are always associated with relative concentration terms x_j that refer to second of two subscripts (in agreement with conventional matrix notation). Hamming distances in this scheme are relative to master copy. It is important to note that diagonal coefficients are by an order of magnitude $(k - 1)/(q^{-1} - 1)$ larger than off-diagonal coefficients that refer to one-error miscopying, and that these in turn are correspondingly larger than coefficients for two-error miscopying, and so on. Error matrix may be ordered recursively in way similar to that demonstrated for buildup of sequence space (Figure 1) such that the antidiagonal represents mutation terms with extreme errors. Following this procedure explicit solutions could be obtained for certain rate coefficient schemes in uniform error model [8].

This conservation is a consequence of assumption (ii), namely, that mutants originate exclusively through erroneous replication and not through external interferences such as radiation or chemical attack. (If this assumption is relaxed, destruction terms must be subtracted in the conservation law to balance the additional first-order off-diagonal mutation terms.) The non-diagonal elements of the value matrix W_{ik} depend strongly on the Hamming distance $d(i, k)$ between template i and erroneous replica k. For the uniform error rate model the expression reads

$$W_{ik} = \left(\frac{q^{-1} - 1}{\kappa - 1} \right)^{d(i, k)} A_k Q_{kk} \equiv \varepsilon^{d(i, k)} A_k Q_{kk}. \tag{II.2}$$

The conservation relation (II.1) allows for a simplification of the summed rate terms, in terms of the total excess production rates E_k, such that the quality factors Q_{ii} no longer appear: as average excess production $\bar{E}(t)$ we define $\sum E_k n_k(t) / \sum n_k(t)$.

In addition to changes due to chemical reactions we must account also for changes caused by transport processes. Since we are not interested in spatially inhomogeneous distributions (cf. the example of a stirred flow reactor), we introduce a general dilution flux term $\Phi = \bar{E}(t)$ that removes material in proportion to the amount produced.

With these assumptions we can write the kinetic ansatz as shown in Appendix 3. Selection represents a kinetic evaluation of sequences relative to one another. It is therefore appropriate to introduce relative population variables

$$x_i(t) = n_i(t) \bigg/ \sum_k n_k(t). \tag{II.3}$$

As is seen from Appendix 2, the flux terms, in the form introduced in the preceding, do not appear in the rate equations referring to relative population variables. Instead the average excess production enters as a threshold of selection. This form of rate equations is not limited to a steady state but rather holds also for the relative growth behavior in time-variable systems [4].

3. How Realistic Is the Kinetic Ansatz?

Replication is a multiple-step reaction that usually involves sophisticated enzymic machinery. Is such a process adequately described by a straightforward linear autocatalytic model?

The kinetics of RNA replication by virus replicases have been studied recently in some more detail [10–12]. Rates have been measured under a variety of conditions. These include variable concentrations of substrate, enzyme, and template; asymmetries of plus and minus strand concentrations; annealing processes; and the presence of competitor strands. The mechanisms have been elucidated analytically and reproduced through computer simulation. Although the reaction mechanism (Figure 3) turned out to be quite involved, the phenomenological description is largely in agreement with the linear autocatalytic ansatz used in this chapter. This holds generally for low concentrations of template where the linear dependence of rates on template concentrations applies throughout. At larger template concentrations saturation effects typical for enzymic reactions are observed. The upgrowth of new templates, even if the enzyme is saturated by other templates (selected in previous steps), is exponential until the new template has become dominant, that is, until it is selected. Neither the fact that the replication process involves many consecutive steps, of which one or a few may or may not be rate limiting, nor the cross-catalytic nature of the plus–minus strand instruction
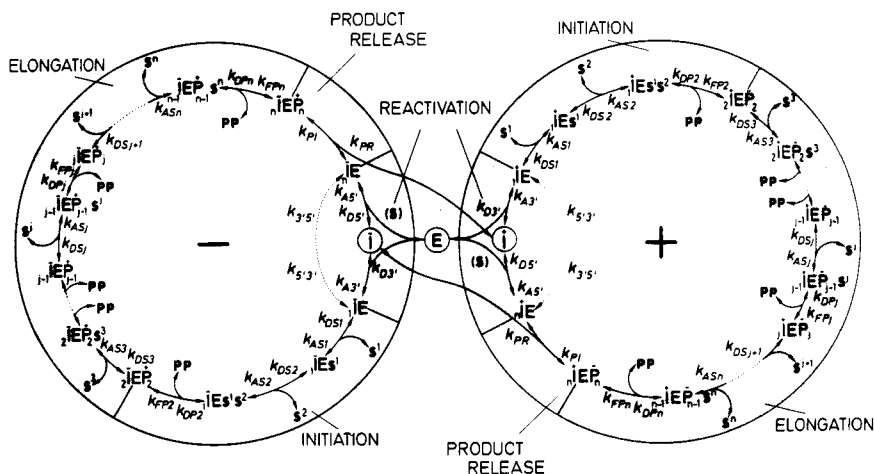
Figure 3. Reaction scheme of complementary replication of single-stranded RNA. Reaction consists of four phases: initiation, elongation, product release, and template reactivation. Reaction product (replica) is complementary to template. Substrates are four nucleoside triphosphates: ATP, GTP, UTP, and CTP. Pyrophosphate (pp) is waste product at each step of incorporation. Symbols: I, RNA template chain; E, enzyme (replicase); P, growing RNA replica chain. Indexes: A, association; D, dissociaton; S, substrate; F, phosphate diester bond formation; PR, product release; the numbers $3'$, or $5'$, refer to end of the RNA chain to which the enzyme binds or from which it dissociates (cf. ref. 10).

alters the overall dependence of the growth rate on template concentration from the model used here. All it does is to superimpose a spectrum of relaxation processes during an induction period of the reaction. During such an induction period the concentrations of the various intermediate states and also their sums for each of the plus and the minus strands assume fixed ratios. All states then grow concomitantly according to the linear autocatalytic rate law. This was verified with natural and artificial RNA sequences. It holds as well for complete virus particles.

DNA replication, although a process that usually depends on much more sophisticated enzymic machinery [13], phenomenologically adheres to the same type of model. The various steps involved contribute to the magnitude of the overall rate, yielding under suitable reaction conditions a defined doubling time of the double-stranded molecule. Since cell division is linked with this replication mechanism, the linear autocatalytic mechanism even holds for autonomous organisms and thus found its way into population biology.

There are, however, reproduction mechanisms that cannot be represented by the linear model. We have mentioned the case of saturation, which in many ways can be materialized, for example, through limited supply, inhibitory interference, or blocking of machinery. The loss of autocatalytic amplification is accompanied by a (partial or total) loss of competition and selection, allowing different sequences to coexist. On the other hand, there are reproductive processes the rates of which depend stronger than linearly on the growing concentration. A virus, for instance, encodes important parts of its reproduction machinery, which is exclusively adapted to the virus template. Hence in the early phase of infection the buildup of the virus population depends on both the virus template and the virus replicase, which is encoded in the template. The growth rate thus depends stronger than linearly on the virus template concentration. We have classified a whole category of such hypercyclic reaction networks that under certain conditions may exhibit hyperbolic rather than exponential growth. The effect is a once-and-for-all type of selection of that compound which managed first to populate the reaction space [4].

Apart from the linear autocatalytic ansatz, the rate equation involves some further assumptions the relevance of which should be discussed. First, the neglect of production terms other than those due to template-instructed reproduction seems straightforward. Whenever template instruction can become effective, it soon will outgrow any other type of production as far as the formation of specific sequences is concerned. Second, the assumption of a buffered level of substrates may seem unnatural. We have studied the effect of exhaustion of substrates by the upgrowth of a more efficiently reproducing mutant in a medium for which the influx of substrate is kept constant. The

more efficient consumption of substrate causes a lowering of its concentration level. Selection under such conditions of constant flows remains qualitatively the same but requires quantitative modifications of the equations involved [14]. The same is true for the assumptions made with respect to fluxes. The "proportional" dilution is most easily realized under experimental conditions. Under this condition the flux terms do not influence at all the relative population number and hence need not be specified further.

4. Solutions of the Rate Equations

In section II.2 we arrived at the following form of the rate equations in relative (i.e., fractional) population variables:

$$\dot{x}_i(t) = [W_{ii} - \bar{E}(t)]x_i(t) + \sum_{k \neq i} W_{ik}x_k(t). \tag{II.4}$$

Due to the fact that the term $\bar{E}(t)$ is inherently nonlinear, these equations have been solved by approximations corresponding to second-order perturbation theory [9]. The solutions reflect the threshold nature of selection and the consequences on length limitations for sequences that can be selected and resist an accumulation of errors (i.e., an error catastrophy).

Exact solutions can be derived for these equations [15, 16]. We follow here essentially the method of Jones et al. [16].

The nonlinear term $\bar{E}(t)$ can be removed through the following substitution:

$$x_i(t) = z_i(t)f(t) \quad \text{with} \quad f(t) = \exp\left(\int_0^t \bar{E}(\tau)\,d\tau\right), \quad i = 1, 2, \ldots, \kappa^\nu, \tag{II.5}$$

yielding the differential equations in the coordinates $z(t)$:

$$\dot{z}_i(t) = \sum_k W_{ik}z_k(t), \tag{II.6}$$

where the sum now includes the diagonal terms too.

We realize from Eqn. (II.4) that with $\sum x_k(t) = 1$, $f(t)$ can be expressed as

$$f(t) = \left(\sum_k z_k(t)\right)^{-1}. \tag{II.7}$$

The solution of Eqn. (II.6) follows the standard procedure of linear algebra.

This procedure and the solutions obtained are presented in Appendix 4.

The solutions $z_i(t)$ have the form shown in Eqn. (A3.10) in Appendix 3. Using Eqs. (II.5) and (II.7), they can be transformed back to the $x_i(t)$ [cf. Eqn. (A4.12) in Appendix 4]. The nature of these solutions becomes more obvious if we express them through the variables $y_i(t)$ as introduced in Appendix 3. The differential equations written in these variables assume the form

$$\dot{y}_i(t) = [\lambda_i - \bar{\lambda}(t)] y_i(t), \tag{II.8}$$

where the mean eigenvalue equals the mean productivity: $\bar{\lambda}(t) = \Sigma_k \lambda_k y_k(t) = \Sigma_k E_k x_k(t) = \bar{E}(t)$. The variables $y_i(t)$ are directly related to the normal modes of the relative population variables $x_i(t)$. They are, like the $x_i(t)$, normalized to 1, that is, $\Sigma_k y_k(t) = 1$.

From the form of Eqn. (II.8) of the differential equations the nature of selection as a consequence of a self-organizing process becomes obvious. The target of selection is a quasi-species defined by the maximum or dominant eigenvalue $\lambda_{\max} \equiv \lambda_0$ and its attributed normal mode, the dominant eigenvector, related to the composite population variable y_0. It is characterized by exclusively positive components, and therefore the quasi-species is not the equivalent of any single sequence. Nevertheless, the quasi-species may be dominated by a single sequence; we call it the master copy. The transformation from x to y coordinates may be physically viewed in the following way: Instead of looking at n single sequences, we look at n different clans, that is, combinations of sequences in which kinship relations are taken into consideration. These clans compete for selection, and the one with the largest eigenvalue λ_0 will win the competition. Equation (II.7) shows that the average of all eigenvalues [being equal to the average net productivity $\bar{E}(t)$] acts as a threshold. All eigenvalues $\lambda_i < \bar{\lambda}(t)$ yield a negative sign for y and hence cause that particular combination to die out. Likewise all eigenvalues $\lambda_i > \bar{\lambda}(t)$ cause the corresponding populations to grow. Hence combinations with small eigenvalues disappear; those with large eigenvalues build up. What happens to the mean eigenvalue $\bar{\lambda}(t) = \bar{E}(t)$ during selection? Let us assume that we started with exclusively nonnegative variables $y_k(0) \geq 0$ and that all eigenvalues λ_k are real. Both conditions need not be fulfilled in realistic systems, but for the sake of coherence we postpone a discussion of the general case to the next section (see also Appendix 5). The increase of $\bar{\lambda}(t)$ will continue until it equals the maximum eigenvalue; the threshold increases until it reaches λ_0, where only one combination is left that can match it. We call this stable combination the quasi-species; it does not change with time any more: $\dot{y}_0(t) \to 0$. Hence selection is a self-organizing process inherently caused by the autocatalytic nature of the formation rates. It can be character-

ized — in analogy to other self-organizing processes, such as equilibration —
by an extremum principle:

$$\bar{E}(t) = \bar{\lambda}(t) \to \lambda_0. \tag{II.9}$$

This principle holds as an absolute principle only for the quasi-linear case
treated in the preceding, that is, for constant rate coefficients W_{ik}. As a local
principle it covers also a larger group of nonlinear systems involving concen-
tration- or time-dependent rate parameters.

Explicit expressions for the eigenvalues λ_i and for the components of the
eigenvectors belonging to λ_i are to be obtained from Eqs. (A4.5)–(A4.7) in
Appendix 4 for any specific set of rate parameters W_{ik}. If the nondiagonal
elements are sufficiently small as compared to the diagonal terms and if the
diagonal elements can be brought into some hierarchical order, where
$W_{mm} > W_{kk}$ for any k, second-order perturbation theory yields the following
approximations for the largest eigenvalue λ_0 and the corresponding eigen-
vector $\mathbf{l}_0$ (see also Section III.1),

$$\lambda_0 = W_{mm} + \sum_{k \neq m} \frac{W_{mk} W_{km}}{W_{mm} - W_{kk}}, \tag{II.10}$$

and the dominant eigenvector $\mathbf{l}_0$,

$$\frac{l_{k0}}{l_{m0}} = \frac{W_{km}}{W_{mm} - W_{kk}} + \sum_{j \neq k, m} \frac{W_{kj} W_{jm}}{(W_{mm} - W_{kk})(W_{mm} - W_{jj})}. \tag{II.11}$$

While in general these expressions are of only limited realistic value — since
many mutants indeed are (nearly) degenerate — they show very clearly one
fact. The surviving quasi-species for which $\lambda_i = \lambda_0$ is dominated by the
sequence with the largest diagonal rate coefficient W_{mm}. We call it the master
species. It is dominant as long as the sum term of Eqn. (II.10) is negligible,
that is, $W_{mm} > W_{kk}$ for any $k \neq m$. Since the products $W_{mk} W_{km}$ are usually
very small, the differences between master I_m and mutants I_k need not be
large. Taking the known mutation rates of RNA viruses and their genome
size of a few thousand nucleotides, an advantage of a tenth of a percent in
W_{mm} values is sufficient to clearly define the master. The mutants then are
grouped around the master in such way that often their average sequence
equals that of the master, which though being the most abundant individual
sequence in the distribution may be present as only a very minor fraction of
the total set of all mutants.

5. Potential Functions, Optimization, and Guided Evolution

The results presented in the preceding section suggest the need for a funda-mental reinterpretation of Darwinian behavior for self-replicating macromol-ecules. First, selection can be shown as a mere (physical) consequence of constrained self-replication. It is a characteristic of biological systems since there is no other way for living entities to come about than through self-reproduction or recombinative reproduction, both being based on replication of DNA. Conversely, however, selection is not inherently linked to life and may occur wherever self-reproduction or complementary-reproduction or autocatalysis in general is involved. One well-studied example from physics is selection in laser modes [17].

In order to visualize the molecular selection process in the more general context of optimization of replication rates, we consider the simple case of replication with ultimate accuracy first. In this case we have $Q_{ik} = \delta_{ik}$, the value matrix W is diagonal ($W_{ik} = W_{kk}\delta_{ik} = W_k = A_k - D_k$) and the corre-sponding system of differential equations is weakly coupled by the $\bar{E}(t)$ term only:

$$\dot{x}_k = x_k(W_k - \bar{E}), \quad k = 1, 2, \ldots, n. \tag{II.12}$$

The solutions of Eqs. (II.12) are of the form

$$x_k(t) = x_k(0)\exp\left(W_k t - \int_0^t \bar{E}(\tau)\,d\tau \right), \quad k = 1, 2, \ldots, n. \tag{II.13}$$

The system converges asymptotically to a homogeneous state in which the most efficiently replicating species is present exclusively:

$$\lim_{t \to \infty} x_0(t) = 1, \quad \lim_{t \to \infty} x_k(t) = 0 \; \forall k \neq m \quad \text{and} \quad W_0 = \max[\, W_j; j = 1, 2, \ldots, n].$$

We calculate the time dependence of the average excess production and find

$$\frac{d\bar{E}}{dt} = \frac{d}{dt}\left(\sum_k W_k x_k \right) = \sum_k x_k\left((W_k - \bar{E})^2 + \frac{dW_k}{dt} \right). \tag{II.14a}$$

In the case of time-independent rate constants the second term vanishes and

$$\frac{d\bar{E}}{dt} = \sum_k x_k(W_k - \bar{E})^2 \geq 0, \tag{II.14b}$$

which is a nonnegative expression since $x_k \geqslant 0$ and W_k, a combination of rate constants, is always real by definition. Thus, the average excess production $\bar{E}$ is a nondecreasing function of time. It represents a quantity that is optimized during selection.

A function $V(x)$, algebraically identical to the average excess production, can be shown to represent a potential for the selection process (Appendix 5):

$$V(x_1, x_2, \ldots, x_n) = \sum_{k=1}^{n} W_{kk} x_k. \qquad (II.15)$$

Now we can visualize evolutionary optimization as a "hill-climbing" process on a "landscape" that is given by an extremely simple potential [Eqn. (II.15)]. This potential, an $(n - 1)$-dimensional hyperplane in n-dimensional space, seems to be a trivial function at first glance. It is linear and hence has no maxima, minima, or saddle points. However, as with every chemical reaction, evolutionary optimization is confined to the cone of nonnegative concentration restricts the physically accessible domain of relative concentrations to the unit simplex $(x_1 > 0, x_2 > 0, \ldots, x_n > 0; \Sigma x_k = 1)$. The unit simplex intersects the $(n - 1)$-dimensional hyperplane of the potential on a simplex (a three-dimensional example is shown in Figure 4). Selection in the error-free scenario approaches a corner of this simplex, and the stationary state corresponds to a "corner equilibrium," as such an optimum on the intersection of a restricted domain with a potential surface is commonly called in theoretical economics.

Let us now consider replication with errors. For sufficiently accurate replication the system approaches a stationary mutant distribution. We do not observe selection of a single species. The target of the selection process is a unique combination of species determined by the dominant eigenvector of the value matrix W. We perform a linear transformation of variables and choose the eigenvectors of the matrix W as the new basis of the coordinate system:

$$\mathbf{x}(t) = (x_1(t), x_2(t), \ldots, x_n(t)) = \sum_{k=1}^{n} x_k(t) \cdot \mathbf{e}_k = \sum_{k=1}^{n} y_k(t) \cdot \mathbf{l}_k. \qquad (II.16)$$

Herein the vectors $\mathbf{e}_k$ are the unit eigenvectors of the Cartesian coordinate system and $\mathbf{l}_k$ the right-hand eigenvectors of the value matrix W as discussed in Appendix 4. Formally the transformed differential equation is of the same general type as (II.12):

$$\dot{y}_k = y_k(\lambda_k - \bar{E}), \quad k = 1, 2, \ldots, n,$$
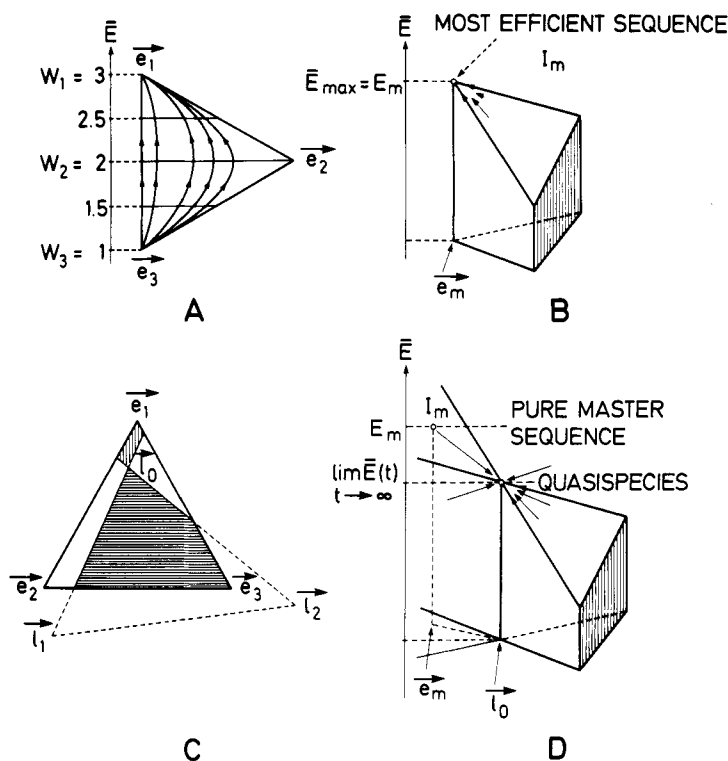
Figure 4. Potential functions and "corner equilibria" in evolutionary optimization. (A) Trajectories and constant level sets of potential V [Eqn. (II.15)] for error-free replication in system with three molecular species. Constants $W_1 = 3$, $W_2 = 2$, and $W_3 = 1$ were chosen. In general, trajectories do not cross constant level sets at right angles. Generalization of definition of gradients based on Riemannian metric allows one to rescale angles such that all trajectories cross constant level sets at right angles (Appendix 5). (B) Demonstrates meaning of "corner equilibrium". "Hill-climbing" process ends at corner of simplex since trajectories cannot cross boundaries. (C) Concentration simplex and triangle spanned by three eigenvectors of value matrix W for one particular choice of constants: $n = 3$, $A_1 = 10$, $A_2 = 3$, and $A_3 = 1$; $Q_{kk} = 0.6$ and $Q_{ik} = 0.2$ for $i \neq k$ and $i, k = 1, 2, 3$. Only dominant eigenvector l_0 lies inside concentration simplex. Entire domain of physically accessible values of relative concentrations is split into four zones. Two are characterized by monotonous behavior of average excess production: in the large, horizontally hatched area $E(t)$ is nondecreasing function. In top zone, hatched vertically, $E(t)$ is nonincreasing. No general predictions can be made for other two areas. (D) Behavior of average excess production $E(t)$ illustrated by means of potential V. Now corner equilibrium lies inside concentration simplex since invariant lines that are not to be crossed by trajectories are given by boundaries of triangle spanned by eigenvectors of value matrix W: l_0, l_1, and l_2.

and the same is true for the solutions,

$$y_k(t) = y_k(0)\exp\left(\lambda_k t - \int_0^t \bar{E}(\tau)\,d\tau \right), \quad k = 1, 2, \ldots, n,$$

and the time dependence of the average excess production,

$$\frac{d\bar{E}}{dt} = \sum_k y_k(\lambda_k - \bar{E})^2.$$

When it comes to details, however, there are substantial differences between the two differential equations and their solutions (see e.g. ref. 18):

1. The variables x_k are relative concentrations and are nonnegative by definition: $x_k \geq 0$. No such relation holds for the variables y_k.
2. The coefficients W_k are obtained from rate constants and mutation frequencies and hence represent real quantities. The λ_k are the eigenvalues of a nonsymmetric matrix and need not be real.

As a consequence of 1 and/or 2, the average excess production is no longer a nondecreasing function of time. Optimization of the average excess production may still occur, but then it is restricted to certain choices of initial conditions (Appendix 5). Jones [19] derived a more complicated function $\bar{\varepsilon}(t)$ shown in Appendix 5 that represents a universal optimization criterion in the replication–mutation system, but the physical meaning of this Lyapunov function is unclear.

The existence of complex eigenvalues of the value matrix W implies that the coefficients in Eqn. (II.15) are complex and rules out the existence of a real-valued potential function. Transient oscillations in the concentrations may occur, but in the limit of long times the system nevertheless converges toward the dominant eigenvector. The corresponding largest eigenvalue is real and positive, and hence all oscillations in concentrations have to fade out inevitably.

The fact that the average excess production may decrease during a selection process can be illustrated by means of a simple example. Consider a homogeneous population consisting exclusively of the master sequence I_m at time $t = 0$. Clearly the excess production is largest since the master sequence is characterized by the maximum selective value. During selection, mutants are formed that have smaller selective values, and finally when the system approaches the stationary distribution, which is the quasi-species, the average excess production reaches the stationary value from above, that is, from higher, total efficiency of replication.

For exclusively real eigenvalues of W the time dependence of the average excess production is determined by the choice of initial conditions. As shown in Appendix 5, optimization of $\bar{E}(t)$ is restricted to initial conditions in the positive orthant $[y_k(0) > 0; \ k = 0, 1, \ldots, n]$. These initial conditions are not difficult to fulfil, and they will apply to many cases in reality. We should keep in mind, nevertheless, that there are other choices of initial conditions, such as the start with a pure master sequence, for which the simple principle does not hold. For one particular type of choice, $y_1(0) > 1$ and $y_k(0) \leqslant 0$ for all $k \neq 1$, the average excess production decreases monotonically.

How likely is it to have a value matrix W with pairs of conjugate complex eigenvalues? Rumschitzki [8] showed that the value matrix W can be converted to a symmetric matrix W' by means of a similarity transformation provided the corresponding mutation matrix Q is symmetric ($Q_{ij} = Q_{ji}$). Then all eigenvalue of W are real. Equal frequency of mutations in both directions, $I_j \rightarrow I_i$ and $I_i \rightarrow I_j$, is a realistic assumption unless the polynucleotide sequences under consideration contain so-called hot spots. These are positions at which point mutations are particularly frequent. It is unlikely that the reverse mutation leading to the sequence with the original hot spot is also an unusually frequent event. Therefore we expect a mutation matrix Q lacking symmetry in these cases.

6. Population Structures

The fundamental reinterpretation of Darwinian behavior to which we referred in the preceding is a consequence of one central fact; namely, that not a single sequence but rather a master sequence and its entire quasi-species distribution appears to be the target of selection. Natural selection being the consequence of self-reproduction has been stressed already in population biology (mainly through the work of Fisher [20], Haldane [21], and Wright [22]). Hence an advantageous mutant appearing to any stochastically significant extent is bound to increase, making selection finally a deterministic event. On the other hand, the appearance of the particular mutant that proves to be selectively advantageous was assumed to be a stochastic event, having no target- or goal-directed bias. This view cannot be sustained any more as a result of the following argument.

The chance of finding an advantageous mutant will increase with increasing Hamming distance (i.e., the mutation distance from the wild type), the main reason being the large increase of the number of mutants and hence possible candidates with increasing distance. Whether a distant mutant can appear will depend on how its precursors are populated. If the precursors of a desired mutant are more populated than those of a nondesired one, there will be guidance of the evolutionary process along preferred routes. The theory

states that it is not the single master sequence representing the wild-type but rather the localized distribution in the sequence space that is the target of selection. Then it is very important to know *how* the various mutants are populated. The equations quoted in the preceding correlate population numbers with kinetic parameters called *selective values*. Our problem now has two aspects: How are selective values distributed in sequence space and how does the value distribution map into a population distribution.

It is immediately seen that guidance to optimal performance requires first that mutant population numbers critically depend on their own selective values (and not only on wild-type selective value) and that the value distribution, as with the altitude distribution on earth, is not entirely random but rather clustered along more or less cohesive routes. The kinetic theory, as previously presented, makes quantitative assertions about the correlations between selective values and population numbers.

According to Eqn. (A1.3) Q_d describes (for a uniform error model) the probability of producing a mutant with Hamming distance d in terms of the normalized error rate ε [Eqn. (A2.1)]. This probability decreases exponentially with increasing Hamming distance. The example of a virus with $v = 4200$ (discussed in Section V) yields an ε value of about 10^{-4}, hence a 10-error mutant would occur only with a probability of 10^{-40}. A mutant, once it has been produced by miscopying of the wild type or of some intermediate precursor, may also replicate itself, possibly with a rate not too much smaller than that of the master copy. Hence the probability of making a given error copy is not identical with the stationary rate of appearance of that copy.

The appearance of a mutant copy I_i of the master copy I_m in a stationary quasi-species can be calculated by the system of rate equations setting $x_i = 0$. In the realm of second-order perturbation theory the explicit expressions can be obtained through recursion. The procedure is best explained with the help of Figure 1. Starting with the master copy, we assume that error copies form only along downhill routes including jumps of any length but neglecting any looped routes (i.e., routes in which Hamming distances do not monotonically increase). This entails the neglect of two kinds of processes:

(i) consecutive change of symbols at a given position (e.g., $A \rightarrow G \rightarrow U$ instead of $A \rightarrow U$) and

(ii) reversion of a change at a given position.

Since any mutational process costs at least a factor ε, those looped routes are negligible unless mutants with increasing distance substantially increase in their population numbers (which would violate the validity of second-order perturbation expressions, for which W_{ii} must not be too close to W_{mm}).

The stationary solutions ($\dot{x} = 0$) then yield for the relative population numbers

$$\frac{\bar{x}_{di}}{\bar{x}_m} = \varepsilon^d \frac{W_{mm}}{W_{ii}} f_{di},$$ (II.17)

where ε is the quotient introduced with Eqn. (A2.1). The index d_i [abbreviated di in Eqn. (II.17)] refers to the ith sequence in the error class defined by Hamming distance d, and f_{di} is obtained through recursion:

$$f_{1i} = \frac{W_{ii}}{W_{mm} - W_{ii}},$$

$$f_{2i} = \frac{W_{ii}}{W_{mm} - W_{ii}} \left(1 + \sum_{j=1}^{\binom{2}{1}} f_{1j} \right),$$

$$f_{3i} = \frac{W_{ii}}{W_{mm} - W_{ii}} \left(1 + \sum_{j=1}^{\binom{2}{1}} f_{1j} + \sum_{j=1}^{\binom{3}{2}} f_{2j} \right),$$ (II.18)

$$\vdots$$

$$f_{ki} = \frac{W_{ii}}{W_{mm} - W_{ii}} \left(1 + \sum_{j=1}^{\binom{k}{1}} f_{1j} + \ldots + \sum_{j=1}^{\binom{k}{k-1}} f_{k-1,j} \right).$$
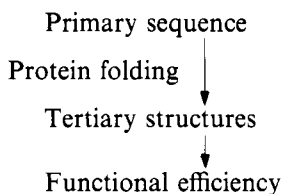
In these expressions W_{ii} always refers to the individual I_i in the corresponding error class under consideration, while the j's in the sums refer to the corresponding precursors of I_i, for example, in f_{1i} to all d one-error precursors or in $f_{d-1,i}$ to all d of the $(d-1)$-error precursors of I_i. Note that due to the iterative nature, the first sum term in f_{di} contains $d!$ d-fold products of $W_{kk}/(W_{mm} - W_{kk})$ terms.

In this approximation neutral mutants (i.e., $W_{ii} = W_{mm}$) and hence singularities in the f_i terms are precluded. Most mutants will have selective values W_{ii} that are small compared to W_{mm} or that may even be zero (nonviable mutants). If all mutants in fact were of this kind (i.e., $W_{ii} \ll W_{mm}$), the mutant appearance would resemble the Poissonian-type distribution, yielding for any individual in error class d the probability ε^d. Expression (II.18) comprises all contributions of mutant states (classes d' ranging from 1 to v, including stepwise mutations as well as any jumps up to length d (occurring in direction $0 \rightarrow d$). Contributions from reverse mutations are neglected in this approximation. They could occur if truly neutral or nearly neutral mutants were present that reversibly are populated with numbers that increase with increasing d. This approximation still allows for quite large terms $W_{ii}/(W_{mm} - W_{ii})$ extending to several orders of magnitude. In fact, all that is required is

that these terms do not become as large as ε^{-1}. Then the x_{di} decrease with increasing Hamming distance d, yet much less drastically than ε^d does (cf. Figure 5 and the more detailed discussion in Section IV.4). In the previously mentioned example of a virus, where for the 10-error mutant, ε^d drops to 10^{-40}, such an individual mutant, if it were in a domain of precursors with W_{ii} values within 1— of W_{mm}, would be raised to a realistic probability of appearance, in contrast to corresponding mutants in low-value domains. This requires not only singular (nearly) neutral mutants in the distribution but also a clustering of such mutants in more or less cohesive domains, as in landscapes on earth where planar and mountainous regions are also clustered.

On the other hand, a comparison with landscapes on earth may be misleading since the sequence space is of high dimension. All distances remain relatively small, while, as Figure 1 demonstrates, all states are highly interconnected. A mountain trip in such a space may proceed along ridges that are multiply interconnected and thereby link up the peaks. Those ridges and peaks are the points in sequence space that are most heavily populated by mutants, which thereby provide guidance to the evolutionary process. Here we come back to the other still unanswered question: how are values distributed in sequence space? If this distribution were entirely random, it could not guide the evolutionary process over any substantial length of path.

Mandelbrot [23] has shown that the "most random" type of height distribution to be expected on earth is of a fractal type. The same should be true for a value distribution in the ν-dimensional sequence space. Such a fractal distribution is highly connective, that is, anything but uncorrelated. Moreover, we know that functional efficiency is clustered around certain sequences. The functional efficiency of an enzyme depends on the correct spatial arrangement of certain amino acid residues that comprise the active center: this is achieved by three-dimensional folding of the polypeptide chain [24]. Hence there exists a correlation:

$$
\begin{array}{c}
\text{Primary sequence} \\
\text{Protein folding} \quad \big\downarrow \\
\text{Tertiary structures} \\
\downarrow \\
\text{Functional efficiency}
\end{array}
$$

Similarities in primary sequences in general will map into similarities of folding, which in turn will lead to similarities in tertiary structures and functional efficiencies.
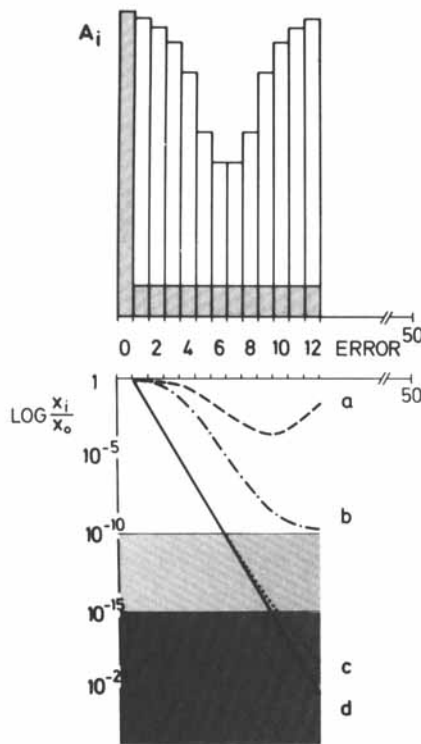
Figure 5. Importance of mutant bias in evolution. Simulation demonstrates role of ridges in selection value landscape in directing route of evolution. Binary sequence of complexity 50 is assumed. In upper part of picture, two value profiles are shown as functions of Hamming distances from reference master 0. Selective value maximum connected to particular mutant at d = 13 (excluded from stationary state calculation) by three different landscapes: (i) lowland where all intermediate mutants have inferiority of 0; (ii) plateau where all mutants have inferiority of 0.2 (open bars); and (iii) smooth mountain ridge with monotonous decrease of inferiority to 0.5: at d = $\{6, 7\}$ followed by rise to 0.98 at $d = 12$ (filled bars). Adding sequence at $d = 13$ that is neutral with master (inferiority 1.0) would produce symmetric distribution between two maxima. For each profile two limiting scenarios have been calculated, and results for mutant populations plotted beneath. Curves b and d: elevated inferiorities (W_{ii}/W_{00}) are found for only one particular route comprising 12 subsequent defined mutation steps. All other mutants have negligible inferiority values. Curves a and c: All mutants with same Hamming distance to nearest maximum have degenerate inferiority values (W_{dd}/W_{00}); i.e., all direct routes are possible because succession of mutation steps is arbitrary. (Consult Figure 1 in order to correlate with routes. Note that routes for $d = 12$ comprise only minute fraction of possible connections in 50-dimensional space, i.e., $2^{12}/2^{50} = 10^{-15}$.) Solid curve d coincides with relative mutant populations according to Poissonian distribution ($W_{dd}/W_{00} = 0$). Curves c, d and a, b refer to inferiority value distributions of plateau and smooth ridge profiles, respectively. Following features are seen: Relative populations, around a peak in selection value landscape that rises abruptly from lowland, drop rapidly with increasing Hamming distance from master. Probabil-

Evidence for such a similarity transformation has been provided by experiments on site-directed mutagenesis. Changes in the primary sequence of enzyme molecules produced through site-specific genetic manipulation in many cases turned out to be of only minor consequence to the catalytic performance. Those changes, of course, must not occur at certain strategic positions that are determinants for the structure of the active site. Otherwise, there are many positions that are relatively insensitive to substitution, especially by related amino acids. The same conclusion can be reached by considering phylogenetic sequence relations. A well-studied case is the enzyme cytochrome-c. Sequences in different species may be dishomologous in more than 70% of the positions, and yet the same reaction is catalyzed (almost) equally well. Hence there is no doubt that a similarity relation corresponding to a clustered-value distribution in sequence space does exist. However, the relation itself is anything but simple, and no straightforward method of calculating those relations seems to be at hand. For the quasi-species model the existence of a clustered-value distribution as such is of paramount importance. Since the theory shows how value distributions map into population distributions, a quantitative analysis of mutant distributions in selection experiments could provide the evidence and allow for quantitatively estimating the effect of mutational guidance in evolution.

III.   ERROR THRESHOLD FOR QUASI-SPECIES LOCALIZATION

The organizing principle embodied in the kinetic equations presented in Section II is given a global character by the appearance of the off-diagonal mutational terms, which provide the variation necessary to the Darwinian logic. Thus, under the physical conditions, where every sequence may be reached by a succession of mutations, the model describes an evolutionary process capable of finding the fittest from arbitrary initial conditions and of returning to it after perturbation. The special feature of the organization

---

ity of reaching another peak with Hamming distance of 12 is very low for typical laboratory populations. In case where sharp crest with low selection values is attached to peak, probability of finding 12-error mutant is only slightly higher. If peak is surrounded by low plateau (between $d = 0$ and $d = 12$), this probability rises by nearly seven orders of magnitude. Effect of higher selective values is drastic: Just one sharp crest with profile depicted above brings probability of occurrence of 12-error mutant up to such order of magnitude that population necessary to find one strand can be realized readily in laboratory. If all routes leading away from summit have same profile, large multiplicity of 12-error mutants will occur (deterministically) in laboratory population. Bias of selective values guiding evolutionary path along ridges in selective value landscape is clearly demonstrated. Note also cumulative effect of connected high-value regions in v-dimensional landscape.

allowing the selection of a particular sequence type is the relative weakness of mutational interactions. The nonequilibrium Darwinian organizing principle has a parallel in the existence of particular ordered states at equilibrium, such as the low-temperature phases of ferro- and antiferromagnets and ferro- and antiferroelectrics or the condensed states of nucleoside bases in stacks. There the temperature represents the equivalent of the mutation rate. At certain critical temperatures we observe order–disorder equilibrium transitions. Likewise, there is a threshold for selection of a particular wild-type sequence in the nonequilibrium Darwinian organizing principle.

In this section we demonstrate the existence of a sharp error threshold in the kinetic model developed in the preceding and explore its dependence on the value topology of Darwinian fitnesses W_i and the mean fidelity $q = \bar{q}$ of single-digit replication. The error threshold result delineates the domain where evolution proceeds to the survival of a stably localized quasi-species. With higher error rates, the class of sequences distinguished as fittest, by the stationary solution of the deterministic kinetic equations, becomes so large that it cannot be sampled by any biological population. A stochastic interpretation of the dynamics as a random drift covering this class is then called for. The character of the present discussion is then intermediate between a deterministic and stochastic account, and although we shall refer to statistical distributions of replication rates, the dynamics will be that of the deterministic kinetics.

1. Error Threshold and Selective Advantage

For a particular sequence I_m to compete successfully with all the other mutants, it has to be produced at a (net) rate W_{mm} faster than the excess production rate $\bar{E}_{k \neq m}$ of these other mutants. The latter rate is enhanced by the inclusion of inexact replication (leading mostly to further mutants and only rarely back to the wild type). Thus, not only does a sequence I_m generally need to have the maximum net rate W_m for exact replication in order to be conserved indefinitely in a population, but it also needs to satisfy the more stringent requirement

$$W_{mm} > \bar{E}_{k \neq m} = \frac{\sum\limits_{k \neq m} E_k \bar{x}_k}{\sum\limits_{k \neq m} \bar{x}_k}, \tag{III.1}$$

where the $\bar{x}_k$ are the stationary relative population numbers.

Indeed, the neglect of backflow of mutants to the wild type, that is, identification of W_{mm} with the diagonal rate coefficient of the master se-

quence, is equivalent to a second-order perturbation result, and to this order a calculation of the wild-type population yields

$$\frac{\bar{x}_m}{\sum_k \bar{x}_k} = \frac{W_{mm} - \bar{E}_{k \neq m}}{E_m - \bar{E}_{k \neq m}},$$ (III.2)

where $E_m = A_m - D_m$ is the excess productivity of the wild type.

According to Eqn. (III.2), the relative equilibrium concentration of the wild type $\bar{x}_m/\sum \bar{x}_k = \bar{x}_m/c_0$ vanishes for some critical error rate at which $W_{mm} = \bar{E}_{k \neq m}$. This, of course, is merely a consequence of the perturbation approach that neglects backflow of mutations to the wild type. The phenomenological calculation of $\bar{x}_m$, based on the use of $\bar{E}_{k \neq m}$, can be carried out exactly if we introduce a mean backflow rate:

$$W_M = \frac{\sum_{k \neq m} W_{mk} \bar{x}_k}{\sum_{k \neq m} \bar{x}_k} = \frac{1}{c_0 - \bar{x}_m} \sum_{k \neq m} W_{mk} \bar{x}_k.$$

Then we obtain the relative equilibrium concentration of the wild type as the positive root of a quadratic equation:

$$\frac{\bar{x}_m}{\sum_k \bar{x}_k} = \tfrac{1}{2} \left(\mathscr{A} - \mathscr{B} + \sqrt{(\mathscr{A} - \mathscr{B})^2 + 4\mathscr{B}} \right)$$ (III.3)

with

$$\mathscr{A} = \frac{W_{mm} - \bar{E}_{k \neq m}}{E_m - \bar{E}_{k \neq m}} \quad \text{and} \quad \mathscr{B} = \frac{W_M}{E_m - \bar{E}_{k \neq m}}.$$

It is readily seen that $\bar{x}_m$ does not vanish at the critical point $\mathscr{A} = 0$, or $W_{mm} = \bar{E}_{k \neq m}$, but is given by

$$\frac{\bar{x}_m}{\sum_k \bar{x}_k} = \frac{\mathscr{B}}{2} \left(\sqrt{1 + \frac{4}{\mathscr{B}}} - 1 \right) \quad \text{for} \quad \mathscr{A} = 0,$$

which is small, on the order of

$$\sqrt{\mathscr{B}} = \sqrt{\frac{W_M}{E_m - \bar{E}_{k \neq m}}}.$$

The exact equation (III.3) may be solved if the phenomenological parameters $\mathscr{A}$ and $\mathscr{B}$ can be computed. For the special test case of a single advantageous master with homogeneously poorer competitors, $\mathscr{A}$ may be computed exactly and $\mathscr{B}$ computed as a function of $\bar{x}_m/\Sigma_k\bar{x}_k$ under the assumption that all other mutants are equally populated. The solutions for $\bar{x}_m$ given by (i) second-order perturbation theory Eqn. (III.2), (ii) the phenomenological Eqn. (III.3) with the above assumption for $\mathscr{B}$, and (iii) the exact solution are presented in Figure 6 for sequence lengths 7 and 20. For large sequence lengths the results are not significantly different from the solution of Eqn. (III.3) with $\mathscr{B} = \varepsilon(1 - \bar{x}_m/\Sigma_k\bar{x}_k)$, $\varepsilon \to 0$ (two straight lines).

Equation (III.2) may be rewritten to isolate the dependence on the copying fidelity q in order to demonstrate that for a given set of replication parameters there is an error-rate-dependent threshold sequence length for quasi-species instability. To this end the selective advantage or superiority parameter σ was introduced:

$$\sigma = \frac{A_m}{D_m + \bar{E}_{k \neq m}}, \tag{III.4}$$
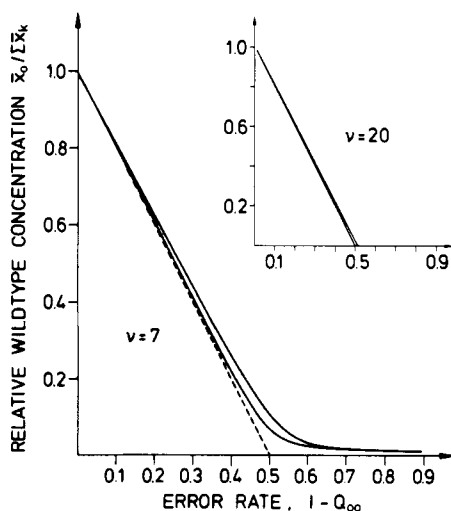
Figure 6. Relative wild-type concentrations as function of error rate $1 - Q_{00}$. We show exact solution curve (upper full line) and compare it with result of perturbation theory [Eqn. (III.2); broken line) and exact solution of Eqn. (III.3) with B calculated as in text (lower full line). Following parameters and rate constants applied: $A_0 = 10$, $\bar{A}_{k \neq 0} = 5$, $D_0 = D_1 = \cdots = D_n$ and $v = 7$, $n = 2^v - 1$.

which is not explicitly dependent on q and in terms of which the inequality (III.1) becomes $Q_m > \sigma^{-1}$, or using Eqs. (A1.1–A1.4) in Appendix 1 to relate the quality factor to the sequence length:

$$v < v_{max} \doteq -\frac{\ln \sigma}{\ln q} \simeq \frac{\ln \sigma}{1 - q}. \tag{III.5}$$

This relationship shows that the amount of information that can be maintained in the form of a specific stable sequence is limited by the copying fidelity q. Thus the price that is paid for the global nature of the organization via variation is a limitation on the discrimination of fitness. Typical variations in fitness can only be selected for up to the critical sequence length v_{max} since the fidelity of copying the whole sequence decreases exponentially with increasing sequence length.

It should be briefly pointed out that simply decreasing the replication rates of other sequences to increase the superiority of the master does not indefinitely postpone the error threshold. Once the replication rates for mutant sequences fall below the destruction rate for the master sequence, the domain of viable existence for the quasi-species is limited by $A_m Q_m > D_m$. In Section III.3 a more fundamental statistical argument restricts the effective superiority to a finite value of order 1.

The dependence of the threshold on the replication rate parameters A_k, D_k expressed through Eqn. (III.1) or (III.4) is dependent on the populations of each of the mutants that appear to weight the average excess productivity $\bar{E}_{k \neq m}$ [cf. Eqn. (III.1)]. In particular, mutants that are distant from the wild type may require a net exact replication rate much closer to that of the wild type than does a nearby mutant in order to reach a substantial fraction in the population and increase $\bar{E}_{k \neq m}$ above W_{mm}. On the other hand, a more detailed analysis of the error threshold in terms of the many rate parameters characterizing the different mutant sequences appears unwieldy, as it requires full solution of an extremely high-dimensional eigenvalue problem.

One possible option is to adopt a statistical description of the kinetic parameters and to ask how likely it is for the quasi-species to be localized about the wild type. This undertaking requires an analysis beyond the second order in perturbation theory since a distant mutant with a selective value very close to that of the wild type may jeopardize the stability of the latter in the population. We were however encouraged by the progress that had been made with a problem of similar difficulty in the very different area of electron or spin localization in disordered solids. Indeed, it turns out that an expression of the form of Eqn. (III.5) may be obtained, with an explicit expression for the superiority parameter σ_0, dependent on the distribution of replication rates but not on any average involving population variables.

More exciting than this, the selective advantage of the wild type turned out to be nonzero even with a continuous distribution of selective values (for the different mutants) up to that of the wild type. These developments are reviewed in the following section.

2. Localization Threshold for Statistically Distributed Replication Rates

We consider once again the deterministic kinetic Eqn. (II.4) with the uniform error rate expression Eqn. (A1.5) relating mutation rates to replication rates:

$$W_{ik} = \left(\frac{q^{-1} - 1}{\kappa - 1} \right)^{d(i, k)} (W_k + D_k) = V_{ik} W_k. \qquad (III.6)$$

If we ignore the destruction terms D_k, the different mutants are fully charac- terized through their rates for exact replication W_k. This assumption simpli- fies the following discussion by allowing a single distribution of value parameters $f(W)$ to completely characterize the mutant spectrum, but this is not essential to the argument. Accordingly, we write

$$V_{ik} = \left(\frac{q^{-1} - 1}{\kappa - 1} \right)^{d(i, k)}, \qquad (III.7)$$

expressing the sharp dependence of mutation rate on the Hamming distance in the mutant space.

In Appendix 6 we summarize the necessary modification of the Rayleigh– Schrödinger perturbation theory, which was applied to replication dynamics [15, 16], in order to obtain explicit expressions for higher order terms. This Brillouin–Wigner approach yields a self-consistent expression for the domi- nant eigenvalue, Eqn. (A6.8), but one involving a perturbation series that may be immediately written down to arbitrarily high order [Eqn. (A6.3) evaluated at $s = \lambda$]. In particular, the wild-type fraction in the population is given by the reciprocal of Eqn. (A6.9) for $k = m$, the result being not restricted to second order. Localization of the quasi-species about the wild type I_m is then a question [29] of the convergence of Eqn. (A6.9), or equivalently of the series

$$T_m = \sum_{i \neq m} T_{im} = \sum_{i \neq m} \frac{{}^c W_{im}(\lambda)}{W_{mm}} \qquad (III.8)$$

determined by Eqn. (A6.3), or more conveniently by Eqn. (A7.1). Here, λ denotes the dominant eigenvalue of the matrix W ($\lambda > W_{ii}$ for all i).

The question of convergence of this series in turn depends on the exact replication rates W_{ii} of all the different mutants. If we describe these statis- tically in the absence of more detailed information, then convergence must be

TABLE I
Effective Superiority of Wild Type for Various Replication Rate Distributions[a]

Distribution	Description	Probability Density	$\langle (W_0 - W)/W \rangle_{\ln}$	$\ln \sigma_{\text{eff}}$
Delta function	Constant value $W = \bar{W} < W_0$	$f(W) = \delta(W - \bar{W})$	$(W_0/\bar{W}) - 1$	$\ln(W_0/\bar{W})$
Rectangular	1. All values $0 \leqslant W \leqslant W_0$ equiprobable	$f(W) = \dfrac{1}{W_0}, \ 0 \leqslant W \leqslant W_0$ (0 otherwise)	1	$\ln 2$
	2. All values, $W_L \leqslant W \leqslant W_0$ equiprobable "near neutral"	$f(W) = \dfrac{1}{W_0 - W_L}, \ W_L \leqslant W \leqslant W_0$ (0 otherwise)	$1 - (W_L/W_0)$ for $W_L \to W_0$	$1 - (W_L/W_0)$ for $W_L \to \bar{W}$
Exponential (*see also gamma distribution*, $\eta = 1$)	Mean $\bar{W}$ truncated at $W_0 \gg \bar{W}$, c_1 is Euler's constant $(0.577 \ldots)$	$f(W) = \dfrac{1}{\bar{W}} e^{-W/\bar{W}}, \ 0 \leqslant W \leqslant W_0$ (0 otherwise)	$(W_0/\bar{W}) e^{c_1}$	$c_1 + \ln(W_0/\bar{W})$
Gamma, truncated at $W_0 \gg \bar{W}$	Parameters β, η, mean $\bar{W} = \beta\eta$, relative variance $\sigma/\bar{W} = \eta^{-1/2}$	$f(W) = \dfrac{W^{\eta - 1}}{\beta^\eta \Gamma(\eta)} e^{-W/\beta}, \ 0 \leqslant W \leqslant W_0$ (0 otherwise)	$(W_0/\bar{W}) e^{c_\eta},$ $c_\eta = c_1 + \ln(\eta - 1)$ $- \sum\limits_{r=1}^{\eta - 1} 1/r$	$c_\eta + \ln(W_0/\bar{W})$

$\eta = 1, \quad \sigma/\bar{W} = 1$
$\eta = 2, \quad \sigma/\bar{W} = 0.71$
$\eta = 3, \quad \sigma/\bar{W} = 0.58$
$\eta = 4, \quad \sigma/\bar{W} = 0.50$
$\eta = 9, \quad \sigma/\bar{W} = 0.33$
$\eta = \infty, \quad \sigma/\bar{W} = 0$

$c_1 = 0.58$
$c_2 = 0.27$
$c_3 = 0.18$
$c_4 = 0.13$
$c_9 = 0.036$
$c_\infty = 0$

| Normal, truncated at 0 and at $W_0 \gg \bar{W}$ | Mean $\bar{W}$, variance σ^2 | $f(W) = (2\pi\sigma^2)^{-1/2} e^{-1/2[(W-\bar{W})/\sigma]^2}$, $0 \leq W \leq W_0$ (0 otherwise) $\sigma/\bar{W} = 0.71$ $\sigma/\bar{W} = 0.50$ $\sigma/\bar{W} = 0.33$ $\sigma/\bar{W} = 0$ | $(W_0/\bar{W})e^{\sigma^2/\bar{W}^2}$ $\sigma^2/2\bar{W}^2 = 0.25$ $\sigma^2/2\bar{W}^2 = 0.125$ $\sigma^2/2\bar{W}^2 = 0.055$ $\sigma^2/2\bar{W}^2 = 0$ | $\frac{1}{2}\dfrac{\sigma^2}{\bar{W}^2} + \ln\dfrac{W_0}{\bar{W}}$ $\ln(W_0/\bar{W})$ |

[a] Here $\underline{f}(W)$ represents an underlying probability distribution from which (in a possibly correlated manner) replication rates of different mutants are sampled. $\bar{W}$ is average of this distribution and W_0 the additionally specified maximum value (which truncates distribution in latter cases). This maximum value is assigned to master sequence m, so we may write $W_{mm} = W_0$.

discussed probabilistically. The question is then how probable localization of the quasi-species is for a given distribution of replication rates.

In order to answer this question, a significant source of statistical correlation arising from mutation paths that visit a particularly advantageous mutant more than once must be considered. In the perturbation theory these paths are represented by products of factors involving the mutant replication rates, and it is necessary to remove the strong correlation that arises between these factors where repeated indices are present in order to obtain a tractable statistical analysis of convergence. The Watson renormalization procedure [29], the application of which to the steady-state quasi-species is summarized in Appendix 7, accomplishes just this [30]. The cost is a consecutive modification of the denominator, which may however be simplified to good approximation, as in Eqn. (A7.5).

The individual contributions to the series T_m may then be shown, as outlined in Appendix 8, to have a narrow probability distribution about their logarithmic mean value provided the replication rates have no long-range correlations with respect to the Hamming distance. The contributions of paths involving products of N mutations may then be summed to an expression that lies with a probability approaching 1 in a narrow range of values about L^N as N becomes large. Thus the convergence of the series T_m, or the nonvanishing of the wild-type population, depends on whether the parameter L is greater than or less than 1. It is further possible to demonstrate that λ is not significantly different from W_{mm} even at threshold, which is then characterized by the equation

$$1 = L(W_{mm}) \simeq (q^{-\nu} - 1) \left\langle \frac{W}{W_{mm} - W} \right\rangle_{\ln},$$ (III.9)

where the logarithmic average defined in Eqn. (A8.9) is over a probability distribution $f(W)$ of replication rates taking values up to that of the wild type W_{mm}. The maximum sequence length corresponding to the preceding threshold is

$$\nu_{max} = \frac{\ln\left[1 + \langle (W_{mm} - W)/W \rangle_{\ln}\right]}{-\ln q} = \frac{\ln \sigma_{eff}}{-\ln q}.$$ (III.10)

In comparison with Eqn. (III.4) the result of the statistical analysis is to provide an expression for the effective superiority parameter σ_{eff} of the wild type in terms of the distribution of replication rates of its mutants.

Some values of the logarithmic average appearing in Eqn. (III.9) and of $\ln \sigma_{eff}$ are recorded for various replication rate distributions in Table I. For

the special case where all the mutants have identical replication rate $\bar{W}$, the superiority parameter of Section III.1 is independent of population variables and may be explicitly evaluated, agreeing with the effective superiority defined by Eqn. (III.10). At the opposite extreme, where all values less than or equal to the wild-type replication rate W_{mm} are equally probable, a finite answer $\ln \sigma_{\text{eff}} = \ln 2 = 0.69 \ldots$ is obtained. As the replication rate distribution becomes more weighted toward the wild-type value, a neutral mutant extreme is approached for which $\ln \sigma_{\text{eff}}$ steadily decreased toward zero. These last two results are at first surprising because the continuous distributions imply the existence of mutants arbitrarily close in selective value to the wild type. That the effective superiority of the wild type is nonzero is a reflection of the crucial effect of mutational distance and the structured population of the quasi-species.

The remaining three distributions in Table I are appropriate when the wild-type replication rate is regarded as the largest value of many sequences sampled. We shall focus on this view in the following section, which allows rather general conclusions to be made with the aid of extreme-value theory.

It will be apparent that the probability distributions, upon which the localization threshold result of Eqn. (III.9) depend, contain no reference to possible correlations between the selective values of neighboring sequences in the mutant space. A consideration of the statistical argument summarized in Appendix 8 shows that the localization threshold depends only on the absence of long-range correlations between the selective values. We illustrate short- and long-range correlation by means of examples. Short-range correlation occurs in systems where the selective value of the mutant has always a fixed probability of a value near that of each of its neighbors. This would lead, for example, to those direct-track mutants, between the wild-type and an unusually good competitor, having a higher probability of a high selective value than others. A long-range correlation, which would jeopardize the result, occurs when the replication rate is dependent upon two properties determined by disjoint subsets of the sequence. Such correlations occur, for example, when a sequence codes for several proteins or when only a fraction of the sequence determines the fitness (cf. Section III.4).

3. Extreme-Value Theory for Effective Superiority

The evaluations of the error threshold and effective superiority discussed so far [Eqn. (III.10)] assume specific knowledge of the wild-type net rate of exact replication W_{mm}. It is clear, however, that the wild type is distinguished from the other mutants only insofar as it has the maximum selective value, and so it is sensible to regard this value as the random extreme of n trials from the probability distribution $f(W)$ [29]. We may then speak of the general

localization threshold properties of a quasi-species belonging to a particular distribution dependent upon n instead of W_{mm}. The results are of more than academic interest in their remarkable insensitivity to n for large populations. Clearly, n is the number of different mutant sequences sampled in the history of the quasi-species.

The extreme value $W_0^{(n)}$ of n trials from a probability density $f(W)$ with the cumulant distribution $F(W)$ satisfying the large W behavior

$$\lim_{W \to \infty} \frac{d}{dW} \left(\frac{1 - F(W)}{f(W)} \right) = 0 \qquad \text{(III.11)}$$

has, asymptotically, the extreme-value distribution of exponential type in that

$$\lim_{n \to \infty} \text{Prob} \{ (W_0 - l_n)nf(l_n) \leq u \} = \exp[-\exp(-u)], \qquad \text{(III.12)}$$

where l_n depends on the underlying distribution $F(l_n) = 1 - 1/n$. The preceding result is one form (due to van Mise, see ref. 32) of the extreme-value theorem, the (cumulant) distribution on the right of Eqn. (III.12) being known as the limiting extreme-value distribution of exponential type.

The physically reasonable distributions in the absence of a specified maximum are not truncated but have diminishing probability for very high values satisfying Eqn. (III.11). The extreme-value parameter l_n, which is the mode of the extreme, and $nf(l_n)$, which is related to its variance, are displayed for exponential, gamma, and normal distributions for large n in Table II. The modal estimate becomes reliable for large n, and the error threshold may be estimated with $W_0 = l_n$, as the table indicates. The n dependence of the threshold through $\ln(\ln n)$ is so extremely weak that rather precise estimates can be made. Thus, in the $Q\beta$ replicase in vitro evolution experiments of the Spiegelman type, the number of molecules limits n to $10^{12 \pm 3}$ even taking into account consecutive sampling of sequences (on a laboratory time scale of 10^2 generations), and $\ln(\ln n)$ is then 3.3 ± 0.3. At the other extreme of the maximum possible number of sequences sampled in the course of the evolutionary history of the earth, $n < 10^{50}$, $\ln(\ln n)$ is only 4.75. The numbers in the table are for $n = 10^{12}$. The results completely confirm the original view that $\ln \sigma$ should be a number of order unity. In comparison with experiments, it should be emphasized that the threshold sequence length must be greater than the observed sequence length for a stably localized quasi-species. For example, a single independent measurement of v and q for a $Q\beta$ viral RNA [34, 35] indicates $\ln \sigma_{\text{eff}} > 1.4$ for this virus.

TABLE II

Extreme-Value Parameters and Corresponding Effective Superiorities for Various Underlying Distributions[a]

Underlying Distribution	Probability Density	l_n	$nf(l_n)$	$\ln \sigma_{eff}$
Exponential (see also gamma distribution, $\eta = 1$)	$f(W) = \dfrac{1}{\bar W} e^{-W/\bar W}$	$\bar W \ln n$	$1/\bar W$	$c_1 + \ln(\ln n)$
Gamma:parameters β, η (c.f. Table I)	$f(W) = \dfrac{W^{\eta-1}}{\beta_\eta \Gamma(\eta)} e^{-W/\beta}$	$\dfrac{\bar W}{\eta} \ln[n/\Gamma(\eta)]$	$\eta/\bar W$	$c_\eta - \ln\eta + \ln\ln[n/\Gamma(\eta)]$
Numerical				
$\eta = 1$ $(\sigma/\bar W = 1)$		$27.6\,\bar W$		3.9
$\eta = 2$ (0.71)		$13.8\,\bar W$		2.9
$\eta = 3$ (0.58)		$8.9\,\bar W$		2.4
$\eta = 4$ (0.50)		$6.5\,\bar W$		2.0
$\eta = 9$ (0.33)		$1.9\,\bar W$[b]		1.2[b]
Normal $(\bar W, \sigma)$	$f(W) = (2\pi\sigma^2)^{-1/2} e^{-1/2[(W-\bar W)/\sigma]^2}$	$\bar W + \sigma(2\ln n)^{1/2}$	$\dfrac{1}{\sigma}(2\ln n)^{1/2}$	$\tfrac{1}{2}\sigma^2/\bar W^2 + \ln(\sigma/\bar W) + \tfrac{1}{2}\ln(\ln n)$
$\sigma/\bar W = 0.71$		$6.3\,\bar W$		1.5
$\sigma/\bar W = 0.50$		$4.7\,\bar W$		1.1
$\sigma/\bar W = 0.33$		$3.5\,\bar W$		0.6

[a] Here n is number of trials from probability distribution $f(W)$ of rates for exact replication having mean W. l_n is value of W for which there is probability $1/n$ of sampling higher replication rate. Numerical values of l_n and $\ln\sigma_{eff}$ are for $n = 10^{12}$ (see text).

[b] Approximation ill founded.

4. Relaxed Error Threshold and Gene Duplication

So far we considered every part of the polynucleotide sequence as being equally important for the replication process. This very restrictive assumption will not be justified for most systems in reality. This is quite obvious for eukaryotes, where the larger fraction of the genome consists of noncoding intervening regions (introns). Though introns, depending on their functional role, may well require conserved sequences, the fidelity requirements will vary greatly and differ from those in expressed regions (exons). Even in the coding regions, mutations in some parts of the polymer have little or practically no influence on the "fitness" of the polynucleotide. In test tube replication experiments such unimportant parts are represented by bases outside the regions of enzymic recognition, when, in addition, they do not contribute to the secondary structure required for replication. In the genomes of organisms these mutations concern changes in the base sequence, which, when translated into protein, yield enzymes that differ very little in catalytic efficiency from those of the wild type. Other mutations are "silent" and lead to identical proteins upon translation. Others may concern parts of the DNA that are not translated at all and that have no influence on replication as well. Such mutations are commonly considered as being *selectively neutral.*

Selectively neutral mutants were observed in test tube evolution experiments [36]. They are much harder to detect with small viruses, which often have highly condensed genomes including overlapping genes, read-through positions, and so on. Neutral mutants were found in chemostat experiments with bacteria [37–39] and in higher organisms [2]. The latter have in addition large fractions of their genomes apparently unused for coding protein or regulation. Changes within these regions of the DNA will be selectively neutral provided they do not interfere with recombination.

Given that such regions of little or no influence on the fitness of a polynucleotide sequence do exist, we expect intuitively a relaxation of the error threshold. Typical situations in which this relaxation appears to be of major importance are gene duplication, multigene messengers, and segmented genes. In this section we present the results of a quantitative analysis of replication with partially relaxed constraints on accuracy [40]. The same approach can be applied also to related biological problems such as the occurrence of "selfish" RNA or DNA or the exon–intron structure of eukaryotic genes. We consider a polynucleotide of chain length v that is built up from two segments A and B. The two segments are v_A and v_B bases long, respectively: $v = v_A + v_B$. Both segments exist in various mutant forms, $A_1, A_2, \ldots, A_r$ and $B_1, B_2, \ldots, B_s$ (Figure 7). We may visualize such a polynucleotide as a primitive genome consisting of two genes only. Both genes are present in several alleles. Segments A and B need not be joined
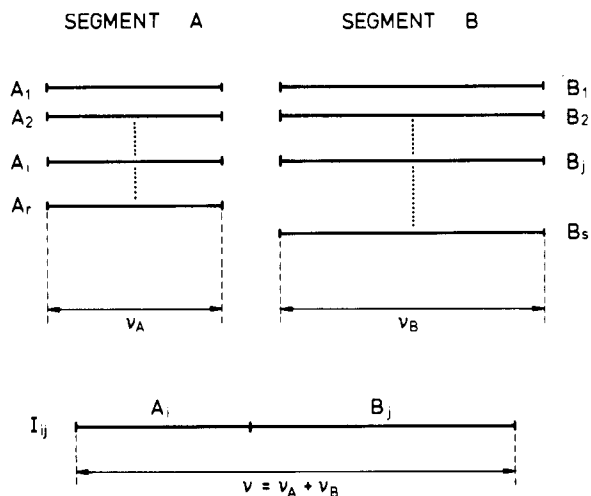
Figure 7. Structure of segmented polynucleotide sequences. Sequence is built from two segments of chain lengths v_A and v_B, respectively; rs different sequences can be formed by combination of r segments of type A and s segments of type B.

physically. For instance, segment B could be split into some or many parts that are inserted into A at different positions. We shall only assume that changes in all parts of A have a strong influence on the fitness of the polynucleotide, whereas those in the parts of B are of little or no influence.

An individual polynucleotide sequence is labeled by two indices, i and j: I_{ij} denotes the combination $A_i - B_j$. There are, of course, rs individual combinations of this kind, that is, rs polynucleotide sequences on which selection acts. We introduce relative concentrations (which carry double indices accordingly), $c_{ij} = [I_{ij}]$:

$$x_{ij} = c_{ij} \left/ \sum_{kl=1}^{rs} c_{kl} \right.$$

(III.13)

with $0 \leqslant x_{ij}$ and $\Sigma x_{ij} = 1$. The selection equation is of the form

$$\dot{x}_{ij} = \sum_{kl=1}^{rs} W_{ij,\,kl} x_{kl} - \bar{E} x_{ij},$$

(III.14)

The selective values and the rate constants are defined as before; in particular, we have

$$W_{ij,\,kl} = A_{kl} Q_{ij,\,kl} - D_{ij} \delta_{ij,\,kl}.$$

In principle, Eqn. (III.14) does not differ from the cases treated above in the preceding.

Now, we shall focus upon individual regions of the polynucleotide sequence $A_{(i)}$ and $B_{(j)}$. We introduce new variables that represent summations over the distribution of the alleles in the other region:

$$x_{i0} = \sum_{j=1}^{s} x_{ij} \quad \text{and} \quad x_{0j} = \sum_{i=0}^{r} x_{ij}. \tag{III.15}$$

Usage of these variables requires definition of average selective values and average rate constants, which are summarized in Appendix 9.

The analysis of the selection equation is based on two assumptions:

1. We consider single-step mutations only. Precisely, we assume that either part A or part B may be changed during one act of replication. Simultaneous changes in both parts are neglected.
2. We shall assume that the quality factors for replication of A and of B do not depend on the particular base sequences. We denote these general factors by Q_A and Q_B.

The second assumption is consistent with the uniform mutation frequency model discussed in previous sections. It is, however, less restrictive than that: Error rates need not be uniformly distributed in A or B as long as the total quality factors remain the same.

With these two assumptions we derive separate selection equations for part A and part B (see Appendix 9 and ref. 39):

$$\dot{x}_{i0} = \left[\left(1 - \sum_{k \neq i} Q_{ik}^A \right) \bar{A}_{i0} - \bar{D}_{i0} - \bar{E} \right] x_{i0} + \sum_{k \neq i} Q_{ik}^A \bar{A}_{k0} x_{k0}, \tag{III.16}$$

$$\dot{x}_{0j} = \left[\left(1 - \sum_{l \neq j} Q_{jl}^B \right) \bar{A}_{0j} - \bar{D}_{0j} - \bar{E} \right] x_{0j} = \sum_{l \neq j} Q_{jl}^B \bar{A}_{0l} x_{0l}, \tag{III.17}$$

with $i, k = 1, \ldots, r$ and $j, l = 1, \ldots, s$.

These two differential equations describe how selection acts on parts of the replicating unit. The equations are equivalent to the selection Eqn. (III.15) in so far as no simplifying assumptions were made except the two concerning the structure of the mutation matrix Q. The interaction between the two parts of the polynucleotide occurs via the average rate constants $\bar{A}_{i0}$, $\bar{D}_{i0}$, $\bar{A}_{0j}$ and $\bar{D}_{0j}$ as well as implicitly through the common average excess production:

$$\bar{E} = \sum_{i=1}^{r} (\bar{A}_{i0} - \bar{D}_{i0}) x_{i0} = \sum_{j=1}^{s} (\bar{A}_{0j} - \bar{D}_{0j}) x_{0j}. \tag{III.18}$$

Let us now compare the mathematical structures of the selection Eqn. (III.15) and the coupled systems of Eq. (III.16) and (III.17): The original equation had rs variables and one conservation relation and was linear apart from the mild nonlinearity caused by $\bar{E}$. Equations (III.16) and (III.17) contain $r+s$ variables only; they fulfil two conservation relations but are highly nonlinear through the coupling terms. We recall from Appendix 9 that, For example,

$$\bar{A}_{i0} = \frac{1}{x_{i0}} \sum_{j=1}^{s} A_j x_{ij}.$$

Hence, solution of these equations will be rather involved in the general case, and one has to rely on numerical procedures. These are readily available because closely related problems are frequently encountered in numerical mathematics [41]. We dispense here with the details of the corresponding equation for the general case where the sequences in part B contribute differently to the fitness. As expected, it is not possible to remove the coupling terms completely.

Let us now consider an important special case in which the coupling terms have a particularly simple structure. We shall assume that the individual base sequences in part B have no specific influence on the replication rate of the whole polynucleotide $A-B$. They are selectively neutral, and we may write

$$A_{ij} = A_i f_B,$$

wherein A_i is the sequence specific effect of part A and f_B is the unspecific effect of part B on the rate of replication. This ansatz has the consequence that the two mean values are of different form:

$$\bar{A}_{i0} = A_1 f_B \quad \text{and} \quad \bar{A}_{0j} = f_B \frac{1}{x_{0j}} \sum_{j=1}^{r} A_i x_{ij},$$

which implies highly unsymmetric coupling of the two differential equations.

For the sake of simplicity we assume equal rates of degradation for all sequences ($D_{ij} = D$). Then the contributions of D to W and $\bar{E}$ cancel in both differential equations, and we may simply neglect them. The selection equation of part A can now be written in a form that closely resembles the original differential equation (Appendix 3):

$$\dot{x}_{i0} = f_B \left((A_i Q_{ii}^A - \bar{E}') x_{i0} + \sum_{k \neq i} A_k Q_{ik}^A x_{k0} \right) \tag{III.19}$$

with

$$E' = \sum_i A_i x_{i0} \quad \text{and} \quad Q_{ii}^A = 1 - \sum_{k \neq i} Q_{ik}^A \qquad (i, k = 1, \ldots, r).$$

In principle, the common factor f_B can be absorbed into the time axis by means of a linear transformation: $\tau = t f_B$. Apart from this scaling the differential Eqs. (III.19) are completely independent of part B. The equations are now confined to part A of the polynucleotide, and the error threshold is determined by the quality factor Q_{ii}^A. This quality factor is substantially larger than the quality factor of the entire sequence,

$$Q_{ij, ij} = 1 - \left(\sum_{k \neq i} Q_{ik}^A + \sum_{l \neq j} Q_{jl}^B \right) \quad \text{and} \quad Q_{ii}^A - Q_{ij, ij} = \sum_{l \neq j} Q_{jl}^B.$$

In other words, the formula derived for the maximum number of bases in the polynucleotide sequence ($\nu_{\max}$) applies now to ν_A only and not to the total length.

In general, chain lengths ν are not conserved in polynucleotide replication. Insertions and deletions are common. Duplications of longer parts of sequences occur as well. Every polynucleotide thus has to compete not only with molecules of the same chain length but also with longer and shorter sequences.

Whether or not a given polynucleotide can compete successfully with competitors of different length is determined largely by the chain length dependence of the rate of replication. This dependence is rather complicated [10–12] and varies with environmental conditions. As an example, we consider the limit of long polynucleotide chains. If template concentrations are in vast excess of polymerase concentrations, a polynucleotide will bind one enzyme molecule or none. The mean time for the complete replication process increases linearly with the chain length, and the rates of replication are roughly proportional to ν^{-1}. If the enzyme is available in excess of templates, however, many polymerase molecules can bind to the same template. The number of polymerase molecules operating on one polynucleotide strand is proportional to its length, and then the rate for the replication of sufficiently long chains becomes independent of the chain length.

The factor f_B is an appropriate measure of the fitness of polynucleotides that carry unused or redundant parts and compete with shorter and longer sequences. In case the rate of replication does not depend on the chain length, f_B will be close to 1 and longer sequences may compete successfully with shorter ones. Rates of replication that are largely insensitive to chain lengths are particularly important for the process of gene duplication. A coding

sequence may be duplicated accidentally by a replication error. The longer polynucleotide formed has to compete with shorter sequences for some intermediate period of time during which its fitness remains practically unchanged ($f_B \approx 1$). One part of the sequence is not required for replication and hence may vary unrestrictedly. If variation of the unused part leads to a sequence that has a positive effect on the fitness of the whole polynucleotide, it will proliferate and eventually become dominant in the population. An increase in fitness may be caused simply by changes in the secondary structure of the polynucleotide. More interesting is a case when the unused part is translated and the translation product gains catalytic activity. Successive gene duplications may lead to whole families of proteins with different substrate specificities.

The new catalytic function may eventually improve the accuracy of replication. Then gene duplication represents a powerful evolutionary mechanism to increase the catalytic capacity of a replicating system, which makes replication more precise whenever this is necessary and possible.

Present-day structures of proteins provide some hints on the role of gene duplications in evolution. Proteins commonly occur in families. These are structurally related enzymes catalyzing reactions of the same class with different substrate specificities. Examples are families of proteases or dehydrogenases. In addition to this, one observes interesting regularities in the structures of many globular proteins: Substructures (so-called motifs) are often repeated exactly or with minor modifications only. Such repetitions were found in the same protein molecule as well as in different protein molecules. Both the modular structure of polymers as well as the existence of protein families can be explained by a gene duplication mechanism.

Last but not least, we may mention the exon–intron structure of eukaryotic genomes. Though error rates are generally adapted to the full genome size and throughout found to be smaller than 10^{-10}, fidelity requirements may differ for exons and introns and even vary within both. For instance, part of the intron regions may have signal character and therefore ought to be highly conserved. The relations referred to in this section will allow a quantitative treatment of such cases.

5. Analogies to Phase Transitions

Numerical studies on quasi-species localization [42]—as they will be presented and discussed in Section IV—showed a sharpening of error thresholds with increasing chain lengths v. *Sharpening* means here that the transition zone from localized quasi-species to delocalized sequence distributions becomes narrower. This phenomenon is reminiscent of cooperative transitions observed with conformational equilibria in biopolymers. We shall investigate here whether or not error thresholds show analogy to phase transitions in the

limit $v \to \infty$. The equilibrium statistical mechanics of lattice models and the dynamics of Markov processes can be described within the same mathematical discipline of "*Markov random fields*" [43], and therefore, similarities in global behavior such as the existence of cooperative phenomena or phase transitions are not completely unexpected. Spin lattice models are particularly well-suited candidates in the search for analogies between equilibrium properties and replication dynamics because several of them are sufficiently simple to allow derivations of analytical expressions of thermodynamic functions [44].

Analogies between replication dynamics and spin lattice models were investigated in recent publications by Demetrius [45] and Leuthäusser [46, 47]. Both approaches are based on a common concept, and we shall discuss them here together. Replication dynamics is considered as a dynamical system in discrete time and modeled by the difference equation

$$x_i(t + \Delta t) - x_i(t) = \sum_{k=1}^{n} W_{ik} x_k(t) - [D_i + \Phi(t)] x_i(t), \quad i = 1, \ldots, n, \qquad \text{(III.20)}$$

which is defined on a time axis $\ldots, t - \Delta t, t, t + \Delta t, \ldots$. The discrete dynamics of the replication–mutation process is, of course, not identical to the dynamics of the kinetic Eqn. (A4.1) since it assumes synchronized replications. In the limit of infinite time $(t \to \infty)$, however, the distributions of polynucleotide sequences in the discrete system converge asymptotically toward the stationary solutions of the differential equation [48].

The discrete dynamical system can be modeled by a multitype branching process provided the flux term $\Phi(t)$ is neglected. Then the distribution of sequences is derived from proper statistics of *genealogies*. A genealogy is an individual time-ordered series of sequences that represents one particular recording of successive descendants: each member of a given genealogy was synthesized by correct or erroneous copying of its precursor as template (Figure 8). The individual genealogies are compared with one-dimensional lattices of generalized spins. Every spin has $n = \kappa^v$ different states, 2^v for binary sequences and 4^v for polynucleotides. Interactions on the lattice are restricted to nearest neighbors. This restriction follows naturally from replication dynamics or from the Markov property of the branching process: the probability distribution of copies—correct replicas and mutants—depends on the template to be copied and *not* on its precursors in the genealogy. For the spin lattices the nearest-neighbor approximation is a more serious restriction since magnetic interactions may reach much further. It is nevertheless the basis of a well-studied class of models known as Ising lattices.

Probability distributions on the genealogies correspond to probability distributions on the lattice of spins. The connection between the discrete

(a)
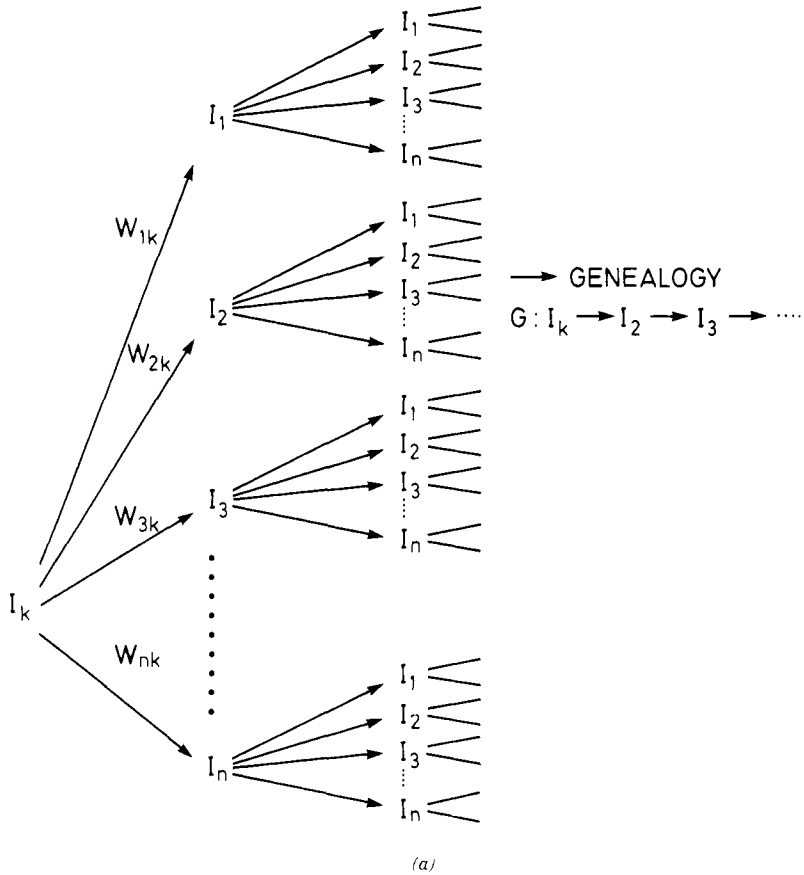
Figure 8. Polynucleotide replication as multitype branching process is compared with spin lattice models: (A) One-dimensional model is based on generalized one-dimensional Ising lattice: Every spin is assumed to exist in n different states corresponding to n different polynucleotide sequences. Genealogy of branching process is considered as analog of particular one-dimensional arrary of spins.

dynamics and the lattice model is based on the following fact: The dominant eigenvalue of the value matrix W, λ, satisfies a variational principle formally identical to the minimization of the free energy in statistical mechanics. As a consequence, the macroscopic quantities in replication dynamics can be compared with the thermodynamic variables in lattice models. The free energy, for example, corresponds to the logarithm of the dominant eigenvalue of the value matrix W: $F \propto \log \lambda$. The product of entropy S and temperature T is the negative analog of the *complexity parameter* H of the stationary
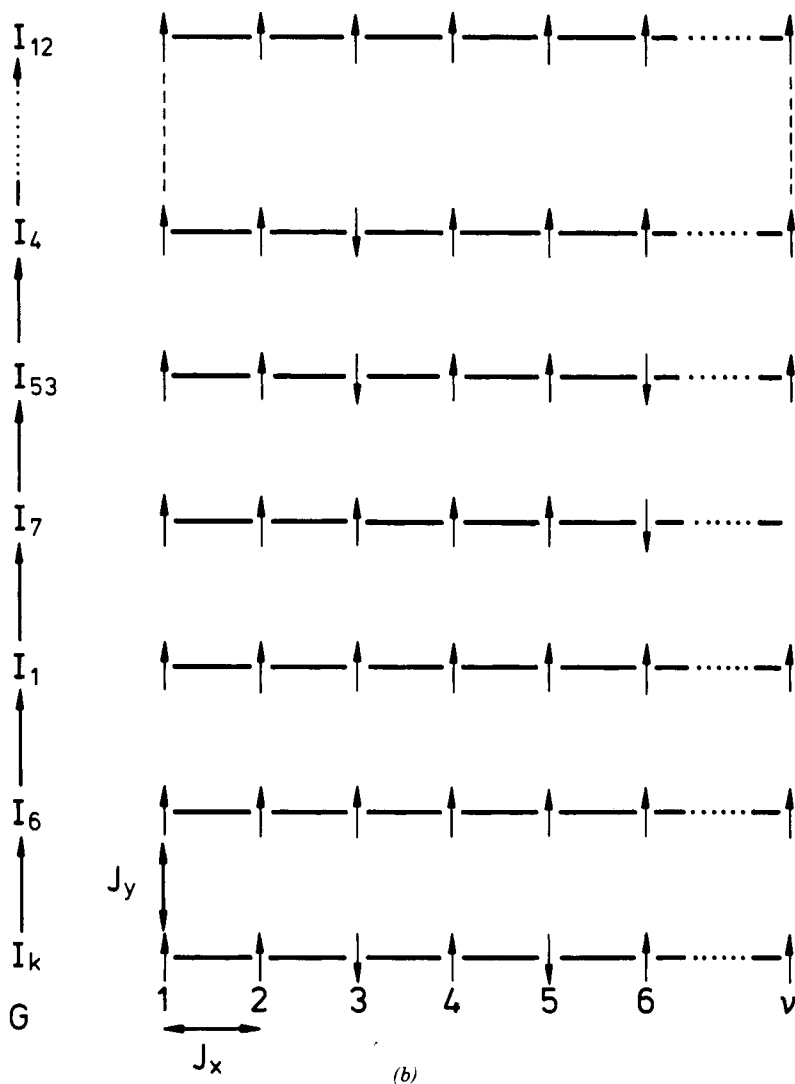
Figure 8. (B) In two-dimensional model genealogy is represented by two-dimensional spin lattice. Individual spins exist in two states ($\uparrow$ and $\downarrow$) corresponding to two digits in binary sequences. Every row of lattice consists of v digits and corresponds to polynucleotide sequence. "In-row" interaction, described by spin–spin coupling constant J_x, is property of individual sequence and contributes to rate constant of replication. "Vertical" coupling constant J_y, on the other hand, is measure of mutation frequency.

sequence distribution: $TS \propto - H$. This complexity parameter is closely related to the Shannon entropy of the sequence distribution, $H_{(Shannon)} = -\Sigma p_k \log p_k$, where p_k is the probability of finding the system in state k [49]. Since the product TS is fixed by analogy to the spin system, the choice of the function corresponding to temperature allows some freedom and requires physical intuition.

In order to search for an interpretation of the error threshold relation by analogy to the spin system, we shall be more specific and consider binary sequences replicating with uniform error rates. Individual sequences are identified with the rows of a two-dimensional spin lattice. A genealogy corresponds to an entire, two-dimensional array of spins. We assign *spin values* ($s = \pm 1$) to the digits (0, 1). A sequence of v digits is identified with a string of spin values:

$$I_k \equiv (s_1^{(k)}, s_2^{(k)}, \dots, s_v^{(lk)}). \tag{III.21}$$

The Hamming distance $d(i, k)$ and the elements of the value matrix W can be expressed in terms of these spin values:

$$d(i, k) = \frac{1}{2}\left(v - \sum_{j=1}^{v} s_j^{(i)} s_j^{(k)} \right) \tag{III.22}$$

and

$$W_{ik} = A_k [q(1-q)]^{v/2} \left(\frac{1-q}{q} \right)^{-1/2\Sigma_j s_j^{(i)} s_j^{(k)}}$$

$$= A_k [q(1-q)]^{v/2} \exp\left(-K \sum_{j=1}^{v} s_j^{(i)} s_j^{(k)} \right), \tag{III.23}$$

where $K = \frac{1}{2} \log [(1-q)/q]$.

Now we consider a two-dimensional spin lattice of dimension $v \times N$. The vector $\sigma^{(i)} = (\sigma_1^{(i)}, \sigma_2^{(i)}, \dots, \sigma_v^{(i)})$ describes the spins in the ith row of the lattice. The spin variables can adopt the two values $\sigma_j^{(i)} = \pm 1$. The nonvanishing terms of the spin Hamiltonian in the nearest-neighbor approximation and in absence of an external magnetic field are

$$H = J_x \sum_{i=1}^{N} \sum_{j=1}^{v-1} \sigma_j^{(i)} \sigma_{j+1}^{(i)} + J_y \sum_{j=1}^{v} \sum_{i=1}^{N-1} \sigma_j^{(i)} \sigma_j^{(i+1)}. \tag{III.24}$$

Here we denote the constant of the nearest-neighbor spin interaction in the horizontal direction, that is, within a row, by J_x and that in the vertical direction by J_y. The partition function of the spin lattice can be factorized

within the Ising model by means of a $2^v \times 2^v$ transfer matrix. The elements of this transfer matrix are obtained as Boltzmann weights of the energies of interaction between individual spin distributions:

$$T_{\sigma\sigma'} = \exp(-\beta E_{\sigma\sigma}) \exp\left(-\beta J_y \sum_{j=1}^{v} \sigma_j \sigma'_j\right). \tag{III.25}$$

Herein, two neighboring rows of the two-dimensional lattice are labeled by $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_v)$ and $\sigma' = (\sigma'_1, \sigma'_2, \ldots, \sigma'_v)$, respectively. The energy $E_{\sigma\sigma}$ describes the interaction of spins within a row. It need not be restricted to nearest-neighbor terms and may also include interactions with a magnetic field. Temperature enters into the expression for $T_{\sigma\sigma'}$ through $\beta = 1/kT$. Macroscopic properties and spin distributions of the spin lattices can be obtained from eigenvalues and eigenvectors of the transfer matrix.

In the two-dimensional Ising model the critical temperature T_c for the order–disorder transition of the infinite system $(N \to \infty, v \to \infty)$ is given by

$$\sinh\frac{2J_x}{kT_c}\sinh\frac{2J_y}{kT_c} = 1.$$

A plot of this relation is shown in Figure 9. The transition temperature
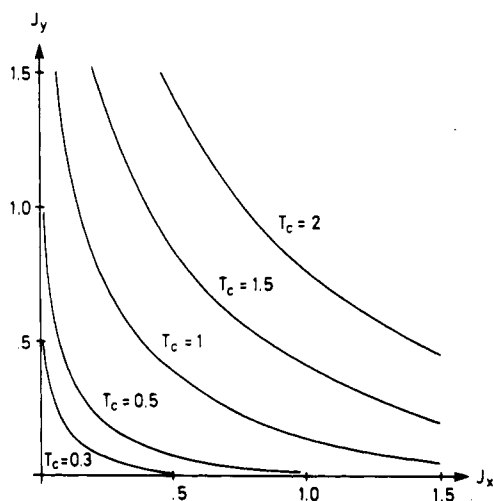
Figure 9. Critical temperture T_c of order–disorder transition in two-dimensional Ising lattice as function of coupling constants J_x and J_y. Note that transition temperature approaches zero when either of two coupling constants vanishes.

converges to $T_c = 0$ in case either of the two spin–spin coupling constants vanishes. In the one-dimensional Ising model we have only a "*degenerate*" phase transition at absolute zero.

Let us now turn to the analogy between the two matrices W and T on which the comparison between replication dynamics and spin lattice is based. The two last factors in Eqs. (III.23) and (III.25) show the same dependence on the spin variables and allow to identify K and βJ_y. This analogy provides a measure for "*temperature*" in replication dynamics:

$$\frac{1}{T} \propto \left| \log \frac{1-q}{q} \right|.$$

Replication degenerates to a random production of sequences in the limit $q \to \frac{1}{2}$ and corresponds to the limit $T \to \infty$, the case of maximum disorder. Direct and complementary replication (see also part C) are the analogs of ferro- and antiferromagnetic cases of the spin system. In the range $\frac{1}{2} < q \leqslant 1$ we have $K < 0$, which corresponds to the condition $\beta J < 0$ for ferromagnetic interaction. For complementary or plus–minus replication, $0 \leqslant q < \frac{1}{2}$, we have $K > 0$ and $\beta J > 0$, respectively, which is characteristic of an antiferromagnet.

A comparison of the first two factors on the right side of Eqn. (III.23) with the first factor in (III.25) is less straightforward. In the transfer matrix of the Ising model we have an "*in-row*" interaction energy of the type

$$E_{\sigma\sigma} = J_x \sum_{j=1}^{v-1} \sigma_j \sigma_{j+1},$$

which represents the first Boltzmann weight in the absence of a magnetic field. The energy $E_{\sigma\sigma}$ is an extensive quantity, and thus $|E_{\sigma\sigma}|$ increases linearly with v. In addition this term shows a unique dependence on the sequence of digits, which is characteristic of spin–spin interactions. The analogy of the "*horizontal*" interaction factors of the spin lattice with replication kinetics requires

$$\exp(-\beta E_{\sigma\sigma}) = \exp\left(-\beta J_x \sum_{j=1}^{v-1} \sigma_j \sigma_{j+1} \right) \propto \exp(\log A_k)$$

apart from constant terms that can be annihilated by appropriate energy scaling; any constant term in the energy enters as a common factor into the partition function and hence has no influence on spin distributions or order–disorder transitions.

In the case of ultimate neutrality, all rate constants A_k are equal and analytical solutions can be derived, as is shown in Appendix 2. In terms of the

Ising model this case corresponds to the limit $J_x \to 0$, where the two-dimensional Ising lattice degenerates to the one-dimensional spin system. We are dealing with a degenerate phase transition:

$$\lim_{\nu \to \infty} T_c(\nu) = 0 \quad \text{or} \quad \lim_{\nu \to \infty} q_c(\nu) = 0,1.$$

There is no phase transition at finite temperatures in the asymptotic limit of infinite chains. This is no surprise since no deterministic selection occurs in the absence of differences in rate constants.

When strictly based on the Ising model the analogy puts a severe restriction on the choice of rate constants, and it is unlikely that any realistic system can meet it. We may, however, extend the comparison to more general spin systems. Many different spin Hamiltonians were proposed and analyzed. To give an example, we mention the Hopfield Hamiltonian [50], which was applied to biological problems, in particular to the study of neutral networks. A strategy proposed by Leuthäusser [46] starts from a given set of replication rate constants—from a "*value landscape*" of the replication system—and tries to find a Hamiltonian that is able to reproduce the rate constants as single row energies: $E_i = -(1/\beta) \ln A_i$. There are several ways to render the expression for the energy more flexible: the presence of an external magnetic field, further neighbor spin interactions, as well as many particle terms may be taken into account. It remains to be shown, however, that physically meaningful value landscapes indeed can be mapped onto more general spin Hamiltonians that will be susceptible to mathematical analysis.

Despite the fact that it may be very difficult to find spin coupling schemes that correspond to realistic value landscapes and can be directly analyzed, the analogy to spin lattices is of great heuristic value. It provides a straightforward explanation of the existence of well-defined error thresholds that sharpen with increasing chain length ν, just as cooperative transitions do in linear biopolymers.

IV. EXAMPLES OF FITNESS LANDSCAPES AND STATIONARY POPULATIONS

The quasi-species was introduced in Section II as a mutant distribution localized in a certain region of sequence space where it is centered around one or several degenerate master sequences. The transition from a nonlocalized to a localized distribution, or the transition among two localized distributions triggered by the appearance of an advantageous mutant, has been shown to be analogous to a phase transition in physical space. In Section III the threshold relations that govern such phase transitions in sequence space have

been discussed in detail. In this section we shall see how population structures and phase transitions depend on the detailed properties of sequence space. Each point in sequence space characterizing a particular sequence is to be assigned a so-called fitness value that expresses the dynamic properties of the particular sequence, such as the replication and decomposition rate. These fitness values will show a more or less continuous short-range order building up to a value landscape that might be characterized by some fractal order, as do landscape topographies on earth. Evolution to optimal fitness may be regarded as a series of phase transitions in which the center of gravity of the localized mutant distribution moves through sequence space, approaching the highest reachable fitness peak. This motion depends not only on the fitness gradients among the 4^ν points in this space but also on the transition probabilities as expressed by the $4^\nu \times 4^\nu$ error or mutation matrix. Both dynamical parameters and the error matrix were used to construct the value matrix W that characterizes the systems of rate equations introduced in Section II.

In the following we discuss examples ranging from the simplest nontrivial cases, for which a full solution is possible, to rather realistic value matrices based on polynucleotide folding, where the analysis of Sections II and III must be applied. The latter is the subject of Section IV.3, while Sections IV.1 and IV.2 concentrate on examples for which an exact solution is numerically feasible.

In principle, stationary mutant distributions may be calculated from the value matrix W by standard diagonalization techniques. Accordingly, we have to compute the dominant eigenvector of a $4^\nu \times 4^\nu$ matrix (for the four bases occurring in present polynucleotides). For most cases of interest (e.g., $\nu > 20$) the problem is clearly intractable (but see the statistical analysis in Section III). This is particularly true for investigations of the error threshold, so in Section IV.1 we study a value matrix with high symmetry [41] for which longer sequences are tractable. In Section IV.2 we use general value matrices for short sequences to investigate the effect of neutral mutants and quasi-species with degenerate master sequences.

For the calculation of stationary mutant distributions we restrict attention to a uniform error rate per digit $(1 - q)$ and assume equal degradation rate coefficients $D_1 = D_2 = \cdots = D_n = D$. Since the addition of a constant to all diagonal elements of a matrix just shifts the spectrum of eigenvalues and has no influence on the eigenvectors, we need only consider the case $D = 0$ without loss of generality. Then the elements of the matrix W are determined by the replication rate coefficients A_k (as in Section III.2) and are of the form

$$W_{ik} = A_k q^{\nu - d(i,k)} (1 - q)^{d(i,k)} = A_k q^\nu \left(\frac{1 - q}{q} \right)^{d(i,k)}, \qquad \text{(IV.1)}$$

where $d(i, k)$ is the Hamming distance introduced in Section II.1. As we mentioned in Section II.5, a value matrix whose elements are defined by Eqn. (IV.1) has exclusively real and positive eigenvalues. Finally, we choose two digits (0, 1) instead of the four bases occurring in natural polynucleotides in the following examples, as this simplifies both the representations and computations without detracting from the main features.

1.   Error Threshold

One approach to calculating the stationary mutant distributions for longer sequences is to form classes of sequences within the quasi-species. These classes are defined by means of the Hamming distance between the master sequence and the sequence under consideration. Class 0 contains the master sequence exclusively, class 1 the v different one-error mutants, class 2 all $v(v-1)/2$ two-error mutants, and so on. In general we have all $\binom{v}{k}$ k-error mutants in class k. In order to be able to reduce the 2^v-dimensional eigenvalue problem to dimension $v+1$, we make the assumption that all formation rate constants are equal within a given class. We write A_0 for the master sequence in class 0, A_i for all one-error mutants in class 1, A_2 for all two-error mutants in class 2, and in general A_k for all k error mutants in class k.

New variables ξ_k are introduced for the relative concentrations of the sum of all k-error mutants:

$$\xi_k = \sum_j x_j, \quad I_j \in \text{Class } j, \quad \text{and hence} \quad \sum_k \xi_k = 1.$$

Finally, it remains to calculate the elements of the matrix Q' consisting of the mutation frequencies from one class of mutants into another. By straight-forward combinatorics, we find

$$Q'_{ik} = q^v \sum_{j=0}^{m} \left(\frac{1-q}{q} \right)^{2j - |k - i|} \binom{v - k}{j + \frac{1}{2}(|k - i| - k + i)} \binom{k}{j + \frac{1}{2}(|k - i| + k - i)}$$

with $m = [\frac{1}{2}(\min\{k + i, 2v - (k + i)\} - |k - i|)]$. (IV.2)

Note that the matrix Q' is not symmetric, in contrast to the matrix of mutation frequencies between individual sequences [cf. Eqn. (IV.1)]. Multi-plication by the corresponding rate constant A'_k yields the corresponding element of the matrix W'. This matrix has to be diagonalized in order to obtain the dominant eigenvector.

The fact that the matrix W' is related to the matrix W by lumping classes of mutants together suggests the existence of a method to relate W' to a

symmetric matrix and directly exploit the positivity of the eigenvalue spec-
trum. In fact, the matrix Q' may be readily symmetrized by normalizing the
classes by their number of elements $\binom{v}{k}$, and then the matrix W' is related to a
symmetric matrix by the standard Jost procedure outlined by Rumschitzky
[8]. This reduces the size of the computational problem.

The first special case to be discussed here considers oligonucleotide
sequences of chain length $v = 5$. The stationary mutant distribution as a
function of the mean single-digit accuracy, $\bar{\xi}_k(q)$ $(k = 1, \ldots, v)$, is plotted in
Figure 10 for a particular choice of the A_ks. We distinguish three ranges of q
with different properties of the stationary distribution:
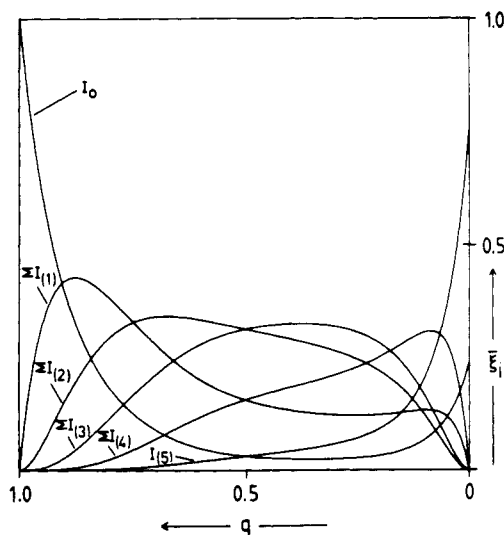
Figure 10. Quasi-species as function of single-digit accuracy of replication (q) for chain v
$= 5$. We plot relative stationary concentration of master sequence ($\bar{\xi}_0$), sum of relative stationary
concentrations of *all* one-error mutants ($\bar{\xi}_1$), of all two-error mutants ($\bar{\xi}_2$), etc. Note that we have
only one five-error mutant $I_{(5)} = I_{31}$ in this particular example. We observe selection of master
sequence at $q = 1$. Then relative concentration of master sequence decreases with decreasing q.
At value $q = 0.5$ all sequences are present in equal concentrations. Hence, sums of concentrations
of two- and three-error mutants are largest—they have statistical weight of 10—those of the one-
and four-error mutants are half as large—they have statistical weight of 5—and finally master
sequence I_0 and its complementary sequence, the five-error mutant I_{31}, are present in relative
concentration of $\frac{1}{32}$ only. At $q = 0$ we have selection of *"master pair"*, which consists of I_0 and I_{31}
in our example. Thus we have direct replication with errors in range $1 > q > 0.5$ and comp-
lementary replication with errors in range $0 < q < 0.5$. Rate constants chosen as $A_0 = 10[t^{-1}]$
and $A_k = 1[t^{-1}]$ for all mutants $k \neq 0$. Here we denote arbitrary reciprocal time unit by $[t^{-1}]$.
All degradation rate constants were put equal: $D_0 = D_1 = D_2 = \cdots = D_{31} = 0$.

1. At large values of the accuracy of replication ($q \simeq 1$) we observe a quasi-species characteristic for "direct" replication, $I_k \rightarrow 2I_k$ predominantly. The master sequence I_0 is most frequent, followed by some one-error mutants, two-error mutants, and so on.

2. In the middle of the plot shown in Figure 10, around the value $q = 0.5$, incorporations of correct and incorrect digits occur with equal probabilities. Inheritance breaks down, and polynucleotide sequences are present according to their statistical weights: The sum of the relative concentrations of one- and four-error mutants, $\bar{\xi}_1$ and $\bar{\xi}_4$, is five times as large as that of the master sequence $\bar{\xi}_0$ or that of the five-error mutant $\bar{\xi}_5$. The statistical weights of two- and three-error mutants is 10. In order to point out the lack of sequence correlations between template and copy we called this process *random replication* [41].

3. At $q = 0$ complementary digits are incorporated with complete accuracy. Thus, we notice "complementary" replication, $I_k^+ \rightarrow I_k^+ + I_k^-$ and $I_k^- \rightarrow I_k^- + I_k^+$ (where I_k^+ and I_k^- denote the plus and minus strand of the sequence I_k). Sequences are selected in pairs of complementary strands. We have a "master pair" that replaces the master sequence of direct replication. In our particular example this is the pair (I_0, I_5). At low nonzero values of the mean single-digit accuracy ($q \simeq 0$) we observe complementary replication with mutation.

For oligonucleotides of very small chain lengths, such as the ones we considered in Figure 10, the three ranges are not well separated. Random replication in a strict sense occurs at $q = 0.5$ only. Indeed, there is practically no correlation between template and copy in the entire flat range $0.4 < q < 0.7$.

Increasing chain length changes the general features of the $\bar{\xi}_1$, q plots rather drastically. For $\nu = 10$ the range of random replication appears to be substantially wider (Figure 11). The $\bar{\xi}_1$, q curves are almost horizontal on both sides of the maximum irregularity condition at $q = 0.5$. In addition, the transitions from direct to random replication and from random to complementary replication are rather sharp. We are now in a position to compare the minimum accuracy of replication that we derived in Section III by perturbation theory with the exact population dependence on q. From Eqs. (III.1) and (III.4) we find ($D_k = 0$; $k = 0, 1, \ldots, n$)

$$q_{\min} = (\sigma_0)^{-1/\nu} = \left(\frac{\bar{A}_{-0}}{A_0} \right)^{1/\nu} \quad \text{with} \quad \bar{A}_{-0} = \sum_{k=1}^{\nu} A_k \bar{\xi}_k \bigg/ (1 - \bar{\xi}_0), \quad \text{(IV.3)}$$

and we may evaluate this directly because the A_k were chosen equal for $k \neq 0$.

The value of $q_{\min}$ calculated by perturbation theory in lowest order falls right into the center of the transition zone that separates the organized quasi-species from the uniform distribution. It represents a good approximation to
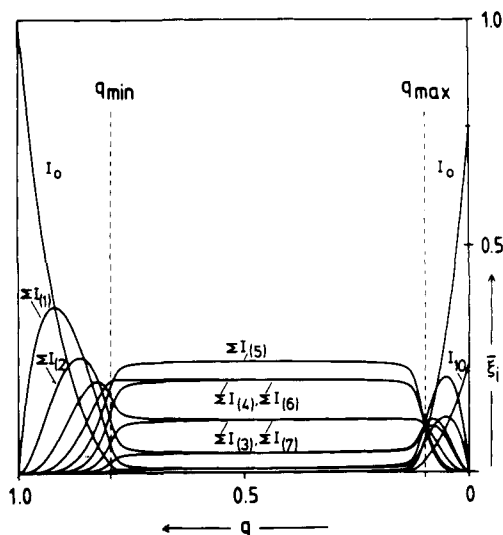
Figure 11. Quasi-species as function of single-digit accuracy of replication (q) for chain length $v = 10$. Computations were performed in complete analogy to those shown in Figure 10. Note that range of *"random replication"* has increased substantially compared to case $v = 5$. We observe fairly sharp transitions between direct and random replication at critical value $q = q_{min}$ and between random and complementary replication at $q = q_{max}$.

the exact error threshold. Indeed, the high-order perturbation result from Eqn. (III.10) agrees well with Eqn. (IV.3) in this special case.

On the other hand, high accuracy in *complementary* replication is characterized by low q values. Here, we have a maximum q value that sets a limit to the errors tolerated in complementary replication. From perturbation theory we derive

$$1 - q_{max} = (\sigma_{0,\bar{0}})^{-1/v} = \left(\frac{\bar{A}_{-0,\bar{0}}}{\sqrt{A_0 A_{\bar{0}}}} \right)^{1/v}$$

$$\text{with } \bar{A}_{-0,\bar{0}} = \sum_{k \neq 0} \sqrt{A_k^+ A_k^-} \, (\bar{x}_k + \bar{x}_{\bar{k}}) \bigg/ \sum_{k \neq 0} (\bar{x}_k + \bar{x}_{\bar{k}}). \tag{IV.4}$$

Concentrations, rate constants, and other quantities referring to the complementary sequence are indicated by the minus symbol ($I_k^+ : \bar{x}_k, A_k^+, \ldots,$ and $I_k^- : \bar{x}_{\bar{k}}, A_k^-$). The value of q_{max} coincides with the center of the transition zone from random to complementary replication.

At still larger chain lengths (we used $v = 50$ in the calculations shown in Figure 12), the transitions become exceedingly sharp. The shape of the plots
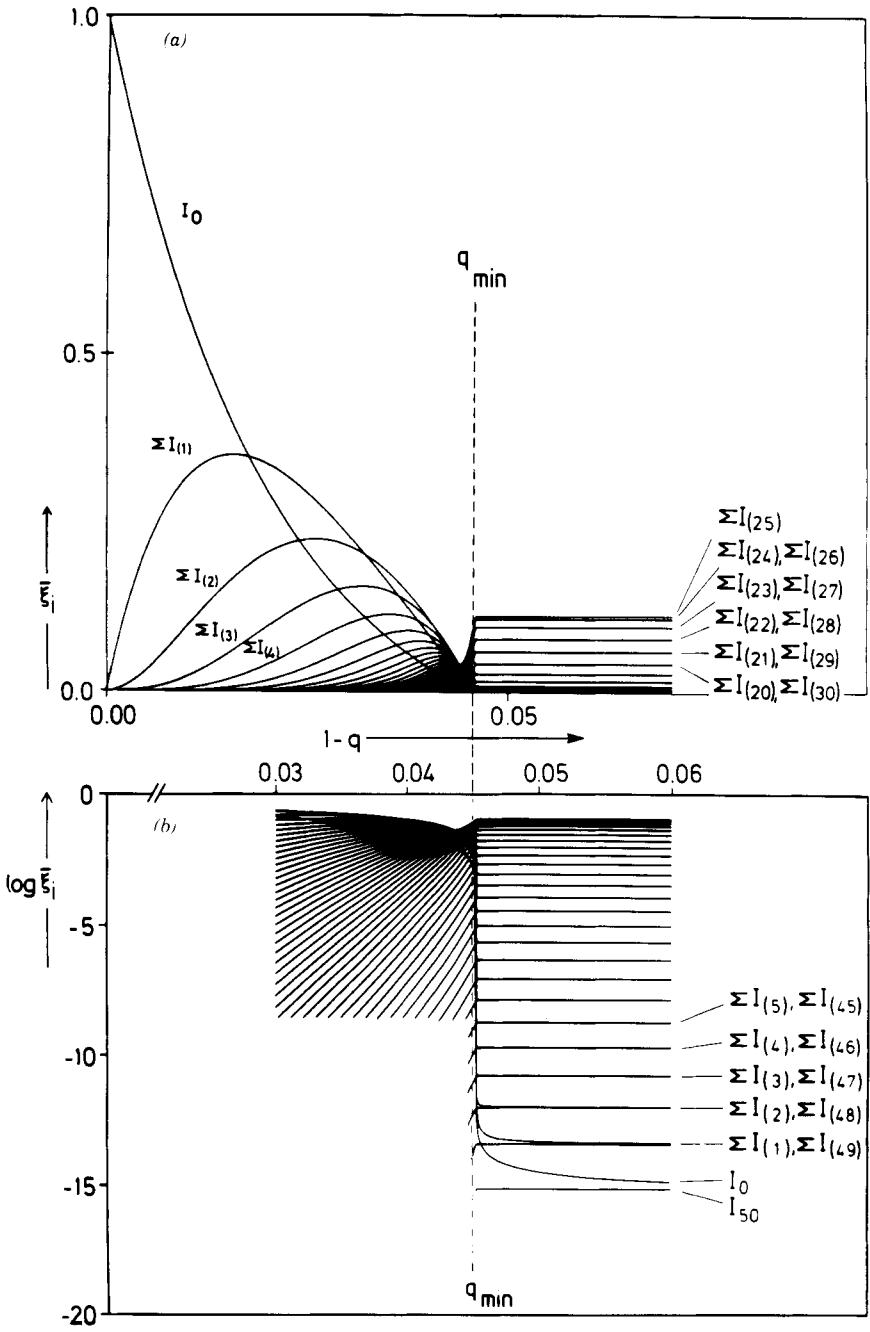
Figure 12. *(Caption overleaf)*

shown in the Figures 10–12 suggests that we are dealing here with an analog of a phase transition as discussed in detail in Section III.5.

2. Degenerate Quasi-species and Neutral Mutants

Mutants that have the same selective value as the master sequence are called (selectively) neutral. We distinguish two classes of neutral mutants:

1. neutral mutants that are "closely related" to the master sequence and
2. distant neutral mutants.

Closely related neutral sequences are present simultaneously in stationary populations. Distant sequences, on the other hand, occur together only occasionally in transient populations. Their mutant distributions are disjoint in the sense that they have no common mutant. Whether or not two selectively neutral sequences share mutants depends both on the Hamming distance and population size. In the first case, the mutant distribution is centered around a dominant, more or less connected group of polynucleotides (replacing the single master sequence considered so far). They are members of the same connected set of mutants and share a common stationary mutant distribution, which we call a *degenerate* quasi-species. However, for two or more neutral master sequences to be equally represented, their mutants have to give identical contributions too. This phenomenon is not considered in the classical study of neutral mutants, which neglects the mutant distribution. Symmetry is unlikely, so some degeneracies will depend on the error rate via a change in the mutant distribution. At different error rates, different neutral sequences are favored. In any case, the dominant eigenvector of the matrix W is representative for the actually observed distribution of sequences.

The second case shows very different behavior: The relative concentrations of the degenerate master sequences are subject to random drift, and the dominant eigenvector of W represents at best a time average of the mutant distribution. Then the dynamics can be modelled only by a stochastic process requiring careful choice of the appropriate mathematical technique and approximations in a hierarchy of equations (see refs. 48 and 51 and Section V.2). One difficulty here is that even very distant mutants contribute if sufficiently neutral. The results of Section III.2 indicate that there is (almost

---

Figure 12. Quasi-species as function of single-digit accuracy of replication (q) for chain length $v = 50$. Computations were performed in complete analogy to those shown in Figure 10. Transition is very sharp at this chain length already. We show linear plot (a) and, in order to demonstrate sharpness even more clearly, a logarithmic plot (b) of relative concentrations around critical point $q = q_{min}$.

surely) a trade-off between distance and accurate neutrality that results in a preservation of the localization threshold behavior.

Here we restrict the discussion to examples of the first class (for further details see ref. 52) and begin with a simple example. Two sequences $I_0^{(1)}$ and $I_0^{(2)}$ have the same selective value W_0. This value is a maximum, $W_0 = \max[W_k; k = 1, \ldots, n]$. We start the discussion of the mutant distribution around the two master sequences with the limit $q \to 1$. For $q = 1$ we have a degenerate eigenvalue: $\lambda_1 = \lambda_2 = W_0$, and hence every combination of concentrations of the two master sequences with $\bar{x}_0^{(1)} + \bar{x}_0^{(2)} = 1$ represents an eigenvector of the matrix W. A small perturbation $(q = 1) \to (q = 1 - \delta)$ destroys this degeneracy. From perturbation theory of degenerate states, which becomes exact in the limit considered, we derive that there exists a unique pair of stationary concentrations $(\bar{x}_0^{(1)}, \bar{x}_0^{(2)})$ that is the proper reference state. These relative stationary concentrations of the two master sequences are determined by the Hamming distance $d(1, 2)$ and by the selective values of the neighboring mutants. It is useful to distinguish three subcases shown in Figure 13:

1. The two master sequences are nearest neighbors in sequence space ($d(1, 2) = 1$):

$$\lim_{q \to 1} \bar{x}_0^{(1)}(q) = \lim_{q \to 1} \bar{x}_0^{(2)}(q) = \tfrac{1}{2}.$$

   Both sequences are represented with equal frequency in the degenerate quasi-species.

2. The two master sequences have a Hamming distance $d(1, 2) = 2$. The normalized concentrations are then

$$\lim_{q \to 1} \bar{x}_0^{(1)}(q) = \alpha, \quad \lim_{q \to 1} \bar{x}_0^{(2)}(q) = 1 - \alpha \quad \text{with } 0 < \alpha < 1.$$

   The value of α is determined by the selective values of the mutants in the neighborhoods of $I_0^{(1)}$ and $I_0^{(2)}$.
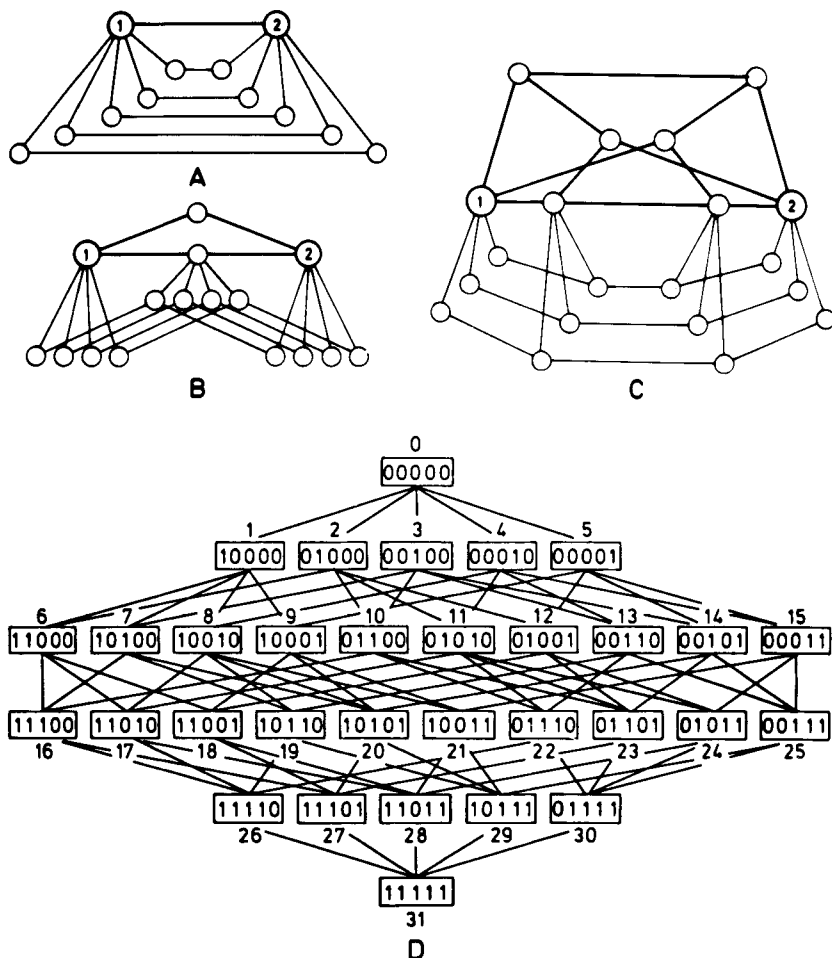
3. The two master sequences have Hamming distances $d(1, 2) \geqslant 3$. This leads to selection in the limit considered:

$$\lim_{q \to 1} \bar{x}_0^{(1)}(q) = 1, \quad \lim_{q \to 1} \bar{x}_0^{(2)}(q) = 0$$

   or

$$\lim_{q \to 1} \bar{x}_0^{(1)}(q) = 0, \quad \lim_{q \to 1} \bar{x}_0^{(2)}(q) = 1.$$

   The decision of which sequence is selected is made again by the neighboring mutants.

Figure 13. Neighborhood relations in sequence space. In A, B, and C we show neighborhoods of pair of sequences (I_1, I_2) with Hamming distances $d(1, 2)$ of 1, 2, or 3, respectively $(v = 6)$. Part drawn in thick lines is general; connections in thin lines depend on chain length v. To give example, for $v = 5$ one connection of equivalent set has to be eliminated. (D) Assignment of numbers to individual sequences of sequence space for $v = 5$ shown as used in Figures 14–30.

In the limit $q \to 0$ we observe an analogous situation. Here the degenerate master sequences are to be replaced by degenerate master pairs. A pair of complementary sequences always has a Hamming distance $d(I_k, I_k^-) = v$. As the Hamming distance between two pairs we use the smallest of the three

distances computed for the four sequences. Then the selection behavior derived for degenerate sequences in the limit $q \to 1$ applies also for degenerate pairs in the limit $q \to 0$.

Around the point of maximum irregularity ($q = 0.5$) the replication rates have no influence on the sequence distribution, and hence degeneracy of rate constants causes no change in the results. In the intermediate ranges, $0.5 < q < 1$ and $0 < q < 0.5$ we have to rely on computer calculations. There are two different scenarios: the "simple" case, where we observe gradual changes in the mutant distribution between error-free and error-prone replication, and the "complicated" case, which does not allow straightforward interpolation because drastic rearrangements of the stationary sequence distributions occur within narrow ranges of q values.

First we present examples of the simple case. In the domain of direct replication ($1 > q > 0.5$) the mutant distributions unfold smoothly around the degenerate sequences. The influence of the neighborhood of the two sequences grows with increasing error rates, that is, with decreasing q values. In the domain of complementary replication ($0 < q < 0.5$) we observe the analogous formation of mutant distributions around the two degenerate master pairs with increasing q values. We have chosen three characteristic examples of two degenerate master sequences with chain length $v = 5$ (Figures 14–16).

In the first example (Figure 14) the sequences I_0 and I_6 are the master sequences. The Hamming distance between them is $d(1, 6) = 2$. One master sequence (I_0) is connected to one-error mutants that have a higher average selective value ($\bar{W} = 5q^v$) than those surrounding the second master sequence (I_6; $\bar{W} = 2.6q^v$). Consequently, the stationary concentration of I_0 is larger than that of I_6. The one-error mutants of I_0 fall into two groups: mutants (I_1, I_2) that have a Hamming distance $d = 1$ to both master sequences and hence are present at higher stationary concentrations than those of the second group (I_3, I_4, and I_5). They have Hamming distance $d = 1$ to I_0 but Hamming distance $d = 3$ to the other master sequence I_6. The mutant distribution thus nicely reflects the details of the structure of the matrix W. In the limit of ultimately precise complementary replication ($q = 0$) we are also dealing with a case of degeneracy. We have two master pairs, (I_0, I_{31}) and (I_6, I_{25}) with identical mean rates of replication: $\sqrt{A_0 A_{31}} = \sqrt{A_6 A_{25}} = \sqrt{10}$. The two master pairs are separated by a Hamming distance $d(\{ \begin{smallmatrix} 0 \\ 31 \end{smallmatrix}, \begin{smallmatrix} 6 \\ 25 \end{smallmatrix} \}) = 2$. They have different neighborhoods and therefore provide an example of case 2 in the limit $q \to 0$.

The second example (Figure 15) considers two distant degenerate sequences I_0 and I_{30}, with $d(0, 30) = 4$. Accordingly, we observe selection in the limit $q \to 1$. The sequence with more efficient one-error mutants (I_{30}) is selected. In the domain of complementary replication we are dealing with two
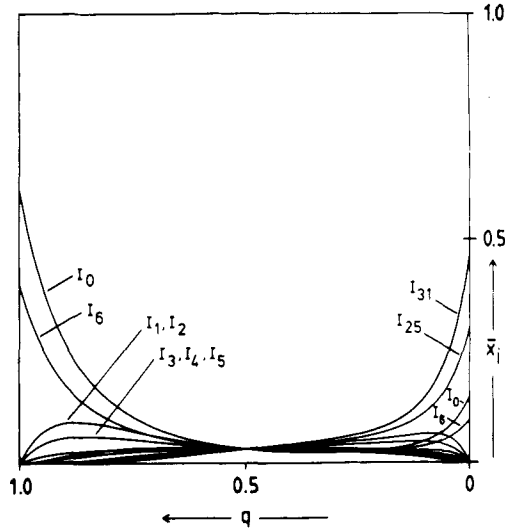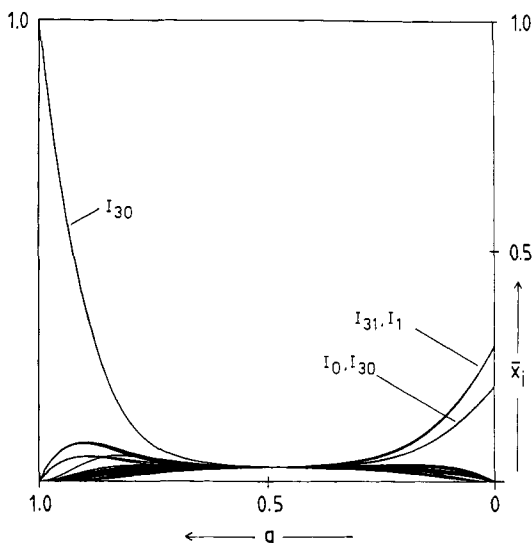
Figure 14. Degenerate quasi-species with $v = 5$. We plot relative concentrations $(\bar{x}_i)$ as functions of single-digit accuracy of replication (q). Two degenerate master sequences I_0 and I_6 are of Hamming distance $d(1, 6) = 2$ in this example—for numbering of sequences see Figure 13. Rate constant: $A_0 = 10$, $A_1 = A_2 = \cdots = A_5 = 5$, $A_6 = 10$, $A_7 = \cdots = A_{31} = 1$. In limit $q \to 1$ we are dealing with two master sequences, whose concentrations approach certain ratio $0 < \alpha < 1$. Concentration of I_0 is higher since this sequence has more efficient neighbors. In limit $q \to 0$ we find two master pairs (I_0, I_{31}) and (I_6, I_{25}) that show essentially same qualitative behavior as two degenerate master sequences. They have Hamming distance $d = 2$ as well.

different selective values, and there the more efficient complementary pair (I_0, I_{31}) wins.

The third example (Figure 16) behaves in a very similar manner to the second in the range of direct replication. But the rate constants have been changed such that we now have two degenerate pairs of complementary sequences in the limit $q \to 0$, (I_0, I_{31}) and (I_1, I_{30}). These two pairs have a Hamming distance $d = 1$, and we expect equal concentrations of I_0 and I_{30} and of I_1 and I_{31}, respectively. It is interesting to note that these concentrations remain almost the same nearly for the whole domain of complementary replication.

Our last case (Figure 17) is an example of "complicated behavior" where the internal structure of the quasi-species is completely reorganized at some critical value of q. In the limit $q \to 1$ we are dealing with two nearly degenerate, distant sequences: $A_0 = 10$ and $A_{50} = 9.9$. Accordingly, we observe selection of the more efficient sequence I_0. The sequences in the neighborhood of I_{50},
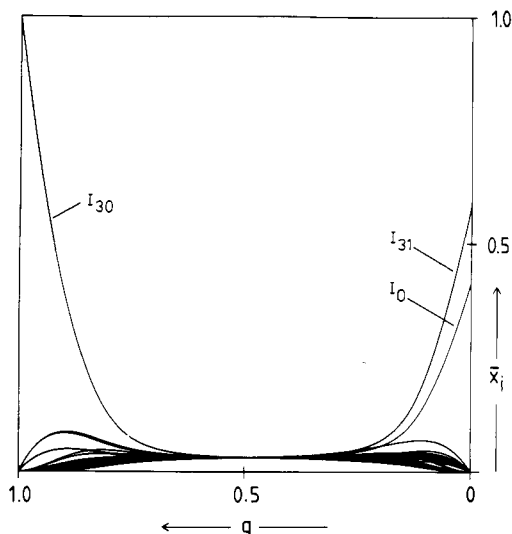
Figure 15. Degenerate quasi-species with $v = 5$. Plot is analogous to that shown in Figure 14. Rate constants A_k, in vectorial notation, are $\mathbf{A} = (10, 5, 5, 1, 1, 1, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 5, 2, 2, 1, 2, 2, 2, 5, 2, 5, 5, 2, 2, 1, 10, 5)$. Two master sequences I_0 and I_{30} have Hamming distance $d(0, 30) = 4$. In limit $q \to 1$ sequence with more efficient neighborhood, I_{30} is selected. In limit $q \to 0$ we observe interesting phenomenon that both sequences are present in equal concentrations since two master pairs (I_0, I_{31}) and (I_{30}, I_1) are degenerate and have Hamming distance $d = 1$.

however, were chosen to be more efficient in replication than those in the neighborhood of I_0: $A_1 = 1$ and $A_{49} = 2$, respectively. Around the point $q \simeq 0.964$ the influence of the neighboring sequences becomes so strong that the less efficient sequence I_{50} outgrows I_0 and dominates the quasi-species. Hence for special choices of rate constants a situation where the most abundant sequence in the quasi-species is different from that with the highest selective value can occur. One may notice that this also means a new concept of neutrality that does not depend on simple degeneracy of the fitness values of two sequences but rather rates the complete fitness topography.

3. Conformation-dependent Value Functions and Fitness Landscapes

As outlined in previous sections, much depends on the properties of physically realistic sets of replication rate constants, despite the comparative insensitivity of error thresholds to details in their distributions. At present it is not feasible to measure or estimate real-valued landscapes empirically, not even in the most simple experimental systems like RNA replication in the $Q\beta$
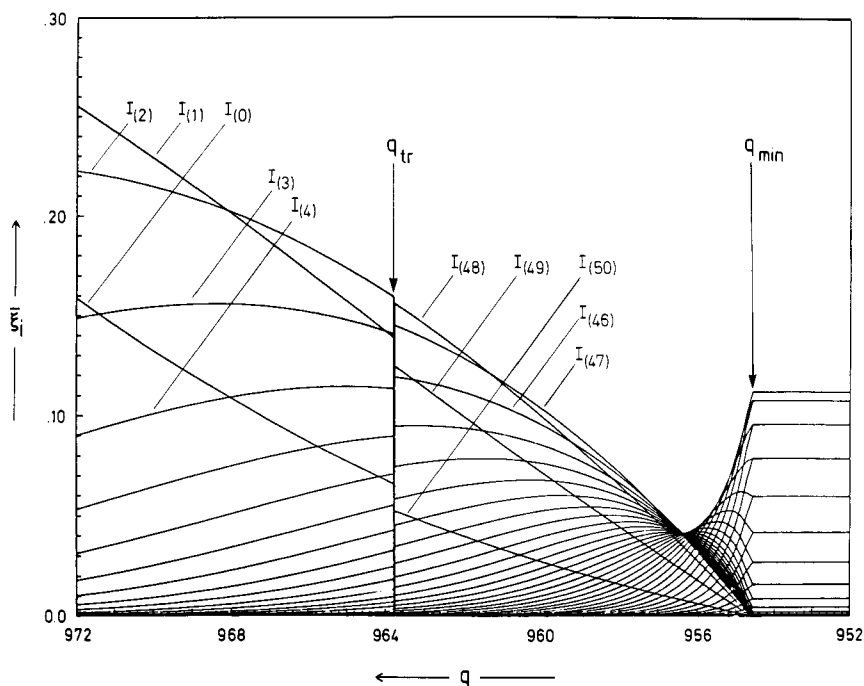
Figure 16. Degenerate quasi-species with $v = 5$. Plot is analogous to that shown in Figure 15. Rate constant of I_6 is reduced to $A_6 = 2$. Situation in limit $q \to 1$ is qualitatively same as that reported in Figure 15: Sequence with more efficient neighborhood, I_{30}, is selected. But in limit $q \to 0$ we are not dealing with degenerate master pairs any more: More efficient pair (I_0, I_{31}), which does not contain sequence I_{30}, wins.

system [10–12]. In this section we present some results of computer calcula-tions on model fitness landscapes. In order to render the model tractable, simplifications are inevitable. Binary sequences I_k are considered instead of natural four-letter RNA sequences since this makes the algorithms much faster. We choose chain lengths of $v \approx 70$ and compute unknotted two-dimensional structures of folding patterns, $\mathscr{F}_k = \mathscr{F}(I_k)$, by means of a folding algorithm. Details are given in ref. 53. Any folding pattern of this kind is decomposed uniquely into ordered combinations of three structural ele-ments:

1. double helical regions or stems containing "base pairs", which consist of complementary symbols on opposite strands, $0 \cdots 1$ or $1 \cdots 0$, re-spectively;

2. loops or reverse turns through which the string folds back on itself in order to form a stem; and

3. other unpaired regions of the sequence.

Folding, in essence, is based on a thermodynamic minimum free-energy criterion. The base pair is the most important stabilizing element of the

Figure 17. Example of almost degenerate quasi-species. For system of binary sequences with chain length $v = 50$ relative stationary concentrations of 51 mutant classes ($\bar{\xi}_i$) are shown in region $0.972 > q > 0.952$. Notice extremely sharp transition at $q_{tr} = 0.9638$ when new master sequence $I_{(50)}$ and its neighbors become dominant. At $q_{min} = 0.9546$ usual error threshold observed. Selective values used in this example are $A_{(0)} = 10$, $A_{(50)} = 9.9$, $A_{(49)} = 2$, and all other $A_{(k)} = 1$ for $k = 1, \ldots, 48$ (cf. figure 5).
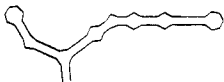
pattern—more precisely there are two major contributions, base pairing and stacking of base pairs aligned in parallel planes—and hence the thermodynamic stability of the secondary structure is roughly proportional to the number of base pairs. In addition there are destabilizing contributions from loops due to their constrained configuration compared with other unpaired elements (element 3). The individual steps of the computer simulation are outlined in Table III.

The folding patterns or conformations, $\mathscr{F}_k$, are evaluated according to certain rules. Here we present three different examples of criteria for the rating of patterns:

1. similarity of folding patterns in comparison to a target structure,
2. thermodynamic stability of folding patterns, and

TABLE III
Computation of a Model Value Landscape

System	Realization	Process
Digit	$\{0, 1\}$	
	$\Downarrow$	*Ligation to String*
Binary Sequence String of $\nu \approx 70$ Digits	$\{2^{70} \approx 10^{21}$ Sequences$\}$	
Primary Structure	$I_k :\ 011000110100\ldots0$	
	$\Downarrow$	*Folding through Base Pairing*
Secondary Structure	$\mathcal{F}_k(I_k) :$	

	$\Downarrow$	*Evaluation*
Molecular Properties	$A_k,\ D_k\ ;\ W_{kk}$	

3. evaluation of replication and degradation rate constants according to
   model assumptions based on the folding patterns.

In the first case a measure of similarity between the conformation
$\mathcal{F}_k = \mathcal{F}(I_k)$ and a predetermined "ideal" reference conformation $\mathcal{F}_0$ was
employed to obtain selective values that depend on the folded conformations.
In determining this function both the specific sequence *and* its folding play a
role. We investigated several joint dependencies. At first it is important to
obtain a workable similarity measure. This was achieved simply by encoding
the two-dimensional folding pattern as a new sequence, the fold sequence f,
which is a string built from symbols $(0, 1, 2)$. Starting from the 5' end, as usual,
a 0 is assigned to a monomer if it is unbound, a 1 if it is part of the first strand
of a double-helical region, and a 2 if it is part of the second strand. This
suffices to determine the conformation because of the no-knots constraint.
The map $I_k \rightarrow \mathcal{F}_k = \mathcal{F}(I_k) \rightarrow f(\mathcal{F}_k)$ is unique, and we define

$$f_k \doteq f(\mathcal{F}(I_k)). \tag{IV.5}$$

Note that the inverse is not true: Several sequences I_j may lead to the same
fold sequence f_k. A suitable measure of similarity of two folding patterns $\mathcal{F}_1$

and $\mathscr{F}_2$ is now just the Hamming distance between the two fold sequences $d(f_1, f_2)$, as defined in Section II.1. In fact, a more general distance function involving insertions and deletions [54] could be used, but the Hamming distance will suffice for the present discussion.

The most straightforward measure of the fitness of a conformation is to identify the dissimilarity from the reference conformation as an activation barrier for replication. Two such fitness assignments were applied: the first, indicated by superscript 1 in parentheses, is of a global nature, and it is based on the Hamming distance of fold sequences; the second, denoted by superscript 2 in parentheses, in addition evaluates similarity of structurally defined subsequences. Implicitly, we assume equal degradation rate constants in this model. Degradation then contributes just an additive constant to the selective values, and we may put $D_k = 0$ without losing generality. For the first assignment

$$A_k^{(1)} = A_0 \exp[-\alpha d(f_0, f_k)] \qquad \text{(IV.6)}$$

is the rate constant of replication of the sequence I_k whose fold sequence is f_k. The reference fold sequence is f_0 here, and α is a real, positive constant determining the precision with which the reference fold sequence must be approximated for a given fitness to be achieved. The rate constant of replication of the reference sequence I_0 is A_0. This simple fitness function has a maximum of $A_k = A_0$ when $f_0 = f_k$, that is, when the sequence I_k has exactly the folding pattern specified by the fold sequence f_0. Because of the nontrivial relationship between sequence I_k and folding pattern or conformation (many sequences may yield the same conformation on folding), Eqn. (IV.6) sustains many local optima.

An additional dependence on segments of the base sequence defined relative to the conformation (such segments are exposed by certain loops, stems, branching points, etc.) may be introduced via the Hamming distance from reference segments in order to model more highly constrained fitness functions. Then the sequence plays a specific role relative to the conformation, and we obtain a second fitness function:

$$A_k^{(2)} = A_0 \exp\left\{-\alpha_1 d(f_0, f_k) - \alpha_2 \sum_{i=1}^{n} d(s_i(f_0), s_i(f_k))\right\}, \qquad \text{(IV.7)}$$

where the segments of the sequence I_k are denoted by $s_i(f_k)$ and those of the reference fold sequence by $s_i(f_0)$; α_1 and α_2 are positive constants and A_0 is the replication rate constant of the reference sequence I_0. We also found it of interest of deal directly with the linearized forms of Eqs. (IV.6) and (IV.7),

which radically changed the shape of the distribution of replication rate constants as well as the effective superiority of the wild type.

We compared the linearized form Eqn. (IV.6) with the long-range Ising-model-like fitness function (cf. the Hamiltonian in Section III.5)

$$A^{(s)} = \bar{w} + \sum_i h_i \sigma_i + \sum_{i<j}^{v} J_{ij} \sigma_i \sigma_j \qquad (\text{IV.8})$$

by generating for a larger number (10,000) of randomly chosen sequences the average values of the folding-based fitness function $A^{(1)}$ with

1. no sequence positions held fixed, $\bar{w} = \langle A^{(1)} \rangle$;
2. one sequence position k held fixed at σ_k,

$$h_k \sigma_k = \langle A^{(1)} | k \rangle - \bar{w};$$

   and

3. two sequence positions k, l held fixed at σ_k, σ_l,

$$J_{kl} \sigma_k \sigma_l = \langle A^{(1)} | k \rangle - \bar{w} - h_k \sigma_k - h_l \sigma_l.$$

In this way the parameters $\bar{w}$, h_i, J_{ij} were deduced from the form of $A^{(1)}$. Using these values, the value of $A^{(s)}$ in Eqn. (IV.8) could be compared with $A^{(1)}$, and no significant correlation was found. The conclusion is that the folding fitness function cannot be decomposed into pairwise contributions from sequence positions. This was confirmed by the fact that the values of h_i, J_{ij} calculated using the different choices of σ_i, $\sigma_j = \pm 1$ gave uncorrelated results. Similar results pertain to $A^{(2)}$ in linearized form (and, of course, for the exponential versions).

In order to gain more insight into the nature of the fitness functions (IV.6) and (IV.7) as relevant for quasi-species structure, we employed a variety of representational tools:

1. The probability density for the varying fitness of sequences ($\sim 20,000$, chosen at random from the 2^v possible with length $v = 72$) was calculated and is displayed in Figure 18 for the linearized form of Eqn. (IV.6) (with $\alpha = 1$). The density appears as a slightly skewed Gaussian about the mean distance from the goal fold. In general, other effects (including stochastic selection) require a truncation of the distribution above the generic or modal value when the fitness function has an exponential or other nonlinear (cooperative) form.

2. The correlation function reveals much more of the special nature of conformation-dependent fitness. However, in this case, just choosing a large
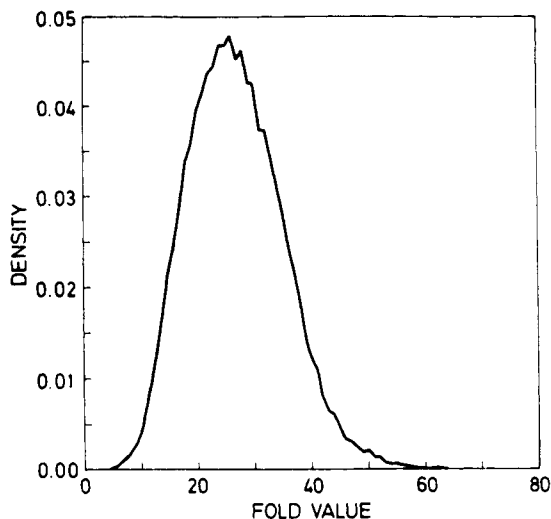
Figure 18. Probability density of folding fitness function. Curve is smoothed frequency plot showing probability density for integer-valued fitness function $A^{(1)}_{\text{linear}}(s) = v - d(f_0, f(s))$, for 21,610 sequences of length $v = 72$ (chosen at random). Here $d(f_0, f(s))$ measures distance between calculated folded form and reference fold f_0, which was taken to be particular nested pair of hairpin loops.

number of sequences at random can only give very well-separated sequences (Hamming distance) and does not provide information about shorter range correlations. We solved this sampling problem by examining sequences along v random mutation pathways from a particular centre sequence. The resulting correlation functions when the center sequence was a local optimum for the two linearized value functions (using the same set of 72 mutant pathways of length 24) are shown in Figure 19. The important features are finite range and an approximately exponential form for both functions. Typical individual trajectories for the two fitness functions are shown in Figure 20.

3. The preceding results demonstrate that along a random mutant pathway the typical situation is a rapid decrease in the fitness level of sequences. However, the issue of the rare occurrence of long extended high-value ridges is left open. Indeed, one would like to know whether the high-value regions are isolated or whether they form a sparse but interconnected network through the mutant space? For the fitness function of type (IV.6), only those monomers directly involved in bonding (and a few additional ones near the boundaries of binding regions) were found to be important in determining fitness near an optimally folded sequence. Constraints on the remaining monomers are limited, and the picture of selection restricted to a roughly $v/2$
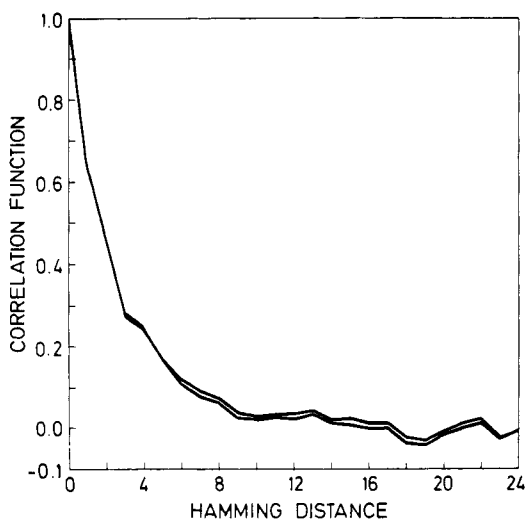
Figure 19. Correlation functions for fitness by folding. Correlation functions for linearized forms of folding fitness functions of type 1 [Eqn. (IV.6), upper curve] and type 2 [Eqn. (IV.7), lower curve] are displayed. Both are normalized by value at local optimum at distance 0. Seventy-two random paths radiating outward 24 shells and starting with distinct first mutants employed. Same paths and reference fold employed in both cases.

dimensional subspace of the ν-dimensional sequence space is qualitatively correct. For the fitness function of Eqn. (IV.7) the situation is different as the degeneracy has been removed. Here a plot of neutral mutants near a local optima reveals long ridges extending into the surrounding sequence space (Figure 21). This confirms, for the present conformational-dependent fitness function, the features that distinguish evolution in the quasi-species as guided in the choice of mutants and not purely random (cf. Section II.5).

The second model evaluation of folding patterns is based on the thermodynamic stability of polynucleotide conformations. Here, not only does this stability decide which of the many possible structures is formed (as in the preceding) but it also determines the fitness. The more stable is the optimal folding pattern $\mathscr{F}_k$ for a particular sequence, the higher is its selective value:

$$A_k^{(3)} = A_0 \exp(-\alpha \Delta(\Delta G)) \approx A_0(1 - \gamma \Delta(\Delta G)) \qquad \text{(IV.9)}$$

with $\Delta(\Delta G) = \Delta G(\mathscr{F}_k) - \Delta G_0$ and α and γ being some real and positive constants. The absolute minimum free energy of the folding patterns for all sequences of the given length, denoted by ΔG_0, is chosen as reference here. In
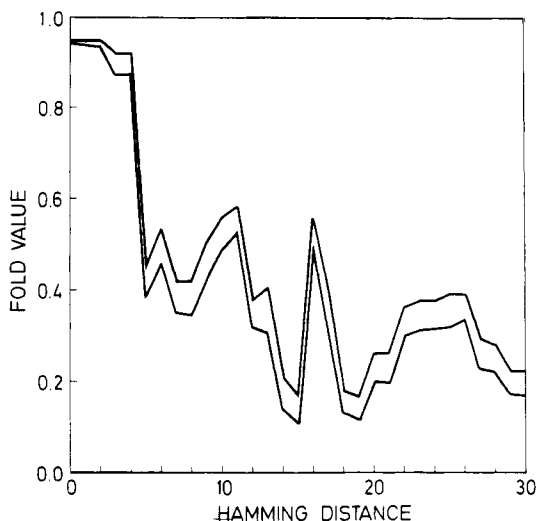
Figure 20. Values along radiating trajectory. Folding values for sequences lying on parti-
cular mutant path randomly radiating from locally optimal sequence. Linearized fitness func-
tions of type 1 (upper curve) and type 2 (lower curve) employed as in Figure 19. Different paths
yield very different jagged profiles.

contrast to the previous example, there is no unique reference conformation
$\mathscr{F}_0$ because several folding patterns may have the lowest possible free energy.
Since we are interested in qualitative features, we only use the linear approxi-
mation as indicated in the preceding. The basis for approximate free-energy
estimates was worked out during the past two decades (see, e.g., ref. 55), and
we are now in a position to assign energy increments to the basic sub-
structures. In essence, these increments are based on the number of base pairs
and their orientations. Here we use the energy weights for all (G, C) sequences
as tabulated in ref. 56 and assign G to the digit 1 and C to 0, respectively.
Again we assume that all degradation rate constants are equal, and we put D_k
$= 0$ without losing generality. This fitness assignment to folding patterns, as
for the previous one, is manifestly incomplete with respect to biophysical
insight into replication kinetics since higher stability or smaller dissimilarity
to a reference pattern do not alone entail faster replication. From the
viewpoint of the mapping $I_k \rightarrow \mathscr{F}_k \rightarrow (A_k, D_k)$, however, we are dealing with a
problem of the same "*universality class*" [57] as in case of the physically more
appealing assignment of the third model. Moreover, we have the advantage of
knowing a priori what the fittest folding patterns look like: Obviously these
are the 32 (inverse) palindromes with complete parallel stacks and variations
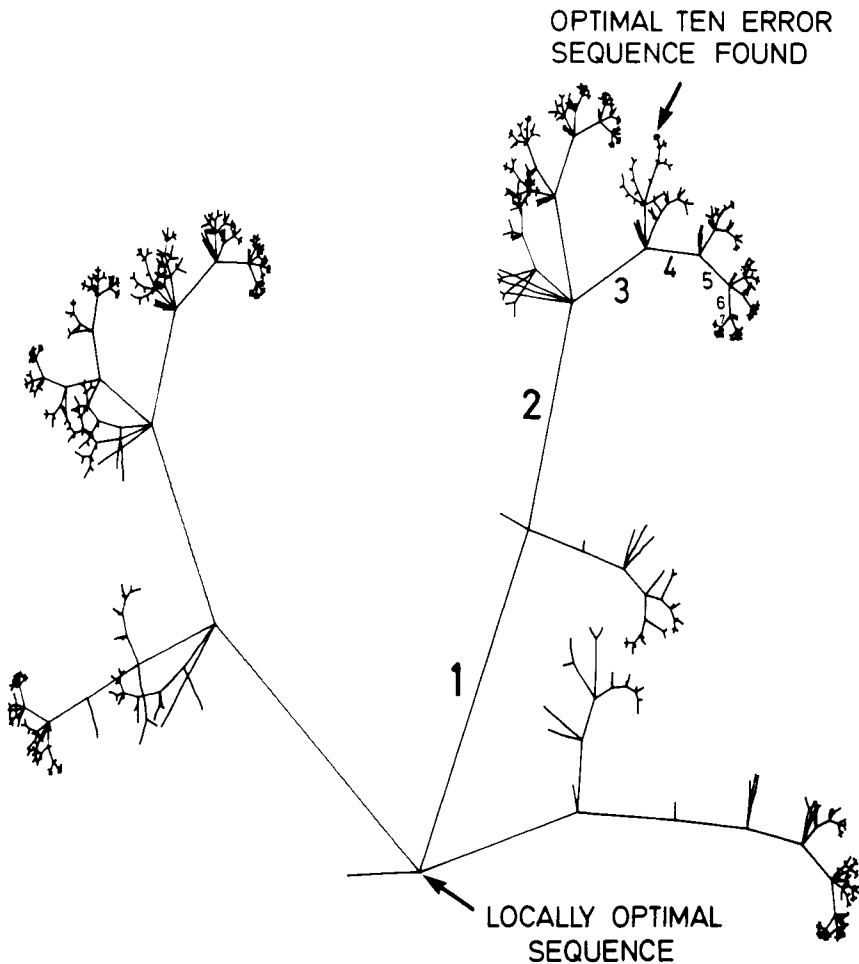
Figure 21. Mutant space high-value contour near local optimum. Diagram is multiply branched tree with different macromolecular sequences at vertices. Each line joins neighboring sequences whose values are within 0.5 of locally optimum sequence at lower center for linearized fitness function of type 2 [Eqn. (IV.7)] and reference fold that is cruciform, like tRNA, for sequence of length 72. Over 1300 branches shown extending up to 10 mutant shells away from central optimum. Better sequence (labeled optimum) was found in tenth mutant shell. Nonrandom sampling of mutant sequences demonstrated typical of population sampling in quasi-species model. Note small number of ridges that penetrate deeply into surrounding mutant space. (Additional connected paths due to hypercube topology of mutant space not shown.)

only in the four digits of the hairpin turn that form the most stable, "perfect" hairpin.

Since the third model is closely related to the second, we present its underlying assumptions first and discuss the results obtained for both value landscapes thereafter. In contrast to the two previous model landscapes we calculate both replication and degradation rate constants independently. Since up to now nobody knows how to compute replication rate constants from folding patterns of single-stranded polynucleotides, our model is essentially of heuristic value. It just tries to meet some biophysical constraints of replication and centers around the fact that virus-specific RNA replication operates on single strands [10] and that unzippering of double helical regions is cooperative [58]. We assume that every stacking region slows down the overall replication process in an additive fashion and that the term each stack contributes is a sigmoidal function that is reminiscent of the Monod–Wyman–Changeux model:

$$A_k^{(4)} = A_0 - A_\infty \sum_j \frac{v_j^{(k)}(1 + v_j^{(k)})^3}{(1 + v_j^{(k)})^4 + \mathscr{L}}. \tag{IV.10}$$

The sum is to be taken over all stacking regions of the folding pattern $\mathscr{F}_k$; the $v_j^{(k)}$ values refer to the individual stacks; in particular, $v_j^{(k)}$ denotes the number of base pairs in the jth stack of the folding pattern $\mathscr{F}_k$. The parameter $\mathscr{L}$ is some large positive constant that determines the detailed shape of the sigmoidal curve. The parameters A_0 and A_∞ represent rate constants: A_0 is the replication rate constant of the linear unstacked chain that serves as the reference here. The difference $A_0 - A_\infty$ refers to an infinitely long, completely stacked conformation. Should a folding pattern yield a negative rate constant ($A_k < 0$), we would declare the sequence I_k as lethal and put $A_k = 0$. The two parameters A_0 and A_∞ may be used to control the fraction of lethal mutants.

The rate constants of the hydrolytic degradation are also made up of additive contributions—here from unpaired regions. Heavier penalties are assigned to free ends or joins than to loops. No cooperativity is assumed:

$$D_k^{(4)} = D_0 + D_l \sum_j \frac{l_j^{(k)}}{l_m} \exp\left[(l_j^{(k)} - l_m)/l_m\right] + D_u \sum_i \frac{v_i^{(k)}}{v}. \tag{IV.11}$$

The first sum is taken over all loops; the second over all other external elements: tails, joins, and so on. Three rate constants D_0, D_l, and D_u are introduced to determine stability against hydrolysis. The number of unpaired bases in the jth loop of the folding pattern $\mathscr{F}_k$ is denoted by $l_j^{(k)}$, the length of the ith unpaired element in the same pattern by $v_j^{(k)}$; l_m is a weighting factor and v the length of the sequence.

The selective values or fitness factors are readily obtained from the equation $W_{kk} = A_k q^v - D_k$. This simply introduces a weighting of replication against stability by means of the single-digit replication accuracy. For high accuracy it is better to replicate faster, whereas longevity or high stability against hydrolysis pays more at high error rates. The third model is an example of an assignment in which high fitness is based on a compromise between two contradictory trends: Long double-helical segments stabilize against hydrolysis, but they also reduce the rate of replication and vice versa. Such contradictory trends seem to be the rule in real biological systems.

In order to characterize the distributions of selective values in the second and the third model, we explored the value landscape by a Monte Carlo search. We created three random samples of 38,000 different sequences each (one repeat with 76,000 sequences gave essentially the same results) with predetermined ratios of probabilities for (0/1) digits, $\rho_1 = 0.2857$, $\rho_2 = 0.5$, and $\rho_3 = 0.7143$, which led to mutant distributions centered at the 20-, 35-, and 50-error mutants of the *all-zero* sequence I_0. Three different parts of the value landscapes determined by Eqs. (IV.9)–(IV.11) were explored in that way. The results are shown in Figure 22.

The value lanscapes created by the second and third model differ significantly. The selective values of the simple stability assignment (IV.9)—here computed for $q = 1$—show roughly a "noisy" Gaussian distribution. The mean selective values for ρ_1 and ρ_3 are about the same: $\bar{W} \approx 500 \ [t^{-1}]$; selective values are rate constants by definition and are therefore given in reciprocal time units. The distribution around the 35-error mutant of I_0, obtained with ρ_2, is significantly superior: $\bar{W} = 750 \ [t^{-1}]$. The interpretation of this finding is straightforward: The 35-error mutants have as many digits 0 as digits 1 and, provided the sequence admits it, form more base pairs than do the 20- or 50-error mutants. Thermodynamic stability in essence counts the numbers of base pairs, and therefore the 35-error mutants of I_0 are the most stable on average.

The selective values for the third model are shown for $q = 1$ in Figure 22 and hence represent excess productions, $E_k = A_k - D_k$. The distributions around the 20- and 50-error mutants of I_0 show interesting bimodal shapes. Further analysis shows that this bimodality is a result of subtle relations between the A_k and D_k according to Eqs. (IV.10) and (IV.11), and we shall not pursue this issue further here. It is useful to split the value landscape of the third model into different contributions resulting from replication and degradation rate constants (Figure 23). The distributions of the degradation rate constants have much in common with the thermodynamic stabilities used in the second model: They also resemble noisy Gaussian distributions. The distributions of replication rate constants, however, differ rather drastically from the others. They look much more bizarre. Due to the sigmoidal
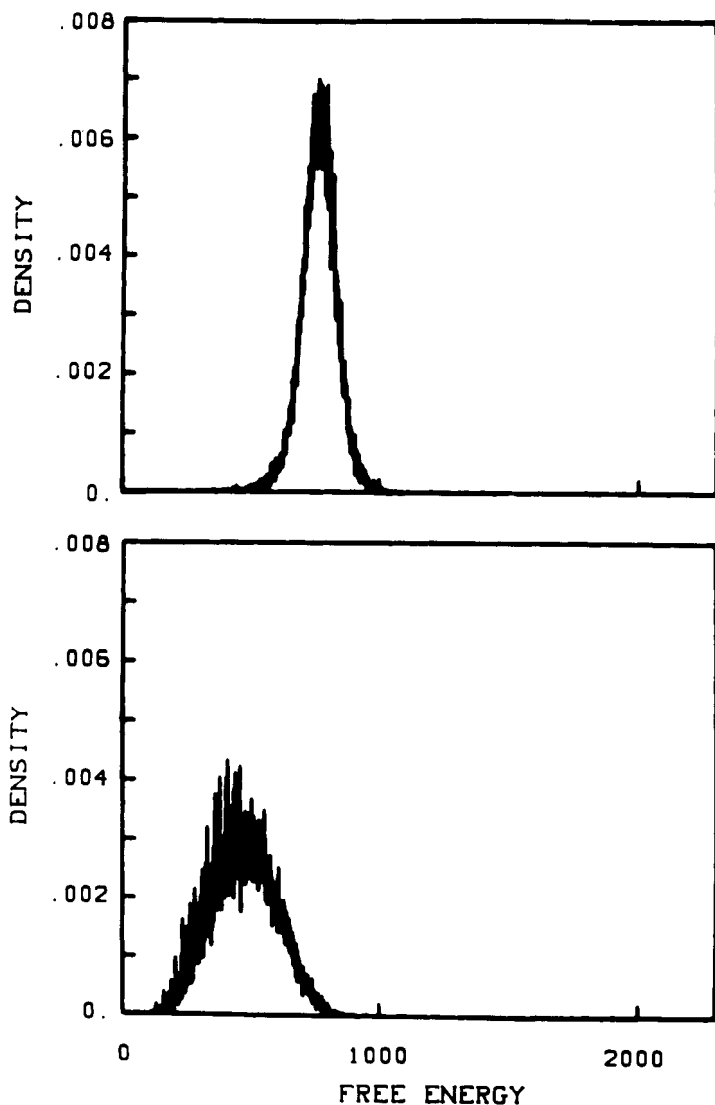
Figure 22. *(Continued overleaf)*

contribution of the lengths of double-helical regions, replication rates are extremely sensitive to minor details of the folding patterns, and this leads to enormous scatter in the distribution. Long stems are a disadvantage for replication, and hence, 20- or 50-error mutants of the reference sequence I_0 replicate faster on the average than do the 35-error mutants. The fraction of nonviable sequences in the sample centered around Hamming distance 35
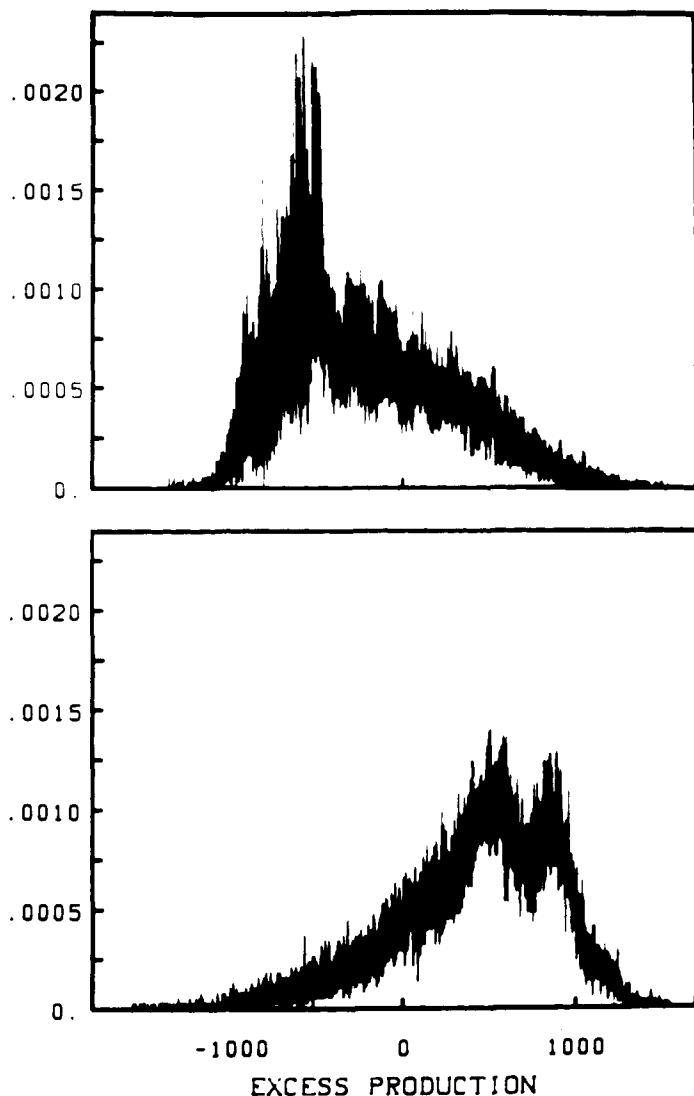
Figure 22. Distribution of selective values in sequence space. Here 38,000 different sequences of length $v = 70$ generated by introducing digits 1 at random with probability $\rho_1 = 0.2857$, $\rho_2 = 0.5$, or $\rho_3 = 0.7143$ into all-zero sequence I_0. This produces Gaussian shape samples centered around 20-, 35-, and 50-error mutants of all-zero sequence. Distributions of free energy $\Delta G(\mathscr{F}_k)$ and excess productions $E_k = A(\mathscr{F}_k) - D(\mathscr{F}_k)$ shown for regions located at mean Hamming distances 35 (upper plots) and 20 (lower plots) from I_0. Apart from scaling factor free-energy distribution constitutes fitness landscape of second model evaluation. Distribution of excess productions is representative for model 3 at high accuracy of replication. Densities in neighborhood of 50-error mutants (not shown) are essentially same as those around 20-error mutants.

Figure 23.   (*Continued overleaf*)
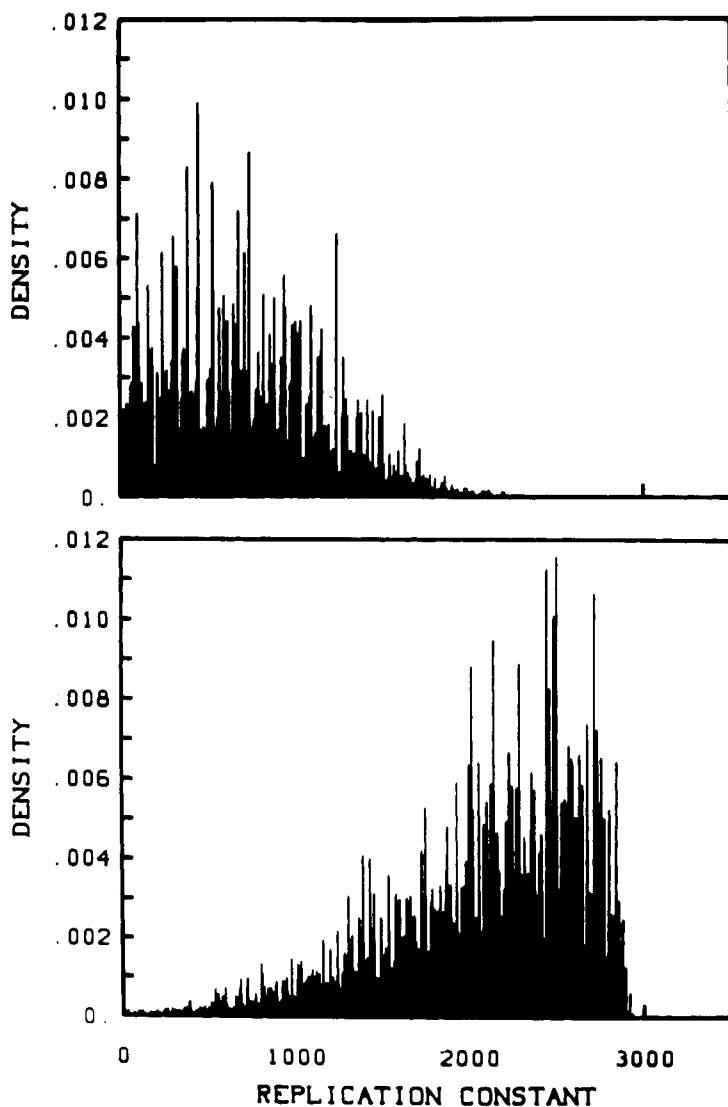
from I_0 is therefore higher than in the other two regions. It amounts to 0.18. As we mentioned in the preceding, the fraction of lethals is a consequence of the choice of A_0 and A_∞ in Eqn. (IV.10) and hence arbitrary. But, we believe this reflects the well-known, naturally arising situation where parts of the sequence space are particularly rich in lethal mutants.

0   1000   2000   3000

DEGRADATION CONSTANT
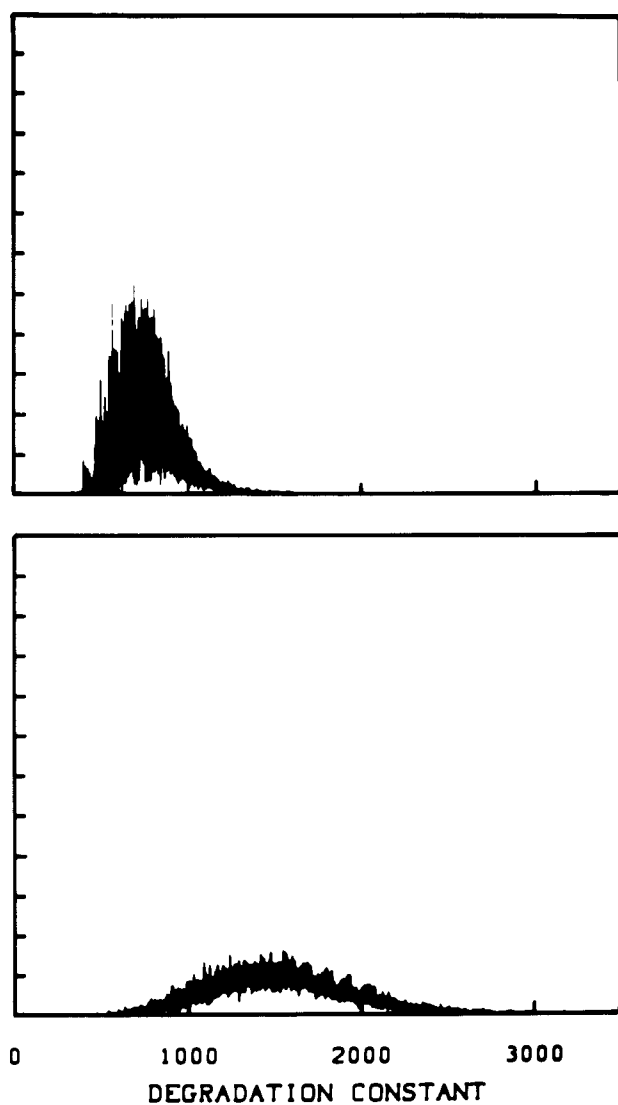
Figure 23. Partitioning of excess productions $E(\mathscr{F}_k)$ into replication (A_k) and degradation (D_k) rate constants. Upper and lower plots show distributions sampled from neighborhoods of 35- and 20-error mutants, respectively, as explained in the caption to Figure 22. Fraction of lethal variants in replication landscape A (this is the fraction of sequences with $A_k = 0$); it amounts to 0.18 in this particular case; has been cut off in order to show details of distribution at positive rate constants.

The overall scans of value landscapes presented in Figures 22 and 23 reveal only general trends persisting in certain regions. Nothing is said about the relative location of high-valued configurations and their connectivity. Investigations aiming at a better understanding of the topological details of the value landscapes of the second and third model, similar to those shown in Figure 21, are presently under way. A simple hint on this topology can be easily obtained, however, by exploring the local shape of the fitness surface. For this goal we calculated selective values of all sequences I_k surrounding some reference sequence I_j with Hamming distance $d(k, j) = 1$. The local scan through sequence space reflects the bizarre structure of realistic value landscapes (Figure 24). Nearby sequences may have very small, almost the same or higher selective values compared to a reference sequence of average properties. From molecular genetics we know that this is also the case in nature. A point mutation might be lethal or harmless since the translation into phenotype may depend crucially on single digits.

4. Asymmetry of Fitness Landscapes: Apparent Guidance of Evolution

The appearance of a mutant moderately distant from the wild type depends on the population of precursor mutants, that is, mutants that are closer to the wild type. It will be seen that this introduces an element of determinism into the generation of mutants despite the fact that the elementary process of mutation is of an entirely stochastic nature. "New" mutants always appear at the periphery of the populated mutant spectrum, which will be shown to be highly asymmetric with respect to the wild type, in a manner depending on the structure of the value landscape. Along ridges, the populated mutant spectrum may reach to regions quite distant from the wild type. Moreover, these are mountainous regions, that is, regions where, due to the correlated structure of the landscape, advantageous mutants are most likely to appear. Hence, the generation of mutants is a process that is highly biased toward success and, presuming some knowledge about the value landscape, appears to be by far less unpredictable than has usually been assumed in population genetics. In the light of the examples discussed thus far, this statement does not sound too revolutionary. However, more quantitative estimates will show how effective this "value guidance" is, turning it almost into a new principle rather than just some quantitative modification of evolutionary theory.

In Figure 5 two landscapes are shown that start at a binary master sequence ($v = 50$) and are followed up to the 12-error copies. One of the landscapes (left side) is the low-value plateau that has been considered already in previous examples, while the other resembles a mountain saddle as typical for any fractal type of hill country. The population numbers of mutants, relative to the population number of the wild type, were calculated by means of second-order perturbation theory [Eqs. (II.17) and (II.18)] for
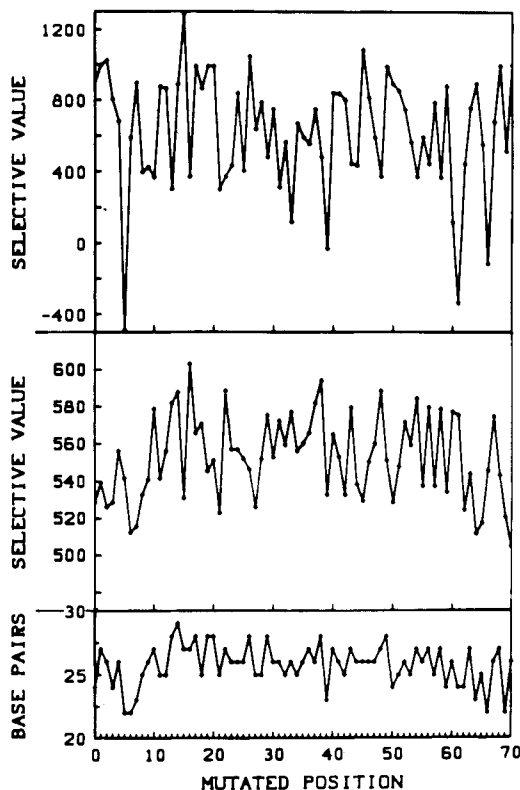
Figure 24. Local shape of value landscapes. Selective values as defined by Eqs. (IV.9)–(IV.11) of 70 nearest neighbors surrounding a given reference sequence I_R are shown. Upper curve refers to evaluation according to third model for $q = 1$ as in Figure 22. Middle curve represents free energies according to second model evaluation. Lower curve counts numbers of base pairs in different folding patterns $\mathscr{F}_k$. Free energy follows approximately number of base pairs, but excess production shows roughly opposite trend.

two cases: (i) The profile shown applies to only one individual chain of mutants up to the 12-error copy, all other mutants having a zero fitness value; and (ii) the profile applies to all precursors of the 12-error mutant, that is, all mutants directly between the wild type and a given 12-error mutant.

The results are very instructive. For the low-value plateau the probability of mutant appearance is essentially Poissonian for the single-mutant chain (i) and only slightly modified for the all-intermediate mutant plateau (ii), indicating that for typical population numbers of 10^{10}–10^{15} any given mutant that has a Hamming distance from the wild type larger than 4–5 is rarely populated. In contrast, the mountain saddle landscape shows probabilities

that suggest in both cases a deterministic population for all mutants under consideration. In particular, the 12-error mutant here—supposing the extended landscape (case ii)—appears by 15–20 orders of magnitude more frequently than in the case of the low-value plateau, suggesting that the evolutionary process would be deterministically guided to the mountainous region surrounding the 12-error mutant, in which a more advantageous new mutant might appear.

In order to appreciate this effect adequately, we must be aware that a binary sequence of length 50 has $\binom{50}{12} \doteq 10^{11}$ twelve-error copies and hence correspondingly many (overlapping) precursor regions, among which all sorts of value profiles are to be expected. The consequence of value guidance is that large areas of sequence space will never be touched by the evolutionary process. Thus the trial-and-error process of mutation is concentrated on those regions in which the advantageous copy is most likely to appear. The large orders of magnitude involved offer a most efficient solution of the complexity problem that initially instigated those who have made quantitative estimates on mutant generation. What counts is not the number of possible mutants (which generally is hyperastronomically high) but rather the value topography of sequence space, that is, the presence or absence of sufficiently branched mountain ridges that pervade major parts of sequence space. They guarantee the teleonomic nature of evolution that was hard to imagine with a "blind filter" of selection acting upon the trial and error of mutation.

It is also obvious that the problem of neutral mutants appears in a new light. Neutral mutants are not restricted any more to the chance coincidence of degenerate wild types that differ in their relative fitness by less than the reciprocal population size. Mutants with a quite broad degree of variability of fitness are potential candidates when rated together with the fitness topography of their neighbors. Only in small populations of species of large genomes—that is, only where every mutant really is unique—would the stochastic trial-and-error model apply. Even in these cases, the long-term behavior would depend on the fitness topography of the kinship neighborhood of the wild type. The deterministic model, of course, represents a limiting case that has to be complemented by stochastic approaches (cf. Section V.5), just as the random-drift model of Kimura represents another limiting case.

V.   CONCLUSION: REVIEW AND OUTLOOK

1.   The Physical Basis of the Model

In this Chapter we have been dealing with the simplest model of a self-reproductive molecular system consistent with known [59–61] kinetic prop-

erties of RNA or DNA: the (molecular) quasi-species. In so far as they satisfy the prerequisites we have assumed, the model may well apply to more complex systems dealt with in molecular biology, including viruses [62, 63] or (vegetatively reproducing) cellular systems that we may collectively subsume as systems of individual "replicators." The model in its simple form does not apply to networks of interacting replicators, but it may be extended so as to include such properties. It is also not intended in this review to describe any realistic scenario of the early phase of the origin of life. These were "historical" rather than truly physical problems, requiring the knowledge of historical initial and boundary conditions. Although we may well guess about the general chemical nature of prebiotic earth, we entirely lack detailed knowledge about particular environmental conditions such as the presence of specific catalysts, be they of chemical (e.g., molecular complex) or of physical (e.g., interfacial) nature.

Hence we are dealing with a conditional model that becomes relevant wherever its prerequisites are fulfilled. Moreover, the model provides a framework for designing experiments dealing with (molecular) biological evolution that can be conducted under defined laboratory conditions.

Being one of the simplest molecular models that display Darwinian properties, it provides some features that may be taken as generally characteristic of Darwinian self-organization. We do not pretend that this simple model applies to all situations for which Darwinian reasoning has been used. This model, for instance, neglects any horizontal exchange of information such as regular and irregular crossing over, gene conversion, transposition, and so on, which are typical of higher organisms. Nevertheless, the rigorous treatment of the model reveals a much more subtle interpretation of what is usually referred to as Darwinian behavior, commonly described as an interplay of stochastic mutation and deterministic selection of statistical significant advantages (including the important modifications brought about by the theory of neutral selection). Evolutionary optimization based on neutral drift and natural selection as a hill-climbing process in the rugged terrain of a value space becomes understandable in a quantitative sense only by the subtleties the model involves. These effects have been tested and identified directly by experiments using molecular replicator systems. We are sure that these effects will carry through to more complex population systems of organisms and therefore must be taken into account for any quantitative description of Darwinian behavior based on natural selection and neutral drift, although they may be complemented by various other phenomena considered to be of importance in population biology, such as fluctuating environments, isolation, migration, and so on, in addition to the recombinative events typical of Mendelian populations.

The quasi-species model rests on essentially three prerequisites:

1. The major constituents of the system, for example, sequences of RNA or DNA, have to be inherently self-reproductive. This is expressed through positive diagonal terms (W_{ii}) in the kinetic equations.
2. Self-reproduction of the constituents is not entirely precise but is rather prone to mutation, being expressed by positive nondiagonal terms W_{ik}. (In other words; species i can come about not only through self-copying but also by miscopying of species k.)
3. The system has to stay far from equilibrium, or better; the turnover of the major constituents is not at equilibrium, where formation and decomposition would be microscopically reversible and detailed balance would hold. This requires a metabolism sustained by a supply of energy-rich monomeric building blocks of the polymeric constituents. (It should be stressed that at equilibrium the first two conditions are not sufficient to cause selective evolutionary self-organization.) As a consequence of microscopic reversibility, all eigenvalues will always turn out to be real and *negative*.

Apart from these three prerequisites, leading to the general form of kinetic equations as represented in Appendix 3, no specific interactions among the constituents—such as mutual enhancement or suppression, recombinative exchange, or other regulative couplings—have been assumed.

2. The Emerging Concepts of Molecular Evolution

The deterministic treatment of the model reveals three major regularities that may be best encompassed by the following terms:

Selection. Local stabilization of a quasi-species distribution in sequence space around one (or several degenerate) master species.

Evolution. Destabilization of the local quasi-species upon arrival of an advantageous mutant that establishes a new quasi-species.

Optimization. Tendency toward global stabilization guided by the population of nearly neutral mutants and their nonrandom distribution in sequence space.

To consider evolutionary optimization as a sequence of discrete selection steps requires some explanation. In fact, it is justified only on the basis of a hierarchical order of the rate terms, in which the off-diagonal terms usually are much smaller than the diagonal ones, decreasing in binomial or Poissonian progression with increasing Hamming distance between template and (erroneous) replica. As a consequence, the deterministic order around an

advantageous master copy, that is, a local quasi-species, is established relatively quickly, while the arrival of a new mutant with selective advantage is a fairly slow process. Such a mutant originates from a precursor mutant at the periphery of the quasi-species distribution, that is, at a relatively large Hamming distance to the master copy, where population numbers relative to that of the master copy are rather small. Strictly speaking, we are dealing here with a stochastic rather than a deterministic event, and the deterministic model at this point needs a stochastic supplement (cf. what follows). However, deterministic population numbers near the periphery of the quasi-species distribution, even for a given error class, may differ by many orders of magnitude depending on the fitness (relative to the master copy) of that particular region in sequence space. A cumulative fitness distribution greatly enhances this effect, which provides a strong deterministic bias for the production of particular mutants at the periphery, preferentially those that are of high fitness or selectively advantageous. This effect provides such a strong guidance toward broad global fitness maxima (the examples given in Figure 5 showing how many orders of magnitude actually are involved) that the classical picture of an evolutionary process as a pure trial-and-error search for advantageous mutants (with its consequently low success probability) turns out to be obsolete. The target of selection is *not* the single advantageous copy, that is, an individual wild type. Instead, it is the total quasi-species distribution that is the target, and this is not of some simple narrow-banded Poissonian type but rather shows a fairly large nonsymmetrical dispersion with protrusions reaching far into the sequence space. These protrusions of the population distribution are found along the ridges of the fitness landscape bridging those areas with fitness values closely resembling that of the wild type.

A tremendous change in the deterministic concept of a Darwinian system is brought about by the quasi-species concept. Let us confront the current interpretation with the classical one.

1. The *wild type*, despite the fact that the whole system may be represented by a unique sequence, is not a single individual but rather a distribution having a defined consensus or master sequence. The individual (or group of individuals) that coincides with the consensus sequence may represent a peak in the population distribution. Its fraction of the total population, however, usually remains fairly small (in cases where it has been tested, even undetectably small).

2. The term *fittest*, correspondingly, is not related to any individual but rather to the complete quasi-species distribution acting as the target of selection. Mathematically, the fittest is characterized by a dominant eigenvector to which all populated states contribute. Only if one sequence clearly

dominates in fitness will its fitness value (diagonal coefficient W_{mm}) determine the maximum eigenvalue to a close approximation.

3. The *population of mutant states* in the quasi-species is strongly modulated by the fitness distribution. The effect is particularly strong for those mutants (i) whose fitness values resemble closely that of the wild type (m) due to the hyperbolic form $W_{ii}/(W_{mm} - W_{ii})$. Moreover, the population number of a given mutant is related not only to the wild-type fitness but also to the fitness values of its neighboring mutants, especially those situated in the subspace referring to all positions that differ from the master sequence. If, in such a domain, all fitness values (or a major portion thereof) are close to that of the master sequence, there will be a tremendous amplification of population numbers relative to domains of low fitness. Optimization routes are to this degree deterministically ordained.

4. *Neutral* or nearly neutral mutants appear in a new light. First, they appear to be of utmost importance in determining the population of states at the periphery of the mutant spectrum and hence in fixing the route of evolution. Second, being part of the quasi-species distribution, they are rated not only by their fitness relative to the master but also with respect to their own mutant environment. The uncertainty of classical theory, how closely the fitness of a mutant has to resemble that of the wild type in order to be called *neutral*, is now replaced by quantitative expressions. It turns out that mutants the fitness values of which deviate considerably more than by $1/n$ (n being the population number) from the wild-type fitness value still may play a very important role, while at the same time true degeneracies due to the rating of the mutant environment become quite rare. As a consequence, in macroscopic populations comprising something like 10^9–10^{12} individual replicators, we usually find only *one* defined consensus sequence, although many (nearly) neutral mutants tend to be populated.

5. *Stochastic theory* must supplement deterministic theory as in the classical wild-type model. This is a formidable task for the complete quasi-species model due to the nonlinear type of relations involved (cf. section 5.).

The results of *classical neutral theory* are valid only for systems of relatively low population numbers and large genomes. If the genome is large enough that even the $3v$ one-error mutants cannot be populated because the population number n is smaller than $3v$, one may expect the results of so-called neutral theory to be representative. Otherwise, modifications due to the reproducible population of (nearly) neutral mutants, as indicated by the deterministic quasi-species model, pertain and finally destroy the basic assumption of the "blind" production of mutants at the periphery of the mutant spectrum.

6. In the limit of large populations and sufficiently small genomes (usually

fulfilled for single RNA or DNA molecules or viruses) we can now give a physical characterization of the terms *selection* and *evolutionary adaptation*.

Selection may be characterized physically as a condensation phenomenon (mathematically; localization of a distribution in sequence space). The analogy to ferromagnetism has been stressed, except that we are dealing with entirely different variables. The role of the Curie temperature is played by an error threshold. Far below the error threshold, the distribution contracts to populate the master species only. This state is determined by the fitness landscape's highest peak rather than cooperative forces, owing to the simple structure of the model. Surpassing the threshold means the "melting" of the quasi-species distribution due to accumulation of errors. Such an error catastrophy means a sharp loss of genetic information. Hence selection is a kind of phase transition in information space. The error threshold has been tested for various virus populations (Table IV) and found to be effective in nature (cf. what follows).

Evolution then may be viewed as a series of stabilizations and destabilizations of quasi-species distributions (i.e., a series of phase transitions) that, in a constant environment, are associated with an increase of fitness. This process is guided by the mutant distribution within each quasi-species. The advantageous mutant appears at the periphery of the populated mutant distribution usually in an area of relatively high fitness. The evolutionary route thus avoids the vast space with a low degree of fitness, and hence optimization proceeds along fairly limited pathways. How discrete the single steps are depends on the proportion of neutral and advantageous mutants

TABLE IV
Replication Error Rates (per Nucleotide) of RNA Viruses[a]

Coliphage $Q\beta$	3×10^{-4}
Foot-and-mouth disease	5×10^{-4}
Influenza	$7 \times 10^{-5} - 2.7 \times 10^{-4}$
Poliomyelitis	3×10^{-5}
Vesicular stomatitis	$3 - 5 \times 10^{-4}$

For Comparison (DNA Replication)

Coliphage λ	2.4×10^{-8}
Coliphage T_4	1.7×10^{-8}
Escherichia coli	2×10^{-10}

[a] Values (literature cf. ref. 63) determined from reversion frequencies of mutants and hence represent *true* error rates (to be distinguished from population numbers of mutants, i.e., so-called mutant frequencies).

within a given quasi-species. If this degree is appreciable (relative to population size), selection and evolution may coalesce into one process approaching the global maximum of fitness. Again, the error threshold relation is of utmost importance. Evolutionary progress is expected to be greatest near the error threshold. Occasionally surpassing this limit may assist in the escape from metastable distributions to find a higher fitness peak, just as *simulated annealing* is used for efficient optimization.

Evolutionary optimization then is not just a blind stochastic trial-and-error search for a better adapted mutant but rather follows an inherent logic:

1. Selective advantage usually is to be expected in a distant mutant, that is, as the periphery of the distribution.

2. Mutants far distant from an established master sequence arise from those that are less distant, that is, from precursors "en route".

3. The probability of producing a distant mutant therefore depends critically on the population numbers of its precursors.

4. Population numbers of mutants are high relative to equidistant competitors if they are situated in a domain of high fitness.

5. A fractal clustering of the fitness distribution in domains is likely to occur.

6. Precursors of distant advantageous mutants therefore have a higher chance to be populated than precursors of deleterious mutants.

7. The high dimensionality of sequence space aids the connectivity of fitness domains of (nearly) neutral mutants.

8. Evolutionary optimization proceeds along defined pathways in sequence space. There are alternative routes, but their number is so highly restricted that one has the impression of a phenomenon of automatic guidance to higher fitness.

This logic, rather than the simplifying interpretation of Darwinian behavior as an interplay of chance (random) mutation and necessity (selective fixation of the advantageous mutant once produced), is the basis of evolutionary optimization or adaptation.

3. Experimental Evidence

Many of the classical observations of molecular genetics, especially those related to viruses, have to be reexamined in the light of the new theory. Virologists have accepted the concept of quasi-species (cf. articles by Domingo and Holland and others [64, 65]). What once had been interpreted as the wild type showing a defined individual sequence now has been identified to be a widely dispersed mutant spectrum with defined consensus

sequence. One of the classical experiments was carried out by Weissmann and co-workers [66]. They cloned single particles of the Coli phage $Q\beta$, amplified the clones, and determined sequences or fingerprints. The result was that none of the different clones exactly agreed in their sequences but rather showed one or several differences. Weissmann estimated from his experiments that a true wild type as an individual sequence was present to an extent of less than 5%. Moreover, by evaluating the rate of revertant formation (occurring in artificially produced *site-specific* mutants) as well as the relative growth rates of the revertant (master copy) and the mutants, he could correlate sequence length with error and wild-type superiority, in agreement with the error threshold relation [Eqn. (III.5)]. The experiment is schematically represented in Figure 25. One may wonder why such an experiment works at all, that is, how a mutant clone can grow up before it is replaced by the revertant "wild type", which must appear soon in the mutant's quasi-species distribution. All clones would then show identical sequences, that is, the wild-type sequence. The explanation that mutant sequences can be identified in macroscopic amounts lies in the fact that what is being fished out in cloning single particles are nearly neutral mutants (being present in relatively high abundance). Since such a mutant (i) has a diagonal coefficient $W_{ii} \approx W_{mm}$ [i.e., being almost neutral with the master (m)], it grows
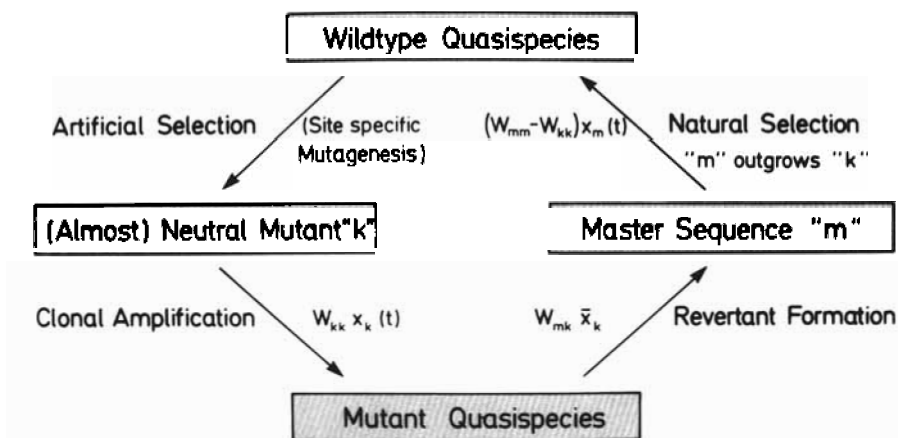
Figure 25. Scheme of Weissmann's experiment [66] for determining error rates of phage $Q\beta$. Success critically depends on finding nearly neutral mutant k, which grows up about as quickly as master copy m, i.e., $W_{kk} \approx W_{mm}$, while being replaced by revertant master copy with relatively small rate (i.e., $(W_{mm} - W_{kk}) \ll W_{kk}$). In natural population such nearly neutral mutants may be picked up frequently because they are abundantly populated. Weissmann chose an extracistronic site-specific mutant.

up in clone amplification as $e^{W_{ii}t} \approx e^{W_{mm}t}$, while the revertant replaces the mutant only according to $e^{(W_{mm} - W_{ii})t}$.

Error rates have also been determined for foot-and-mouth disease virus by Domingo [67], for influenza A virus by Palese [68], and for vesicular stomatitis virus by Holland [69] and their co-workers. All results show the correlation of error rate with sequence length. However, the factor $\ln \sigma$ may considerably influence the results if $\sigma \to 1$. In fact, some of those experiments—where single cloned mutants were amplified directly for the sake of sequence analysis—may produce misleading results since in these cases near-neutral mutants only are to be detected. (Mutants with W_{ii} considerably smaller than W_{mm} would appear with unduly large delay.) Influenza A seems to be a case where a large fraction of the infectious virus mutants are indeed (almost) neutral, possibly as a consequence of the fact that this virus contains eight separate RNA chains in its genome. On the other hand, poliovirus, which has been tested by the same method, did not show detectable (neutral) mutants. In these cases a quantitative study of the kinetics of revertant formation from site-directed mutants may offer a more direct access to the average error rate $(1 - \bar{q})$.

In carrying out such experiments on error rates, one has to specify carefully the conditions under which the results were obtained. In particular, one has to distinguish the (average) probability $(1 - \bar{q})$ of producing a substitution per site (in replicating a sequence) from the rate of appearance of a mutant in an established quasi-species distribution. Hence it makes a difference whether one observes one or a few replication rounds in an essentially nonstationary distribution or whether one counts frequencies of appearance at a steady state. Hot spots, that is, single sites showing extremely high rates of mutation, have been reported frequently in the literature, and their first discovery certainly may be considered a cornerstone in the rise of molecular virology [70, 71]. However, frequency of appearance may mean either a high rate of production (from wild type) or a high (almost neutral) rate of replication or both. Which interpretation is correct, that is, which fraction is to be associated with either of the two effects, remains to be analyzed for most cases reported.

In any case, working with complete virus particles requires some caution with respect to an evaluation of results. The observable amplification of many viruses appears in the form of single bursts defining infectious cycles. In each burst, however, many copies of virus particles are set free, of which only a minor fraction turns out to be infectious. Within a replication cycle inside the host cell the replication machinery is available to both viable and nonviable (or less efficiently) replicating virions, while after the burst only those particles will survive that have encoded the correct machinery for further infection. Moreover, one may start under conditions where single infections (i.e., one

virus infecting a cell) are guaranteed. After several bursts, such conditions cannot be easily maintained. In the neighborhood of the site of a burst multiple infections will take place. It is important to carry out experiments in such a way that their interpretation either is not sensitive to the details of this mechanism or such that the exact mechanism, if known, is taken into account.

These difficulties can be avoided if replication experiments are carried out in vitro, that is, under cell-free conditions. The late Sol Spiegelman pioneered this kind of experiment [72–74] using the Coli phage $Q\beta$. The basis of these experiments was the observation that the isolated genome of phage $Q\beta$ can be replicated in a suitable reaction mixture, including the monomeric building blocks in an energy-rich form, that is, the four nucleoside triphosphates, the purified enzyme, and the $Q\beta$—replicase (which was isolated from infected cells), the mixture being adjusted to suitable pH and ionic strength conditions [72]. The first replication rounds produced infectious particles, that is, RNA chains that upon introduction to host cells were able to produce infectious phages. If, however, serial upgrowth and dilution was carried out over many generations, the acting selection pressure for efficient replication caused the sequences to delete those portions that are not essential for recognition by the replication enzyme and there grew out a particle only about one-tenth of the length of the intact phage genome but having a replication rate per site far in excess of the wild type [73]. One may call this process *degeneration* rather than *evolution* because the resulting particle entirely lost its capacity of infection. Nevertheless, with respect to replication by the $Q\beta$ enzyme, these particles were better adapted to their new environment, which did not require them to be infectious.

In another experiment using a small RNA variant of about 220 nucleotides (which, again, was not infectious but was efficiently recognized by the replication enzyme), Spiegelman et al. [74] demonstrated a true evolution process. They added ethidium bromide, an inhibitor of replication, to the reaction mixture and found the successive outgrowth of mutants with one, two, and three single-site substitutions, respectively (cf. Figure 26). The final product after about 40 generations was a three-error mutant that was about a factor of 2 more efficient in replication than the original wild type. When these experiments were carried out, the kinetics and mechanism of replication were not yet known in any detail. Such studies were recently done in collaboration with Biebricher and Gardiner [59–61]. The kinetic experiments were supplemented by computer simulations and by analytical treatment of simplified mechanisms. These studies led to a complete identification of all major reaction steps of enzyme RNA replication. The simplified scheme shown in Figure 27 indicates how complicated the true mechanism is and how important it is in evolutionary experiments to choose appropriate conditions. There are essentially three regimes of replication one has to
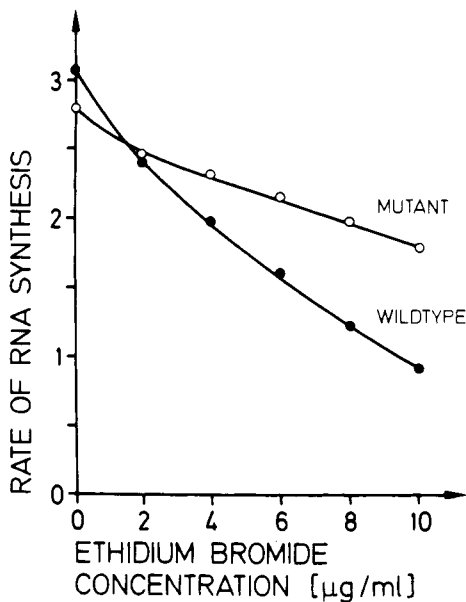
Figure 26. In serial dilution experiment, Spiegelman [73, 74] and co-workers obtained three successive mutants with increased adaptation to presence of ethidium bromide, a drug that interferes with replication. Experiment starts with population of 10^6 MDV strands (variant of $Q\beta$-RNA comprising about 220 nucleotides that is well-adapted to $Q\beta$-RNA–replicase). Population is amplified to about 10^{11} copies and subsequently diluted to initial concentration. Iteration of this procedure led to final product, three-error mutant that was obtained after about 40 iterations. As replication rate data show, mutant is slightly inferior to wild type in absence of ethidium bromide but twice as efficient as wild type at final concentration of ethidium bromide.

distinguish: (i) at low template-to-enzyme ratios where the reaction proceeds exponentially, rating the growth constant of plus and minus strands as a geometric mean; (ii) at template-to-enzyme ratios of 1 and moderately larger, where the enzyme gets saturated by template and the reaction proceeds essentially linearly with time according to some Michaelis–Menten scheme; and finally (iii) at large template-to-enzyme ratios where inhibitions (binding of template to replica site and template double-strand formation) cause the rate to reach a plateau (Figure 28). Selection experiments were carried out in the different regimes leading to differing results [75]. In the exponential growth range the most efficiently growing species (according to its overall rate) is selected while in the linear growth range a species grows out that wins the competition of binding to the enzyme (regardless of growth efficiency). It turns out that, with regard to the reaction mechanism, the original evolution
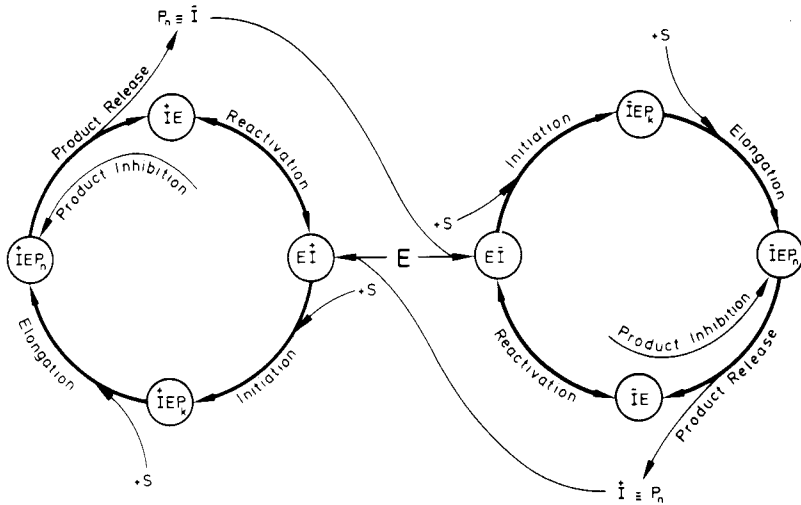
Figure 27. Detailed in vitro mechanism of RNA replication by $Q\beta$–replicase [59]. RNA grows exponentially as long as template concentration is below enzyme concentration. Growth rate becomes constant and hence RNA concentration rises linearly when template concentration exceeds that of enzyme, while, finally, at large template excess, rate decreases down to zero due to enzyme inhibition and template double-strand formation. In these in vitro experiments, $Q\beta$–replicase is present as environmental factor. In vivo the enzyme is formed during the first 20 min after infection of host cell followed by RNA replication during second half of infectious cycle. After about 40 min, about a thousand infectious phage particles per cell are released in burst. These thousand infectious particles usually are minor part of total burst size.

experiments were not carried out under uniquely defined (nor incidentally under optimal) conditions. It is important to optimize the population size with respect to mutant expectancy. Spiegelman's experiments were done under conditions far below the error threshold and therefore resembled quite inefficient adaptation.

The discovery of de novo synthesis [76–79] of RNA by $Q\beta$ replicase opened a new way of studying evolution processes in the laboratory under more favorable conditions. In the process of de novo synthesis many different templates are being formed by individual enzyme molecules and then compete for growth, leading finally to the outgrowth of a fittest sequence. These experiments can be done in the presence of conditions exerting some selection pressure. Ethidium bromide provides such a selection pressure [76], as in Spiegelman's experiments [74], and de novo synthesis in the presence of ethidium bromide does indeed reveal much more efficient products, as seen in Figure 29.
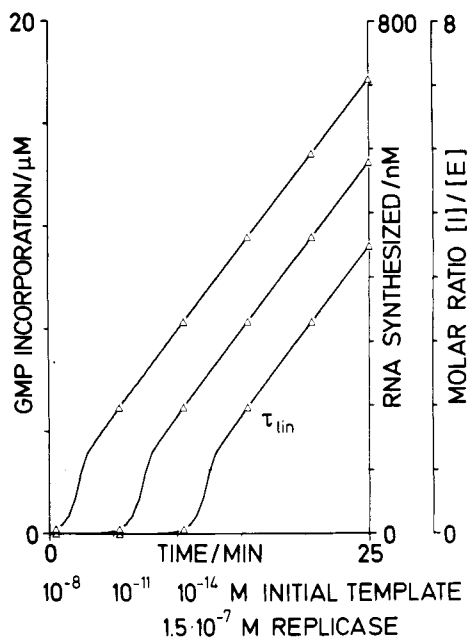
Figure 28. Rate profiles obtained in vitro experiments of RNA replication by $Q\beta$–replicase [59]. Exponential growth range manifests itself by constant shift of growth curves (lag periods) for constant ratios of template numbers. Detectable growth range refers to template concentrations that are comparable to or higher than enzyme (replicase) concentration, while high template-to-enzyme ratios cause leveling of growth curve due to enzyme inhibition and template double-strand formation. Evolution experiments have been carried out with differing results in both exponential and linear growth range.

At present we are endeavoring to optimize these techniques so as to work close to the error threshold and use annealing procedures to speed up the evolutionary process. Automated machines for evolutionary adaptation of RNA sequences and of viruses have been constructed [80, 81].

In summary, experimental evidence available confirms the results of theory and opens some interesting aspects for future work on evolutionary optimization. In fact, very recently, Biebricher [82] was able to clone a quasi-species distribution of RNA and to demonstrate most directly the widely spread mutant spectrum that is characteristic of a molecular quasi-species.

4. Limitations of the Model

Major limitations have been mentioned already, as far as they are related to the particular mechanism of RNA replication or virus infection. The linear
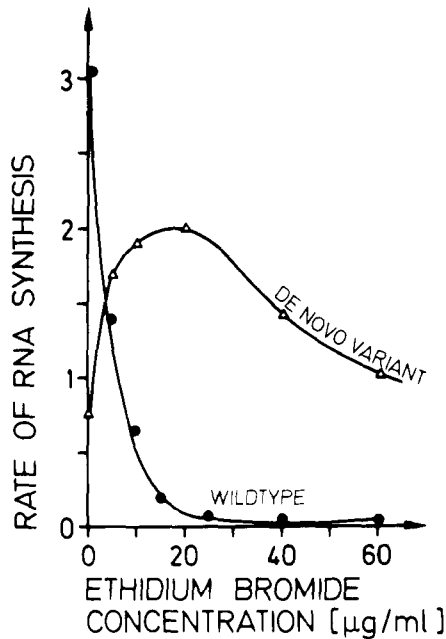
Figure 29. RNA sequences, obtained by de novo synthesis using $Q\beta$–replicase, can be optimally adapted to presence of inhibitors [76]. Growth rate of such variant as function of ethidium bromide concentration indicates *not only much higher degree* of adaptation (as to be expected from broader quasi-species-like mutant spectrum obtained under de novo synthesis conditions, i.e., near error threshold) but also a drug addiction, i.e., higher synthesis rate at finite ethidium bromide concentration.

autocatalytic ansatz of replication (i.e., the term $W_{ii}x_i$) with uniform error rates ($W_{ik}x_k$) represents only the simplest possible model. Nevertheless, it does yield all the experimental features that have been observed in corresponding reaction systems. There is a wide range of conditions where the simple ansatz indeed proves to be a perfectly satisfactory approximation, as computer simulations of the more complex mechanisms have shown [73–75]. Moreover, knowing the particulars of any mechanism in more detail would allow one to adapt the equations correspondingly.

There are, however, more principal limitations, and it is important to stress that the quasi-species model is a particular model holding only where its prerequisites are fulfilled. The linear autocatalytic rate law is one of the prerequisities typical for the quasi-species nature. If rates become independent of the concentrations of growing substraces, coexistence may replace competition, or if, on the other hand, the autocatalytic rate law depends more

strongly than linearly on the concentration of the growing entity (as is the case for hypercycles), selection may become of an entirely different, non-Darwinian nature than is normal for a quasi-species [4].

The typical outcome of quasi-species behavior is natural selection in the Darwinian sense. We believe that it should be possible to adapt the quasi-species model to any situation where natural selection in the Darwinian sense plays a major role. Hence the model may be generalized so as to include all kinds of horizontal gene transfer typical for recombination. The model as presented is essentially a deterministic model that holds only for a sufficiently large population size. This limitation and a possible way to overcome it will be the subject of the final section.

5.   Replication as a Stochastic Process

Evolutionary phenomena by their very nature call for a stochastic description. The source of the undeniable, inherent importance of fluctuations can be located in the molecular structure of biopolymers. The numbers of potentially different polynucleotide sequences, as we pointed out in Section II, are always much larger than the numbers of molecules in realistic populations. So it came as rather a surprise that we were able to characterize stationary mutant distributions by the application of conventional reaction kinetics to polynucleotide replication. Error thresholds, derived by means of the quasi-species concept, brought us indeed to the limits of justification of the deterministic approach. Consider a population replicating with an error rate that exceeds the critical value. The model predicts an approximately uniform distribution of sequences at the stationary state. Such a state can never occur since we cannot have less than one molecule of every type. What happens at error rates above threshold is the breakdown of the stationary state. We are dealing with a changing ensemble of sequences that migrates through sequence space as a random walker. Around the error threshold—this is the range that appears to be most relevant for evolution—this walk is anything but trivial. The underlying value landscapes, as we pointed out in Section IV.3, are highly complicated, bizarre objects, and they determine where and how fast the population moves. The success of the quasi-species model is documented by the powerful localization concept [29] as outlined in Section III.2. It allowed us to extend the approach to such complicated cases where only statistical information on the distribution of selective values is available (Section III.3).

Other important questions, however, remain unanswerable within the deterministic theory. Among them are phenomena directly related to finite populations such as the dependence of error thresholds on population size or the mean lifetimes of mutants in populations. The latter quantity is of particular importance for highly fluctuating mutant distributions such as

those occurring at high error rates or those encountered with distant, selectively neutral mutants. Then the distributions of polynucleotide sequences are essentially nonstationary and are characterized in terms of mean lifetimes rather than by relative concentrations. Another relevant problem concerns the probability of finite populations being caught in local optima of fitness landscapes. In principle, this question can be answered appropriately by means of first passage times that allow one to compute the average time a population will spend in a given region of sequence space before passing to a different region.

Several attempts to describe replication–mutation networks by stochastic techniques were made in the past. We cannot discuss them in detail here, but we shall briefly review some general ideas that are relevant for the quasi-species model. The approach that is related closest to our model has been mentioned already [51]: the evolutionary process is viewed as a sequence of stepwise increases in the populations' mean fitness. Fairly long, "quasi-stationary" phases are interrupted by short periods of active selection during which the mean fitness increases. The approach towards optimal adaptation to the environment is resolved in a manner that is hierarchical in time. Evolution taking place on the slow time scale represents optimization in the whole of the sequence space. It is broken up into short periods of time within which the quasi-species model applies only locally. During a single evolutionary step only a small part of sequence space is explored by the population. There, the actual distributions of sequences resemble local quasi-species confined to well-defined regions. Error thresholds can be defined locally as well.

In the case of selective neutrality—this means that all variants have the same selective values—evolution can be modeled successfully by diffusion models. This approach is based on the analysis of partial differential equations that describe free diffusion in a continuous model of the sequence space. The results obtained thereby and their consequences for molecular evolution were recently reviewed by Kimura [2]. Differences in selective values were found to be prohibitive, at least until now, for an exact solution of the diffusion approach. Needless to say, no exact results are available for value landscapes as complicated as those discussed in Section IV.3. Approximations are available for special cases only. In particular, the assumption of rare mutations has to be made almost in every case, and this contradicts the strategy basic to the quasi-species model.

Chemical reaction networks are frequently modeled by Markov processes and can be formulated as master equations. Commonly, it is straightforward to write down the master equation, but when it comes to derive solutions, hard-to-justify approximations are inevitable; see, for example, ref. 83. In essence, the same is true for polynucleotide replication described by a master

equation. The master equation corresponding to the kinetic ansatz of Appendix 3 was first formulated by Ebeling and Feistel [84]. Several attempts were made to derive results from this master equation [85–89] or from a closely related Langevin equation [90], but none of these was successful in deriving a stochastic version of the error threshold.

A more restricted specification of the Markov process used in the modeling of replication and mutation yields more results for special cases. Birth-and-death processes were applied to error-free replication. We mention two examples [91, 92]. Probabilities of mutant extinction and first-passage times for the selectively neutral case can be derived immediately. Multitype branching processes allows mutations to be included [48, 93]. This approach corresponds to Eqn. (A3.1) without the flux term $\Phi(t)$. A stochastic error threshold was derived for this unconstrained replication–mutation system [48]. It is formally identical to the deterministic expressions derived in Section III.1, but the interpretation is different: the value matrix W is replaced by the mean matrix of the multitype branching process. At error rates smaller than the threshold the system sustains a master sequence and its mutant distribution with a certain probability that manifests itself in a finite probability of survival to infinite time. If the error rate exceeds the threshold value, the probability of survival for the master sequence and for all other sequences is zero. This implies that we are dealing with a changing ensemble of sequences just as in the deterministic model, where the quasi-species fails to localize under these conditions. The stochastic error threshold sharpens with increasing chain length v—as it does in the deterministic case—and this implies here that the probability of survival to infinite time decreases sharply from almost 1 at error rates just below critical to almost zero just above the threshold. Attempts to incorporate the constraint $\Phi(t)$ into the multitype branching model by means of a combined branching and sampling technique are presently in progress.

A recent attempt at a direct stochastic theory by Weinberger [94] using the deterministic flow term as an external (precomputed) constraint should be mentioned here. The intractability of a large coupled system of second-order partial differential equations for the generating function is then reduced to a (nonlinearly coupled) system of ordinary differential equations. The price is the loss of proper population regulation and possible extinction.

At present the most that can be said analytically about the error threshold in a finite population is the necessary condition [50]

$$\frac{\sigma q^{v}}{\sigma q^{v} - 1} = \varepsilon \bar{x}_{m}, \qquad \varepsilon \ll 1, \qquad (V.1)$$

which demands small fluctuations in the master sequence population. Master

equations can be studied by computer simulation provided particle numbers are not too large. An algorithm particularly well suited for the investigation of stochastic chemical reactions networks was worked out by Gillespie [95]. It was used to compute error thresholds for finite populations on the simple value landscape we analyzed deterministically in Section IV.1. Chain lengths up to $v = 20$ and populations up to $N = 2000$ were simulated [96]. Some results are shown in Figure 30 for a superiority, $\sigma = 10$. In small populations the error threshold occurs at higher q values than in the infinite population, which is implicitly assumed in the deterministic approach. Fluctuations endanger the stability of stationary sequence distributions, and hence the smaller the population, the fewer errors it can tolerate. A comparison of Eqn. (V.1) with the numerical results with $\bar{x}_m$ determined by the deterministic second-order perturbation result,

$$\bar{x}_m = \frac{q^v - q^v_{\text{det, cr}}}{1 - q^v_{\text{det, cr}}} N, \qquad (V.2)$$

is also shown in Figure 30 and yields good agreement for the conventional choice $\varepsilon = 0.1$. Further work along the lines suggested in the preceding will
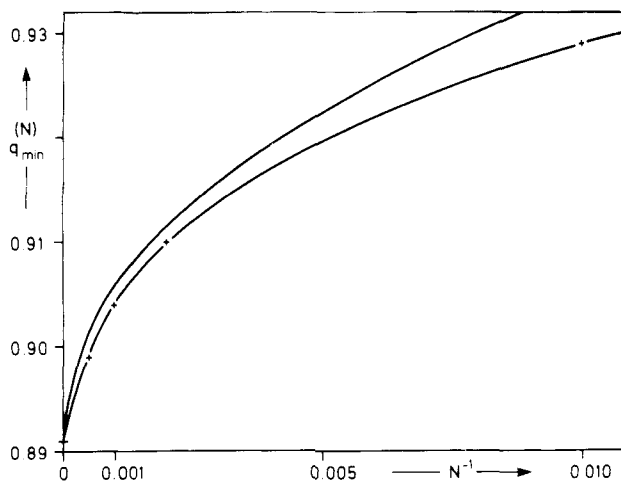
Figure 30. Error threshold as function of population size. Stochastic replication–mutation dynamics in ensemble of polynucleotide sequences with chain length $v = 20$ simulated by Gillespie's algorithm [95]. Critical single-digit accuracy of replication (q_{min}) at which ordered quasi-species is converted into changing population of sequences with finite lifetimes is plotted as function of $1/N$, reciprocal population size (lower curve). For further details see ref. 96. Upper curve is theoretical prediction of Eqn. (V.1) based on ref. 51.

provide more information about this and the lifetime of metastable quasi-species. The above brief review of stochastic theory suffices to provide a suitable context for the embedding of the deterministic quasi-species model while substantiating its basic claims.

APPENDIX 1.   MUTATION FREQUENCIES AND HAMMING DISTANCES

The Hamming distance $d(i, k)$ of two sequences I_i and I_k counts the number of digits in which they differ. The number of sequences in a mutant class, that is, the number of sequences with Hamming distance $d(0, k) = d$ from a reference sequence I_0, is

$$N_d = \binom{v}{d}(\kappa - 1)^d, \tag{A1.1}$$

where v is the number of symbols in the sequence class, that is, the length of the chain, and κ is the number of symbol classes, that is, the number of different digits. In binary sequences we have $\kappa = 2$, and in polynucleotides we have $\kappa = 4$. Summation over all possible Hamming distances $0 \leqslant d \leqslant 0$ yields

$$N = \sum_{d=0}^{v} \binom{v}{d}(\kappa - 1)^d = \kappa^d. \tag{A1.2}$$

The frequency at which the set of sequences with Hamming distance d is produced as error copies of the template I_0 is

$$Q_d = \binom{v}{d}q^{v-d}(1-q)^d = \binom{v}{d}q^v(q^{-1}-1)^d. \tag{A1.3}$$

The frequency of correct replication, Q_0, represents the quality factor of the replication process:

$$Q_0 = q^v. \tag{A1.4}$$

The frequency with which a given polynucleotide is produced as an error copy of the reference sequence I_0 is then obtained from

$$Q_{i0} = \frac{Q_{d(0, i)}}{N_{d(0, i)}} = Q_0 \left(\frac{q^{-1} - 1}{\kappa - 1} \right)^{d(0, i)}$$

or, in general, the frequency of producing I_i as an error copy of any sequence I_k is

$$Q_{ik} = \frac{Q_{d(k,i)}}{N_{d(k,i)}} = Q_0 \left(\frac{q^{-1} - 1}{\kappa - 1}\right)^{d(k,i)}. \tag{A1.5}$$

APPENDIX 2.   RECURSIVE CONSTRUCTION OF THE MUTATION MATRIX Q

The mutation matrix for binary sequences of length $v + 1$, $Q(v + 1)$, is to be constructed from $Q(v)$ in such a way that the eigenvalues and eigenvectors are obtained recursively. We start with $v = 1$ and set

$$\varepsilon = (q^{-1} - 1) \quad \text{and} \quad B(v) = q^{-v} Q(v). \tag{A2.1}$$

Then we obtain for the two smallest B matrices (on the top and in the rightmost column of the matrix we indicate to which sequence the corresponding elements refer)

$$\begin{array}{cc} 0 & 1 \end{array}$$
$$B(1) = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix} \begin{array}{c} 0 \\ 1 \end{array} \tag{A2.2}$$

$$\begin{array}{cccc} 00 & 01 & 10 & 11 \end{array}$$
$$B(2) = \begin{pmatrix} 1 & \varepsilon & \varepsilon & \varepsilon^2 \\ \varepsilon & 1 & \varepsilon^2 & \varepsilon \\ \varepsilon & \varepsilon^2 & 1 & \varepsilon \\ \varepsilon^2 & \varepsilon & \varepsilon & 1 \end{pmatrix} \begin{array}{c} 00 \\ 01 \\ 10 \\ 11 \end{array}. \tag{A2.3}$$

This recursion consists of duplication of sequence space as indicated in Figure 2. Note that the sequences in the rows and columns of the Q matrix are arranged in lexicographical order. The recursion can be generalized and yields

$$B(v + 1) = \begin{pmatrix} B(v) & \varepsilon B(v) \\ \varepsilon B(v) & B(v) \end{pmatrix} \tag{A2.4}$$

The eigenvalues $\lambda_k(v + 1)$ and the corresponding eigenvectors $\eta_k(v + 1)$ are computed recursively too:

$$\lambda_k(v + 1) = \lambda_k(v)(1 + \varepsilon), \qquad \lambda_{k+2^v}(v + 1) = \lambda_k(v)(1 - \varepsilon) \tag{A2.5}$$

and

$$\eta_k(v+1) = \binom{\eta_k(v)}{\eta_k(v)}, \qquad \eta_{k+2^v}(v+1) = \binom{\eta_k(v)}{-\eta_k(v)}, \qquad k = 1, \ldots, 2^v.$$

(A2.6)

Hence, all eigenvalues are of the same general form:

$$\lambda_k(v) = \prod_{j=1}^{v} (1 + \gamma_j \varepsilon) \quad \text{with} \quad \gamma_j \in \{-1, +1\}.$$

(A2.7)

They are all real and positive and fulfill the inequality

$$\lambda_{\max}(v) \geqslant \lambda_k(v) \geqslant \lambda_{\min}(v)$$

with

$$\lambda_{\max}(v) = (1 + \varepsilon)^v \quad \text{and} \quad \lambda_{\min}(v) = (1 - \varepsilon)^v.$$

(A2.8)

Both the largest and the smallest eigenvalue are nondegenerate (the former result also following directly from the Frobenius theorem).

APPENDIX 3.   THE RATE EQUATIONS

The differential equation for the time dependence of the population variable $n_i(t)$ is of the form (cf. Figure 2)

$$\dot{n}_i(t) = W_{ii} n_i(t) + \sum_{k \neq i} W_{ik} n_k(t) + \frac{n_i(t)}{\sum_k n_k(t)} \Phi(t).$$

(A3.1)

The number of variables is extremely large. If we admit all point mutations of sequences with length v, we have $i = 1, 2, \ldots, \kappa^v$. The number of sequences per unit volume of a given type, I_i, are denoted by $n_i(t)$. The rate parameters are collected in the value matrix W. The diagonal elements W_{ii} are the rate parameters for the net or excess production of I_i through (correct) self-copying. They are composite quantities,

$$W_{ii} = A_i Q_{ii} - D_i$$

(A3.2)

with A_i and D_i being the rate constants of total (correct and erroneous) template- (I_i) induced replication and degradation, respectively. The frequency of correct copying, Q_{ii}, corresponds to the quality factor Q_0 as defined

in Appendix 1. The off-diagonal elements W_{ik} are the rate parameters for mutational production of the sequence I_i through (imperfect) copying of the template I_k. They can be expressed in the form

$$W_{ik} = A_k Q_{ik}, \tag{A3.3}$$

where A_k is the total rate of replication on the template I_k and Q_{ik} is the frequency of the mutation $I_k \rightarrow I_i$ (Appendix 1). The (time-dependent) dilution flux is given by $\Phi(t)$. It represents the constraint for the particular model under consideration. The conservation law

$$\sum_k Q_{ik} = 1 \tag{A3.4}$$

allows one to define an average excess production $\bar{E}(t)$ that is of the simple form

$$\bar{E}(t) = \sum_k n_k(t)(A_k - D_k) \bigg/ \sum_k n_k(t). \tag{A3.5}$$

The total production rate of the population can then be written as

$$\sum_k \dot{n}_k(k) = \bar{E}(t)\sum_k n_k(t) + \Phi(t) \tag{A3.6}$$

It will turn out to very useful to introduce relative fractions of population variables: $x_i(t) = n_i(t)/\Sigma_k n_k(t)$. Then, the equation for the relative growth rate of sequence I_i is

$$\dot{x}_i(t) = [W_{ii} - \bar{E}(t)]x_i + \sum_{k \neq i} W_{ik} x_k(t). \tag{A3.7}$$

APPENDIX 4. THE SOLUTION OF THE SYSTEM OF DIFFERENTIAL EQUATIONS (II.6)

Consider each variable $z_i(t)$ as the ith component of a column vector $\mathbf{z}$ and each rate coefficient W_{ik} as an element of the quadratic value matrix W. The system of rate Eqs. (II.6) then reads

$$\frac{d\mathbf{z}(t)}{dt} = W\mathbf{z}(t). \tag{A4.1}$$

Provided the matrix W is diagonalizable (which we shall assume throughout this Chapter; cf. also Appendix 2 and Section IV), the system of differential equations can be decoupled by means of the transformation

$$\mathbf{z}(t) = L\boldsymbol{\zeta}(t) \quad \text{and} \quad \boldsymbol{\zeta}(t) = L^{-1}\mathbf{z}(t), \tag{A4.2}$$

where $\boldsymbol{\zeta}(t)$ represents again a column vector and L as well as its inverse L^{-1} are quadratic matrices. Combination of (A4.1) and (A4.2) yields

$$\frac{d\boldsymbol{\zeta}(t)}{dt} = L^{-1}WL\boldsymbol{\zeta}(t). \tag{A4.3}$$

The matrix $\Lambda = L^{-1}WL$ is diagonal as assumed in the preceding. Its elements are the eigenvalues of W: λ_k ($k = 0, 1, \ldots$). The corresponding column vectors $\mathbf{l}_k = \{l_{1k}, l_{2k}, \ldots\}$ are stationary compositions (or "normal modes") for $\boldsymbol{\zeta}(t)$, which grow in magnitude at the rates λ_k. They are found as (right-hand) eigenvectors of the matrix W. Provided the off-diagonal elements of the value matrix are strictly positive, the Frobenius theorem applies. Then the largest eigenvalue $\lambda_{\max} = \lambda_0$ is nondegenerate. Without loss of generality, we may order the eigenvalues

$$\lambda_0 > \lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \lambda_4 \geqslant \cdots.$$

The eigenvector corresponding to λ_0 is called the dominant eigenvector $\mathbf{l}_0$. All its components are strictly positive: $l_{i0} > 0$ ($i = 1, 2, \ldots$). The left-hand eigenvectors are given by the rows of the inverse matrix L^{-1}.

For each component of $\boldsymbol{\zeta}(t)$ we obtain a solution curve of the general form

$$\zeta_k(t) = \zeta_k(0)\exp(\lambda_k t) \tag{A4.4}$$

wherein initial conditions are denoted by $\zeta_k(0) = \zeta_k(t = 0)$. The eigenvalues λ_k of the value matrix W are the solutions λ of the determinantal equation

$$\det(W - \lambda I) = 0, \tag{A4.5}$$

where det means determinant and I here denotes the unit matrix. Evaluation of Eqn. (A4.5) yields a polynomial in λ whose degree is given by the number of concentration coordinates. In order to determine the components of the right and left eigenvectors of the matrix W, we start from

$$WL = L\Lambda \tag{A4.6}$$

and

$$L^{-1}W = \Lambda L^{-1} \quad \text{or} \quad W'(L^{-1})' = (L^{-1})'\Lambda. \tag{A4.7}$$

The prime means here the transposed matrix. Explicitly in matrix elements, $L = \{l_{ij}\}$ and $L^{-1} = \{h_{ij}\}$, the last two equations read

$$\sum_j W_{ij}l_{jk} = l_{ik}\lambda_k, \tag{A4.8}$$

$$\sum_j W_{ji}h_{kj} = h_{ki}\lambda_k. \tag{A4.9}$$

The explicit solutions $z_i(t)$ of the linear differential Eqn. (A4.1) are obtained as superpositions of individual normal modes,

$$z_i(t) = \sum_k l_{ik}c_k\exp(\lambda_k t), \quad i = 1, 2, \dots . \tag{A4.10}$$

The constants c_k can be obtained from the initial conditions, $z_i(0) = x_i(0)$:

$$c_k = \sum_i h_{ki}x_i(0). \tag{A4.11}$$

[Note that the initial value of the function $f(t)$ in Eqn. (II.5) is $f(0) = 1$ by definition.] Explicit solutions for the relative populations variables $x_i(t)$ can be obtained from Eqs. (II.5) and (II.7):

$$x_i(t) = \frac{\sum_k l_{ik}c_k\exp(\lambda_k t)}{\sum_j\sum_k l_{jk}c_k\exp(\lambda_k t)}. \tag{A4.12}$$

Formally they can also be expressed as

$$x_i(t) = f(t)\sum_k l_{ik}c_k\exp(\lambda_k t) = \sum_k l_{ik}b_k y_k(t), \tag{A4.13}$$

where

$$y_k(t) = (c_k/b_k)f(t)\exp(\lambda_k t) \quad \text{or} \quad c_k/b_k = y_k(0) \tag{A4.14}$$

with

$$b_k = \left(\sum_i l_{ik}\right)^{-1} \quad \text{and} \quad \sum_k y_k(t) = 1. \tag{A4.15}$$

Note that the variables $y_k(t)$ are normalized, since

$$\sum_k \frac{c_k}{b_k} = \sum_k \sum_i h_{ki} x_i(0) \sum_j l_{jk} = \sum_k x_k(0) = 1 \qquad \text{(A4.16)}$$

due to $LL^{-1} = I$.

APPENDIX 5. GRADIENT SYSTEMS AND SELECTION

Optimization processes are easily visualized in the case where the dynamics can be described by a "gradient system":

$$\dot{x}_k = \left(\frac{\partial V}{\partial x_k}\right)_{x_{j \neq k}}, \qquad k = 1, \ldots, n, \qquad \text{(A5.1)}$$

or in vector notation, $\mathbf{x} = (x_1, \ldots, x_n)$:

$$\frac{d\mathbf{x}}{dt} = \text{grad } V.$$

The potential is not an explicit function of time but depends on it implicitly: $V = V[\mathbf{x}(t)]$. The potential function V is nondecreasing along a trajectory $\mathbf{x}(t)$ since

$$\frac{dV}{dt} = \sum_k \frac{\partial V}{\partial x_k} \frac{dx_k}{dt} = \sum_k \left(\frac{\partial V}{\partial x_k}\right)^2 \geq 0. \qquad \text{(A5.2)}$$

Gradient systems do not sustain oscillations in the variables $x_k(t)$: the Jacobian matrix $A = \{a_{ij}; i, j = 1, \ldots, n\}$ where

$$a_{ij} = \frac{\partial}{\partial x_j} \frac{dx_i}{dt} = \left(\frac{\partial^2 V}{\partial x_j \, \partial x_i}\right)_{x_{k \neq i,j}} = \left(\frac{\partial^2 V}{\partial x_i \, \partial x_j}\right)_{x_{k \neq i,j}} = a_{ji} \qquad \text{(A5.3)}$$

is symmetric and hence has no complex eigenvalues. Reaction–diffusion equations derived from gradient systems,

$$\frac{\partial x_k}{\partial t} = D_k \nabla^2 x_k + \left(\frac{\partial V}{\partial x_k}\right)_{x_{j \neq k}}, \qquad k = 1, \ldots, n,$$

do not form stable spatial patterns under no-flux boundary conditions [24].

First we consider error-free replication. Does it represent a gradient system? The differential equation for n competing sequences reads

$$\dot{x}_k = x_k(W_k - \bar{E}), \qquad k = 1, \ldots, n \quad \text{and} \quad W_k = A_k - D_k. \quad (A5.4)$$

Here $\bar{E}$ is a nondecreasing function of time provided the rate constants W_k are time independent:

$$\left(\frac{\partial \bar{E}}{\partial x_k}\right)_{x_{j \neq k}} = W_k - \bar{E},$$

and hence

$$\frac{d\bar{E}}{dt} = \sum x_k(W_k - \bar{E})^2 \geqslant 0. \quad (A5.5)$$

(This equation is always true since concentrations x_k are positive or zero and rate constants are real by definition.) The mean excess production $\bar{E}$, however, is not a potential function in the sense of Eqn. (A5.1) since the trajectories of Eqn. (A5.4) do not intersect with the constant level sets of $\bar{E}$ at right angles (Figure 4). Based on techniques originally introduced by Shah-shahani [25], it is possible to transform Eqn. (A5.4) into a generalized gradient system [26, 27],

$$\frac{d\mathbf{x}}{dt} = \text{grad } V \quad (A5.6)$$

This generalized gradient is based on a Riemannian metric defined on the interior of the concentration space $\{x_k > 0; \Sigma_k x_k = 1, k = 1, \ldots, n\}$, which replaces the conventional Euclidean metric. We compare the definitions of the two inner products:

$$\text{Euclidean metric:} \qquad \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k=1}^{n} x_k y_k,$$

$$\text{Shahshahani metric:} \qquad [\mathbf{x}, \mathbf{y}]_z = \sum_{k=1}^{n} \frac{1}{z_k} x_k y_k.$$

In the Shahshahani metric every component of the inner product is weighted by the corresponding coordinate of the point at which the inner product is formed. This weighting distorts the direction of the gradient of the potential

$$V(x_i, \ldots, x_n) = \bar{E} = \sum_{k=1}^{n} W_k x_k \quad (A5.7)$$

such that it coincides with the direction of the trajectories of Eqn. (A5.4). In Figure 4 we show a simple example of dimension $n = 3$.

Next we consider the equation with mutations in a coordinate system whose axes are spanned by the eigenvectors of the matrix W according to Eqn. (II.16). Then the selection equation is of the general form

$$\dot{y}_k = y_k(\lambda_k - \bar{E}), \qquad k = 1, \ldots, n, \tag{A5.4'}$$

wherein we denote the coefficient of the kth eigenvector ($\mathbf{l}_k$) by y_k and the corresponding eigenvalue by λ_k. Equation (A5.4'), although formally identical, differs from (A5.4) with respect to the domains of variables as well as the range of allowed parameter values. The relative concentrations $x_k(t)$ were nonnegative, whereas the y_k may be positive or negative depending on the initial conditions $y_k(0)$. The eigenvalues of the matrix W may be real or complex. Only the largest eigenvalue λ_0 is real and positive by the Frobenius theorem. Consequently, the mean excess production ($\Sigma x_k = \Sigma y_k = 1$)

$$\bar{E} = \sum_k x_k W_{kk} = \sum_k y_k \lambda_k \tag{A5.8}$$

is no longer a nondecreasing function of time since

$$\frac{d\bar{E}}{dt} = \sum_k y_k(\lambda_k - \bar{E})^2$$

may be either positive or negative. It is illustrative to split the concentration space around the asymptotically stable point ($y_0 = 1$, $y_k = 0$, $k = 1, \ldots, n - 1$) into orthants according to the signs of the variables (Figure 4). Since the differential Eqn. (A5.4') is invariant on the boundaries of the orthants,

$$y_k = 0 \rightarrow \dot{y}_k = 0 \rightarrow \ddot{y}_k = 0, \ldots$$

trajectories cannot cross these boundaries. The system thus will always stay inside the orthant to which it was assigned by the choice of initial conditions. There is one orthant, namely, the positive orthant ($y_k \geqslant 0$; $k = 0, 1, \ldots, n - 1$), in which $\bar{E}(t)$ is a nondecreasing function for real eigenvalues λ_k, and one orthant, the orthant ($y_0 > 1$, $y_k \leqslant 0$; $k = 1, \ldots, n - 1$), in which $\bar{E}(t)$ is nonincreasing for real eigenvalues λ_k. In the remaining orthants $\bar{E}$ may decrease, increase, or pass through an extremum.

Jones [19] derived a function

$$\bar{\varepsilon}(t) = \sum_k p_k \operatorname{Re} \lambda_k \tag{A5.9}$$

with $p_k(t) = |y_k(t)|/\Sigma_j|y_j(t)|$ and Re λ_k being the real part of a potentially complex pair of eigenvalues. It is easy to visualize that

$$\frac{d\bar{\varepsilon}}{dt} = \sum_k p_k(\text{Re }\lambda_k - \bar{\varepsilon})^2 \geqslant 0 \qquad (A5.10)$$

holds in full generality.

APPENDIX 6.   BRILLOUIN–WIGNER PERTURBATION THEORY OF THE QUASI-SPECIES

The coordinates $z(t)$ of the differential Eqn. (II.6) can be expressed in the Laplace transform variable s, which is conjugate to the time t. With a single copy of sequence type I_k as initial conditions, they take the form

$$\tilde{z}_{ik}(s) = \frac{\delta_{ik}}{s - W_{ii}} + \sum_j \frac{1}{s - W_{ii}} W_{ij}\tilde{z}_{jk}(s). \qquad (A6.1)$$

Here δ_{ik} denotes the Kronecker delta.

The perturbation series solution is then obtained by iteration as

$$\tilde{z}_{ik}(s) = \frac{\delta_{ik}}{s - W_{ii}} + \frac{1}{s - W_{ii}}\,{}^c W_{ik}\tilde{z}_{kk}(s), \qquad (A6.2)$$

where

$${}^c W_{ik}(s) = \sum_j \frac{W_{ij}W_{jk}}{s - W_{jj}} + \sum_{j,l} \frac{W_{ij}W_{jl}W_{lk}}{(s - W_{jj})(s - W_{ll})} + \cdots. \qquad (A6.3)$$

In particular,

$$\tilde{z}_{kk}(s) = [s - W_{kk} - {}^c W_{kk}(s)]^{-1}, \qquad (A6.4)$$

and hence the perturbation solution is determined in closed form by the perturbation series of Eqn. (A6.3). The asymptotic behavior is exponential growth with the rate given by the dominant eigenvalue λ_0,

$$z_{ik}(t) \simeq a_{ik}\exp(\lambda_0 t) \quad \text{for large } t, \qquad (A6.5)$$

where

$$a_{ik} = \lim_{s \to 0} s\tilde{z}_{ik}(s + \lambda_0) = \frac{1}{\lambda_0 - W_{ii}}\,{}^c W_{ik}(\lambda_0)\,a_{kk} \qquad (A6.6)$$

and

$$a_{kk} = \lim_{s \to 0} \frac{s}{s + \lambda_0 - W_{kk} - {}^cW_{kk}(s + \lambda)}, \tag{A6.7}$$

which results from Eqs. (A6.2) and (A6.4). If $a_{kk} \neq 0$, then Eqn. (A6.7) yields directly the result of the Brillouin–Wigner perturbation theory for the dominant eigenvalue

$$\lambda_0 = W_{kk} + {}^cW_{kk}(\lambda_0), \tag{A6.8}$$

and the population variables x_k may be obtained in terms of Eqn. (A6.3) evaluated at $s = \lambda$:

$$(x_k)^{-1} = \sum_i \frac{a_{ik}}{a_{kk}} = 1 + \sum_{i \neq k} \frac{{}^cW_{ik}(\lambda)}{\lambda - W_{ii}}. \tag{A6.9}$$

APPENDIX 7.   RENORMALIZATION OF THE PERTURBATION THEORY

Equation (A6.3) may be rewritten using the factorization of Eqn. (III.7) in the form

$$^cW_{ik}(\lambda_0) = W_{kk}\left[\sum_j \frac{V_{ij}V_{jk}}{a_j} + \sum_{j,l} \frac{V_{il}V_{lj}V_{jk}}{a_l a_j} + \cdots\right], \tag{A7.1}$$

where

$$a_i = \frac{\lambda_0 - W_{ii}}{W_{ii}}, \tag{A7.2}$$

and V_{ik} is given by Eqn. (III.7) for $i \neq k$. Although the factors V_{ii} are trivially zero, the series may be renormalized to remove less immediate repetitions following the Watson [30] procedure exploited by Anderson [31]. The result is

$$^cW_{ik}(\lambda_0) = W_{kk}\left[\sum_{j \neq k} \frac{V_{ij}V_{jk}}{e_j^k} + \sum_{j \neq k;\ l \neq j,k} \frac{V_{il}V_{lj}V_{jk}}{e_l^{kj}e_j^k} + \cdots\right], \tag{A7.3}$$

where

$$e_j^k = [\lambda_0 - W_{jj} - {}^cW_{jj}^k(\lambda_0)]/W_{jj},$$

$$e_l^{k,j} = [\lambda_0 - W_{ll} - {}^cW_{ll}^{k,j}(\lambda_0)]/W_{ll}, \ldots. \tag{A7.4}$$

Here the superscripts are suppressed in the summations involved in the corresponding perturbation expressions. Repeated indices have been removed (affecting first the fourth-order terms) at the expense of corrections to the denominators. The renormalized denominators may be simplified approximately to

$$e_i = \frac{\lambda_0}{W_{mm}} \frac{W_{mm} - W_{ii}}{W_{ii}} \tag{A7.5}$$

using a kind of mean-field approximation valid for high-dimensional mutant spaces [29]. The eigenvalue λ_0 is then to be obtained from a series of the self-consistent form:

$$\frac{\lambda_0}{W_{mm}} = 1 + \frac{W_{mm}}{\lambda_0} \left(\sum_{j \neq m} \frac{V_{mj} V_{jm}}{(W_{mm} - W_{jj})/W_{jj}} + \frac{W_{mm}}{\lambda_0} \right.$$

$$\left. \times \sum_{j \neq m;\, l \neq j, m} \frac{V_{ml} V_{lj} V_{jm}}{(W_{mm} - W_{ll})(W_{mm} - W_{jj})/W_{ll} W_{jj}} + \cdots \right), \tag{A7.6}$$

and the remaining expressions depend on λ_0.

APPENDIX 8. STATISTICAL CONVERGENCE OF PERTURBATION THEORY

Considering a typical high-order form in the renormalized perturbation theory expression for $^c W_{im}(\lambda)$, in particular for the sequence $\zeta = (j, k, \ldots, l)$ of $N - 1$ indices,

$$T_{mi\zeta}^{(N)} = \frac{V_{il} \cdots V_{kj} V_{jm}}{e_l \cdots e_k e_j} = V_{il} \cdots V_{kj} V_{jm} t_\zeta^{(N)}, \tag{A8.1}$$

direct averaging diverges. The logarithm of such a term, however, has mean and variance

$$\langle \ln t_\zeta^{(N)} \rangle = Nt \tag{A8.2}$$

and

$$\sigma^2 \{\ln t_\zeta^{(N)}\} = N\sigma^2, \tag{A8.3}$$

where t and σ are asymptotically independent of N, provided there were no

long-range correlations between the denominators e_i. Correlations as a result of repeated indices have been removed by the renormalization procedure outlined in Appendix 7. Equations (A8.2) and (A8.3) imply a sharpening probability distribution for $t_\zeta^{(N)}$:

$$\text{Prob}\{\exp[t(N - N^p)] < t_\zeta^{(N)} < \exp[t(N + N^p)]\} \to 1 \quad \text{as } N \to \infty,$$
$$(A8.4)$$

where p is a parameter satisfying $\frac{1}{2} < p < 1$. The preceding result summarizes the essential statistical argument employed by Economou and Cohen [33].

Since all the terms are positive rather then of random sign, Eqn. (A8.4) may be used to prove that the sum of all terms of Nth order has a value for larger N near

$$T_m^{(N)} = \sum_{i,\zeta} T_{mi\zeta}^{(N)} = e^{Nt} \sum_{i,l,\ldots,j} V_{il} \cdots V_{kj} V_{jm} = e^{Nt}(q^{-\nu} - 1)^N \quad (A8.5)$$

with probability 1. The last equality employs the closure summation (cf. Appendix 1):

$$\sum_{i \neq m} V_{im} = \sum_{d=1}^{\nu} N_d \left(\frac{q^{-1} - 1}{\kappa - 1}\right)^d = q^{-\nu} - 1. \quad (A8.6)$$

In the absence of systematic correlations between the replication rates of nearby mutants the average t may be obtained,

$$\langle t \rangle = -\langle \ln e_j \rangle = -\left\langle \ln \frac{W_{mm} - W_{jj}}{W_{jj}} \right\rangle - \langle \ln \lambda_0 \rangle + \ln W_{mm}, \quad (A8.7)$$

and the convergence of the perturbation series depends on the inequality

$$1 > (q^{-\nu} - 1) \frac{W_{mm}}{\langle \lambda_0 \rangle_{\ln}} \left\langle \frac{W_{jj}}{W_{mm} - W_{jj}} \right\rangle_{\ln}, \quad (A8.8)$$

where the logarithmic averages are defined by

$$\left\langle \frac{W_{jj}}{W_{mm} - W_{jj}} \right\rangle_{\ln} = \exp\left\{\left\langle \ln \frac{W_{jj}}{W_{mm} - W_{jj}} \right\rangle\right\} \quad (A8.9)$$

and

$$\langle \lambda_0 \rangle_{\ln} = \exp\{\langle \ln \lambda_0 \rangle\}. \quad (A8.10)$$

APPENDIX 9.   VARIABLES, MEAN RATE CONSTANTS, AND MEAN SELECTIVE VALUES FOR THE RELAXED ERROR THRESHOLD

The composite sequences (Figure 7) and its concentrations are defined by

$$I_{ij} = A_i - B_j; \quad [I_{ij}] = x_{ij}n; \quad n = \sum_{ij} n_{ij}; \quad \sum_{ij} x_{ij} = 1. \tag{A9.1}$$

This leads to the conventional selection equation

$$\dot{x}_{ij} = (W_{ij,ij} - \bar{E})x_{ij} + \sum_{ij \neq kl} W_{ij,kl}x_{kl}, \tag{A9.2}$$

with the selective value

$$W_{ij,kl} = A_{kl}Q_{ij,kl} - D_{ij}\delta_{ij,kl}. \tag{A9.3}$$

Averaging leads to

$$I_{i.} \equiv A_i - B: \quad [I_{i.}] = \left(\sum_{k=1}^{s} x_{ik} \right)n = x_{i0}.n; \quad \sum_{i} x_{i0} = 1 \tag{A9.4}$$

and

$$I_{.j} \equiv A - B_j: \quad [I_{.j}] = \left(\sum_{k=1}^{r} x_{kj} \right)n = x_{0j}.n; \quad \sum_{j} x_{0j} = 1 \tag{A9.5}$$

The same procedure for the rate constants yields

$$\bar{A}_{i0} = \frac{1}{x_{i0}} \sum_{k=1}^{s} A_{ik}x_{ik} \quad \text{and} \quad \bar{A}_{0j} = \frac{1}{x_{0j}} \sum_{k=1}^{r} A_{kj}x_{kj}, \tag{A9.6}$$

$$\bar{D}_{i0} = \frac{1}{x_{i0}} \sum_{k=1}^{s} D_{ik}x_{ik} \quad \text{and} \quad \bar{D}_{0j} = \frac{1}{x_{0j}} \sum_{k=1}^{r} D_{kj}x_{kj}, \tag{A9.7}$$

and

$$\bar{E} = \sum_{ij}(A_{ij} - D_{ij})x_{ij} = \sum_{i=1}^{r}(\bar{A}_{i0} - \bar{D}_{i0})x_{i0} = \sum_{j=1}^{s}(\bar{A}_{0j} - \bar{D}_{0j})x_{0j}. \tag{A9.8}$$

The only assumption made here concerns the mutation matrix: simultaneous

mutations in both parts A and B are excluded,

$$
Q_{ij,kl} = \begin{cases}
Q_{ij,ij} = Q_{ii}^A Q_{jj}^B & \text{if } i = k \quad \text{and} \quad j = l, \\
Q_{ik}^A & \text{if } j = l \quad \text{and} \quad i \neq k, \\
Q_{jl}^B & \text{if } i = k \quad \text{and} \quad j \neq l, \\
0 & \text{if } i \neq k \quad \text{and} \quad j \neq l,
\end{cases}
\tag{A9.9}
$$

where

$$
Q_{ii}^A = 1 - \sum_{k \neq i} Q_{ik}^A \quad \text{and} \quad Q_{jj}^B = 1 - \sum_{l \neq j} Q_{jl}^B.
\tag{A9.10}
$$

Finally we obtain the coupled selection equations

$$
\dot{x}_{i0} = (\bar{A}_{i0} Q_{ii}^A - \bar{D}_{i0} - \bar{E}) x_{i0} + \sum_{k \neq i} Q_{ik}^A \bar{A}_{ik} x_{k0}
\tag{A9.11}
$$

and

$$
\dot{x}_{0j} = (\bar{A}_{0j} Q_{jj}^B - \bar{D}_{0j} - \bar{E}) x_{0j} + \sum_{l \neq j} Q_{jl}^B \bar{A}_{0l} x_{0l}
\tag{A9.12}
$$

with $i, k = 1, 2, \ldots, r$ and $j, l = 1, 2, \ldots, s$.

Acknowledgment

References

1. E. P. Wigner, Proc. 11th Conf. Robert A. Welch Found. Houston (Texas) 1967.
2. M. Kimura, *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, U.K., 1983.
3. E. Domingo, D. Sabo, T. Taniguchi, and C. Weissmann, *Cell 13*:735–744 (1978).
4. M. Eigen and P. Schuster, *The Hypercycle—a Principle of Natural Selforganization*. Springer-Verlag, Berlin 1979. A combined reprint of *Naturwissenschaften 64*:541–565 (1977): *65*:7–41, 341–369 (1978).
5. M. Eigen, *Chem. Scrip. 26B*:13–26 (1986).
6. P. Schuster, *Chem. Scrip. 26B*:27–41 (1986).
7. M. Eigen and C. K. Biebricher, "Sequence Space and Quasispecies Distributions." In E.

Domingo, P. Ahlquist, and J. J. Holland (eds.), *RNA-Genetics, Vol. II: RNA Variability*. CRC Press, Baton Rouge, FL (in press).

8. D. S. Rumschitzki, *J. Math. Biol. 24*:667–680 (1987).

9. M. Eigen, *Naturwissenschaften 58*:465–526 (1971).

10. C. K. Biebricher, M. Eigen, and W. C. Gardiner, Jr., *Biochemistry 22*:2544–2559 (1983).

11. C. K. Biebricher, M. Eigen, and W. C. Gardiner, Jr., *Biochemistry 23*:3186–3194 (1984).

12. C. K. Biebricher, M. Eigen, and W. C. Gardiner, Jr., *Biochemistry 24*:6550–6560 (1985).

13. A. Kornberg, *DNA Replication*. Freeman, San Francisco, 1980.

14. I. R. Epstein, *J. Theor. Biol. 78*:271–298 (1979).

15. C. J. Thompson and J. L. McBride, *Math. Biosci. 21*:127–142 (1974).

16. B. L. Jones, R. H. Enns, and S. S. Rangnekar, *Bull. Math. Biol. 38*:15–18 (1975).

17. H. Haken and H. Sauermann, *Z. Physik 173*:261–275 (1963); *176*:47–62 (1963).

18. R. Feistel and W. Ebeling, *Stud. Biophys. (Berlin) 71*:139 (1976).

19. B. L. Jones, *J. Math. Biol. 6*:169–175 (1978).

20. R. A. Fisher, *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, 1930 [2nd rev. ed., Dover Publications, New York, 1958].

21. J. B. S. Haldane, *The Causes of Evolution*. Harper and Row, New York, 1932.

22. S. Wright, *Evolution and the Genetics of Populations*, Vols. I–IV. The University of Chicago Press, Chicago, 1968, 1969, 1977, and 1978.

23. B. B. Mandelbrot, *The Fractal Geometry of Nature*. Freeman, New York, 1983.

24. T. E. Creighton, *Proteins—Structures and Molecular Principles*. Freeman, New York, 1983.

25. R. G. Casten and C. J. Holland, *J. Diff. Eqs. 27*:266–273 (1978).

26. S. Shahshahani, *A New Mathematical Framework for the Study of Linkage and Selection*. Memoirs AMS, *211*, 1979.

27. K. Sigmund, "The Maximum Principle for Replicator Equations." In W. Ebeling and M. Peschel (eds.), *Lotka-Volterra-Approach to Cooperation and Competition in Dynamic Systems*. Akademie-Verlag, Berlin, 1985, pp. 63ff.

28. P. Schuster and K. Sigmund, *Ber. Bunsenges. Phys. Chem. 89*:668–682 (1985).

29. J. S. McCaskill, *J. Chem. Phys. 80*:5194 (1984).

30. K. M. Watson, *Phys. Rev. 105*:1388 (1957).

31. P. W. Anderson, *Phys. Rev. 109*:1492 (1958).

32. H. A. David, *Order Statistics*, Wiley, New York, 1970.

33. E. N. Economou and M. H. Cohen, *Phys. Rev. 135*:2931 (1972).

34. E. Batschelet, E. Domingo, and C. Weissmann, *Gene 1*:27 (1976).

35. E. Domingo, R. A. Flavell, and C. Weissman, *Gene 1*:3 (1976).

36. C. K. Biebricher, *Evolut. Bio. 16*:1 (1983).

37. R. K. Selander and B. R. Levin, *Science 210*:545 (1980).

38. D. E. Dykhuizen and D. L. Hartl, *Genetic 96*:801 (1980).

39. D. L. Hartl and D. E. Dykhuizen, *Proc. Natl. Acad. Sci. U.S.A. 78*:6344 (1981).

40. M. Eigen and P. Schuster (unpublished).

41. H. F. Schaefer III (ed.), *Modern Theoretical Chemistry, Vol. 3: Methods of Electronic Structure Theory*, Plenum Press, New York, 1977.

42. J. Swetina and P. Schuster, *Biophys. Chem. 16*:329 (1982).

43. R. Kindermann and J. L. Snell, *Contemporary Mathematics, Vol. 1: Markov Random Fields and Their Applications*, American Mathematical Society, Providence, R.I., 1980.

44. R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London, 1982.

45. L. Demetrius, Polynucleotide Replication and Statistical Mechanics, preprint, 1987.

46. I. Leuthäusser, *J. Chem. Phys. 84*:1884 (1986).

47. I. Leuthäusser, Statistical Mechanics of Eigen's Evolution Model, *J. Stat. Phys. 48*:343 (1987).

48. L. Demetrius, P. Schuster, and K. Sigmund, *Bull. Math. Biol. 47*:239 (1985).

49. L. Demetrius, *J. Stat. Phys. 30*:709 (1983).

50. J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A. 79*:2554 (1982).

51. J. S. McCaskill, *Biol. Cybern. 50*:63 (1984).

52. P. Schuster and J. Swetina, *Bull. Math. Biol. 50*:635 (1988).

53. W. Fontana and P. Schuster, *Biophys. Chem. 26*:123 (1987).

54. M. S. Waterman, T. F. Smith, and W. A. Beyer, *Adv. Math. 20*:367 (1976).

55. C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry*, Vol. 3, Freeman, San Francisco, 1980, pp. 1183ff.

56. M. Zuker and D. Sankoff, *Bull. Math. Biol. 46*:591 (1984).

57. P. W. Anderson, *Proc. Natl. Acad. Sci. U.S.A. 80*:3386 (1983).

58. D. Pörschke, in I. Pecht and R. Rigler (eds.), *Chemical Relaxation in Molecular Biology*, Springer-Verlag, Heidelberg, 1977, pp. 191ff.

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

60. C. K. Biebricher, M. Eigen, and W. C. Gardiner, *Biochemistry 23*:3186 (1984).

61. C. K. Biebricher, M. Eigen, and W. C. Gardiner, *Biochemistry 24*:6550 (1985).

62. C. K. Biebricher, M. Eigen, in *RNA Genetics*, Vol. I, E. Domingo, J. J. Holland, and Ahlquist (eds.), CRC Press, Boca Raton, FL 1–21 (1988).

63. M. Eigen and C. K. Biebricher, in *RNA Genetics*, Vol. III, E. Domingo, J. J. Holland, and Ahlquist (eds.), CRC Press, Boca Raton, FL 211–245 (1988).

64. E. Domingo, E. Martina-Salas, F. Sobrino, J. C. de la Torre, A. Portela, J. Ortin, C. Lopez-Galindez, P. Perez-Brena, N. Villanueva, R. Najera, S. Van de Pol, S. Steinhauer, N. DePolo, and J. Holland, *Gene 40*:1 (1986).

65. J. Holland, K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. Van de Pol, *Science 215*:1577 (1982).

66. E. Domingo, D. Sabo, T. Taniguchi and C. Weissmann, *Cell 13*:735 (1978). E. Domingo, A. Flavell, and C. Weissmann, *Gene 1*:3 (1976). E. Batschlelet, E. Domingo, and C. Weissmann, *Gene 1*:27 (1976).

67. E. Domingo, M. Davila, and J. Ortin, *Gene 11*:333 (1980).

68. J. D. Parvin, A. Moscona, W. T. Pan, J. Lieder, and P. Palese, *J. Virol. 59*:377 (1986).

69. K. R. Spindler, F. M. Horodyski, and J. J. Holland, *Virology 119*:96 (1982).

70. S. Benzer, *Proc. Natl. Acad. Sci. U.S.A. 47*:403 (1961).

71. S. Benzer, *Proc. Natl. Acad. Sci. U.S.A. 45*:1607 (1959).

72. D. R. Mills, R. L. Peterson, and S. Spiegelman, *Proc. Natl. Acad. Sci. U.S.A. 58*:217 (1967).

73. F. R. Kramer, D. R. Mills, P. E. Cole, T. Nishihara, and S. Spiegelman, *J. Mol. Biol. 89*:719 (1974).

74. D. R. Mills, F. R. Kramer, and S. Spiegelman, *Science 180*:916 (1973).

75. C. K. Biebricher, M. Eigen, and W. C. Gardiner, *Biochemistry* (submitted).

76. M. Sumper and R. Luce, *Proc. Natl. Acad. Sci., U.S.A. 72*:162 (1975).

77. C. K. Biebricher, M. Eigen, and R. Luce, *J. Mol. Biol. 148*:361 (1981).

78. C. K. Biebricher, M. Eigen, and R. Luce, *J. Mol. Biol. 148*:391 (1981).

79. C. K. Biebricher, M. Eigen, and R. Luce, *Nature 321*:89 (1986).

80. H. Otten, Dissertation, Braunschweig, 1988.

81. A. Schwienhorst, Diplom-Arbeit, Bielefeld, 1988.

82. C. K. Biebricher, *Cold Spring Harbor Symp., Quant. Biol. 52*: (in press).

83. C. W. Gardiner, *Handbook of Stochastic Methods*, Springer-Verlag, Berlin, 1983.

84. W. Ebeling and R. Feistel, *Ann. Physik 34*:81 (1977).

85. P. H. Richter, *Bull. Math. Biol. 37*:193 (1975).

86. B. L. Jones and H. K. Leung, *Bull. Math. Biol. 43*:665 (1981).

87. R. Heinrich and I. Sonntag, *J. Theor. Biol. 93*:325 (1981).

88. H. K. Leung, *Bull. Math. Biol. 46*:399 (1984).

89. H. K. Leung, *Bull. Math. Biol. 47*:231 (1985).

90. H. Inagaki, *Bull. Math. Biol. 44*:17 (1982).

91. A. F. Bartholomay, *Bull. Math. Biophys. 20*:97 (1958).

92. P. Schuster and K. Sigmund, *Bull. Math. Biol. 46*:11 (1984).

93. B. M. Pötscher, *Bull. Math. Biol. 47*:263 (1985).

94. E. D. Weinberger, Dissertation, New York University, 1987.

95. D. Gillespie, *J. Comp. Phys. 22*:403 (1976).

96. M. A. Nowak, Eine numerische Simulation der RNA-Replikation zur Berechnung der Stochastischen "Error-Threshold." Diploma Thesis, Universität Wien, 1987.