Physical aspects of evolutionary optimization and adaptation

Walter Fontana, Wolfgang Schnabl, and Peter Schuster*
Institut für Theoretische Chemie der Universität Wien, Währingerstrasse 17, A 1090 Wien, Austria
(Received 2 February 1989; revised manuscript received 5 May 1989)

A model of an objective function based on polynucleotide folding is used to investigate the dynamics of evolutionary adaptation in finite populations. Binary sequences are optimized with respect to their kinetic properties through a stochastic process involving mutation and selection. The objective function consists in a mapping from the set of all binary strings with given length into a set of two-dimensional structures. These structures then encode the kinetic properties, expressed in terms of parameters of reaction probability distributions. The objective function obtained thereby represents a realistic example of a highly "rugged landscape." Ensembles of molecular strings adapting to this landscape are studied by tracing their escape path from local optima and by applying multivariate analysis. Effects of small population numbers in the tail of the sequence distribution are discussed quantitatively. Close upper bounds to the number of distinct values produced by our objective function are given. The distribution of values is explored by means of simulated annealing and reveals a random scatter in the locations of optima in the space of all sequences. The genetic optimization protocol is applied to the "traveling salesman" problem.

I. OPTIMIZATION OF COMPLEX SYSTEMS

New optimization algorithms derived from physics and biology were applied within the past two decades to systems with highly "rugged cost functions"[1]—for a recent review summarizing applications to the "traveling salesman" problem see Ref. 2. Methods from statistical physics of spin glasses—as "simulated annealing"[3]—from neural network dynamics[4] and from the genetic theory of evolution[5,6] became widely known. These methods—though being apparently distinct and having their undeniable individual advantages and merits—share common features. To give an example: replication and mutation dynamics in populations multiplying without recombination (i.e., asexually) was shown to correspond to statistical mechanics of equilibrium spin systems.[7–9] An extremal principle which is formally identical to the minimization of free energy in statistical mechanics holds for the optimization of replication performance in infinite populations.[7,8] Techniques originating in the study of electronic localization in random potentials can be applied to identify changes in the "localization" behavior of mutant distributions.[10,11] "Localization" refers here to an abstract space which houses all possible solutions of the combinatorial optimization problem. For the optimization of properties encoded by strings of two symbols this abstract space is known as sequence space.[12]

Optimization becomes a hard problem in finite populations, when available resources—the product of time and population capacity, so to say—are vanishingly small compared to the set of possible candidates for solutions to the problem. This is generally so for combinatorial tasks. The optimization of properties encoded by the structure of polynucleotides represents such a task that also might have played a role in prebiotic evolution. The evolutionary adaptation of a finite ensemble of molecular species must be described by stochastic means. Every finite distribution of polynucleotides created by replication and mutation has a periphery of scarcely populated mutants that forces a stochastic treatment. In addition, mutation as such is an intrinsically stochastic event.

In this contribution we shall be concerned with an optimization technique characterized as "evolution reactor."[13] It is a variant of conventional genetic algorithms[6] and mimics adaptation of viruses in its most simple form. Experimental efforts to determine distributions of sequences in real populations are enormous, not to mention the formidable task of tracking population dynamics in phase space. Computer experiments capturing essential features are an alternative. The usage of configuration landscapes reflecting qualitatively correct features of the realistic cases is crucial for the relevance of the results. Biological systems typically exhibit relatively simple modular components tied into a web of interactions that result in surprising global properties. On the scale of cellular regulation such systems might be modeled by Boolean networks.[14] Evolutionary dynamics on landscapes resulting from such a network[15] or from the more specialized case of a spin-glass function with binary interactions[16] have been studied. On a molecular scale, which we consider here, modules might be as simple as monomers whose global interactions produce a complex two- or three-dimensional structure.

The present contribution is organized as follows. In Sec. II we introduce the reaction network. Section III describes the mapping that assigns to each binary string an integer valued net productivity—defining its "fitness"—by first computing the string's secondary structure. We proceed, then, to derive an algorithm based on counting partitions of integers for establishing close upper limits on the number of different fitness values in our model. This will guide us in discussing the density distribution of values. However, it is the distribution of fitness values in sequence space which is pivotal to

evolutionary dynamics. In Sec. IV we focus by means of genealogies on two scenarios of optimization dynamics related to distinct topological features of the fitness landscape. Effects of low occupation numbers are made quantitative and result in a threshold below which accurate evaluation of fitness parameters is jeopardized. Section V uses tools from multivariate analysis to provide information of the changes occurring in the distribution of sequences during the escape from a local fitness trap. In Sec. VI we present strong evidence for a random scatter in sequence space of clustered optima. This is in contrast to ultrametric distributions found in mean-field spin-glass models[17] or in the "traveling salesman" problem.[18] Section VII reports an application of the evolutionary reaction network to the traveling salesman problem. Section VIII concludes the paper.

II. AN OPTIMIZATION ALGORITHM DERIVED FROM MOLECULAR EVOLUTION

Recently, we conceived an evolutionary computer game[13]—for a schematic illustration see Fig. 1—and applied it to a model of polynucleotide replication. The model system consists of a reaction network describing

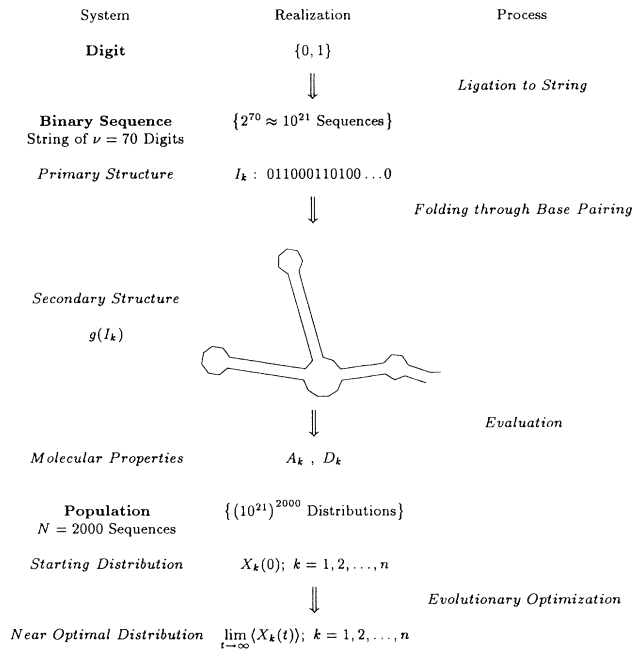

FIG. 1. A computer model of evolutionary adaptation. The model mimics RNA folding and evaluation of secondary structures according to some predetermined rules computing replication (A_k) and degradation rate constants (D_k). Real polynucleotides are replaced by binary sequences which fulfill a complementary pairing principle, $0 \equiv 1$, in analogy to the base pairing rules. Secondary structures are formed by means of a minimum free-energy criterion. Optimization is performed on the level of populations according to the Darwinian rules of mutation and natural selection.

error-free replication, mutation, degradation, and dilution as parallel processes. It has been introduced by Eigen[19] and was analyzed further by means of deterministic differential equations.[11,20–25] The reaction scheme provides n^2 polymerization channels of n polynucleotides I_k $(k=1,\ldots,n)$, n degradation pathways, and the unspecific dilution flux $\phi(t)$:

$$I_k \xrightarrow{A_k Q_{kk}} 2I_k \ , \tag{1}$$

$$I_k \xrightarrow{A_k Q_{jk}} I_k + I_j \ , \tag{2}$$

$$I_k \xrightarrow{D_k} 0 \ , \tag{3}$$

$$I_k \xrightarrow{\phi(t)} 0, \quad j,k=1,\ldots,n \ . \tag{4}$$

Herein, A_k is the rate constant of replication using I_k as template, D_k its degradation rate constant, and Q_{jk} denotes the probability of producing I_j as a mutant of I_k. Here we model the polynucleotides as binary strings of fixed length $\nu, I_k \in \{0,1\}^\nu$, and consider exclusively mutations that interchange digits, hence leaving the chain length ν constant. The probability q of correctly copying a digit is assumed to be constant along the chain and independent of the digit type. The mutation probabilities then take the simple form

$$Q_{jk} = q^{\nu - d(j,k)}(1-q)^{d(j,k)} \ , \tag{5}$$

where $d(j,k)$ denotes the distance in the Hamming metric of the ν-dimensional cubic sequence space. We define the all-0-sequence to be the origin of the hypercube. The expression "k-error class" denotes the set of $\binom{\nu}{k}$ sequences with k 1 digits ("errors" with respect to the origin) and $\nu-k$ 0 digits. The expression "k-error environment" denotes the k-error class together with neighboring error classes according to context. In order to achieve canonical selection constraints, the flux $\phi(t)$ is adjusted to yield conservation of the number of polymer molecules. From this it follows that the flux compensates the mean excess production $\bar{E}(T)$ of the network, $\phi(t) = \bar{E}(t) = \sum_k E_k x_k(t)$, with $E_k = A_k - D_k$ being the net productivity of I_k and $x_k(t)$ its fraction in the system $-\sum_{j=1}^{n} x_j = 1$. These relative concentrations change according to

$$\dot{x}_k(t) = [w_{kk} - \bar{E}(t)]x_k(t) + \sum_{j\,(\neq k)} w_{kj}x_j(t) \ , \qquad k,j=1,\ldots,n \ . \tag{6}$$

In the deterministic scenario the system (6) relaxes towards a unique stationary sequence distribution, which is given by the dominant eigenvector of the matrix $\underline{W} = \{w_{kj}\}$, $w_{kj} = A_j Q_{kj} - D_j \delta_{kj}$. This stationary distribution was characterized as "quasispecies"[20] in order to point at the analogy to the notion of species in biology. The asymptotic value of the mean productivity $\lim_{t\to\infty} \bar{E}(t) = \phi(\infty)$ matches the corresponding largest eigenvalue $\lambda_{\max}$.

Because of nonzero coupling terms Eq. (6) is nonlocal in the relative sequence concentrations. It describes the

selection process and its outcome, the quasispecies in the limit of infinite populations. It addresses neither the question of how and under which conditions the fittest sequence I_m, with $w_{mm}=\max_k\{w_{kk}\}$, is found in a finite population, nor the question of how sequence distributions move in sequence space. In particular, the occurrence of local traps and the mechanism by which populations escape from them is a typical stochastic feature.

A single stranded RNA sequence is known to acquire a three-dimensional structure by folding back on itself due to intramolecular pairing of complementary bases. Here we restrict ourselves to the two-dimensional folding pattern termed "secondary structure" (see, for example, Fig. 3). A secondary structure is a planar graph whose vertices are the bases constituting the particular sequence. The set of edges consists of $\nu-1$ "backbone" edges connecting adjacent bases in the sequence and a number of edges that represent hydrogen bonded Watson-Crick pairs. The structures resulting upon folding display typical features. Regions of unpaired digits that are closed by one or more base pairs are termed loops. They appear in several forms (see Figs. 1 and 3): one-sided bulges, hairpin loops, and internal loops with more than one branch emanating from them. Regions of contiguous base pairs are referred to as stacks or helical regions. In our figures they are drawn as parallel lines, leaving out the edges corresponding to the pairings for the sake of less congested images. Unpaired ends or open junctions connecting two stacks are denoted as external elements.

RNA secondary structure is known to influence the replication velocity along the strand[26] and the overall stability against hydrolytic degradation. Our interest in mimicking RNA folding derives from the fact that RNA is the simplest example for combining the genotypic (sequence) level and the phenotypic (structure) level into a single molecular object. The mapping from one level to the other is provided by sequence-specific intramolecular interactions.

Each feature of the secondary structure contributes additively an energy increment to the overall free energy of the folding pattern. Stacks are the only stabilizing elements. The exponentially many structures that are possible for a given sequence differ therefore in their thermodynamic stability. To a particular sequence I_k (the genotype) we assign the energetically most stable secondary structure (the phenotype), $g(I_k)$, using an energy optimization algorithm derived from secondary structure prediction.[27,28] Ties are resolved deterministically. We consider binary sequences of length $\nu=70$ and a single pairing rule, where 0 is defined to be complementary to 1. The physical parameters of base pairing and stacking were adopted from data on G (guanine) and C (cytosine) containing polyribonucleotides.

The folded sequence is subsequently parsed into its loops, stacks, and external elements—like junctions and unpaired ends. This information is then used to build heuristic replication (A_k) and degradation (D_k) parameters (see Sec. III). More precisely, $R_k^{(1)}=A_k$, $R_k^{(2)}=D_k$, and $R_k^{(3)}=\phi(t)$ are parameters of exponential reaction probability densities $P_k^{(\alpha)}(t)$, $\alpha=1,2,3$, for sequence I_k:

$$P_k^{(\alpha)}(t)=R_k^{(\alpha)}\exp(-R_k^{(\alpha)}t),\quad t>0,\quad \alpha=1,2,3\ . \tag{7}$$

$P_k^{(\alpha)}(t)dt$ is the probability that sequence I_k reacts during the time interval $(t,t+dt)$ according to the replication channel [$\alpha=1$, Eqs. (1) and (2)], the degradation channel [$\alpha=2$, Eq. (3)] or disappears through dilution [$\alpha=3$, Eq. (4)]. The parameters A_k and D_k correspond to the average number of replication and degradation events, respectively, occurring in a unit time interval and determine the "fitness" of I_k. Hence, in our stochastic model, fitness is a property which makes sense only in terms of an expectation value (see Sec. IV). Nevertheless, we refer to A_k and D_k as "rate constant" for convenience.

Evolutionary optimization dynamics is simulated by means of an algorithm which was designed for stochastic chemical reactions by Gillespie.[29,13] A logistically constrained population of $\Theta=2000$ sequences then reacts according to the network described by Eqs. (1)–(4), whereby each currently existing reaction channel is chosen according to an exponential density with corresponding parameter A_k, D_k, or

$$\phi(t)=\sum_k N_k(t)(A_k-D_k)/\Theta\ ,$$

where $N_k(t)$ denotes the number of sequences of type I_k at time t.

III. THE FITNESS LANDSCAPE

The kinetic assignment defines replication and degradation rate densities separately. Each stacking region is weighted sigmoidally depending on its length and contributes additively to the overall replication slowdown due to base pairing,

$$A_k'=\kappa_R-\kappa_1\sum_j\frac{n_j^{(k)}(1+n_j^{(k)})^3}{(1+n_j^{(k)})^4+L},\quad k=1,2,\ldots,2^\nu \tag{8}$$

whereas each unpaired loop or external region increases in a noncooperative way the overall degradation rate:

$$D_k'=\kappa_D+\kappa_2\sum_j\frac{u_j^{(k)}}{u_m}\exp[(u_j^{(k)}-u_m)/u_m]+\kappa_3\frac{1}{\nu}\sum_l\nu_l^{(k)}\ ,$$
$$k=1,2,\ldots,2^\nu\ . \tag{9}$$

In both Eqs. (8) and (9) κ_R, κ_D, κ_1, κ_2, κ_3, and L are scaling constants, $n_j^{(k)}$ is the number of base pairs in the jth stack of sequence I_k, $u_j^{(k)}$ is the number of unpaired digits in the jth loop of I_k, u_m is the maximum number of weighted loop digits, $\nu_l^{(k)}$ is the length of the lth external element (free end, junction), and ν is the length of the sequence. In the work reported here, the following numerical values were used: $\kappa_R=30$, $\kappa_D=0$, $\kappa_1=30$, $\kappa_2=20$, $\kappa_3=40$, $L=10\,000$, $u_m=25$, and $\nu=70$. These parameter values were chosen such that they reflect experimental data when available.

In our implementation A_k is set to zero whenever it turns out negative, and both A_k and D_k are truncated to integer after shifting two decimal places:

$$A_k=[\max(0,A_k')100],\quad D_k=[D_k'100]\ , \tag{10}$$

where $[x]$ denotes the largest integer less or equal to x. The density of fitness states was analyzed numerically by

sampling in different regions of the hypercubical sequence space. The density distribution of replication and degradation rate constants is shown in Fig. 2—as an illustrative example we sampled sequences containing twenty 1-digits on the average. Note that Fig. 2 is not a bar chart. The densities of similar values of the rate constants are so irregular that connecting the individual points results in such a bizarre plot of the function. Moreover, the densities do not provide any information about the relative locations of configurations realizing particular values. The densities of replication rate constants, $\rho(A_k)$, consist of intensive discrete bands and reveal vast scatter as a result of their sensitivity to structural details. The densities of degradation rate constants, $\rho(D_k)$, on the other hand, show roughly Gaussian shapes and have a significantly more dense support. We proceed now to develop a simple argument that explains this behavior and gives a means for computing upper bounds on the possible number of distinct fitness values in our model.

Let $S^{(k)}$ be a multiset (i.e., a set in which elements can appear in multiple copies) of stacking regions of a given secondary structure $g(I_k)$ with n base pairs,

$$S^{(k)}=\{n_1^{(k)},n_2^{(k)},\ldots,n_j^{(k)}\},\quad n_i^{(k)}>0\ .$$

$S^{(k)}$ represents a partition of n: $n=\sum_i n_i^{(k)}$. Let $U^{(k)}$ be a multiset of unpaired regions of $g(I_k)$, which sum up to

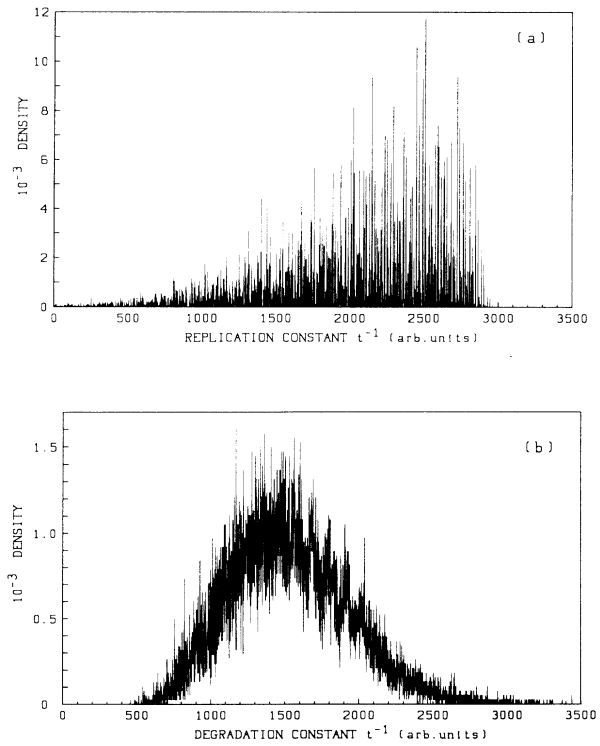

FIG. 2. Distribution of kinetic parameters. (a) Density of replication rate constants A_k obtained by sampling binomially 38 000 sequences of length $\nu=70$ with on average 20 1-digits. (b) Corresponding density of degradation constants D_k (Ref. 13).

the corresponding total length of $m=\nu-2n$. Clearly,

$$A_k=f_A(S^{(k)}),\quad D_k=f_D(U^{(k)})\ , \tag{11}$$

where f_A and f_D denote the functions given in Eqs. (8) and (9), respectively. Note that with respect to Eq. (9) the sum over all free ends and/or junctions is equivalent to one external unpaired region. By virtue of Eq. (11) structures with the same number of base pairs (n) have different rate constants, if they realize different partitions of n and/or m. For fixed n the total number of possible partitions $p(n)$ is given by Euler's recursion

$$p(n)=\frac{1}{n}\sum_{m(<n)}\sigma(n-m)p(m)\ ,$$

with $\sigma(k)$ being the sum over all divisors of k and $p(0)=1$.

The mappings from sequence I_k to A_k and D_k contain the following hierarchical levels of degeneracy.

(1) Many I_k lead to the same two-dimensional (2D) structure $g(I_k)$.

(2) Many structures $g(I_k)$ containing n base pairs are characterized by the same $S^{(k)}$ and/or $U^{(k)}$.

(3) Different $S^{(k)}, U^{(k)}$ are mapped by Eqs. (11) onto the same real-valued A_k and D_k, respectively.

(4) The truncation to non-negative integers in Eq. (10) causes a "coarse graining" degeneracy.

(5) Structures are actually mapped on (A_k,D_k) pairs and distinct pairs may result in the same net productivity $E_k=A_k-D_k$.

Levels 1 and 2 depend on the folding procedure. This phenotype function contains an "unpredictable" element, in the sense that there is no shortcut to the application of the folding algorithm in obtaining the structure of I_k. Levels 3 and 4 depend on the actual implementation of the functions f_A and f_D. Their contribution to the overall degeneracy can be taken into account by explicitly computing f_A and f_D for all possible arguments $S^{(k)}$ and $U^{(k)}$.

It is therefore sensible to ask how many $S^{(k)}$ and $U^{(k)}$ for structures with exactly n base pairs do exist, that will be treated as distinct arguments in the functions f_A and f_D. In other words, how many different A_k and D_k values are possible before entering level 3 as a function of the number n of base pairs in the structures?

Let b denote the number of bifurcation points in $g(I_k)$ [i.e., structural elements where contiguous pairs branch into more "arms" without the need of intervening unpaired digits, see Fig. 5(d) structure C], moreover let $e=1$, if $g(I_k)$ has any external digits and $e=0$ otherwise. Then we have

$$|S^{(k)}|-b=|U^{(k)}|-e\quad\forall k\ . \tag{12}$$

On large samples in different parts of the sequence space we found that usually $b=0$ and in 87–92 % of the cases $e=1$, i.e., the number of stacks equals the number of loops. If Eq. (12) is the only constraint, then—referring to structures with n base pairs and $m=\nu-2n$ unpaired digits—we get for the number of elements in the sets

$\{S^{(k)}\}_n$ and $\{U^{(k)}\}_n$, which represent arguments to the f_A and f_D functions, respectively:

$$l=\min(n,m) ,$$

$$\mathcal{S}_n=|\{S^{(k)}\}_n|=\sum_{s=1}^{l} p(n,s) , \tag{13a}$$

$$\mathcal{U}_n=|\{U^{(k)}\}_n|=\sum_{u=1}^{l} p(m,u)+\sum_{u=1}^{l+1}\sum_{i=1}^{p(m,u)} t_i(m,u) , \tag{13b}$$

where $p(k,l)$ is the number of partitions of k with length l (i.e., $k=\sum_{i=1}^{l} n_i$), $p(k,k+1)=0$, and $t_i(m,u)$ is the number of distinct terms in the ith partition of m with length u.

The function f_A is symmetric with respect to all terms of a partition, but f_D breaks this symmetry with respect to terms that represent loops and the one specifying the size of the external region [$\sum_l \nu_l$ in Eq. (9)]. This is the reason for the second sum in (13b); it refers to $e=1$, i.e., $u-1$ loops and 1 external region, whereas the first sum of (13b) refers to $e=0$, i.e., u loops.

Stacking regions consist of two adjacent base pairs at least, and since they are the only stabilizing elements a structure will very rarely contain isolated pairs (this occurred in 1–1.5 % of our samples). Thus the overwhelming majority of structures realize those partitions $p^*(n)$ of n whose terms are all >1. This greatly reduces the number of actually realized partitions. Furthermore, our model requires at least four unpaired digits in each hairpin turn. Taking into account all these constraints we obtain in analogy to Eq. (13)

$$\hat{s}=\max_{p^*(n,s)>0} s=[n/2] ,$$

$$\hat{u}=\max_{p^{(1)}(m,u)>0} u=m-3 ,$$

$$l=\min(\hat{s},\hat{u}) ,$$

$$\mathcal{S}_n=\sum_{s=1}^{l} p^*(n,s) , \tag{14a}$$

$$\mathcal{U}_n=\sum_{u=1}^{l} p^{(1)}(m,u) + \sum_{u=1}^{l+1}\sum_{i=1}^{p^{(1)}(m,u)} [t_i(m,u)-H(1-\theta_i(m,u))] , \tag{14b}$$

where $p^{(i)}(m,u)$ denotes the number of partitions of m with length u having at least i terms ≥ 4, $H(x)$ is the Heaviside function—$H(x)=1$ for $x\geq 0$, $H(x)=0$ otherwise—and $\theta_i(m,u)$ counts the actual number of terms ≥ 4 in the ith partition of m with length u. Equation (14) is valid independently of the number of hairpins in the structure, since the D_k function does not distinguish between hairpin loops and internal loops. Hence b does not enter Eqs. (13) and (14). We cannot express $\mathcal{S}_n$ or $\mathcal{U}_n$ in terms of elementary functions, but it is easy to generate the required partitions on a computer[30] and hence to obtain p^*, $p^{(i)}$, and t_i. As an example with $\nu=70$ let us take $n=19$ base pairs, which is frequently realized in the 20-error environment (most of the 1-digits will pair) of the all-0-sequence (Fig. 2). The number of possible arguments $\mathcal{S}_{20}$ is 105. They are mapped by f_A after truncation to an interval ranging from $\min_k f_A(S^{(k)})=29$ to $\max_k f_A(S^{(k)})=2815$. In this case all arguments result in distinct replication constants. On the other hand, we get $\mathcal{U}_{20}=25\,690$ arguments $U^{(k)}$ mapped by f_D on an interval of the natural numbers ranging from $\min_k f_D(U^{(k)})=1088$ to $\max_k f_D(U^{(k)})=2400$. Here the truncation to integers produces a lot more degeneracy. The 25 690 possible arguments result in 1052 different degradation constants. Equation (14) represents a very reasonable upper bound with respect to our model. Some partitions may correspond to structures that are actually not optimal for any binary sequence I_k of length $\nu=70$. This cannot be taken into account because of the complexity of the interactions that are responsible for the secondary structure $g(I_k)$. However, the effect is rather small, since it is highly probable that out of many structures represented by a single partition at least one will be realized by some I_k.

The above argument explains why in Fig. 2 the density of the degradation constants has a much more dense support than that of the replication constants. It reproduces also the difference in the densities of one order of magnitude. Clearly, sampling binomially about the 20-error environment of the all-0-sequence will also produce structures with more or less than $n=19$ base pairs. The data for $n=18$ or 20 lead to a completely analogous result. We conclude that in our model two major factors are responsible for the distinct statistics of replication and degradation constants for long chains.

- Isolated base pairs occur very rarely, thus reducing the set of possible $S^{(k)}$ (different replication constants).
- The distinction between unpaired regions due to loops and due to external digits significantly increases the number of distinct arguments $U^{(k)}$ (different degradation constants).

We note that the secondary structures of a given sequence I and its polarity inverted (indicated by a superscript dot, which means a swap end for end) and complemented ($\tilde{}$) sequence, $\tilde{I}^{\cdot}$, are identical. The same holds for the polarity inverted sequence $I^{\cdot}$ and the complementary $\tilde{I}$. Self-complementary regions are palindromes. They are by definition invariant under complementation and inversion. If an algorithm claims that a certain helical region is optimal, then it must also be optimal in the inverted and complemented sequence. This is clearly valid for any sequence alphabet, provided that structures are constructed by complementarity rules. Thus our fitness surface contains at most $2^{\nu-1}$ structurally distinct states, where every set $\{I_k, I_k^{\cdot}, \tilde{I}_k, \tilde{I}_k^{\cdot}\}$ with $I_k \neq I_k^{\cdot}$ contributes at most two structures. An example is shown in Fig. 3. The Hamming distance between two sequences I_k and $\tilde{I}_k^{\cdot}$ with degenerated structures is given by

$$d(I_k,\tilde{I}_k^{\cdot})=\nu-d(I_k,I_k^{\cdot})=|\nu-2d(0,I_k)|+2\lambda$$

$$\lambda=\begin{cases}0,2,4,\ldots,d(0,I_k) & \text{if } \nu \text{ is even}\\ 0,1,2,\ldots,d(0,I_k) & \text{if } \nu \text{ is odd.}\end{cases}$$

$d(0,I_k)$ defines the error class of I_k relative to the all-0-sequence. Densities in the regions centered around the d- and $(\nu-d)$-error classes are necessarily identical: the numbers of potential base pairs are the same and so are the sets of partitions, leading to A_k or D_k states with the same realization frequencies—provided the sample was chosen large enough to represent the error-class properly.

Next we count the maximum number of possible distinct $(S^{(k)},U^{(k)})$ combinations—remember that $E_k = f_A(S^{(k)}) - f_D(U^{(k)})$. The rate constants of any structure with more than one hairpin are identical to those of a suitable single-hairpin structure. Thus it suffices to consider only the latter. The combinations will depend on the number of possible bifurcation points, since each such point adds one term in $S^{(k)}$ without adding one in $U^{(k)}$. However, any structures containing bifurcations splitting into more than two hairpin arms have rate constants that are again identical to suitable structures with bifurcations resulting in exactly two hairpins. Equations (8) and (9) do not depend on the types of loops and stacks, nor on their relative position. This simplifies the counting procedure enormously. Let $\mathcal{C}_n$ denote the number of possible combinations for structures with n base pairs. Then

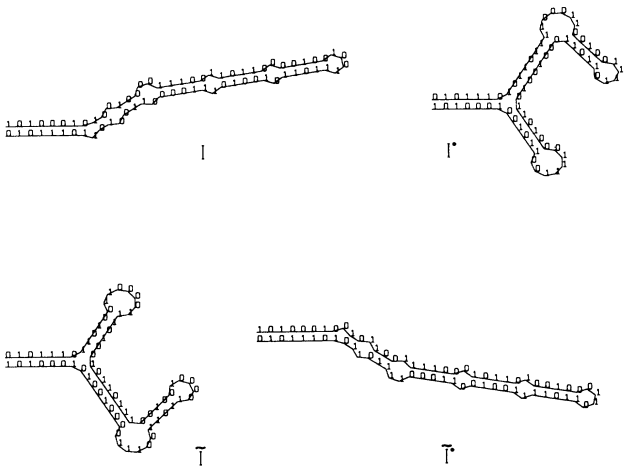

FIG. 3. Symmetry of sequence space with respect to secondary structures. Every sequence I at distance d from all-0-sequence origin has a structure identical to that of the complemented and inverted sequence $\bar{I}^{\cdot}$ at distance $\nu-d$ from the origin.

$$\mathcal{C}_n = |\{(S^{(k)},U^{(k)})\}_n| = \sum_{s=1}^{l}\left\{p^*(n,s)\left[p^{(1)}(m,s)+\sum_{i=1}^{p^{(1)}(m,s+1)}[t_i - H(1-\theta_i)] + \sum_{b=1}^{\lfloor (s-1)/2 \rfloor}\left[p^{(1+b)}(m,s-b)+\sum_{i=1}^{p^{(1+b)}(m,s-b+1)}[t_i - H(1+b-\theta_i)t_{\theta_i}]\right]\right]\right\}, \tag{15}$$

with l as in Eq. (14) and where the not yet defined t_{θ_i} denotes the number of distinct terms among those θ_i terms that are ≥ 4 (the minimum number of unpaired digits in a hairpin loop). Equation (15) counts all possible combinations by freely combining stack partitions of length $s, p^*(n,s)$, with partitions of unpaired regions, $p^{(i)}(m,u)$, that belong to structures as classified by (e,b) in Eq. (12). The four terms multiplied by $p^*(n,s)$ correspond to structures with $(e=0,b=0)$, $(e=1,b=0)$, $(e=0,b>0)$, and $(e=1,b>0)$, respectively. As with Eq. (14), $\mathcal{C}_n$ represents a very good upper bound. $\mathcal{C}(\nu)=\sum_{n=1}^{[(\nu-4)/2]}\mathcal{C}_n$ is the number of all possible combinations for chain length ν. Figure 4 shows a semilogarithmic plot of $\mathcal{C}(\nu)$ versus ν (upper curves). Curve 1 is a calculation of Eq. (15), curve 2 neglects bifurcation points (i.e., $b=0$). Computing $E_k = A_k - D_k$ (truncation to integer) for all partition combinations $\mathcal{C}(\nu)$ we obtain curve 3. In the case $\nu=70$ at most 5 752 290 different $(S^{(k)},U^{(k)})$ combinations enter the above-mentioned level 3. Levels 3–5 subsequently lead to an enormous degeneracy resulting in at most 4328 different fitness values, E_k, for $\nu=70$. Surprisingly, despite this degeneracy, the resulting fitness landscape is very rugged (see Ref. 13) and the realization frequencies of similar selective values fluctuate significantly (Fig. 2). For a discussion of the distri-

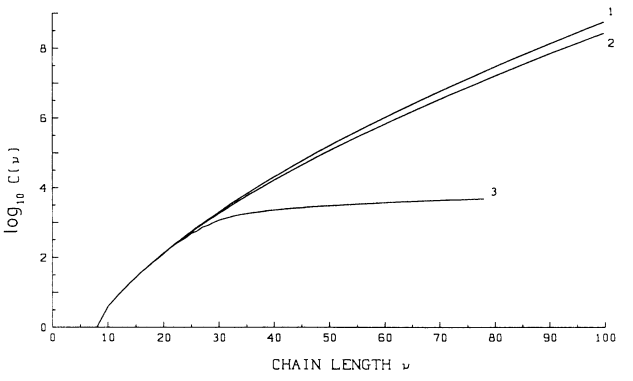

FIG. 4. Diversity of values in the model landscape. The logarithm of the number of partition combinations $\mathcal{C}(\nu)$ that are distinguished by Eqs. (8) and (9) is plotted vs the sequence length ν. Curve 1 corresponds to Eq. (15), curve 2 neglects the terms involving bifurcation points ($b=0$). Curve 3 represents the number of distinct net productivities $E_k = A_k - D_k$ obtained by computing the rate constants from all $\mathcal{C}(\nu)$ combinations. The bending off to a nearly constant function is due to the truncation to integer [Eq. (10)] and to the fact that most of the partitions are mapped via f_A and f_D on a fixed interval determined by the parameters in Eqs. (8) and (9).

bution of extreme values see Sec. VI.

It is interesting to note that the partition that maximizes E_k (no external elements, 12 loops of size 1, 1 loop of size 4, 12 stacks of size 2, 1 stack of size 3: $E_k = 2245$) corresponds to a structure that is not realized by any sequence. All strings that could fulfill the requirements for such a partition form other secondary structures on folding. The phenotype $g(I_k)$ is the result of a free-energy optimization that tends to minimize stack interruptions—a criterion which is exactly opposite to the E_k maximization. This captures a typical situation arising in the adaptation of complex systems. On the one side the phenotype is frequently itself the result of an optimization process that follows a certain set of rules (cf. polynucleotide or protein folding). On the other side, the same phenotype "codes" for an objective function, termed fitness, by determining the kinetic or functional properties of the molecule. This "code" is usually established by a different set of rules. Both sets often lead to conflicting goals by competing for available interactions. In the present case the mapping from sequence to structure is dominated by the stability of the latter. This is achieved essentially by seeking to establish as much intramolecular interactions (i.e., base pairs) as possible. But these interactions must be available during the replication process and hence stacking regions have to be opened by the copying machinery thereby causing a slowdown of the overall replication rate. As in real systems the phenotype represents a serious constraint with respect to fitness optimization.

IV. DYNAMICS ALONG GENEALOGIES

An alternate view providing some information on connectivity in sequence space can be obtained by following population dynamics along individual trajectories of the evolutionary process. For this goal all genealogies have been recorded during the computer experiments. In general, such genealogies will look more like nets than single lines, since a given species can be produced by many others at different times. For simplicity of presentation we restrict a genealogy to be a time series,

$$I_1 \rightarrow I_2 \rightarrow \cdots \rightarrow I_k \rightarrow I_{k+1} \rightarrow \cdots \rightarrow I_n \ ,$$

in which species I_{k+1} was not present in the population at the instant of its production from species I_k. In most of the cases this means that species I_{k+1} has been produced for the first time during the history of the system by species I_k. However, it may also be the case that I_{k+1} had been in the system for some time in the past and was lost thereafter. It is important to keep in mind that the genealogies we are reporting here record the history of successful sequences. They show in backwards view how the ultimate survivors of the "sequence competition game" came into existence.

In the following we consider two simulations with single digit accuracy $q = 0.999$ starting from homogenous populations of $\Theta = 2000$ all-0-sequences with chain lengths of $\nu = 70$ digits. One of them is the evolutionary jump reported in Ref. 13 and shown in Fig. 5(a). This jump is the consequence of a special rule for large loops in the degradation function D_k (9). Loops with more than $u_m = 25$ unpaired positions are evaluated as having exactly u_m hydrolytically accessible digits: the exponentials in Eq. (9) are identical to unity if $u_j^{(k)} > u_m$. Now imagine that, due to mutation, a loop with u_1 unpaired digits is tied off from a loop having $u > u_m$ free bases. If the number of digits in the remaining loop, $u - u_1$, continues to be larger than u_m, the new structure will have a lower selective value, since the replication rate has been decreased due to the additional helical region and the degradation rate has been increased due to the new loop. There is a particular structure which takes full advantage from the "large-loop rule." It has exactly three stacked pairs sticking together the ends and closing a large loop with $u = 64$ digits. Indeed, three pairs are the minimum required from the thermodynamic data available to the folding algorithm in order to stabilize such a large loop. It is easily seen that every structure derived from the best large-loop structure by a few mutations must possess a lower productivity $A_k - D_k$.

The large-loop rule has no biophysical relevance. It is even somewhat inconsistent to introduce a cutoff into the loop-size-dependent properties. Smooth bending off to a linear or logarithmic dependence upon u would be more reasonable. This inconsistency, however, does not change the many-valley characteristic of the fitness landscape and is in line with the naive translation from structure to kinetic parameters. In our model its purpose is to create an isolated fitness peak which serves as a potential "evolutionary trap."

The ensemble in experiment A evolves from a homogenous population of all-0-sequences and is trapped precisely by the large-loop structure discussed above. By "master sequence" we denote henceforth the species that is currently the most frequent. The effectiveness of this trap—a consequence of its depth—is readily seen in Fig. 6, where we have plotted the changes occurring in the Hamming distance between the current master sequences at all times and the virtual consensus sequences, i.e., the integer valued centers of mass, of the population. The digits of this virtual sequence $\sigma_i(t)$, $i = 1, 2, \ldots, \nu$, reflect the majorities at the positions i in the ensemble at time t:

$$\sigma_i(t) = \left[0.5 + \frac{1}{N(t)} \sum_k \sigma_i^{(k)} n_k(t) \right] .$$

During the 600 time units, which the population spends in the large-loop trap, the consensus sequence is identical to the master, indicating the presence of at most a few low populated neutral mutants and therefore a very narrow sequence distribution (see also Sec. V).

Now we focus on the genealogical path from the all-0-sequence to the trap and the escape from there to the master sequence at $t = 960$. Figure 7 shows the corresponding evolution of the structures and Fig. 8 displays a characteristic window in the kinetics of species succession in a resolution of one time unit. The sequence representing the large-loop trap is labeled A7 (this label refers to the simulation run). The complete succession dynamics of most frequent sequences from A7 to A27 is

shown in Table I. During the spreading of the initial population the large-loop master (M1=A7) is produced and rapidly dominates the population. After M1 is established, it lasts 555 time units until a "right" position happens to be mutated and a structure is built which, although having a lower selective value, is good enough to reach a significant number of copies for a short time. During this time many different paths were tried out in order to get out of the trap, but they led nowhere, one and all.

The main features of the escape from the large-loop trap are as follows.

- Many species of the genealogical path from A7 to the master at $t=960$ (M2=A27) are short-living fluctuations. The lifetime of the third intermediate between M1 and M2, A10, was even shorter than the resolution. Twelve of 19 intermediates never reach more than 20 particles. Six of them remain strictly below 10 individuals.
- The sharp transition in Fig. 5(a) is a reflection of the fast takeover of the new advantageous mutant M2 as soon as it has been produced.
- A list of the binary strings shows that the whole path from M1 to M2 has been covered by a series of one-digit mutations apart from a single event, A9→A10, which in-

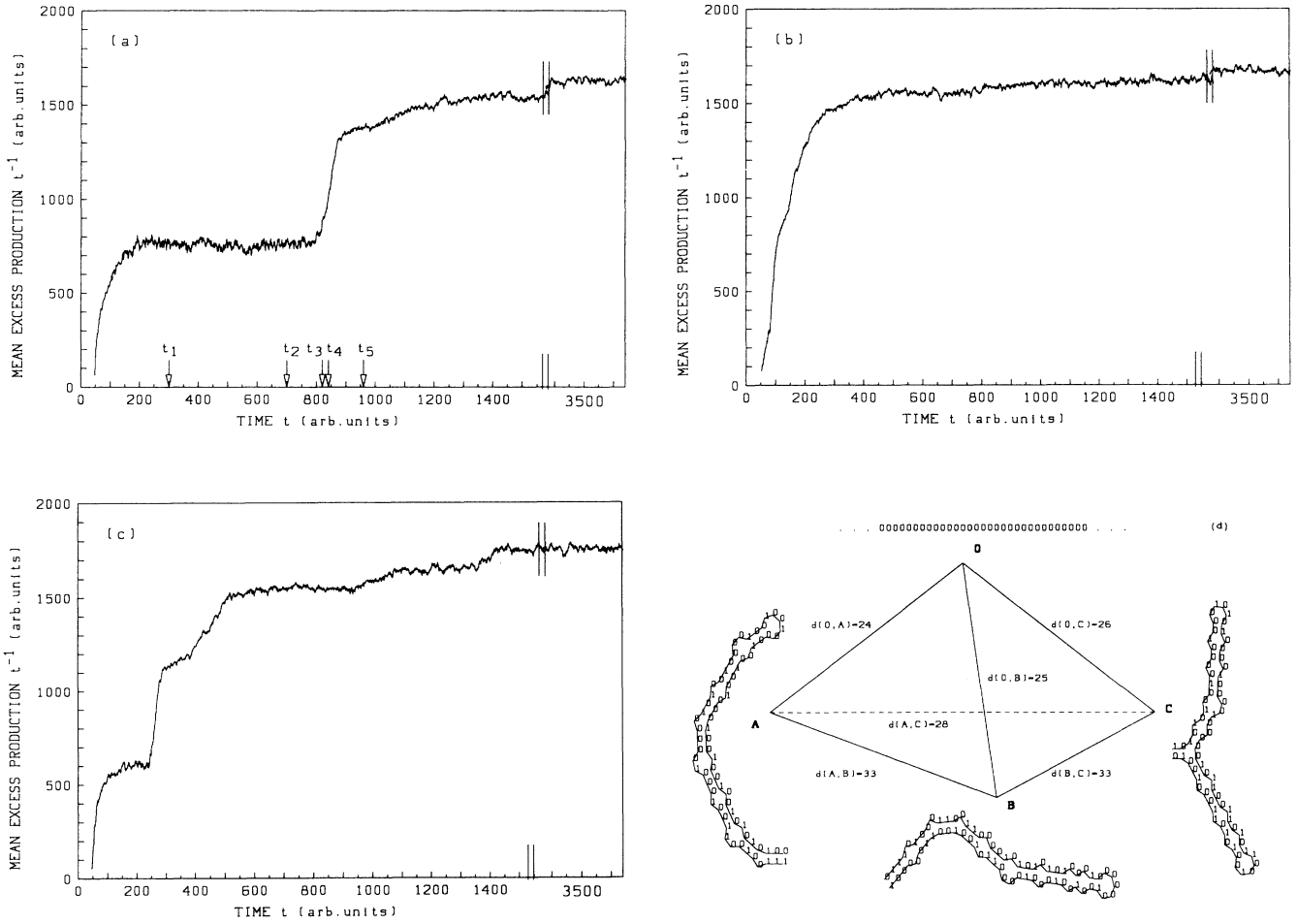

FIG. 5. Mean excess productivity during optimization. Three runs (A,B,C) starting all from identical initial conditions (homogeneous population of $\Theta=2000$ all-0-sequences with chain length $\nu=70$, single-digit replication accuracy $q=0.999$), but controlled by different pseudo-random-number series, show markedly distinct time evolutions of $\bar{E}(t)$. Run A was reported in Ref. 13. The arrows indicate the times at which a cluster analysis of the population distribution support is discussed in Sec. V. For details see text. (d) shows the best and most frequent sequences at termination time of the three experiments A,B,C. The three quasispecies centered about these sequences are pairwise disjoint and localized in completely different regions of the sequence space as indicated by their Hamming distances.

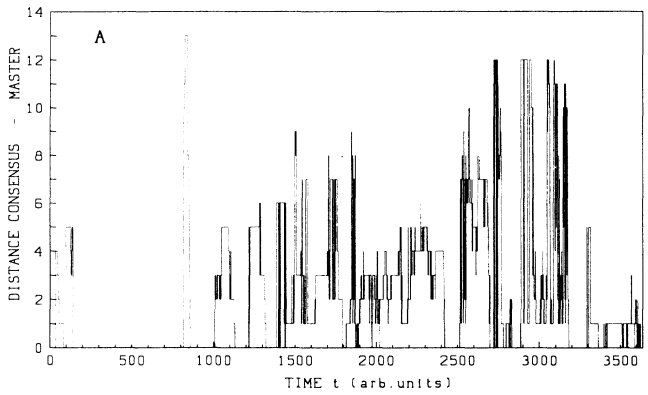

FIG. 6. Broadness of distributions. From the Hamming distance between the actual consensus sequence and master sequence we obtain information about the sharpness of a distribution. Large distances indicate that different species reach population levels comparable to the master sequence. Hence the master sequence no longer coincides with the integer-valued center of the population.

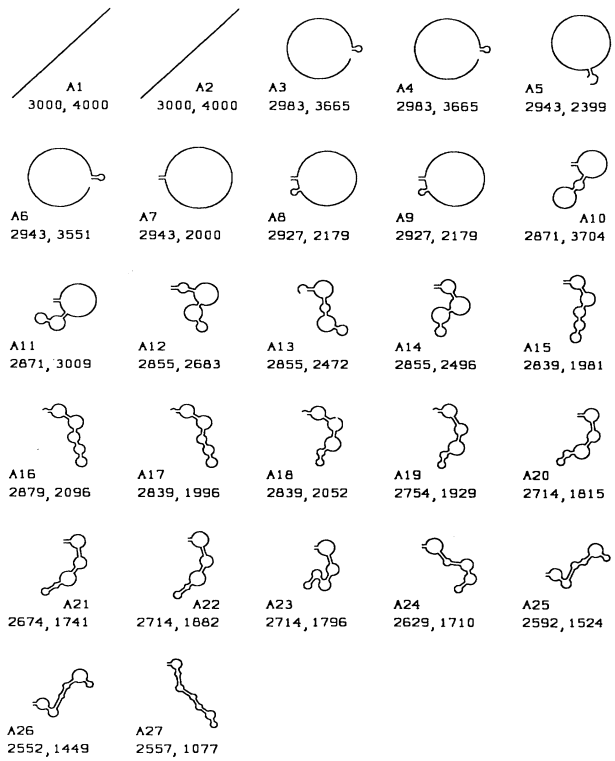

FIG. 7. Evolution of structures. The evolution of the structures encoded by the sequences along the genealogy escaping successfully the large-loop trap are shown. A7 is the large-loop master sequence M1 which is displaced by A27. The kinetic parameters are displayed as A_k, D_k.

volves the change of two digits. Thus a path of length 21 in the Hamming metric has been established stepwise purely by stochastic noise in the periphery of the quasispecies, while the latter was sharply localized in the large-loop trap.

We want to stress that in the present frame of a stochastic reaction network the replication and degradation constants derived from the secondary structures are the parameters of exponential densities of reaction probabilities,[13] as explained in Sec. II. In the deterministic case optimization takes place because of the selection term in Eq. (6), where the selective value of species k, $w_{kk} = A_k q^{\nu} - D_k$, is compared against the average productivity $\bar{E}(t)$ of the whole ensemble. The deterministic rate constants A_k and D_k that determine the selective value w_{kk} or the net productivity $E_k = A_k - D_k$ relate to

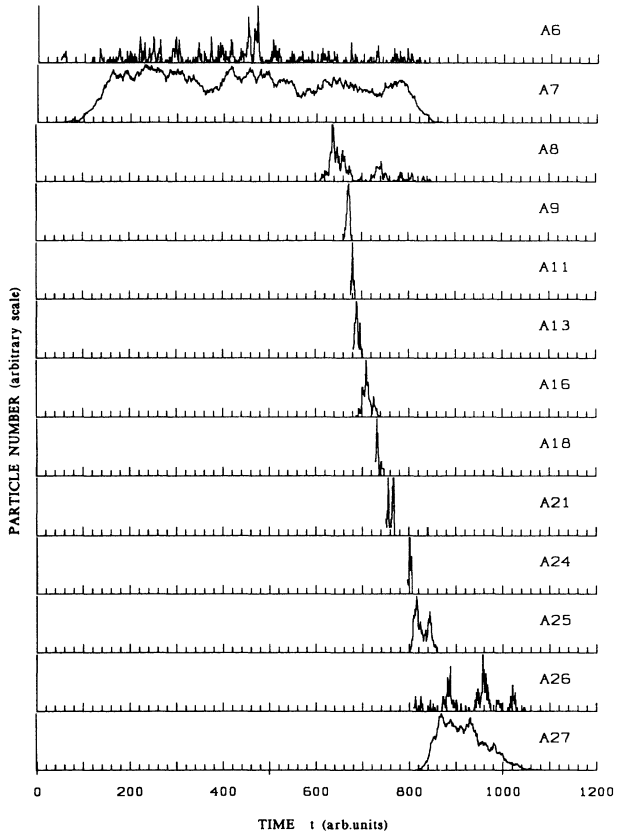

FIG. 8. Dynamics of the escape from a fitness trap. The kinetics of selected intermediates along the escape path out of the large-loop trap, A7, are shown. The resolution is of one time unit. Each ordinate is blown up to the maximum particle number reached by the corresponding species—see Table I. The structures of the binary strings are shown in Fig. 7. Note the extremely short residence time of the intermediates A9–A25. The precursor species A6 and the successor A8, both at distance 1 from the master A7, are continuously maintained in the population through back mutation from the master. The same holds for A26, the distance 1 precursor of the next master A27. They all disappear together with their "supporters."

the parameters A_k and D_k in the stochastic version by representing quantities averaged over many particles populating species k. Since the probability density that sequence I_k undergoes a replication event is exponential, the probability distribution that n replication events occurred up to time t is Poissonian

$$\exp(-A_k t)(A_k t)^n/n! \ .$$

Hence A_k is the average number of replications of species k per unit time. In the stochastic network selection operates on the level of expectation values. As long as a molecular species is populated by a few particles with finite lifetimes, these averages are not well-defined quantities within the system. The fluctuations in the occurrence of reaction events will dominate, hence establishing an "evaluation threshold" below which no selection takes place.

More precisely, consider a newly produced mutant present initially in a single copy. Let this mutant have a replication rate parameter A below the average $\bar{E}(t)$ of the current population ensemble. Without loss of generality we set its degradation rate parameter D to zero [otherwise it can be included in $\bar{E}(t)$]. Hence the mutant duplicates according to an exponential probability density with parameter A and disappears (dilution flow) according to an exponential probability density with parameter $\bar{E}(t)$. It is a reasonable approximation to assume $\bar{E}(t)=\bar{E}=\text{const}$ during the lifetime of the mutant species k, provided the time scale for changes in the mutant population is shorter than the time scale for changes in the bulk population. The stochastic process governing the dynamics of the mutant species is a linear birth and death process with birth rate A and death rate $\bar{E}$, $A<\bar{E}$ by assumption, and starting with a single copy of I_k at time $t=0$. Let $T_0(n)$ be a stochastic variable denoting the time elapsed until the species disappears from the system under the condition that it began with n copies. By a tedious calculation[31] we obtain in the case $A<\bar{E}$ for the expectation value $E(T_0(n))$ in units $1/\bar{E}$:

$$E(T_0(n))=\frac{\bar{E}}{A}n\sum_{l=1}^{n}\binom{n-1}{l-1}\left(\frac{\bar{E}}{A}-1\right)^{l-1}\frac{1}{l}\left[\sum_{i=1}^{l-1}\frac{1}{i}\left(1-\frac{\bar{E}}{A}\right)^{-i}-\ln\left(1-\frac{A}{\bar{E}}\right)\right] . \tag{16}$$

TABLE I. Recording of most frequent sequences during the escape from the "large-loop trap" in the computer experiment A [Fig. 5(a)]. All entries are given in time steps of the simulation and in particle numbers.

Species	Time of first appearance	Time span	Maximum number	Average number	Cumulative number
A6	46	403	20	2.8	1107
A7[a]	54	818	1076	646.5	528 800
A8	609	188	79	12.9	2428
A9	658	20	26	10.8	215
A10[b]	674	0			
A11	674	10	12	5.2	52
A12	678	22	19	6.7	147
A13	680	20	13	5.7	113
A14	681	7	6	3.3	23
A15	686	2	1	1.0	2
A16	686	50	47	12.5	623
A17	708	23	32	8.8	203
A18	727	16	8	2.8	44
A19	729	49	23	11.2	549
A20	749	2	1	1.0	2
A21	750	20	7	3.0	60
A22	758	34	14	5.6	191
A23	789	69	43	17.1	1182
A24	795	10	12	6.0	60
A25	799	59	62	24.5	1443
A26	808	121	15	3.2	392
A27[a]	812	274	784	307.6	84 282

[a]Master sequence of a quasispecies.
[b]This species appeared and disappeared between two successive recordings.

When starting with a single copy, $n=1$, Eq. (16) simplifies to

$$E(T_0(1))=-\frac{\bar{E}}{A}\ln\left[1-\frac{A}{\bar{E}}\right], \quad A<\bar{E} . \tag{17}$$

As is well known, the mutant faces extinction with probability one if and only if $A \leq \bar{E}$. Note that for $A=\bar{E}$ the expected time to extinction, $E(T_0(n))$, is infinite. Indeed, Eq. (16) diverges for all n in the limit $A \rightarrow \bar{E}$.

The point, however, is that the mean time to extinction becomes relatively large already in the range $0.9 \leq A/\bar{E} < 1$. This implies together with a large variance of $E(T_0(1))$ in the same range (which we found numerically), that there is a significant probability for the mutant species to reach a modest number of copies—even if $A<\bar{E}$—before eventually disappearing. It is precisely this stochastic effect of small population numbers[32] which enables many less advantageous species to establish a small population and to give rise to further mutants. The same holds for species that drop into the small population regime from above. In this case Eq. (16) applies for suitable n. We derived analogous expressions[31] for the expected time until the kth individual of a species disappears if the initial number of individuals was n. It becomes apparent that the first copies disappear quite quickly for $A<\bar{E}$. However, the last few (2 or 3, say) copies have long average residence times. Thus there is an occupation regime, which has to be crossed by every newly produced species and in which fitness is not a sharply defined attribute. A realistic quasispecies always has such a stochastic boundary. The population dynamics of the evolutionary jump (Fig. 8) show that series of short-living fluctuations in that boundary are able to bridge very large distances in sequence space, thereby enabling the population to escape deep local traps.

It is interesting to compare this scenario with the kinetic resolution (Table II) of the genealogical path from the start sequence to the final master sequence at $t=3710$ of experiment B, shown in Fig. 5(d), which has been obtained with analogous settings to the previous one but under the control of a different random number sequence. In this run the population is not caught by the large-loop trap and evolves rather smoothly. The analysis of the genealogies—partly summarized in Table II—shows the following characteristics.

● After the initial spreading of the homogenous population—after $t=200$, say—the intermediates gen-

TABLE II. Recording of most frequent sequences in the computer experiment B [Fig. 5(b)]. All entries are given in time steps of the simulation and in particle numbers.

Species	Time of first appearance	Time span	Maximum number	Average number	Cumulative number
B10	56	57	49	27.8	1586
B11	58	50	53	13.8	688
B12	66	6	6	2.8	17
B13	67	93	122	37.0	3440
B14	93	36	10	3.2	115
B15	99	40	27	10.9	434
B16	114	14	6	2.0	28
B17	117	76	39	12.2	925
B18	129	110	458	148.1	16285
B19	169	101	88	26.7	2692
B20	174	72	28	7.1	509
B21	208	205	208	70.8	14512
B22	237	390	124	20.9	8159
B23	248	692	629	222.9	154229
B24	333	376	66	19.7	7388
B25	352	252	195	63.2	15916
B26	485	57	15	2.6	148
B27	489	370	209	76.5	28289
B28	757	150	69	22.1	3312
B29	812	431	77	19.1	8226
B30	909	737	109	21.7	15959
B31	1058	981	263	50.4	49426
B32	1101	1019	292	53.0	53988
B33	1120	2458	965	300.3	738223
B32′	2122	1401	804	206.6	289484
B34	2477	665	147	46.6	31002
B35	2603	618	260	62.8	38829
B36	2951	665	249	50.0	33234
B37	3132	579	939	577.2	334173

erally reach higher occupation numbers than in the previous experiment. The residence time is correspondingly larger.

- The time lags between the production of successive intermediates are much larger.
- During the quasistationary phase from $t \approx 1100$ to ≈ 3000 extensive drift of neutral mutants is observed. The kinetic of B32′ is identical to the last part of B32. In fact, it is the same sequence that happened to die out just before $t=2000$ and shortly after reappears again as the "successor of its own successor (B33)." Species B32=B32′ and B33 are selectively neutral.

The different dynamics of the two described examples are due to different topologies of the visited regions on the fitness surface. In the first case the distribution is trapped at an isolated fitness peak. There are no high valued ridges that lead away from there. The metastable quasispecies remains localized for a very long time, since it has to wait for a suitable series of fluctuations in its low occupied periphery that bridge the distance to regions with higher fitness. In the second case, the evolutionary trajectory happens to bypass the isolated peak and proceeds quickly along a connected net of ridges on the fitness surface. Most of the intermediates do not belong to the stochastic periphery at the time when they produce their successors. Nevertheless the chosen route still depends at certain times on species whose occupation density is at or below the evaluation threshold.

Figure 5(c) is the record of a third experiment, C, with the same parameter settings, but differing again in the random number series. The master sequence at the end of simulation is shown in Fig. 5(d) together with the most populated sequences at the termination of runs A and B. Simulation C is not caught by the deep trap mentioned above, though it is slowed down by other structures partially taking advantage of the large-loop rule (they happen to have free ends). The final master sequence of run C carries a peculiar T-like element (bifurcation point). This structural feature splits 7 base pairs into 2+2+3 thus reducing the replication slowdown, without the need of local bulges or internal loops that would score for hydrolysis. On the other hand, it forces a second hairpin turn, hence making four additional positions accessible to hydrolysis. This type of structure is already present in the master at time $t=447$. The sequences sharing this feature at that time were then steadily optimized in the two hairpin arms, therefore making the elimination of this structure type more and more costly, hence "freezing in" this feature and thereby prejudging the following route of adaptation. Indeed, the structure that emerged finally is optimal with respect to the T-like element. This demonstrates how the outcome of an evolution experiment depends sensitively on its previous history.

The final master sequences of runs A, B, and C [Fig. 5(d)] have almost the same Hamming distances from the initial all-0-sequence: $d(0,A)=26$, $d(0,B)=25$, and $d(0,C)=24$. They reside in neighboring error classes with respect to the starting sequence. Apart from this "similarity" they are quite different: the three sequences A, B, and C span an almost equilateral triangle of Hamming distance $d \approx 31$ in the sequence space: $d(A,B)=33$, $d(A,C)=28$, and $d(B,C)=33$. Their quasispecies distributions are pairwise disjoint, hinting to a "glassy" behavior of the replication-mutation strategy at a high replication accuracy. In the context of spin glasses Palmer[33] stressed the fact that some systems may be ergodic in a technical sense in the limit of infinite times, but are far from showing this behavior on reasonable physical time scales.

Lack of localization in sequence space was observed with less accurately replicating populations. In one particular computer experiment we started from a homogeneous population of the best sequence obtained from simulated annealing ($E=2045$, cf. Sec. VI). In the time interval $0<t<530$ a replication accuracy of $q=0.995$ was applied and a mutant distribution of quasispecies type formed. Then, we reduced the replication accuracy below the threshold,[11,13] $q=0.993$, and continued the simulation for another 500 timesteps. The optimum sequence was quickly lost from the population and after that, a succession of "master" sequences was observed. These most frequent sequences are hardly more than random fluctuations in particle numbers: in a short interval of 50 time steps ($950<t<1000$) as many as ten different master sequences were recorded.

V. PICTURES FROM AN EVOLVING POPULATION

Here we study the microscopic structure of the sequence distribution and its changes during the escape from the large-loop trap using cluster analysis. We look at the support of a given distribution, i.e., a swarm of points that occupies vertices of the ν-dimensional hypercube. It makes no difference if a given point is heavily populated or present only in a single copy, since we are primarily interested in the similarity structure of this "mutant cloud."

Let us denote by $\sigma_{kit} \in \{0,1\}$ the ith digit of the kth sequence in cluster t. If we do not distinguish between clusters we leave off the index t. Let

$$\langle \boldsymbol{\sigma} \rangle = (\langle \sigma_1 \rangle, \langle \sigma_2 \rangle, \ldots, \langle \sigma_\nu \rangle)$$

with $\langle \sigma_i \rangle = (1/N)\sum_{k=1}^{N} \sigma_{ki}$ be the real-valued center of mass of the system consisting of N vertices on the hypercube. Let

$$\underline{S} = (s_{ij}), \quad s_{ij} = \sum_{k=1}^{N} (\sigma_{ki} - \langle \sigma_i \rangle)(\sigma_{kj} - \langle \sigma_j \rangle) , \tag{18}$$

be the covariance matrix of the whole system,

$$\begin{aligned} &\underline{V} = (v_{ij}) , \\ &v_{ij} = \sum_{t=1}^{l} \sum_{k=1}^{N_t} (\sigma_{kit} - \langle \sigma_{it} \rangle)(\sigma_{kjt} - \langle \sigma_{jt} \rangle) , \end{aligned} \tag{19}$$

represents the covariance contribution within the l clusters of a given partition and

$$\begin{aligned} &\underline{B} = (b_{ij}) , \\ &b_{ij} = \sum_{t=1}^{l} n_t (\langle \sigma_{it} \rangle - \langle \sigma_i \rangle)(\langle \sigma_{jt} \rangle - \langle \sigma_j \rangle) , \end{aligned} \tag{20}$$

is the covariance due to the separation between clusters. n_t denotes the number of points in cluster t. As is well known in multivariate statistics,

$$\underline{S}=\underline{V}+\underline{B} \ . \tag{21}$$

To achieve an informative partition of the ensemble into clusters of similar sequences one might minimize $\mathrm{Tr}\underline{V}$, i.e., the total variance inside the clusters, thereby maximizing $\mathrm{Tr}\underline{B}$, i.e. the separation of the groups.[34] $\mathrm{Tr}\underline{V}=\sum_{t=1}^{l}V_t$ with V_t denoting the variance in a given cluster t,

$$\begin{aligned} V_t &= \sum_{k=1}^{n_t}\sum_{i=1}^{\nu}(\sigma_{kit}-\langle\sigma_{it}\rangle)^2 \\ &= \sum_{k=1}^{n_t}\sum_{i=1}^{\nu}\sigma_{kit}^2-n_t\langle\boldsymbol{\sigma}\rangle_t^2 \ . \end{aligned} \tag{22}$$

With a little algebra we find that V_t depends on the Hamming distances $d(i,j)$ between the sequences of cluster t as

$$V_t=\frac{1}{n_t}\sum_{i>j}d(i,j) \ . \tag{23}$$

We partitioned the sequence ensemble according to the variance criterion[35] minimizing $\mathrm{Tr}\underline{V}$. The results are in perfect agreement with a completely different but much more visual way of looking at the detailed structure of a sequence distribution consisting in the construction of a minimum spanning tree (MST). For the ease of presentation we report only the latter.

Consider the complete graph whose vertices are the N occupied points in sequence space. The MST is the tree which connects all vertices and whose total length over all $N-1$ edges is a minimum. Of course this is a graph in a subspace of the ν cube and has to be projected onto some 2D plane, that minimizes the loss of details.

Let s_i^2 denote the variance at position i in the sequences and let N be the size of the sample,

$$s_i^2=\frac{1}{N-1}\left[\sum_{k=1}^{N}\sigma_{ki}^2-\frac{1}{N}\langle\sigma_i\rangle^2\right] \ . \tag{24}$$

If a position is conserved throughout the sample, we set its variance to unity. Let $\underline{C}=\{c_{ij}\}$ be the (symmetric) covariance matrix of the normalized variables $\sigma_i^*=(\sigma_i-\langle\sigma_i\rangle)/s_i$,

$$c_{ij}=\frac{1}{N-1}\sum_{k=1}^{N}\sigma_i^*\sigma_j^* \ , \tag{25}$$

and let $\underline{T}$ be the matrix that diagonalizes $\underline{C}$, $\underline{T}^{-1}\underline{C}\,\underline{T}=\underline{\Lambda}$. The trace of $\underline{\Lambda}$ is obviously the total variance. Sequence $\boldsymbol{\sigma}^*=(\sigma_1^*,\sigma_2^*,\ldots,\sigma_\nu^*)^T$, expressed in normalized components is written in the new coordinate system, $\boldsymbol{\sigma}'=\underline{T}^{-1}\boldsymbol{\sigma}^*$, and we retain $\sigma_i'=\sum_{k=1}^{\gamma}\sigma_k^*v_k^i$, $i=1,2$, where v_k^1,v_k^2 are the eigenvectors associated with the two largest eigenvalues of $\underline{C}$. Thus we project onto the hyperplane that incorporates the two largest variance shares of the system.

Several snapshots are taken at characteristic times T_k ($k=1,\ldots,5$) before, during, and after the transition. Species are labeled according to their frequencies in the population: the most frequent species is denoted by "1," the next frequent by "2," and so on.

(1) $T_1=300$. $N=143$ different species are present in the network. Figure 9(a) shows the MST of the distribution support that includes only the 100 most frequent species. It is characterized by a few nodes that display a high degree of branching. These nodes are highly populated species and thus represent centers of high replication activity. This in turn implies that they are at the origin of most of the mutants. The nucleation point at the center of the tree is sequence 1, the current master sequence. The MST visualizes nicely that the overwhelming part of the exploration activity of the network concentrates at the highest populated species, i.e., at the currently fittest ones.

(2) $T_2=700$. $N=220$ different species are present at this time. The distribution has broadened into two clusters a ($|a|=142$) and b ($|b|=78$) centered at the sequences 1 and 4, respectively. We report the distance between the real-valued mass centers of clusters a and b in terms of the "city block (L_1) metric",

$$d_1(a,b)=\sum_i|\langle\sigma_i\rangle_a-\langle\sigma_i\rangle_b| \ .$$

This becomes the Hamming distance in the case of single sequences. The center-to-center distance d_1 between clusters a and b is 3.38. The groups are easily recognizable on the right margin of the MST [Fig. 9(b)]. On the left at a distance 5 from the bulk we find the sequences 12 and 45 with Hamming distance $d(12,45)=4$. Sequence 12 is a member of the genealogy towards the future master sequence. It is the intermediate A16 of Sec. IV.

(3) $T_3=820$. $N=279$ different species are present. The system is just before the jump. Figure 9(c) is a clean-cut example for two totally separated clusters. The groups a ($|a|=98$) and b ($|b|=181$) with center-of-mass distance $d_1=16.64$ are nucleated by species 1 and 3. The right island (b) displays three distinct subgroups b_1, b_2, and b_3 (centers: 3, 6, and 12).

(4) $T_4=840$. $N=307$ different species are present. The sequence ensemble is in the jump phase just after the change of the master sequences. The MST [Fig. 9(d)] signals four clusters a, b, c, and d ($|a|=66$, $|b|=126$, $|c|=76$, $|d|=39$) with centers at 1,5,3,2 and distances $d_1(a,b)=10.53$, $d_1(a,d)=20.83$, $d_1(b,c)=9.87$, $d_1(a\cup b,c)=10.93$, and $d_1(a\cup b\cup c,d)=17.61$. d is the local distribution containing the large-loop sequence (2), while a is centered at the new master. The selective values $(A_iQ_{ii}-D_i)$ are $w_a=1307$, $w_d=743$, $w_c=1073$, $w_b=1044$. Inspection of the data allows us to relate the clusters of the MST at $T_3=820$ to those at $T_4=840$:

$$a^{T_3}\rightarrow d^{T_4}, \quad b_1^{T_3}\rightarrow c^{T_4}, \quad b_2^{T_3}\rightarrow b^{T_4} \ ,$$

$$b_3^{T_3}\rightarrow(a^{T_4}\text{ and }b^{T_4}) \ .$$

The new (current) master sequence $1\in a^{T_4}$ was assigned as sequence 45 to cluster b_3 at time $T_3=820$. The ensemble was evolving towards b^{T_4} and c^{T_4}, when suddenly the

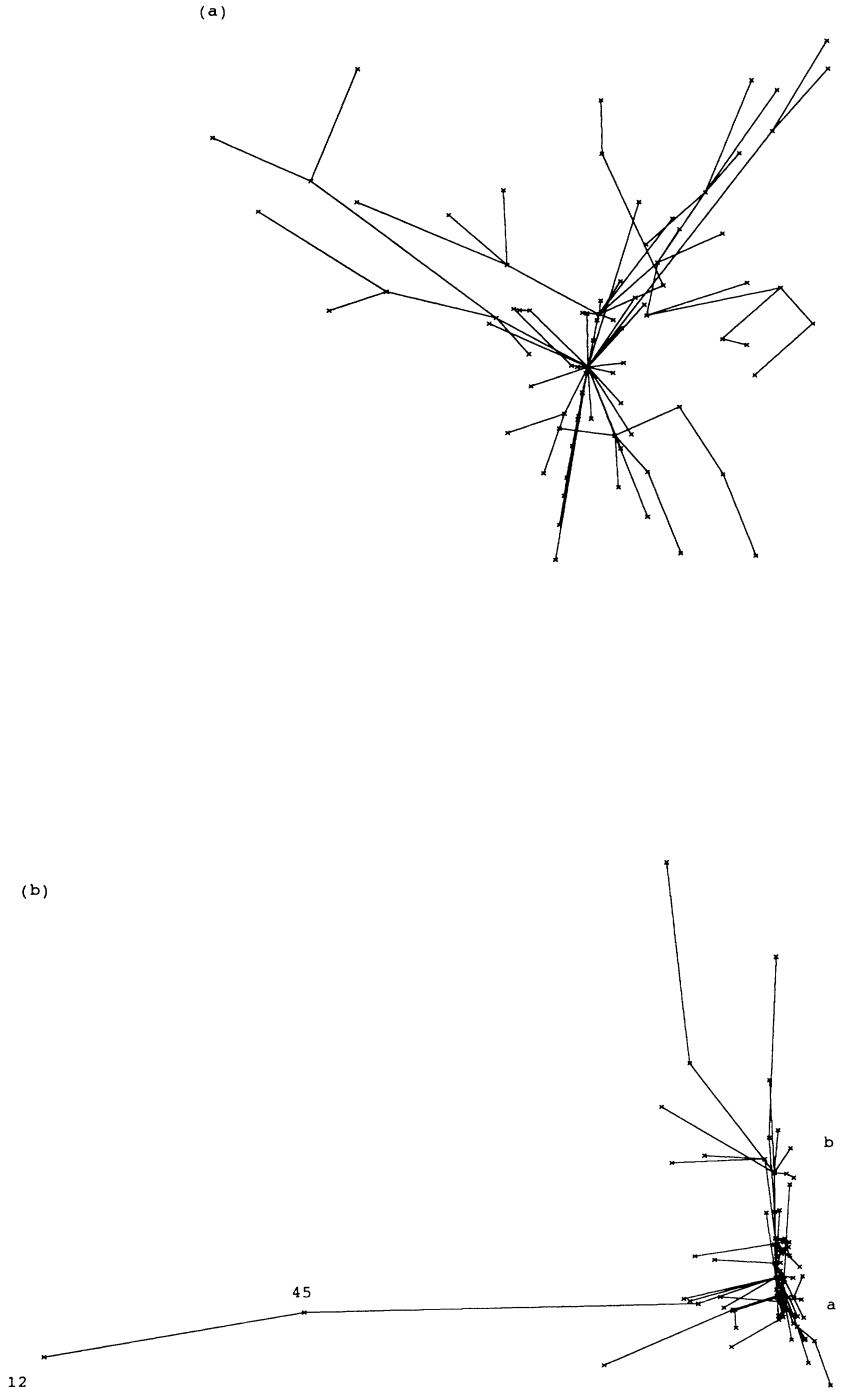

FIG. 9. Cluster analysis of a species distribution. Similarity structure of the distribution of species in run A according to minimum spanning tree clustering of the 100 most frequent species present. The truncation to 100 merely enhances the details of the distribution core without affecting its structure. The plots refer to the indicated times in Fig. 5(a). $T_1 = 300$ (a), $T_2 = 700$ (b), $T_3 = 820$ (c), $T_4 = 840$ (d), $T_5 = 960$ (e).

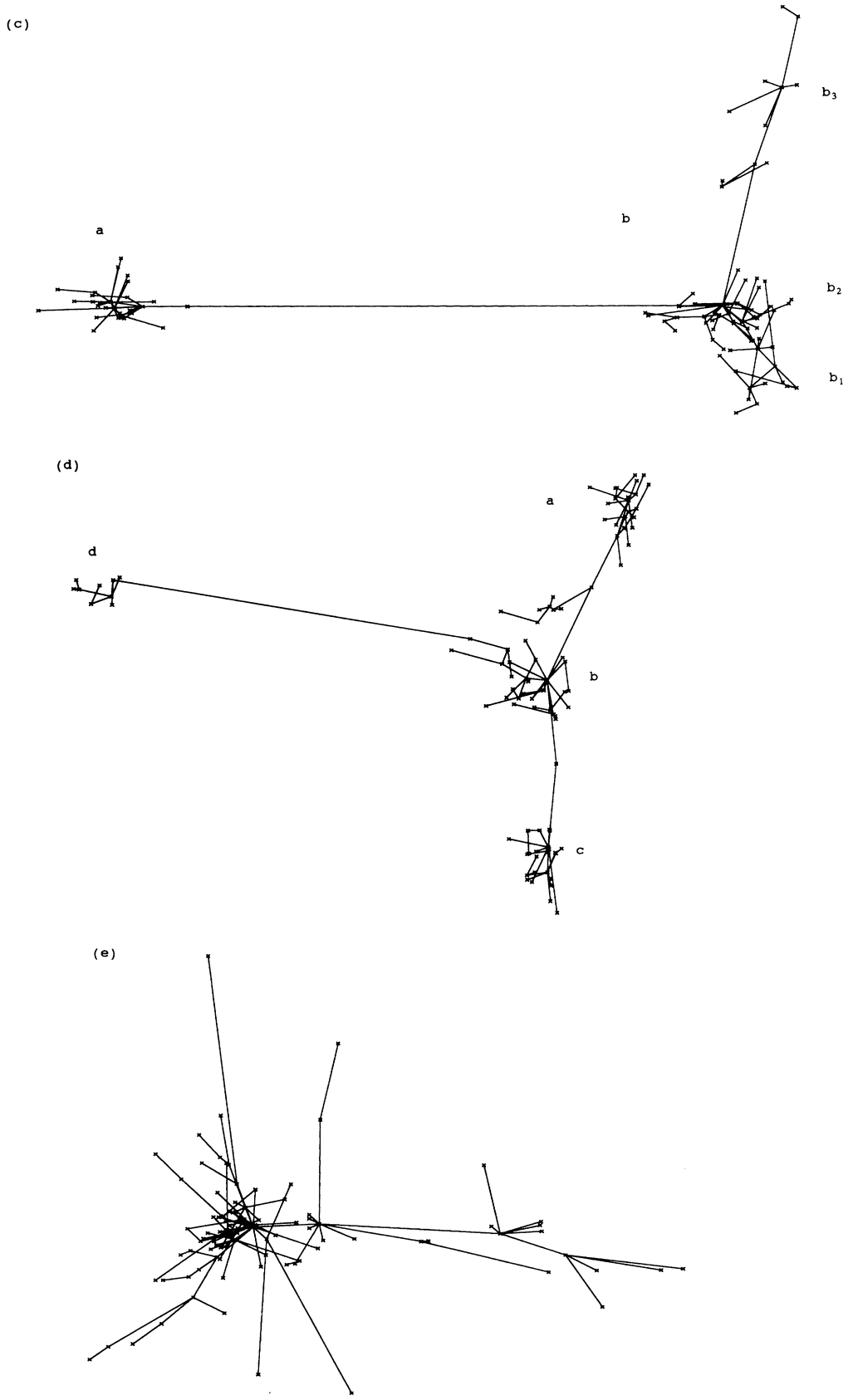

FIG. 9. (*Continued*).

current master sequence appeared leading to a change of direction.

(5) $T_5 = 960$. $N = 276$ different species are present. The ensemble has condensed around the new master sequence [Fig.9(e)]. The support is again localized in a narrow region of the sequence space. The new region clearly has a different topology from the previous one. Many selectively neutral or quasineutral species share the population resources and lead to a broadening of the distribution. The analysis identifies many modes with dense centers, all connected with distance 1.

The study of this time series demonstrates that a population of mutants and its changes can be analyzed by focusing on the structure of the distribution support. The fitness surface is mapped by means of a population into a probability distribution that controls the allocation of replication events. The scheme by which the currently best species get most of the resources implies that the majority of the mutants are direct relatives of the fittest sequences. This population version of a best first search heuristic is the simplest conceivable evolutionary strategy. The multivariate cluster analysis thereby provides an efficient yet simple tool for tracking the motion of high-dimensional sequence ensembles.

VI. DISTRIBUTION OF MAXIMA

For a better understanding of optimization dynamics we need information on the distribution of positions and heights of extreme values in the configuration-space landscape. Conceptual advances in statistical physics of systems with quenched-in randomness and frustration, as exemplified by the paradigm of spin glasses[36] and by its link to optimization theory,[37,38] stress the role played by the structure of the system's objective function in behaviors like ergodicity breaking[33] or freezing. The ultrametric organization of Gibbs states in the long-range spin glass[17] carries over to those combinatorial optimization problems that can be mapped onto it. Graph bipartitioning[38] serves an an example. Some evidence has been accumulated in favor of an ultrametric structure of equally weighted locally stable configurations in random symmetric "traveling salesman" problems.[18]

In this section we proceed along these lines and investigate the distribution of the maxima of the net productivity, $E_k = A_k - D_k$, in our folding based model landscape (8) and (9). For this purpose we used the simulated annealing technique[3] to generate a set of structures with locally optimal E values. As is well known, this strategy uses a Markov chain to sample configurations according to a Boltzmann distribution. In the present case a given sequence is modified with probability p_1 by a single or with probability p_2 by a double mutation. Either both events occurred with the same frequency $p_1 = p_2 = 0.5$, or each trial configuration was chosen randomly out of all 2485 one- and two-error mutants relative to the actual one. Runs allocating the same resources but operating with only single-digit exchanges never reached very high regions—characterized by $E_k \geq 2000$. Let the net productivity of the trial sequence be E_t and let E_a be that of the currently valid one. The acceptance probability at the ith temperature step then is $\min\{1, \exp(E_t - E_a)/T_i\}$. T_i was decreased by a factor 0.85 whenever either 100–120 configurations were accepted or 20 000–40 000 trials had been performed. The dynamic programming approach to folding is rather time consuming. The implementation of a search tree which continuously memorized the last few thousand generated sequences along with their selective values (structures) improved in our case the time performance by a factor up to 10 at low temperatures. However, the limited resources forced us to compromise between fast computation and an optimal annealing schedule.

In 32 runs we accumulated 3921 distinct sequences with $E \geq 1950$. Simulated annealing thus leads to somewhat higher values of the net productivity E than the evolution reactor, whose best achievement under the operating conditions described in (Ref. 13) was a master sequence with $E = 1919$—computer experiment C shown in Fig. 5(d). When compared with the starting value of the all-0-sequence, $E_0 = -1000$, however, the final achievement of both techniques differs by less than 10% only.

The series of all sequences accepted during an annealing run produces a "trajectory" that can be conveniently visualized by projection on the hyperplane spanned by the first two eigenvectors of the corresponding covariance matrix, as explained in Sec. V for the minimum spanning tree. A typical result is shown in Fig. 10 where we projected the history of an annealing run comprising a series of 2237 sequences. The search path terminates at low temperature on a high plateau consisting of 121 configurations with $E = 2019$ and one with $E = 2045$.

FIG. 10. Simulated annealing trajectory. The series of sequences accepted during the course of a simulated annealing optimization is visualized by projection on the plane catching the largest variance (see Sec. V, minimum spanning tree). The bulk on the right corresponds to the high-temperature phase. The frozen optima are visible as an island on the left that has been reached at low temperatures along a ridge connecting it to the bulk.

This set of closely related sequences—seen in the upper left corner of the figure—is well separated from the regions visited at high temperatures and is connected to them by a partially visible long ridge. This ridge, protruding out of the high-temperature bulk and leading to the frozen out optima, was visible very frequently in our runs. The starting basin consists of 75 configurations with $E=2019$, whose location in Fig. 10 is easily identified by the cubelike loop. Its "city block distance" from the all-0-corner amounts to 27.03. Forty-six digit positions are strictly conserved in this basin. The above-mentioned final set has a distance from the zero corner of 38.13, while 40 positions are left invariant. The walk bridging the two areas with high E includes 300 distinct sequences, all having the same $E=1993$. Along that ridge only two positions are conserved. The city block distance between both areas amounts to 25.85, which is as much as one third of the diameter of the sequence space.

The example considered here clearly demonstrates the existence of sets of closely related high-valued configurations—corresponding to "valleys" in the language of spin-glass physics. These sets are connected by long channels of slightly less advantageous neutral mutants.

Five sequences, S_1,S_2,S_3,S_4,S_5, with $E=2045$—the highest value obtained—were found. S_1 and S_2 as well as S_3 and S_4 are close relatives at a Hamming distance 2. They have been obtained during two runs. The triangle S_2,S_3,S_5 has edges of length $d(S_2,S_3)=32$, $d(S_3,S_5)=43$, and $d(S_5,S_2)=31$.

As many as 879 sequences had a net productivity $E\geq 2011$. Figure 11 shows the minimum spanning tree of that sample. Again the clustered structure of our configuration space becomes visible. The MST indicates seven clearly separated groups, each of which corresponds to a set of maxima produced in a run. Note that we did not consider the symmetric set of maxima obtained by inversion and complementation. The MST of the two symmetric sets lumped together shows that the individual clusters are not affected and do not merge, being separated by edges of length ranging between 20 and 24.

In order to check whether or not there is some order underlying the distribution of highly efficient species in sequence space, we computed pair and triangle statistics. It seems obvious to compare the distribution of optimal E values with the ultrametric distribution of the (weighted) minima of spin Hamiltonians. In a Euclidean space of dimension ν, at most $\nu+1$ points can form an ultrametric set, which is defined by $d(i,j)\leq\max\{d(i,k),d(j,k)\}$ for all triples of points (i,j,k). Indeed the metric between two microconfigurations in the spin glass is expressed in terms of overlap,

$$q(i,j)=(1/\nu)\sum_{k=1}^{\nu}\sigma_{ki}\sigma_{kj}\ ,$$

where $\sigma_{ki}\in\{-1,1\}$ is the ith spin in configuration k. In the "$\{-1,1\}^\nu$ cube" the Hamming distance $d(i,j)$ is given by $\frac{1}{2}(\nu-\sum_{k=1}^{\nu}\sigma_{ki}\sigma_{kj})$. In sequence space, the $\{0,1\}^\nu$ cube, we use the Hamming distance $d(i,j)=\sum_{k=1}^{\nu}|\sigma_{ki}-\sigma_{kj}|$, $\sigma_{ki}\in\{0,1\}$. Its relation to the overlap $q(i,j)=(1/\nu)[\nu-2d(i,j)]$ holds obviously on both cubes.

Figure 12 shows the frequency of distances $d(i,j)$ in several sets of configurations of high E values. The excess density at low distances is clearly due to the "intravalley" contributions, whereas the part centered at distance 35 stems from the edges between the clusters. If we draw an edge in the hypercube at random the probability $P(k)$ that it will have length k is binomial, $P(k)=\binom{\nu}{k}/(2^\nu-1)$. Thus the most probable distance is

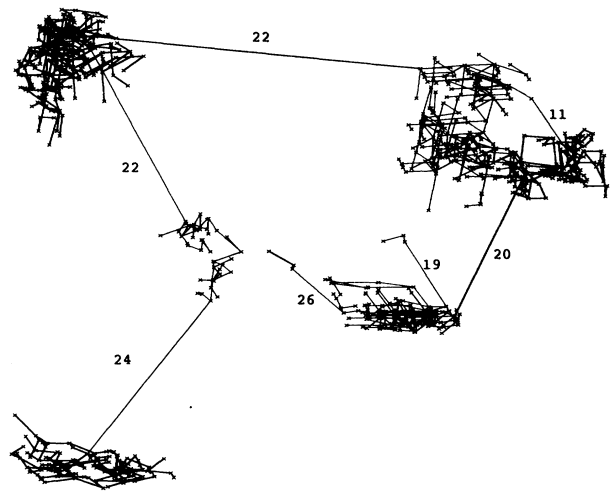

FIG. 11. Clustering of extreme fitness values. The projection shows the minimum spanning tree of the frozen out optima of seven annealing runs operating under identical conditions (except random number seeds). The optimal regions reached by each run consist of an ensemble of selectively neutral mutants and qualify as clusters in the minimum spanning tree. The shortest Hamming distances between clusters are indicated.

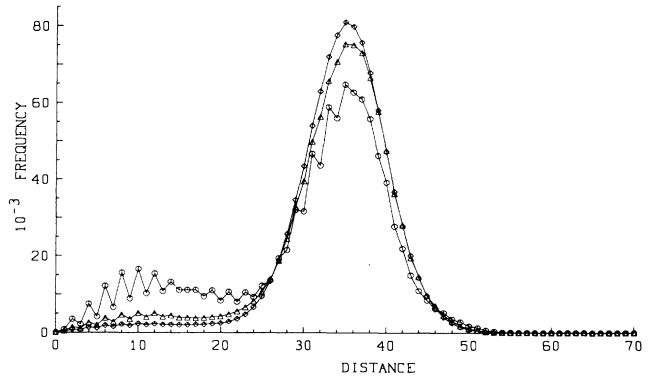

FIG. 12. Pair statistics of optima. The distribution of distances in the set of optima obtained during annealing runs is shown. Circles refers to the set of 879 sequences with $E\geq 2011$, triangles to the set of 2071 sequences with $E\geq 1955$, and squares to 3921 sequences with $E\geq 1950$.

$\nu/2$. The distribution of distances between groups indicates that the optima in our model landscape are scattered at random.

It could well be the case that configurations with optimal or near optimal net productivities are concentrated with a high bias on triangles that are equilateral or isosceles with small basis. The reference formulas for randomly distributed triangles are derived in the Appendix. From the Appendix—Eqs. (A3) and (A4)—we obtain for $\nu=70$ the following probability that an arbitrary triangle is either equilateral or isosceles with small basis: $P_E+P_I=0.0052+0.0958=0.101$. As much as every tenth random triangle is expected to be ultrametric. We computed all 112 805 879 triangles in the sample of 879 maxima and found 11 802 278 ultrametric ones, corresponding to a frequency of 0.105. From the equation for $P(m,l,k)$ we calculate the probability of a random triangle whose longest sides are k and l, $\sum_m P(m \le k \le l)$, which allows us to concentrate the statistics of triangles on a single plot [Fig. 13(a)]. Figure 13(b) shows the analogous statistics for our sample. We recognize a ridge of small triangles along the diagonal which is—as in the previous analysis—due to contributions inside the groups, whereas the distribution of intergroup triangles closely resembles the random case. This indicates that the best configurations in our value landscape are evenly distributed throughout the sequence hypercube in accordance with the results derived in Sec. III. The assignment of only 4328 different selective values to $2^{70}\approx 10^{21}$ sequences according to Eq. (11) produces many regions at large distances from each other that assume similar fitness values. Nevertheless the density of maxima is still too low for providing an appreciable probability of hitting them by chance. These two facts make optimization on such a surface a nontrivial problem.

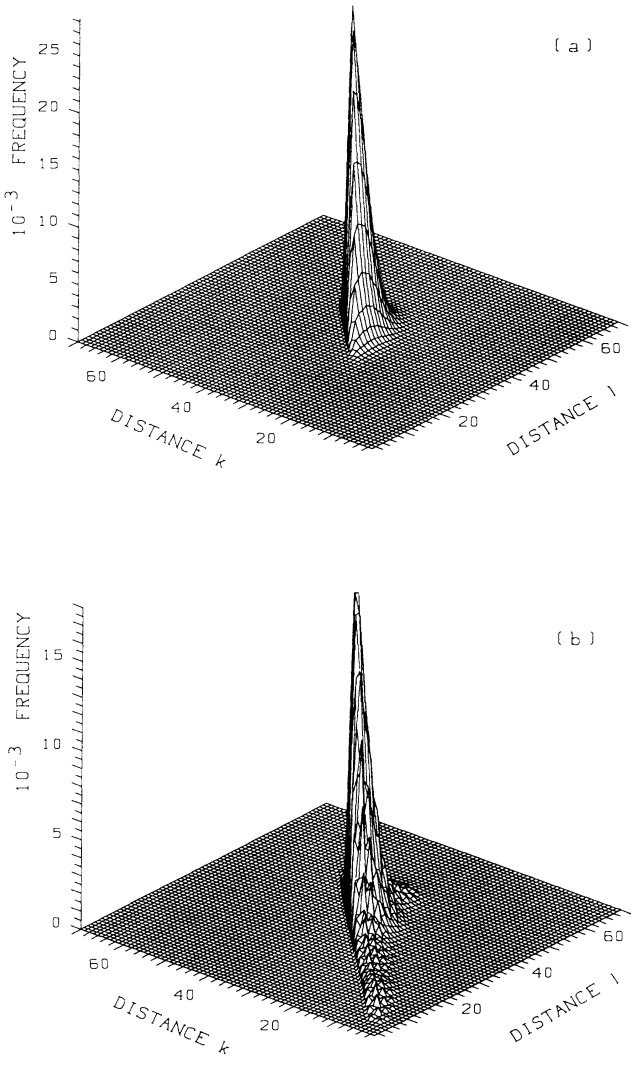

FIG. 13. Triangle statistics of optima. (a) Probability distribution for drawing a random triangle whose longest sides are k and l from a hypercube with $\nu=70$ using Eq. (A2) (see Appendix). (b) Analogous frequency distribution among all triangles in the set of 879 optima obtained with simulated annealing.

VII. THE TRAVELING SALESMAN

During the last years some areas of physics began to pay much attention to difficult (i.e., NP-complete[39]) optimization problems (NP represents nondeterministic polynomial). This was essentially due to ideas originating in spin-glass research[3,40] and neural networks.[4] It seems obvious to apply the evolutionary strategy to problems of the "traveling salesman" type, since it makes no difference if we optimize fitness values of model RNA molecules or cost functions of other given combinatorial objects. The approach goes back at least to Rechenberg[5] and has been advocated independently in computer science areas, where it is known as "genetic algorithm."[6]

We give here simply an example of the evolution reactor acting on the traveling salesman problem (TSP). In that problem one has to search for the shortest Hamiltonian circuit in a set of "cities," i.e., in some distribution of points i, $i=1,2,\ldots,\nu$. Thus an instance of the problem is represented by a (usually symmetric) distance matrix $\underline{L}=(l_{ij})$. A circuit is given by a permutation $\Pi_i=\pi(1)\cdots\pi(\nu)$ of the points and its length is $L(\Pi_i)=\sum_{k=1}^{\nu} l_{\pi(k)\pi(k+1)}$, where $\pi(\nu+1)=\pi(1)$.

Of course we focus on the $(\nu-1)!/2$ permutations that represent different circuits, since the starting point and the direction of walk are irrelevant. A permutation Π_i now takes the role of a sequence I_i. Its "replication rate" A_i is simply defined to be

$$A_i=\lfloor c_1\{c_2-L(\Pi_i)\}+0.5\rfloor ,$$

with $c_1>0$, $c_2>L_{\max}$ as scaling constants, $D_i\equiv 0\ \forall i$.

Maximizing A_i clearly implies minimizing $L(\Pi_i)$. The replication process is analogous to the sequence case. At each $\pi(k)$, $k=1,\ldots,\nu$ of Π_i a mutation event happens with probability $1-q$. As a mutation event we chose the inversion of a random subtour beginning at $\pi(k)$, i.e., a transformation exchanging two connections and approaching a two-optimal solution in the terminology of

Lin.[41] We did experiments with $\nu=36$ cities on a regular grid in the unit square, $q=0.999$, $\Theta=2000$ and starting with a homogenous population of a random configuration. A grid with an even number of cities on a side always allows for optimal circuits whose sections are all parallel to the sides of the lattice, thus having length $L_{\min}=\nu/(\sqrt{\nu}-1)$. The ensemble, however, starting with lengths about 21.2, is caught in a local trap with $L=8.3-8.5$ to be compared with the optimal length $L_{\min}=7.2$. It seems not to be easy to remove the "frozen in" defects of the better tours via inversion moves. Similar results are obtained with transpositions and/or random grids. Improving the power of the evolutionary strategy implies optimizing the population resources Θ, the replication accuracy q, and the underlying cost topology (via different A_i functions and mutation operators). This requires a local stochastic theory of the mutation-selection model—a theory that continues to remain a challenge in this field.

VIII. CONCLUSIONS

A fitness landscape (objective function) based on a model for folding of RNA molecules and evaluating the resulting secondary structures has been analyzed by combinatorial means and by applying the simulated annealing technique. The main result is that clusters of high-valued configurations are evenly scattered throughout the sequence space. These clusters are connected by long ridges of slightly less advantageous mutant configurations.

Adaptive walks of populations of $N=2000$ replicating and mutating individuals were studied on this "rugged" fitness landscape. For this goal we built a computer model which simulates replication kinetics in a kind of flow reactor called the "evolution reactor." Optimization dynamics in the evolution reactor was analyzed by means of genealogies. The dynamics along selected genealogies suggest two basic scenarios for adaptation.

- Populations may be trapped by isolated fitness peaks. Escaping from there implies waiting for a suitable series of fluctuations that percolates through to better mutants. Our simulations show that fluctuations in the periphery of a sharply localized distribution are able to bridge long distances. Isolated traps lead to long-living metastable states. Fluctuations result in a sudden breakdown of stability, which is monitored as an evolutionary jump.
- In a second scenario it is the existence of connections between high-valued regions which enables the population to proceed smoothly along these paths. Intermediates on the trajectory to the final master sequence generally consist of well populated species.

The distribution of finitely many molecular strings always has a low populated boundary in which stochastic effects are dominant, hence preventing a reliable evaluation of averaged quantities like fitness. Below this evaluation threshold the fate of mutants appears to be weakly dependent on their rate parameters giving rise to fluctuations that reach out very far in sequence space even when no nearby local optima are available.

Simulation experiments show that it may become increasingly costly to eliminate shape features that happened to be advantageous at an early time instant. This results in a "decision" for certain structural elements, thus constraining the subsequent optimization process.

Population dynamics maps the fitness surface into a probability distribution that controls the allocation of replication events. The Darwinian scheme—by which the "best" get most of the resources—implies that most of the mutants are direct relatives of the fittest types, thus embodying the heuristic of a "best first" search with some degree of parallelism. Trying to make improvements by modifying the currently best solutions is the simplest way of dealing with incomplete information about the underlying quality function. Clearly, it cannot be guaranteed that an optimal solution will be found on an accessible physical time scale.

The dynamic allocation of resources is well reflected in the similarity structure of the distribution support. The minimum spanning tree is a useful tool for making simple visual displays of this structure. Conventional cluster techniques can be used to track the motion and to resolve the microstructure of a population in the space of sequences of constant chain length.

The traveling salesman problem was studied in the evolution reactor in order to obtain information on the performance of this genetic algorithm on a different problem. Without any special modification the evolution reactor worked well on this standard study case.

Interestingly, simulated annealing reached higher peaks on the fitness landscape than the evolution reactor. The higher optimization efficiency of simulated annealing came out to our surprise. Accordingly, we searched for the causes of the different optimization behavior. "Efficiency" has two different aspects.

- The time required to reach a "quasistationary," hopefully "near optimal" state corresponding to a maximum of the fitness landscape.
- The fitness value of the "near optimal" state which can be reached within the disposable time span, or in the limit of long—but finite—times.

As far as the first point is concerned, the evolution reactor is disfavored by conventional computing techniques. Genetic algorithms are based on massively parallel processing of potential solutions in the population, but the individual replication steps are excuted sequentially—one after the other—on the computer. Simulated annealing, however, is not only faster on the computer, but it proceeds also towards higher peaks. Why does nature not use a procedure analoguous to simulated annealing then? Of course we can only give a highly speculative answer: it seems that variable environments make the "memory" of the genetic algorithm—achieved by its inheritance mechanism—superior to the rather "memoryless" simulated annealing procedure. On the other hand, it might well be the case that nature's choice of a particular optimization technique merely reflects the constraints ultimately imposed by the physics and chemistry of life.

ACKNOWLEDGMENTS

This work was supported financially by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (Projects No. 5286 and No. 6864) and the Hochschuljubiläumsstiftung Wien. IBM Austria provided personal computer equipment for this study. Generous supply with computer time on the IBM 3081 and 3090/400E mainframe computers by the EDV Zentrum, Universität Wien is gratefully acknowledged.

APPENDIX: TRIANGLE STATISTICS ON SEQUENCE SPACE

We compute the probability of getting by chance an ultrametric triangle (equilateral or isosceles with small basis) out of all possible ones. Let A,B,C be configurations on the $\{0,1\}^\nu$ hypercube. Let their mutual Hamming distances be $d(A,B)=l$, $d(A,C)=m$, and $d(B,C)=k$ with $m \le l$. If l and m are given, then $k=l+m-2g$, with $g=0,1,\ldots,m$, where g denotes the number of common positions that have to be changed in obtaining B and C from A. Then for given g

$$P(k=l+m-2g)=\frac{\binom{l}{g}\binom{\nu-l}{m-g}}{\binom{\nu}{m}} . \tag{A1}$$

Without loss of generality we can fix A. Then the probability for a triangle with specifically assigned sides m,l,k is

$$P^*(m,l,k)=\frac{1}{N}\frac{1}{2^{\delta_{l,m}}}\binom{\nu}{m}\binom{\nu}{l}P(k=l+m-2g)$$

$$=\frac{1}{N}\frac{1}{2^{\delta_{l,m}}}\binom{\nu}{l}\binom{l}{\frac{l+m-2g}{2}}\binom{\nu-l}{\frac{m+l-k}{2}},$$

$$m \le l, \quad 0<k\le \nu , \tag{A2}$$

where $N=(2^\nu-1)(2^\nu-2)/2$ is the number of triangles involving point A. To obtain the probability for choosing at random any triangle whose sides are m,l,k with $m \le l$, we must sum P^* over all distinct permutations $\pi(1)\pi(2)\pi(3)$ of m,l,k for which $\pi(1)\le\pi(2)$. If, for example, $m<l<k$, then $P(m,l,k)=P^*(m,l,k)+P^*(m,k,l)+P^*(l,k,m)$. We note that the side lengths of equilateral triangles and the small basis of isosceles triangles must be even. The probability $P_E(\nu)$ for an equilateral triangle in the ν cube is then given by

$$P_E(\nu)=\frac{1}{(2^\nu-1)(2^\nu-2)}\sum_{0<d\in G\le\nu}\binom{\nu}{d}\binom{d}{\frac{d}{2}}\binom{\nu-d}{\frac{d}{2}} , \tag{A3}$$

with G denoting the set of even numbers. For an isosceles triangle with small basis we get

$$P_I(\nu)=\frac{3}{(2^\nu-1)(2^\nu-2)}\sum_{d\in G}\sum_{d<l\le\nu-(d/2)}\binom{\nu}{l}\binom{\nu-l}{\frac{d}{2}}\times\binom{l}{\frac{d}{2}} . \tag{A4}$$

*Author to whom correspondence should be sent. Electronic address: a8441dam@awiuni11.bitnet.

[1]S. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).
[2]D. G. Bounds, Nature (London) **329**, 215 (1987).
[3]S. Kirkpatrick, D. C. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[4]J. J. Hopfield and D. W. Tank, Biol. Cybern. **5**, 141 (1985).
[5]I. Rechenberg, *Evolutionsstrategie* (Frommann-Holzboog, Stuttgart und Bad Cannstatt, 1973).
[6]J. H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975).
[7]L. Demetrius, Phys. Scr. **36**, 693 (1987).
[8]L. Demetrius, J. Stat. Phys. **30**, 709 (1983).
[9]I. Leuthäusser, J. Stat. Phys. **48**, 343 (1987).
[10]J. S. McCaskill, J. Chem. Phys. **80**, 5194 (1984).
[11]M. Eigen, J. McCaskill, and P. Schuster, J. Phys. Chem. **92**, 6881 (1988).
[12]R. W. Hamming, *Coding and Information Theory*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1986).
[13]W. Fontana and P. Schuster, Biophys. Chem. **26**, 123 (1987).
[14]S. Kauffman, J. Theor. Biol. **22**, 437 (1969).
[15]G. Weisbuch, C. R. Acad. Sci. Paris **298**, 375 (1984).
[16]C. Amitrano, L. Peliti, and M. Saber, C. R. Acad. Sci. Paris **307**, 803 (1988).
[17]M. Mézard, G. Parisi, N. Sourlas, and M. Virasoro, J. Phys. (Paris) **45**, 843 (1984).
[18]S. Kirkpatrick and G. Toulouse, J. Phys. (Paris) **46**, 1277 (1985).
[19]M. Eigen, Naturwissenschaften **58**, 465 (1971).
[20]M. Eigen and P. Schuster, *The Hypercycle—a Principle of Natural Self-Organization* (Springer-Verlag, Berlin, 1979).
[21]J. Swetina and P. Schuster, Biophys. Chem. **16**, 329 (1982).
[22]M. Eigen, Ber. Bunsenges. Phys. Chem. **89**, 658 (1985).
[23]M. Eigen, Chem. Scr. **26B**, 13 (1986).
[24]P. Schuster, Phys. Scr. **35**, 402 (1987).
[25]P. Schuster and J. Swetina, Bull. Math. Biol. **50**, 635 (1988).
[26]D. R. Mills, C. Dobkin, and F. R. Kramer, Cell **15**, 541 (1978).
[27]M. Zuker and P. Stiegler, Nucleic Acids Res. **9**, 133 (1981).
[28]M. Zuker and D. Sankoff, Bull. Math. Biol. **46**, 591 (1984).
[29]D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[30]E. M. Reingold, J. Nievergelt, and N. Deo, *Combinatorial Algorithms. Theory and Practice* (Prentice-Hall, Englewood Cliffs, NJ, 1977), pp. 190–195.
[31]W. Fontana, Diploma thesis, Universität Wien, 1984.
[32]P. Schuster, in *Complex Systems—Operational Approaches*, edited by H. Haken (Springer-Verlag, Berlin, 1985).
[33]R. G. Palmer, Adv. Phys. **31**, 669 (1982).
[34]J. H. Ward, J. Am. Stat. Assoc. **58**, 236 (1963).
[35]D. Wishart, Biometrics **28**, 165 (1969).

[36]S. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975).
[37]C. Bachas, J. Phys. A **17**, L709 (1984).
[38]Y. Fu and P. W. Anderson, J. Phys. A **19**, 1605 (1986).
[39]M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, San Francisco, 1979).
[40]*Heidelberg Colloquium on Glassy Dynamics,* Vol 275 of *Lecture Notes in Physics,* edited by J. L. van Hemmen and I. Morgenstern (Springer-Verlag, Berlin, 1987).
[41]S. Lin and B. W. Kernighan, Oper. Res. **21**, 498 (1973).