Error thresholds for molecular quasispecies as phase transitions: From simple landscapes to spin-glass models

P. Tarazona
*Institut für Theoretische Chemie der Universität Wien, A-1090 Wien, Austria and Departamento de Fisica de la Materia Condensada, Universidad Autonoma de Madrid, E-28049, Madrid, Spain**

(Received 19 June 1991)

The correspondence between Eigen's model [Naturwissenschaften **58**, 465 (1971)] for molecular quasispecies and the equilibrium properties of a lattice system proposed by Leuthäusser [J. Chem. Phys. **84**, 1884 (1986); J. Stat. Phys. **48**, 343 (1987)] is used to characterize the error thresholds for the existence of quasispecies as phase transitions. For simple replication landscapes the error threshold is related to a first-order phase transition smoothed by the complete wetting of the time surface. Replication landscapes based on the Hopfield Hamiltonian for neural networks allow for the tuning of the landscape complexity and reveal the existence of two error thresholds, bracketing a region of spin-glass quasispecies between the simple quasispecies and the fully disordered mixture of sequences.

PACS number(s): 87.10.+e, 64.60.Cn, 05.50.+q

I. INTRODUCTION

The correspondence between Eigen's model for the molecular evolution of replicating molecules [1–3] and the equilibrium properties of a surface lattice system, proposed by Leuthäusser [4,5] a few years ago, opened a route towards the understanding of the error threshold for the existence of molecular "quasispecies," within the well-established framework of equilibrium phase transitions. But so far, this analogy has not been explored beyond the estimations done in the original work [5]. It remained to be shown that the method may be applied to a general type of landscape, to compare the results with those obtained directly from the time-evolution ordinary differential equations, and, what is more important, to explore the global phase diagram for different replication landscapes, taking advantage of the powerful methods developed for equilibrium phase transitions. The latter prospect is particularly interesting for systems with frustrated interactions, like the spin glasses, which were first suggested by Anderson [6] as an appropriate model for the stability and diversity characterizing biological systems.

The aim of this work is to advance along these lines. We start with a review of Leuthäusser's approach. In Sec. III we describe the method used to obtain the exact (numerical) solution of the problem for simple landscapes and the results obtained with this method, with particular stress in the difference between the behavior of the bulk and the surface, which had been somehow overlooked, and also in how the nature of the error thresholds, regarded as equilibrium phase transitions, depends on the structure of the replication landscape. The "selection transitions;" which have been described for replication landscapes with quasidegenerate maxima are also analyzed within this framework. In Sec. IV we apply the method to landscapes derived from the Hopfield Hamiltonian for neural networks [7], which proves to be a very appropriate model to explore the transition from "simple" to "complex" replication landscapes. Analytic results, using the replica symmetric theory in the limit of infinite size, together with the numerical solution of the problem for finite sizes, give some new insight into how the structure of the quasispecies may be qualitatively changed by the complexity of the replication landscape, leading to the presence of two error thresholds and the coexistence of different molecular types between them. In Sec. V we present a general discussion of the results and the conclusions.

II. MOLECULAR EVOLUTION AS EQUILIBRIUM STATISTICAL MECHANICS

The model for the evolution of self-replicating molecules, in the mathematical formulation given by Eigen and Schuster [2], starts with the description of the molecule as a sequence $(s_1,s_2,\ldots,s_N)$ of a fixed number N of variables s_i, representing the different kinds of monomers which are used to build the macromolecules. There would be four possible values for the s_i, to represent the four different bases in the RNA or the DNA molecules, but here we follow Leuthäusser [4] and take them to be binary variables, with values ± 1, which may be regarded as the purine and pyrimidine basis in a polynucleotide, or just as a (more or less artificial) binary code to represent a complex variable structure. We use S_k, with $k=1,2^N$, as a shorthand for the sequences, and x_k for the concentration of each sequence in population with a very large number of molecules. The time evolution of this concentration follows from the self-replicating nature of the molecules,

$$\frac{dx_k}{dt}=\sum_j W_{kj}x_j-[D_k+\Phi(t)]x_k \tag{1}$$

where W_{kj} are the elements of the replication matrix giv-

ing the probability of getting the sequence S_k as the result of the replication of molecule S_j. This matrix would be diagonal if each sequence produces only perfect copies of itself, but the model introduces the possibility of mutations in the replication process, with low but positive nondiagonal elements in W. The constants D_k in (1) represent the death probability of each molecule and the "flux" term $\Phi(t)$ may be added to include any overall restriction on the total number of molecules. If this term is neglected, the constants D_i may be absorbed into the diagonal terms of W, and we get a simple set of linear differential equations. The largest eigenvalue of the matrix W gives the net production rate of molecules in the stationary state, and the corresponding eigenvector gives the composition of the population in terms of the relative concentrations of each type of molecules. If there is a particular molecular structure or "master sequence" S_0, with a value of W_{00} much higher than any other diagonal term in W, and the replication is very accurate so that the nondiagonal terms are small, then it is clear that the stationary population will be mainly made of S_0 molecules. If the mutation rate is increased, leading to larger values of W_{ij} for $i \neq j$, a cloud of mutants around the master sequence will appear in the stationary population, in what Eigen and Schuster called the "quasispecies" [2]. If the mutation rate is increased still further, there is an "error threshold" at which the structure of the population changes sharply to a random distribution all over the sequence space (the hypercube), with the loss of the genetic information carried by the "quasispecies." Extensive analysis by Eigen, Schuster, McCaskill, and co-workers (see [3] and references therein), has shown that this information catastrophe is a quite general phenomenon which should appear in a broad variety of systems, represented by different structures of the matrix W.

The work of Leuthäusser [4,5], to map the time evolution given by the ordinary differential equations (ODE) into a system at thermodynamic equilibrium, starts with the discretization of time in "generations," so that instead of $x_k(t)$ in (2.1), we have $x_k(i)$, with $i=0,1,\ldots$ to represent the concentration of molecule S_k, i generation times after a given initial population $x_k(0)$. The matrix element W_{jk} is interpreted as the probability for a molecule S_k of producing a molecule S_j in the next generation. The matrix W becomes the discrete time propagator, i.e., the population n generations after $i=0$ is given by

$$X(n)=W^n X(0) , \tag{2}$$

where we use the vector notation $X(i)=(x_1(i), x_2(i), \ldots, x_{2^N}(i))$. This suggests the interpretation of W as a transfer matrix which gives the probability of configuration $X(i+1)$ in a canonical ensemble in terms of the configuration of the nearest-neighbor site $X(i)$ [8]. If we go back from the shorthand notation S_k to the full representation of the molecules as binary sequences $(s_1, s_2, \ldots, s_N)$, we arrive at the description of the population time evolution as a lattice system, with Ising variables $s_j(i)=\pm 1$, at the sites of a two-dimensional square lattice, with j running from 1 to N to represent the position along the molecular chain and, in the time direction, i representing the generations from $i=0$, at the fixed initial state $X(0)$, up to the final population at $i=n$ at the other end (Fig. 1). If the Hamiltonian acting on the two-dimensional lattice may be written as a sum over pairs of nearest-neighbor rows in the time direction:

$$H=\sum_{i=0}^{n-1} h[S(i),S(i+1)] , \tag{3}$$

with an arbitrary function h of the two neighbor sequences, the matrix W is given by

$$W_{jk}=\exp(-\beta h[S_j,S_k]) , \tag{4}$$

with the usual inverse temperature $\beta=1/k_B T$. Any replication matrix may, in principle, be transformed according to (3) and (4) into an effective Hamiltonian H acting on the two-dimensional lattice representation of the population dynamics. However, this would be of little use if the form of this effective Hamiltonian is so awkward that it allows no hope for the statistical analysis of the lattice model. This would be the case if the time evolution of the system follows nearly any rule other than the replication with random point mutations.

If each site along the binary chain is copied exactly with a probability q, independently of the other sites (point mutations), so that $1-q$ is the error rate per site and generation, the general form for the matrix W is given by

$$W_{jk}=A[S_k]q^N\left[\frac{1-q}{q}\right]^{d_H[S_j,S_k]} \tag{5}$$

where $d_H[S_j,S_k]$ is the Hamming distance between the two sequences,

$$d_H[S,S']=\tfrac{1}{2}\left[N-\sum_{k=1}^{N} s_k s'_k\right] \tag{6}$$

and $A[S]$ is the "replication landscape," which is a function of the N variables s_j along the chain represented by

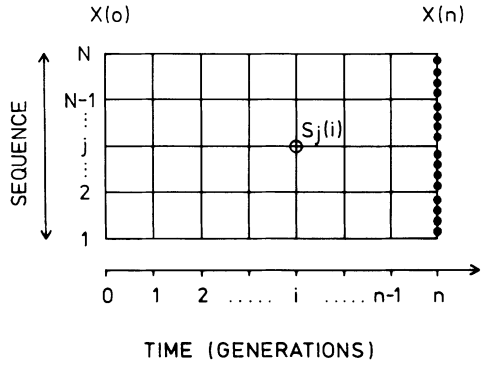

FIG. 1. Sketch of the lattice system used to map the molecular evolution. Time runs along the horizontal axis in generations and the vertical axis represents the molecular sequence. The relevant properties come from the distribution of the last column, $i=n$, or "surface." The dots along this line represent the lack of the vertical interaction given by the replication landscape.

S. This function appears in (5) only with the sequence at the right-hand side, because it gives the measure of the fitness of the "mother molecule" to produce copies (either accurate or mutants) in a generation time. By comparing (5) with (3) and (4) we arrive at the following effective Hamiltonian [9]:

$$-\beta H=\sum_{i=0}^{n-1}\left[\beta\sum_{j=1}^{N}s_j(i)s_j(i+1)+\ln A\,[S(i)]\right]+\frac{nN}{2}\ln[q(1-q)] \tag{7}$$

and the inverse temperature is $\beta=\ln\sqrt{q/(1-q)}$ for $q>0.5$ (the case with $q<0.5$ corresponds to the copy into complementary basis and it is given by the ferro-antiferromagnetic symmetry $\beta\rightarrow-\beta$). The index i runs along the time direction from the initial time, $i=0$, at which the system is set at a given initial configuration, $X(0)$, to the final time, $i=n$, at which we wish to know the molecular population $X(n)$. The Hamiltonian (7) has nearest-neighbor interactions along the time direction and an unspecified interaction, $\ln A\,[S]$, between all the sites along the chain for each value of i, excluding the final $i=n$. This exclusion comes from the asymmetry of the matrix W in (5) and it reflects the intrinsic irreversibility of the original time evolution (1). The structure of the population after n generation times may only be obtained from the structure of the "surface" in a slab of length n. The preceding layer in the slab does not give the structure of the population at time $n-1$, which may only be obtained from the "surface" of a slab with only $n-1$ rows. The limit $n\rightarrow\infty$ gives the steady state of the system, independent of the initial configuration $X(0)$, as the structure of the "free surface" with an infinite "bulk" at one side and nothing at the other side. The properties of this surface will not depend on its particular position, but they may be quite different from those of the bulk, not only because of the lack of neighbors on the right-hand side (reflecting the irreversibility of the original process, with the population at any time tied to the population at the past but not to the future), but also because of the lack of interactions along the surface chain.

The analogy between the original ODE for the time evolution and the surface properties of a lattice system in thermodynamic equilibrium may be applied to any replication landscape A, with the only restriction imposed by our capability to solve the resulting problem in equilibrium statistical mechanics. Although the short-time dynamics of the population may have been changed by the discretization of time, the stationary state, which is going to be our only concern here, should not be affected [10].

III. ERROR THRESHOLDS FOR SIMPLE LANDSCAPES

To study the steady state we take the limit $n\rightarrow\infty$ and the lattice system represented in Fig. 1 becomes semi-infinite in one direction and it has a large but finite size N in the other direction. It may in principle be exactly analyzed with transfer-matrix techniques [11], although this will become computationally difficult as N grows, much in the same way as the direct analysis of the ODE in terms of the eigenvector of the matrix W [2,3]. Our main concern here is to characterize the error threshold as a quasiphase transition, i.e., as the signature of a phase transition in the limit $N\rightarrow\infty$. The analysis of the bulk phase with translational invariance is always much simpler than the study of the surface, and it may be thought [5] that any phase transition will be observed in the bulk phase as well as in the surface. However, the results presented here show that the characteristics of the error threshold are strongly affected by its "surface character," so that a full solution of the surface system may be required. We should also notice that, despite the representation of our system in Fig. 1 as a lattice in two dimensions, the Hamiltonian (7) does not in general have a two-dimensional structure. This would only be the case if the interaction along the chain, given by the replication landscape $A\,[S]$, has only short-range couplings. Any realistic model for $A\,[S]$ includes the interactions between sites which, being far away from each other along the chain, may become close in the folded polymer. The result of these long-range interactions is that the models and, in particular, their phase transitions will have no two-dimensional character whatsoever.

We have obtained the exact (numerical) solution for a few cases, with simple landscapes, using the cluster-variational method, originally derived by Kikuchi [12], in which the free energy of the system is written in terms of the distribution of probabilities for a basic cluster, of p sites, which tiles the full lattice. The results become exact if the basic cluster is taken to be the "ladder" formed by two full consecutive columns in Fig. 1, that is, the full chains for two consecutive generations. This is equivalent to the Bethe solution for a one-dimensional system with a variable represented by the N-component vector $S=(s_1,\ldots,s_N)$, with 2^N different values instead of a simple Ising chain. Of course, the number of values for S grows so fast with N that it makes impossible the direct computation for N larger than say 10 or 12, unless we may use some symmetry of the landscape to reduce the number of independent variables.

A. Replication landscapes with a single sharp maximum

The simplest possible case corresponds to a replication landscape which only depends on the distance (6) to a single "master sequence" $S_0=(\xi_1,\xi_2,\ldots,\xi_N)$. In that case the steady-state concentration of any sequence depends only on its distance to S_0, which may take only $N+1$ different values instead of 2^N. Landscapes with this symmetry have been explored with the original ODE [2,3] and the simplest possible type within the class is

$$A\,[S_0]=A_0 \text{ and } A\,[S]=A_1<A_0 \text{ for any } S\neq S_0\ , \tag{8}$$

which has been studied in detail [2,3,13,14], including the effects of finite population and the limit $N\rightarrow\infty$.

To study this landscape we have to obtain the equilibrium surface structure of a lattice system with the Hamiltonian

$$-\beta H=\sum_{i=0}^{n-1}\left[\beta\sum_{j=1}^{N}s_j(i)s_j(i+1)+\ln(A_0/A_1)\prod_{j=1}^{N}\frac{\xi_j s_j(i)+1}{2}\right], \tag{9}$$

where we have dropped constant terms which only add a trivial shift to the free energy. This Hamiltonian is quite different from the usual two-spin additive interactions of most lattice models; there are up to N-spin interactions which produce strong many-site correlations and make the mean-field approximation inaccurate. Probably for this reason, this type of landscape was not considered by Leuthäusser [5], who turned to other models leading to a more "natural" effective Hamiltonian in the lattice.

Within our "ladder-cluster" method, there is no particular difficulty in studying this problem. The minimization of the free energy with respect to the probability distribution of the possible configurations for the "ladder," $z_i[S(i),S(i+1)]$, leads to a system of coupled equations for the chain distributions, $y_i[S(i)]$, and the Hamiltonian (9) is simple to handle within this description. The equations (see the Appendix) may be solved numerically, with the usual procedure for surface problems in statistical mechanics. First we get the properties of the "bulk" phase, with translational invariance along the "time" direction, and then we study a slab coupled to the bulk solution at one side and with a free surface at the other. The slab has to be large enough to have a surface structure independent of its size. The "surface layer" approximation, used in [5] for a different landscape, is equivalent to taking the slab width of only one lattice site, so that the surface is forced to follow closely the structure of the bulk.

In Figs. 2 and 3 we give the usual representation of the quasispecies [3]: the relative concentration of the sum of all the sequences at the same distance of the master sequence. The data correspond to a system with $N=20$ and a replication landscape (8) with $A_0/A_1=10$. Figure 2 was calculated with the bulk distributions and Fig. 3 with the surface distribution, using a slab width of 40 sites. The differences between the two figures are evident. Both of them show a clear error threshold, at $1-q=0.11$, beyond which the population is randomly distributed over the sequence space (so that most of the sequences are at a distance of about $N/2$ from S_0). For any error rate below the threshold the distribution of the bulk has an overwhelming majority of master sequences and the error threshold appears as a very sharp change in the structure of the system, clearly reminiscent of a first-order phase transition. In Fig. 3, the surface shows a very different behavior, the structure of the quasispecies changes in a smooth way when approaching the error threshold, the quasispecies becomes a broad distribution, in which the master sequence only provides a small number of sequences to the total population. This is precisely the most celebrated characterization of the error threshold for molecular quasispecies [3], which makes clear that the correspondence between the ODE and the equilibrium lattice system is well founded, but only if in the latter we obtain the properties of the free surface. The bulk phase gives the correct value for the error threshold but a wrong structure of the quasispecies in its neighborhood.

To understand the nature of the error threshold as a phase transition and to analyze the difference between bulk and surface, it is better to shift to a representation in terms of an order parameter. The natural choice for lattice systems with Ising variables is the projection of the population on the master sequence:

$$m=\frac{1}{N}\sum_{j=1}^{N}\xi_j\langle s_j\rangle \tag{10}$$

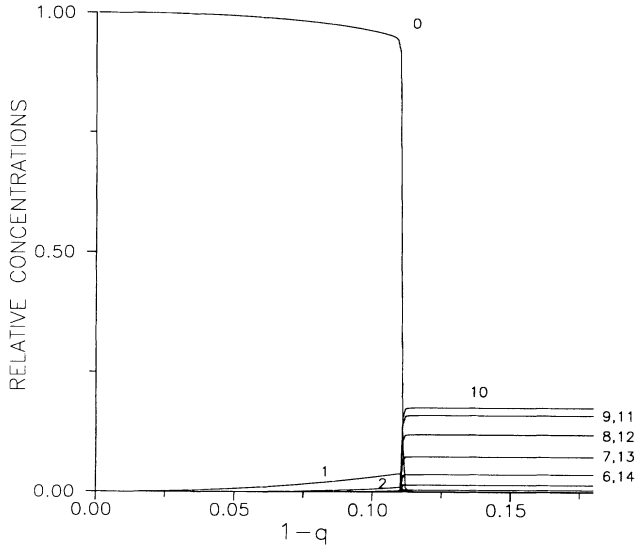

FIG. 2. Relative concentration of the sequences at given Hamming distance from the master sequence, as a function of the error rate $1-q$. The data correspond to the bulk distribution for a sequence with $N=20$ and a single-peak landscape, as in Eq. (8), with $A_0/A_1=10$.

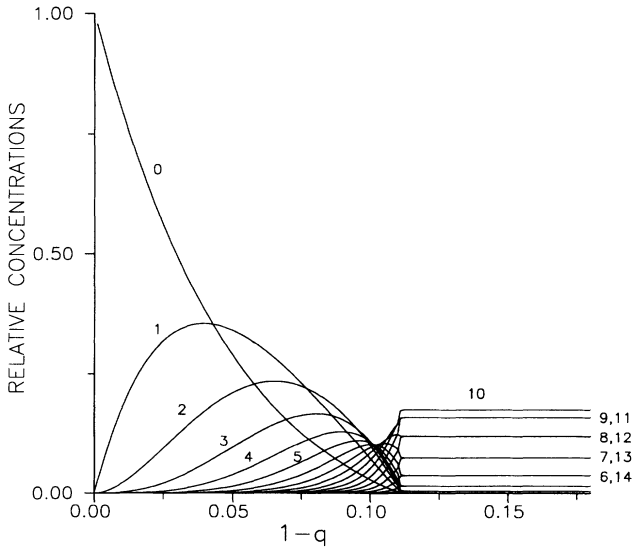

FIG. 3. Distribution of the steady-state population, given by the relative concentrations for the same system as in Fig. 2, but obtained from the surface layer.

where $\langle\ \rangle$ represents statistical average over the population at a given generation. For $m=1$ the population is fully made of master sequences, for $m=0$ it is dispersed over the sequence space, and for $m=-1$ it would be made of the sequence complementary to the S_0, i.e., in the opposite corner of the hypercube [with the landscape (8) this would never appear, but is becomes useful with other replication landscapes]. An alternative, but fully equivalent representation would be given by the average distance of the population to the master sequence, $\langle d\rangle = N(1-m)/2$.

In Fig. 4 we present the order parameter m evaluated at the bulk, m_b, and at the surface, m_s. At the error threshold both m_b and m_s go to error, but m_b shows the sharp decay of a first-order phase transition (only slightly rounded by the finite size of the sequence) while m_s goes continuously to zero with a divergence of the derivative at the transition (again rounded by the finite N). The analysis of the order parameter in the chains next to the surface makes clear that we are in the presence of what is known as "complete wetting" of the surface by the disordered phase, a phenomenon which has been extensively analyzed over the past decade with many theoretical models and experiments [15]. Having a surface with complete wetting, some of the most typical characteristics of first-order transitions like hysteresis and nucleation disappear. The wet surface provides a barrierless nucleation site so that the transition is smoothed on the wet side of the phase diagram. We may conclude the analysis of the replication landscape (8) by stating that the error threshold, in this case, corresponds to a quasi-transition of first order in the bulk with complete wetting at the surface, giving the characteristic one-sided smoothness which has been carefully analyzed with the ODE [2,3,13,14]. The same applies to replication landscapes with a broader peak around the master sequence as far as $A[S]$ goes flat as the distance to the master sequence goes to $N/2$.

B. Quasidegenerate quasispecies

Our second example of simple landscape has two maxima at opposite corners of the hypercube, so that both $S_0=(\xi_1,\ldots,\xi_N)$ and its complement, $S_N=(-\xi_1,\ldots,-\xi_N)$, have larger values of $A[S]$ than any other sequence. If $A[S_0]=A[S_N]$ we have degenerate quasispecies and the population, below the error threshold, is a mixture of both of them. However, any small difference between $A[S_0]$ and $A[S_N]$ would be enough to erase one of the quasispecies from the steady-state population. The reason for this instability is that the two quasispecies come from different eigenvectors of the matrix W, so that they compete with each other, and only if they have strictly equal eigenvalues do they coexist. Schuster and Swetina [16] have proved that (in the limit of perfect replication, $q\to 1$) the two peaks in the landscape may only cooperate to build up a quasispecies if the two maxima are very close to each other, at a distance less than 3.

When the two quasispecies are competing with each other and they are nearly but not quite degenerate, an interesting behavior may be observed [16] if one of them has a high but narrow maximum and the other is not as high but broader. The simplest example is

$$A[S_0]=A_0\ ,\quad A[S_N]=A_N\ ,\quad A[S_{N-1}]=A_{N-1}$$

and for any other S

$$A[S]=A_1<A_{N-1}<A_N<A_0\ . \tag{11}$$

For high replication accuracy the quasispecies is centered at S_0 but as the error rate, $1-q$, increases the broader quasispecies would be at advantage around S_N. The transition from one to the other was found to be very sharp [16], in contrast with the relative smoothness of the error threshold.

We have studied a quasidegenerate landscape of this kind with $N=20$, $A_0/A_1=10$, $A_N/A_1=9.9$, and $A_{N-1}/A_1=2$. The same order parameter (10) may be used to follow the two possible quasispecies, with $m=1$ for S_0 and $m=-1$ for S_N. The results for the surface distribution show the same qualitative features as those studied by Schuster and Swetina [16], and the order parameters shown in Fig. 5 give a clear understanding of the differences between the "selection" transition, at

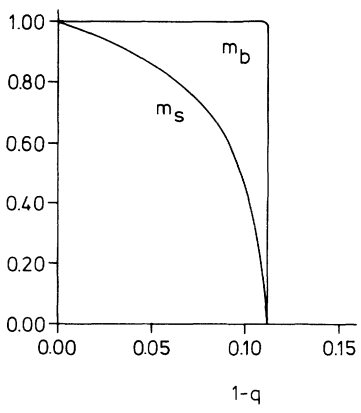

FIG. 4. The order parameter as defined in Eq. (10), for the bulk, m_b, and the surface, m_s, of the same system as in Figs. 2 and 3.

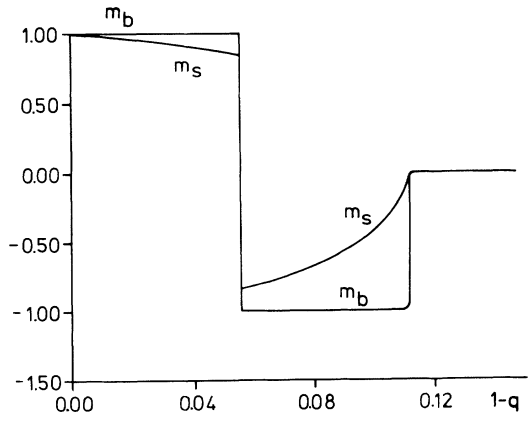

FIG. 5. The order parameters in the bulk, m_b, and at the surface, m_s, for sequences with $N=20$ in a quasidegenerate double-peak replication landscape, as described in Eq. (11) with $A_0/A_1=10$, $A_{20}/A_1=9.9$, and $A_{19}/A_1=2$.

$1-q=0.056$, and the error threshold at $1-q=0.113$. The bulk order parameter m_b shows sharp changes at both transitions, indicating their first-order character, but the surface, m_s, goes through the "selection" transition without any signature of wetting. Thus the qualitative difference between the two transitions corresponds to the presence or absence of surface wetting in first-order phase transitions.

C. Mattis landscape

Our last example of simple landscapes was proposed by Leuthäusser [5] and it is derived from the Mattis Hamiltonian [17] for long-range spin systems. The replication landscape still depends only on the distance to a master sequence $S_0=(\xi_1,\ldots,\xi_N)$, but as a smooth function,

$$A[S]=\exp\left[\frac{K}{2N^2}\sum_{j\neq j'}^{N}\xi_j\xi_{j'}s_js_{j'}\right]. \tag{12}$$

The equivalent Hamiltonian in the lattice system (7) becomes

$$-\beta H=\sum_{i=0}^{n-1}\left[\beta\sum_{j=1}^{N}s_j(i)s_j(i+1)+\frac{K}{2N^2}\sum_{j\neq j'}^{N}\xi_j\xi_{j'}s_j(i)s_{j'}(i)\right] \tag{13}$$

where we have again dropped the last constant term in (7). The original Mattis Hamiltonian is only the last term in the brackets of (13) but with N instead of N^2 in the denominator, to have a difference between the maximum and the minimum energy of order N and give a nontrivial thermodynamic limit $N\to\infty$. As pointed out by Leuthäusser [5], this would correspond to a maximum fitness landscape growing exponentially with the length of the sequence, which is in contradiction with experimental results for the RNA of coliphages [18]. The landscape (12) gives $A_{\max}/A_{\min}\approx\exp(K/2)$, as $N\to\infty$ and the nontrivial thermodynamic limit of the Hamiltonian (13) may only appear through the nearest neighbors coupling along the time direction.

With the introduction of Gaussian variables [5,19], $m(i)$, $i=0$, $n-1$, the partition function with (13) is exactly transformed into

$$Z_N=\exp\left[-\frac{nK}{2N}\right]\left[\frac{K}{2\pi}\right]^{n/2}N^n\times\int_{-\infty}^{\infty}\prod_{i=0}^{n-1}dm(i)\exp\left[-\frac{N^2K}{2}\sum_{i=0}^{n-1}m(i)^2+N\ln Z_1\right] \tag{14}$$

where Z_1 is the partition function of a one-dimensional Ising chain

$$Z_1=\mathrm{Tr}_{\{s\}}\left[\exp\left[\sum_{i=0}^{n-1}[\beta s(i)s(i+1)+Km(i)s(i)]\right]\right] \tag{15}$$

with the usual definition of the trace over the spin configurations. The Gaussian variables $m(i)$ provide independent fields $Km(i)$ acting at each site along the chain. For large values of N, the interals in (14) may be obtained by the saddle-point method leading to

$$\ln Z_N=-\frac{nN^2K}{2}\sum_{i=0}^{n-1}m(i)^2+N\ln Z_1 \tag{16}$$

evaluated at its minimum with respect to $m(i)$ which, rescaled as $m\to m/N$, become precisely the order parameters defined in (10).

The bulk order parameter m_b is easily obtained from (16) as the solution of the following equation:

$$m=\frac{e^{\beta}\sinh(Km/N)}{[e^{-2\beta}+e^{2\beta}\sinh^2(Km/N)]^{1/2}}, \tag{17}$$

which has only the trivial solution, $m=0$, unless e^{β} is of order $N^{1/2}$, in which case the coupling along the "time direction" is strong enough to compensate for the extra factor N in the denominator in (13). The relevant temperature parameter is

$$\Theta=\frac{N}{Ke^{2\beta}}=\frac{(1-q)N}{qK}\approx\frac{(1-q)N}{2q\ln(A_{\max}/A_{\min})}, \tag{18}$$

and there is a critical point at $\Theta_c=1$ at which the system becomes ordered in one of the two degenerate states around the master sequence S_0, or its complement S_N, with $m=\pm(1-\Theta^2)^{1/2}$. That is, the error threshold appears in this system as a continuous, second-order bulk phase transition, in the limit of $N\to\infty$. For any finite N it would be rounded and it only appears in (17) as an artifact of the saddle-point approximation for (14).

As in the previous cases, the population near the error threshold has to be obtained from the surface structure. The expectation for a normal surface with a critical bulk [20] is that the order parameter m_s will go to zero like $\Theta_c-\Theta$, and the population distribution would go through the error threshold in a much smoother way than in Fig. 3. These trends are confirmed by calculations of the surface structure within the saddle-point approximation for (14) and also by the exact solution of the problem with the method described above (Fig. 6).

IV. COMPLEX LANDSCAPES WITH HOPFIELD HAMILTONIAN

In the preceding section we have explored different kinds of simple replication landscapes to characterize the error threshold as a phase transition. However, the main interest of Leuthäusser's formalism is to take advantage of the powerful methods developed, in equilibrium statistical physics, to deal with complex systems [21]. A possible route towards complex replication landscapes was already suggested by Leuthäusser [5]: the Mattis landscape (12) may be considered as a particular case, with $p=1$, of a Hopfield landscape,

$$A[S]=\exp\left[\frac{K}{2N^2}\sum_{j\neq j'}^{N}\sum_{\mu=1}^{p}\xi_j^{\mu}\xi_{j'}^{\mu}s_js_{j'}\right], \tag{19}$$

where there are p different master sequences or overlap-

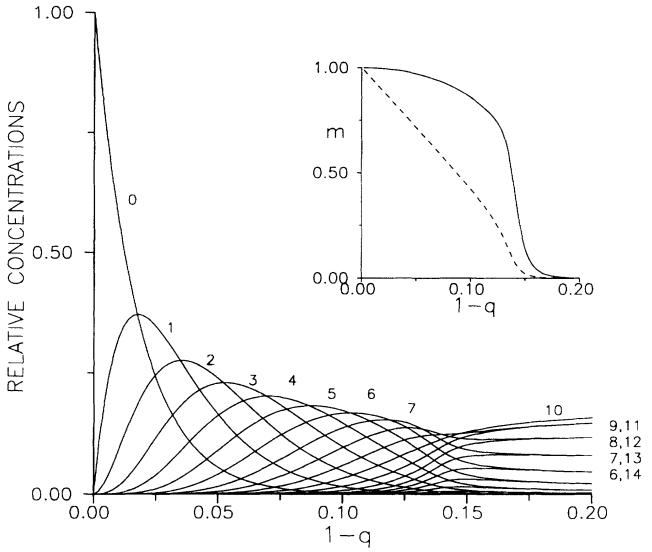

FIG. 6. Relative concentrations in the steady-state populations for molecules at the distance given by the numbers, from the master sequence, as function of the error rate $1-q$ for sequence with $N=20$ in a Mattis replication landscape, Eq. (12), with $A_{\max}/A_{\min}=10$. The degeneracy between the master sequence and its complement has been broken by taking $A(20)=0$, to get a simpler representation. The inset represents the bulk (full line) and surface (dashed line) order parameters. These are results from the exact numerical solution and, because of the relatively small value of N, they still show some quantitative difference with the results of the saddle-point approximation described in Eqs. (16)–(18), which becomes increasingly accurate for larger N.

ping "patterns." The equivalent Hamiltonian in our lattice system is the direct generalization of (13). It is similar to the original Hopfield Hamiltonian for neural networks [7] but, as in (13) it contains an extra factor N in the denominator which may only be compensated by the nearest-neighbor interactions along the time direction. The solution may be obtained following the same procedure as in the Mattis Hamiltonian [5,19], but using p different Gaussian variables, $m^{\mu}(i)$, for each site along the chain. The saddle-point equations give self-consistency equations for these variables, which in the bulk phase and neglecting terms in $1/N$, become

$$m^{\mu}=\frac{1}{N}\sum_{j=1}^{N}\xi_j^{\mu}\frac{\sum_{\nu=1}^{p}m^{\nu}\xi_j^{\nu}}{\left[\Theta^2+\left[\sum_{\nu=1}^{p}m^{\nu}\xi_j^{\nu}\right]^2\right]^{1/2}}\ . \tag{20}$$

Each of the m^{μ} variables corresponds to the averaged overlap with one of the master sequences,

$$m^{\mu}=\frac{1}{N}\sum_{j=1}^{N}\xi_j^{\mu}\langle s_j\rangle \tag{21}$$

and the results depend on the particular set of patterns which have been chosen. If the number of patterns p is kept finite as $N\to\infty$, then the random fluctuations are suppressed and the order parameters become self-averaged, i.e, the average over the chain in (20) may be substituted by the average over the random pattern, ξ^{μ}, denoted by $\langle\langle\ \rangle\rangle$:

$$m^{\mu}=\left\langle\!\!\left\langle \xi^{\mu}\frac{\sum_{\nu=1}^{p}m^{\nu}\xi^{\nu}}{\left[\Theta^2+\left[\sum_{\nu=1}^{p}m^{\nu}\xi^{\nu}\right]^2\right]^{1/2}}\right\rangle\!\!\right\rangle . \tag{22}$$

This was the stage reached by Leuthäusser [5] and, as discussed there, the solutions of (22) would follow the qualitative trend of the Hopfield Hamiltonian, which leads to equations similar to (22) but with $\tanh(x)$ instead of $x/(1+x^2)^{1/2}$ (or the equivalent function used in Ref. [5]). The system still has the same critical temperature, $\Theta_c=1$, as in the $p=1$ case and below this temperature the system has $2p$ degenerate states. But the coexistence of the different master sequences in the steady-state population is an artifact of the absolute symmetry assumed for the replication landscape. As discussed in Sec. III B, any small difference between the maxima for the different master sequences will destroy the balance and only one of the $2p$ sequences would contribute to the population. The landscape (19) also has secondary maxima, created by the overlap of any odd number of patterns [19], but in the self-averaging limit, $p/N\to 0$, they only appear as metastable states below the critical value of Θ, so that this seems to add nothing new to the simple landscapes of the preceding section.

However, if one actually carries out the computation of the order parameters m^{μ} from (21) (without assuming self-averaging) for large but finite N ($N=100-1000$ would be the relevant values for simple strands of RNA), the result show something very different. In Fig. 7 we present a set of results for $N=300$ and $p=5$, which correspond to different random choices of the patterns ξ_j^{μ}. The differences between them make clear that they are not self-averaging, but there are some features present in all them. At low Θ (high replication accuracy) there is always a pattern which, because of the random overlaps with the others, has a small but crucial advantage in building up the stationary state quasispecies. The left part of the diagram is very similar to the result for the Mattis landscape and the constant lower values for the other order parameters just reflect their random overlap with the ground state. As the error rate increases there is a region of strong changes. The structure of this region depends on the particular choice of the random patterns, in some cases there are sharp changes reminiscent of phase transitions, in other cases the changes are smoother, but in general they lead to a region in which several order parameters have similar (absolute) values. A further increase of Θ leads to an error threshold in which all the m^{μ} become zero indicating the end of the quasispecies. These features are enhanced by increasing p, the transition between the two regimes becomes sharper and the structure of the intermediate region becomes more and more complex.

Some understanding of the phenomenon may be achieved by the analysis of the average with respect to the patterns, in the limit $N\to\infty$, but keeping a constant

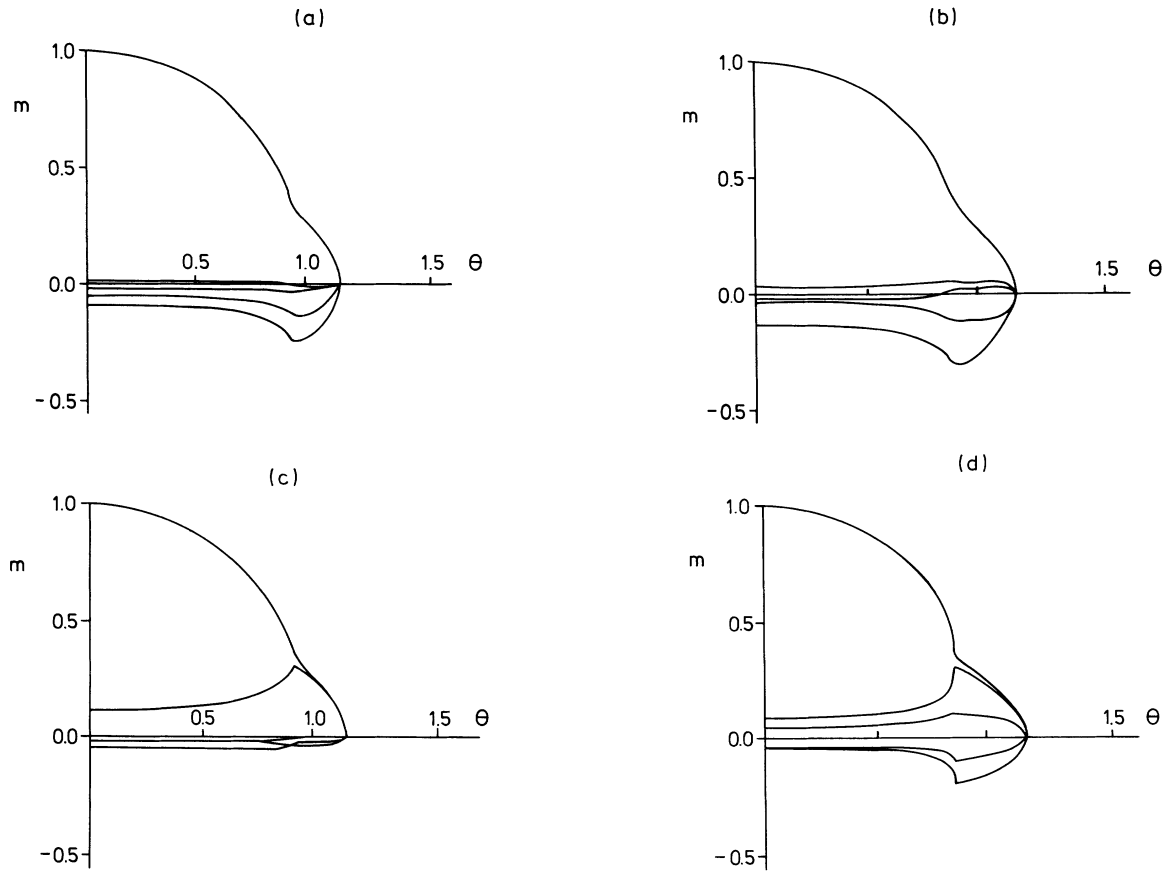

FIG. 7. The order parameters m^{μ}, for a Hopfield landscape with $N=300$, $p=5$ master sequences and $A_{\max}/A_{\min}=10$, as a function of the effective temperature defined in Eq. (18). The four different sets correspond to different random choices of the patterns $\xi_j^{\mu}=\pm 1$.

ratio $\alpha=p/N$. The case considered by Leuthäusser corresponds to $\alpha=0$, in which the self-averaging makes it unnecessary to average over the patterns. For $\alpha>0$, the average may be done with the replica symmetric theory following a procedure similar to that done by Amit, Gutfreund, and Sompolinsky [22] for the original Hopfield model for neural networks. As in the simple Mattis landscape [(12) and (13)] the extra factor $1/N$ would give only the trivial solution $m^{\mu}=0$ unless the coupling along the time direction is strong enough to compensate it. This leads to a mapping into the phase diagram for original Hopfield model [22], with Θ defined in (18), as the relevant temperature. For finite but small values of α there are two different phase transitions. At low Θ we recover the Mattis states, as fully degenerate (because of the limit $N\to\infty$) ground states. The lower transition line, $\Theta_C(\alpha)$, is a first-order phase transition leading to a spin-glass phase in which the system has no macroscopic (of order 1) overlap with any of the patterns but it has nonzero Edwards-Anderson order parameter [21]. The transition line at the higher $\Theta_G(\alpha)$ is the second-order transition to the fully disordered (paramagnetic) phase.

To make sure that $\alpha=p/N$ is the appropriate parameter to measure the "complexity" of the replication landscape for finite but large N, we have analyzed the case $\alpha=0.02$ for different values of N from 50 to 1000 (i.e., p from 1 to 20). For each random choice of the patterns, $\xi_j^{\mu}=\pm 1$, we have solved (20) as a function of Θ, with different initial conditions to get in each case the absolute equilibrium state. The upper error threshold Θ_u was easily determined by the vanishing of all the m^{μ}. To get a criterion for the lower one was more difficult, because its nature is very variable. Finally we used the value Θ_l giving the maximum of

$$r=\sum_{\mu=1}^{p}(m^{\mu})^2-\max(m^{\nu})^2\ , \tag{23}$$

i.e., the sum of the squared order parameters excluding the largest, which was in good agreement with the apparent location of the threshold by inspection of Fig. 7. The values of Θ_u and Θ_l depend on the random patterns, and were averaged over a series (typically of 20 sets) to get the mean value and the standard deviation. These are shown in Fig. 8, which makes quite clear that α (and not any other combination of p and N) is the relevant parameter to describe the "complexity" of the landscape (19). Of course, as $p\to 1$ we go back to the simple Mattis case, no matter the value of N, but for $p>5$ ($N>250$) the only

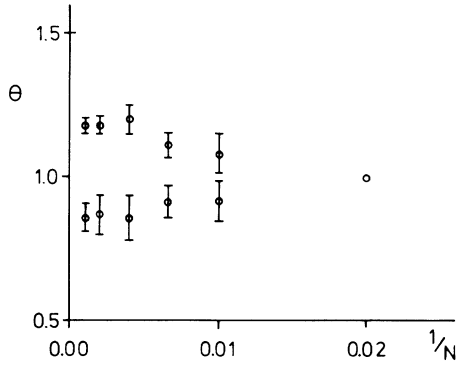

FIG. 8. The effective temperature parameter Eq. (18) evaluated at the upper, Θ_u, and lower, Θ_l, error thresholds for a series of Hopfield landscapes with $\alpha=p/N=0.02$, as a function of $1/N$. The central points and the error bars correspond to the mean value and the mean quadratic deviation in series of 20 random choices of the p master sequences.

effect of increasing N keeping constant α is to decrease the error bars without appreciable shift of the mean values.

Repeating the process for different values of α we locate the two transition temperatures, or error thresholds, in terms of the complexity parameter α. Both lines shown in Fig. 9 go to $\Theta=1$ as $\alpha\to 0$, giving back the single error threshold of simple replication landscapes. As α increases Θ_u goes above 1 and Θ_l goes below 1, opening a gap between the two error thresholds. The comparison with the predictions of the $N\to\infty$ limit show that Θ_u follows the track of Θ_G but always above it, while Θ_l

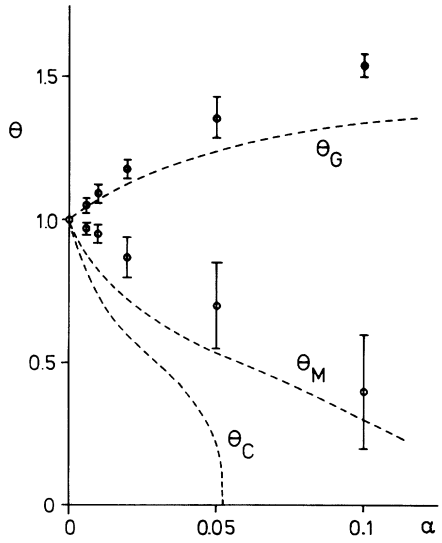

FIG. 9. The upper and lower error thresholds, Θ_u and Θ_l, for Hopfield landscapes for different values of $\alpha=p/N$. The central point and the error bars correspond to the mean value and the mean quadratic deviation in series of 20 random choices of the master sequences. THe sequence length N was adapted to the value of p to get mean values independent of N, as shown in Fig. 8. The dashed lines correspond to the averaged transition lines (as labeled) for the pure Hopfield Hamiltonian in the limit $N\to\infty$ as given in Ref. [22].

seems to follow not the first-order transition line between the Mattis states and the "spin-glass" Θ_C, but the line Θ_M at which the Mattis states become unstable [22], and again with $\Theta_l>\Theta_G$. This is easy to understand because the results for $N\to\infty$ correspond to the transition temperatures averaged over all the Mattis states (and there are an infinite number of Mattis states), while the results for finite N correspond to the transition from the most stable of the possible states (the one giving the equilibrium configuration at each Θ), so that the transition lines are necessarily pushed towards higher temperatures.

To check how robust the results are with respect to changes in the model landscape, we have studied the case $\alpha=0.01$ (with $N=500$ and $p=5$) with a modified version of (19) including a weight factor w^μ, for each of the master sequences, $\mu=1$ to p, which are taken as random variables with unit mean value. Increasing the dispersion of the w^μ increases the difference between the main peaks in the landscape, which otherwise comes only from their random overlaps. The results show similar mean values for the error thresholds. The only apparent effect is to produce a larger dispersion between the different realizations of the system, as may be expected.

Altogether, we may confirm the existence of the two error thresholds as a result of the complexity of the replication landscape. The region between the two thresholds becomes a spin-glass phase in the limit of $N\to\infty$ and is characterized, for large but finite N, by the simultaneous overlap with many of the patterns. The implications for the structure of the quasispecies in the spin-glass phase are discussed in the next section. To end this one, we should just remember that Fig. 9 corresponds to the bulk phase diagram and the real population structure may only be obtained from the surface layer. However, as in all the cases above, the location of the error threshold may be obtained from the bulk phase diagram and we may predict that the structures across the lower (first-order) transition are not going to be changed qualitatively, while near the upper error threshold the surface would have a smoothed behavior as in the simple Mattis case.

V. DISCUSSION

We have applied the formalism proposed by Leuthäusser to map Eigen's model of molecular evolution into a lattice system in thermodynamic equilibrium, with a direction to represent the time. The steady-state population is given by the surface structure along this "time" direction. The exact solution for simple replication landscapes allow us to characterize the quasispecies error threshold as a phase transition. For landscapes with a single sharp peak, which have been extensively analyzed within the ODE of the original model, we found that (for sequence length $N\to\infty$) the error threshold is a first-order phase transition, with complete wetting of the "time surface" by the disordered phase. It is this wetting behavior that produces the characteristic one-sided smoothness of the changes in the quasispecies approaching the error threshold. The contrast with the sharp "selection transition," which had been described for

double-peak landscapes, is due to the lack of complete surface wetting in this case. We expect that this wet first-order transition is the most common case for simple models of the replication landscape. The requirement for this seems to be that, in terms of the distance d to the master sequence, the replication landscape goes flat around $d=N/2$, with $\ln[A(d)/A(N/2)]\sim|1-2d/N|^{\nu}$ and $\nu>2$ (the case in Sec. III A corresponds to $\nu\rightarrow\infty$). For the Mattis landscape proposed by Leuthäuser [5], $\nu=2$ and the error threshold becomes a continuous (second-order) transition. Also in this case the surface structure changes much more smoothly than the bulk near the error threshold, so that the order parameter (or the average distance of the population to the master sequence) varies linearly with the error rate.

The extension to a landscape based on the Hopfield model allows us to study the effects of the landscape complexity. As in the original Hopfield model for neural networks, the ratio $\alpha=p/N$, between the number of master sequences and the length of the sequence, is the relevant parameter to measure the "complexity" of this model landscape. The self-averaging regime studied by Leuthäusser [5] is valid in the limit $\alpha=0$, even if p is large but finite as $N\rightarrow\infty$, but it gives no qualitative change with respect to the simple $p=1$ case. For $\alpha>0$ there is not self-averaging, so that the results depend on the patterns of the master sequences. The relevant properties come out of the "quenched average" over these patterns as usual in spin glasses [21] and which has already been described by Demetrius [23] in the context of chemical kinetics. This has been done here with the replica symmetric theory for $N\rightarrow\infty$ and with the solution for finite N (between 100 and 5000) of series with random patterns, with p between 5 and 50. The result is that for small but finite α a new global structure appears. The system presents two different error thresholds, which depend on α, rather than on any other combination of p and N, and they only collapse as $\alpha\rightarrow 0$. Below the lower threshold error, Θ_l in terms of the effective temperature (18), the system is in a Mattis state, so that the quasispecies is centered around one of the master sequences. Between Θ_l and the upper error threshold Θ_u, the system has a structure which, for $N\rightarrow\infty$, becomes a spin-glass phase. The structure of this phase may be understood, following the close connection with the original Hopfield Hamiltonian [22]. Besides the main $2p$ minima of the Hamiltonian (maxima of the replication landscape) at the master and their complementary sequences, there are secondary minima (maxima) produced by the overlap of any odd number of sequences. For large p (and finite N) these secondary peaks of the replication landscapes form a complex network, connected through the saddle points which appear at the overlap of any even number of patterns. The secondary peaks are always lower than those of the master sequences, but as the quasispecies becomes broader, by increasing the error rate (or Θ), the quasispecies percolates through the network of secondary peaks and this provides a collective broad peak which becomes more favorable than the narrow and higher peaks at the master sequences, much as in the simple case of "quasidegenerate" quasispecies of Sec. III B. The use of the term "percolation" implies that it is not just the spreading of the quasispecies over a small subspace of the full sequence space, the network of secondary maxima contains a tiny fraction of the possible sequences but it spans over the full sequence space in a very complex network which gives to the spin-glass quasispecies a qualitative difference with the simple quasispecies centered around a single maximum of the replication landscape. The upper error threshold corresponds to the usual spreading of the quasispecies over the full sequence space with the loss of any genetic information.

McCaskill and co-workers [3,24] proposed a Watson renormalization scheme to study the error threshold in terms of statistical distributions of replication rates, following the work of Anderson [25] for the localization spins or electrons in the impurity conduction band. The idea is that those sites in the sequence space connected by a mutation path with high replication rates may be treated as a cluster, so that the Brillouin-Wigner perturbation analysis of the replication matrix is not messed up by the multiple connections within the clusters. This seem to lead directly to the spin-glass quasispecies described above, with a cluster of secondary maxima competing with the isolated main peaks. For the simple Mattis landscape, $p=1$, the predictions for the error threshold obtained with this approximation are in fair agreement with the exact results. However, this treatment fails to show any hint about the presence of the intermediate spin-glass quasispecies in the Hopfield landscape, even for fairly large p, and the error threshold is still predicted to be around $\Theta=1.0$, nearly independently of the value of p. This failure is probably due to the mean-field approximation used in the derivation of the statistical treatment [25]. In this context it is worth pointing out that the spin-glass quasispecies only appears as the result of the random overlaps between the p master sequences. If these overlaps are made zero, either by taking the limit $N\rightarrow\infty$ with finite p or by selecting the patterns to be orthogonal even with finite N, then the "complexity" of the landscape disappears and we recover the results of the simple Mattis landscape independently of the number of patterns. It is the overlap among the random patterns that creates the network of secondary maxima and the spin-glass quasispecies. This implies a kind of correlations that is missed by the statistical treatment [24] which only contains the information about the distribution of landscape values but not the correlation between the values at neighbor sites.

The use of a "spin-glass" analogy in molecular evolution was first proposed by Anderson [6], as an alternative to the model of Eigen and Schuster [2] which could provide for both the stability and the diversity in the population. As discussed in Sec. III B, for simple landscapes the coexistence of quasispecies around different master sequences may only be achieved with the precise tuning of the landscape which would never appear in real systems. Without this coexistence it is hard to imagine how prebiotic systems could evolve and form cross-catalytic networks linking the genetic information to metabolism, and the formation of quasispecies of polynucleotides would be

the end rather than the beginning of self-organization. The "escape" through the nonlinearity of the ODE defining the molecular evolution was already proposed by Eigen and Schuster [2] leading to the concept of hypercycles. These ideas have been extensively explored in the past decade, but the overall picture which has emerged is that most of these systems will again evolve in a unique stable molecular structure unless the cross-catalytic effects are tailored in a very precise form, which is unlikely to appear spontaneously.

The alternative proposed by Anderson [6] uses the conjugation of the sequences to allow for increasing the sequence length, leading to a sequence space more complex than Eigen's hypercube and also to a mutation mechanism which may direct the evolution along a preferential path much more efficiently than the independent point mutations [26]. The final ingredient in Anderson's model is a death function which depends on the molecular sequence through a spin-glass Hamiltonian, with random long-range couplings between the sites along the sequence, which will allow for stability and diversity at the same time. The complexity of the mutation-replication-growth mechanism through conjugation of sequences only allows for computer simulations of the model [27].

What comes out of this work is that the spin-glass quasispecies may also be obtained within the model of Eigen and Schuster [2], even in its most simple version with fixed sequence length and uncorrelated point mutation. Leuthäusser's mapping makes it possible to analyze it as a system in thermodynamic equilibrium, with the powerful theoretical techniques developed for these systems [21]. The Hopfield landscape seems also to be a better (theoretical) choice than the Sherrinton-Kirpatrick model used by Anderson [6]. The parameter α allows one to turn on the complexity of the replication landscape and shows that for moderate values, $\alpha < 0.14$, there are two error thresholds. Too accurate replication will drive the system into the Mattis region, with the whole quasispecies around one single master sequence. Only in the intermediate range of mutation rates is there a spin-glass quasispecies, with a complex structure of coexisting sequences, which is stable at the linear order for the replication equations. We may speculate that small nonlinear terms may now be included, acting as a perturbation but on the basis of the intrinsic "stability of the diversity" granted, as the linear level, by the complexity of the landscape.

This work has only been concerned with the deterministic limit of the molecular evolution, as given by the limit of infinite population. In the map to the spin-lattice model this corresponds to the perfect thermodynamical equilibrium. The stochastic effects in finite populations have been studied before for simple landscapes [13], showing clear evidence of the error threshold (at a slightly shifted value of the error rate) even for quite small populations. The stochastic effects may be more important for complex landscapes, because the existence of metastable states may freeze the system in configurations quite out of the true equilibrium. We know that this is the case for the original Hopfield model and work is in progress to investigate these effects in the present context. A particularly appealing idea has been raised very recently by Derrida and Peliti [29,30] about the spin-glass structure of the population produced by a purely stochastic process for a completely flat replication landscape. We have two very different sources of population diversity with spin-glass structure: the landscape complexity for infinite population and the finite population effects in the simplest possible landscape, corresponding to opposite and clearly oversimplified limits of real systems. The crosseffects between these two sources, as well as the role of geographical separation, should be explored.

ACKNOWLEDGMENTS

At different stages in the realization of this work I have benefited from useful discussions with M. Eigen, L. Peliti, P. Stadler, J. McCaskill, and very particularly with P. Schuster whom I also acknowledge for the careful reading of the manuscript and the hospitality at the Institut für Theoretische Chemie in Vienna. Financial support for this work was provided by the Direccion General de Investigacion Cientifica y Tecnica of Spain under grant SEUI: PB0237 and the "Programa de Perfeccionamiento y Movilidad de Personal Investigador."

APPENDIX: BETHE SOLUTION FOR SIMPLE REPLICATION LANDSCAPES

Using the vector notation, $S=(s_1,s_2,\ldots,s_N)$, the system in Fig. 1 may be treated as a one-dimensional chain with a variable S at each site i, taking any of the 2^N values which represent the full configuration of the binary sequence. The Bethe approximation for this system becomes exact as a variational expression for the free energy in terms of the probability $Z_{i,i+1}[S,S']$ of finding the configurations S and S' at the neighboring sites i and $i+1$ along the chain [12]. The probability of finding a single site i in the configuration S is represented by $Y_i[S]$ and it may be obtained from the two-site probability on the left or on the right of the site i:

$$Y_i[S]=\sum_{\{S'\}} Z_{i,i+1}[S,S']=\sum_{\{S'\}} Z_{i-1,i}[S',S] \tag{A1}$$

where $\{S\}$ represents the sum over all the configurations of S. The free energy for a chain of M sites, without the trivial last constant term, the Hamiltonian (7), may be written as

$$\beta F=\sum_{i=0}^{M-1}\sum_{\{S,S'\}} Z_{i,i+1}[S,S']\ln Z_{i,i+1}[S,S']-\beta S\cdot S') + \sum_{i=0}^{M-1}\sum_{\{S\}} Y_i[S](\ln Y_i[S]-\ln A[S]) \tag{A2}$$

where $S\cdot S'=\sum_{j=1}^{N} s_j s_j'$ and the expression (A2) has to be minimized with respect to all the variables $Z_{i,i+1}$ with the restriction that they satisfy (A1) and are properly normalized,

$$\sum_{\{S,S'\}} Z_{i,i+1}[S,S']=1\ . \tag{A3}$$

For the "bulk" phase, with translational invariance the

minimization may be done with respect to a single and symmetric function $Z_b[S,S']=Z_b[S',S]$, and it leads to

$$Z_b[S,S']=C_b p_b[S]p_b[S']\exp(\beta S\cdot S') , \tag{A4}$$

where C_b is a Lagrange multiplier used to satisfy (A3) and we define the auxiliary function, $p_b[S]=\sqrt{Y_b[S]A[S]}$. The Euler-Lagrange equation is transformed in an equation for the one-site probability:

$$Y_b[S]=C_b p_b[S]\sum_{\{S'\}} p_b[S']\exp(\beta S\cdot S') , \tag{A5}$$

which is easily solved by iterations.

To study the surface structure we have to keep each of the functions $Z_{i,i+1}$ as a different variable and to introduce the restriction (A1) with a set of Lagrange multipliers $\Phi_i[S]$. The final equations follow the same scheme as for other lattice surface problems [28],

$$Z_{i,i+1}[S,S']=C_i p_i[S]q_{i+1}[S']\exp(\beta S\cdot S') \tag{A6}$$

with the auxiliary functions

$$p_i[S]=(Y_i[S]A[S]\Phi_i[S])^{1/2} \tag{A7a}$$

and

$$q_i[S]=(Y_i[S]A[S]/\Phi_i[S])^{1/2} . \tag{A7b}$$

The one-site functions $Y_i[S]$ and $\Phi_i[S]$ are given by the solutions of

$$Y_i[S]=\{C_i C_{i+1}\exp(\beta S\cdot S')P_{i-1}[S]Q_{i+1}[S']\}^{1/2} \tag{A8}$$

and

$$\Phi_i[S]=\frac{C_{i-1}P_{i-1}[S]}{C_i Q_{i+1}[S]} \tag{A9}$$

where

$$P_i[S]=\sum_{\{S'\}} p_i[S']\exp(\beta S\cdot S') ,$$

$$Q_i[S]=\sum_{\{S'\}} q_i[S']\exp(\beta S\cdot S') . \tag{A10}$$

The constant C_i are determined by the normalization of (A8). This set of equations (A7)–(A10) is not the most compact representation of the solution but it provides a simple way to solve the problem by iterations.

So far the equations may be applied to any replication landscape represented by the function $A[S]$, but as N grows the sum over the configurations $\{S\}$ soon becomes unfeasible unless we may use some symmetry to reduce it. The simplest case is that studied in Sec. III in which $A[S]$ is only a function of the distance $d=d_h[S,S_0]$ to master sequence S_0. All the functions depend only on that distance and the sums in (A10) may be done as

$$P(d)=\sum_{\{S'\}} p[S']\exp(\beta S\cdot S')=\sum_{d'=0}^{N} p(d')M_{d,d'} , \tag{A11}$$

where the $(N+1)\times(N+1)$ matrix M is given by

$$M_{d,d'}=\sum_{k=\max(0,d+d'-n)}^{\min(d,d')}\frac{d!(N-d)!}{k!(d-k)!(d'-k)!(N-d-d'+k)!}\exp[\beta(4k-2d-2d'+n)] . \tag{A12}$$

Similar expressions may be found for replication landscapes which depend on the distances to several master sequences, but the computational difficulty grows very fast with the number of different sequences.

*Permanent address to be used for correspondence.

[1] M. Eigen, Naturwissenchaften **58**, 465 (1971).
[2] M. Eigen and P. Schuster, Naturwissenchaften **64**, 541 (1977); **65**, 7 (1978); **65**, 341 (1978). These works are also published in M. Eigen and P. Schuster, *The Hypercycle—A Principle of Natural Self-Organization* (Springer-Verlag, Berlin, 1979).
[3] M. Eigen, J. McCaskill, and P. Schuster, J. Phys. Chem. **92**, 6881 (1988); Adv. Chem. Phys. **75**, 149 (1989).
[4] I. Leuthäusser, J. Chem. Phys. **84**, 1884 (1986).
[5] I. Leuthäusser, J. Stat. Phys. **48**, 343 (1987).
[6] P. W. Anderson, Proc. Natl. Acad. Sci. USA **80**, 3386 (1983).
[7] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
[8] W. A. Little, Math. Biosci. **19**, 101 (1974).
[9] Our β is equivalent to $J_y\beta=\beta/2$ in Ref. [5].
[10] L. Demetrius, P. Schuster, and K. Sigmund, Bull. Math. Biol. **47**, 239 (1985).
[11] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, London, 1982).
[12] R. Kikuchi, Phys. Rev. **81**, 988 (1951).
[13] M. Nowak and P. Schuster, J. Theor. Biol. **137**, 375 (1989).
[14] J. Swetina and P. Schuster, Biophys. Chem. **16**, 329 (1982).
[15] D. E. Sullivan and M. M. Telo da Gama, in *Fluid Interfacial Phenomena,* edited by C. A. Croxton (Wiley, New York, 1986); S. Dietriech, in *Phase Transitions and Critical Phenomena,* edited by C. Domb and J. L. Lebowitz (Academic, New York, 1988), Vol. 12.
[16] P. Schuster and J. Swetina, Bull. Math. Biol. **50**, 635 (1988).
[17] D. C. Mattis, Phys. Lett. **56A**, 421 (1976).
[18] C. Biebricher, in *Evolutionary Biology,* edited by M. K. Hecht, B. Wallace, and G. T. Prance (Plenum, New York, 1983).
[19] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).
[20] K. Binder, in *Phase Transition and Critical Phenomena,* edited by C. Domb and J. L. Lebowitz (Academic, New York, 1983), Vol. 8.
[21] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[22] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. (N.Y.) **173**, 30 (1987).

[23] L. Demetrius, J. Chem. Phys. **87**, 6939 (1987).
[24] J. S. McCaskill, J. Chem. Phys. **80**, 5194 (1984).
[25] P. W. Anderson, Phys. Rev. **109**, 1492 (1958).
[26] T. Ikegami and K. Kanedo, Phys. Rev. Lett. **65**, 3352 (1990).
[27] D. S. Rokhsar, P. W. Anderson, and D. L. Stein, J. Mol. Evol. **23**, 119 (1986).
[28] M. M. Telo da Gama, P. Tarazona, M. P. Allen, and R. Evans, Mol. Phys. **71**, 801 (1990).
[29] B. Derrida and L. Peliti, Bull. Math. Biol. **53**, 355 (1991).
[30] M. Serva and L. Peliti, J. Phys. A **24**, L705 (1991); P. G. Higgs and B. Derrida, *ibid.* **24**, L985 (1991).