Quasispecies theory for multiple-peak fitness landscapes

David B. Saakian,[1,2] E. Muñoz,[3] Chin-Kun Hu,[1] and M. W. Deem[3]
[1]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[2]*Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia*
[3]*Department of Physics and Astronomy, Rice University, Houston, Texas 77005-1892, USA*
(Received 15 September 2005; revised manuscript received 13 December 2005; published 11 April 2006)

We use a path integral representation to solve the Eigen and Crow-Kimura molecular evolution models for the case of multiple fitness peaks with arbitrary fitness and degradation functions. In the general case, we find that the solution to these molecular evolution models can be written as the optimum of a fitness function, with constraints enforced by Lagrange multipliers and with a term accounting for the entropy of the spreading population in sequence space. The results for the Eigen model are applied to consider virus or cancer proliferation under the control of drugs or the immune system.

DOI: 10.1103/PhysRevE.73.041913 PACS number(s): 87.23.Kg, 02.50.−r, 87.10.+e, 87.15.Aa

I. INTRODUCTION

Methods of statistical physics have been applied successfully to understand phase transitions of various physical systems in the past few decades [1]. Molecular models of biological evolution also exhibit phase transition behaviors and such models have received much attention in recent decades [2–12]. In particular, the notion of adaptive evolution on a fitness (replication rate) landscape has proven very fruitful [2–5]. In the last decade, several exact results [6–11] have been derived for the Eigen [2,3] and Crow-Kimura [4,6] quasispecies models of biological evolution and their generalizations [12] for a single peak fitness landscape.

However, it is widely accepted that biological evolution proceeds on a rugged fitness landscape [13,14]. In this paper, we consider a multiple peak replication rate landscape as a means to model a rugged fitness landscape. To date, there are few rigorous results for multiple peak fitness landscapes. Such results begin to make the connection with the biologically-relevant case of a rugged fitness landscape. We derive here exact error thresholds by means of a path integral representation for both the Eigen and Crow-Kimura mutation-selection schemes with an arbitrary number of replication rate peaks.

We first generalize the Crow-Kimura model to the multiple peak case. The solution of the one-peak version of this model, where the replication rate is a function of Hamming distance from one configuration, was provided by a path integral representation in [9,10]. We provide here the solution to this model for K peaks, where the replication rate is a function of Hamming distances from K configurations, again by means of a path integral. We find that the mean distances from the peaks maximize the replication rate, with constraints provided by Lagrange multipliers, and with an additional term that represents the entropy of the population in sequence space. Explicit solutions to this maximization task are given for the two-peak case.

We then generalize and solve the continuous-time Eigen model for K peaks, where the replication and degradation rates are functions of Hamming distances from K configurations. A solution of the discrete-time, single-peak Eigen model, which in a sense interpolates between the Crow-Kimura and continuous-time Eigen model [15], was provided in [16]. We solve here the continuous-time Eigen model for K peaks, again by means of a path integral representation. The mean distances from the peaks maximize an excess replication rate with an effective mutation rate, with constraints provided by Lagrange multipliers, and with an additional term that represents the entropy of the population in sequence space under the effective mutation rate.

The Eigen model was first developed to study viral evolution [2], and we use our solution of the two-peak Eigen model to consider viral propagation in the presence of either immune system suppression or an antiviral drug. The preferred viral genome exists at one point in genome space. Conversely, the drug or immune system suppresses the virus most strongly at some other point in genome space. These two points in genome space are the two peaks of the model. The viral growth rate and the suppression rate both decrease with the Hamming distance away from these two unique points.

The rest of the paper is organized as follows: In Sec. II we describe the generalization of the Crow-Kimura, or parallel, model [4] to multiple peaks and provide a solution of this model for an arbitrary replication rate function that depends on distances from K peaks. In Sec. III, we describe the Eigen model and provide a solution for arbitrary replication and degradation rate functions that depend on distances from K peaks. In Sec. IV, we use the Eigen model with two peaks to address the interaction of the immune system with a drug. We consider both adaptable viruses and the original antigenic sin phenomena [17]. We also consider tumor suppression by the immune system. We discuss these results and conclude in Sec. V. We provide a derivation of the path integral representation of the continuous-time Eigen model in the Appendix .

II. CROW-KIMURA MODEL WITH MULTIPLE PEAKS

Here we first briefly introduce the Crow-Kimura model [4] and its quantum spin version [6] so that it is easier to understand its generalizations to be studied in the present paper. In the Crow-Kimura model, any genotype configura-

tion i is specified a sequence of N two-valued spins $s_n=\pm 1$, $1\leqslant n\leqslant N$. We denote such configuration i by $S_i \equiv (s_1^i,\ldots,s_N^i)$. That is, as in [3], we consider $s_n=+1$ to represent the purine (A,G) and $s_n=-1$ to represent the pyrimidine (C,T). Two-values spin models have also been used to study long-range correlations in DNA sequences [18] and DNA unzipping [19,20] and valuable results have been obtained. The difference between two configurations S_i and S_j $\equiv (s_1^j,\ldots,s_N^j)$ is described by the Hamming distance $d_{ij}=(N-\Sigma_n s_n^i s_n^j)/2$, which is the number of different spins between S_i and S_j. The relative frequency p_i of the configuration S_i, $1\leqslant i\leqslant 2^N$, satisfies

$$\frac{dp_i}{dt} = p_i\left(r_i - \sum_{j=1}^{2^N} r_j p_j\right) + \sum_{j=1}^{2^N} \mu_{ij} p_j. \tag{1}$$

Here r_i is the replication rate or the number of offspring per unit period of time (the fitness) of the sequence S_i, and μ_{ij} is the mutation rate to move from sequence S_i to sequence S_j per unit period of time. In the Crow-Kimura model, only single base mutations are allowed: $\mu_{ij}=\gamma\Delta(d_{ij}-1)-N\gamma\Delta(d_{ij})$. Here $\Delta(n)$ is the Kronecker δ function.

The fitness of an organism with a given genotype is specified in the Crow-Kimura model by the choice of the replication rate function r_i, which is a function of the genotype: $r_i=f(S_i)$. It has been observed [6,7] that the system (1), with $r_i\equiv f(s_1^i,\ldots,s_N^i)$ evolves according to a Schrödinger equation in imaginary time with the Hamiltonian

$$-H = \gamma\sum_{n=1}^{N}(\sigma_n^x - 1) + f(\sigma_1^z,\ldots,\sigma_N^z). \tag{2}$$

Here σ^x and σ^z are the Pauli matrices. The mean replication rate, or fitness, of the equilibrium population of genotypes is calculated as (see Reference [3]):

$$\lim_{t\to\infty}\sum_i p_i(t) r_i = \lim_{\beta\to\infty}\frac{1}{\beta}\ln Z \equiv \lim_{\beta\to\infty}\frac{1}{\beta}\ln \mathrm{Tr}\exp(-\beta H).$$

In this way it is possible to find the phase structure and error threshold of the equilibrium population. In the generalized setting, the Crow-Kimura model is often called the parallel model.

A. The parallel model with two peaks

We consider two peaks to be located at two configurations $v_n^1, v_n^2, 1\leqslant n\leqslant N$, where $v_n^1=\pm 1, v_n^2=\pm 1$, and the two configurations have l common spins: $\Sigma_{n=1}^N v_n^1 v_n^2 = 2l-N$. The value of l determines how close the two peaks are in genotype space. Now the replication rate r_i of configuration S_i is a function of the Hamming distances to each peak,

$$r_i = f(2L_1/N-1, 2L_2/N-1), \tag{3}$$

where $\Sigma_{n=1}^N v_n^1 s_n = 2L_1 - N$ and $\Sigma_{n=1}^N v_n^2 s_n = 2L_2 - N$.

Due to the symmetry of the Hamiltonian, the equilibrium frequencies are a function only of the distances from the two peaks: $p_i\equiv p(L_1,L_2)$. We define the factors x_{α_1,α_2} that describe the fraction of spins a configuration S_i has in common with the spins of configurations v^1, v^2. In particular, we define the fraction of spins that are equal to α_k times the value in peak configuration v^k. For K peaks, the general definition is $x_{\alpha_1\cdots\alpha_K}=(1/N)\Sigma_{n=1}^N \delta[s_n,\alpha_1 v_n^1]\cdots\delta[s_n,\alpha_K v_n^K]$. For the two peak case, x_{α_1,α_2} satisfy the relations $x_{++}+x_{+-}+x_{-+}+x_{--}=1$, $x_{++}+x_{+-}=L_1/N$, $x_{++}+x_{-+}=L_2/N$, and $x_{++}+x_{--}=l/N$. Thus these factors are related to the distances from the configuration to each peak and to the distance between the peaks;

$$\begin{aligned}
x_{+-}(L_1,L_2) &= (L_1 - L_2 + N - l)/(2N),\\
x_{++}(L_1,L_2) &= (L_1 + L_2 - N + l)/(2N),\\
x_{--}(L_1,L_2) &= (-L_1 - L_2 + N + l)/(2N),\\
x_{-+}(L_1,L_2) &= (-L_1 + L_2 + N - l)/(2N).
\end{aligned} \tag{4}$$

With these factors, we find the following equation for the total probabliity at a given value of L_1 and L_2, $P(L_1,L_2)$:

$$\begin{aligned}
\frac{dP(L_1,L_2)}{dt} &= f\left(\frac{2L_1}{N}-1,\frac{2L_2}{N}-1\right)P(L_1,L_2) - \gamma N P(L_1,L_2)\\
&+ \gamma\sum_{\alpha_1=\pm1,\alpha_2=\pm1} N x_{\alpha_1,\alpha_2}(L_1+\alpha_1, L_2+\alpha_2)\\
&\times P(L_1+\alpha_1, L_2+\alpha_2) - P(L_1,L_2)\\
&\times \sum_{L_1',L_2'=0}^{N} f\left(\frac{2L_1'}{N}-1,\frac{2L_2'}{N}-1\right)P(L_1',L_2').
\end{aligned} \tag{5}$$

Only the values of L_1 and L_2 satisfying the conditions $0\leqslant L_i\leqslant N$, $|L_1+L_2-N|\leqslant l$, $|L_1-L_2|\leqslant N-l$ are associated with nonzero probabilities. Equation (5) can be solved numerically to find the error threshold and the average Hamming distance of the population to the two peaks. In the next section we solve this equation, and its generalization to K peaks, analytically.

B. Exact solution of the K peak case by a path integral representation

We consider the case of K peaks. We consider the replication rate to depend only on the distances from each peak

$$r_i = f\left(\frac{2L_1}{N}-1,\cdots,\frac{2L_K}{N}-1\right) \equiv N f_0(u_1,\ldots,u_K), \tag{6}$$

where $Nu_k=\Sigma_{n=1}^N v_n^k s_n = 2L_k - N$, $1\leqslant k\leqslant K$. The observable value $\langle u_k\rangle$ is called the surface magnetization [21], or surplus [6], for peak k.

Characterization of the fitness function that depends on K peaks through the K values of u_k requires more than the $K(K-1)/2$ Hamming distances between the peaks. It proves convenient to define the 2^K parameters $y_{\alpha_1\cdots\alpha_K}\equiv y_i, 1\leqslant i\leqslant 2^K$. These are defined by $y_i=(1/N)\Sigma_{n=1}^N\Pi_{k=1}^K\delta(\alpha_{ik},v_n^k)$. Here α_{ik} is the set of indices $\alpha_1\cdots\alpha_K$, and $\alpha_{ik}=\alpha_k$ in the ith set of indices $\alpha_1\cdots\alpha_K$. The introduction of the 2^K parameters y_i is one principle point of this paper.

The Suzuki-Trotter method has been applied in [9,10] to convert the quantum partition function for a single peak

model into a classical functional integral. While calculating $Z \equiv \exp[-\beta H]$, intermediate spin configurations are introduced. We find Z is a functional integral, with the integrand involving a partition function of a spin system in the 2D lattice. In the spin system, there is a nearest-neighbor interaction in horizontal direction and a mean-field-like interaction in the vertical direction. This spin system partition function was evaluated in [9,10] under the assumption that the field values are constant. A path integral representation of the discrete time Eigen model, which is quite similar to the parallel model, was introduced by Peliti [16].

Here we generalize this procedure to K peaks and calculate the time-dependent path integral and Ising partition function. Since the replication rate is a function of K distances, the functional integral is over K fields that represent the K magnetizations. The path integral form of the partition function is

$$Z = \int \mathcal{D}M_k \mathcal{D}H_k \exp\left\{ N \int_0^\beta d\beta' \left[f_0[M_1(\beta'), \ldots, M_K(\beta')] - \sum_{k=1}^{K} H_k(\beta') M_k(\beta') - \gamma \right] + N \sum_{i=1}^{2^K} y_i \ln Q_1 \right\}, \tag{7}$$

where

$$Q_1 = \mathrm{Tr}\, \hat{T} e^{\int_0^\beta d\beta' [\sigma^x \gamma + \sigma^z \sum_{k=1}^{K} \alpha_{ik} H_k(\beta')]}. \tag{8}$$

Here $\beta = t$ is the large time to which Eq. (1) is solved, and the operator $\hat{T}$ denotes time ordering [10], discussed in the Appendix in the context of the Eigen model. Using that N is large, we take the saddle point. Considering $\delta \ln Z / \delta M_k(\beta') = 0$ and $\delta \ln Z / \delta H_k(\beta') = 0$, we find $M_k(\beta')$ and $H_k(\beta')$ independent of β' is a solution. At long time, therefore, the mean replication rate, or fitness, per site becomes

$$\frac{\ln Z}{\beta N} = \max_{M_k, H_k} \left\{ f_0(M_1, \ldots, M_K) - \sum_{k=1}^{K} H_k M_k - \gamma + \sum_{i=1}^{2^K} y_i \left[\gamma^2 + \left(\sum_{k=1}^{K} \alpha_{ik} H_k \right)^2 \right]^{1/2} \right\}. \tag{9}$$

We take the saddle point in H_k to find

$$M_k = \sum_i y_i \alpha_{ik} \frac{\sum_{k'=1}^{K} \alpha_{ik'} H_{k'}}{\sqrt{\gamma^2 + \left(\sum_{k'=1}^{K} \alpha_{ik'} H_{k'} \right)^2}}. \tag{10}$$

We note that the observable, surface magnetization given by $\langle u_k \rangle$, is not directly accessible in the saddle point limit, but is calculable from the mean replication rate [6]. In the one peak case one defines the observable surface magnetization for a monotonic fitness function as follows [7]: one solves the equation $f_0(\langle u \rangle) = (\ln Z)/(\beta N)$. For multiple peaks, we use this same trick, considering a symmetric fitness function and assuming $\langle u_1 \rangle = \langle u_2 \rangle \cdots = \langle u_K \rangle$.

C. Explicit results for the two peak case

For clarity, we write the expression for the case of two peaks. In this case, $y_{++} + y_{--} = (1+m)/2$ and $y_{+-} + y_{-+} = (1-m)/2$, where $m = (2l - N)/N$. We solve Eq. (10) for the fields H_k and put the result into Eq. (9). We find that for a pure phase, the bulk magnetizations maximize the function

$$\frac{\ln Z}{N\beta} = f_0(M_1, M_2) + \frac{\gamma}{2} \sqrt{(1+m)^2 - (M_1 + M_2)^2} + \frac{\gamma}{2} \sqrt{(1-m)^2 - (M_1 - M_2)^2} - \gamma, \tag{11}$$

with the constraints

$$-1 \leqslant M_1 \leqslant 1, \quad -1 \leqslant M_2 \leqslant 1,$$
$$-(1+m) \leqslant M_1 + M_2 \leqslant 1 + m,$$
$$-(1-m) \leqslant M_1 - M_2 \leqslant 1 - m. \tag{12}$$

In the case of a quadratic replication rate, $f_0 = k_1 M_1^2 + k_2 M_2^2 + k_3 M_1 M_2$, Eq. (11) becomes

$$\frac{\ln Z}{N\beta} = k_1 M_1^2 + k_2 M_2^2 + k_3 M_1 M_2 + \frac{\gamma}{2} \sqrt{(1+m)^2 - (M_1 + M_2)^2} + \frac{\gamma}{2} \sqrt{(1-m)^2 - (M_1 - M_2)^2} - \gamma, \tag{13}$$

with the constraints of Eq. (12).

As an example, we consider the replication rate function $f_0 = k(M_1^2 + M_2^2 + M_1 M_2)/2$. When $m \geqslant 0$, and the two peaks are within a Hamming distance of $N/2$ of each other, there is a solution with $M_1 = M_2 = M$ for which

$$\frac{3kM^2}{2\gamma} + \left[\left(\frac{1+m}{2} \right)^2 - M^2 \right]^{1/2} - \frac{1+m}{2} = \frac{k}{2\gamma} (\langle u_1 \rangle^2 + \langle u_2 \rangle^2 + \langle u_1 \rangle \langle u_2 \rangle), \tag{14}$$

where the observable, surface magnetization, is given by $\langle u_i \rangle = \langle 2L_i/N - 1 \rangle$. We find

$$M_1 = M_2 = M = \sqrt{(1+m)^2/4 - \gamma^2/(9k^2)}. \tag{15}$$

We have for the mean replication rate, or fitness, per site

$$\frac{\ln Z}{\beta N} = \frac{3k}{2} \left(\frac{1+m}{2} - \frac{\gamma}{3k} \right)^2, \tag{16}$$

so that [6]

$$\langle u_1 \rangle = \langle u_2 \rangle = \frac{1+m}{2} - \frac{\gamma}{3k}. \tag{17}$$

When $m < 0$, and the two peaks are greater than a Hamming distance of $N/2$ of each other, there is a solution with $M_1 = -M_2 = M$ for which

$$\frac{kM^2}{2\gamma}+\left[\left(\frac{1-m}{2}\right)^2-M^2\right]^{1/2}-\frac{1-m}{2}$$
$$=\frac{k}{2\gamma}(\langle u_1\rangle^2+\langle u_2\rangle^2+\langle u_1\rangle\langle u_2\rangle). \quad (18)$$

One solution is

$$M_1=-M_2=\sqrt{(1-m)^2/4-\gamma^2/k^2}, \quad (19)$$

which gives for a mean replication rate, or fitness, per site

$$\frac{\ln Z}{\beta N}=\frac{k}{2}\left(\frac{1-m}{2}-\frac{\gamma}{k}\right)^2, \quad (20)$$

so that [6]

$$\langle u_1\rangle=-\langle u_2\rangle=\frac{1-m}{2}-\frac{\gamma}{k}. \quad (21)$$

Numerical solution is in agreement with our analytical formulas, as shown in Table I.

TABLE I. Comparison between the analytical formulas Eqs. (17), (21) for the two peak landscape in the parallel model and results from a direct numerical solution of the system of differential equations, Eq. (5), for sequences of length $N=1000$, with $p(L_1,L_2,t=0)=\delta(L_1,N)\delta(L_2,l)$.

m	k	$\langle u_1\rangle$	$\langle u_2\rangle$	$\langle u_1\rangle_{\text{analytic}}$	$\langle u_2\rangle_{\text{analytic}}$
0.93	3.0	0.85	0.85	0.853	0.853
0.93	2.0	0.80	0.80	0.798	0.798
0.7	3.0	0.74	0.74	0.738	0.738
0.7	2.0	0.68	0.68	0.683	0.683
−0.7	3.0	0.52	−0.52	0.517	−0.517
−0.7	2.0	0.35	−0.35	0.35	−0.35
−0.93	3.0	0.63	−0.63	0.631	−0.631
−0.93	2.0	0.46	−0.46	0.465	−0.465

III. EIGEN MODEL WITH MULTIPLE PEAKS

A. Exact solution by a path integral representation

In the case of the Eigen model, the system is defined by means of replication rate functions, r_j, as well as degradation rates, D_j,

$$\frac{dp_i}{dt}=\sum_{j=1}^{2^N}[Q_{ij}r_j-\delta_{ij}D_j]p_j-p_i\left[\sum_{j=1}^{2^N}(r_j-D_j)p_j\right]. \quad (22)$$

Here the frequencies of a given genome, p_i, satisfy $\Sigma_{i=1}^{2^N}p_i=1$. The transition rates are given by $Q_{ij}=q^{N-d_{ij}}(1-q)^{d_{ij}}$, with d_{ij} being the Hamming distance between two genomes S_i and S_j. The parameter $\gamma=N(1-q)$ describes the efficiency of mutations. We take $\gamma=O(1)$. As in Eq. (6), we take the replication and degradation rate to depend only on the spin state, in particular on the Hamming distances from each peak: $r_i=f(S_i)$ and $D_i=D(S_i)$ where

$$f(S)=Nf_0(u_1,\ldots,u_k),\quad D(S)=Nd_0(u_1,\ldots,u_k). \quad (23)$$

We find the path integral representation of the partition function for the Eigen model for the K peak case in the limit of long time as

$$Z=\int \mathcal{D}M_k\mathcal{D}H_k\mathcal{D}m_0\mathcal{D}h_0\exp\left\{N\int_0^\beta d\beta'\left[f_0(M_1,\ldots,M_K)e^{-\gamma(1-m_0)}-h_0m_0-\sum_{k=1}^{K}H_kM_k-d_0(M_1,\ldots,M_K)\right]+N\sum_{i=1}^{2^K}y_i\ln Q_1\right\}, \quad (24)$$

where

$$Q_1=\mathrm{Tr}\,\hat{T}e^{\int_0^\beta d\beta'\left[\sigma^x h_0(\beta')+\sigma^z\Sigma_{k=1}^{K}\alpha_{ik}H_k(\beta')\right]}. \quad (25)$$

The M_k are the values of the magnetization, and γm_0 is an effective mutation rate. This form is derived in the Appendix . Using that N is large, we take the saddle point. As before, we find the mean excess replication rate per site, $f_m=\lim_{t\to\infty}\Sigma_i p_i(t)(r_i-D_i)/N$, from the maximum of the expression for $Z=\mathrm{Tr}\exp(-\beta H)$. We find $Z\sim\exp(\beta N f_m)$, where

$$f_m=f_0(M_1,\ldots,M_K)e^{-\gamma(1-m_0)}-d_0(M_1,\ldots,M_K). \quad (26)$$

Here m_0, M_k are defined through the fields H_k,

$$M_k=\sum_i y_i\alpha_{ik}\frac{\sum_{k'=1}^{K}\alpha_{ik'}H_{k'}}{\sqrt{h_0^2+\left(\sum_{k'=1}^{K}\alpha_{ik'}H_{k'}\right)^2}},$$

$$m_0=\sum_i y_i\frac{h_0}{\sqrt{h_0^2+\left(\sum_{k=1}^{K}\alpha_{ik}H_k\right)^2}}. \quad (27)$$

We define

$$m_i = \frac{\sum_{k=1}^{K} \alpha_{ik} H_k}{\sqrt{h_0^2 + \left(\sum_{k=1}^{K} \alpha_{ik} H_k \right)^2}}. \tag{28}$$

We thus find $m_0 = \Sigma_{i=1}^{2^K} y_i \sqrt{1-m_i^2}$, giving Eq. (26).

B. Simple formulas for the two peak case

In the two peak, $K=2$, case we can define the m_i from Eq. (28) from the system

$$M_1 = \frac{1+m}{2}(m_1+m_2) + \frac{1-m}{2}(m_1-m_2),$$

$$M_2 = \frac{1+m}{2}(m_1+m_2) - \frac{1-m}{2}(m_1-m_2), \tag{29}$$

where m is the overlap between two peaks and we have defined m_1, m_2 in terms of the m_i from Eq. (28) by $m_{++} = -m_{--} = m_1 + m_2$ and $m_{+-} = -m_{-+} = m_1 - m_2$. We have for the mean excess replication rate per site

$$f_m = f_0(M_1, M_2) \exp\left[-\gamma \left(1 - \frac{1+m}{2} \sqrt{1-(M_1+M_2)^2/(1+m)^2} - \frac{1-m}{2} \sqrt{1-(M_1-M_2)^2/(1-m)^2} \right) \right] - d_0(M_1, M_2). \tag{30}$$

C. Eigen model with quadratic replication rate without degradation

We apply our results to model qualitatively the interaction of a virus with a drug. In some situations, one can describe the action of a drug against the virus simply as a one peak Eigen model; that is, the replication rate is a function of the Hamming distance from one peak. The virus may increase its mutation rate, and at some mutation rate there is an error catastrophe [22]. Let us define the critical γ for the replication rate function

$$f_0(M) = \frac{kM^2}{2} + 1. \tag{31}$$

According to our analytical solution, Eq. (26), we consider the maximum of the mean excess replication rate per site,

$$f_m = f_0(M) \exp[-\gamma(1 - \sqrt{1-M^2})],$$

which can also be obtained from Eq. (30) by taking $m=1$ and $M_1 = M_2$. The error catastrophe occurs and leads to a phase with $M=0$ when $k < \gamma$. The error threshold for this quadratic case is the same as in the case of the Crow-Kimura model Eq. (1). The average of u, $\langle u \rangle$, satisfies the equation

TABLE II. Comparison between the analytical formula Eq. (32) for the quadratic landscape (31) in the Eigen model and results from a direct numerical solution of the system of differential equations Eq. (22), for sequences of length $N=4000$, with $p(u, t=0) = \delta(u, 1)$. We set $\gamma = 5$.

k/γ	M	$\langle u \rangle$	$\langle u \rangle_{\text{analytic}}$
1.2	0.24	0.065	0.068
1.4	0.31	0.112	0.113
1.6	0.35	0.146	0.147
1.8	0.38	0.172	0.172
2.0	0.41	0.192	0.193

$$f_0(M) \exp[-\gamma(1 - \sqrt{1-M^2})] = f_0(\langle u \rangle) = \frac{k \langle u \rangle^2}{2} + 1. \tag{32}$$

This equation gives $\langle u \rangle_{\text{analytic}}$ shown in Table II, which are in agreement with numerical solutions.

IV. BIOLOGICAL APPLICATIONS

The Eigen model is commonly used to consider virus or cancer evolution. We here consider an evolving virus or cancer and its control by a drug or the immune system, using the $K=2$, two-peak version of the Eigen model. To model this situation, we consider there to be an optimal genome for virus replication, and we consider the replication rate function $f_0(M_1, M_2)$ to depend only on the Hamming distance of the virus or cancer from this preferred genome, $N(1-M_1)/2$. Conversely, there is another point in genome space that the drug or immune suppresses most strongly, and we consider the degradation rate function $d_0(M_1, M_2)$ to depend only on the Hamming distance from this point, $N(1-M_2)/2$. While each of the functions f_0 and d_0 depends only on one of the two distances, this is multiple-peak problem, because both distances are needed to describe the evolution of the system.

A. Interaction of virus with a drug

We first consider a virus interacting with a drug. We model this situation by the Eigen model with one peak in the replication rate function and one peak in the degradation rate function. The virus replicates most quickly at one point in genome space, with the rate at all other points given by a function that depends on the Hamming distance from this one point. That is, in Eq. (30) we have

$$f_0(M_1, M_2) = \begin{cases} A, & M_1 = 1 \\ 1, & M_1 < 1. \end{cases} \tag{33}$$

At another point in genome space, a drug suppresses the virus most strongly. We consider the case of exponential degradation, a generic and prototypical example of recognition [17],

$$d_0(M_1, M_2) = e^{-b(1-M_2)}. \tag{34}$$

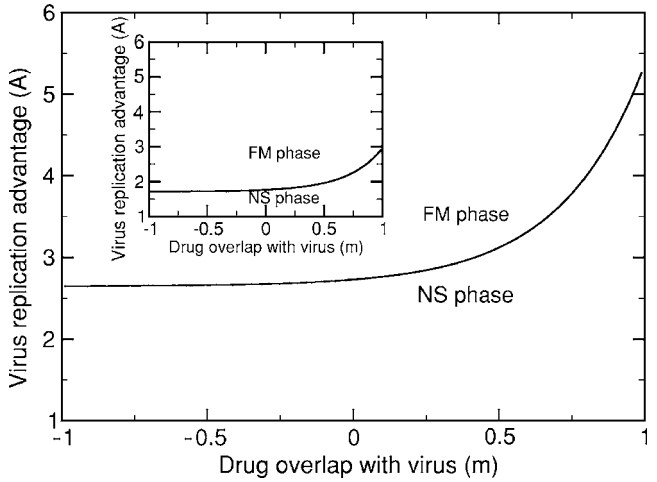

FIG. 1. A phase diagram for the interaction of virus and drug, according to the narrow replication advantage model, Eqs. (34) and (33). We set $b=3.5$ in the exponential degradation function Eq. (34), and $\gamma=1$. As the drug overlaps more with the virus, a higher viral replication advantage is required for the virus to survive. In the NS phase, the drug eliminates the virus. In the inset is shown the phase diagram for the interaction of an adaptable virus and a drug, according to the flat peak replication advantage model, Eqs. (34) and (36). We use $M_0=0.9$ to represent a broad peak for the virus replication rate.

Applying the multiple-peak formalism, we find two phases. There is a selected, ferromagnetic (FM) phase with $M_1=1$, $M_2=m$ and mean excess replication rate per site

$$f_m = Ae^{-\gamma} - \exp[-b(1-m)]. \tag{35}$$

There is also a nonselective (NS) phase, with $M_1<1$. The values of M_1 and M_2 in the NS phase are those which maximize Eq. (30) given the constraints of Eq. (12). The error threshold corresponds to the situation when the mean excess replication rate of the FM and NS phases are equal. The phase diagram as a function of the optimal replication rate of the virus and the distance between the points of optimal virus growth and optimal virus suppression is shown in Fig. 1. The optimal replication rate is A, and the distance between the points of optimal virus growth and optimal virus suppression is $N-l$, where the parameter m is defined as $m=(2l-N)/N$. As the point in genome space at which the drug is most effective moves toward the point in genome space at which the virus grows most rapidly, the virus is more readily eradicated. Alternatively, one can say that as the point in genome space at which the drug is most effective moves toward the point in genome space at which the virus grows most rapidly, a higher replication rate of the virus is required for its survival.

B. Interaction of an adaptable virus with a drug

We now consider a virus that replicates with rate A when the genome is within a given Hamming distance from the optimal genome and with rate 1 otherwise. That is, in Eq. (30) we have

$$f_0(M_1,M_2) = \begin{cases} A, & M_0 \leqslant M_1 \leqslant 1 \\ 1, & -1 \leqslant M_1 < M_0, \end{cases} \tag{36}$$

where $M_0>0$ and is close to 1. We consider the suppression of the virus by the drug as expressed in Eq. (34). There is again a ferromagnetic (FM) phase with a successful selection. In the FM phase, one has $M_0 \leqslant M_1 \leqslant 1$. The evolved values of M_1 and M_2 maximize

$$f_m = A\exp\left[-\gamma\left(1-\frac{1+m}{2}\sqrt{1-(M_1+M_2)^2/(1+m)^2}\right.\right.$$
$$\left.\left.-\frac{1-m}{2}\sqrt{1-(M_1-M_2)^2/(1-m)^2}\right)\right] - d_0(M_1,M_2). \tag{37}$$

There is also a NS phase where the virus has been driven off its advantaged peak, $M_1<M_0$. In this case, one seeks a maximum of Eq. (30) with $f_0=1$ via M_1 and M_2 in the range $-1 \leqslant M_1 \leqslant M_0$, $-1 \leqslant M_2 \leqslant 1$, subject to the constraints of Eq. (12).

A phase diagram for this case is shown in the inset to Fig. 1. The broader range of the virus fitness landscape allows the virus to survive under a greater drug pressure in model Eq. (36) versus Eq. (33). That is, as the drug overlaps more with the favored virus genotypes, the adaptable virus is still able to persist due to the greater range of genotype space available in the FM phase. For such an adaptable virus, a more specific, multidrug cocktail might be required for eradication. A multidrug cocktail provides more suppression in a broader range of genome space, so that the adaptable drug may be eradicated under a broader range of conditions.

C. Original antigenic sin

The immune system acts much like a drug, as a natural protection against death by infection. Prior exposure, such as vaccination, typically increases the immune control of a virus. In some cases, the immune control of a virus is nonmonotonic in the overlap between the vaccine and the virus [17]. This phenomenon is termed the original antigenic sin. To model original antigenic sin, we consider a nonmonotonic degradation function, centered around the second peak, which represents the nonmonotonic behavior of the binding constant, as in our previous model [17]. We fit the binding constant data [17] to a sixth order polynomial in p, where $p=(1-M_2)/2$ is the relative distance between the recognition of the antibody and the virus. The degradation function is shown in the inset in Fig. 2. We consider a single peak virus replication rate, Eq. (33).

There is an interesting phase structure as a function of m. From Eq. (30), we have a FM phase with $M_1=1$, $M_2=m$. We also have a nonselective NS phase, with $M_1<1$, where M_1, M_2 are determined by maximization of Eq. (30) with $f_0=1$ under the constraints of Eq. (12). The phase diagram for typical parameters [17] is shown in Fig. 2. A continuous phase boundary is observed between the FM and NS phases. The virus replication rate required to escape eradication by the adaptive immune system depends on how similar the virus and the vaccine are. When the vaccine is similar to the

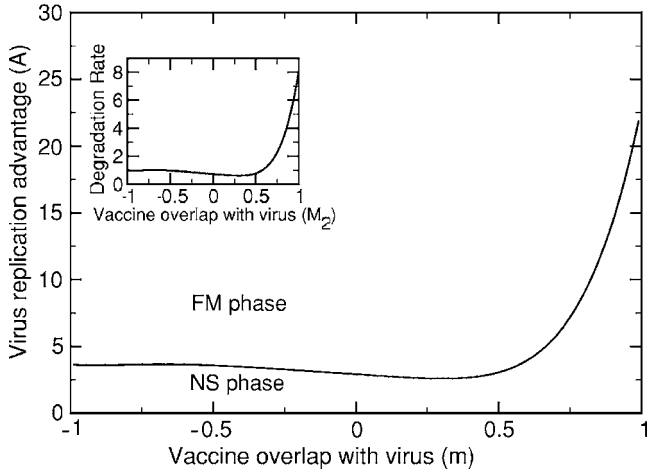

FIG. 2. A phase diagram corresponding to the original antigenic sin model is shown. The degradation function (shown in the inset) is chosen to closely reproduce the binding constant behavior in [17], with a limiting degradation rate of $d_0(M_1, M_2=-1)=1$. We use $\gamma=1$. The virus replication advantage required to escape immune system control is a non-monotonic function of the overlap of the vaccine with the virus.

virus, m near 1, a large virus replication rate is required to escape eradication. This result indicates the typical usefulness of vaccines in protection against and eradication of viruses. When the vaccine is not similar to the virus, $m<0$, the vaccine is not effective, and only a typical virus replication rate is required.

When the vaccine is somewhat similar to, but not identical to, the virus, the replication rate required for virus survival is nonmonotonic. This result is due to the nonmonotonic degradation rate around the vaccine degradation peak. The minimum in the required virus replication rate, $m \approx 0.30$, corresponds to the minimum in the degradation rate, $M_2 \approx 0.30$. The competition between the immune system, vaccine, and virus results in a nontrivial phase transition for the eradication of the virus.

D. Tumor control and proliferation

We consider cancer to be a mutating, replicating object, with a flat replication rate around the first peak, Eqs. (36) and (30). We consider the immune system to be able to eradicate the cancer when the cancer is sufficiently different from self. Thus, the T cells have a constant degradation rate everywhere except near the self, represented by the second peak,

$$d_0(M_1,M_2) = \begin{cases} B, & -1 \leqslant M_2 < M_b \\ 0, & M_b \leqslant M_2 \leqslant 1. \end{cases} \tag{38}$$

To be consistent with the biology, we assume $M_b>0$. We also assume $M_0>1/2$. Typically, also, the Hamming distance between the cancer and the self will be small, m will be positive and near unity, although we do not assume this.

There are four possible selective, ferromagnetic phases. We find the phase boundaries analytically, as a function of $m=(2l-N)/N$. For $mM_0<M_b$, there is a FM4 phase with $M_1=M_0$ and $M_2=mM_0$. The mean excess replication rate per site is $f_m=Ae^{\gamma(\sqrt{1-M_0^2}-1)}-B$. There is a FM3 phase with $M_1=M_0$ and $M_2=M_b$. The mean excess replication rate per site is $f_m=Ae^{\gamma(\sqrt{(1+m)^2-(M_0+M_b)^2}+\sqrt{(1-m)^2-(M_0-M_b)^2}-2)/2}$. This phase is chosen over the FM4 phase when the mean excess replication rate is greater. For $mM_0 \geqslant M_b$ there is a FM2 phase with $M_1=M_0$ and $M_2=mM_0$. The mean excess replication rate per site is $f_m=Ae^{\gamma(\sqrt{1-M_0^2}-1)}$. For $mM_b \geqslant M_0$ there is a FM1 phase with $M_1=mM_b$ and $M_2=M_b$. The mean excess replication rate per site is $f_m=Ae^{\gamma(\sqrt{1-M_b^2}-1)}$.

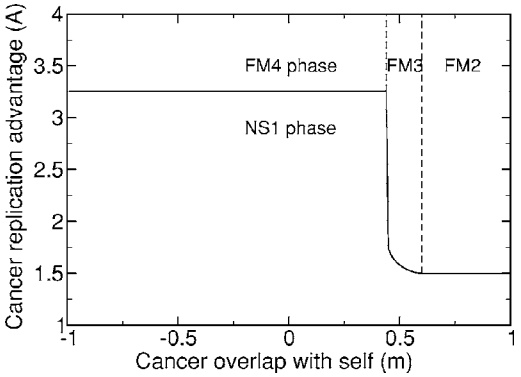

FIG. 3. A phase diagram corresponding to the immune control of cancer is shown. We use $\gamma=1$, $B=1$, $M_0=0.9$, and $M_b=0.54$. The cancer replication advantage required to escape immune system control decreases with the overlap of the cancer with the self. Three of the four selective and one of the two nonselective phases are present for the chosen parameters.

There are two nonselective phases. There is a NS1 phase with $M_1=mM_b$ and $M_2=M_b$. The mean excess replication rate per site is $f_m=e^{\gamma(\sqrt{1-M_b^2}-1)}$. There is a NS2 phase with $M_1=M_2=0$. The mean excess replication rate per site is $f_m=1-B$.

In Fig. 3 is shown the phase diagram for cancer proliferation. According to our previous model [17,23], we choose $(1-M_b)/2=0.23$. We choose $M_0=0.9$ for the width of the advantaged cancer phase. We choose the immune suppression rate as $B=1$. As the cancer becomes more similar to the self, the immune control becomes less effective, and the replication rate required for the cancer to proliferate becomes less. Three of the four selective and one of the two nonselective phases are present for this set of parameters.

V. DISCUSSION AND CONCLUSION

We have used the Eigen model to consider the interaction of a virus or cancer with a drug or the immune system. One can also use the parallel model to represent the replication dynamics of the virus or the cancer. This would be an interesting application of our formalism.

Another application of the formalism would be to consider explicitly the degradation induced by multidrug cocktails. That is, one would consider one peak to represent the preferred virus genome and $K-1$ degradation peaks to represent the $K-1$ drugs. We note that in the general case, the y_i parameters depend on the explicit location of the drug degradation peaks, not simply the distance between them. Re-

sults from this application of the formalism could be quite illuminating as regards the evolution of multidrug resistance.

In conclusion, we have solved two common evolution models with general fitness, or replication and degradation rate, landscapes that depend on the Hamming distances from several fitness peaks. Why is this important? First, we have solved the microscopic models rather than assuming a phenomenological macroscopic model. As is known in statistical mechanics, a phenomenological model may not always detect the fine structure of critical phenomena. Second, approximate or numerical solutions, while useful, do not always explicitly demonstrate the essence of the phenomenon. With analytical solutions, the essence of the phenomenon is transparent. Third, we have derived the first path integral formulation of the Eigen model. This formulation may prove useful in other studies of this model of molecular evolution.

Our results for cancer are a case in point. There are four stable selective phases and two stable nonselective phases. These results may help to shed light on the, at present, poorly understood phenomena of interaction with the immune system, and on why the immune response to cancer and to viruses differs in important ways. These phases could well also be related to the different stages, or grades, through which tumors typically progress.

Our results are a first step toward making the connection with evolution on rugged fitness landscapes, landscapes widely accepted to be accurate depictions of nature. We have applied our solution of these microscopic complex adaptive systems to model four situations in biology: how a virus interacts with a drug, how an adaptable virus interacts with a drug, the problem of original antigenic sin [17], and immune system control of a proliferating cancer.

ACKNOWLEDGMENTS

This work has been supported by the following Grant Nos. CRDF ARP2–2647–YE-05, NSC 94–2112–M-001–014, NSC 94–2119–M-002–001, NSC 94–2811–M001–014, AS–92-TP-A09, DARPA HR00110510057, and 1R90 DK071504–01.

APPENDIX

In this appendix we derive the path integral representation for the solution to the Eigen model. For simplicity, we show the derivation for the $K=1$ case. To our knowledge, this is the first path integral expression representation of the solution to the Eigen model. This path integral representation allows us to make strong analytic progress. We start from the quantum representation of the Eigen model [8]. The Hamiltonian is given by

$$-H=\sum_{l=0}^{N} Ne^{-\gamma}\left(\frac{\gamma}{N}\right)^{l}\sum_{1\leqslant i_1<i_2<\cdots<i_l\leqslant N}\sigma_{i_1}^{x}\sigma_{i_2}^{x}\cdots\sigma_{i_l}^{x}\times f_0(\sigma^z)$$
$$-Nd_0(\sigma^z)\approx Ne^{-\gamma}e^{(\gamma/N)\Sigma_i\sigma_i^x}f_0(\sigma^z)-Nd_0(\sigma^z), \quad \text{(A1)}$$

where we have used the fact that with γ/N small, we need to consider only $l\ll N$ spin flips. The partition function is decomposed by a Trotter factorization,

$$Z=\mathrm{Tr}\,e^{-\beta H}=\mathrm{Tr}\langle S_1|e^{-\beta H/L}|S_2\rangle\langle S_2|e^{-\beta H/L}|S_3\rangle\langle S_L|e^{-\beta H/L}|S_1\rangle. \quad \text{(A2)}$$

Here

$$\langle S_{l-1}|e^{-\beta H/L}|S_l\rangle=\langle S_{l-1}|e^{(\beta N/L)}\left[e^{-\gamma}e^{(\gamma/N)\Sigma_i\sigma_i^x}f_0(\sigma^z)-d_0(\sigma^z)\right]|S_l\rangle$$
$$=\langle S_{l-1}|I+\frac{\beta N}{L}\left[e^{-\gamma}e^{(\gamma/N)\Sigma_i\sigma_i^x}f_0\left(\sum_n s_n^l/N\right)-d_0\left(\sum_n s_n^l/N\right)\right]|S_l\rangle. \quad \text{(A3)}$$

We use the notation $M_l=\Sigma_n s_n^l/N$. We find

$$\alpha_l=\langle S_{l-1}|I+\frac{\beta N}{L}e^{-\gamma}f_0\left(\sum_n s_n^l/N\right)e^{(\gamma/N)\Sigma_i\sigma_i^x}-\frac{\beta N}{L}d_0\left(\sum_n s_n^l/N\right)|S_l\rangle$$
$$=\langle S_{l-1}|S_l\rangle\left[1-\frac{\beta N}{L}d_0(M_l)\right]+\frac{\beta N e^{-\gamma}}{L}f_0(M_l)e^{B\Sigma_n(s_n^{l-1}s_n^l-1)}, \quad \text{(A4)}$$

where $e^{-2B}=\gamma/N$. We thus find

$$\alpha_l=\Delta(d_l)\left[1-\frac{\beta N}{L}d_0(M_l)\right]+\frac{\beta N}{L}e^{-\gamma}f_0(M_l)e^{Bd_l}, \quad \text{(A5)}$$

where $d_l=\Sigma_n(s_n^{l-1}s_n^l-1)$. To represent this in path integral form, we consider

$$\frac{1}{(2\pi)^2}\int dhdmd\psi e^{\Delta tNe^{-\gamma}f_0(M_l)e^{Bm}e^{-\psi m}-\Delta tNd_0(M_l)}\times e^{\psi d-h(m-d)}$$
$$=\frac{1}{2\pi}\int dhdm\left[\delta(d)e^{-\Delta tNd_0(M_l)}+\Delta tNe^{-\gamma}f_0(M_l)\times e^{Bm}\delta(m-d)\right]e^{-h(m-d)}+O[(\Delta t)^2]$$
$$=\int dm\left[\delta(d)\delta(m-d)e^{-\Delta tNd_0(M_l)}+\Delta tNe^{-\gamma}f_0(M_l)\times e^{Bm}\delta(m-d)\delta(m-d)\right]$$
$$=\delta(d)e^{-\Delta tNd_0(M_l)}+\Delta tNe^{-\gamma}f_0(M_l)e^{Bd}\delta(0)$$
$$=\delta(0)\left[\Delta(d)e^{-\Delta Nd_0(M_l)}+\Delta tNe^{-\gamma}f_0(M_l)e^{Bd}\right], \quad \text{(A6)}$$

where $\Delta t=\beta/L$. We note that had we used a Fourier representation of the δ function on the finite domain $[-A/2,A/2]$ instead of the infinite domain $(-\infty,\infty)$, the expression $2\pi\delta(0)$ simply becomes A; moreover, such a finite representation of the δ function is a sufficiently accurate representation of the $\Delta(d_l)$ constraint when $A\gg N$. Ignoring the constant prefactor $\delta(0)$ terms, we can write the full partition function as

$$Z = \mathrm{Tr} \int \mathcal{D}\psi \mathcal{D}h \mathcal{D}m e^{\Sigma_l \psi_l d_l + (\beta N/L)\Sigma_l [-h_l m_l + h_l d_l]}. \quad \text{(A7)}$$

We now introduce the integral representation of the constraint $\delta[(\beta/L)(NM_l - \Sigma_n s_n^l)]$. After rescaling $Bm_l \to m_l$, $h_l \to Bh_l$ we find

$$Z = \mathrm{Tr} \int \mathcal{D}\psi \mathcal{D}h \mathcal{D}m \mathcal{D}H \mathcal{D}M e^{(\beta N/L)\Sigma_l [e^{-\gamma} f_0(M_l) e^{m_l}} \times e^{e^{-\psi_l m_l/B} - d_0(M_l) - h_l m_l - H_l M_l} \times e^{(\beta/L)\Sigma_l H_l \Sigma_n s_n^l + \Sigma_l (\psi_l + \beta N B h_l/L)\Sigma_n (s_n^{l-1} s_n^l - 1)}. \quad \text{(A8)}$$

We note by an expansion of the

$$\exp[(\beta N/L) e^{-\gamma} f_0(M_l) e^{m_l} \exp(-\psi_l m_l/B)] = \sum_{k_l=0}^{\infty} [(\beta N/L) e^{-\gamma} f_0(M_l) e^{m_l}]^{k_l} \exp(-k_l \psi_l m_l/B)/k_l!$$

term in Eq. (A8) to first order in $\beta N/L$ that the integral over ψ_l gives nothing more than $\delta(-k_l m_l/B + d_l)$ for $k_l = 0, 1$. This condition, however, is already enforced by the h_l field when $k_l = 1$ and by the m_l field when $k_l = 0$ if we take as a rule to disallow mutations when $h_l = 0$. We can, thus, remove the integral over ψ, removing the $\delta(0)$ that we anticipated, to find

$$Z = \int \mathcal{D}h \mathcal{D}m \mathcal{D}H \mathcal{D}M e^{(\beta N/L)\Sigma_l [e^{-\gamma} f_0(M_l) e^{m_l} - d_0(M_l)} \times e^{-h_l m_l - H_l M_l} Q, \quad \text{(A9)}$$

where

$$Q = \mathrm{Tr}\, e^{(\beta/L)\Sigma_l H_l \Sigma_n s_n^l + (\beta N B/L)\Sigma_l h_l \Sigma_n (s_n^{l-1} s_n^l - 1)} F. \quad \text{(A10)}$$

Here Q is the partition function of N 1D Ising models of length L. Here F enforces the constraint of disallowing mutations when $h_l = 0$: $F = \Pi_{l=1}^{L} \Delta\{\Delta(h_l)[\Sigma_n (s_n^{l-1} s_n^l - 1)]\}$. We note that $Q = Q_1^N$, where Q_1 is the partition function of one of these models. We are not, at this point, allowed to assume that the H_l or h_l fields are constant over l. Indeed, by Taylor series expanding the first term in Eq. (A9) and integrating over m_l, we find that $h_l = 0$ or $h_l = L/(\beta N)$. For $h_l = 0$, we disallow mutations, as formalized by F. Thus,we can replace e^{-2B} by $\gamma t_l/N$, where $t_l = 0$ if $h_l = 0$, and $t_l = 1$ if $h_l = L/(\beta N)$. We evaluate the partition function Q_1 with an ordered product of transfer matrices. To first order in β/L the matrix at position l is given by $T_l = I + \epsilon_l$ where

$$\epsilon_l = \begin{pmatrix} \frac{\beta H_l}{L} & \frac{\gamma t_l}{N} \\ \frac{\gamma t_l}{N} & -\frac{\beta H_l}{L} \end{pmatrix} = \begin{pmatrix} \frac{\beta H_l}{L} & \frac{\beta \gamma h_l}{L} \\ \frac{\beta \gamma h_l}{L} & -\frac{\beta H_l}{L} \end{pmatrix}. \quad \text{(A11)}$$

We find

$$Q_1 = \mathrm{Tr} \prod_l T_l \sim \mathrm{Tr} \prod_l e^{\epsilon_l}. \quad \text{(A12)}$$

We rescale $h \to h/\gamma$ and $m \to m\gamma$ and take the continuous limit to find

$$Q_1 = \mathrm{Tr}\, \hat{T} e^{\int_0^\beta d\beta' [\sigma^z H(\beta') + \sigma^x h(\beta')]}, \quad \text{(A13)}$$

where the operator $\hat{T}$ indicates (reverse) time ordering, and $\beta' = \beta(L-l)/L$. We find the form of the partition function to be

$$Z = \int \mathcal{D}h \mathcal{D}m \mathcal{D}H \mathcal{D}M e^{N \int_0^\beta d\beta' [e^{-\gamma} f_0(M) e^{\gamma m} - d_0(M) - hm - HM] + N \ln Q_1}. \quad \text{(A14)}$$

Noting the N prefacing the entire term in the exponential, we take the saddle point. We note that

$$\delta Q_1/\delta H(\beta')|_{H(\beta')=H, h(\beta')=h} = (\beta H/\sqrt{H^2+h^2}) 2 \times \sinh(\beta\sqrt{H^2+h^2})$$

and

$$\delta Q_1/\delta h(\beta')|_{H(\beta')=H, h(\beta')=h} = (\beta h/\sqrt{H^2+h^2}) 2 \sinh(\beta\sqrt{H^2+h^2}).$$

We, thus, find a solution of the saddle point condition to be fields H, M, h, m independent of β that maximize

$$\frac{\ln Z}{N} = \beta\left[f_0(M) e^{-\gamma} e^{\gamma m} - d_0(M) - hm - HM\right] + \ln\left[2 \cosh(\beta\sqrt{H^2+h^2})\right], \quad \text{(A15)}$$

when the fields are averaged over a range $\Delta\beta = O(1/N)$ by the saddle point limit. In the limit of large β, we find

$$\frac{\ln Z}{\beta N} = \max_{M,H,m,h} \left[f_0(M) e^{-\gamma} e^{\gamma m} - d_0(M) - hm - HM + (H^2+h^2)^{1/2}\right]. \quad \text{(A16)}$$

One can also derive Eq. (A16) by means of a series expansion in β, a "high temperature" expansion.

The generalization of the path integral representation to the multiple peak Eigen case proceeds as in the parallel case. One introduces K fields for the magnetizations, M^k, and K fields enforcing the constraint, H^k. One also finds in the linear field part of the Ising model the sum $\Sigma_{k=1}^K H_l^k \Sigma_n v_n^k s_n^l$ instead of simply $H_l \Sigma_n s_n^l$. The definition of the y_i and the α_{ik} allows one to rewrite this in the form that leads to Eqs. (7) and (24).

[1] See M. E. Fisher, Rev. Mod. Phys. **70**, 653 (1998); H. E. Stanley, *ibid.* **71**, S358 (1999), and references therein; S.-K. Ma, C. Dasgupta, and C.-K. Hu, Phys. Rev. Lett. **43**, 1434 (1979); C. Dasgupta, S.-K. Ma, and C.-K. Hu, Phys. Rev. B **20**, 3837 (1979); C.-H. Hu, *ibid.* **29**, 5103 (1984); C.-H. Hu, *ibid.* **46**, 6592 (1992); C.-Y. Lin and C.-K. Hu, Phys. Rev. E **58**, 1521 (1998).
[2] M. Eigen, Naturwiss. **58**, 465 (1971).
[3] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[4] J. Crow and M. Kimura, *An Introduction to Population Genetics Theory* (Harper and Row, New York, 1970).
[5] R. E. Lenski, C. Ofria, T. C. Collier, and C. Adami, Nature (London) **400**, 661 (1999).
[6] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997); **79**, 1782 (1997).
[7] E. Baake, M. Baake, and H. Wagner, Phys. Rev. E **57**, 1191 (1998).
[8] D. Saakian and C. K. Hu, Phys. Rev. E **69**, 021913 (2004).
[9] D. B. Saakian and C. K. Hu, Phys. Rev. E **69**, 046121 (2004).
[10] D. B. Saakian, C.-K. Hu, and H. Khachatryan, Phys. Rev. E **70**, 041908 (2004).
[11] D. B. Saakian and C.-K. Hu, Proc. Natl. Acad. Sci. U.S.A. **103**, 4935 (2006).
[12] P. W. Messer, P. F. Arndt, and M. Lassig, Phys. Rev. Lett. **94**, 138103 (2005).
[13] S. Wright, Genetics **16**, 97 (1931).
[14] S. Kauffman and S. Levin, J. Theor. Biol. **128**, 11 (1987).
[15] K. Jain and J. Krug, in *Structural Approaches to Sequence Evolution: Molecules, Networks and Populations*, edited by H. R. U. Bastolla, M. Porto, and M. Vendruscolo (Springer-Verlag, Berlin, 2005); q-bio, PE/0508008.
[16] L. Peliti, Europhys. Lett. **57**, 745 (2002).
[17] M. W. Deem and H. Y. Lee, Phys. Rev. Lett. **91**, 068101 (2003).
[18] C. K. Peng, S. V. Buldyrev, A. L. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Nature (London) **356**, 168 (1992).
[19] D. K. Lubensky and D. R. Nelson, Phys. Rev. Lett. **85**, 1572 (2000).
[20] A. E. Allahverdyan, Z. S. Gevorkian, C.-K. Hu, and M. C. Wu, Phys. Rev. E **69**, 061908 (2004).
[21] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).
[22] M. Eigen, Proc. Natl. Acad. Sci. U.S.A. **99**, 13374 (2002).
[23] J. M. Park and M. W. Deem, Physica A **341**, 455 (2004).