Quasispecies theory for multiple-peak fitness landscapes

David B. Saakian, E. Muñoz, Chin-Kun Hu, and M. W. Deem

1Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan
2Yerevan Physics Institute, Alikhanian Brothers St. 2, Yerevan 375036, Armenia
3Department 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 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-
frequencies are a function only of the distances from the two peaks. In the Crow-Kimura model, only single base mutations are allowed: \( \mu_{ij} = \gamma \Delta(d_{ij} - 1) \). Here \( \Delta(n) \) is the Kronecker \( \delta \) 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 \( r_i = f(\sigma_1^i, \ldots, \sigma_N^i) \) evolves according to a Schrödinger equation in imaginary time with the Hamiltonian

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

Here \( \sigma^i \) and \( \bar{\sigma}^i \) are the Pauli matrices. The mean replication rate, or fitness, of the equilibrium population of genotypes is calculated as (see Reference [3]):

\[
\lim_{t \to \infty} \sum_{i} p_i(t) r_i = \lim_{\beta \to \infty} \frac{1}{\beta} \ln Z = \lim_{\beta \to \infty} \frac{1}{\beta} \ln \text{Tr} \exp(-\beta H).
\]

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 \( v^1, v^2 \), \( 1 \leq n \leq N \), where \( v^i_n = \pm 1, \bar{v}^i_n = \pm 1 \), and the two configurations have \( l \) common spins: \( \sum_{n=1}^{N} v^1_n v^2_n = 2l \). The value of \( l \) determines how close the two peaks are in genotype type. 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 \( \sum_{n=1}^{N} v^1_n = 2L_1 - N \) and \( \sum_{n=1}^{N} \bar{v}^1_n = 2L_2 - N \).

Due to the symmetry of the Hamiltonian, the equilibrium frequencies are a function only of the distances from the two peaks: \( p_i = p(L_1, L_2) \). We define the factors \( x_{\alpha_1, \alpha_2} \) 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 \( \alpha_k \) times the value in peak configuration \( v^k \). For \( K \) peaks, the general definition is \( x_{\alpha_1, \cdots, \alpha_K} = (1/N) \sum_{n=1}^{N} (S^{\alpha_k}_{\alpha_k} v^k_{n}) \). For the two peak case, \( x_{\alpha_1, \alpha_2} \) satisfy the relations \( x_{++, ++} + x_{++, +\bar{v}^1} + x_{++, \bar{v}^1} + x_{++, \bar{v}^2} = 1 \), \( x_{++, v^1} = L_1/N \), \( x_{++, v^2} = L_2/N \), and \( x_{++, v^2} = -L_1/N \). Thus these factors are related to the distances from the configuration to each peak and to the distance between the peaks;

\[
x_{++}(L_1, L_2) = \left( L_1 - L_2 + N - 1 \right)/(2N),
\]

\[
x_{++}(L_1, L_2) = \left( L_1 + L_2 - N + 1 \right)/(2N),
\]

\[
x_{++, 1}(L_1, L_2) = \left( -L_1 - L_2 + N + 1 \right)/(2N),
\]

\[
x_{++, 1}(L_1, L_2) = \left( -L_1 + L_2 + N - 1 \right)/(2N). \tag{4}
\]

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

\[
\frac{dP(L_1, L_2)}{dt} = \gamma \left[ \sum_{\alpha_1=1,2,\alpha_2=1} N x_{\alpha_1,\alpha_2} (L_1 + \alpha_1, L_2 + \alpha_2) \right] P(L_1, L_2)
\]

\[
\times \left[ \sum_{l_1, l_2=0}^{N} f\left( \frac{2L_1}{N} - 1, \frac{2L_2}{N} - 1 \right) \right] P(L_1, L_2). \tag{5}
\]

Only the values of \( L_1 \) and \( L_2 \) satisfying the conditions \( 0 \leq L_1 \leq N, |L_1 + L_2 - N| \leq 1, |L_1 - L_2| \leq N - 1 \) 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

\[
r_i = f\left( \frac{2L_1}{N} - 1, \cdots, \frac{2L_K}{N} - 1 \right) = N f_0(\alpha_1, \cdots, \alpha_K), \tag{6}
\]

where \( N f_0 = \sum_{n=1}^{N} v^k_n = 2L_k - N, 1 \leq k \leq K \). The observable value \( f_0 \) 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 \( f_0 \) requires more than the \( K(K-1)/2 \) Hamming distances between the peaks. It proves convenient to define the \( 2^K \) parameters \( y_{\alpha_1, \cdots, \alpha_K} = c_k, l \leq i \leq 2^K \). These are defined by \( y_i = (1/N) \sum_{n=1}^{N} \delta_{\alpha_k, v^k_n} \). Here \( \alpha_k \) is the set of indices \( \alpha_1, \cdots, \alpha_K \) and \( \alpha_k = 0 \) 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 $Z=\exp[-\beta H]$, 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

$$Z = \int \mathcal{D}M_1 \mathcal{D}H_k \exp \left\{ \sum_{k=1}^{K} H_k(\beta') M_k(\beta') \beta - \gamma \right\},$$

where

$$Q_1 = \operatorname{Tr} \hat{T} e^{i \int_0^\beta d\beta' \left[ f_0[M_1(\beta'), \ldots, M_K(\beta')] \right. \left. - \sum_{k=1}^{K} H_k(\beta') M_k(\beta') - \gamma \right] + \sum_{i=1}^{N} \gamma_i \ln Q_1},$$

with the constraints

$$-1 \leq M_1 \leq 1, \quad -1 \leq M_2 \leq 1,$$

$$-1 \leq N \leq 1 + m, \quad -1 \leq N - M_1 \leq 1 - m, \quad (12)$$

In the case of a quadratic replication rate, $f_0=k_1M_1^2+k_2M_2^2+k_3M_1M_2$, Eq. (11) becomes

$$\ln Z / N \beta = k_1M_1^2 + k_2M_2^2 + k_3M_1M_2 + \frac{\gamma}{2}(1+m)^2 - (M_1 + M_2)^2$$

$$+ \frac{\gamma}{2}(1-m)^2 - (M_1 - M_2)^2 - \gamma, \quad (13)$$

with the constraints of Eq. (12).

As an example, we consider the replication rate function $f_0=k(M_1^2+M_2^2+M_1M_2)/2$. When $m=0$, 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

$$3kM^2 - \left[ \frac{1+m}{2}M^2 + \frac{1+m}{2} \right]^{1/2} = \frac{1+m}{2} \quad (14)$$

where the observable, surface magnetization, is given by $M_1=M_2=M = \sqrt{(1+m)^2/4 - \gamma^2/9k^2}$.

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

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

so that

$$\langle u_1 \rangle = \frac{1+m}{2} - \frac{\gamma}{3k}, \quad (17)$$

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
effective mutation rate. This form is derived in the
Here the frequencies of a given genome, 
which gives for a mean replication rate, or fitness, per site
so that
Numerical solution is in agreement with our analytical for-
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
Here the frequencies of a given genome, \( p_i \), satisfy \( \sum_{i=1}^{N} p_i = 1 \). The transition rates are given by 
with 
\[
Z = \int D\hat{M}_k D\hat{H}_k D\hat{m}_0 D\hat{h}_0 \exp \left\{ N \int_0^\beta d\beta \left[ f_0(M_1, \ldots, M_K) e^{-\gamma(1-m_0)} - h_0 m_0 - \sum_{k=1}^{K} H_k M_k - d_0(M_1, \ldots, M_K) \right] + N \sum_{i=1}^{K} y_i \ln Q_i \right\},
\]
where
\[
Q_i = \langle e^{\frac{1}{\beta} \sum_{j=1}^{N} \sum_{i=1}^{K} \alpha_{ik} \mu H_j} \rangle.
\]
The \( M_k \) are the values of the magnetization, and \( \gamma m_0 \) is an effective mutation rate. This form is derived in the
Using that \( N \) is large, we take the saddle point. As before, we find the mean excess replication rate per site, 
\[
f_m = \lim_{t \to \infty} \frac{1}{N} \sum_{i=1}^{K} p_i(t) (r_i - D_i) / N, \]
from the maximum of the expression for \( Z = \text{Tr} \exp(-\beta H) \). We find \( Z \sim \exp(\beta N f_m) \), where
\[
f_m = f_0(M_1, \ldots, M_K) e^{-\gamma(1-m_0)} - d_0(M_1, \ldots, M_K).
\]
Here \( m_0, M_k \) are defined through the fields \( H_k \),

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\( m \) & \( k \) & \( \langle u_1 \rangle \) & \( \langle u_2 \rangle \) & \( \langle n \rangle_{\text{analytic}} \) \\
\hline
0.93 & 3.0 & 0.85 & 0.85 & 0.853 \\
0.93 & 2.0 & 0.80 & 0.80 & 0.798 \\
0.7 & 3.0 & 0.74 & 0.74 & 0.738 \\
0.7 & 2.0 & 0.68 & 0.68 & 0.683 \\
-0.7 & 3.0 & 0.52 & -0.52 & 0.517 \\
-0.7 & 2.0 & 0.35 & -0.35 & 0.35 \\
-0.93 & 3.0 & 0.63 & -0.63 & 0.631 \\
-0.93 & 2.0 & 0.46 & -0.46 & 0.465 \\
\hline
\end{tabular}
\caption{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 \( p(L_1, L_2, t=0) = \delta(L_1, N) \delta(L_2, l) \).}
\end{table}
\[ m_i = \frac{\sum_{k=1}^{K} \alpha_{ik} H_k}{\sqrt{h_0^2 + \left( \sum_{k=1}^{K} \alpha_{ik} H_k \right)^2}}. \]  

(28)

We thus find \( m_0 = \sum_{i=1}^{K} \gamma_i \sqrt{1 - m_i^2} \), 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

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

\[ M_2 = \frac{1 + m}{2} (m_1 + m_2) - \frac{1 - m}{2} (m_1 - m_2), \]  

(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

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

(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 \( \gamma \) for the replication rate function

\[ f_0(M) = \frac{KM^2}{2} + 1. \]  

(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

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

(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

\[ f_0(M_1,M_2) = \begin{align*}  
A, & \quad M_1 = 1 \\
1, & \quad M_1 < 1. 
\end{align*} \]  

(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],

\[ d_0(M_1,M_2) = e^{-b(1-M_2)}. \]  

(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

\[
f_m = A e^{-\gamma} \exp[-b(1-m)].
\]

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

\[
f_0(M_1, M_2) = \begin{cases} A, & M_0 \leq M_1 \leq 1, \\ 1, & -1 \leq M_1 < M_0, \\ 0, & M_1 < -1, \\ 1, & -1 \leq M_1 \leq M_0, \\ 0, & M_1 > M_0, \\ 1, & -1 \leq M_1 < M_0, \\ 0, & M_1 < -1, \\ 1, & M_1 > M_0, \\ 0, & M_1 < -1, \\ 1, & M_1 > M_0, \\ 0, & M_1 < -1, \\ 1, & M_1 > M_0, \\ 0, & M_1 < -1, \\ 1, & M_1 > M_0, \\ 0, & M_1 < -1, \\ 1, & M_1 > M_0, \\ 0, & M_1 < -1, \\ 1, & M_1 > M_0, \end{cases}
\]

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 \( M_0 \leq M_1 \leq 1 \). The evolved values of \( M_1 \) and \( M_2 \) maximize

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

There is also a NS phase where the virus has been driven off its advantageous 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 \(-1 \leq M_1 \leq M_0, -1 \leq M_2 \leq 1\). We use \( M_0 = 0.9 \) to represent a broad peak for the virus suppression.

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

\[
p = (1 - M_3)/2
\]

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
QUASISPECIES THEORY FOR MULTIPLE-PEAK...

we find the phase boundaries analytically, as a function of \( m \) 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 ineffective, 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, \( m = 0.30 \), corresponds to the minimum in the degradation rate, \( M_2 = 0.30 \). 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,

\[
d_0(M_1, M_2) = \begin{cases} B, & -1 \leq M_2 < M_b \leq 1, M_b \leq M_b \leq 1. \end{cases}
\]

To be consistent with the biology, we assume \( M_b > 0 \). We also assume \( M_b > 1/2 \). 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 \( m M_0 < M_b \), there is a FM4 phase with \( M_1 = M_0 \) and \( M_2 = m M_0 \). The mean excess replication rate per site is \( f_m = A e^{\gamma (1-M_b-1)-B} \). There is a FM3 phase with \( M_1 = M_0 \) and \( M_2 = m M_b \). The mean excess replication rate per site is \( f_m = A e^{\gamma (1+m)-B} \). There is a FM2 phase with \( M_1 = m M_0 \) and \( M_2 = m M_b \). The mean excess replication rate per site is \( f_m = e^{\gamma (1-M_b-1)} \). There is a FM1 phase with \( M_1 = m M_b \) and \( M_2 = m M_0 \). 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-
results 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, N.SC. 94–2112–M-001–014, N.SC. 94–2119–M-002–001, N.SC. 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

$$
H = \sum_{i=0}^{N} Ne^{-\gamma N} \sum_{1 \leq i_1 < i_2 < \cdots < i_N \leq N} \sigma_{i_1}^x \sigma_{i_2}^x \cdots \sigma_{i_N}^x \times f_0(\sigma^x)
$$

$$
- Nd_0(\sigma^x) = Ne^{-\gamma N} \sum_{1 \leq i_1 < \cdots < i_N \leq N} \sigma_{i_1}^x \sigma_{i_2}^x \cdots \sigma_{i_N}^x \times f_0(\sigma^x) - Nd_0(\sigma^x),
$$

(A1)

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

$$
Z = \text{Tr} e^{-BH} = \text{Tr}(S_1 | e^{-BH/2} S_2 | e^{-BH/2} | S_3 | e^{-BH/2} | S_4 | e^{-BH/2} | S_1).$

(A2)

Here

$$
\langle S_{t-1} | e^{-BH} | S_t \rangle = \langle S_{t-1} | e^{H(t-1)} e^{-\gamma N} \sum_{i} d_0(\sigma^x) - d_0(\sigma^y) | S_t \rangle
$$

$$
= \langle S_{t-1} | I + \frac{\beta N}{L} e^{-\gamma N} \sum_{i} d_0(\sigma^x) \left( \sum_n s_{i}^x/N \right) - d_0(\sum_n s_{i}^x/N) | S_t \rangle.
$$

(A3)

We use the notation $M_t = \sum_i s_{i}^x/N$. We find

$$
\alpha_t = \langle S_{t-1} | I + \frac{\beta N}{L} e^{-\gamma N} \sum_{i} s_{i}^x/N \sigma_i \sigma_i^z - d_0(\sum_n s_{i}^x/N) | S_t \rangle
$$

$$
= \langle S_{t-1} | S_t \left[ 1 - d_0(M_t) \right] + \frac{\beta N e^{-\gamma N}}{L} f_0(M_t) e^{\beta N s_{i}^x(s_{i}^x-1)/2} - d_0(M_t) \rangle,
$$

(A4)

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

$$
\alpha_t = \Delta(d) \left[ 1 - d_0(M_t) \right] + \frac{\beta N e^{-\gamma N}}{L} f_0(M_t) e^{\beta d_0(1-d_0/2)},
$$

(A5)

where $d_0 = \sum_i (s_{i}^x-1)$. To represent this in path integral form, we consider

$$
\frac{1}{(2\pi)^2} \int dhdm \text{det} \left( \begin{array}{c} \Delta Ne^{-\gamma N} f_0(M_t) e^{\beta m} e^{-\Delta t d_0(M_t)} \\ \Delta Ne^{-\gamma N} f_0(M_t) e^{\beta m} e^{-\Delta t d_0(M_t)} \end{array} \right) \times e^{i d_0(m-d)}
$$

$$
= \frac{1}{2\pi} \int dhdm \text{det} \left( \begin{array}{c} \Delta Ne^{-\gamma N} f_0(M_t) e^{\beta m} e^{-\Delta t d_0(M_t)} \\ \Delta Ne^{-\gamma N} f_0(M_t) e^{\beta m} e^{-\Delta t d_0(M_t)} \end{array} \right)
$$

$$
= \int dm \text{det} \left( \begin{array}{c} \Delta Ne^{-\gamma N} f_0(M_t) e^{\beta m} e^{-\Delta t d_0(M_t)} \\ \Delta Ne^{-\gamma N} f_0(M_t) e^{\beta m} e^{-\Delta t d_0(M_t)} \end{array} \right)
$$

$$
= \Delta(d) e^{-\Delta t d_0(M_t)} + \Delta t Ne^{-\gamma N} f_0(M_t) e^{\beta d_0(1-d_0/2)},
$$

(A6)

where $\Delta t = \beta/L$. We note that had we used a Fourier representation of the $\delta$ 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 $\delta$ function is a sufficiently accurate representation of the $\Delta(d)$ constraint when $A \gg N$. Ignoring the constant prefactor $\delta(0)$ terms, we can write the full partition function as
We now introduce the integral representation of the constraint $\delta((\beta/L)(NM_l-\sum s_{m_l}^i))$. After rescaling $Bm_l \rightarrow m_l$, $h_l \rightarrow Bh_l$ we find

$$Z = \text{Tr} \int D\psi Dh Dm e^{N[(\beta/L)\sum_i (\psi_{L_l} + (\beta Nh_l) s_{m_l})] - Nm H}$$

(A7)

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

$$Q_1 = \text{Tr} e^{\int [d\beta' \phi(\beta')]} Q$$

(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

$$Z = \int Dh Dm DHDMe^{N\int [d\beta' \phi(\beta')]} e^{\sum_{l \neq 0} (\beta Nh_l s_{m_l}) - d_0(M) - d_0(M) - HMM} \ln Q_1.$$  

(A14)

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

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

and

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

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

$$\frac{\ln Z}{N} = \beta \int (f_0(M) e^{-x e^{ym} - d_0(M) - hm - HM})$$

$$+ \ln \left[ 2 \cosh(\beta \sqrt{H^2 + h^2}) \right].$$

(A15)

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

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

(A16)

One can also derive Eq. (A16) by means of a series expansion in $\beta$, 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, $M^K_t$, and $K$ fields enforcing the constraint, $H^K$. One also finds in the linear field part of the Ising model the sum $\sum_{k=1}^K H^K s_{m_l} e_{m_l}^k$ instead of simply $H s_{m_l} e_{m_l}$. The definition of the $y_i$ and the $\alpha_{ik}$ allows one to rewrite this in the form that leads to Eqs. (7) and (24).


