Host-parasite coevolution and optimal mutation rates for semiconservative quasispecies

Yisroel Brumer and Eugene I. Shakhnovich
Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA
(Received 20 January 2004; published 4 June 2004)

In this paper, we extend a model of host-parasite coevolution to incorporate the semiconservative nature of DNA replication for both the host and the parasite. We find that the optimal mutation rate for the semiconservative and conservative hosts converge for realistic genome lengths, thus maintaining the admirable agreement between theory and experiment found previously for the conservative model and justifying the conservative approximation in some cases. We demonstrate that, while the optimal mutation rate for a conservative and semiconservative parasite interacting with a given immune system is similar to that of a conservative parasite, the properties away from this optimum differ significantly. We suspect that this difference, coupled with the requirement that a parasite optimize survival in a range of viable hosts, may help explain why semiconservative viruses are known to have significantly lower mutation rates than their conservative counterparts.

DOI: 10.1103/PhysRevE.69.061909 PACS number(s): 87.14.Gg, 87.23.Kg, 87.10.+e

I. INTRODUCTION

Introduced over 30 years ago, the quasispecies model of evolution [1,2] has provided an invaluable tool for the study of complex evolutionary behaviors. In the model, a fitness landscape is introduced, which accounts, often in a highly approximate manner, for the complex interplay between genotype, phenotype, and environment by assigning a relative fitness for each genomic sequence (and thus associating phenotype with genotype, an approximation that must be treated with care). Through the consideration of numerous individually mutating copies of a genome, evolutionary systems can be studied analytically and numerically on these fitness landscapes, which has provided enormous insight into the process of evolution and the nature of mutation rates in real biological systems. In particular, it was found that a phase transition (known as the "error catastrophe") occurs as the mutation rate increases, and a marked crossover can be observed from the existence of a quasispecies (wherein most individuals in the population contain genomes close to a fitness peak) to a near-random walk in genome space with no discernible quasispecies present [2].

The vast majority of the literature on the quasispecies model involve studies of asymptotic behavior on numerous stationary landscapes [3–6]. This corresponds to a situation where static environmental conditions are considered to be the dominant evolutionary pressure on a species. However, this picture fails to describe the cornucopia of evolutionary pressures in nature. Many organisms, parasites, survive through the detrimental use of host biochemical processes. The parasite requires the host to live. The host survives better if it can avoid or destroy the parasite, providing an intriguing scenario: the host must evolve to defeat the parasite and the parasite must evolve to evade the host's defenses. This creates a nonlinear feedback cycle as both species scour a time-dependent fitness landscape that changes as the other species mutates.

Parasites are ubiquitous in nature, ranging from the microscopic (e.g., viruses, bacteria, protozoa) to fungi, helminths, and arthropods. The interaction between parasites and hosts is very complex, with parasites exhibiting multistage life cycles, inert phases, and the use of multiple intermediate hosts, while hosts employ a wide variety of behavioral and immune defenses. This ongoing struggle has been well documented in mammals, birds, fish, bacteria, and other organisms.

Recent work on time-dependent quasispecies landscapes [7,8] has allowed for the study of a simple model of coevolution by Kamp and Bornholdt [9,10], discussed in detail in Sec. III. They derived a parameter-independent expression for the optimal mutation rate for a host genome, which compared admirably with experimental results on B-cell mutation rates [9]. An expression was also derived for optimal viral mutation rates [10] which, although dependent on the parameters of the model, explained numerous phenomena including the constancy of mutation rates within a viral class. However, this model considers the interaction only between a conservatively replicating parasite and host.

In its conservative formulation, the quasispecies model considers single stranded genomes that produce multiple copies of itself, each possessing a set of point mutations, while the original genome is conserved. While this model is obviously applicable to numerous RNA-based viruses, the vast majority of organisms, including many viruses and other parasites, store genetic information in double stranded DNA. DNA replicates semiconservatively through a series of steps discussed in Section II. In a recent work, Tannenbaum *et al.* [11] reformulated the quasispecies model to accurately represent semiconservative systems, which were found to display fundamentally different behavior than conservative systems with respect to the error catastrophe in the infinite time limit on a static landscape. Thus, to properly model the coevolution of a parasite and its host, the host system must replicate semiconservatively, while the parasite can be modeled as either conservative, as in the case of many riboviruses, or semiconservative, as by many lysogenic double stranded DNA viruses or higher parasites. Retroviruses, such as HIV, likely display characteristics of both modes of replication, as do immune systems that undergo somatic hypermutation.

In this paper, we extend Kamp and Bornholdt's model of coevolution to the case of a semiconservative host interact-

ing with either a conservative or semiconservative parasite. We consider the optimal behavior for both the host and parasite, and demonstrate the similarities and differences between the conservative and semiconservative models.

The paper is organized as follows: in Sec. II we present the quasispecies model and its extension to semiconservative replication. In Sec. III we discuss the model of host-parasite coevolution for both conservative and semiconservative organisms. Section IV presents the results and discussion and Sec. V presents our conclusions.

II. THE QUASISPECIES MODEL

In this section, we present some necessary background on the conservative and semiconservative quasispecies models for the purpose of a self-contained discussion. Greater detail may be found in the original papers.

A. Conservative replication

The quasispecies model studies the evolution of a population of organisms, each with a genome $\phi=s_1s_2\cdots s_n$, where each s_i represents a "letter" chosen from an alphabet of size S. Often, S is chosen to be two to model the pyrimidine and purine groups or four to model the nucleotides. Assuming first-order growth kinetics and associating phenotype with genotype (i.e., that the growth rate of an individual is directly determined by ϕ), it can be shown that

$$\frac{dx_\phi}{dt}=\sum_{\phi'}A(\phi')W(\phi,\phi')x_{\phi'}-f(t)x_\phi, \tag{1}$$

where x_ϕ denotes the fraction of the population with genome ϕ, $A(\phi)$ represents the fitness, or growth rate, of sequence ϕ, $W(\phi,\phi')$ is the likelihood of creating sequence ϕ from ϕ' by mutations, and $f(t)=\Sigma_\phi A(\phi)x_\phi$ is the average fitness of the population, holding the population size constant and introducing competition. If only point mutations are allowed and a genome-independent mutation probability ϵ is assumed, then $W(\phi,\phi')$ can be written in terms of the genome length n and the number of bases at which ϕ and ϕ' differ, the Hamming distance $HD(\phi,\phi')$, as

$$W(\phi,\phi')=\left(\frac{\epsilon}{S-1}\right)^{HD(\phi,\phi')}(1-\epsilon)^{n-HD(\phi,\phi')}. \tag{2}$$

These equations can be greatly simplified in the case of a single fitness peak landscape, where a master sequence, ϕ_0, has a fitness much greater than all other sequences. The rest of the genomes are assumed to be equally fit, which can be described by the growth rates

$$A(\phi)=\begin{cases}\eta, & \phi\neq\phi_0\\ \sigma\gg\eta, & \phi=\phi_0.\end{cases} \tag{3}$$

The sequences can then be grouped into Hamming classes based on their distance from the master sequence by defining

$$w_l=\sum_{\phi\in\{\phi|HD(\phi,\phi_0)=l\}}x_\phi \tag{4}$$

and

$$A(l)\equiv A(\phi)\quad \phi\in\{\phi|HD(\phi,\phi_0)=l\}. \tag{5}$$

This reduces the problem from S^n dimensions to $n+1$ dimensions. If mutations that lead from higher to lower Hamming distances are ignored (an approximation that becomes exact as $n\to\infty$),

$$\frac{dw_l}{dt}=\sum_{l'=0}^{l}\frac{(n-l')!}{(n-l)!}A(l')(\epsilon)^{l-l'}(1-\epsilon)^{n-(l-l')}w_{l'}-f(t)w_l, \tag{6}$$

where $f(t)=\Sigma_l A(l)w_l=\sigma w_0+\eta(1-w_0)=(\sigma-\eta)w_0+\eta$. Defining $y_i=w_i\exp[\int_0^t f(s)ds]$ removes the nonlinearity in these equations and the linear set of differential equations can be solved for any Hamming class. The solution for the master sequence is

$$y_0(t)=y_0(0)e^{q^n\sigma t} \tag{7}$$

and, for the first Hamming class,

$$y_1(t)=y_0(0)n\left(\frac{(e^{q^n\sigma t}-e^{q^n\eta t})(1-q)\sigma}{(\sigma-\eta)q}\right), \tag{8}$$

where $q=1-\epsilon$, a definition we shall use throughout the paper.

B. Semiconservative replication

In order to properly model a semiconservative system, a double stranded molecule generated from an alphabet of size S must be considered, where each letter i uniquely pairs with $(i+S/2)\mathrm{mod}\,S$. DNA requires $S=4$, where the letters can be assigned as $A\equiv1, G\equiv2, T\equiv3, C\equiv4$. A single DNA molecule of length n consists of a strand $\phi=s_1s_2\cdots s_n$ and a complementary strand $\underline{\phi}=\underline{s}_1\underline{s}_2\cdots\underline{s}_n$ where $\underline{s}_i$ denotes the complement of s_i. Hence, each DNA molecule may be represented by the pair $\{\phi,\underline{\phi}\}\equiv\{\underline{\phi},\phi\}$.

When a semiconservative molecule replicates, it undergoes a three step process shown schematically in Fig. 1. First, each genome $\{\phi,\underline{\phi}\}$ unzips to form two single stranded genomes, ϕ and $\underline{\phi}$. Each strand is then copied to produce two new pairs, $\{\phi,\underline{\phi}'\}$ and $\{\underline{\phi},\phi'\}$, where the primes denote the fact that the two fresh strands may contain replication errors. At this point, proofreading mechanisms can distinguish between the new and old strands and may fix all or some of the replication errors, which can be spotted by the fact that $s_i'\neq\underline{s}_i$. All of these repair mechanisms are included in the base-independent error probability ϵ. In the last step, the new and old strands become indistinguishable. Various maintenance enzymes repair the remaining mismatches, but cannot determine which of the strands ϕ and $\underline{\phi}'$ is the newly replicated strand. Hence, the repair is made in the new strand with 50% probability and in the old strand with 50% probability. The final result is that the original strand $\{\phi,\underline{\phi}\}$ is replicated to create two new strands, $\{\phi'',\phi''\}$ and $\{\phi''',\underline{\phi}'''\}$ [12].

The quasispecies equations for this system can be written as [11]

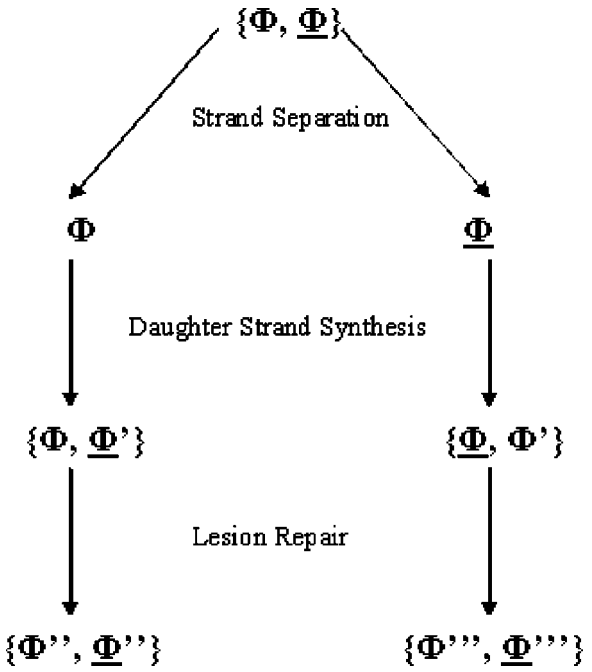

FIG. 1. A schematic model of DNA replication. The original, double stranded genome unzips to create two single stranded genomes. Each of these is copied to produce two new complementary strands. Methyl-directed and post-methylation DNA repair keep the effective error rate low. Adapted from Tannenbaum *et al.* [11].

$$\frac{dx_{\{\phi,\underline{\phi}\}}}{dt} = \sum_{\{\phi',\underline{\phi}'\}} A(\{\phi',\underline{\phi}'\})x_{\{\phi',\underline{\phi}'\}}[p(\phi',\{\phi,\underline{\phi}\}) + p(\underline{\phi}',\{\phi,\underline{\phi}\})] - [A(\{\phi,\underline{\phi}\}) + f(t)]x_{\{\phi,\underline{\phi}\}}, \quad (9)$$

where $f(t)=\Sigma_{\phi}A(\{\phi,\underline{\phi}\})x_{\{\phi,\underline{\phi}\}}$ and $p(\phi',\{\phi,\underline{\phi}\})$ represents the probability that the unzipped strand ϕ' will produce the pair $\{\phi,\underline{\phi}\}$. To make these equations more useful, we can define $A(\phi)\equiv A(\phi,\underline{\phi})$ and $x_{\phi}\equiv\frac{1}{2}x_{\{\phi,\underline{\phi}\}}$ if $\phi\neq\underline{\phi}$ and $x_{\phi}\equiv x_{\{\phi,\underline{\phi}\}}$ if $\phi=\underline{\phi}$. After some manipulation, we obtain

$$\frac{dx_{\phi}}{dt} = 2\sum_{\phi'} A(\phi')x_{\phi'}\left(\frac{\epsilon}{2}\right)^{HD(\phi,\phi')}\left(1-\frac{\epsilon}{2}\right)^{n-HD(\phi,\phi')} - [A(\phi) + f(t)]x_{\phi}, \quad (10)$$

where $f(t)=\Sigma_{\phi}A(\phi)x_{\phi}$. This differs from Eq. (1) by a change in $W(\phi,\phi')$ to reflect the unzipping and repair properties of the genome, and the additional term $-A(\phi)x_{\phi}$, which represents the destruction of the initial genome.

We now turn our attention to semiconservative replication on a single fitness peak landscape. This case is more complicated than for a conservative system, since viability genes often exist on both strands in nature. Hence, if there exists a sequence ϕ_0 with fitness σ, it stands to reason that the sequence $\underline{\phi}_0$ should have fitness σ as well, effectively creating a double fitness peak landscape (this assumption is by no means fundamental to the work). However, noting that $x_{\phi}=x_{\underline{\phi}}$ for all times, both by definition and by conservation in Eq. (10), this difficulty can be sidestepped. As long as n is not too small, the area around each fitness peak can be locally treated as a single fitness peak landscape as the two peaks are distant in sequence space. Hence, ignoring back mutations, the two master sequences obey the equations

$$\frac{dw_0}{dt} = 2(1-\epsilon/2)^n\sigma\underline{w}_0 - [\sigma+f(t)]w_0 = 2(1-\epsilon/2)^n\sigma w_0 - [\sigma+f(t)]w_0, \quad (11)$$

$$\frac{d\underline{w}_0}{dt} = 2(1-\epsilon/2)^n\sigma w_0 - [\sigma+f(t)]\underline{w}_0 = 2(1-\epsilon/2)^n\sigma\underline{w}_0 - [\sigma+f(t)]\underline{w}_0, \quad (12)$$

where w_i represents the concentration of the ith Hamming class as before. Therefore, we can redefine the concentration of the master sequence to include both w_0 and $\underline{w}_0$ and use Eq. (11) for the sum of the two. While this is not strictly necessary and has no effect on the results, it does reduce the bookkeeping, and the characteristics of the individual peaks can be obtained by simply dividing by two. A similar procedure yields

$$\frac{dw_1}{dt} = 2(1-\epsilon/2)^{n-1}\left(\frac{\epsilon}{2}\right)n\sigma w_0 + 2(1-\epsilon/2)^n\eta w_1 - [\eta+f(t)]w_1, \quad (13)$$

where we include sequences of Hamming distance one away from both master sequences. The definition $y_i = w_i\exp[\int_0^t f(s)ds]$ once again removes the nonlinearity. The solutions for the first two Hamming classes are

$$y_0(t) = y_0(0)e^{2\sigma(1-\epsilon/2)^n-\sigma}, \quad (14)$$

$$y_1(t) = y_0(0)n\left(\frac{\sigma\epsilon(1-\epsilon/2)^{n-1}}{(S-1)(\sigma-\eta)[2(1-\epsilon/2)^n-1]}\right) \times (e^{\sigma[2(1-\epsilon/2)^n-1]t} - e^{\eta[2(1-\epsilon/2)^n-1]t}). \quad (15)$$

III. HOST-PARASITE CO-EVOLUTION

Historically, the main focus of research on the quasi species model has related to static and equilibrium properties of the system [5,3,13–17]. A number of recent works, however, have explored the dynamics of the system under various conditions [7,8,18,19], which has allowed the study of the simple model of coevolution described here. Following the work of Kamp and Bornholdt [9,10], we envision a population of host and parasite organisms (which we shall refer to as the immune system and virus), each described by a set of quasispecies equations. Ignoring the interspecies interaction, the immune and viral genomes, of length n_{is} and n_v, respectively, evolve independently on a single fitness peak landscape, where the master sequences have fitness $\sigma_{is}\gg\eta_{is}$ and $\sigma_v\gg\eta_v$. To model the deleterious effect of the immune system on the virus, the dominant immune genome imposes a large death rate δ on the corresponding viral sequence. If this dominant immune genome matches the viral master se-

quence, the viral fitness peak will move to an arbitrary sequence of the first Hamming class. The viral quasispecies then adapts to this new fitness peak on a timescale τ_v, the time required for the population of the new master sequence to overtake that of the old. At this point, the immune system fitness peak adjusts to match the new viral peak, and adapts on a similarly defined timescale τ_{is}. Thus, through the iteration of these steps, the viral fitness peak scours sequence space in an attempt to avoid the immune system, which follows on its heels. Applying recent results on dynamic fitness landscapes [7], regions of stability can be defined for both the viral and immune quasispecies by determining a characteristic timescale for regrowth of a new master sequence. If the landscape moves slowly enough, the master sequence has time to regenerate to the master sequence concentrations reached before the peak shift and the species will survive for all time. If, however, the master sequence cannot regenerate rapidly enough, a second peak shift will occur before the new master sequence reaches the concentration held by the old master sequence before the first shift. The third master sequence cannot reach the levels of the second, and this continues until, eventually, there is no discernible master sequence in the population. For the conservative case, this can be stated rigorously by comparing the growth of a single member of the first Hamming class described by Eq. (8) with $e^{\eta\tau}$, the uninhibited growth of a random sequence far from the fitness peak (as mutations in and out of this sequence should cancel). Using Eq. (8) this ratio can be defined, for both the immune and viral quasispecies, as [8,9]

$$\kappa \equiv \frac{w_1(\tau)}{ne^{\eta\tau}w_0(0)(S-1)} \equiv \left(\frac{(e^{(q^n\sigma-\eta)\tau}-e^{(q^n\eta-\eta)\tau})(1-q)\sigma}{(S-1)(\sigma-\eta)q}\right), \tag{16}$$

where τ is the lag time between peak shifts and the parameters $\{q,\sigma,\cdots\}$ represent the parameters for either species. The quasispecies survives only when $\kappa \geqslant 1$.

The last piece necessary to complete the coevolution model, then, is the speed with which the landscape moves. By the definition of our model, τ is the sum of the time required for the regeneration of the virus, τ_v, plus the time required for the regeneration of the immune system, τ_{is}. Hence, we must solve for $\tau=\tau_{is}+\tau_v$, where

$$e^{(q_v^n\eta_v-\delta)\tau_v}w_{0,v}(\tau) = e^{(q_v^n\sigma_v\tau_v)}\frac{w_{1,v}(\tau)}{n(S-1)}, \tag{17}$$

$$e^{q_{is}^n\eta_{is}\tau_{is}}w_{0,is}(\tau) = e^{(q_{is}^n\sigma_{is}\tau_{is})}\frac{w_{1,s}(\tau)}{n(S-1)}. \tag{18}$$

This can be solved to obtain

$$e^{(q_v^n\eta_v-\delta)\tau_v}e^{q_v^n\sigma_v\tau} = e^{q_v^n\sigma_v\tau_v}\frac{(e^{q_v^n\sigma_v\tau}-e^{q_v^n\eta_v\tau})(1-q_v)\sigma_v}{(S-1)(\sigma_v-\eta_v)q_v}, \tag{19}$$

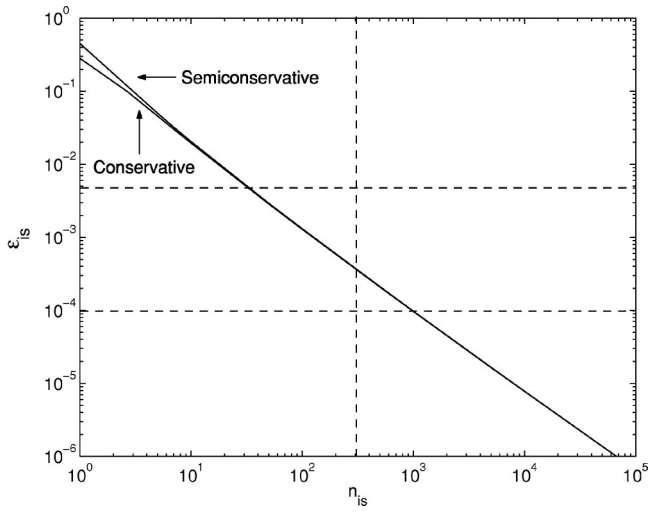

FIG. 2. Optimal immune system mutation rate vs n_{is}. The dashed lines represent experimental values for somatic hypermutation of B-cell complementary determining regions, adapted from Ref. [9].

$$e^{q_{is}^n\eta_{is}\tau_{is}}e^{q_{is}^n\sigma_{is}\tau} = e^{q_{is}^n\sigma_{is}\tau_{is}}\frac{(e^{q_{is}^n\sigma_{is}\tau}-e^{q_{is}^n\eta_{is}\tau})(1-q_{is})\sigma_{is}}{(S-1)(\sigma_{is}-\eta_{is})q_{is}}, \tag{20}$$

which yields, with the reasonable approximations that $q\approx 1$ and $\sigma \gg \eta$ (the latter of which is used throughout the paper),

$$\tau_v \cong -\frac{\ln\left(\frac{1-q_v}{S-1}\right)}{q_v^n(\sigma_v-\eta_v)+\delta}, \tag{21}$$

These equations can be applied to determine the optimal mutation rate for both the host and the parasite. The host can minimize the region of viability for the parasite by evolving a mutation rate such that

$$\frac{\partial \kappa_v}{\partial \epsilon_{is}} = 0, \tag{23}$$

yielding [9]

$$\epsilon_{is} - 1 - n_{is}\ln\left(\frac{\epsilon_{is}}{S-1}\right) = 0. \tag{24}$$

This equation has the nice quality of being independent of the parameters of the immune model, as well as the properties of the virus. The solution to this equation is shown in Fig. 2 and compared to the experimentally verified mutation rate for human B-cell receptors. This is discussed at length in Sec. IV.

Optimizing the viral mutation rate requires solving for

$$\frac{\partial \kappa_v}{\partial \epsilon_v} = 0, \tag{25}$$

or [10]

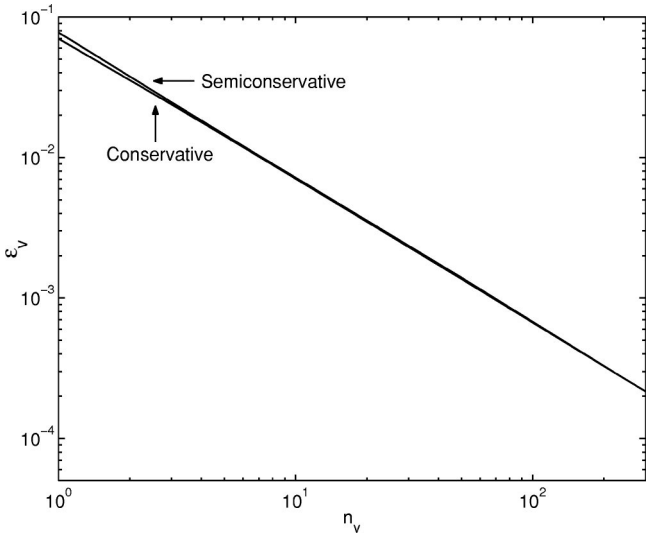

FIG. 3. Optimal viral mutation rate vs n_v for a conservative and semiconservative virus interacting with a semiconservative immune system; $n_{is}=100$, $\sigma_{is}=\sigma_v=100$, $\eta_{is}=\eta_v=1$, $\delta=200$, $\epsilon_{is}=0.001$.

$$[q_v^{n_v}(\sigma_v-\eta_v)+\delta]\{n_v(q_v-1)q_v^{2n_v}\sigma_v^2\tau_{is} + \delta[q_v+(q_v-1)n_v q_v^{n_v}\sigma_v\tau_{is}]+\tau_v[q_v-q_v^{n_v+1} - (q_v-1)n_v q_v^{2n_v}\sigma_v\tau_{is}]\}+n_v q_v^{n_v}(q_v-1) \times(\eta_v^2-\delta\sigma_v-\eta_v\sigma_v)\ln\left(\frac{1-q_v}{S-1}\right)=0, \tag{26}$$

the solution of which is shown in Fig. 3 for a chosen set of parameters.

We now turn our attention to the central theme of this paper, the coevolution of semiconservative organisms. Applying the results of Sec. II and following the procedure outlined above, we find, for a semiconservatively replicating host,

$$\kappa_{is}=\left(\frac{\sigma_{is}\epsilon_{is}(1-\epsilon_{is}/2)^{n_{is}-1}}{(S-1)(\sigma_{is}-\eta_{is})[2(1-\epsilon_{is}/2)^{n_{is}}-1]}\right) \times\left(e^{[2\sigma_{is}(1-\epsilon_{is}/2)^{n_{is}}-\sigma_{is}-\eta_{is}]\tau}-e^{[2\eta_{is}(1-\epsilon_{is}/2)^{n_{is}}-2\eta_{is}]\tau}\right), \tag{27}$$

$$\tau=\tau_{is}+\tau_v, \tag{28}$$

$$\tau_{is}=-\frac{\ln\left(\frac{(1-\epsilon_{is}/2)^{n_{is}}\epsilon_{is}}{[2(1-\epsilon_{is}/2)^{n_{is}}-1](S-1)}\right)}{[2(1-\epsilon_{is}/2)^{n_{is}}-1](\sigma_{is}-\eta_{is})}. \tag{29}$$

A conservatively replicating virus interacting with this host will still follow the behavior described by Eqs. (16) and (21), albeit with the proper, semiconservative τ_{is} defined above. For the case of a semiconservative virus we obtain

$$\kappa_v=\left(\frac{\sigma_v\epsilon_v(1-\epsilon_v/2)^{n_v-1}}{(S-1)(\sigma_v-\eta_v)[2(1-\epsilon_v/2)^{n_v}-1]}\right) \times\left(e^{[2\sigma_v(1-\epsilon_v/2)^{n_v}-\sigma_v-\eta_v]\tau}-e^{[2\eta_v(1-\epsilon_v/2)^{n_v}-2\eta_v]\tau}\right), \tag{30}$$

$$\tau_v=-\frac{\ln\left(\frac{(1-\epsilon_v/2)^{n_v}\epsilon_v}{[2(1-\epsilon_v/2)^{n_v}-1](S-1)}\right)}{[2(1-\epsilon_v/2)^{n_v}-1](\sigma_v-\eta_v)+\delta}. \tag{31}$$

We now proceed to find the optimal mutation rates for both organisms. Differentiating κ_v by ϵ_{is} and setting the result to zero gives us a criterion for the optimal immune mutation rate,

$$\frac{-2+\epsilon_{is}+n_{is}\epsilon_{is}-2(1-\epsilon_{is}/2)^{n_{is}}\left[-2+\epsilon_{is}+n_{is}\epsilon_{is}\ln\left(\frac{(S-1)[2-(1-\epsilon_{is}/2)^{-n_{is}}]}{\epsilon_{is}}\right)\right]}{[1-2(1-\epsilon_{is}/2)^{n_{is}}]^2(\epsilon_{is}-2)\epsilon_{is}}=0. \tag{32}$$

This equation has all of the nice properties of Eq. (24), defining an optimal mutation rate for any genome length, independent of the parameters of the system. The solution to this equation is plotted in Fig. 2, along with the conservative solution and the experimental range for observed rates per base pair per generation of somatic hypermutation in the complementary determining regions (CDR's) found in B-cell antigen receptors.

To maximize the stability of the viral quasispecies we set $\partial\kappa_v/\partial\epsilon_v=0$ as before. After a fair bit of work, we obtain an unwieldy expression omitted here in the interest of space [20]. The expression simplifies immensely in the limit $\delta\to\infty$, the limit of an ideally efficient immune system. In this limit,

$$\frac{n_v\epsilon_v}{2[1-2(1-\epsilon_v/2)^{n_v}]^2}+\frac{n_v\sigma_v\epsilon_v(1-\epsilon_v/2)^{n_v}\tau_{is}-1}{1-2(1-\epsilon_v/2)^{n_v}}=0. \tag{33}$$

The ideally efficient immune system is not an unreasonable approximation, as immune systems are highly efficient in destroying invaders once a suitable antibody is produced. The full expression as well as the above limiting form are dependent on both the parameters of the model and the prop-

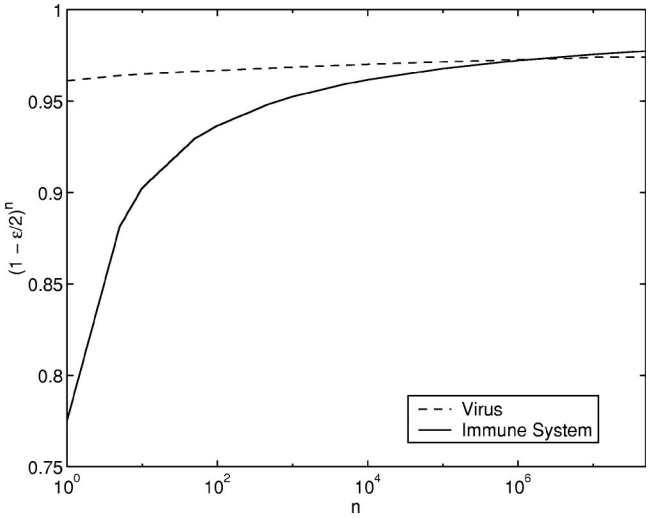

FIG. 4. $(1-\epsilon/2)^n$ for the optimal mutation rate of a semiconservative immune system and virus. This parameter can be used as a measure of the "conservativeness" of a semiconservative system; $\sigma_{is}=\sigma_v=100$, $\eta_{is}=\eta_v=1$, $\delta=200$, $\epsilon_{is}=0.001$.

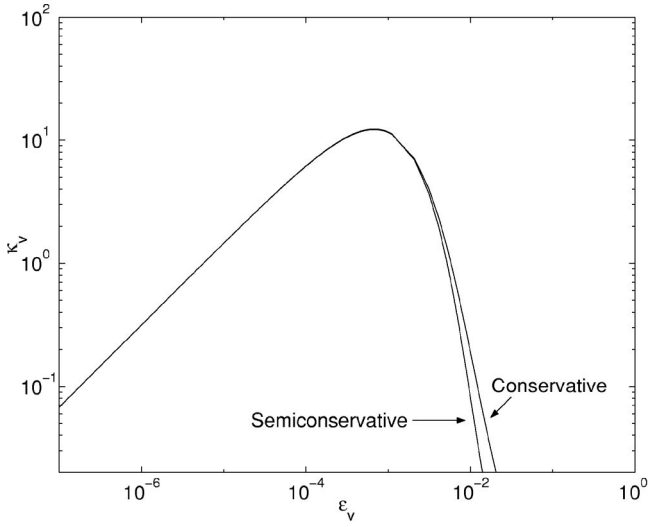

FIG. 5. κ_v vs ϵ_v for a conservative and semiconservative virus interacting with a semiconservative immune system; $n_{is}=n_v=100$, $\sigma_{is}=\sigma_v=100$, $\eta_{is}=\eta_v=1$, $\delta=200$, $\epsilon_{is}=0.001$.

erties of the immune system as in the conservative case. The solution of the full expression for a particular set of parameters is shown in Fig. 3.

IV. RESULTS AND DISCUSSION

Given the fundamental differences between semiconservative and conservative modes of replication, the most striking aspect of Figs. 2 and 3 is the similarity between the conservative and semiconservative optimal mutation rates at high n, particularly for the viral species. This is most easily understood by noting that, as $(1-\epsilon/2)^n \to 1$ for any semiconservatively replicating organism, the probability that a mutation will be found in the *original* strands after replication vanishes. Hence, in this limit, semiconservative and conservative replication are expected to mimic each other. This parameter is shown in Fig. 4 for the optimal viral and immune mutation rates. Clearly, with the exception of small immune genomes, the conservative system can be used as a good approximation for semiconservative replication. It is important to note, however, that this knowledge could not have been extracted from the data for the conservative system. A large value for $(1-\epsilon/2)^n$ in the conservative system is a necessary but not sufficient criterion to justify the use of a conservative model, and the full semiconservative calculation is required.

Equation (26) remains dependent on the parameters of the model, but general trends are obvious when biologically reasonable parameters are employed. While the extremal behavior of Eqs. (27) and (30) differs little from Eq. (16) for genome lengths that are not too small, the behavior away from the maxima differs greatly. Figure 5 displays κ_v vs ϵ_v for a given set of parameters for both the conservative and semiconservative models. It is immediately clear that, while the two models coincide at small ϵ (with a slightly higher peak height for either species for some parameters), their behavior differs greatly otherwise, with the semiconservative model displaying a more drastic dropoff in viability as ϵ increases, true for all biologically reasonable parameters studied. The parameters shown in Fig. 5 were chosen as a representative, rather than extreme, example of this behavior. The importance of this result is best understood in light of the evolutionary pressures one would expect a viral population to encounter. The independence of Eq. (32) from the properties of the viral system suggests that there exists an optimal mutation rate for an immune receptor *independent* of the qualities of the parasite against which it is defending. Thus, it is reasonable to expect (within the limitations imposed by additional evolutionary pressures, such as the need to distinguish between self- and foreign antigens) an immune receptor to evolve this mutation rate nearly exactly. However, in the viral case, the optimal mutation rate depends strongly on the nature of the immune system it is attacking. Thus, the virus must evolve the mutation rate that maximizes its overall viability against the range of immune systems it is likely to infect, including both inter-species and intra-species viability. The mutation rate that optimizes defense against one host may be a poor choice for another, and the virus must find the mutation rate that affords the best protection against all hosts, even if this is not the best mutation rate for evading any particular immune system. Such a compromise clearly involves the behavior of κ_v over a wide range of ϵ, rather than just at the maximum. One would therefore expect the more drastic dropoff at higher ϵ to force the semiconservative virus to develop a lower mutation rate so as to increase its viability against immune systems that lower the ϵ_v with the maximal value of $\partial\kappa_v/\partial\epsilon_v$. Quantifying this statement requires an intelligent estimate of the distribution of immune properties, a subject of future research. Qualitatively, this agrees well with the experimentally verified fact that semiconservative viruses display significantly lower mutation rates than their conservative counterparts [21,22].

V. CONCLUSIONS

In this paper, we have extended Kamp and Bornholdt's model of coevolution to incorporate the semiconservative nature of DNA replication for both species. A parameter-independent expression was derived for the optimal mutation rate of an immune receptor, which agrees well with experimental data. Convergence of the conservative and semiconservative results was demonstrated for realistic genome sizes, justifying the use of a conservative model in this case.

Optimizing the stability of the immune species yielded a maximum that coincides with the conservative model for realistic genome sizes. A similar correspondence exists for the virus, albeit with a dependence on the parameters of the model. Away from the maximum, the conservative and semiconservative models display different behaviors that provide a possible explanation for the high mutation rates found in conservative viruses. It is always dangerous to extrapolate from a simplified model of this kind to the complex systems found in nature. A true virus and immune system must contend with innumerable evolutionary pressures, biological, chemical and otherwise, such as the requirement that T-cells recognize and do not bind host proteins. The work represented in this paper describes a generalized model which we feel captures the robust qualitative features of host-parasite coevolution, providing insight into the complex workings of nature.

ACKNOWLEDGMENTS

The authors would like to thank Emmanuel Tannenbaum, Brian Dominy, Eric Deeds, and Stefan Bornholdt for useful and interesting discussions. This research was partially supported by an NIH postdoctoral fellowship.

[1] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[2] M. Eigen, J. McCaskill, and P. Schuster, Adv. Chem. Phys. **75**, 149 (1989).
[3] S. Galluccio, Phys. Rev. E **56**, 4526 (1997).
[4] P. Schuster and P. F. Stadler, Comput. Chem. (Oxford) **18**, 295 (1994).
[5] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).
[6] L. Peliti, e-print cond-mat/9712027.
[7] C. O. Wilke, C. Ronnewinkel, and T. Martinetz, Phys. Rep. **349**, 395 (2001).
[8] M. Nilsson and N. Snoad, Phys. Rev. Lett. **84**, 191 (2000).
[9] C. Kamp and S. Bornholdt, Phys. Rev. Lett. **88**, 068104 (2002).
[10] C. Kamp, C. O. Wilke, C. Adami, and S. Bornholdt, Complexity **8**, 28 (2003).
[11] E. Tannenbaum, E. J. Deeds, and E. I. Shakhnovich, Phys. Rev. E
[12] D. Voet and J. Voet, *Biochemistry* (Wiley, New York, NY, 1995).
[13] R. Pastor-Satorras and R. Sole, Phys. Rev. E **64**, 051909 (2001).
[14] S. Franz and L. Peliti, J. Phys. A **30**, 4481 (1997).
[15] P. R. A. Campos and J. F. Fontanari, Phys. Rev. E **58**, 2664 (1998).
[16] D. Alves and J. F. Fontanari, Phys. Rev. E **57**, 7008 (1998).
[17] S. Altmeyer and J. McCaskill, Phys. Rev. Lett. **86**, 5819 (2001).
[18] M. Nilsson and N. Snoad, Phys. Rev. E **65**, 031901 (2002).
[19] C. O. Wilke, C. Ronnewinkel, and T. Martinetz, Phys. Rep. **349**, 395 (2001).
[20] Expression available upon request.
[21] J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow, Genetics **148**, 1667 (1998).
[22] J. W. Drake and J. J. Holland, Proc. Natl. Acad. Sci. U.S.A. **96**, 13910 (1999).