Coevolution of Quasispecies: B-Cell Mutation Rates Maximize Viral Error Catastrophes

Christel Kamp and Stefan Bornholdt*

Institut für Theoretische Physik, Universität Kiel, Leibnizstraße 15, D-24098 Kiel, Germany

(Received 23 August 2001; published 28 January 2002)

Coevolution of two coupled quasispecies is studied, motivated by the competition between viral evolution and adapting immune response. In this coadaptive model, besides the classical error catastrophe for high virus mutation rates, a second "adaptation" catastrophe occurs, when virus mutation rates are too small to escape immune attack. Maximizing both regimes of viral error catastrophes is a possible strategy for an optimal immune response, reducing the range of allowed viral mutation rates to a minimum. From this requirement, one obtains constraints on B-cell mutation rates and receptor lengths, yielding an estimate of somatic hypermutation rates in the germinal center in accordance with observation.

DOI: 10.1103/PhysRevLett.88.068104 PACS numbers: 87.10.+e, 87.19.Xx, 87.23.Cc, 87.23.Kg

During the past 30 years, the concept of quasispecies [1,2] has developed into a valuable tool for modeling key features of molecular and viral evolution. In its most general form it describes sequence evolution under error-prone replication whose individual rate is given by a fitness function in sequence space. Although it is known that the topology of a subregion in sequence space with extraordinary high replicative power affects the selective advantage of its sequences and the robustness of the subsystem [3,4], essential features of the model can already be studied in the presence of a simple peaked fitness function, resulting in the formation of a central master sequence surrounded by a cloud of mutant sequences. A prominent feature of such systems is the occurrence of an error catastrophe, a sudden breakdown of stability when mutation rates get large. Recent developments of quasispecies models include their formulation within a statistical mechanics framework and a characterization of the error catastrophe as a phase transition [5–9].

While traditionally defined on static fitness landscapes, the concept of quasispecies recently has been extended to nonstationary fitness environments [10]. This has important applications, as viruses usually face quickly changing environments in the tight and temporal niches of their hosts. This new approach allows for studies of the adaptive response of a quasispecies to changing external conditions. This has been studied for different choices of time dependent fitness functions [10,11].

Here, we further extend this approach and study a virus in the environment of an adaptive immune system taking a further step beyond a simple time dependent viral fitness function. An interesting observation is that motifs of immune receptors often form something very similar to quasispecies: The presence of a viral epitope motif induces the proliferation of the corresponding immune receptor sequence. This "master" sequence is surrounded by a cloud of closely related receptor sequences that emerge from somatic hypermutation in the germinal centers [12,13]. Interpreting the conditions that lead to proliferation of specific immune receptors as their fitness, we can formally consider the B-cell population induced by a specific viral epitope as a quasispecies. In this paper we therefore study the coevolution of two asymmetrically coupled quasispecies under competition. While the immune quasispecies is strongly attracted by the virus, the viral quasispecies is driven away from its current master sequence by the immune system. This results in a migration through sequence space as observed in many infectious diseases as in HIV [14,15]. In the following, let us first define the model in detail. Then, dynamical regimes and stability bounds are discussed which occur as a result of the selective forces acting on both sides of the system. Finally, from the perspective of an optimal immune response, a relationship between receptor size and mutability is derived.

Consider a model with two quasispecies of genetic sequences, one of them coding for a virus and the second one coding for the variable part of an immune receptor. Sequence lengths are n_v and n_{is}, respectively, with bases taken from an alphabet of length $\lambda = 4$. Mutation rates are quantified by the copy fidelities per base $q_v, q_{\mathrm{is}} < 1$. The time evolution of the two distributions, the concentration $z_k(t)$ of viral sequence k, as well as the concentration of immune cells $y_k(t)$, representing receptor coding sequence k, is described by two sets of coupled differential equations of the type introduced by Eigen [1]

$$\dot{y}_k = \sum_l W_{kl}^{is} A(z_l) y_l - f y_k \,, \tag{1}$$

$$\dot{z}_k = \sum_l [W_{kl}^{v} B_l - \delta_{kl} C(y_l)] z_l \,, \tag{2}$$

$$W_{kl}^{i} = \frac{q_i^{n-d(k,l)} (1 - q_i)^{d(k,l)}}{(\lambda - 1)^{d(k,l)}} \,, \qquad i \in \{is, v\}; \qquad k, l \in \{1, \ldots, \lambda^n\} \,. \tag{3}$$

Locally in sequence space, we assume a simple one-to-one map between viral sequence and the sequence coding for

the immune receptor that maximally fits the viral epitope (within the local neighborhood of mutated receptors). For simplicity, both sequences share the same subscript l in the above formulation. Therefore, $A(z_l)$ denotes the growth rate of the B-cell clone corresponding to receptor sequence l and depends on the concentration of its complementary viral sequence z_l. The viral replication rate is B_l, and its decay rate $C(y_l)$, which depends on the associated immune cell concentration. For viral, as well as immune receptor evolution, a transition probability W_{kl}^i from a sequence l to sequence k by mutation is assumed, depending on the respective copy fidelity q_i, sequence length n_i, and the Hamming distance of the two sequences $d(k,l)$. For simplicity let us assume $n_{\text{is}} = n_v = n$ with the comparable complexity of viral epitopes and corresponding immune receptors in mind. Equation (1) models the relative concentrations of immune receptor coding sequences, with constant overall population size normalized by $f = \sum_l A(z_l)y_l$. The viral population (2), on the other hand, does not reach a constant population size, as the immune system usually works efficiently enough for the virus not to enter the regime of saturation. Therefore, absolute concentration is considered here and is the adequate quantity to quantify viral feedback to immune cell proliferation.

As viral existence in sequence space often is restricted to narrow niches with high fitness, let us assume a fitness landscape with a single peak with $B_l = \sigma_v \gg B_{m\neq l} = \eta_v$. This consequently neglects viral antigenic diversity but may be justified as a null hypothesis, assuming that there is generally a dominant strain among viable strains. This master sequence moves as a result of immune system pressure. Vice versa, as the immune response represents a very specific answer to a pathogen, let us also model the immune fitness landscape to have a single peak, corresponding to the receptor matching the current antigenic master sequence. This simplifies the coevolution model to a few discrete alternatives: $A(z_l) = \sigma_{\text{is}}$ if and only if z_l represents the concentration of the viral master sequence, otherwise $A(z_l) = \eta_{\text{is}} \ll \sigma_{\text{is}}$. Analogously, $C(y_l) = \delta$ if and only if y_l represents the dominant immune receptor's concentration, otherwise $C(y_l) = 0$. This makes the complicated couplings in the above equations tractable and allows us to neglect mutational backflow to the respective master sequences. Let us first write down simplified equations that apply to both quasispecies (1) and (2). For this purpose we use non-normalized concentrations [16,17] for quasispecies (1) also, and neglect the decay term in (2) for the moment. Then, each of the two quasispecies can be written in terms of the concentrations of a master sequence x_0 and of an arbitrary sequence of the first error class $\Gamma_1 x_1$.

$$\dot{x}_0(t) = q^n \sigma x_0(t), \tag{4}$$

$$\dot{x}_1(t) = \frac{1-q}{\lambda - 1} q^{n-1} \sigma x_0(t) + q^n \eta x_1(t). \tag{5}$$

The interaction between the two systems has to be specified by extra rules on the basis of the above definition of growth and decay rates. To keep the model as simple as possible, the decay rate δ affects only the viral sequence matching the dominant immune receptor. If this happens to coincide with the viral master sequence, the viral fitness peak will effectively move. Depending on its strength, the former fitness peak eventually will drop below the environmental growth rate. In this situation, sequences in the suppressed viral master sequence's neighborhood will now be selected with respect to positive deviations in growth rate from their neighbors representing an effective movement of the fitness peak. To be specific, the dynamical rules of this process are defined as follows: (1) Once the immune system imposes a decay rate $\delta > 0$ on the viral master sequence (so far stabilized at the viral fitness peak), the narrow niche of the virus is assumed to move to an arbitrary sequence of the first error class. (2) The viral quasispecies adapts to the new fitness peak on a time scale τ_v given by the dynamical equations above. (3) The fitness peak of the immune quasispecies is adjusted to the new maximum of the viral distribution. (4) The immune system adapts to the new fitness peak on a second time scale τ_{is} determined as above.

These steps are then iterated. While this is a strongly simplified picture of the coevolutionary dynamics of two coupled quasispecies, it allows a simple estimate of the dynamical regimes of two coupled sets of equations of type (4) and (5). Each of the fitness peaks is adjusted once during each cycle of duration $\tau = \tau_v + \tau_{\text{is}}$ (in steps 1 and 3, respectively). This allows us to follow the arguments of Nilsson and Snoad [10] to determine the growth of the respective future master sequences over a full cycle τ relative to the environmental growth $e^{\eta\tau}$ as a criterion for the quasispecies' survival [10]

$$\kappa = \frac{1}{e^{\eta\tau}} \frac{x_1(\tau)}{x_0(0)} = \frac{[e^{(q^n\sigma-\eta)\tau} - e^{(q^n\eta-\eta)\tau}](1-q)\sigma}{(\lambda-1)(\sigma-\eta)q}. \tag{6}$$

This expression is applied to each one of the two quasispecies (with the respective variables), defined over the full interval $\tau = \tau_v + \tau_{\text{is}}$ between two adjustments of its fitness peak. For a relative growth coefficient $\kappa > 1$ a species will survive and for $\kappa \leq 1$ it will get extinct.

Now consider a coupled system of viral and immune quasispecies where the immune part exerts a selective pressure on the virus in the form of a nonvanishing kill or decay rate δ. To estimate the migration time scale of the virus τ_v, let us iterate the model for a full cycle τ starting at the moment of the move of the fitness peak at $t = 0$. The relative size of the old and new master sequence peaks is then subsequently determined for another time interval τ_v. Let us assume $x_1(0) = 0$ since the new error class one sequence members are mainly recruited from the former, weakly populated error class two. The time scale τ_v is given by the waiting time until the new master sequence

population exceeds the old one:

$$e^{(q_v^n\eta_v-\delta)\tau_v}x_0(\tau) \stackrel{!}{=} e^{q_v^n\sigma_v\tau_v}x_1(\tau) \Rightarrow e^{(q_v^n\eta_v-\delta)\tau_v}e^{q_v^n\sigma_v\tau} = e^{q_v^n\sigma_v\tau_v}\,\frac{(e^{q_v^n\sigma_v\tau}-e^{q_v^n\eta_v\tau})(1-q_v)\sigma_v}{(\lambda-1)(\sigma_v-\eta_v)q_v}\,. \tag{7}$$

Mutational flows between the involved sequences can be neglected due to the small growth of the former master sequence and the small size of the initial new master sequence population. Assuming $\sigma_v \gg \eta_v$ and $q_v \approx 1$ the viral adaptation time scale can be estimated to

$$\tau_v \approx -\frac{\ln(\frac{1-q_v}{\lambda-1})}{q_v^n(\sigma_v-\eta_v)+\delta}\,. \tag{8}$$

Similarly, for the migration time for the immune quasispecies $\tau_{\rm is}$ we obtain

$$\tau_{\rm is} \approx -\frac{\ln(\frac{1-q_{\rm is}}{\lambda-1})}{q_{\rm is}^n(\sigma_{\rm is}-\eta_{\rm is})}\,. \tag{9}$$

Both, τ_v as well as $\tau_{\rm is}$ exhibit a local minimum at specific values of their copy fidelities q_v and $q_{\rm is}$, mainly determined by the balance between the requirement of a sufficiently large initial population for the formation of a future master sequence and sufficiently low mutational losses of the new master sequence. Inserting τ into the expressions for viral stability κ_v and immune stability $\kappa_{\rm is}$ according to (6), one obtains estimates for the regimes of viral and immune (co)existence. Because of the resulting intricate dependence of κ_v on q_v we were not able to derive a simple expression for an optimal mutation rate within the viral regime of existence as done in [10], to discuss its scaling behavior in dependence of genome length in comparison with former results [18,19].

Nonetheless we can get some qualitative information from Fig. 1 that shows viral and immunological regimes of (co)existence in terms of the respective mutation rates $\mu = 1-q$. The classical error catastrophe occurs at lower mutation rates in comparison to the static error threshold $\mu_{\rm err}^{\rm stat} = 1-(\frac{\eta}{\sigma})^{1/n} = 0.045 \approx \frac{\ln(\frac{\sigma}{\eta})}{n}$ (cf. [1]). This effect is due to additional mutational losses by migration and becomes large for small τ. In addition, Fig. 1 shows different limiting behaviors for κ_v and $\kappa_{\rm is}$ for $\mu_v \to 0$ and $\mu_{\rm is} \to 0$, respectively, which can be summarized as

$$\kappa_v \stackrel{\mu_v\to 0}{\to} 0\,,$$
$$\kappa_{\rm is} \stackrel{\mu_{\rm is}\to 0}{\approx} e^{(\sigma_{\rm is}-\eta_{\rm is})\tau_v}.$$

For the viral quasispecies one observes a second error ("adaptability") catastrophe at small viral mutation rates, because a minimum viral mutation rate is needed to escape the decay rate δ induced by the immune response at the viral master sequence.

The κ_v surface in the μ_v-$\mu_{\rm is}$ plane, whose $\kappa_v = 1$ contour lines are shown among others in Fig. 1, is dominated by a saddle point: $\kappa_v(\mu_v)$ exhibits a local maximum while $\kappa_v(\mu_{\rm is})$ shows a local minimum. An optimal strategy for viral suppression is, therefore, to adjust the mutation rate $\mu_{\rm is}$ of the immune quasispecies such that κ_v operates in its valley, with maximum regions of error catastrophes on both sides. One obtains the condition

$$\frac{\partial\kappa_v}{\partial\mu_{\rm is}} \stackrel{!}{=} 0\,, \tag{10}$$

which can be written as

$$\mu_{\rm is} - 1 + n_{\rm is}\mu_{\rm is}\ln\left(\frac{\mu_{\rm is}}{\lambda-1}\right) = 0\,. \tag{11}$$

This mutation rate minimizes the regime of possible existence of the viral quasispecies in Fig. 1. Depending on the involved viral and immune growth rates, this range of allowed viral mutation rates μ_v may vary (and even vanish for some values). Note that the relationship between optimal $\mu_{\rm is}$ and $n_{\rm is}$ is independent of $\sigma_{v/\rm is}$, $\eta_{v/\rm is}$, and δ. It depends only weakly on the length of the alphabet λ, which however is fixed here.

How does this compare with experimental results? Let us focus on B cells and their antibodies. Each antibody has at least two antigen receptors located in the variable regions of the antibody's heavy and light chains, each of which contains about 110 amino acids. Each receptor is coded by approximately 660 nucleotides. Antigen detection takes place in 6 subregions, the complementarity determining regions (CDRs) that represent 20%–30% of the antibody's variable(V) regions [20,21]. In the course of the primary immune response one observes somatic hypermutation in

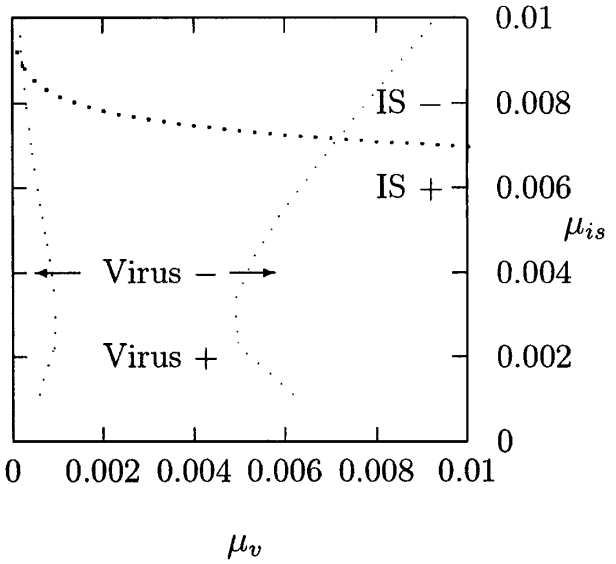

FIG. 1. Regimes of viral and immune quasispecies (co)existence, with $+/-$ denoting stable/unstable regions of the respective quasispecies in dependence on mutation rates μ_v and $\mu_{\rm is}$. Parameters are $\sigma_v = \sigma_{\rm is} = 10$, $\eta_v = \eta_{\rm is} = 1$, $\delta = 200$, $n_v = n_{\rm is} = 50$, and $\lambda = 4$. A large value of δ is chosen to get a good qualitative view of the system's behavior.

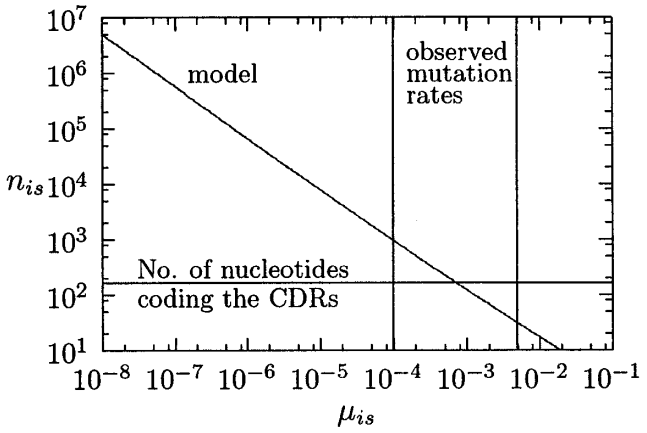

FIG. 2. Predicted mutation rates vs receptor coding lengths for an optimal immune response ($\lambda = 4$, $\sigma_{v/\mathrm{is}}$, $\eta_{v/\mathrm{is}}$, δ arbitrary), in comparison to observed rates of somatic hypermutation and observed CDR coding lengths.

the recombined V-region genes, with mutational hot spots at the CDRs, resulting in an enhanced affinity towards the invading antigen [12,22,23]. Observed mutation rates are in the range of 10^{-4}–10^{-3} mutations per base pair per generation [24–26]. Mutation rates in the CDRs are approximately twice to tenfold higher than those found in the entire V region [25,27]. These observations are quite universal to adaptive immune systems that are common to jawed vertebrates differing only in effectivity of selection due to varying stages of germinal centers' expression [28].

As Fig. 2 shows, the model prediction agrees well with the observed somatic hypermutation rates and CDR receptor lengths.

To summarize, the dynamics of the coevolution of two coupled quasispecies has been studied. In particular, this model was formulated to provide a simple toy model for the coadaptive system of viral evolution and immune adaptation. The model characterizes the different regimes of (co)existence of viral and immune quasispecies and predicts the correct range of somatic mutation rates in accordance with observation. Possible extensions of this work are numerous, as this is only a first account of basic principles of coevolving quasispecies. Analytical approaches beyond the simple approximation presented here, as well as numerical extensions, may provide a more accurate picture of the dynamics and further possibilities to relate to biological data. Further applications include modeling HIV dynamics, e.g., by adding an overall decay rate representing the HIV-induced loss of $CD4^+$ T cells.

C. K. would like to thank the Stiftung der Deutschen Wirtschaft for financial support.

*Email address: bornholdt@theo-physik.uni-kiel.de

[1] M. Eigen, Naturwissenschaften **58**, 465 (1971).
[2] M. Eigen and P. Schuster, *The Hypercycle—A Principle of Natural Self-Organization* (Springer-Verlag, Berlin, 1979).
[3] E. van Nimwegen, J. P. Crutchfield, and M. Huynen, Proc. Natl. Acad. Sci. U.S.A. **96**, 9716 (1999).
[4] C. O. Wilke, J. L. Wang, C. Ofria, R. E. Lenski, and C. Adami, Nature (London) **412**, 331 (2001).
[5] M. Nowak and P. Schuster, J. Theor. Biol. **137**, 375 (1989).
[6] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).
[7] S. Bonhoeffer and P. Stadler, J. Theor. Biol. **164**, 359 (1993).
[8] E. Baake, M. Baake, and H. Wagner, Phys. Rev. Lett. **78**, 559 (1997).
[9] S. Altmeyer and J. McCaskill, Phys. Rev. Lett. **86**, 5819 (2001).
[10] M. Nilsson and N. Snoad, Phys. Rev. Lett. **84**, 191 (2000).
[11] C. Wilke, C. Ronnewinkel, and T. Martinetz, Phys. Rep. **349**, 395 (2001).
[12] R. S. Harris, Q. Kong, and N. Maizels, Mutat. Res. **436**, 157 (1999).
[13] M. Meyer-Herrman, A. Deutsch, and M. Or-Guil, J. Theor. Biol. **210**, 265 (2001).
[14] S. Ganeshan, R. E. Dickover, B. Korber, Y. Bryson, and S. Wolinsky, J. Virol. **71**, 663 (1997).
[15] T. Allen *et al.,* Nature (London) **407**, 386 (2000).
[16] B. Jones, R. Enns, and S. Rangnekar, Bull. Math. Biol. **38**, 15 (1976).
[17] C. Thompson and J. McBride, Math. Biosci. **21**, 127 (1974).
[18] M. Nowak, Nature (London) **347**, 522 (1990).
[19] A. Sasaki, J. Theor. Biol. **168**, 291 (1994).
[20] I. Roitt, *Essential Immunology* (Blackwell Scientific Publ., Oxford, 1994).
[21] A. Aigner and S. Neumann, *Immunchemie* (Fischer, Jena, 1997).
[22] M. Weigert, I. Cesari, S. J. Yonkovich, and M. Cohn, Nature (London) **228**, 1045 (1970).
[23] G. Nossal, Cell **68**, 1 (1992).
[24] D. McKean *et al.,* Proc. Natl. Acad. Sci. U.S.A. **81**, 3180 (1984).
[25] C. Berek and C. Milstein, Immunol. Rev. **105**, 5 (1988).
[26] N. S. Levy, U. V. Malipiero, S. G. Lebecque, and P. J. Gearhart, J. Exp. Med. **169**, 2007 (1989).
[27] T. Dörner *et al.,* J. Immunol. **158**, 2779 (1997).
[28] M. F. Diaz and M. Flajnik, Immunol. Rev. **162**, 13 (1998).