Error catastrophe and phase transition in the empirical fitness landscape of HIV

Gregory R. Hart[1] and Andrew L. Ferguson[2,*]

[1]*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

[2]*Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

(Received 9 August 2014; published 10 March 2015)

We have translated clinical sequence databases of the p6 HIV protein into an empirical fitness landscape quantifying viral replicative capacity as a function of the amino acid sequence. We show that the viral population resides close to a phase transition in sequence space corresponding to an "error catastrophe" beyond which there is lethal accumulation of mutations. Our model predicts that the phase transition may be induced by drug therapies that elevate the mutation rate, or by forcing mutations at particular amino acids. Applying immune pressure to any combination of killer T-cell targets cannot induce the transition, providing a rationale for why the viral protein can exist close to the error catastrophe without sustaining fatal fitness penalties due to adaptive immunity.

DOI: 10.1103/PhysRevE.91.032705 PACS number(s): 87.19.xd, 87.15.ak, 87.15.Zg, 87.23.Cc

HIV has killed more then 30 million persons worldwide, with another 30 million infected [1]. Antiretroviral drugs have rendered HIV infection a manageable condition [1], but their high cost makes them effectively inaccessible in the developing world [2] and rates of drug-resistant mutations in persons on therapy more than 36 months exceed 20% [3]. A vaccine remains unavailable [4,5]. The high mutation rate and sequence diversity of HIV present significant challenges for therapy, but recent computational advances offer new ways to identify susceptible targets to guide the design of new drugs and vaccines [6–9].

We recently presented an approach to translate clinical databases of HIV sequences into models of the viral "fitness landscape" that quantify viral replicative capacity as a function of its amino acid sequence [7,8,10]. Most viral mutations are deleterious, compromising fitness, but certain patterns of mutations enable the virus to escape immune surveillance while maintaining high fitness [6]. Empirical fitness models can explicitly resolve these high fitness pathways, and reveal vulnerabilities where drug therapy or vaccine-induced immune pressure can cripple viral fitness [7–9]. Regarding sequences in the database as observations from an underlying probability distribution, the least structured (i.e., maximum entropy) model capable of reproducing the two lowest order moments of the amino acid frequencies in the clinical database (i.e., frequency with which each single amino acid is observed in each position, and pairs of amino acids in pairs of positions) is the infinite range Potts model [8,11–13],

$$P(\vec{z},T)=\frac{e^{-\beta E(\vec{z})}}{Z(T)},\quad E(\vec{z})=\sum_{i=1}^{m}h_i(z_i)+\sum_{i=1}^{m}\sum_{j>i}^{m}J_{ij}(z_i,z_j), \tag{1}$$

where $E(\vec{z})$ is the infinite range Potts Hamiltonian, m is the number of positions in the protein, $\vec{z}$ is a m-element vector encoding the protein sequence, and each element of $\vec{z}$ is an integer in the range 1–21 corresponding to the 20 natural amino acids plus the gap or blank [14]. As in statistical thermodynamics, we term E the dimensionless "energy," $Z(T)=\sum_{\vec{z}}e^{-\beta E(\vec{z})}$ the partition function, and $\beta=1/T$ the dimensionless inverse temperature, that we set to unity for the purposes of parameter inference [7,8,11,13]. (In contrast to the standard expressions, our energy and temperature are dimensionless and we have eliminated Boltzmann's constant.) Determination of the model parameters—the one-body external fields $\{h_i\}$ and two-body pairwise couplings $\{J_{ij}\}$—that reproduce the observed one and two-body amino acid frequencies constitutes solution of a canonical inverse problem that may be tackled in many ways [8,13].

Under the ansatz that highly prevalent viral strains should also be highly fit, the prevalence of a strain in the population of infected hosts $P(\vec{z})$ should be a good proxy for its replicative fitness $f(\vec{z})$, and thus the energy assigned by our model $E(\vec{z})$ should be negatively correlated with the logarithmic fitness $\log(f(\vec{z}))$ [7–9]. This relationship can be derived exactly under restrictive assumptions [15], and we have shown that under mild conditions on the sequences in the clinical databases, this relationship holds generically [9]. Furthermore, we have shown that if these sequences are drawn from a genetically diverse population of infected hosts, the effects of host immune responses are averaged out, such that the model predicts the intrinsic viral fitness, uncontaminated by "footprints" of adaptive immunity [9]. We previously reported models for the HIV proteins p17 (matrix) and p24 (capsid), and empirically validated them against experimental measurements of *in vitro* replicative fitness [8].

Since the model parameters are fitted at a temperature of unity, $T_{\mathrm{op}}=1$ is the effective "biological operating temperature" [13]. Modulating T away from unity corresponds to a uniform scaling of the $\{h_i\}$ and $\{J_{ij}\}$ parameters [13]. As $T\to 0$, the fitness landscape predicts the lowest energy (highest fitness) strain to dominate the viral population, whereas for $T\to\infty$, all strains become equally represented independent of energy (fitness).

It is the aim of this work to show that the fitness landscape for the p6 protein in HIV-1B predicts that the viral population is poised close to a phase transition at $T=1.20$ between a high-fitness, low-diversity (i.e., low-energy, low-entropy) and a low-fitness, high-diversity (i.e., high-energy, high-entropy) population, and that the transition may be induced—and viral fitness crippled—by elevating the mutation rate or forcing mutations at particular amino acid positions.

*alf@illinois.edu

Model inference. An ensemble of 1824 DNA sequences of the HIV-1 p6 protein were downloaded from the LANL HIV database [16]. Sequences were restricted to subtype B—the most prevalent form in Europe and the Americas—treatment-naïve hosts, and not classified as "problematic." Sequences were aligned to the HXB2 reference sequence [17], and translated to the cognate 53 amino acid protein sequence. Ambiguous codons were translated as a blank. We fitted the $\{h_i\}$ and $\{J_{ij}\}$ parameters in Eq. (1) using the approach in Ref. [8]. Using the approach detailed in Ref. [9], we have estimated that no individual position in p6 should be subjected to CTL immune pressure in more than 17% of the human population. This lends empirical support to our assertion that signatures of adaptive immunity should be averaged out in a sequence ensemble drawn from immunologically diverse hosts.

Model validation. Protein p6 is less well experimentally and clinically studied than the other HIV Gag proteins [7,8], but analysis by Brumme *et al.* of the IHAC cohort of chronically infected hosts identified two statistically significant p6 escape mutations directly associated with T-cell immune pressure: B40-E34D [mutation from E (Glu) to D (Asp) at p6 position 34 in an epitope presented by the human leukocyte antigen (HLA) B40] and A68-R42K [18]. Particular contiguous groups of amino acids known as *epitopes* are recognized by T cells when all positions contain wild-type (most probable) amino acids [19]. The virus escapes immune recognition by mutating at one or more of the positions in the epitope. Cross referencing the two escape mutations with the LANL T-cell epitope maps [16], we identified the former as associated with the B40-K_{33}ELYPLTSL$_{41}$ cytotoxic T-cell (CTL) epitope, and the latter with the A68-K_{33}ELYPLTSLRS$_{43}$ and A68-E_{29}PIDKELYPLTSLRS$_{43}$ helper T-cell (Th) epitopes. If we assume that all polymorphisms within the targeted epitope are equally efficient at mediating escape, then the virus should make mutations incurring the smallest energy costs (lowest fitness penalty). Our model predicts that of all the p6 polymorphisms in the clinical sequence database (i) the E34D and R42K escapes are the single lowest energy polymorphisms at these two positions, and (ii) the E34D escape is the 2/69 lowest energy polymorphism within the B40-K_{33}ELYPLTSL$_{41}$ epitope, and the R42K escape as the 1/73 and 1/118 lowest energy polymorphism within the A68-K_{33}ELYPLTSLRS$_{43}$ and A68-E_{29}PIDKELYPLTSLRS$_{43}$ epitopes, respectively. That our model predicts the clinically observed escape mutations to incur the smallest energy cost (fitness penalty) provides support that it reflects intrinsic viral fitness. We observe that we have validated analogous models for HIV Gag p17 and p24 against the more comprehensive clinical and experimental data available for these proteins [7,8].

Density of states. Computing the partition function in Eq. (1) "solves" the Potts model in the sense that nearly all thermodynamic quantities are calculable from $Z(T)$ [20]. We may express the partition function as a sum over energy levels, $Z(T) = \sum_{\vec{z}} e^{-\beta E(\vec{z})} = \sum_E g(E) e^{-\beta E}$, where $g(E)$ is the density of states. We developed a parallel implementation of the Wang-Landau algorithm [21,22] described in Refs. [23,24] to estimate the density of states, $\hat{g}(E) = C\, g(E)$, up to a multiplicative constant C, that we fix by asserting the uniqueness of the lowest energy (highest fitness) wild-type strain [Fig. 1(a)]. From $Z(T)$

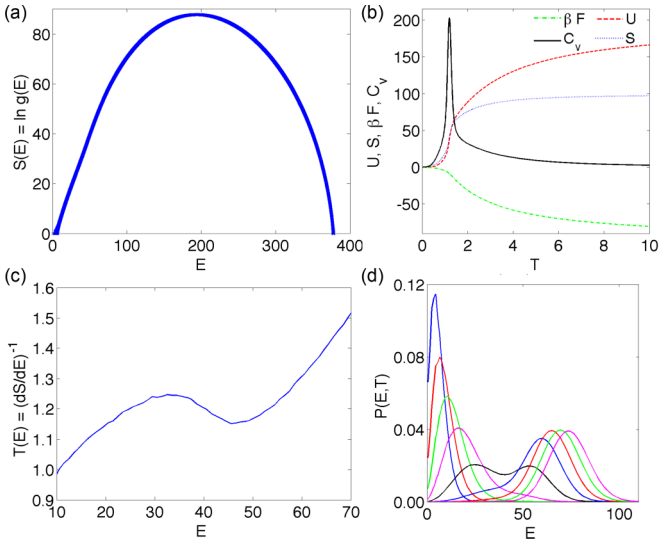

FIG. 1. (Color online) Thermodynamics of HIV-1B protein p6. (a) Density of states estimated by Wang-Landau sampling. (b) Dimensionless energy U, entropy S, free energy F, and heat capacity C_v, as a function of the dimensionless temperature $T = \beta^{-1}$. (c) The microcanonical caloric curve $T(E)$, which has a negative gradient over the region E=35–45. (d) The canonical distribution $P(E,T)$ at (left to right) T=0.8:0.1:1.6 exhibits a bimodal distribution at T_{coex}=1.20.

we computed the energy $U(T) = -\partial \ln Z(T)/\partial\beta$, free energy $\beta F(T) = -\ln Z(T)$, entropy $S(T) = \beta U(T) - \beta F(T)$, heat capacity $C_v(T) = \partial U/\partial T$, microcanonical temperature prescribed by the caloric curve $T(E) = [\partial S(E)/\partial E]^{-1} = [\partial(\ln g(E))/\partial E]^{-1}$, and canonical distribution $P(E,T) = g(E)e^{-\beta E}/Z(T)$ [Figs. 1(b)–1(d)].

Phase transitions are formally defined in infinite systems as nonanalyticities in the equation of state, but there have been many studies of finite system analogs of bulk transitions [25]. We observe a jump in $U(T)$ and $S(T)$, sharp peak in $C_v(T)$, bimodality in $P(E,T)$, and a negative gradient in the microcanonical caloric curve—indicative of positive curvature in $S(E)$ and a negative microcanonical heat capacity $C_v(E) = -(\partial S/\partial E)^2 / (\partial^2 S/\partial E^2)$—at the coexistence temperature $T_{\text{coex}} = 1.20$, defined by equal areas below the peaks of $P(E,T)$ [21] (Fig. 1). These are all (equivalent) indicators of a finite system phase transition in the sequence space of p6 at $T_{\text{coex}} = 1.20$ [25,26].

Prevalence of mutant strains. In performing Wang-Landau sampling, we also collected estimates of the average fraction of the 53 amino acids that were non-wild-type for all sequences of a particular energy $f_{\text{nwt}}(E)$, and the fraction of sequences containing precisely $k = \{0,1,2\}$ non-wild-type amino acids $H_k(E)$. Combining these distributions with our density of states, we computed these quantities as a function of temperature using the identity, $X(T) = \sum_E X(E) g(E) e^{-\beta E}/Z(T)$ [27] (Fig. 2). Consistent with the sharp increase in the energy and entropy of the viral population (i.e., decrease in fitness and increase in diversity) at T_{coex}=1.20, we observe a concomitant jump in the fraction of mutant residues per strain, $f_{\text{nwt}}(T)$. At T_{op}=1, the fraction of strains in the viral population that are wild type, contain a single mutation, and contain

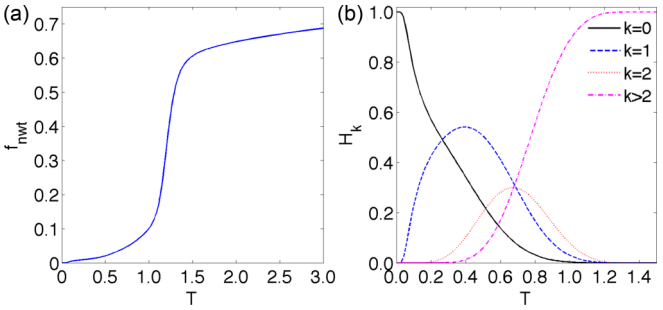

FIG. 2. (Color online) Prevalence of mutant strains. (a) The average fraction of non-wild-type residues per strain f_{net} exhibits a sharp jump at T_{coex}=1.20. (b) The fraction of strains in the population that are a Hamming distance of k=0,1,2 from the wild-type strain as a function of temperature.

exactly two mutations are $H_0(T)$=0.34%, $H_1(T)$=3.0%, and $H_2(T)$=7.8%. At T_{coex}=1.20 these fractions are 0.0079%, 0.11%, and 0.45%, with 99.4% of strains containing three or more point mutations.

Permutation test. To determine whether the observed phase transition is a generic property of our Potts Hamiltonian, we computed the density of states for 10 artificial Hamiltonians constructed by random shuffling of the $\{J_{ij}\}$ parameters leaving the $\{h_i\}$ parameters intact, and 10 in which both the $\{h_i\}$ and $\{J_{ij}\}$ parameters were independently shuffled. In all cases the signatures of the phase transition—jump in $U(T)$ and $S(T)$, bimodality in $P(E,T)$, back bending in $T(E)$, and sharp peak in $C_v(T)$—vanished (Fig. 3), indicating that the phase transition is contingent on the precise structure of external fields and mutational couplings within the p6 protein. We are currently working to understand the properties of the $\{h_i\}$ and $\{J_{ij}\}$ that mediate the presence or absence of a phase transition by studying the structure of the energy landscape prescribed by the Potts Hamiltonian.

Interpretation of T. The high replication and mutation rates of HIV cause it to exist as a *quasispecies*, or cloud of closely related mutant strains, whose nonequilibrium evolution is described by Eigen's equation [28–31]. Leuthäuser mapped Eigen's equation to the equilibrium properties of a two-dimensional lattice model [32,33]. We have shown that the

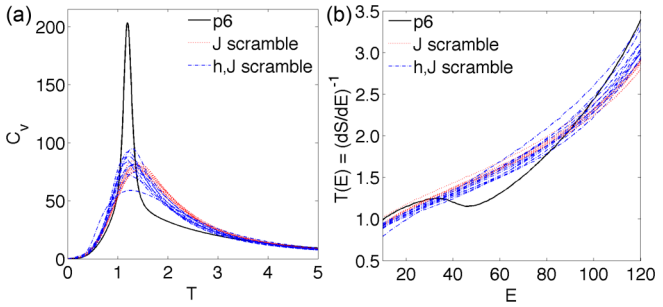

FIG. 3. (Color online) Artificial Hamiltonians generated by 10 random shuffles of the $\{J_{ij}\}$ parameters (dotted lines), and 10 shuffles of the $\{h_i\}$ and $\{J_{ij}\}$ parameters (dot-dashed lines), do not show signatures of the phase transition exhibited by p6. (a) The sharp peak in C_v disappears, and (b) the caloric curve does not possess any region of negative gradient.

Ising models inferred using our approach [i.e., Eq. (1) with binary $\vec{z}$ elements denoting wild-type and mutant amino acids] well approximate the equilibrium distribution of strains in Leuthäuser's formulation [9]. Under this correspondence, $\beta = T^{-1} = \ln\sqrt{q/(1-q)}$, where q is the per position probability of correctly copying an amino acid in a viral replication event. In the limit of perfect replication fidelity $q \to 1$ and $T \to 0$, whereas for random copying $q \to 0.5$ and $T \to \infty$. Extending this correspondence to the Potts model, we may interpret T as an increasing function of the viral mutation rate, and the proximity of the phase transition at T_{coex}=1.20 to T_{op}=1 is precisely analogous to the observation that the HIV error rate lies very close to the *error catastrophe* beyond which there is lethal error accumulation, and the quasispecies collapses [30,31,34–37]. The existence of the viral error catastrophe is well established in theoretical models of viral mutational dynamics [30,38,39], and a recent study of HIV employing theoretical fitness models motivated by experimental data detected an error catastrophe in both quasispecies models and stochastic population genetics-based simulations [39]. To the best of our knowledge, the present work represents the first time that the error catastrophe has been detected in an empirical model of viral fitness. Experimental reports supporting an HIV error catastrophe *in vivo* have motivated clinical trials of an HIV mutagen [40,41], but confounding factors, such as cellular resource scarcity at high mutation rates and the existence of a competing "extinction threshold," have so far precluded unambiguous experimental confirmation of this transition [39,42,43].

The proximity of the phase transition is thought to convey survival advantage by providing a phenotype reservoir [35], maximizing adaptability [36], and optimizing immune escape [30]. Sella and Hirsh have proposed that a viral quasispecies should minimize a "free fitness," $G = U - T_{\text{op}}S$, as an optimal balance between fitness and diversity [15]. We find that $G(T) = U(T) - (T_{\text{op}} = 1)\,S(T)$ exhibits a minimum at T=1.00, suggesting that the quasispecies is structured in accordance with this principle. Immunologically, we may exploit the proximity of the transition to induce a large increase in the average energy (decrease in fitness) of the viral population $U(T)$ by a small increase in T. In practice, T_{op} may be pushed beyond T_{coex}=1.20 by drug therapies that elevate the viral mutation rate [44,45]. Promising results have emerged in recent years [40,41], but the clinical translation of mutagens designed to exploit the purported error catastrophe warrants careful further study [38].

One- and two-point mutations. We hypothesized that it may also be possible to induce the phase transition by forcing the virus to make mutations at particular positions in the protein. To test this conjecture, we recomputed the density of states under the constraint that the virus was forbidden from mutating to strains in which the amino acid at a single position i was wild type for each of the i=1...53 positions. Rejection of trial moves to these forbidden states in the Wang-Landau algorithm were treated according to Ref. [46]. As illustrated in Fig. 4, forcing single point mutations depresses T_{coex} with an attendant increase in the average energy (decrease in fitness) of the viral population. Remarkably, forcing a point mutation in the amino acid at position i=11 depresses the coexistence temperature to T_{coex}=1.01, and more than doubles the population energy at the

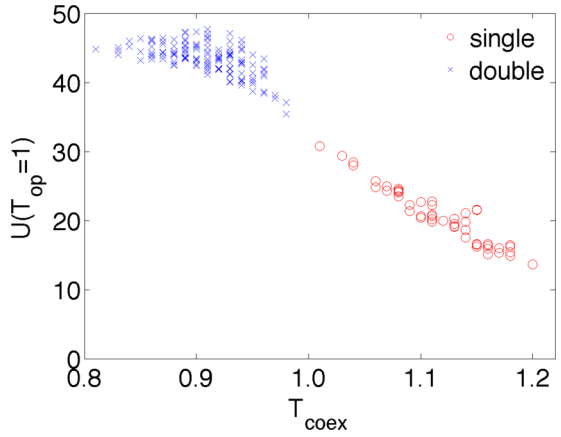

FIG. 4. (Color online) Impact on the average energy of the viral population at the operating temperature $U(T_{\rm op}=1)$ by forcing (a) single mutations away from the wild-type residue at each of the 53 amino acids (circles), and (b) double mutations at pairs of the 17 positions at which single mutations caused a $T_{\rm coex}$ to fall below T=1.1 (crosses).

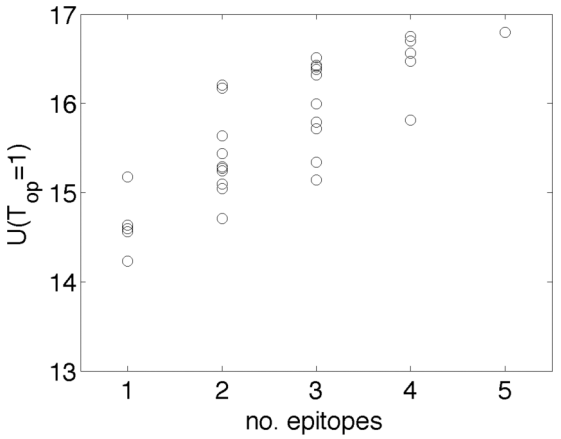

FIG. 5. Impact on the average energy of the viral population at the operating temperature, $U(T_{\rm op}=1)$, upon applying immune pressure to all possible {1,2,3,4,5}-epitope combinations of the five known CTL epitopes in p6.

operating temperature from $U\,(T_{\rm op}=1)$=13.5 [cf. Fig. 1(b)] to 30.7.

Application of mutational pressure to any single position does not, however, depress $T_{\rm coex}$ below $T_{\rm op}$=1. Accordingly, we identified the 17 positions in the protein at which mutational pressure pushed $T_{\rm coex}<1.1$, and recomputed the density of states for the $(17\times16)/2=136$ systems constrained such that pairs of amino acids at positions i and j within this set were both forbidden to be wild type. As illustrated in Fig. 4, all 136 instances of two-point mutational pressure induced the phase transition (i.e., $T_{\rm coex}<1$), with the coexistence temperature maximally depressed to $T_{\rm coex}=0.81$ by positions $(i=11, j=37)$, and the average energy maximally elevated to $U\,(T_{\rm op}=1)$=47.7 by positions (i=44, $j=53$).

CTL immune pressure. As an intracellular viral protein, the T cells of the adaptive immune system are primarily responsible for recognizing p6 as a pathogenic protein by binding to epitopes comprising a small number of contiguous amino acids. The LANL T-cell epitope maps list five known CTL epitopes within p6: T_{23}PSQKQEPI$_{31}$, S_{25}QKQEQIDK$_{33}$, E_{29}PKDREPL$_{38}$, K_{33}ELYPLTSL$_{41}$, and Y_{36}PLASLRSLF$_{45}$ [16]. Eight Th epitopes have been mapped to within a region of 18 amino acids or less, but only one was confirmed as an exact epitope. Motivated by interest in the design of CTL-based HIV vaccine immunogens [5,7,47], we sought to assess whether the application of CTL immune pressure can also induce the phase transition. To simulate CTL targeting at each single epitope, we recomputed the density of states under the constraint that the virus was forbidden to contain all wild-type amino acids within the epitope [46]. We performed analogous calculations for all possible doubles, triples, quads, and quints of epitopes.

As illustrated in Fig. 5, targeting of any single epitope only raises the average energy of the viral population from $U\,(T_{\rm op}=1)=13.5$ to 14.2–15.2. Simultaneous targeting of all five epitopes corresponds to $U\,(T_{\rm op}=1)$=16.8. All possible combinations of epitope targeting result in $T_{\rm coex}$=1.17–1.19, indicating that CTL immune pressure at any known epitopes is unable to induce the phase transition, and can only effect relatively modest increases in population energy. Indeed, we find that it is not possible to induce the phase transition by targeting *any* contiguous group of nine amino acids within p6.

In summary, we have translated sequence databases of the HIV-1 clade B p6 protein into an empirical fitness landscape quantifying the replicative capacity of the virus as a function of its amino acid sequence. Using Wang-Landau sampling, we have identified a phase transition in the sequence space of the p6 viral protein corresponding to an error catastrophe. Our model predicts that the transition can be induced by elevating the operating temperature beyond the coexistence temperature, $T_{\rm op}>T_{\rm coex}$, using drug therapies that increase the viral mutation rate [44,45]. Alternatively, the coexistence temperature may be suppressed below the operating temperature, $T_{\rm coex}<T_{\rm op}$, by forcing particular pairs of mutations, suggesting a means to cripple the HIV virus by applying mutational pressure at carefully selected positions. In principle, this might be achieved by drugs or small molecule inhibitors with localized binding sites on the p6 protein, similar to the anti-HIV drugs that bind protease, integrase, and reverse transcriptase [48,49]. In practice, development of such molecules is a challenging problem in drug design. Interestingly, CTL immune pressure at any combination of known epitopes in p6 cannot induce the transition, providing an empirical rationalization for why the virus can exist in close proximity to the error catastrophe without sustaining catastrophic fitness costs due to adaptive immune pressure. We are extending our analysis to empirical fitness landscapes for the larger HIV Gag proteins p17 and p24, and the nonstructural proteins of hepatitis C virus to determine the generality of our findings to other viral proteins.

G.R.H. acknowledges support through a Computational Science and Engineering Graduate Fellowship from the University of Illinois. We thank Dr. Tom Butler and Dr. John Barton for useful discussions.

[1] K. M. De Cock, H. W. Jaffe, and J. W. Curran, Emerg. Infect. Dis. **17**, 1044 (2011).

[2] K. A. Gebo, J. A. Fleishman, R. Conviser, J. Hellinger, F. J. Hellinger, J. S. Josephs, P. Keiser, P. Gaist, R. D. Moore *et al.*, AIDS **24**, 2705 (2010).

[3] K. M. Stadeli and D. D. Richman, Antivir. Ther. **18**, 115 (2013).

[4] B. M. Baker, B. L. Block, A. C. Rothchild, and B. D. Walker, Expert Opin. Biol. Th. **9**, 55 (2009).

[5] B. D. Walker and D. R. Burton, Science **320**, 760 (2008).

[6] V. Dahirel, K. Shekhar, F. Pereyra, T. Miura, M. Artyomov, S. Talsania, T. M. Allen, M. Altfeld, M. Carrington, D. J. Irvine *et al.*, Proc. Natl. Acad. Sci. USA **108**, 11530 (2011).

[7] A. L. Ferguson, J. K. Mann, S. Omarjee, T. Ndung'u, B. D. Walker, and A. K. Chakraborty, Immunity **38**, 606 (2013).

[8] J. K. Mann, J. P. Barton, A. L. Ferguson, S. Omarjee, B. D. Walker, A. K. Chakraborty, and T. Ndung'u, PLOS Comput. Biol. **10**, e1003776 (2014).

[9] K. Shekhar, C. F. Ruberman, A. L. Ferguson, J. P. Barton, M. Kardar, and A. K. Chakraborty, Phys. Rev. E **88**, 062705 (2013).

[10] S. Wright, in *Proceedings of the Sixth International Congress of Genetics*, Vol. 1 (Genetics Society of America, Bethesda, 1932), pp. 356–366.

[11] F.-Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

[12] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

[13] T. Mora and W. Bialek, J. Stat. Phys. **144**, 268 (2011).

[14] F. Morcos, A. Pagiani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa, and M. Weigt, Proc. Natl. Acad. Sci. USA **108**, E1293 (2011).

[15] G. Sella and A. E. Hirsh, Proc. Natl. Acad. Sci. USA **102**, 9541 (2005).

[16] Los Alamos National Laboratory HIV Database [http://www.hiv.lanl.gov] (Accessed: July 12, 2014).

[17] T. Leitner, B. Korber, M. Daniels, C. Calef, and B. Foley, in *HIV Sequence Compendium 2005 (Report No. LA-UR 06-0680)*, edited by T. Leitner, B. Foley, B. Hahn, P. Marx, F. McCutchan, J. W. Mellors, S. Wolinsky, and B. Korber (Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, 2005), pp. 41–48.

[18] Z. L. Brumme, M. John, J. M. Carlson, C. J. Brumme, D. Chan, M. A. Brockman, L. C. Swenson, I. Tao, S. Szeto, P. Rosato *et al.*, PLOS ONE **4**, e6687 (2009).

[19] K. Murphy, P. Travers, and M. Walport, *Janeway's Immunobiology*, 8th ed. (Taylor & Francis, Philadelphia, 2011).

[20] D. P. Landau, S.-H. Tsai, and M. Exler, Am. J. Phys. **72**, 1294 (2004).

[21] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).

[22] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[23] M. O. Khan, G. Kennedy, and D. Y. C. Chan, J. Comput. Chem. **26**, 72 (2005).

[24] L. Zhan, Comput. Phys. Commun. **179**, 339 (2008).

[25] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses* (Cambridge University Press, Cambridge, 2003).

[26] P. Chomaz and F. Gulminelli, AIP Conf. Proc. **970**, 175 (2008).

[27] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, Proteins: Struct., Funct., Bioinf. **62**, 232 (2006).

[28] M. Eigen, Naturwissenschaften **58**, 456 (1971).

[29] M. Eigen and P. Schuster, Naturwissenschaften **64**, 541 (1977).

[30] M. Nowak and R. M. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology* (Oxford University Press, Oxford, 2000).

[31] P. Tarazona, Phys. Rev. A **45**, 6038 (1992).

[32] I. Leuthäusser, J. Chem. Phys. **84**, 1884 (1986).

[33] I. Leuthäusser, J. Stat. Phys. **48**, 343 (1987).

[34] M. Eigen, Proc. Natl. Acad. Sci. USA **99**, 13374 (2002).

[35] J. Alonso and H. Fort, Philos. T. R. Soc. A **368**, 5569 (2010).

[36] R. V. Solé, Eur. Phys. J. B **35**, 117 (2003).

[37] J. W. Drake and J. J. Holland, Proc. Natl. Acad. Sci. USA **96**, 13910 (1999).

[38] M. Nowak and P. Schuster, J. Theor. Biol. **137**, 375 (1989).

[39] K. Tripathi, R. Balagam, N. K. Vishnoi, and N. M. Dixit, PLOS Comput. Biol. **8**, e1002684 (2012).

[40] M. J. Dapp, C. L. Clouser, S. Patterson, and L. M. Mansky, J. Virol. **83**, 11950 (2009).

[41] J. I. Mullins, L. Heath, J. P. Hughes, J. Kicha, S. Styrchak, K. G. Wong, U. Rao, A. Hansen, K. S. Harris, J.-P. Laurent *et al.*, PLOS ONE **6**, e15135 (2011).

[42] A. Grande-Pérez, E. Lázaro, P. Lowenstein, E. Domingo, and S. C. Manrubia, Proc. Natl. Acad. Sci. USA **102**, 4448 (2005).

[43] J. J. Bull, R. Sanjuan, and C. O. Wilke, J. Virol. **81**, 2930 (2007).

[44] R. A. Smith, L. A. Loeb, and B. D. Preston, Virus Res. **107**, 215 (2005).

[45] J. Summers and S. Litwin, J. Virol. **80**, 20 (2006).

[46] B. J. Schulz, K. Binder, M. Müller, and D. P. Landau, Phys. Rev. E **67**, 067102 (2003).

[47] M. J. McElrath and B. F. Haynes, Immunity **33**, 542 (2010).

[48] E. J. Arts and D. J. Hazuda, Cold Spring Harb. Perspect. Med. **2**, a007161 (2012).

[49] E. De Clercq, Nat. Rev. Drug Discov. **6**, 1001 (2007).