THE PROTEIN SOCIETY

Evolutionary dynamics of viral escape under antibodies stress: A biophysical model

Nicolas Chéron,[1,2†] Adrian W.R. Serohijos,[1†] Jeong-Mo Choi,[1] and Eugene I. Shakhnovich[1*]

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138

[2]Département de Biochimie et Centre Robert-Cedergren en Bioinformatique et Génomique, Université de Montréal, Montréal, Quebec, Canada H3T 1J4

Received 18 December 2015; Accepted 2 March 2016
DOI: 10.1002/pro.2915
Published online 00 Month 2016 proteinscience.org

Abstract: Viruses constantly face the selection pressure of antibodies, either from innate immune response of the host or from administered antibodies for treatment. We explore the interplay between the biophysical properties of viral proteins and the population and demographic variables in the viral escape. The demographic and population genetics aspect of the viral escape have been explored before; however one important assumption was the *a priori* distribution of fitness effects (DFE). Here, we relax this assumption by instead considering a realistic biophysics-based genotype-phenotype relationship for RNA viruses escaping antibodies stress. In this model the DFE is itself an evolvable property that depends on the genetic background (epistasis) and the distribution of biophysical effects of mutations, which is informed by biochemical experiments and theoretical calculations in protein engineering. We quantitatively explore *in silico* the viability of viral populations under antibodies pressure and derive the phase diagram that defines the fate of the virus population (extinction or escape from stress) in a range of viral mutation rates and antibodies concentrations. We find that viruses are most resistant to stress at an optimal mutation rate (OMR) determined by the competition between supply of beneficial mutation to facilitate escape from stressors and lethal mutagenesis caused by excess of destabilizing mutations. We then show the quantitative dependence of the OMR on genome length and viral burst size. We also recapitulate the experimental observation that viruses with longer genomes have smaller mutation rate per nucleotide.

Keywords: viral evolution; optimal mutation rate; neutralizing antibodies; folding stability

Introduction

Viruses face the stress of various environments.[1] It has been hypothesized that an adaptive strategy among viral populations to fight this pressure is to modulate their mutation rates[2] and previous works have studied the existence of an optimal mutation rate for viruses.[3] An upper bound to the rate of mutation exists due to lethal mutagenesis whereby mutational load leads to a population decline.[4–6] Lethal mutagenesis is the presumed mode of action of some anti-retroviral drugs, such as ribavirin, that increase the mutation rate to lethal levels.[7,8] A lower bound for mutation rate has been hypothesized for several reasons. First, there is the requirement for an elevated mutation rate to increase the supply of beneficial mutations. These mutations could help the virus to adapt to the changing

Additional Supporting Information may be found in the online version of this article.

†These authors contributes equally to this work.

Grant sponsor: Defense Advanced Research Projects Agency (DARPA); Grant number: HR0011-11-C-0093.

*Correspondence to: Eugene I. Shakhnovich, Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138. E-mail: shakhnovich@chemistry.harvard.edu

Published by Wiley-Blackwell. © 2016 The Protein Society

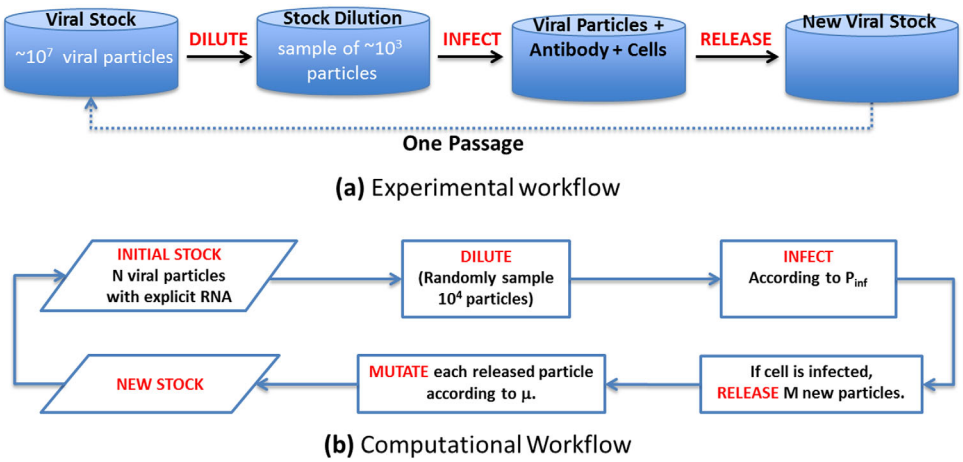


Figure 1. Experimental and computational methodologies to study evolution of viral escape. The computational workflow starts with a stock of N viral particles (with explicit sequences) from which 10^4 are randomly chosen and infect cells according to fitness (which is assumed to map to the probability of infection P_{inf}). If the cell is infected, M new particles are released and each one of them is mutated according to a mutation rate μ, forming a new stock of particles. (a) Experimental workflow. (b) Computational Workflow.

environmental stress that can come from the immune response of the host cell. This scenario is supported by the observed hypermutability of bacteria isolated from natural populations.[9,10] Another potential explanation of the lower bound on the mutation rate is the high cost of maintaining replication fidelity,[2,11,12] although such cost remains to be experimentally demonstrated. The existence of an optimal mutation rate due to the balance for the requirement for beneficial mutations and the avoidance of the lethal mutagenesis have been postulated in the past.[3] There have been several theoretical models on how such an optimal mutation rate could arise.[3,12–15] However, these models are largely phenomenological and lack molecular realism. Where this balance occurs and what is the underlying molecular (or cellular) mechanism is still being debated both from theoretical and experimental standpoints because the structure of the fitness landscape, and consequently the distribution of fitness effects (DFE), is unclear. Assuming the DFE *a priori* is challenging because the DFE itself can change during evolution, as demonstrated in experiments[16] and in forward evolutionary simulations on biophysics-based fitness landscapes (in the context of protein folding and binding landscape[17,18] or in the context of transcription factor binding[19]).

We present here an alternative approach where the fitness landscape of the virus is projected on the biophysical properties of the essential proteins. Indeed, Bloom *et al*. provided direct experimental support of the mapping between folding stability and the fitness of a virus in the case of the neuraminidase of Influenza.[20] The biophysics-based fitness landscape has also been used to explain the measured DFE among viruses[21] and evolution of mutation rates.[18,22] More importantly, the distribution of changes induced by random mutations on key biophysical properties such as folding stability are well-defined and established.[6,23] Thus, the DFE does not need to be an assumption in the evolutionary model. Several groups including our own recently showed that a biophysical approach could account for the influences of mutations on proteins and could have an effect on the distribution of fitness effects and the dynamics of adaptation.[5,6,17,21,24–26] Our goal here is to provide the simplest model, although sophisticated enough at the level of biophysics, to determine the underlying molecular mechanism to investigate the contribution of biophysics to evolution. We investigate in this report the balance between two competing forces (the lethal mutagenesis and the maintenance of supply of beneficial mutations) during the evolution of viral populations under antibodies stress by providing biophysical details into the molecular effects of mutations.

Methods

Biophysical model of viral evolution

To determine the interplay between the viral mutation rate, its demography, and the biophysical properties of viral proteins, we have decided to mimic *in silico* a typical "serial passaging" experiment [Fig. 1(A)]. During these experiments, a fraction of an initial viral stock is added to a medium containing cells and other stressors (e.g., drugs or antibodies). The virus can then infect the cells under an external pressure (the drugs or the antibodies) and new viral particles are released, giving rise to a new stock. These steps constitute a single passage. After multiple passages, the virus titer can then be measured or the virus can be sequenced and compared to the initial viral stock. Our primary concern here is to model experiments of viral evolution that use serial

passaging; a premise is that the results from these kinds of studies are applicable to the evolution of natural populations. What is the appropriate demographic description of natural viral evolution can be debated. There is a claim that periodic bottlenecking resembles the inter-host infection, whereby only a few viral particles are transmitted between hosts.[27–31]

We start the computational model with a stock of N viral particles, from which 10^4 particles are randomly chosen to infect cells according to their fitness.[32] The fitness of a virus is considered to be proportional to its probability of infection P_{inf} which itself is assumed to be a function of the folding stability of all viral proteins and of the affinity to the antibodies. Thus, this model applies to lytic cycles but not to lysogenic cycles where the integration of the viral genome in the host genome would require other parameters in the fitness function. We also assume that the multiplicity of infection (MOI) is equal to 1, meaning that each viral particle can encounter at least one host cell. If a cell is infected, M new particles are released. In our definition, the "burst size" M is the number of progenies after a successful infection of a mammalian cell. During replication, each new particle can be mutated at a specified per genome rate μ.

In our model, the fitness is expressed as the probability of infecting a cell (P_{inf}). We assumed that for a viral particle to infect a cell, the viral proteins must be folded (to be functional) and be free from the neutralizing antibodies (while all proteins encoded by the viral genome must be folded, only one is considered to bind to the antibodies). Considering a thermodynamic equilibrium between the antibodies-binding protein, the antibodies and the complex (with an equilibrium constant of K_{d}), one can write if p viral proteins are encoded by the viral genome (with $\Delta G = G_{\mathrm{folded}}\text{-}G_{\mathrm{unfolded}}$ and $\beta = 1/RT$):

$$P_{inf} = \left(\frac{e^{\beta \Delta G_1}}{1 + e^{\beta \Delta G_1} + e^{\beta \Delta G_1} \left(\frac{[Ab]}{K_d} \right)} \right) \times \prod_{k=2}^{p} \left(\frac{1}{1 + e^{\beta \Delta G_k}} \right) \quad (1)$$

assuming a two-state folding thermodynamics.[33] Here we distinguish between epitope-containing capsid protein, which we denote as the first viral protein, and the rest of viral proteins which have to be folded to function but do not interact with antibodies. The first factor represents the probability that the epitope-containing capsid protein is folded and not sequestered by the antibodies Ab. Assuming that the antibodies can bind only to the folded capsid protein, there are three distinct species: folded and free capsid, folded and bound capsid, and unfolded capsid. Thus, the fraction of folded and bound is determined by the ratio of the Boltzmann probability of being folded and the total probability of being in any of the three states. The second factor is the probability that the remaining *p-1* viral proteins are in their native (functional) states under thermodynamic equilibrium. Thus, we do not rely on an assumption regarding the distribution of fitness effects as it has often been done before. Instead, we rely on a biophysical model of the fitness, for which data can be derived from theory or experiments. Here, the viral fitness depends on two biophysical traits of the viral proteins –the Gibbs free energy of folding of each protein (ΔG) and the binding constant of the capsid protein to the antibodies (K_d)– and on the external pressure (the concentration of antibodies in the medium [Ab]).

In this model, we don't explicitly consider the capsid assembly, which would require additional assumptions on the strength of the interaction between individual capsid domains. The equation also assumes that the antibodies could target only one of the capsid domains. The motivation for the latter assumption is that each antibody is large and it probably occludes other antibodies from binding.

Biophysical effects of mutations

When a mutation occurs in the viral genome, we estimate its effect on the folding Gibbs free energy in the following way: $\Delta G_{\mathrm{mutant}} = \Delta G_{\mathrm{wildtype}} + \Delta\Delta G$. The free energy of folding of the wild type ($\Delta G_{\mathrm{wildtype}}$) for all proteins is set initially to -5.0 kcal/mol, which is close to the average folding stability of real proteins.[34] This initial choice is not relevant to the overall results because the population will equilibrate to the folding stability dictated by mutation-selection balance.[35] The biophysical effect $\Delta\Delta G$ is drawn from a normal distribution with a mean value of $\Delta\Delta G_{mean} = -0.13(\Delta G_{mean}) + 0.23$ kcal/mol and a standard deviation of $\Delta\Delta G_{\mathrm{sd}} = 1.7$ kcal/mol (Ref. 6 and 23). These values are derived from the distribution of experimentally measured ΔG and $\Delta\Delta G$ values collated in the ProTherm database.[34,36] The linear dependence between $\Delta\Delta G_{\mathrm{mean}}$ and ΔG is also derived from the ProTherm database.[36] It models the observation that when the wild-type protein is more stable, there are fewer sequences that can lead to stabilization so arising mutations are on average more destabilizing (sequence depletion). That is, the more stable the protein is, the more likely it is to find destabilizing mutations.[36]

Whereas all mutations can affect the free energy of folding, it is not the case for the binding affinity. The average number fraction of residues that participate in protein-protein interactions is ~15% according to a bioinformatics study.[37] Thus, we assumed that the probability that a mutation hits a residue belonging to the epitope of the viral capsid epitope-containing protein is 0.15. We also assumed that the perturbation of the protein−protein interaction is statistically analogous to mutational effect on the

folding Gibbs free energy (i.e., the distribution of $\Delta\Delta G_{\text{binding}}$ is the same as the one for $\Delta\Delta G_{\text{folding}}$):

$$K_{d,\text{mutant}} = K_{d,\text{wildtype}} \exp\left(\beta \Delta\Delta G_{binding}\right) \quad (2)$$

This equation reflects the assumption that the energetics that drive protein−protein interaction are largely similar to energetics that drive protein folding. Indeed, the two distributions of $\Delta\Delta G_{\text{binding}}$ and $\Delta\Delta G_{\text{folding}}$ have similar features in that most mutations are deleterious—they mostly destabilize proteins and perturb protein-protein interaction. Miller *et al*. found through exhaustive computational mutagenesis of the viral capsid of bacteriophage QX174 that the two distributions are similar.[38] Thus, the central part of our algorithm is an assumption regarding the distribution of biophysical effects of mutations. The average influence of ΔG and of K_{d} on P_{inf} is shown in Supporting Information (see Supporting Information Fig. S1). One must note that there can be a trade-off between binding and folding free energies. Manhart *et al*. found that during evolution different regimes can be observed depending on the values of $\Delta G_{\text{binding}}$ and $\Delta G_{\text{folding}}$. For example, when $\Delta G_{\text{binding}}$ is low and $\Delta G_{\text{folding}}$ is high, there is a need to preserve the folding stability and epistasis plays a strong role to evolve between sequences.[39,40] This can also help to stabilize marginally stable proteins.

Model system

Our model is not explicitly sequence-dependent because the effects of mutations are inferred from the distributions of $\Delta\Delta G_{\text{folding}}$ and $\Delta\Delta G_{\text{binding}}$, but we account for the number of proteins and the genome length. We have used values for the murine norovirus for which three proteins are important during infection: VP_1 (541 amino-acids), VP_2 (208 amino-acids) and VP_g (124 amino-acids).[41] The remaining proteins are only expressed once the virus has penetrated into the cell and were not considered in the model [thus, a value of $p = 3$ was used, see Eq. (1)]. Should VP_1 or VP_2 be unfolded, the capsid would be degraded. Should VP_g be unfolded, the genome would not be able to replicate. Thus, the three considered proteins are essential for the virus. Realistic values of burst size for RNA viruses are in the range of 50 to ~1700 (Ref. 44); however, to save computational time a value of $M = 100$ was used unless otherwise noted.

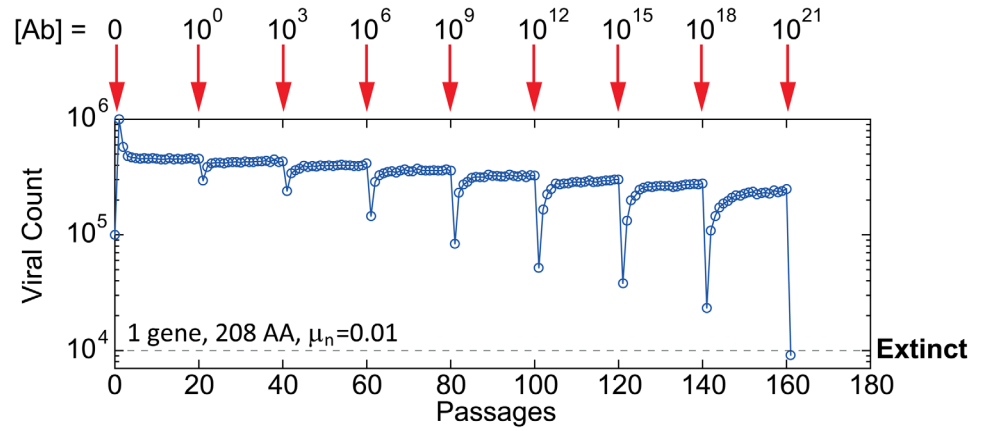
Results

Evolution under antibodies stressor

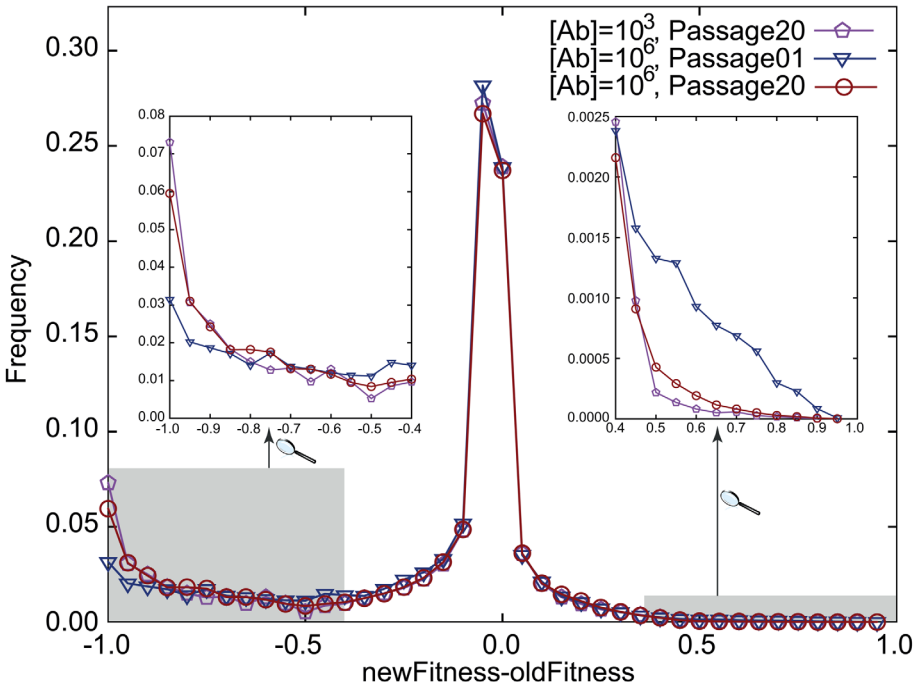
We first show the dynamics of the viral population aimed at mimicking the serial passaging experiments (Supporting Information Fig. S2). Starting with 10^5 viral particles and no external pressure ($[Ab]=0$), the viral count first increases and quickly equilibrates. Since we start the simulations with very stable proteins ($\Delta G_{\text{folding}} = -5$ kcal/mol), P_{inf} is close to 1 and all cells are infected; thus, the viral stock quickly increases. However, the dynamics of this initial equilibration is irrelevant because the following results are a consequence of the evolutionary dynamics already under equilibrium. After some time, the population starts migrating towards mutation-selection balance by acquiring deleterious mutations and stays stable for the 100 first passages (Supporting Information Fig. S2). Upon addition of antibodies (at passage 101), the viral count dramatically drops, but then some particles find beneficial mutations that eventually lead to viral escape from antibodies. After around 10 passages, the viral count reaches again a new steady state (Supporting Information Fig. S2). When the external antibodies pressure is too high, the population goes extinct (see Fig. 2). We consider the population extinct when the titer is below the "bottleneck" count, which is the number of particles passaged to the next generation (10^4 particles). For a burst size of $M = 100$, a population goes extinct when less than 100 cells are infected since it will give rise to less than 10^4 released virus particles. Expectedly, as the concentration of antibodies is increased, the viral titer keeps dropping and equilibrating. However, the cycle of administration of the antibodies affects the population dynamics of the viruses. Progressive increase of the antibodies concentration allows the population to adapt before being stressed by another increase of the amount of antibodies. Thus the population can tolerate higher antibodies concentration.

The novelty of our model lies in the nonfixed distribution of fitness effects. Thus, the distribution of the selection coefficient (defined as the probability of infection before and after mutation, $s = P_{inf}^{after} - P_{inf}^{before}$) at different passages is different. We present in Figure 2(B) the distribution of selection coefficient for the last passage with $[Ab]=10^3$, the first passage with $[Ab]=10^6$, and the last passage with $[Ab]=10^6$ from Figure 2(A). When the concentration of antibodies increases (from purple pentagons to blue triangles), the system is no more equilibrated. Thus, the fraction of lethal mutations with $s < 0$ decreases (see insert on the left) and the fraction of beneficial mutations with $s > 0$ increases (see insert on the right). After 20 passages without changing the environment (from blue triangles to red circles), the system had enough opportunity to adapt and the distribution of selection coefficient recovers its shape prior to the subsequent change of antibodies concentration. The two equilibrated distributions (purple pentagons and red circles) are very similar, but they are not exactly the same. This explains why the virus titer is not completely recovered after each antibodies addition, and keeps decreasing progressively.

To generalize this result, we explored the onset of extinction for various combinations of mutation rate and antibodies concentration. For a given value



(a) Trajectory of viral evolution with an increasing pressure of [Ab]=10^3.



(b) Evolution of the distribution of selection coefficient (inserts are small regions of the main graph).

Figure 2. Viral evolution under an increasing pressure. In panel (a), the viral count first increases because we start the simulations with very stable proteins and P_{inf} is close to 1. The viral count then quickly equilibrates. When antibodies are "added", the viral count dramatically drops under the pressure before reaching again a steady-state. As the concentration of antibodies is increased, the viral titer keeps dropping and equilibrating. When *[Ab]* is too high, the viral count drops to the extinction. In panel (b), we can observe that the distribution of selection coefficient is not fixed depending on the conditions. When the concentration of antibodies increases (last passage with [*Ab*]=10^3 in purple pentagons to first passage with [*Ab*]=10^6 in blue triangles), the system is no more equilibrated. Thus, the fraction of lethal mutations decreases (insert on the left) and the fraction of beneficial mutations increases (insert on the right). After 20 passages without changing the environment (last passage with [*Ab*]=10^6 in red circles), the system had enough opportunity to adapt and the distribution of selection coefficient recovers its shape prior to the subsequent change of antibodies concentration. (a) Trajectory of viral evolution with an increasing pressure of [*Ab*]=10^3. (b) Evolution of the distribution of selection coefficient (inserts are small regions of the main graph).

of the mutation rate μ, we determined the highest concentration of antibodies that still leads to viral escape. First, we performed evolutionary simulations to allow the population to reach mutation-selection balance at the chosen mutation rate but at zero antibody concentration. This equilibration phase was performed for 20 passages since we observed that this value is enough to equilibrate the virus stock (see for example Supporting Information Fig. S2 where it can be seen that the steady-state has been reached at passage 20). Then, we stressed the population at the chosen antibodies concentration. Extinction or survival is evaluated after 5 passages; we found that 5 passages are enough to observe the

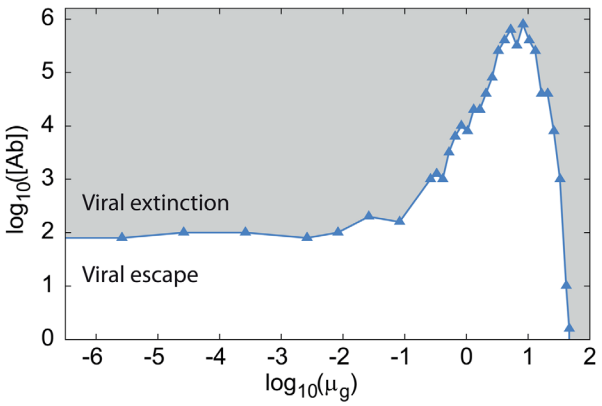


Figure 3. Phase diagram of virus survival under antibodies stressor. At each mutation rate, the highest amount of antibodies that still leads to viral escape after five passages is determined, indicating how much pressure the virus can endure and still survive. The optimal mutation rate for an RNA virus is found to be around $\log_{10}(\mu_g)=0.92$ or 8.3 mutations per genome per replication.

survival of a viral population at a given antibodies concentration. If the virus has escaped, the viral stock increases at each step and there is no need to go beyond 5 passages.

The value of antibodies concentration at which the virus escapes is an indication of how much pressure it can endure and still survive. By varying the mutation rate μ, we determined the phase diagram outlining the conditions at which we observed extinction and escape of the virus under antibodies stress (see Fig. 3). Under antibodies pressure, increasing the mutation rate increases the likelihood of acquiring mutations that lower the binding free energy of the protein-antibodies interaction, and then lead to escape. However, at sufficiently high mutation rates, further increase in mutation rate leads to lethal mutagenesis, because arising mutations decrease folding stability, and subsequently fitness by lowering the fraction of folded proteins P_f as given by (1) (Ref. 5,6,17,21,24). Under very low mutation rate, proteins are more stable, but the chance of acquiring escape mutation is decreased. Thus, there is a balance between lethal mutagenesis and maintenance of supply of escape mutations which is quantified by the phase diagram in Figure 3. We find that under our model (which provides biophysically realistic estimates of the molecular effects of mutations) there exists a mutation rate at which virus can sustain highest stress. This mutation rate for an RNA virus in our model is found to be around $log_{10}(\mu_g)=0.92$ or 8.3 mutations per genome per replication (for a burst size of $M=100$). This value is somewhat above the range of experimentally observed mutation rates for RNA virus (see Table S1 in Supporting Information and Discussion below). It must be pointed out that the existence of a mutation rate allowing a maximal amount of stress does not prove that the population can reach this mutation rate, neither that this mutation rate is always optimal in a broad range of conditions; such a study will be the subject of a following paper. The lethal mutagenesis threshold is calculated to be $log_{10}(\mu_g)=1.69$ or 48.8 mutations per genome per replication, which is higher than previously reported maximum value of lethal mutagenesis:[5,6,21] one must note that in these previous models, the burst size corresponding to conservative or semi-conservative replication was fixed at $M=2$. Thus, the lethal mutagenesis occurs here at a higher value of the mutation rate due to much greater burst size compared to previous models.

Influence of the genome size

We next determined the effect of the genome size on the optimal mutation rate. In addition to the norovirus model with three genes detailed previously, we also performed simulations with either only one of the genes or with a model where some of the norovirus genes are duplicated to increase the genome length. The phase diagrams for four representative models are presented in Supporting Information (Supporting Information Fig. S3). The peak in the phase diagram with respect to the mutation rate per genome is almost constant; however, for the mutation rate per nucleotide, an influence of the size of the genome is observed. For example, for 124 amino acids, it is found to be $\log_{10}(\mu_n)=-2.0$; for 208 amino acids, it is $\log_{10}(\mu_n)=-1.75$; for 541 amino acids, it is $\log_{10}(\mu_n)=-2.4$; for 873 amino acids, it is $\log_{10}(\mu_n)=-2.5$. Thus, it appears that the longer the genome is, the lower the optimal mutation rate ($R=-0.90$, p-value $=0.006$). This result reproduces experimental trend, as shown in Figure 4 where we report experimentally observed (Ref. 44 and Supporting Information Table S1) and computed mutation rates allowing maximal stress for RNA virus. Two viruses

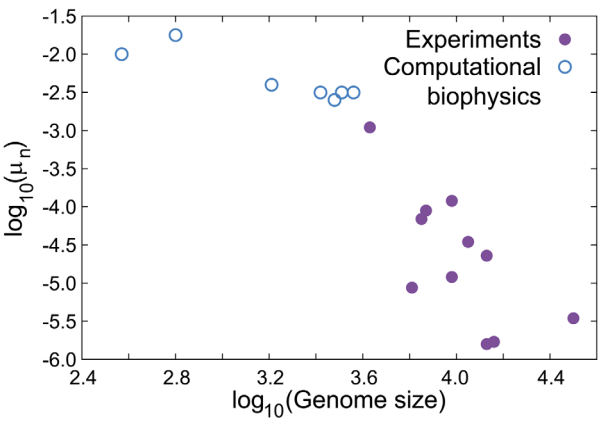


Figure 4. Influence of the genome length on the viral mutation rate. Computed mutation rates with maximal stress (our model) are in blue open-circle whereas experimentally observed mutation rates are in purple filled-circle. For the computational data, $R=-0.90$ and p-value $=0.06$.

reported in reference[42] are excluded in Supporting Information Table S1 and Figure 4 because these are bacteriophages and not interact with antibodies. Nonetheless, there is an equivalent pressure for bacteriophages that manifests as an evolutionary "arms race model".[43,44] Thus, we envision that there is also an optimal mutation rate for these systems, but the underlying biophysics (such as protein-DNA interaction) could be different. Note that the longest genome we simulated is 1206 amino-acids long, and we did not extend our study to longer genomes because of computational cost to perform simulations with longer genomes.

Influence of the burst size

Larger burst sizes increase the number of viral particles, which effectively lowers the chance of extinction. Thus, we also explored the effect of burst size on the phase diagram (Fig. 5, Supporting Information Fig. S4). At higher burst sizes, the onset of lethal mutagenesis is delayed and the viral population is able to tolerate higher concentration of antibodies: when a single viral particle gives rise to more progeny, the chances to find a stabilizing mutation increase, and a higher maximal mutation rate for the lethal mutagenesis is achieved. Interestingly, the effect of burst size on the onset of lethal mutagenesis (from 12.0 to 48.8 mutations per genome replication) is greater than its effect on the optimal mutation rate (Fig. 5). Indeed, it appears that the burst size (M) has a small influence on the value of this mutation rate (μ_g) –ranging from 4.2 to 8.3 mutations per genome replication.

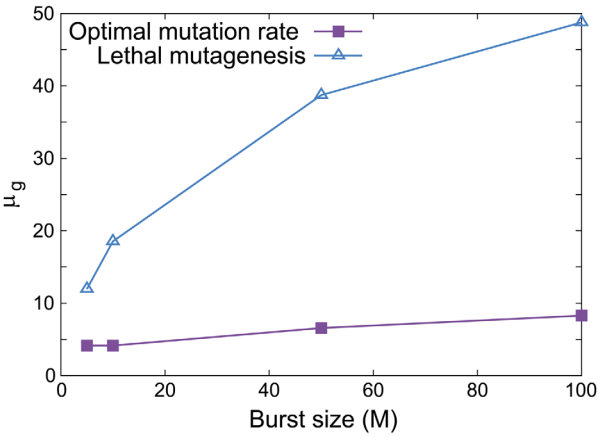


Figure 5. Influence of the burst size on the optimal mutation rate (blue open-triangle) and on the lethal mutagenesis threshold (purple filled-square). The lethal mutagenesis threshold is defined as the value for which the virus cannot perform 20 passages of equilibration with *[Ab]*=0.

Discussion and Conclusion

We presented here a simple biophysics-based model of dynamic of viral evolution that quantitatively explores the balance between lethal mutagenesis and the requirement for beneficial mutations that allows escape from antibodies. Contrary to previous models of viral evolution that *a priori* assume a fixed distribution of fitness effects of mutations,[4,45] we instead base our assumptions on the biophysical genotype-phenotype relationship and distribution of biophysical effects of mutations. As a consequence, the distribution of selection coefficient depends on the status of the system (is it equilibrated or not) and on the environment, reflecting epistasis in the viral fitness landscape. The distributions that we have used to describe folding and binding free energies were derived from a lot of protein types, which can be seen as a limitation. However, the distribution of biophysical effects can be explored computationally[23] and experimentally[34] for some specific proteins. Thus, if one restrains data to one family of proteins, more accurate results can be obtained. First, using this biophysics-based model of evolution, we find that the survival and viability of the viral population under antibodies stress is optimal between 4.2 and 8.3 mutations per genome per replication (for a burst size between 5 and 100). This value can indeed be explained as a balance between lowering the binding affinity with the antibodies and keeping the viral proteins stable enough. At low mutation rate, there is only a small likelihood of finding a mutation that could strongly perturb the binding of the capsid protein to the antibodies (strongly beneficial mutation). On the other hand, when the mutation rate is too high, some mutations could perturb the binding of the capsid to the antibodies while other mutations could compromise the folding stability of any viral protein.

The mutation rate of RNA virus is commonly estimated to be approximately 1 mutation per genome per replication.[46,47] We report in Supporting Information Table S1 the experimentally observed mutation rates of some RNA viruses.[42] The average mutation rate per genome for these RNA viruses is $\mu_g = 0.37$, ranging from 0.024 to 1.2 mutations per genome per replication. Our model of viral evolution is simple and does not account for other biological factors that could influence mutation rates or their fitness effects (for example, cellular quality control response of host cells, the cost of protein production, selection for optimal codon or secondary RNA structures). Thus, while our model predicts a mutation rate that allows a maximal amount of stress, the exact values of these mutation rates may differ from the experimental values of actual viral mutation rates given in Table S1. There is a broad (approximately 2 orders of magnitude) variation in the measured and estimated values of mutation rates (Supporting Information Table S1),[42] presumably due to confounding factors such as effective population size and genome length. Indeed, in agreement with earlier works,[14,42,48–50] we found that the mutation rate per nucleotide that allows a maximal amount of stress decreases when the genome length increases (Fig. 4).

We also showed the dependence of the mutation rate which allows maximal stress on burst size (Fig. 5; see also Supporting Information Fig. S4). The dependence of the optimal mutation rate on burst size is due to the increased likelihood at higher burst sizes of finding a variant that acquired a beneficial mutation. In a similar vein, the higher burst size results in an effective increase in fitness because of higher number of progenies.

Under mutation-selection balance, the evolution against the antibodies should proceed towards the near-neutral regime.[35] In the review article of Wang and Bull,[51] they pointed out that fitness has three components: burst size, adsorption rate at which viruses could interact with cells, and lysis time (time interval between infect and burst). In our model, we account for adsorption rate (which is proportional to the probability of infection) and burst size, but not lysis time. Our main goal here is to immediately connect to biophysics, thus we mapped the fitness to the adsorption rate (P_{inf}). A possible compensation between lysis time and burst size will require additional assumptions,[51] which will be the subject of future works. It must also be pointed out that the current report assumes monoclonal antibodies. A very interesting extension of the study will be to include other critical players: host cells, viruses, and B-cells in the germinal centers.[52] Indeed, a biophysics-based evolutionary model of the immune response could uncover its physical basis. In the future, other stressors could also be considered in the model presented here (such as inhibitors and pH) because effects of mutations on these quantities can be estimated.

Acknowledgments

The authors thank Andrew B. Feldman, Jeffrey S. Lin, Assaf Rotem, Thomas J. Smith and Christiane E. Wobus for fruitful discussions. They also thank Amy Gilson for comments on the manuscript. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

References

1. Meyers LA, Bull JJ (2002) Fighting change with change: adaptive variation in an uncertain world. Trends Ecol E 17:551–557.
2. Combe M, Sanjuan R (2014) Variation in RNA virus mutation rates across host cells. PLoS Pathog 10: e1003855.
3. Kamp C, Wilke CO, Adami C, Bornholdt S (2002) Viral evolution under the pressure of an adaptive immune system: optimal mutation rates for viral escape. Complexity 8:28–33.
4. Bull JJ, Sanjuan R, Wilke CO (2007) Theory of lethal mutagenesis for viruses. J Virol 81:2930–2939.
5. Chen P, Shakhnovich EI (2009) Lethal mutagenesis in viruses and bacteria. Genetics 183:639–650.
6. Zeldovich KB, Chen P, Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. Proc Natl Acad Sci USA 104:16152–16157.
7. Crotty S, Maag D, Arnold JJ, Zhong W, Lau JY, Hong Z, Andino R, Cameron CE (2000) The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. Nat Med 6:1375–1379.
8. Crotty S, Andino R (2002) Implications of high RNA virus mutation rates: lethal mutagenesis and the antiviral drug ribavirin. Microbes Infect 4:1301–1307.
9. Oliver A, Canton R, Campo P, Baquero F, Blazquez J (2000) High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. Science 288:1251–1254.
10. Bjorkholm B, Sjolund M, Falk PG, Berg OG, Engstrand L, Andersson DI (2001) Mutation frequency and biological cost of antibiotic resistance in Helicobacter pylori. Proc Natl Acad Sci USA 98:14607–14612.
11. Furio V, Moya A, Sanjuan R (2005) The cost of replication fidelity in an RNA virus. Proc Natl Acad Sci USA 102:10233–10237.
12. Sniegowski P (2001) Evolution: constantly avoiding mutation. Curr Biol 11:R929–R931.
13. Regoes RR, Hamblin S, Tanaka MM (2013) Viral mutation rates: modelling the roles of within-host viral dynamics and the trade-off between replication fidelity and speed. Proc R Soc B Biol Sci 280:2012–2047.
14. Sniegowski PD, Gerrish PJ, Johnson T, Shaver A (2000) The evolution of mutation rates: separating causes from consequences. Bioessays 22:1057–1066.
15. Raynes Y, Sniegowski PD (2014) Experimental evolution and the dynamics of genomic mutation rate modifiers. Heredity 113:375–380.
16. Silander OK, Tenaillon O, Chao L (2007) Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. PLoS Biol 5:e94.
17. Serohijos AW, Shakhnovich EI (2014) Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding regions. Mol Biol E 31:165–176.
18. Heo M, Shakhnovich EI (2010) Interplay between pleiotropy and secondary selection determines rise and fall of mutators in stress response. PLoS Comput Biol 6: e1000710.
19. Mustonen V, Lassig M (2009) From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. Trends Genet 25:111–119.
20. Gong LI, Suchard MA, Bloom JD (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. eLife 2:e00631.
21. Wylie CS, Shakhnovich EI (2011) A biophysical protein folding model accounts for most mutational fitness effects in viruses. Proc Natl Acad Sci USA 108:9916–9921.
22. Heo M, Kang L, Shakhnovich EI (2009) Emergence of species in evolutionary "simulated annealing". Proc Natl Acad Sci USA 106:1869–1874.
23. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS (2007) The stability effects of protein mutations appear to be universally distributed. J Mol Biol 369:1318–1332.
24. Goldstein RA (2011) The evolution and evolutionary consequences of marginal thermostability in proteins. Proteins 79:1396–1407.

25. Bloom JD, Raval A, Wilke CO (2007) Thermodynamics of neutral protein evolution. Genetics 175:255–266.
26. Sikosek T, Chan HS, Bornberg-Bauer E (2012) Escape from adaptive conflict follows from weak functional trade-offs and mutational robustness. Proc Natl Acad Sci USA 109:14888–14893.
27. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, García-Sastre A, tenOever BR (2014) Influenza A virus transmission bottlenecks are defined by infection route and recipient host. Cell Host Microbe 16:691–700.
28. Artenstein MS, Miller WS (1966) Air sampling for respiratory disease agents in army recruits. Bacteriol Rev 30:571–572.
29. Gerone PJ, Couch RB, Keefer GV, Douglas RG, Derrenbacher EB, Knight V (1966) Assessment of experimental and natural viral aerosols. Bacteriol Rev 30:576–588.
30. Rubin LG (1987) Bacterial colonization and infection resulting from multiplication of a single organism. Rev Infect Dis 9:488–493.
31. Nowak MA, Anderson RM, McLean AR, Wolfs TF, Goudsmit J, May RM (1991) Antigenic diversity thresholds and the development of AIDS. Science 254:963–969.
32. Wobus CE, Thackray LB, Virgin HWT (2006) Murine norovirus: a model system to study norovirus biology and pathogenesis. J Virol 80:5104–5112.
33. Privalov PL (1979) Stability of proteins: small globular proteins. Adv Prot Chem 33:167–241.
34. Kumar MD, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, Uedaira H, Sarai A (2006) ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. Nucleic Acids Res 34: D204–D206.
35. Serohijos AW, Shakhnovich EI (2014) Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. Curr Opin Struct Biol 26:84–91.
36. Serohijos AW, Rimas Z, Shakhnovich EI (2012) Protein biophysics explains why highly abundant proteins evolve slowly. Cell Rep 2:249–256.
37. Hu Z, Ma B, Wolfson H, Nussinov R (2000) Conservation of polar residues as hot spots at protein interfaces. Proteins 39:331–342.
38. Miller CR, Lee KH, Wichman HA, Ytreberg FM (2014) Changing folding and binding stability in a viral coat protein: a comparison between substitutions accessible through mutation and those fixed by natural selection. PLoS One 9:e112988.
39. Manhart M, Morozov AV (2013) Path-based approach to random walks on networks characterizes how proteins evolve new functions. Phys Rev Lett 111:088102.
40. Manhart M, Morozov AV (2015) Protein folding and binding can emerge as evolutionary spandrels through structural coupling. Proc Natl Acad Sci USA 112:1797–1802.
41. Hardy ME (2005) Norovirus protein structure and function. FEMS Microbiol Lett 253:1–8.
42. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R (2010) Viral mutation rates. J Virol 84:9733–9748.
43. Labrie SJ, Samson JE, Moineau S (2010) Bacteriophage resistance mechanisms. Nat Rev Microbiol 8: 317–327.
44. Stern A, Sorek R (2011) The phage-host arms race: shaping the evolution of microbes. Bioessays 33:43–51.
45. Badgett MR, Auer A, Carmichael LE, Parrish CR, Bull JJ (2002) Evolutionary dynamics of viral attenuation. J Virol 76:10524–10529.
46. Drake JW (1999) The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Ann NY Acad Sci 870:100–107.
47. Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics 148: 1667–1686.
48. Anderson JP, Daifuku R, Loeb LA (2004) Viral error catastrophe by mutagenic nucleosides. Annu Rev Microbiol 58:183–205.
49. Gupta A, LaBar T, Miyagi M, Adami C (2016) Evolution of genome size in asexual populations. arXiv 1511.05548.
50. Lynch M, Conery JS (2003) The origins of genome complexity. Science 302:1401–1404.
51. Bull JJ, Wang IN (2010) Optimality models in the age of experimental evolution and genomics. J Evol Biol 23:1820–1838.
52. Muyoung H, Zeldovich KB, Shakhnovich EI (2011) Diversity against adversity: how adaptive immune system evolves potent antibodies. J Statist Phys 144:241–267.