Journal of Statistical Physics (2021) 182:38
https://doi.org/10.1007/s10955-021-02716-2

MANUSCRIPT

Studying Viral Populations with Tools from Quantum Spin Chains

Saumya Shivam[1] · Christopher L. Baldwin[2,3] · John Barton[4] · Mehran Kardar[5] · S. L. Sondhi[1]

Received: 19 March 2020 / Accepted: 28 January 2021 / Published online: 10 February 2021

Abstract
We study Eigen's model of quasi-species (Eigen in Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58(10):465, 1971), characterized by sequences that replicate with a specified fitness and mutate independently at single sites. The evolution of the population vector in time is then closely related to that of quantum spins in imaginary time. We employ multiple perspectives and tools from interacting quantum systems to examine growth and collapse of realistic viral populations, specifically considering excessive mutations in certain HIV proteins. All approaches used, including the simplest perturbation theory, give consistent results.

Keywords Biological evolution · Quantum spin chains · HIV

1 Introduction

A central concept in population genetics is an idealized fitness F, which (in the absence of competition, resource limitation, etc.) governs the exponential growth in the number of individuals N according to $(dN/dt) = FN$ [19]. However, mutations diversify the genetic make-up (genotype) of the population, and modify its overall fitness. Eigen introduced the concept of **quasi-species** to describe the 'cloud' of closely related genotypes. Eigen's model considers a population composed of a set of sub-populations (labelled $\{a\}$), each contributing N_a individuals that coexist in a rapidly evolving larger community of (varying) total size $N = \sum_a N_a$ [13]. Mutations between quasi-species further diversify the composition of the

Communicated by Michael Lässig.

✉ Saumya Shivam
sshivam@princeton.edu

1 Department of Physics, Princeton University, Princeton, NJ 08544, USA

2 National Institute of Standards and Technology, Gaithersburg, MD 20899, USA

3 Joint Quantum Institute, University of Maryland, College Park, MD 20742, USA

4 Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA

5 Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

population. If the sub-populations $\{a\}$ are assigned fitness values $\{F_a\}$, and mutations from sub-population a to b occur at rates W_{ba}, the makeup of the population changes over time according to

$$\frac{dN_a}{dt} = F_a N_a + \sum_{b \neq a} [W_{ab} N_b - W_{ba} N_a] . \tag{1}$$

In terms of the fractions $x_a = N_a/N$, the total population size $N = \sum_a N_a$ grows as $(dN/dt) = \overline{F} N$. The *mean fitness* $\overline{F} = \sum_a F_a x_a$, must be non-negative for the quasi-population to survive at long times.[1]

At the molecular level, genetic information is maintained in the sequence of bases of nucleotides (e.g. in the sequence of the DNA or RNA of a virus). For modeling purposes, we assume that this information appears as a sequence of $L \gg 1$ characters (e.g. one of 4 nucleotides of DNA, or 20 amino acids of a protein). For simplicity of presentation (as well as practicality, as in the case of HIV proteins described below) we adapt a binary representation, $n_i = (0, 1)$ or $\sigma_i^z = (-1, +1)$, for $i = 1, 2, \cdots, L$, to indicate whether the site i is in its wild-type (consensus) or a mutated state. In this representation of a genotype, its fitness is some function of its sequence $F_a = F[\{\sigma_i^z\}]$. Note that by our definition of a consensus sequence, it is possible for a sequence to have a higher fitness than the consensus sequence, since the consensus is based only on single-site frequencies.

In a simple model of mutations, the state of each site (independently of other sites) changes away from consensus at the rate μ_f, and reverts back to wild-type at the rate μ_b. Equation (1) is linear, and its right hand side can be regarded as representing the action of a matrix H on a (column) vector $\mathbf{N}$ containing the sub-population sizes $\{N_a\}$. For $\mu_f = \mu_b = \mu$, the action of mutations on site i can be represented by the Pauli matrix $\mu\sigma_i^x$, such that

$$\frac{d\mathbf{N}}{dt} = -H\mathbf{N} , \quad \text{with} \quad H = -F[\{\sigma_i^z\}] - \mu \sum_{i=1}^{L} (\sigma_i^x - \mathbf{1}) . \tag{2}$$

The above equation provides an analogy to the imaginary-time evolution of spins in a quantum chain, governed by the Hamiltonian H. The fitness function $F[\{\sigma_i^z\}]$ corresponds to interactions between spins, while mutations are implemented through the transverse magnetic field μ.

The analogy between Eigen's quasi-species model and quantum chains has been noted and explored in a number of references [2–4,22,32,37]. The behaviour of genetic structure of a quasi-species model and magnetization of the corresponding quantum spin model for some idealized fitness landscapes (like an Ising chain or a mean field Hamiltonian) have been studied in Refs. [2,3,37]. Hermisson et al. [22] generalize the binary representation of DNA sequences to the four nucleotides, leading to a corresponding four state quantum chain. Saakian and Hu [32] consider mutations between sequences separated by more than one Hamming distance (see also Ref. [31]). As is well known in physics literature, a quantum system in d dimensions is equivalent to a classical system in $d + 1$ dimensions with discretized time. Leuthäusser [24] explored a similar analogy between the Eigen model and a two dimensional Ising system.

There are, however, important differences between quasi-species evolution, and the dynamics of a quantum chain (in real time):

- While the Hamiltonian H is real and symmetric, the evolution of $\mathbf{N}$ is not unitary, in that the overall population size $N(t) = \sum_a N_a(t)$ changes as a function of time. By

[1] This simple *mutation–selection* model ignores several issues such as competition between different strains, and stochastic fluctuations. It thus assumes large populations unconstrained by resources.

contrast, time variation according to $(d\mathbf{N}/dt) = -iH\mathbf{N}$ preserves the norm of the vector $\mathbf{N}$, such that $\{|N_a|^2\}$ can be regarded as probabilities. The natural set of probabilities for the evolving population are the proportions $\{x_a(t) = N_a(t)/N(t)\}$.

- Even the symmetry of H is an artifact of the simplification $\mu_f = \mu_b = \mu$. In the biological context, it is more likely that forward and mutation rates are not the same, leading to the replacement of $\mu(\sigma_i^x - \mathbf{1})$ for $\mu_f \neq \mu_b$ with the asymmetric matrix $\mu_f\sigma^+ + \mu_b\sigma^-$ [32].

Despite these differences, there are aspects of the dynamics that are common to the two systems, in certain special cases. For quantum systems at low temperature, one is often interested in the low energy properties, which are determined by the eigenvectors of H corresponding to the lowest eigenvalues. Similarly, the long-time behavior in the quasi-species evolution is governed by the few largest eigenvectors of the matrix $-H$, i.e., the exact same eigenstates. In particular, the ground state energy is (minus) the mean fitness, while the ground state vector characterizes the prevalence of mutations in the quasi-species population. Together, they determine the eventual fate of the quasi-species as described below:

• **Error threshold** Eigen introduced the concept of an error catastrophe [13] by considering a fitness function with a single peak on a uniform background (for one or more mutations), which can be described by

$$F_{\text{Eigen}} = F_\mu + (F_0 - F_\mu)\prod_{i=1}^{L}\frac{1-\sigma_i^z}{2}. \tag{3}$$

In words, the wild type has fitness F_0 while any mutation reduces the fitness to $F_\mu < F_0$. Upon increasing the mutation rate μ, there is a transition when the fraction of population in the fit state decreases dramatically. This is a genuine singularity (phase transition) in the limit $L \to \infty$, although the threshold μ_c^E scales as $1/L$ (the superscript E in μ_c^E is used to distinguish from the fatal mutation rate, discussed next) . More precisely, one can show that for $\mu < \mu_c^E = (F_0 - F_\mu)/L$, the fraction x_ℓ of the population having ℓ mutations falls off exponentially as

$$x_\ell \sim \left(\frac{\mu}{\mu_c^E}\right)^\ell, \tag{4}$$

whereas for $\mu > \mu_c^E$, nearly the entire population has roughly $L/2$ mutations.

The fitness function in Eigen's case is of course highly artificial, and questions remain as to whether the concept of error threshold is applicable to more realistic landscapes (see also Ref. [16]). Nonetheless, sharp transitions like this, in which the relevant quantity is zero on one side and non-zero on the other, are in fact quite common in the context of quantum spin systems. Examples include the transition between ferromagnetic and paramagnetic equilibrium phases [33,36], frozen versus thermalizing spin glasses [5,6,23], and many-body localization-delocalization transitions [1,27,28].

• **Fatal mutation load** In Eigen's model, the quasi-species population can still grow if $F_\mu > 0$, as each sequence is viable. However, the population can also disappear upon increased mutation because of the reduction in overall fitness. As an example, consider the so-called Mount Fuji Landscape [3]:

$$F_{\text{MFL}} = -LF_0 - h\sum_{i=1}^{L}\sigma_i^z, \tag{5}$$

specifically with $0 < F_0 < h$. The most fit sequence has a fitness $L(h-F_0)$, with any mutation (independently) carrying a load of $2h$. With increasing mutation rate, the quasi-species cloud

acquires more mutations, but (unlike the case of error threshold) remains anchored to the wild-type state. Since the problem is equivalent to independent quantum spins, it is easy to identify the ground state as corresponding to all spins tilted along $(-h, \mu)$, leading to the mean fitness of $\bar{F}(\mu) = L(\sqrt{h^2 + \mu^2} - \mu - F_0)$. For any mutation rate μ larger than $\mu_c = (h^2 - F_0^2)/2F_0$, $\overline{F}(\mu)$ is negative and the population will die off. We should also point out that the mutation load can be dramatically altered in asexual populations due to a different mechanism of clonal interference [17,21], where sequences with high fitnesses compete with each other.

The above two routes to population collapse have quite distinct signatures in the quantum perspective. Loss of reproductive capacity in the second case corresponds to the ground state eigenvalue crossing zero. However, eigenstates of the chain are independent of the choice of overall constant (the parameter F_0) and do not exhibit any change at the fatal mutation rate μ_c. Eigen's error threshold (at μ_c^E), however, is accompanied by a dramatic change (delocalization) of the ground state, even though the ground state eigenvalue may have either sign depending on F_0. The change in eigenstates is a function of the full Hamiltonian (fitness function), and it is still not entirely clear to what extent actual viral landscapes exhibit such behavior.

In this work, we instead investigate fatal mutation load and the corresponding rate μ_c for different types of HIV proteins, using traditional tools for interacting quantum spins like perturbation theory and mean field approximation. In the next section we review the process of arriving at the fitness function (Eq. 2) using the observed prevalence of different sequences. In Sect. 3, we use perturbation theory and the mean field approximation to arrive at the fatal mutation rate μ_c for the HIV proteins.

2 Fitness Using Prevalence Landscape

Several other forms of fitness landscape have been proposed and explored in the literature. We would like to explore the quantum analogy, identifying mechanisms that lead to population collapse (e.g. by error catastrophe or high mutation load). To this end, we will focus on a particular class of landscapes constructed from the prevalence of observed sequences for a virus.

The massive size of the space of possible protein sequences makes it impossible to estimate virus prevalence simply by counting sequences. Each site in a protein sequence can be one of 20 amino acids, and typical protein sequences have lengths on the order of hundreds of sites. However, the available data typically consists of only a few thousand sequences.

A common approach to this problem is to search for the simplest probability distribution (by which we mean the one with the largest entropy) over the space of protein sequences that is capable of reproducing the single site and pairwise frequencies of amino acids in the data [12]. To further reduce complexity, one can represent protein sequences with a reduced or even binary (zero for 'wild-type' and one for 'mutant') amino acid alphabet [8,14,30]. In the binary case, the maximum entropy model capable of reproducing the empirical correlations is an Ising model with local fields h_i and pairwise couplings J_{ij}. Finding the appropriate fields and couplings to reproduce the correlations in the data is a challenging statistical problem [12]. For the models presented here, we applied the adaptive cluster expansion method [8,11] to infer field and coupling parameters that reproduced mutant frequencies and correlations observed in sequence data from a variety of HIV proteins [7].

Following the assumption that the most prevalent viral sequences are likely to also have the highest fitnesses, the prevalence landscape can be used as a proxy for fitness. The simple connection between prevalence and fitness could be obscured due to strong interactions between HIV and the immune system, which drives the virus to accumulate mutations rapidly. Careful modeling suggests, however, that prevalence is a good proxy for fitness for viruses such as HIV, which produces chronic infections and stimulates a diverse range of immune responses [34]. This assumption is also well-supported by experiments that tested the effects of different mutations on HIV replication [14,25,26]. These assumptions fail for a non-lethal seasonal virus such as the flu virus, which due to accumulation of similar immune responses in the hosts, evolves from year to year in a somewhat directed fashion.

3 Estimating Mean Fitness for HIV Proteins

We study population collapse for a few HIV proteins for which the fitness has been estimated to have the form [7,10]

$$F[\{\sigma_i^z\}] = F_0 - \sum_i h_i \sigma_i^z - \sum_{i<j} J_{ij} \sigma_i^z \sigma_j^z \,, \tag{6}$$

where σ_i are the Pauli matrices and F_0 is a constant. Including the effect of mutations as described in Eq. 2 leads to a dynamics governed by the Hamiltonian

$$\begin{aligned} H &= -F[\{\sigma_i^z\}] - \mu \sum_i (\sigma_i^x - \mathbf{1}) \\ &= -F_0 + \sum_i h_i \sigma_i^z + \sum_{i<j} J_{ij} \sigma_i^z \sigma_j^z - \mu \sum_i (\sigma_i^x - \mathbf{1}). \end{aligned} \tag{7}$$

We apply to this Hamiltonian three simple techniques, all common for studying quantum spin systems: first-order perturbation theory in μ, second-order perturbation theory, and a (non-perturbative) mean-field approximation. As will be shown, all three give consistent results, and whatever discrepancies exist can be easily understood.

3.1 First-Order Perturbation Theory

While sophisticated methods from quantum spin chains can certainly be applied to such a system, in practice we find that standard perturbation theory provides a more than adequate tool for identifying population collapse.

Let us denote the eigenvalue of the consensus state at $\mu = 0$ by E_0. Recall that the corresponding state is defined to have all $\sigma_i^z = -1$. To first order in perturbation theory, the change in eigenvalue is the expectation value of $\mu \sum_i (\sigma_i^x - \mathbf{1})$ in this state. The expectation value of σ_i^x being zero in any σ_i^z eigenstate, the perturbed eigenvalue is simply

$$E = E_0 + \mu L \,, \tag{8}$$

which becomes positive at the mutation rate (remember that E_0 must be negative).

$$\mu_c = -\frac{E_0}{L}. \tag{9}$$

As discussed in the Introduction, this corresponds to the mean fitness of the population becoming negative. The first order fatal mutation rate looks superficially similar to μ_c^E, the

critical mutation for the case of error threshold (Eq. 3), although in this case population collapse is due to a fatal mutation load.

3.2 Second-Order Perturbation Theory

The second-order correction to the eigenvalue of the consensus sequence, assuming non-degenerate levels, gives an expression

$$E = E_0 + \mu L + \mu^2 \sum_a \frac{1}{E_0 - E_a}\,, \tag{10}$$

where the sum is over configurations with one spin flipped. The fatal mutation rate can then be found by solving the quadratic equation

$$E_0 + \mu_c L + \mu_c^2 \sum_a \frac{1}{E_0 - E_a} = 0. \tag{11}$$

Note from Eq. (10) that if E_0 is less than all E_a (as is often the case, and must be if the consensus is the true ground state), then the second-order terms are necessarily negative. The eigenvalue is lower than the first-order estimate, and thus the fatal mutation rate is larger.

3.3 Mean-Field Approximation

As an alternative to the perturbative calculations, we also employ a mean field approach to directly find an approximation to the ground state of the quantum Hamiltonian. In this procedure the interacting Hamiltonian

$$H = -F_0 + \sum_i h_i \sigma_i^z + \sum_{i<j} J_{ij} \sigma_i^z \sigma_j^z - \mu \sum_i (\sigma_i^x - 1)\,, \tag{12}$$

is approximated by that describing non-interacting spins, as

$$H_{MF} = -F_0 + \sum_i h_i^{eff} \sigma_i^z - \mu \sum_i (\sigma_i^x - 1)\,. \tag{13}$$

The effective field at each site is expressed in terms of the average z component of spins at connected sites as

$$h_i^{eff} = h_i + \sum_{j \neq i} J_{ij} \langle \sigma_j^z \rangle \quad \text{where} \quad \langle \sigma_j^z \rangle = -\frac{h_i^{eff}}{\sqrt{(h_i^{eff})^2 + \mu^2}}\,. \tag{14}$$

To solve the set of coupled equations for $\{h_i^{eff}\}$, we use an iterative procedure, starting with the $\{h_i\}$, alternately computing the expectation values of spins and the corresponding effective fields until they converge. The mean fitness is finally computed as

$$\bar{F}_{MF}(\mu) = F_0 + \sum_{i=1}^{N} \left(\sqrt{(h_i^{eff})^2 + \mu^2} - \mu \right)\,, \tag{15}$$

from which we obtain μ_c by setting $\bar{F}_{MF}(\mu) = 0$.

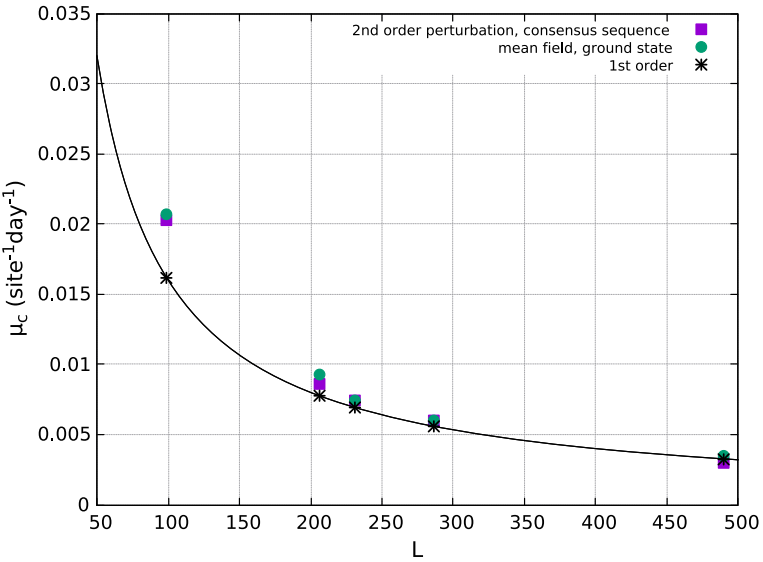

Fig. 1 Fatal mutation rates from mean-field approximation, compared to those from perturbing around the consensus sequence for five HIV proteins gag ($L = 490$), integrase ($L = 286$), nef ($L = 206$), p24 ($L = 231$), and protease ($L = 99$). The solid line is Eq. (9) assuming constant E_0

3.4 Results on Fatal Mutation Rate for HIV Proteins

We considered five HIV proteins of variable lengths: p24 ($L = 231$), gag ($L = 490$), integrase ($L = 286$), nef ($L = 206$), and protease ($L = 99$). The Ising on-site and exchange fields for each protein were estimated previously [7,10] from prevalence of different sequences in collected data. We then multiplied the prevalence landscape by a factor of $\beta = 0.023$/day, and added an overall constant of $F_0 = 1.6$/day to construct a fitness landscape. These choices were dictated by typical viral loads in patients, as discussed in Ref. [10].

In all cases, we both performed perturbation theory around the consensus state and applied the mean-field approximation. It will be important to note that the consensus is not the fittest state for the gag, nef, and protease proteins, whereas it is the fittest for p24 and integrase. The resulting μ_c are shown in Fig. 1.

All estimates for μ_c are quite consistent. There are some discrepancies, but they are all as expected. First, it is indeed the case that the first-order estimates are typically lower than the second-order values (the gag protein is the sole exception out of those considered). For p24 ($\mu_c = 0.007$/site/day) and integrase ($\mu_c = 0.006$/site/day), the consensus sequence corresponds to the ground state and thus we find quite good agreement between the second-order and mean-field results. Yet for gag, nef, and protease, the perturbation theory values only inform us of the mutation rate beyond which the consensus and all similar sequences die out. Alternate quasi-species having lower energies will survive for slightly larger μ. Since the mean field theory is able to (approximately) identify the true ground state, regardless of whether it agrees with the consensus or not, we expect it to predict higher values for μ_c. This is indeed what we find for the three proteins in question. We also note that the fatal mutation rates obtained for those are higher as compared to some obtained for other HIV proteins via alternate methods ($\sim 10^{-4}$/site/day) [18,20].

3.5 Effect of Mutations in Comparison with Experimental Fitness

As discussed in Sect. 2, the prevalence landscape is able to provide us an approximation of the fitness landscape, and is well supported by experiments. One quantity often measured by experiments is the replication capacity (RC) of a sequence, which is an empirical estimate of its growth rate and thus strongly associated with fitness [26]. RC is quantified by the exponential rate of growth in the number of infected cells over time in experiments, measured over multiple rounds of infection, for different mutant viruses. Comparing the model eigenvalues and the RC of some mutant viruses relative to the consensus sequence (for the gag protein, where multiple measurements are available) reveals a good correlation between the two. This leads us to conclude that indeed the Ising landscape derived for the proteins are a good measure of viral fitness (Fig. 2).

In comparisons with experimental growth data (via RC), it may be necessary to include the effect of mutations on the Ising landscape since the observed growth includes mutations as well [26]. We use our perturbation theory approach to compare the corrected eigenvalues of sequences by computing the difference in eigenvalue of a given sequence a and the consensus sequence (having eigenvalue E_0, which is not the lowest eigenvalue for gag protein) using Eq. 10, given by

$$\Delta E_a = E_a - E_0 + \mu^2 \sum_{b \neq a} \frac{1}{E_a - E_b} - \mu^2 \sum_{c \neq 0} \frac{1}{E_0 - E_c} \tag{16}$$

where E_a and E_0 refer to the unperturbed eigenvalues ($\mu = 0$). Considering a mutation rate typically expected in HIV proteins $\mu = 10^{-5}$/site/day, and the RC data observed for some mutations in the gag protein in [26], Fig. 2 shows that the effect of mutations is negligible when comparing to experimental fitness. However, to see any generic effect of mutations, we considered a higher mutation rate of 0.001/site/day (which is unrealistic but still below μ_c) and we find that the corrections can vary both in the sign and magnitude for the mutants considered here.

3.6 Role of Epistasis

Mutations at a given site can increase or decrease the overall fitness depending on the interaction of the mutated site with other site. This phenomenon of *epistasis* has been investigated in different HIV proteins, with evidence for both positive and negative epistasis [9,15,29,38]. Here, we explore the importance of epistasis for fatal mutation rates for population collapse (Fig. 1), by comparing the fatal rates using fitness landscapes obtained using the same methods discussed in Sect. 2, and forcing the interaction terms $J_{ij} = 0$ (in Eq. 7) for all pairs of sequences. The comparison is illustrated in Fig. 3. First, we note that the fatal mutation rate derived using first order perturbation theory (Eq. 9) is independent of J_{ij}. To see the effect of interactions, we compare the exact fatal mutation rate in the independent case with the rate obtained using mean field approximation (Eq. 15) for the interacting fitness landscape. Assuming that the mean field approximation in the interacting case is close enough to the exact answer that the distinction for the smaller length proteins with the non interacting case is real, we find that typically the mutation rate does not change significantly, but there is a small reduction upon including interactions that points toward negative epistasis. The decrease can be attributed to the fact that the collapse is determined by the eigenvalue of the ground state (or close to the consensus state), and typically mutations lower the fitness of the consensus sequence.

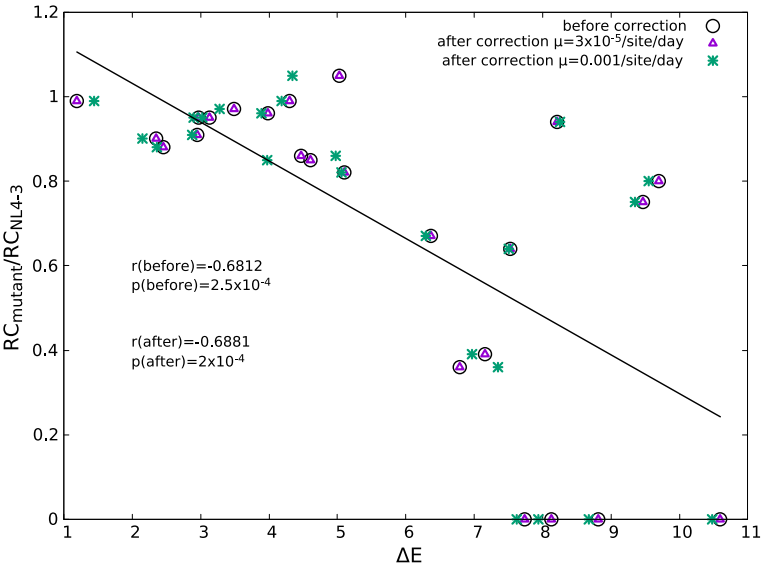

Fig. 2 Relationship between the predicted eigenvalue difference of some mutants relative to the consensus sequence and their replication capacities [26], for the gag protein (the NL4-3 sequence is the consensus sequence). The Pearson's correlation coefficient $r = -0.6812$ and p(two tailed) = 2.5×10^{-4}, with $n = 24$. Correction to eigenvalues using second order perturbation theory (Eq. 16) leads to negligible changes for a realistic mutation rate $\mu = 3 \times 10^{-5}$/site/day. But considering a higher mutation rate (0.001/site/day) shows that the correction can vary both in the sign and magnitude for different sequences. The r(after) and p(after) correspond to the higher mutation rate

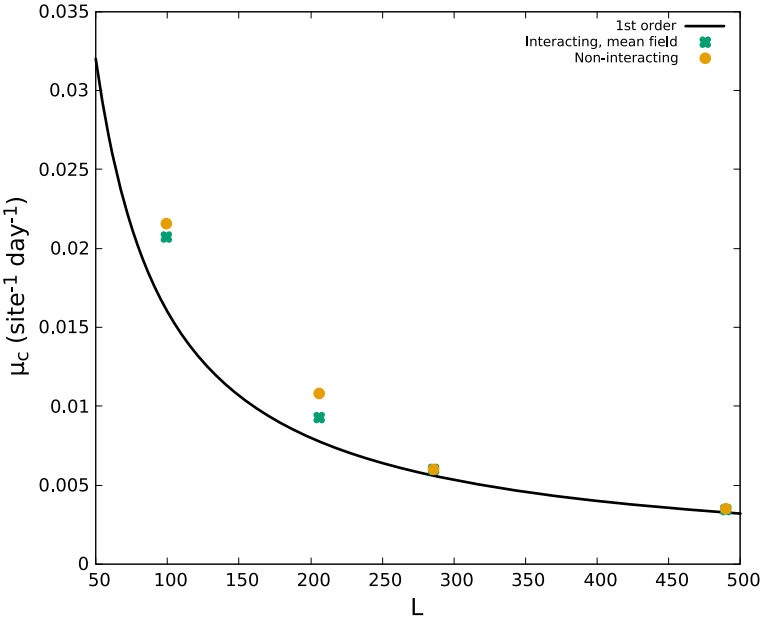

Fig. 3 Comparison of fatal mutation rates for the interacting (in the mean field approximation) and non interacting fitness landscapes for HIV proteins gag ($L = 490$), integrase ($L = 286$), nef ($L = 206$), and protease ($L = 99$). The first order line (Eq. 9) does not depend on interactions

4 Discussion

Here we applied ideas from interacting quantum systems to study Eigen's quasispecies model. Using past estimates for HIV fitness landscapes, which we verified to be in good agreement with experimental data, we computed the fatal mutation rate μ_c beyond which viral populations would be expected to decline due to reduced reproductive capacity. We have found that first-order perturbation theory works remarkably well in predicting μ_c. Even though we do not have an exact result against which to compare, neither higher-order terms nor a non-perturbative mean-field approximation gives significantly different results.

Note that the perturbative expression in Eq. (9) is superficially similar to Eigen's original result [13], even though the population collapse identified here is not due to an error catastrophe but rather due to negative mean fitness.

Hart and Ferguson recently investigated the possibility of an error catastrophe for HIV using estimated fitness landscapes [20]. They found evidence for a phase transition to a low fitness state in the HIV protein p6, which forms a part of gag. The transition occurs at a temperature larger than one, which is interpreted as a signal of a mutation rate that is higher than what is observed in nature. However, their approach does not estimate the critical mutation rate precisely in terms of a probability of mutation per replication cycle.

In fact, the immune system has special defenses that can force viruses to undergo an "error catastrophe." APOBEC proteins cause viral hypermutation, which nearly always results in the production of defective viruses (and for that reason not usefully modelled by the perturbative treatment here since the mutation rates can be orders of magnitude higher) [35]. However, APOBEC proteins can be countered by the HIV protein vif, thus allowing replication of the virus to continue unchecked. It is clear that a detailed understanding of the "phase diagrams" for HIV proteins will be a useful tool for combating the virus.

Acknowledgements The authors would like to thank the Galileo Galilei Institute in Florence, Italy where a part of this work was performed. This research was performed while CLB held an NRC Research Associateship award at the National Institute of Standards and Technology. MK is supported by NSF through Grant No. DMR-1708280.

References

1. Abanin, D.A., Altman, E., Bloch, I., Serbyn, M.: Colloquium: many-body localization, thermalization, and entanglement. Rev. Mod. Phys. **91**(2), 021001 (2019)
2. Baake, E., Wagner, H.: Mutation-selection models solved exactly with methods of statistical mechanics. Genet. Res. **78**(1), 93 (2001)
3. Baake, E., Baake, M., Wagner, H.: Ising quantum chain is equivalent to a model of biological evolution. Phys. Rev. Lett. **78**(3), 559 (1997)
4. Baake, E., Baake, M., Wagner, H.: Quantum mechanics versus classical probability in biological evolution. Phys. Rev. E **57**(1), 1191 (1998)
5. Baldwin, C., Laumann, C., Pal, A., Scardicchio, A.: Clustering of nonergodic eigenstates in quantum spin glasses. Phys. Rev. Lett. **118**(12), 127201 (2017)
6. Bapst, V., Foini, L., Krzakala, F., Semerjian, G., Zamponi, F.: The quantum adiabatic algorithm applied to random optimization problems: the quantum spin glass perspective. Phys. Rep. **523**(3), 127 (2013)
7. Barton, J.P., Kardar, M., Chakraborty, A.K.: Scaling laws describe memories of host-pathogen riposte in the HIV population. Proc. Nat. Acad. Sci. **112**(7), 1965 (2015)
8. Barton, J.P., De Leonardis, E., Coucke, A., Cocco, S.: ACE: adaptive cluster expansion for maximum entropy graphical model inference. Bioinformatics **32**(20), 3089 (2016)
9. Bonhoeffer, S., Chappey, C., Parkin, N.T., Whitcomb, J.M., Petropoulos, C.J.: Evidence for positive epistasis in HIV-1. Science **306**(5701), 1547 (2004)

10. Chen, H., Kardar, M.: Selforganization of matter and the evolution of biological macromolecules. bioRxiv:518704 (2019)
11. Cocco, S., Monasson, R.: Adaptive cluster expansion for inferring Boltzmann machines with noisy data. Phys. Rev. Lett. **106**(9), 090601 (2011)
12. Cocco, S., Feinauer, C., Figliuzzi, M., Monasson, R., Weigt, M.: Inverse statistical physics of protein sequences: a key issues review. Rep. Prog. Phys. **81**(3), 032601 (2018)
13. Eigen, M.: Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften **58**(10), 465 (1971)
14. Ferguson, A.L., Mann, J.K., Omarjee, S., Ndung'u, T., Walker, B.D., Chakraborty, A.K.: Translating HIV sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. Immunity **38**(3), 606 (2013)
15. Flynn, W.F., Haldane, A., Torbett, B.E., Levy, R.M.: Inference of epistatic effects leading to entrenchment and drug resistance in HIV-1 protease. Mol. Biol. Evol. **34**(6), 1291 (2017)
16. Galluccio, S.: Exact solution of the quasispecies model in a sharply peaked fitness landscape. Phys. Rev. E **56**(4), 4526 (1997)
17. Gerrish, P.J., Lenski, R.E.: The fate of competing beneficial mutations in an asexual population. Genetica **102**, 127 (1998)
18. Gupta, V., Dixit, N.M.: Scaling law characterizing the dynamics of the transition of HIV-1 to error catastrophe. Phys. Biol. **12**(5), 054001 (2015)
19. Haldane, J.B.S.: Mathematical Proceedings of the Cambridge Philosophical Society. In A mathematical theory of natural and artificial selection, part V: selection and mutation. Cambridge University Press, Cambridge, vol. 23, pp. 838–844 (1927)
20. Hart, G.R., Ferguson, A.L.: Error catastrophe and phase transition in the empirical fitness landscape of HIV. Phys. Rev. E **91**(3), 032705 (2015)
21. Held, T., Klemmer, D., Lässig, M.: Survival of the simplest in microbial evolution. Nat. Commun. **10**(1), 1 (2019)
22. Hermisson, J., Wagner, H., Baake, M.: Four-state quantum chain as a model of sequence evolution. J. Stat. Phys. **102**(1–2), 315 (2001)
23. Jörg, T., Krzakala, F., Kurchan, J., Maggs, A.: Simple glass models and their quantum annealing. Phys. Rev. Lett. **101**(14), 147204 (2008)
24. Leuthäusser, I.: Statistical mechanics of Eigen's evolution model. J. Stat. Phys. **48**(1–2), 343 (1987)
25. Louie, R.H., Kaczorowski, K.J., Barton, J.P., Chakraborty, A.K., McKay, M.R.: Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies. Proc. Natl. Acad. Sci. **115**(4), E564 (2018)
26. Mann, J.K., Barton, J.P., Ferguson, A.L., Omarjee, S., Walker, B.D., Chakraborty, A., Ndung'u, T.: The fitness landscape of HIV-1 gag: advanced modeling approaches and validation of model predictions by in vitro testing. PLoS Comput. Biol. **10**, 8 (2014)
27. Nandkishore, R., Huse, D.A.: Many-body localization and thermalization in quantum statistical mechanics. Annu. Rev. Condens. Matter Phys. **6**(1), 15 (2015)
28. Pal, A., Huse, D.A.: Many-body localization phase transition. Phys. Rev. b **82**(17), 174411 (2010)
29. Parera, M., Perez-Alvarez, N., Clotet, B., Martínez, M.A.: Epistasis among deleterious mutations in the HIV-1 protease. J. Mol. Biol. **392**(2), 243 (2009)
30. Rizzato, F., Coucke, A., de Leonardis, E., Barton, J.P., Tubiana, J., Monasson, R., Cocco, S.: Inference of compressed Potts graphical models. Phys. Rev. E **101**(1), 012309 (2020)
31. Rumschitzki, D.S.: Spectral properties of Eigen evolution matrices. J. Math. Biol. **24**(6), 667 (1987)
32. Saakian, D., Hu, C.K.: Eigen model as a quantum spin chain: exact dynamics. Phys. Rev. E **69**(2), 021913 (2004)
33. Sachdev, S.: Quantum Phase Transitions. Cambridge University Press, Cambridge (2011)
34. Shekhar, K., Ruberman, C.F., Ferguson, A.L., Barton, J.P., Kardar, M., Chakraborty, A.K.: Spin models inferred from patient-derived viral sequence data faithfully describe HIV fitness landscapes. Phys. Rev. E **88**(6), 062705 (2013)
35. Simon, V., Bloch, N., Landau, N.R.: Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. Nat. Immunol. **16**(6), 546 (2015)
36. Suzuki, S., Inoue, J.I., Chakrabarti, B.K.: Quantum Ising Phases and Transitions in Transverse Ising Models, vol. 862. Springer, New York (2012)
37. Wagner, H., Baake, E., Gerisch, T.: Ising quantum chain and sequence evolution. J. Stat. Phys. **92**(5–6), 1017 (1998)

38. Zhang, T.H., Dai, L., Barton, J.P., Du, Y., Tan, Y., Pang, W., Chakraborty, A.K., Lloyd-Smith, J.O., Sun, R.: Predominance of positive epistasis among drug resistance-associated mutations in HIV-1 protease. PLOS Genet. (2020). https://doi.org/10.1371/journal.pgen.1009009