Lethal Mutagenesis, Error Thresholds, and the Fight Against Viruses

Rigorous Modeling is Facilitated by a Firm Physical Background

PETER SCHUSTER

Peter Schuster, Editor-in-Chief of Complexity is at the Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, 1090 Wien, Austria (e-mail: pks@tbi.univie.ac.at)

Viroids and viruses can be viewed as ultimate parasites, which exploit the host through entering cells and manipulating cellular metabolism for their own reproduction. Proliferation of viruses is bound to successful replication of virus genomes, RNA or DNA, where successful implies viable virus progeny, that is, capable of reproduction. Viroids are naked RNA molecules and therefore entirely dependent on host biochemistry. The genomic RNA of simple bacteriophages[1] mimics messenger RNAs of the host cell and it is recognized and translated by host ribosomes. Simple virus genomes encode for a few proteins only, commonly for (i) a virus specific RNA replicase or at least a subunit of a replicase, which takes care that virus RNA and not cellular RNA is replicated preferentially, (ii) a protein for coating the virus RNA, and (iii) a protein that initiates lyses of the cell. As Charles Weissmann [1] pointed out already in the 1970s, in simple cases the life cycles of phages in bacterial cells are encoded by the structure of the genomic virus RNA and her unfolding–folding dynamics.

Forty years ago Manfred Eigen [2] published a kinetic theory of evolution at the molecular level, that is, directly applicable to replication of viral RNA. Among other things, one major characteristic of this theory is the handling of mutations: Correct replication and mutations are treated as parallel chemical reactions and this has the advantage that the approach applies equally well to the whole range of mutation rates from very small to large. The kinetic mutation selection equations of the model proposed by Eigen for N different variants X_j in the population are:

$$\frac{dx_j}{dt} = \sum_{i=1}^{N} Q_{ji} \cdot f_i x_i - x_j \phi(t); \; j = 1, 2, \ldots, N \qquad \phi(t) = \sum_{i=1}^{N} f_i x_i; \; \sum_{i=1}^{N} x_i = 1$$

[1]*Bacteriophages are viruses, which reproduce in bacterial cells.*

 Vol. 17, No. 2
DOI 10.1002/cplx.20399
Published online 22 October 2011 in Wiley Online Library (wileyonlinelibrary.com)

The variables in the equation are the normalized concentrations of the variants[2]: $[X_j] = x_j$. Two classes of parameters appear in the equation: fitness values f_j,[3] and dimensionless mutation frequencies Q_{ji}, which represent the probabilities to obtain the variant X_j as an error copy of the template X_i, the mutation frequencies are elements of a (stochastic) mutation matrix **Q**, as conservation of probabilities requires $\sum_{j=1}^{N} Q_{ji} = 1$, and they depend on the mutation rate per site and replication, p, which is a property of the RNA replication machinery of the cell, the virus or both. Implicit in the model equations are several assumptions: (i) The material required for replication is available in excess and remains present at constant concentration despite virus RNA synthesis, (ii) the population size is neither infinite nor zero to allow for normalization of variables, $\sum_{i=1}^{N} x_i = 1$, and (iii) the system is well-mixed and fluctuations in the population variables don't play a dominant role. Condition (i) is always fulfilled at the beginning of the virus infection of a cell, although there may be a shortage of components for the synthesis of viral RNA near the end of the infection. Condition (ii) will be discussed later. Condition (iii) is a general assumption in chemical kinetics and population dynamics, which is indispensible unless detailed information on spatial structures is available.

Selection and evolution take place in virus populations, which commonly contain a wide spectrum of variants with different fitness values [3, 4]. For sufficiently accurate replication and long enough time, populations approach stationary states called "quasispecies," which consists of a fittest genotype—the master sequence being present at highest concentration—and its mutants. For reproduction with ultimate accuracy expressed by a vanishing mutation rate, $\lim p \to 0$, the quasispecies contains exclusively the master sequence, and with increasing mutation rate the relative concentration of the master decreases. Eigen introduced an approximation that provided analytical expressions for the stationary mutant distribution as a function of the mutation rate. At some critical replication accuracy p_{cr}, the entire quasispecies vanishes in this approximation. Accurate numerical approximations of the exact solution have shown that the concentrations do not vanish but rather become very small [5]: $p > p_{cr} \Rightarrow x \approx 1/4^n$ with n being the genome length (An illustrative example is shown in Figure 1; for a detailed presentation of quasispecies theory see Eigen et al. [6]). Commonly, genome lengths n are a few hundred nucleotides for viroids and a few thousand nucleotides for viruses. Instead of a vanishing mutant distribution a phase transition-like change from a structured quasispecies to a uniform distribution of all variants is found by numerical computation [5], and the transition was characterized as error threshold: Because of error accumulation through imperfect replication no genome can be conserved over generations, in genealogies there is a zero longtime correlation between template and copy, and inheritance breaks down. The state of the populations beyond the error threshold has been illustratively characterized as random replication, since fitness plays no role for reproduction. The population is nonstationary for reasons to be discussed in the next paragraph.

What means random replication for a virus population? A uniform distribution of genotypes, which—as said above—is predicted by the kinetic model as the longtime outcome of evolution at mutations rates above the error threshold, cannot exist in reality. Because of the enormously large number of possible sequences—amounting to 4^n—concentrations would be smaller than a single molecule per experimental volume by many orders of magnitude. The population inevitably becomes nonstationary and migrates randomly through sequence space.[4] In addition, populations break up into clones, which drift independently as has been verified by numerical computation in case of neutral evolution [7, 8]. Whether a migrating virus population can exist beyond the error threshold or not is primarily a question of the fitness landscape: High degrees of neutrality and small fractions of mutations leading to nonviable variants or variants with zero fitness are clearly supportive for survival. In general, however, occurrence of extinction of the virus population can be expected, because viruses are under strong selection by the host's defense system and hence the degree of neutrality will be relatively small. Accordingly, it is very likely or of probability one that virus populations will become extinct after they passed the error threshold. As the number of imperfectly copied positions increases (linearly) with sequence length, longer genomes require higher replication accuracy. Experimental determination of spontaneous mutation rates [9, 10] yielded an approximate value of one error per reproduction and genome for simple RNA viruses, which implies that these species are reproducing close to the error threshold. This result can be interpreted straightforwardly: To escape the host's defense system the virus has to mutate as fast as possible and progresses towards the maximum tolerable mutation rate, which is given by the error threshold. The idea

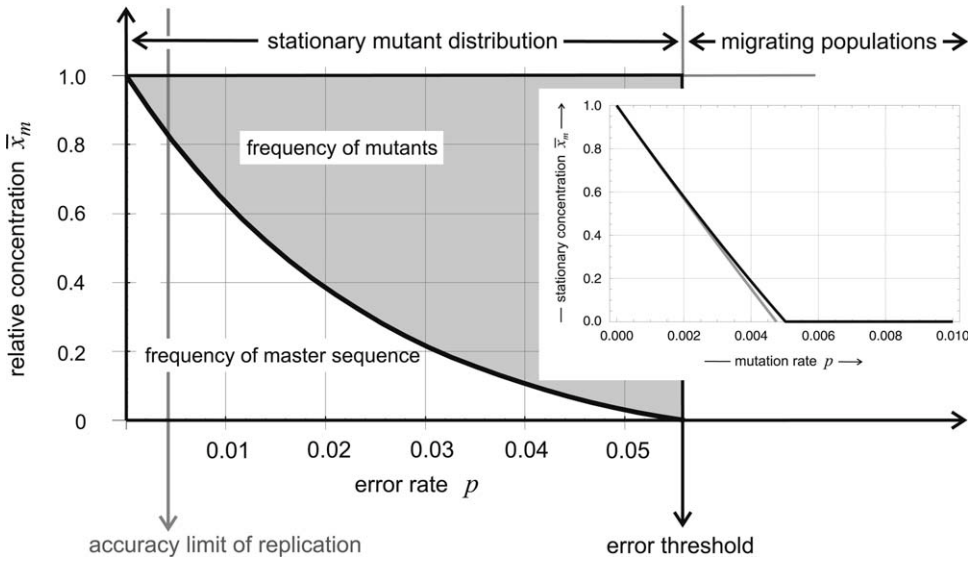
[2] *Concentrations commonly used in chemistry and condensed matter physics are particle numbers per unit volume. The standard unit is moles per liter.*

[3] *In biology the reproductive success is commonly measured in fitness values counting the (average) progeny of a variant in the next generation. In the molecular model the fitness values are functions of replication rate constants, binding constants, mutation rates and/or other physical parameters.*

[4] *The sequence space is an abstract space of all conceivable sequences. The Hamming distance counting the number of positions in which two aligned sequences differ is an appropriate metric in sequence space.*

DOI 10.1002/cplx

FIGURE 1



The error threshold observed with solutions of the mutation selection equation. The stationary frequency of the master sequence X_m, which is the sequence with the highest fitness value f_m, is plotted against the mutation rate p: $\bar{x}_m(p)$. Since no natural or artificial process can occur with ultimate accuracy there is a zone between $p = 0$ and the physical accuracy limit that cannot be accessed in reality. An increasing error rate leads to more mutants in the population and accordingly the fraction of the master sequence decreases. At some critical mutation rate, $p = p_{cr}$, a drastic change in the distribution of mutants is observed. The ordered quasispecies changes into the uniform distribution within a very small range of p values in the manner of a phase transition. Approximating full population dynamics by a neglect of backwards mutations, i.e., mutations from mutants back to the master sequence, the concentration of the master sequence vanishes at the error threshold. The insert shows a comparison of the approximation neglecting backwards mutation (gray) and the exact solution (black; parameters: $n = 20$, $f_0 = 2.2$, $f_n = 2.0$). At mutation rates beyond the error threshold the existence of a uniform stationary population is an artifact of the deterministic approach and instead migrating populations are observed.

of an error threshold and the known fact that mutation rates are tunable by means of small molecular weight compounds encouraged direct practical application in pharmaceutical research: The design of new drugs based on a mechanism of driving virus populations across the error thresholds was conceived as a novel antiviral strategy [11, 12].

Virus populations may become extinct without passing the error threshold as was correctly pointed out in the same year by James Bull et al. [13, 14]. The cause of extinction in this case is easily interpreted: Survival of the population requires a certain fraction of successful offspring that propagate the disease through infecting healthy cells. Bull et al. [14] derived a simple equation for the condition of extinction after mutation and selection have reached a stationary state,

$$e^{-\vartheta} R < 1,$$

wherein ϑ is the rate of deleterious mutations per genome[5] and replication and R is the average number of infectious progeny from a single infected cell. As the mutation rate ϑ increases, the number of infectious progeny goes down until it falls below the extinction threshold and then the virus population dies out. The authors called this phenomenon lethal mutagenesis, which is an unfortunate choice of notion as I shall outline later on.

A direct comparison of the condition of extinction with the error threshold observed in the mutation selection equation, however, is not possible, since the latter in the form reported by Bull et al. [14] contains normalized variables hence cannot describe the time dependence of the population size. A calculation of $C(t)$ is nevertheless easily possible [15]:

$$\frac{dC}{dt} = \sum_{i=1}^{N} f_i c_i - \Phi(t) = \bar{f}(t) \cdot C - \Phi(t).$$

The variables c_i are the concentrations of the individual variants,

[5] *The total genomic mutation rate* $\mu = p \cdot n$ *in the model of Bull et al.* [14] *accounts for neutral and deleterious mutations – advantageous mutations are excluded – and hence we have* $\mu = \nu + \vartheta$ *with* ν *being the rate of neutral and* ϑ *being the rate of deleterious mutations per genome and replication.*

DOI 10.1002/cplx

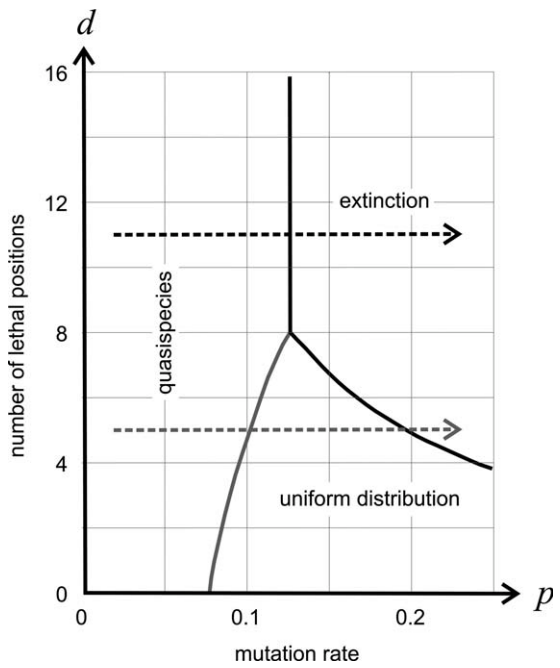
$c_i = x_i \cdot C$, the function $\bar{f}(t)$ is the mean fitness of the population and $\Phi(t)$ represents the constraint that is introduced by the experimental or in nature by the environmental conditions (For the condition $\Phi = \phi$ we obtain $dC/dt = 0$, which is tantamount to constant population size). Extinction is visualized straightforwardly by means of the differential equation for C: Assuming constant Φ that is realized, for example, in a flow reactor and decreasing $\bar{f}(t)$ as more mutants with low or zero fitness values f_i are produced, the right hand side of the equation will become negative for sufficiently high mutation rates p and then the population will die out.

More recently Hector Tejero et al. [16] studied the relation between lethal mutagenesis, extinction and error threshold by means of Eigen's original equations [2]. They distinguish three classes of variants: (i) the master sequence, X_m, (ii) viable and non-neutral mutants, X_k, and (iii) lethal mutants, X_l, with $f_l = 0$, and interpret the constraint as a universal degradation rate parameter, $\Phi = D$. Their most important result is shown in Figure 2: The plane spanned by the number of lethal positions on the sequence d[6] and the mutation rate p is separated into three regions: (i) the domain of the quasispecies, (ii) the domain of extinction, and (iii) the domain of the uniform distribution of variants. For increasing mutation rates two scenarios are possible. At low degree of lethality expressed as by a low value of d the population passes first an error threshold and becomes extinct later, and at sufficiently high d values, the quasispecies goes extinct without previously passing an error threshold.

Coming back to the previously made comments on the nonexistence of uniform distributions of virus populations we may conclude that the increase of the mutation rate leads to extinction of the virus population either by increasing the fraction of nonviable mutants or by accumulation of errors without limits. Who should care how the population goes extinct when a potential drug is effective? Nobody presumably, if the compound is already at hand, but the researchers in medicinal chemistry and pharmacology will be interested when they are confronted with the need to design new drugs. Knowing the target of the compound and the mechanism of action is essential for successful drug development. In this aspect neither of the two models is helpful at the current state of the art, but the Eigen model [2] is more easily extended to including details of the molecular mechanisms than the model of Bull et al. [14] and therefore I would give preference to it.

Finally, I would like to stress that the usage of the term "lethal mutagenesis" exclusively for the extinction threshold is unfortunate, because both mechanisms, accumulation of lethal variants and accumulation of copying errors, lead to extinction of the virus population when realistically considered. My suggestion is therefore to use lethal mutagenesis for both phenomena, to distinguish the extinction threshold from the error threshold, to keep the race open, and to leave it for the future, which of the two concepts provides more successful ideas for progress in the medical treatment of virus infections.

FIGURE 2



Lethal mutagenesis in virus populations. The figure is redrawn from Tejero et al.16 and shows three distinct scenarios for stationary populations as functions of mutation p rates and degrees of lethality d: (i) a domain within which ordered quasispecies are formed, (ii) a domain of population extinction, and (iii) a domain where the uniform distribution prevails. At low degree of lethality the stationary population passes first the error threshold and then goes extinct at higher p values, whereas no error threshold exists at higher degree of lethality. There, the population goes extinct without passing an error threshold.

[6] *A lethal position on a genome is a nucleotide that leads to a lethal variant when mutated.*

DOI 10.1002/cplx

REFERENCES

1. Weissmann, C. Making of a phage. FEBS Lett 1974, 43, S10–S18.
2. Eigen, M. Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 1971, 58, 465–523.
3. Domingo, E.; Sabo, D.; Weissmann, C. Nucleotide-sequence heterogeneity in RNA of phage Qβ. Experientia 1976, 32, 792.
4. Domingo, E.; Sabo, D.; Taniguchi, T.; Weissmann, C. Nucleotide-sequence heterogeneity of an RNA phage population. Cell 1978, 13, 735–744.
5. Swetina, J.; Schuster, P. Self-replication with errors—A model for polynucleotide replication. Biophys Chem 1982, 16, 329–345.
6. (a) Eigen, M.; McCaskill, J.; Schuster, P. Molecular quasispecies. J Phys Chem 1988, 92, 6881–6891; (b) Eigen, M.; McCaskill, J.; Schuster, P. The molecular quasispecies. Adv Chem Phys 1989, 75, 149–263.
7. Derrida, B.; Peliti, L. Evolution in a flat fitness landscape. Bull Math Biol 1991, 53, 355–382.
8. Huynen, M.A.; Stadler, P.F.; Fontana, W. Smoothness within ruggedness: The role of neutrality. Proc Natl Acad Sci USA 1996, 93, 397–401.
9. Drake, J.W. Rates of spontaneous mutation among RNA viruses. Proc Natl Acad Sci USA 1993, 90, 4171–4175.
10. Drake, J.W.; Charlesworth, B.; Charlesworth, D.; Crow, J.F. Rates of spontaneous mutation. Genetics 1998, 148, 1667–1686.
11. Domingo, E. Virus entry into error catastrophe as a new antiviral strategy. Virus Res 2005, 107, 115–228.
12. Biebricher, C.K.; Eigen, M. The error threshold. Virus Res 2005, 107, 117–127.
13. Bull, J.J.; Meyers, L.A.; Lachman, M. Quasispecies made simple. PLoS Comput Biol 2005, 1, e61.
14. Bull, J.J.; Sanjuán, R.; Wilke, C.O. Theory of lethal mutagenesis for viruses. J Virol 2007, 81, 2930–2939.
15. Eigen, M.; Schuster, P. The hypercycle. A principle of natural self-organization. Part B: The abstract hypercycle. Naturwissenschaften 1978, 65, 7–41.
16. Tejero, H.; Marín, A.; Montero, F. Effect of lethality on the extinction and the error threshold of quasispecies. J Theor Biol 2010, 262, 733–741.

DOI 10.1002/cplx