Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Q

QUASISPECIES

Esteban Domingo, Centro de Biología Molecular "Severo Ochoa" (CSIC-UAM), Universidad Autónoma de Madrid, Cantoblanco, Madrid, Spain

Origins of the Concept

Quasispecies describes a type of population structure in which collections of closely related genomes are subjected to a continuous process of genetic variation, competition and selection. Quasispecies has become very important in virology because it provides an interpretation for the extensive plasticity, both genetic and phenotypic (of biological features), displayed by many viruses, in particular RNA viruses. The quasispecies concept originated in theoretical studies on early informational macromolecules by Eigen and colleagues, and in experimental work on bacteriophage $Q\beta$ by Weissmann and colleagues. Eigen was the first to elaborate a quantitative treatment of the generation of error copies of simple replicating molecules (replicons) as those that probably populated the earth at the early stages of life. This treatment represented a link between principles of information theory and Darwinian evolutionary concepts, and was inspired to a great extent by experiments on evolution of simple RNA molecules replicating *in vitro* carried out by Spiegelman and his associates in the 1960s.

Eigen, Schuster and their colleagues defined the quasispecies as an organized, steady-state distribution of error copies of self-replicating macromolecules (nucleic acids). It represents a rated distribution of mutants (mutant spectrum) centered around one or several master sequences (**Table 1** and **Fig. 1**). The copying fidelity with which replication occurs determines the fraction of progeny molecules which will be identical to the parental replicon. This theoretical concept assumes a selection equilibrium in the distribution whose evolution can be described by a system of ordinary differential equations. This selection equilibrium is metastable in that it will collapse when an advantageous mutant appears in the distribution. The previous quasispecies will then be substituted by a new one, characterized by a new master sequence and a new mutant spectrum. The stability of a quasispecies depends on the complexity of the genetic information contained in the genome, the copying fidelity and the superiority of the master sequence. This is expressed mathematically in an error threshold relationship which limits the amount of genetic information that can be transmited when copying fidelity is limited (**Table 1**).

Quasispecies has a physical, a chemical and a biological definition. Physically a quasispecies distribution of genomes can be regarded as a cloud in sequence space. Sequence space refers to all possible sequences that theoretically could be occupied by a genome (**Table 2**). Viral quasispecies can be viewed as clouds that occupy tiny portions of such theoretical space. Occupation of sequence space by viruses is restricted by multiple genetic and functional constraints. The clouds have, at any given time, a defined population structure dependent on the distance of each point to the consensus sequence (that could be taken as the center of the distribution), and on the relative fitness values of the different mutant components. The distance or number of mutated positions between a given sequence and the best adapted one (the master sequence) is called the Hamming distance. This term is also used to indicate the number of mutations between two sequences.

Chemically, the definition of quasispecies is the one most familiar to virologists: a rated distribution of related, nonidentical genomes. Biologically, quasispecies are the target of selection. It is not easy to imagine an ensemble rather than an individual genome is the actual target of selection. However, individual RNA genome sequences may have a fleeting existence, and there are now a number of cases documenting that the evolutionary fate of an individual viral genome is strongly influenced by the mutant spectrum surrounding it.

Table 1 Important equations in quasispecies

• *Basic differential equation describing quasispecies evolution*

$$\frac{dx_i}{dt} = (A_i Q_i - D_i)x_i + \sum_{k \neq i} w_{ik} x_k + \phi_i$$

dx_i/dt = variation of concentration of mutant i as a function of time
x_i = concentration of mutant i
A_i = velocity constant for the replication of i
Q_i (quality factor) = probability that replication of i produces i
D_i = velocity of decomposition of i
W_{ik} x_k = velocity of formation of i as a result of erroneous replication of any other molecule k
ϕ_i = Constant that corrects for changes of concentration of i as a result of flux of molecules

• *Error threshold relationship*

$$\nu_{max} < \ln \sigma_o / (1 - \bar{q})$$

ν_{max} = the maximum length of the sequence (genetic complexity) that can be maintained during replication
σ_0 = selectivity or superiority of the master sequence relative to its mutant spectrum
$\bar{q}$ = average copying fidelity. The average error rate is $(1 - \bar{q})$. (Different positions of a replicating nucleic acid often show different copying fidelities. In this case $\bar{q}$ is the geometric mean of the copying fidelity value at each position.) The expected number of errors in a sequence of length ν replicating with a fidelity of $\bar{q}$ is $\nu\ (1 - \bar{q})$

The quasispecies theory was aimed at interpreting self-organization and evolutionary optimization of primitive genetic material. It was recognized that error rates lower than about 10^{-2} misincorporations per nucleotide and round of copying would be difficult to attain without participation of catalytic functions contained in some types of protein and RNA structures. This gave rise to the extended notion of catalytic hypercycle which combined the coding ability of nucleic acids (types of molecules well suited as a repository of genetic information) with the catalytic potential of proteins (more complex molecules that can combine multiple active surfaces to accelerate a variety of biochemical reactions).

Real Virus Quasispecies

The second origin of the quasispecies concept was experimental. It had long been suspected that RNA viruses were genetically unstable, as suggested by the abundance of mutants in viral stocks and by the difficulty of maintaining some mutants (conditional lethal, plaque-type) free of wild-type revertants. The first evidence that an RNA virus depicted features of quasispecies was obtained by Weissmann and associates in Zürich working with bacteriophage Qβ. Upon replication in its host *Escherichia coli*, a clone of bacteriophage Qβ generated error copies with high frequency. Quantification of the reversion of a site-specific mutant, and of its growth rate relative to the parental wild-type Qβ, allowed an estimate of the mutation rate for a specific purine transition: about 10^{-4} per round of copying, a value about 10^5-fold larger than that estimated for the mutation rate of DNA genomes. Furthermore, genomic RNA from many phage clones was analyzed by T1-oligonucleotide fingerprinting (rapid nucleotide sequencing techniques were not yet available). The results indicated that, assuming a random distribution of mutations among the genomes analyzed, each infectious RNA differed in 1–2 positions from the average or consensus sequence in the population. Interestingly, the T1-oligonucleotide map of the RNA from the uncloned phage population remained unchanged during 50 serial infections, a fact that must be kept in mind before considering stability of a consensus sequence as evidence against quasispecies. A highly dynamic mutant spectrum can nevertheless produce the same average sequence over many generations of viral replication. Weissmann and colleagues concluded that: 'The genome of Qβ phage cannot be described as a defined unique structure, but rather as a weighted average of a large number of different individual sequences.'

Subsequent studies with many human, animal and plant RNA viruses, or viruses which include an RNA

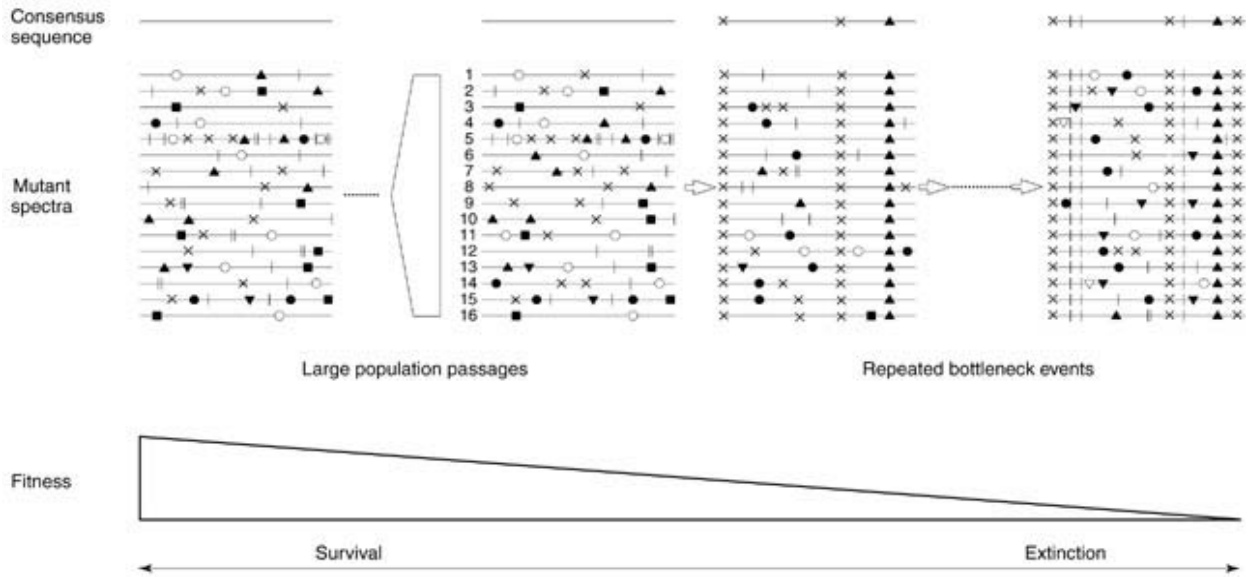

Figure 1 Representation of quasispecies and of possible outcomes of their evolution. Horizontal lines represent individual genomes, and symbols on the lines represent mutations. Large population passages (left) in a constant environment generally lead to fitness gains. The average sequence may or may not be modified as a result of replication. Repeated bottleneck events (right, small arrows) generally result in accumulation of mutations and average fitness losses that may drive viral populations near extinction. For a mathematical expression of quasispecies and definition of terms see **Tables 1** and **2**.

Table 2 Definition of some terms frequently used in the literature of quasispecies

- *Consensus or average sequence (of nucleotides or amino acids)* The sequence that results from taking for each position the most frequent residue (nucleotide or amino acid) found at the corresponding position in a set of aligned sequences. The consensus sequence may not exist physically as a component of the mutant distribution
- *Fitness* A parameter that describes the adaptation of an organism to its environment. For viruses, relative fitness values are quantitated as their ability to produce infectious progeny relative to a reference virus, in a defined environment
- *Master sequence* The dominant sequence in a quasispecies. It often shows the highest selection value among the mutant sequences present, but it may or may not coincide with the consensus or average sequence of the population
- *Mutant spectrum* The ensemble of mutant genomes that constitute a quasispecies
- *Mutation frequency* The proportion of mutants (average for an entire sequence or specific for a defined site) in a genome population
- *Mutation rate* The frequency of occurrence of a mutation event during genome replication
- *Population number* The number of individuals in a population. For viruses, it is the number of infectious genomes replicating in a given biological niche (cell, tissue, organ, organism). The number of genomes quantitated in an infected host is also termed the viral load
- *Rate of fixation of mutations* The frequency of mutations in the viral genome which become dominant per unit time in an infected host or among a set of infected hosts. This rate is generally not constant. The assumption of a molecular clock is not justified for RNA viruses
- *Sequence space* A theoretical representation of all possible variants of a genomic sequence. The sequence space is n^{ν} in which n is the number of symbols in an informational macromolecule and ν is the complexity (total number of residues) of the macromolecule. For a single-stranded RNA virus of 10 000 residues in length the sequence space is $4^{10\,000}$

as a replication intermediate, have documented that they all share high mutation rates or frequencies, and the dynamics of competition and selection predicted by the quasispecies model. The genetic plasticity of RNA replicons led Holland and colleagues to emphasize the multiple biological and evolutionary implications of such a highly dynamic RNA world parasitizing the relatively static DNA world of differentiated organisms.

A generalized quasispecies concept is now used by virologists to describe dynamic distributions of nonidentical but closely related mutant and recombi-

nant viral genomes subjected to a continuous process of genetic variation, competition and selection, and which act as a unit of selection.

Environmental Perturbations and Occupation of Sequence Space

Real virus quasispecies, such as those replicating in the course of acute or chronic infections of differentiated organisms, are permanently perturbed by a complicated array of environmental influences. One is compartmentalization of virus replication within infected organisms: viruses may replicate in disparate cell types, such as epithelial cells of the upper respiratory tract, lymphocytes or brain cells. Different sites within an infected organism will be subjected to different kinds of selective pressures derived either from variations in the internal environment (immune responses, metabolic or nutritional alterations, temperature, etc.) or from externally applied influences such as the presence of antiviral agents. Competition among components of the mutant spectrum will be established within each infected cell. Virus progeny from different cells may engage in subsequent rounds of competition (for entry into target cells, for effective replication, etc.). Such rating events may involve many genomes in the case of acute, systemic infections in which viruses may reach up to 10^{12} infectious units at a given time in the entire organism, and viral RNA loads composed of infectious and noninfectious genomes may reach exceedingly high levels. Evolving mutant distributions are easily shaped by selective forces and perturbed by random sampling events (**Fig. 1**), providing a genetic flexibility that could not be reached if mutations were rare events. Adaptability is facilitated by the limited genome size (generally between 3×10^3 and 3×10^4 nucleotides) and large population numbers. All possible single and double mutants, and decreasing proportions of triple, quadruple, etc. mutants are potentially present in many viral populations. In contrast, only a minute fraction of all possible single mutants is present at any given time in populations of cellular organisms. That is, the ability to explore their allotted share of sequence space is vastly superior for small, abundant replicons than for complex DNA genomes. Furthermore, one or a few nucleotide (or amino acid) replacements are often sufficient to trigger biologically relevant changes (alterations in cell tropism, virulence, escape from immune surveillance, resistance to antiviral inhibitors, etc.). Variant genomes with few mutations have a high probability of occurrence within the mutant spectra of evolving quasispecies. Of course, clouds of mutants are far from being symmetrical, as even single point mutants may have reduced fitness and may not be represented, or represented with low frequency, in the mutant spectra. Mathematical extensions of the initial theoretical quasispecies concept have been developed to cope with some of the perturbations undergone by real virus quasispecies; however, such treatments are complex and difficult to apply to experimental observations with viruses.

Multiple Biological Implications

Quasispecies represents a new description of the population structure of RNA viruses, and it has numerous implications for the biology of this important group of pathogens (**Table 3**). Particularly relevant to viral disease control is the pre-existence and selection of mutants completely or partially resistant to antiviral agents. Examples are influenza virus A mutants resistant to amantadine (1-aminoadamantane) and rimantadine (α-methy-1-adamantane methylamine), human rhinovirus mutants displaying low- or high-level resistance to disoxaril, 5-[7-[4-(4,5-dihydro-2-oxadozolyl)phenoxy]heptyl]-3-methyl-isoxazole, or other related compounds, and, more recently, human immunodeficiency virus mutants resistant to reverse transcriptase and protease inhibitors.

The frequency of antibody-resistant mutants of RNA viruses is generally high (10^{-4}–10^{-6} per infectious genome) both in cell culture and *in vivo*. The frequency of cytotoxic T lymphocyte-escape mutants is more difficult to determine but variations at specific T cell epitopes permitting virus escape *in vivo* have been documented in several systems, such as human hepatitis C virus, human immunodeficiency virus and lymphocytic choriomeningitis virus.

RNA viruses may use a number of alternative receptors or coreceptors for entry into cells, and shifts in receptor specificity may be triggered by amino acid replacements at surface residues on the viral capsid or viral envelope. Mutant viruses able to use either alternative receptors or altered forms of a standard receptor occur in nature and have also been selected in cell culture.

Virulent (disease-causing) variants may be generated in populations of avirulent viruses, or may be present as minority components of attenuated viral preparations. Beck, Levander and associates described that an avirulent coxsackievirus B3 became cardiopathic upon replication in selenium- or vitamin E-deficient mice, as a result of repeated selection of six specific mutations scattered along the viral genome. It is likely that such a phenotypic modification was facilitated by the impaired immune responses, and the larger ensuing viral loads, in infected hosts bearing

Table 3 Some implications of the quasispecies structure of viral populations

- It offers a general and highly effective adaptive strategy
- In particular, it explains the pre-existence and selection in viral populations of mutants with altered phenotypic properties:
  - resistant to antiviral agents
  - antibody and cytotoxic T lymphocyte escape
  - with altered cell receptor specificity
  - with different ability to induce interferon
  - with increased virulence
- It implies the existence of thresholds for phenotypic expression
- There is a possible association between the pathogenic potential of a virus and its mutant spectrum complexity

some nutritional deficiencies. Chumakov and colleagues quantitated the presence of virulent poliovirus in the mutant spectrum of attenuated poliovirus used for vaccine preparation. When the proportion of virulent poliovirus variants in the vaccine exceeded a certain value, the virulent phenotype was manifested and the vaccine failed the safety tests in monkeys. This, and additional results with other viral systems, have established that expression of a variant phenotype encoded in a subset of genomes from the mutant spectrum depends on their proportion in the mutant distribution. These results agree with theoretical predictions on the behaviour of viral quasispecies in that the fate of each individual component of the mutant spectrum is strongly dependent on the mutant cloud surrounding it.

Recent evidence suggests that the complexity of the mutant spectrum may influence the outcome of viral infections. The complexity of the coronavirus mouse hepatitis virus quasispecies may contribute to its pathogenic potential. Likewise, the nonresponse to treatment with interferon α in chronic hepatitis C virus infections may relate to the number of viral molecular species detected in the infected patients. Model experiments with the animal pathogen foot-and-mouth disease virus showed that indeed the repertoire of viral mutants that became dominant in an evolving quasispecies depended on the population size of the virus. Thus, it is not surprising that both the population size and the complexity of the mutant spectrum may be important determinants of the pathogenic potential of some viruses, although the number of well-documented cases is still limited.

Connections with Population Genetics and with Current Concepts of Complexity

The main departure of the quasispecies concept from previous models of population genetics is the emphasis on mutation (or, more generally, in genetic variation) to the point of invalidating the concept of the wild type as a defined genomic sequence. RNA viral genome sequences are statistically defined, but individually indeterminate. Recognition of such indetermination, together with the emphasis on continuous mutant generation, contributed to the success of quasispecies, rather than other quantitative models of population biology, as a descriptor of RNA viruses.

The rapid evolution of RNA viruses has been used to explore some principles and theories of population biology. A concept of increasing interest in virology is fitness (**Table 2**). This is a complex parameter which has been the object of considerable research and debate in biology. Fitness of a virus is measured as its relative ability to produce infectious progeny under a defined set of environmental conditions. Relative fitness values have been determined in tissue culture and in some cases in laboratory mice or other hosts. Cells or organisms are co-infected with the virus to be tested together with a genetically or phenotypically marked reference virus. The progeny virus is passaged for several transfers and the quantification of the proportion of the two viruses relative to the initial mixture yields a relative fitness value. Using this assay a number of observations on the dynamics of viral quasispecies have been made.

When large populations of RNA viruses are allowed to replicate in a defined environment, fitness gains are generally observed when measured in the same environment (**Fig. 1**). However, adaptation of virus to one environment (for example one cell type) may result in profound fitness losses in another environment. Two vesicular stomatitis virus clones with similar relative fitness competing in serial infections coexisted for many passages. However, in a rather unpredictable manner, one of the clones abruptly displaced the other and became dominant in the population. This observation agrees with the competitive exclusion principle of population genetics that states that unless there is a niche differentiation, one of two competing species will eventually outcompete the other. In such competitions both the

Table 4 Some practical implications of the quasispecies structure of RNA viruses

- Vaccines must be multivalent (multiple B and T epitopes)
- Antiviral agents must be used in combination (directed to independent targets; the number depends on viral population size and genome turnover)
- Completely new antiviral strategies should be explored (drugs capable of throwing viral replication into error catastrophe)
- The use of virulent RNA viruses as pest control agents should be avoided

winners and the losers gained fitness in such a manner that their relative position in the fitness landscape was similar at the end and at the start of the competition. This agrees with the Red Queen hypothesis of population biology, as, in the words of the Red Queen in Lewis Carroll's *Through the Looking Glass*, 'it takes all the running you can do, to keep in the same place'.

In contrast to large population passages, repeated bottleneck events such as those mediating serial plaque-to-plaque transfers, resulted in stochastic fitness losses (**Fig. 1**). This is known as the 'Muller's ratchet' effect, according to which asexual populations of organisms will tend to incorporate deleterious mutations in a rather irreversible fashion unless compensatory mechanisms such as recombination can restore the initial, better adapted, genomes. For well-adapted viral quasispecies, repeated plaque isolations will tend to deviate successive mutant distributions from the optimal one, resulting in average fitness losses (**Fig. 1**). The transmission population size needed to maintain viral fitness is dependent on the initial fitness of the viral clone tested. The higher the initial fitness the larger the transmission pool must be to avoid fitness losses. Studies with foot-and-mouth disease virus have documented that viral clones subjected to serial plaque-to-plaque transfers accumulate unusual mutations, providing insight into the types of low-frequency mutants that populate viral quasispecies.

RNA viruses in their evolutionary dynamics constitute attractive experimental systems of complex adaptive behaviour. They display a highly indeterminate fine structure as well as a rather unpredictable behavior. Indetermination arises from the stochastic nature of mutagenesis, in conflict with the need of the system to ensure genetic continuity. Adaptability stems from the variable degrees of success of subsets of genomes. The indetermination of mutagenesis, together with the directionality of selective forces, situate RNA viruses in a subtle border between reproducibility of some observations and unpredictability of others. Thus, RNA virus quasispecies may also become interesting model systems for studies on complexity.

Strategies for Disease Prevention and Control

The great adaptability of viral quasispecies creates difficulties for viral disease control, and may also contribute to the emergence of new viral pathogens. When quasispecies was recognized as the most adequate descriptor of pathogenic RNA viruses, the need to adjust antiviral strategies to the new findings became apparent to some scientists. These recommended strategies (**Table 4**) are not yet generally followed, probably because of an inherent tendency of thinking of RNA viruses still as genetically defined entities. If a single selective pressure is applied to limit virus replication, the highly dynamic mutant spectrum is likely to provide variants capable of overcoming the selective constraints. Research on entire new antiviral strategies, such as the possibility of displacement of virus replication into error catastrophe, seems justified in view of accumulating evidence that the copying fidelity properties of viral replicases can be modified by structural alterations of the enzymes.

See also: **Antivirals; Defective interfering viruses; Emerging and re-emerging virus diseases; Immune escape mechanisms; Persistent viral infection.**

Further Reading

Domingo E and Holland JJ (1997) RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* 51: 151.

Domingo E, Sabo DL, Taniguchi T and Weissmann C (1978) Nucleotide sequence heterogeneity of an RNA phage population. *Cell* 13: 735.

Domingo E, Holland JJ, Biebricher C and Eigen M (1995) Quasispecies: the concept and the word. In: Gibbs AJ, Calisher CH and Garía-Arenal F *et al* (eds) *Molecular Basis of Virus Evolution*, p. 171. Cambridge: Cambridge University Press.

Eigen M and Schuster P (1979) *The Hypercycle – A Principle of Natural Self-Organization*. Berlin: Springer.

Holland JJ, Spindler K, Horodyski F *et al* (1982) Rapid evolution of RNA genomes. *Science* 214: 1577.

Nowak MA (1992) What is a quasispecies? *Trends* in *Ecol. Evol.* 7: 118.