J Mol Evol (1989) 28:524–527

Journal of Molecular Evolution

Evolution of RNA Genomes: Does the High Mutation Rate Necessitate High Rate of Evolution of Viral Proteins?

E.V. Koonin and A.E. Gorbalenya

Institute of Poliomyelitis and Viral Encephalitides of the USSR Academy of Medical Sciences, 142782 Moscow Region, USSR

Summary. RNA genomes have been shown to mutate much more frequently than DNA genomes. It is generally assumed that this results in rapid evolution of RNA viral proteins. Here, an alternative hypothesis is proposed that close cooperation between positive-strand RNA viral proteins and those of the host cells required their coevolution, resulting in similar amino acid substitution rates. Constraints on compatibility with cellular proteins should determine, at any time, the covarion sets in RNA viral proteins. These ideas may be helpful in rationalizing the accumulating data on significant sequence similarities between proteins of positive-strand RNA viruses infecting evolutionarily distant hosts as well as between viral and cellular proteins.

Key words: Covarions — Evolutionary rates — Protein coevolution — RNA genomes

RNA viruses are unusual among autoreproducing biological entities in many respects, beginning with the chemical nature of their genomes. Also characteristic of these agents are the small genome size and extremely high replication error frequency due primarily to the lack of correction mechanisms in viral RNA-dependent RNA polymerases (Holland et al. 1982; Domingo et al. 1985; Smith and Inglis 1987; Steinhauer and Holland 1987).

Specifically, the misincorporation rate of bacteriophage Q_β RNA-dependent RNA polymerase is 10^{-3}–10^{-4} per nucleotide per replication round, i.e., 10^4–10^7 times higher than in DNA replication (Holland et al. 1982; Steinhauer and Holland 1987). Very similar error rates have been demonstrated also for vesicular stomatitis virus and poliovirus polymerases (Steinhauer and Holland 1987; Ward et al. 1988).

The relationship between the misincorporation frequency and actual mutation rate is not fully understood; nevertheless, it seems obvious that the higher the replication error frequency, the higher the mutation rate should be. This is generally confirmed by numerous studies on the variability of RNA viral genomes in tissue culture or in nature, in spite of somewhat contradictory specific values; anyhow, a general consensus exists that RNA genomes mutate much more frequently than DNA genomes (Holland et al. 1982; Domingo et al. 1985; Smith and Inglis 1987; Steinhauer and Holland 1987). Based on these observations, it is often assumed that the evolution of RNA virus proteins also proceeds very rapidly. Thus, there was speculation that the relationships between RNA viruses inferred from protein sequence similarities reflect processes of divergent evolution that proceeded rather recently, in the geological sense of the word at least (Ahlquist et al. 1985; Goldbach 1986, 1987). In the framework of this hypothesis, the similarities between plant and animal viruses are explained by horizontal transmission of ancestral viruses, presumably by insect vectors.

However, this is not the only possible view. Many RNA viral proteins, especially those constituting the replication machinery of positive-strand RNA viruses, are believed to function in close contact with cellular proteins. The most striking example of such cooperation is the bacteriophage Q_β replicase that consists of three cellular subunits, along with the

Offprint requests to: E.V. Koonin

phage-encoded subunit (Blumenthal 1979). Two of these cellular subunits, translation elongation factor Tu and ribosomal protein S1, are slowly evolving proteins (Linz et al. 1986; Regnier et al. 1987). Cellular protein factors have been implicated in picornavirus RNA replication (Koonin and Agol 1983; Andrews et al. 1985), and one of the poliovirus-encoded proteases has been shown to be involved in proteolysis of a specific cellular protein (Krausslich et al. 1987). For other virus–cell systems only indirect evidence is available, such as demonstration of tight association of viral replication complexes with intracellular membranes (cf. Ranki and Kaariainen 1979; Watanabe and Okada 1986). Detailed studies of interactions between viral and cellular proteins is highly desirable, and evidence of their cooperation seems obvious. It is reasonable to hypothesize that this cooperation should impose some degree of functional constraint on the structure of viral proteins, thus slowing down their evolution. In itself, this idea is not new. It has been suggested more than once that viral proteins interacting with cellular components should evolve more slowly than those not involved in such interactions (Ahlquist et al. 1985; Emini et al. 1985). It is tempting, however, to make a stronger proposal, i.e., that due to the intimacy of cooperation, the absolute rates of amino acid substitution in interacting viral and cellular proteins should be within the same order of magnitude.

One may wonder how can this view be reconciled with the high mutation rate characteristic of RNA genomes. The dynamics of protein evolution is probably best described by the so-called covarion model of Fitch (Fitch 1971; Ratner et al. 1985). The key definition of this model is a covarion, i.e., a position in a polypeptide chain where substitutions can be introduced without impairing the function of the protein. The actual evolutionary rate for a protein is determined by the total number of unique covarions available during a given time interval. This value is connected with the amino acid substitution rate by the following equation:

$$Q_\tau = Q_0 + c \cdot m \qquad (1)$$

where Q_τ is the total number of unique covarions during the time interval τ, i.e., the total number of positions in the polypeptide chain where substitutions could be fixed during the interval τ; Q_0 is the true covarion number, i.e., the number of positions in which substitutions can be fixed at any given moment; c is the covarion turnover rate expressed as the fraction of covarions renewed with each substitution fixed; and m is the total number of substitutions fixed during the interval τ.

As shown by Fitch (1971), Q_0 is typically a small number (for example, for hemoglobins and cytochromes c Q_0 = 4–10). The value of Q_0 depends on the degree of structural integrity necessary for the function of the given protein (and/or of a macromolecular complex, the components of which are proteins); the more ordered the structure that is necessary for this function, the smaller is the number of positions where substitutions are permitted. The c parameter values are in the range of 0.7–0.8. This means that, with each substitution fixed, 70–80% of former covarions become nonvariable, whereas an equal number of new covarions arises (see Ratner et al. 1985). The m value is directly proportional to the mutation rate (Kimura 1983).

Compare now the evolution of RNA-encoded and of DNA-encoded proteins. In terms of the covarion model, our statement that the evolutionary rates for both kinds of proteins may be roughly equal is expressed as

$$Q_\tau^{RNA} \approx Q_\tau^{DNA} \qquad (2)$$

X-ray studies of spatial structure of positive-strand RNA virions showed that the general folding principles of RNA-encoded proteins may be similar to those of DNA-encoded proteins (Liljas 1986). Thus, it is natural to assume

$$Q_0^{RNA} \approx Q_0^{DNA} \qquad (3)$$

From the high mutation rate characteristic of RNA genomes it follows that

$$m^{RNA} \gg m^{DNA} \qquad (4)$$

Substituting (1) into (2) gives

$$Q_0^{RNA} + c^{RNA} \cdot m^{RNA} \approx Q_0^{DNA} + c^{DNA} \cdot m^{DNA} \qquad (5)$$

or, taking into account the equality (3), then

$$c^{RNA} \cdot m^{RNA} \approx c^{DNA} \cdot m^{DNA} \qquad (6)$$

As the inequality (4) holds for the m values, we obtain

$$c^{RNA} \ll c^{DNA} \qquad (7)$$

Thus, according to our hypothesis, in the evolution of RNA viral proteins the high mutation rate is counter-balanced by a very low covarion turnover rate; in other words, with each substitution fixed, only a very small (if any) fraction of the covarion set is renewed. This means that a characteristic of the evolution of such proteins should be multiple (including reverse) substitutions in the same positions of the polypeptide chain. One may speak of "covarion persistence" resulting in relatively slow apparent evolutionary rates determined by the evolutionary rates of the cellular proteins interacting with the given RNA viral protein. Gibbs suggested (1980) that the evolutionary rates of RNA viral proteins should not differ significantly from those of proteins of DNA organisms as functional con-

straints are probably similar for both kinds of proteins. From the above discussion it is explicit, however, that this condition alone (i.e., $Q_0^{RNA} \approx Q_0^{DNA}$, in terms of the covarion model) is insufficient to account for the approximate equality.

Recently some data have accumulated on sequence similarity between cellular and positive-strand RNA viral proteins (Zimmern 1983; Gorbalenya et al. 1986a, 1988a,b,c; Hodgman 1988). Particularly impressive are similarities between proteases and putative helicases of eukaryotic viruses and respective eubacterial enzymes (Gorbalenya et al. 1986a, 1988a,b,c; Hodgman 1988). Though it is possible that these similarities are due to recent divergence of viral proteins from highly conserved eukaryotic homologs of the bacterial enzymes, we feel that they are best rationalized in terms of the present hypothesis.

The present concept relies on the presumed cooperation and coevolution of viral and cellular proteins. It is well known that the rates of evolution of the latter vary by as much as three orders of magnitude (Kimura 1983). Clearly, according to our hypothesis, the evolutionary rates of positive-strand RNA viral proteins interacting with the cellular ones should fall within the same range. It is also important to realize that the ideas developed here are directly applicable only to selectively neutral but not to adaptive amino acid substitutions. The rate of the latter depends not only on the mutation rate, but also on the effective population size (Kimura 1983; Ratner et al. 1985). The value of the latter parameter is very uncertain in the case of RNA viruses. Well-substantiated evidence demonstrates that, under positive selection, evolution of viral proteins may be extremely rapid (Holland et al. 1982; Domingo et al. 1985; Steinhauer and Holland 1987; Smith and Inglis 1987). Thus, the present concept in no way contradicts the generally accepted notion of high adaptability of RNA viruses of which high mutation rate is a prerequisite. These rapid alterations occur, however, upon shifting virus/milieu equilibrium (i.e., when a virus is transferred to a new environment) and involve mainly virion proteins, or even external domains thereof, interacting with cellular receptors, antibodies, etc. In general, however, the vast majority of amino acid substitutions in proteins are probably neutral, constituting the basis for the similarity and diversity that is revealed upon protein sequence comparison (Kimura 1983).

The ideas developed here suggest that sequence comparisons between proteins of positive-strand RNA viruses belonging to different families may be helpful in elucidating not only relatively recent but also early events in the evolution of this virus class. Moreover, as positive-strand RNA viruses possess many features reminiscent of most primitive genetic systems (Eigen et al. 1981; Gilbert 1986; Gorbalenya et al. 1986b), one may even hope that such comparisons should shed some light on what the hypothetical primordial "RNA world" was like.

Acknowledgments. We thank Drs. V.I. Agol and K.M. Chumakov for critical reading of the manuscript.

References

Ahlquist P, Strauss EG, Rice CM, Strauss JH, Haseloff J, Zimmern D (1985) Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. J Virol 53:536–542

Andrews NC, Levin D, Baltimore D (1985) Poliovirus replicase stimulation by terminal uridylyl transferase. J Biol Chem 260: 7628–7635

Blumenthal T (1979) Q_β replicase and protein synthesis elongation factors EF-Tu and EF-Ts. Meth Enzymol 60:628–638

Domingo E, Martinez-Salas E, Sobrino F, de la Torre JC, Portela A, Ortin J, Lopez-Galindez C, Perez-Brena P, Villanueva N, Najera R, Van de Pol S, Steinhauer D, DePolo N, Holland JJ (1985) The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. Gene 40:1–8

Eigen M, Gardiner W, Schuster P, Winkler-Oswatitsch R (1981) The origin of genetic information. Sci Am 244:88–118

Emini E, Schleiff WA, Colonno RJ, Wimmer E (1985) Antigenic conservation and divergence between the viral-specific proteins of poliovirus type 1 and various picornaviruses. Virology 140:13–20

Fitch WM (1971) Rate of change of concomitantly variable codons. J Mol Evol 1:84–96

Gibbs A (1980) How ancient are the tobamoviruses? Intervirology 14:101–108

Gilbert W (1986) The RNA world. Nature 319:618

Goldbach R (1986) Molecular evolution of plant RNA viruses. Annu Rev Phytopathol 24:289–310

Goldbach R (1987) Genome similarities between plant and animal RNA viruses. Microbiol Sci 4:197–202

Gorbalenya AE, Blinov VM, Donchenko AP (1986a) Poliovirus-encoded proteinase 3C: a possible evolutionary link between cellular serine and cysteine proteinase families. FEBS Lett 194:253–257

Gorbalenya AE, Donchenko AP, Blinov VM (1986b) Poliovirus proteins with different functions are probably of common origin. Molek Genetika No. 1:36–41 [in Russian]

Gorbalenya AE, Koonin EV, Blinov VM, Donchenko AP (1988a) Sobemovirus genome appears to encode a serine protease related to cysteine proteases of picornaviruses. FEBS Lett 236:287–290

Gorbalenya AE, Koonin EV, Blinov VM, Donchenko AP (1988b) A conserved NTP-binding motif in putative helicases. Nature 333:22

Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1988c) A novel superfamily of nucleoside triphosphate-binding motif containing proteins probably involved in duplex unwinding in DNA and RNA replication and recombination. FEBS Lett 235:16–24

Hodgman TC (1988) A protein superfamily involved in nucleic acid replication and recombination. Nature 333:22–23

Holland JJ, Spindler K, Horodyski F, Grabau E, Nichol S, Van de Pol S (1982) Rapid evolution of RNA genomes. Science 215:1577–1585

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, England

Koonin EV, Agol VI (1983) Encephalomyocarditis virus replication complexes that prefer nucleoside diphosphates as substrates for viral RNA synthesis. Virology 129:309–318

Krausslich HG, Nicklin MJH, Toyoda H, Etchison D, Wimmer E (1987) Poliovirus proteinase 2A induces cleavage of eukaryotic initiation factor 4F polypeptide p220. J Virol 61: 2711–2718

Liljas L (1986) Structure of spherical viruses. Prog Biophys Mol Biol 18:1–36

Linz JE, Lira LE, Sypherd PS (1986) The primary structure and the functional domains of an elongation factor-1 from *Mucor racemosa*. J Biol Chem 261:15022–15029

Ranki M, Kaariainen L (1979) Solubilized RNA replication complex from Semliki forest virus-infected cells. Virology 98: 298–307

Ratner VA, Zharkikh AA, Kolchanov NA, Rodin SN, Soloviev VV, Shamin VV (1985) Problems of the theory of molecular evolution. Nauka, Novosibirsk, USSR [in Russian]

Regnier P, Grunberg-Manago M, Ortier C (1987) Nucleotide sequence of the pnp gene of *Escherichia coli* encoding polynucleotide phosphorylase. Homology of the primary structure of the protein with the RNA binding domain of ribosomal protein S1. J Biol Chem 262:63–68

Smith DB, Inglis SC (1987) The mutation rate and variability of eukaryotic viruses: an analytical review. J Gen Virol 68: 2729–2740

Steinhauer D, Holland JJ (1987) Rapid evolution of RNA viruses. Annu Rev Microbiol 41:409–433

Ward CD, Stokes MAM, Flanegan JB (1988) Direct measurement of the poliovirus RNA polymerase error frequency in vitro. J Virol 62:558–562

Watanabe Y, Okada Y (1986) In vitro viral RNA synthesis by a subcellular fraction of TMV-inoculated tobacco protoplasts. Virology 149:64–73

Zimmern D (1983) Homologous proteins encoded by yeast mitochondrial introns and by a group of RNA viruses from plants. J Mol Biol 171:345–352

Received May 27, 1988/Revised September 2, 1988