Gross rearrangements within the 5'-untranslated region of the picornaviral genomes

Evgeny V.Pilipenko, Vladimir M.Blinov and Vadim I.Agol
Institute of Poliomyelitis and Viral Encephalitides, The USSR Academy of Medical Sciences, Moscow Region 142782 and Moscow State University, Moscow 119899, USSR

Received November 9, 1989; Revised and Accepted January 15, 1990

ABSTRACT

An analysis of reported nucleotide sequences revealed several cases of gross rearrangements in the 5'-untranslated region (5-UTR) of picornaviral genomes. A large (>100 nt) duplication was discovered in a downstream region of poliovirus 5-UTR involved in the translational control. Properties of the poliovirus mutants with large deletions [Kuge and Nomoto (1987) J. Virol. 61, 1478–1487] show that a single copy of the appropriate repeating unit is compatible with a wild type phenotype of the virus. In contrast to poliovirus and another enterovirus genomes, human rhinovirus RNAs contain only a single copy of this repeating unit. Another similarly large repeat was found in an upstream segment of the bovine enterovirus 5-UTR. A comparison of the primary and secondary structures of cardio- and aphthovirus 5-UTRs demonstrated the existence of a large (ca. 250 nucleotides) insertion/deletion in a region preceding the poly(C) tract. The two latter rearrangements appear to involve elements of the viral genome replication machinery. Possible origin as well as evolutionary and functional implications of these structural peculiarities are discussed.

INTRODUCTION

Picornaviruses are small naked icosahedral animal viruses comprising four genera, *Enterovirus* (polioviruses, coxsackieviruses, echoviruses etc.), *Rhinovirus, Cardiovirus* (encephalomyocarditis virus, Theiler's murine encephalomyelitis virus etc.), and *Aphthovirus* (foot-and-mouth disease viruses). The genome of these viruses is represented by a single-stranded RNA molecule of positive (mRNA-like) polarity. This RNA possesses an unusually long 5'-untranslated region (5-UTR) ranging in length from about 600 nucleotides in human rhinoviruses to more than 1200 nucleotides in aphthoviruses. This part of the viral genome embodies several essential *cis*-acting control elements involved in its replication, translation, and possibly other functions. As could be expected, it is highly conserved, both in terms of the primary and secondary structures, but, surprisingly, there are two quite different sets of such conserved structures, one for entero- and rhinoviruses (1–3), and the other for cardio- and aphthoviruses (4; cf. also, Tiley and King, *Abstr. of 6th Meet. of the Europ. Study Group on Mol. Biol. of Picornaviruses*, B12, Bruges, September 10–16, 1989). A closer inspection revealed, however, several examples of individual or group-specific gross rearrangements within each of the two consensus structures. These rearrangements, which are described in this note, should have important evolutionary implications. Moreover, they could help explain some puzzling experimental observations.

A POLIOVIRUS REPEAT

The first example of such an interesting feature concerns the poliovirus 5-UTR and is represented by directly repeated sequences located in the segment preceding, and partly intruding into, the polyprotein coding region (Fig. 1). In fact, there are two pairs of such repeats. The first repeating unit (Fig. 1, a) is over 100 nucleotides-long, and occupies, in poliovirus type 1 Mahoney strain, positions 533 to 645 and 670 to 772 (the initiator AUG starts at residue 743); 61% of nucleotides in the 2 units are identical. The stretch separating these repeats (positions 646–669) is in turn a unit of another tandem repeat (Fig. 1, b); the second, downstream, unit of this repeat corresponds to a 5'-terminal portion of a larger, just mentioned repeating element. Here, 15 nucleotides out of 27 are identical. A similar pattern of the repeating units arrangement could be revealed in the 5-UTRs of other poliovirus serotypes as well (Fig. 1), although the extent of similarity among the respective pairs of repeating units appeared in these cases to be somewhat lower (41–46% and 48–54% for the larger and smaller repeating units, respectively).

The repeating units overlap an essential *cis*-acting translational control element located inside the poliovirus 5-UTR (5–11). Among other features, this element contains a highly conserved stem-loop structure between nucleotides 584 and 614 (using the poliovirus type 1 numbering) (1–3). This structure was proposed to specifically interact with a host-cell translation initiation factor (12, 13). In an apparent contradiction with the proposed importance of this conserved structural element, Kuge and Nomoto (14) reported that extended deletions intruding into this region could result in viable poliovirus mutants, some of which even having no obvious phenotypic changes, at least in tissue culture cells. This amazing observation could possibly be explained just by the existence of the repeating sequences in the appropriate segment of the poliovirus genome. As shown in Fig. 2, due to the repeat, the removal of nucleotides 600–726, as in one of Kuge and Nomoto's viable mutants (IC-DH), should

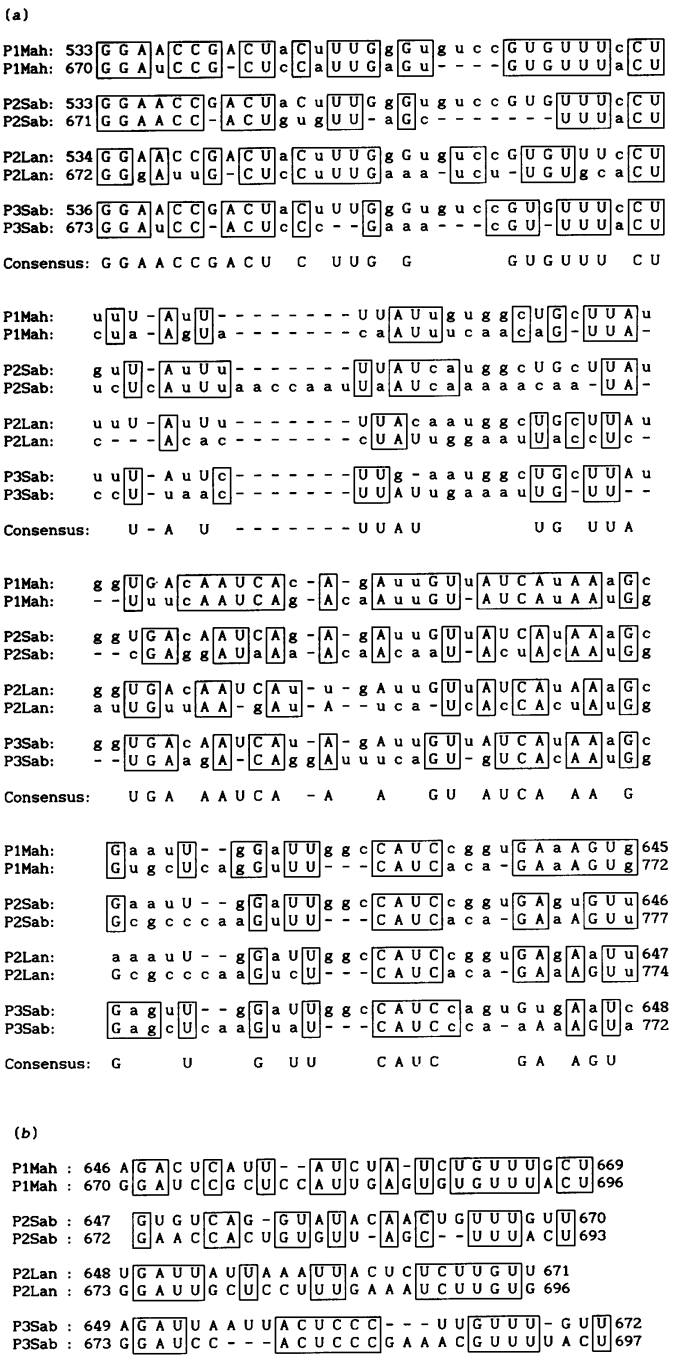

Figure 1. Repeating units in the 5-UTR of poliovirus genomes. Type 1, Mahoney strain (P1Mah; the primary structure was taken from refs. 30 and 31); type 2, Sabin strain (P2Sab; ref. 32); type 2, Lansing strain (P2Lan; ref. 33); type 3, Sabin strain (P3Sab; refs. 32 and 34). Identical nucleotides within a given pair of units are boxed. (a) Larger repeating units. The repeats in all four strains are aligned with each other, and a consensus sequence (capitals) has been derived. (b) Smaller repeating units. No attempt to align the repeats and to derive a consensus sequence for different strains has been made because of a considerable sequence divergence within the region separating the larger units.

merely slightly alter (by destroying a single G-U pair) the critical stem-loop structure of the putative *cis*-acting translation control element. This fact is the more remarkable as the potential to form such a structure in the downstream repeating element itself is

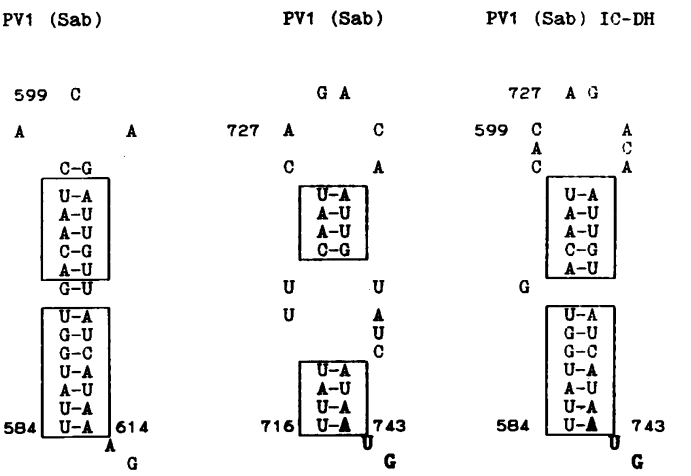

Figure 2. Proposed secondary structures of an element of the poliovirus genome. From right to left, the 584–614 segment (1) and 716–743 segment of the poliovirus type 1, Sabin strain, genome, and the 584–743 segment of a viable mutant (IC-DH; cf. ref. 14), lacking nucleotides 600–726. The primary structure was taken from (32). The initiator AUG is printed in bold letters.

not so well conserved (Fig. 2). In addition, the data of Kuge and Nomoto (14) demonstrate that a single copy of the repeating unit is fully compatible with the viability as well as wild-type plaque phenotype of poliovirus (more extended deletions, which truncated the remaining copy of the repeating unit, resulting in small-plaque mutants or unviable progeny).

Similarly, the RNAs of human rhinoviruses also possess only a single copy of this repeating unit (cf., ref. 1), and this fact explains why rhinovirus 5-UTRs are shorter than poliovirus ones. Interestingly, the alignment of the poliovirus downstream repeating unit and rhinovirus 5-UTRs shows that the initiator AUGs occupy nearly identical (shifted only by a single codon) positions in these two types of genomes (not shown), whereas the corresponding triplet in the poliovirus upstream repeating unit is mutated to AAG (Figs. 1 and 2). A possible evolutionary relationship between the polio- and rhinovirus 5-UTRs will be discussed below.

A BOVINE ENTEROVIRUS REPEAT

Another repeating element was found in the 5-UTR of the genome of bovine enterovirus (BEV). The alignment of its primary structure with those of other enteroviruses was suggestive of the presence of an insertion in an upstream segment of the BEV RNA (15). Actually, a direct tandem repeat >100 nucleotides in length could be revealed here (positions 7–115 and 116–230) (Fig. 3, a). Although the appropriate repeating units have about 43% divergent nucleotides, they may fold into nearly identical secondary structure elements (Fib. 3, b). These elements are very similar to a single-copy element found by Rivera *et al.* (2) to be conserved among different entero- and rhinovirus genomes; mutations in the appropriate nucleotide sequence were reported to affect replication of poliovirus RNA (16; Andino, Rieckhof, Trono and Baltimore, *Abstr. of 6th Meet. of the Europ. Study Group on Mol. Biol. of Picornaviruses*, A17, Bruges, September 10–16, 1989). It seems very likely that at least the upstream repeating unit of the BEV 5-UTR should be specifically recognized by the viral genome replication machinery.

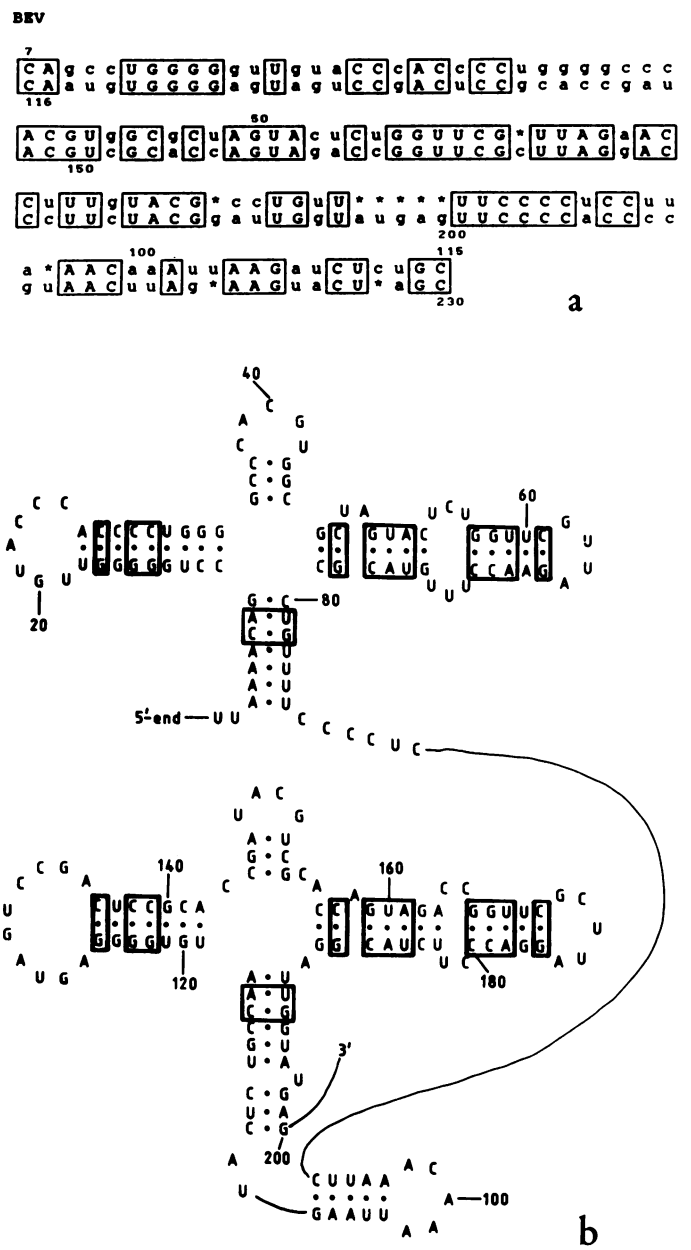

Figure 3. Repeating units in the 5-UTR of the BEV genome. (a) The alignment of the two tandem repeats. (b) The proposed secondary structure of the repeats. The primary structure was taken from (15).

AN APHTHOVIRUS/CARDIOVIRUS INSERTION/ DELETION

One more example of a large rearrangement was detected upon a comparison of the primary and secondary structures of the encephalomyocarditis (EMC) and foot-and-mouth disease virus (FMDV) 5-UTRs. The RNAs of these viruses are known to have a long poly(C) tract, which is located 150–400 nucleotides from the 5′ end (17–21). Models for the secondary structure of the region preceding the poly(C) tract were suggested for both groups of viruses (22, 23). A similar, though not identical, folding shown in Fig. 4 demonstrates that these regions could form a kind of stem-loop structure, with the long imperfect stems having several

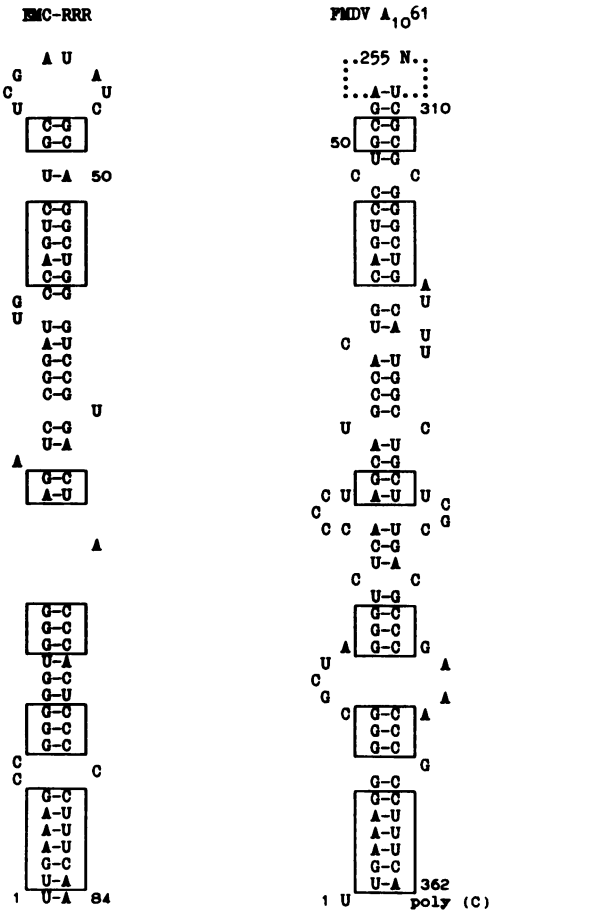

Figure 4. Conserved 5′-terminal secondary structure elements of the EMC virus, strain RRR, and FMDV, strain $A_{10}61$, genomes. The primary structures were taken from refs. 22 and 23, respectively.

conserved, among EMC and FMDV RNAs, secondary structure elements and a ca. 250 nucleotides-long insertion into the loop in the latter case. In other words, despite a huge insertion into the terminal structure that could safely be assumed to be directly involved in the EMC genome replication, the essential recognition elements appear to be preserved in the FMDV RNA due to long-range interactions. The origin of this insertion/deletion is unknown, although a weak similarity to a region of FMDV RNA downstream from poly(C) could be revealed (not shown).

It may be noted that in the genome of another cardiovirus, Theiler's murine encephalomyelitis virus, there is a long segment which appears to replace the poly(C) tract characteristic of such cardioviruses as encephalomyocarditis or Mengo viruses (24). This could be regarded as one more example of gross rearrangements in the picornaviral 5-UTRs.

POSSIBLE ORIGIN AND FUNCTIONAL SIGNIFICANCE OF THE REPEATING UNITS

Thus, the length heterogeneity of the picornaviral 5-UTRs appears to be primarily caused by duplications and large insertions/deletions of nucleotide stretches. It should be mentioned that relatively short repeating elements 3′ from FMDV poly(C) were described previously (25). Moreover, it was speculated that the entire poliovirus genome originated through the multiplication of short genetic elements (26).

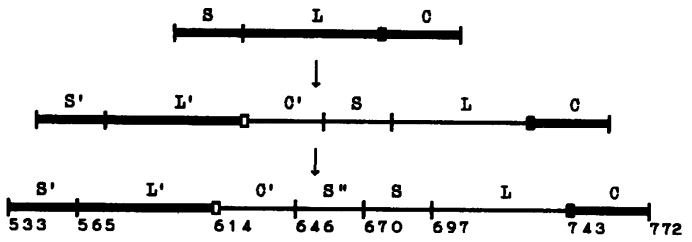

Figure 5. A hypothetical reconstruction of the origin of the enterovirus genome. S and L denote short and long segments of the repeating unit, and C corresponds to the region encoding an N-terminal portion of the viral polyprotein. Heavy lines correspond to the most conserved and functionally most important segments. Full rectangles denote actual initiator AUGs of the polyprotein reading frame. Open rectangles correspond to mutated or otherwise inactivated initiation codons. The numbers correspond to the nucleotide positions in the poliovirus type 1 RNA, from which appropriate segments are started. For other explanations, see text.

The mechanism of generation of repeats and other rearrangements is unknown, but it is most likely related to template switches postulated to occur during the replication of picornaviral RNA (27–29). The generation of adjacent (tandem) direct repeating elements requires a single 'jump' of the nascent chain, perhaps in association with the RNA-dependent RNA polymerase, from one template to another (or from one locus of a template to another locus of the same template) (cf. 27), whereas for the appearance of noncontiguous repeating elements (or the insertion of 'foreign' sequences) at least two such jumps are needed. The duplication-generating jumps should obviously be facilitated if the template already contains short direct repeats (27). It may be noted that some short repeats existing in poliovirus and bovine enterovirus genomes could be related to the generation of the large repeating units.

A closer comparison of the structures of the two repeating units of the polio genome, on the one hand, and the appropriate segment of the rhinovirus genome, on the other, allowed us to propose a model for the possible evolutionary relationship between the 5-UTRs of the two picornaviral genera (Fig. 5). According to this reconstruction, a precursor of the polio genome had a 'rhino-like' structure, i.e., it possessed a unique sequence element encompassing the beginning of the polyprotein reading frame and a 5'-adjacent segment (in the present-day human rhinovirus (HRV)-2 RNA, this element occupies positions from 524 to 640 with the initiator AUG starting at position 611). The untranslated upstream sequence was essential for the initiation of the polyprotein synthesis. The next step of the poliovirus genome evolution could be represented by a duplication of the entire element (that is of its both noncoding and coding sequences). As a consequence, an additional, and perhaps functional, upstream initiator triplet was created, which might interfere with the activity of the genuine AUG that opened the polyprotein reading frame. Such interference was abolished by a mutation in the upstream AUG (a point mutation here is actually present in the poliovirus RNA); in addition, possible deleterious effects of the newly generated and interfering reading frame could be minimized by the appearance of translation terminators. Then, a relatively short 5' segment of the downstream repeating unit was also duplicated. At this point, the structure of the relevant repeat-containing 5-UTR segment was represented by the following formula: S'-L'-C'-S''-S-L-C, where S and L are short and long noncoding sequences, respectively, and C is the coding sequence of the repeating unit; if we accept that recombinant events take place largely during the minus strand synthesis (28), then the downstream repeating unit (S-L-C) of the original recombinant molecule should directly correspond to the appropriate sequence of the first template RNA, whereas the upstream element (S'-L'-C'-S'') evolved through reiterative copying of the same or another template.

The most important portions of this structure were S' and L' (involved in the initial interaction with ribosomes, or initiation factors, or both; see references 5–11 and 14) as well as C (encoding an N-terminal sequence of the viral polyprotein). Indeed, these 3 portions were markedly more conserved among polio- and rhinovirus genomes as compared with the C'-S''-S-L 'insert' (not shown). Thus, this reconstruction shows how the close proximity of the control element (S-L) and the coding region (C) characteristic of the rhinovirus RNA was lost in the course of generation of the poliovirus genome.

The fact that the large repeats, while apparently being nonessential (at least in the case of poliovirus), are nevertheless genetically stable suggests that they could perform a useful function. One such function was evident from the previous discussion — the repeating unit may serve as a 'spare part'. A more interesting possibility is the involvement of the repeats in some kinds of regulatory phenomena. For example, they may participate in rearrangements of the secondary or tertiary structures of *cis*-acting elements, or in an interaction with some *trans*-acting factors. The availability of such powerful experimental tools like virus-specific cDNA clones could make these speculations amenable for direct testing.

ACKNOWLEDGEMENTS

We are grateful to Drs D.Baltimore, E.Ehrenfeld, V.Racaniello and N.Sonenberg for preprints, and to A.Gmyl for help in the manuscript preparation.

REFERENCES

1. Pilipenko,E.V., Blinov,V.M., Romanova,L.I., Sinyakov,A.N., Maslova,S.V. and Agol,V.I. (1989) *Virology* **168**, 201–209.
2. Rivera,V.M., Welsh,J.D. and Maizel,J.V.,Jr. (1988) *Virology* **165**, 42–50.
3. Skinner,M.A., Racaniello,V.R., Dunn,G., Cooper,J., Minor,P.D. and Almond,J.W. (1989) *J. Mol. Biol.* **207**, 379–392.
4. Pilipenko,E.V., Blinov,V.M., Chernov,B.K., Dmitrieva,T.M. and Agol,V.I. (1989) *Nucl. Acids Res.* **17**, 5701–5711.
5. Svitkin,Y.V., Maslova,S.V. and Agol,V.I. (1985) *Virology* **147**, 243–252.
6. Svitkin,Y.V., Pestova,T.V., Maslova,S.V. and Agol,V.I. (1988) *Virology* **166**, 394–404.
7. Pelletier,J., Kaplan,G., Racaniello,V.R. and Sonenberg,N. (1988) *Mol. Cell Biol.* **8**, 1103–1112.
8. Pelletier,J. and Sonenberg,N. (1988) *Nature* **334**, 320–325.
9. Trono,D., Andino,R. and Baltimore,D. (1988) *J. Virol.* **62**, 2291–2299.
10. Bienkowska-Szewczyk,K. and Ehrenfeld,E. (1988) *J. Virol.* **62**, 3068–3072.
11. Pestova,T.V., Maslova,S.V., Potapov,V.K. and Agol,V.I. (1989) *Virus Res.* **14**, 107–118.
12. Meerovich,K., Pelletier,J. and Sonenberg,N. (1989) *Genes Develop.* **3**, 1026–1034.
13. Del Angel,R.M., Papavassiliou,A.G., Fernandez-Tomas,C., Silverstein,S.J. and Racaniello,V. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8299–8303.
14. Kuge,S. and Nomoto,A. (1987) *J. Virol.* **61**, 1478–1487.
15. Earle,J.A.P., Skuce,R.A., Fleming,C.S., Hoey,E.M. and Martin,S.J. (1988) *J. Gen. Virol.* **69**, 253–263.
16. Racaniello,V.R. and Meriam,C. (1986) *Virology* **15**, 498–507.
17. Brown,F., Newman,J., Stott,J., Porter,A., Frisby,D., Newton,C., Carey,N. and Fellner,P. (1974) *Nature (London)* **251**, 342–344.
18. Chumakov,K.M. and Agol,V.I. (1976) *Biochem. Biophys. Res. Commun.* **71**, 551–557.
19. Harris,T.J.R. and Brown,F. (1976) *J. Gen. Virol.* **33**, 493–501.
20. Chumakov,K.M., Chichkova,N.V. and Agol,V.I. (1979) *Dokl. Akad. Nauk SSSR* **246**, 994–996.

21. Rowlands,D.J., Harris,T.J.R. and Brown,F. (1978) *J. Virol.* **26**, 335–343.
22. Vartapetian,A.B., Mankin,A.S., Skripkin,E.V., Chumakov,K.M., Smirnov,V.D. and Bogdanov,A.A. (1983) *Gene* **26**, 189–195.
23. Newton,S.E., Carroll,A.R., Campbell,R.O., Clarke,B.E. and Rowlands,D.J. (1985) *Gene* **40**, 331–336.
24. Pevear,D.C., Calenoff,M., Rozhon,E. and Lipton,H.L. (1987) *J. Virol.* **61**, 1507–1516.
25. Clarke,B.E., Brown,A.L., Currey,K.M., Newton,S.E., Rowlands,D.J. and Carroll,A.R. (1987) *Nucl. Acids Res.* **15**, 7067–7079.
26. Gorbalenya,A.E., Donchenko,A.P. and Blinov,V.M. (1986) *Mol. Genet. Mikrobiol. Virusol.* **1**, 36–41.
27. Romanova,L.I., Blinov,V.M., Tolskaya,E.A., Viktorova,E.G., Kolesnikova,M.S., Guseva,E.A. and Agol,V.I. (1986) *Virology* **155**, 202–213.
28. Kirkegaard,K. and Baltimore,D. (1986) *Cell* **47**, 433–443.
29. Kuge,S., Saito,I. and Nomoto,A. (1986) *J. Mol. Biol.* **192**, 473–487.
30. Kitamura,N., Semler,B.L., Rothberg,P.G., Larsen,G.R., Adler,C.J., Dorner,A.J., Emini,E.A., Hanecak,R., Lee,J.J., van der Werf,S., Anderson,C.W. and Wimmer,E. (1981) *Nature* **291**, 547–553.
31. Racaniello,V.R. and Baltimore,D. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4887–4891.
32. Toyoda,H., Kohara,M., Kataoka,Y., Suganuma,T., Omata,T., Imura,N. and Nomoto,A. (1984) *J. Mol. Biol.* **174**, 561–585.
33. La Monica,N., Meriam,C. and Racaniello,V.R. (1986) *J. Virol.* **57**, 515–525.
34. Stanway,G., Cann,H.J., Hauptmann,R., Hughes,P.J., Clarke,L.D., Mountford,R.C., Minor,P.D., Schild,G.C. and Almond,J.W. (1983) *Nucl. Acids Res.* **11**, 2111–2126.