A model for rearrangements in RNA genomes

Evgeny V. Pilipenko, Anatoly P. Gmyl and Vadam I. Agol*

M. P. Chumakov Institute of Poliomyelitis and Viral Encephalitides, The Russian Academy of Medical Sciences, Moscow Region 142782 and Moscow State University, Moscow 119899, Russia

Received March 14, 1995; Accepted April 18, 1995

ABSTRACT

Engineered mutants of Théler’s murine encephalomyelitis virus (TMEV) and poliovirus having altered spacing between the oligopyrimidine and AUG moieties of a translational control element are known to generate pseudorevertants with deletions or insertions that tend to restore the wild-type structure of this element. The primary structure of the rearranged region of these pseudorevertants suggests that short direct repeats are strongly preferred as parting and anchoring sites during the jumps of the nascent strand 3’ end. When the parting and anchoring sites are separated by a long RNA segment, they can be brought in close proximity by an appropriate folding of the template strand. On the basis of evidence derived from the analysis of the pseudorevertant genomes, it is proposed that a class of RNA rearrangements (some recombinations, deletions, insertions) proceed through the following steps: (I) pausing of the nascent strand caused by misincorporations (or other reasons); (II) dissociation of the RNA polymerase together with the 3’ end of the nascent strand (a kind of proof-reading); and (III) re-annealing of the nascent and template strands (precise or imprecise, but with the 3’ base paired) and resumption of the synthesis.

INTRODUCTION

Viral RNA genomes are known to undergo a variety of rearrangements, such as deletions, insertions and recombinations (1,2). Although the possibility that some of such rearrangements occur post-replicatively cannot rigorously be ruled out, the prevailing belief is that they are largely due to replication errors. The process is believed to be initiated by binding the ribosome (or a ribosomal subunit) to internal cis-acting elements located in the lengthy 5’-untranslated region (5’ UTR) (7-10). Among these cis-elements, a tandem consisting of a conserved oligopyrimidine (or box A) and an AUG (11) was demonstrated to be physiologically significant, the spacing between these two elements being of prime importance (12-14). In some picorna-viruses (polioviruses and other enteroviruses), this tandem lies far upstream of the initiator codon of the polyprotein, whereas in others (TMEV and other cardioviruses as well as foot-and-mouth disease virus), the AUG moiety of the tandem is represented by the initiator codon.

When the distance between box A and the appropriate AUG has been changed by site-directed mutagenesis, viable poliovirus (12,14) and TMEV (15) mutants could be generated. However, the in vitro or in vivo growth potential of the poliovirus and TMEV mutants, respectively, was lowered, and fitter (pseudo)revertants could readily be selected under appropriate conditions (12-14). The phenotypic reversions were due to genetic changes of several types. Of interest for the present work were deletions from, or insertions into, the modified box A/AUG spacer. These naturally occurring rearrangements varied in size and boundaries, but tended to return the spacing between the two elements to an approximately wild-type value. The availability of a variety of rearrangements within limited segments of the viral genomes permitted a detailed analysis of the parting and anchoring sites. In turn, the knowledge of these sites provided insights into a possible mechanism of jumps of the nascent RNA strands and RNA polymerase. Preliminary data were reported in part at a Workshop on Genetic Recombination and Defective Interfering Particles in RNA Viruses (Madrid, 21-23 March, 1994) and at a Meeting on Molecular Biology of Picornaviruses EUROPIC-94 (Helsinki, 6-11 August, 1994).

Short deletions in TMEV RNA

A set of deletions found among the progeny of the engineered TMEV mutant GD/I.27-40 already described (13) and identified more recently (RIO and R11; unpublished) is shown in Figure 1. The distance between the conserved oligopyrimidine element

* To whom correspondence should be addressed
UUUUC (box A) and the initiator AUG1068 in the mutant RNA was 48 nt compared with the wild-type value of 21 nt. The mutant proved to be unstable upon reproduction in the mouse brain, readily producing (pseudo) revertants. Appropriate segments of 12 revertant genomes detected in the homogenates of the infected mice brains were sequenced (in five cases, where attempts to recover viable viruses from these homogenates were undertaken, the isolate genomes were also sequenced and the homogenate findings were confirmed). Nine different deletions were observed; their lengths varied between 20 and 28 nt, shortening thereby the box A/AUG distance to an acceptable value of 20–28 nt. Remarkably, the boundaries of the deletions manifested a strong correlation with the locations of direct repeats.

Within the box A/AUG spacer of GD/I.27-40, there were three copies of each of 4 (or 5) nt-long repeats GAUC(U) (α-repeat) and UUAU(U) (β-repeat), and four copies of a related repeat UAAU(G) (γ-repeat) (Fig. 1). Should the parting and anchoring sites be represented by direct repeats, one could then predict eight major variants of the permitted boundaries for 20–30 nt-long deletions. Five of the observed deletions had precisely the theoretically expected boundaries, involving α1 and α2 (R5 and one variant of R9), β1 and β3 (R10), β2 and β3 (R2), γ1 and γ3 (R3 and R8), as well as γ2 and γ3 (R4). In R6, the jump appeared to involve the same combination of repeats as in R10 (β1 and β3) but to create a run of six consecutive Us one should postulate an insertion of two additional Us. The sixth pair of repeats (γ2 and γ4) appeared to be used for the generation of R11 deletion, a U residue being lost from the relevant oligouridylate. It should be remarked that U insertions/deletions were not infrequently found in the revertant RNAs (see below). Thus, six out of the eight predicted boundaries have been observed among a relatively small set of revertants.

In the two remaining cases, the nascent strand, before jumping, appeared to be elongated a few nucleotides beyond the repeat (indeed, there is no obvious reason why the detachment should occur just at a repeat). However, if the jump takes place outside a repeat, the pairing between the 3' end of the nascent strand and the anchoring site cannot be precise. At least two general possibilities exist. First, the complementarity could be ensured by G–U pairing. This was likely the case with R7, where dissociation of the nascent strand occurred perhaps after copying 4 nt beyond γ3 (Fig. 2a). After landing at γ2, the 3' end turned out to be paired thanks to G–U pairs. In this case, the parting and anchoring sites might be longer than in the previous examples. Such jumps involving G–U pairs in addition to the canonical Watson–Crick pairs will be considered as precise.

During generation of the R1 RNA, the synthesis of the complementary strand very likely went also beyond a repeat (γ4). However, there should be a problem with the subsequent re-annealing of the dissociated 3' terminal oligonucleotide of the nascent strand, which lacked homology to the putative anchoring site. As a solution to this problem, we postulate that an error
cannot be analyzed here. In the cases of insertions of AU include the pause-promoted dissociation, imprecise re-annealing, initiate the rearrangement. The subsequent events might again appears to be the most likely, if not the only possible, explanation for the actual primary structure in the rearranged segment. In turn, such an error should facilitate pausing, termination and eventual dissociation of the 3' end of the nascent strand.

Being attracted by the very idea that replication errors might stimulate polymerase jumps, we propose that such errors may contribute even to the cases, where the jumps appear to involve identical parting and anchoring direct repeats (Fig. 4a). Once the polymerase makes a mistake, its further movement should be hindered. As a consequence, the probability of the enzyme dissociation together with the 3' end of the nascent strand increases. This mechanism can well have a physiological meaning serving as a kind of proof-reading: the error-triggered dissociation of the nascent strand would limit replication of the aberrant genome. It may parenthetically be noted that we never detected point mutations immediately adjoining the repeat. This failure can be interpreted to mean that a mismatch at the 3' position of the jumping end is indeed a serious obstacle to its re-utilization as a primer.

Before being discarded, the dissociated nascent chain may attempt to re-bind a template in such a way that the 3' terminal base of the former becomes paired. The attempts may be particularly successful, if repeats of the parting site can be found, and the perfect repeats are of obvious preference. If the misincorporation occurs immediately after copying the first segment, it might merely a speculation. On the other hand, a misincorporation may take place a few bases after passing the first repeat. The occurrence of such errors can be deduced from the sequencing data (see above). Imprecise pairing after the jump is expected in such cases (Fig. 4b). Thus, a repeat and the deletion boundary may be separated by several bases.

The error-promoted jumping could result in generation of deletions, insertions and recombinant genomes. If a viable genome is created, the rearrangement can contribute to the viral evolution.

**Short deletions in poliovirus RNA**

Some additional genome rearrangements found in our collection of revertants can also be readily explained in the framework of the model proposed above; on the other hand, they also point to some of its limitations.

The engineered poliovirus mutant P23M had a 19 base insert, which increased the length of the box A/AUG spacer to 41 nt instead of the wild type value of 22 nt (Fig. 5). There was an 8 base-long repeat (UUUUAUUG) and numerous shorter, 3
Figure 4. A general model for RNA rearrangements involving direct repeats (underlined) as parting (R2) and anchoring (R1) sites upon precise (a) and imprecise (b) jumping. The complementarity between the 3' end of the nascent strand and the template strand are shown in a key-lock manner. The black symbol corresponds to an erroneously incorporated 3' terminal nucleotide of the nascent strand.

Figure 5. Genome structure of the revertants of P23M. For details see the legend to Figure 1.

Extended deletions in polio and TMEV genomes

Extended deletions (>100 nt) were observed in both the TMEV and poliovirus RNAs. Due to an engineered insertion, the distance between the oligopyrimidine and the initiator AUG in two mutant TMEV RNAs (GD/PV-22 and -12) was made 129 bases-long (Fig. 7a) (15). The mutants proved to be attenuated and generated revertants in mice brains. Four deletions were observed ranging from 103 to 110 bases, approximately restoring the wild-type spacing between the oligopyrimidine and the initiator AUG (13). Again, direct repeats appeared to serve as parting and anchoring sites; in the cases of GD/PV-22R1 and -12R1, the duplex formed by the 3'-end of the nascent strand and the anchoring site might contain 5 and 7 bp, including 2 G-U pairs in each case. Insertion (in GD/PV-22R1) or loss (in GD/PV-12R1) of an U residue has also to be postulated.

Of course, there was a plethora of short repeats between the widely separated parting and anchoring sites. However, the length of deletions was dictated by physiological requirement to have a proper distance between the oligopyrimidine and AUG moieties of the tandem. But how could an appropriate short repeat at such
Figure 7. Genome structure of the GD/PV-22 and -12 revertants (a) and the model explaining their origin (b). In panel (a), 84 nt in the GD/PV-22 and -12 genomes are not shown for the sake of brevity. The parting and anchoring sites of GD/PV-12 for revertants R1 and R2 are marked by thin and thick arrows, respectively. For details see the legend to Figure 1.

(a)

GD/PV-22 1040 129 nt 1071
GD/PV-22R1 GGUGUGAAGUGGCGCGUGCGUUGG::CUAUUAUUGCAUAUUGG
GD/PV-22R2 GGUGUGAAGUGGCGCGUGCGUUGG::CUAUUAUUGCAUAUUGG

GD/PV-12 1046 anchoring
GD/PV-12R1 GGUGUGAAGUGGCGCGUGCGUUGG::CUAUUAUUGCAUAUUGG
GD/PV-12R2 GGUGUGAAGUGGCGCGUGCGUUGG::CUAUUAUUGCAUAUUGG

(b)

GD/PV-12

Figure 8. Genome structure of the PV1/A8 revertants with extended deletions. Nucleotides 586-745 (the PV1 numbering) in the PV1/A8 genome are not shown for the sake of brevity. For details see the legend to Figure 1.

Figure 7b shows the proposed secondary structure near the parting and anchoring sites. These models suggest that the sites may not be as separated as it could appear.

In poliovirus RNA, extended deletions emerged as a response to an 8 base deletion engineered into the spacer between box A and AUG, making the spacer unacceptably short, 14 bases instead of 22. Among the revertants selected from the mutant progeny, there were a few that had lost >150 bases, including the parental cryptic AUG. The revertants did, however, acquire a novel box A/AUG tandem by recruiting the genuine initiator AUG instead of the cryptic AUG (14). Three such deletions were observed (Fig. 8). In one case (PV1/A8-74), the putative jump was precise and involved a tetranucleotide repeat. In another case (PV1/A8-95), the repeat was 3 nt long, and insertion of an additional U residue was required. Only a dinucleotide repeat could be discerned in the last case (PV1/A8-90), which might be converted into a trinucleotide, if a point mutation at the 3' end of the jumping strand is assumed. Again, coming together of the parting and anchoring sites due to the RNA folding could be suggested in at least two out of the three instances (not shown).

Concluding remarks

The TMEV and polio mutant genomes with the engineered insertions into (deletions from) the spacer between the oligopyrimidine and AUG moieties of an important translational control element were under strong pressure to restore the length of the spacer to an approximately wild-type value. This pressure resulted in the selection of revertants having a deletion (insertion) within a limited length range. If the region where deletions should be generated had an array of short direct repeats, these repeats were strongly preferred as parting and anchoring sites by the jumping 3' end of the nascent strand (Fig. 1). In many cases, the evidence for the use of homologous parting and anchoring sites is straightforward. However, in some other cases, the jumps appeared to involve non-homologous sequences adjoining the repeats, suggesting the possibility of imprecise pairing between the nascent and template strands. Moreover, a replication error should be invoked to explain the structure of the genome of certain deletion mutants.

We propose that in all these diverse cases the genome rearrangements may involve the following common steps:
Step I. Pausing of the synthesis of the nascent strand caused by misincorporations (or other reasons).

Step II. Dissociation of the RNA polymerase together with the 3' end of the nascent strand (substitutes for proof-reading).

Step III. Re-annealing of the nascent and template strand (precise or imprecise, but with the 3' base paired) and resumption of the synthesis.

Obviously, the proposed mechanism could explain not only deletions but also insertions as well as recombinations.

Three specifications appear to be warranted. First, the parting (and anchoring) sites encountered in the mutants described here are generally AU-rich, the only exception from this rule being GD-4OR5 (Fig. 1). It could be speculated that the pause-promoted dissociation of the 3'-end of the nascent strand from the parting site was facilitated by a low thermodynamic stability of the local duplex. Secondly, insertions or deletion of one or two U residues in the rearranged area (and outside it; e.g., GD/PV-22R2) appeared to be quite common. Usually, the U insertions/deletions affected oligouridylate stretches. Thus, certain homopolymeric sequences may be sensed by the viral polymerase as 'slippery', allowing back and forth movements of the enzyme uncoupled from the formation of a phosphodiester bond. Thirdly, the apparently beneficial role of G-U pairs in the annealing of the jumping end of the nascent strand and the anchoring site on the template might be interpreted to mean that the rearrangements take place, at least in some cases, during the synthesis of the (-) strand as shown on our schematic figures and as suggested earlier (4,5).

It should be admitted that the system described here had peculiarities not necessarily shared by other RNA genomes experiencing rearrangements: first, there was a very strong pressure for deletions having a certain length, and secondly, the template had several sets of multiple direct repeats. These peculiarities made it possible to analyze in detail factors affecting the choice of parting and anchoring sites. However, rearrangements in picornavirus genomes do not necessarily involve direct repeats (4,5,17-19; A. P. Gmyl et al., unpublished observations). Therefore, the proposed model intends to explain the mechanism of only a class of rearrangements, which occur preferably if some requirements are met.

ACKNOWLEDGEMENTS

This study was supported in part by grants from the Human Frontier Science Program Organization (RG 470/93-M), International Science Foundation (RA1000) and Russian Foundation for Basic Research (93-04-07594). VIA is a Soros Professor.

REFERENCES