

## Frequent Polymerase Errors Observed in a Restricted Area of Clones Derived from the Attachment (G) Protein Gene of Respiratory Syncytial Virus

PATRICIA A. CANE,\* DAVID A. MATTHEWS, AND CRAIG R. PRINGLE

*Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom*

Received 14 July 1992/Accepted 2 November 1992

**Sequence analysis of a large number of clones derived from the carboxy-terminal one-third of the attachment (G) protein gene of subgroup A respiratory syncytial viruses revealed a region very prone to polymerase errors which resulted mainly in frameshifts because of the insertion or deletion of adenosine residues in some but not all runs of such residues. Such mutations were detected in 14% of clones derived from mRNA, 58% of clones derived from genomic-sense RNA, and 50% of clones derived from in vitro-transcribed RNA. This phenomenon appears to be dependent on the template sequence.**

Respiratory syncytial virus (RSV) is a major cause of lower respiratory tract infection in infants and vulnerable adults. The virus causes repeated infections of individuals, and it is unclear whether this is due to genetic variability, an inadequate immune response, or both. RSV isolates can be divided into two serological subgroups, A and B. Between these subgroups the greatest antigenic and sequence differences occur in the attachment (G) protein, which has only 53% amino acid identity between prototype strains of subgroups A and B (8). Within the subgroups, the G protein is also highly variable, with up to 20% amino acid variability observed between subgroup A strains (3) and up to 9% variability observed between subgroup B strains (12).

Characterization of RSV mutants selected for resistance to neutralizing anti-G monoclonal antibodies revealed a novel mechanism involving frameshift mutations for the generation of such mutants (7). Between nucleotides 588 and 654, using the nucleotide numbering of Wertz et al. (16) for the mRNA from the G gene, there are three runs of six or seven adenosine (A) residues. Some neutralization escape mutants showed deletions or insertions of A's into these areas, causing frameshift mutations which altered the amino acid sequence of the carboxy-terminal one-third of the protein. One mutant showed two frameshift mutations, the second mutation restoring the original amino acid sequence after seven residues that had been changed (7). This work indicated that viable (in cultured cells at least) RSV could exist despite radical alteration of the carboxy-terminal portion of the G protein. The attachment function of the G protein in culture does not appear to depend on the integrity of the carboxy-terminal one-third of the protein. Also, it has been shown that vaccinia virus recombinants expressing truncated G proteins could still induce neutralizing antibodies even when missing the 68 carboxy-terminal amino acids (9). The biological significance of the escape mutant work was underlined by Sullender et al. (12), who, when examining variation in natural isolates of subgroup B RSV, found that in one strain a mutation caused by a single A deletion and subsequent A insertion caused a shift in reading frame for seven amino acids relative to other G genes analyzed.

Frameshift mutations changing the entire carboxy-termi-

nal region of the G protein have not been reported for natural isolates (3, 12). However, in this paper we report the presence of such mutations in a minority of clones generated by reverse transcription and amplification by the polymerase chain reaction (PCR) from mRNAs of the G-protein gene of many virus isolates. These mutations involve mainly two oligo(A) tracts between nucleotides 588 and 630, together with occasional mutations in a run of four A's between these tracts, suggesting that this region of the G gene is very prone to polymerase error; frameshift mutations were not detected in a third run of six A's starting at 649. Where cDNA was derived from genomic-sense RNA, over 50% of clones showed frameshift mutations affecting the same oligo(A) tracts; a similar high incidence of such mutations was also observed in clones derived from in vitro-transcribed RNA, indicating that the phenomenon was not related to a specific polymerase. The biological significance of these observations is unclear, but they may indicate that there is a hot spot for mutation in this part of the RSV G gene.

Clones of the G-gene mRNA between nucleotides 531 and 917 were derived by reverse transcription (avian myeloblastosis virus reverse transcriptase) with an oligo(T)-based primer and by polymerase chain reaction (PCR) as previously described (3). The virus strains were isolated in Birmingham (those prefixed RSB) or Liverpool (those prefixed RSL), United Kingdom, during the period from 1989 to 1991, except for strain RSS2, which was isolated in Newcastle, United Kingdom, in 1976. Sequencing was carried out with either a Sequenase kit (U.S. Biochemicals) or a double-stranded DNA cycle sequencing kit (Bethesda Research Laboratories). For screening large numbers of clones to look for frameshifts, initially only one nucleotide type was examined for each clone.

To obtain clones derived from genomic-sense RNA from strain RSB89-6190, we primed reverse transcription with a primer homologous to nucleotides 531 to 549. Following inactivation of the reverse transcriptase by heating at 94°C for 10 min, the RNA was digested with RNase A (Sigma) at a final concentration of 40  $\mu\text{g ml}^{-1}$ . PCR was carried out with the same primer as that used for cDNA synthesis, together with a primer complementary to nucleotides 889 to 909 (3), extended at the 5' end with GGAATT to act as a linker for cloning. The PCR product was restricted with

\* Corresponding author.

TABLE 1. Mutations observed in mRNA clones of the G genes of RSV isolates

Strain <sup>a</sup>	Clone no.	Mutation	Frameshift	Amino acid effect
RSB89-6190	31	+1A at 594	+1	194–298 changed; no termination
	33	-1G at 598	-1	196–255 changed; 256–298 lost
	34	+1A at 630	+1	207–298 changed; no termination
	51	G to A at 564	None	Termination at 183
RSB90-8106	21	+2A at 630	+2	207–298 changed; +3 extra
RSB90-8260	5	-1A at 663	-1	216–297 changed; +3 extra
	11	-1A at 607	-1	199–297 changed; +3 extra
	18	-1A at 594	-1	193–297 changed; +3 extra
RSB91-115	3	+3A at 594	None	+K after 194
	16	-1A at 594	-1	193–208 changed; 209–297 lost
RSL459	9	+1A at 607	+1	198–270 changed; 271–297 lost

<sup>a</sup> We examined 23 clones from RSB89-6190, 7 clones from RSB90-8106, 20 clones from RSB90-8260, 5 clones from RSB91-115, and 2 clones from RSL459.

*EcoRI* and *HaeIII*, gel purified, and ligated into Bluescribe (Vector cloning systems) digested with *EcoRI* and *SmaI*.

To examine the frequency of mutations in a gene other than the G gene, we generated clones of part of the phosphoprotein (P) gene. The P gene was selected because the P gene of other paramyxoviruses is a site of transcriptional editing (13, 14). The P gene of strain RSS2 was amplified between nucleotides 1 and 747 by previously described methods (4). The PCR product was restricted with *EcoRI* and *HincII*, and the fragment between nucleotides 37 and 144 was gel purified and inserted into Bluescribe restricted with *EcoRI* and *SmaI*. This portion of the P gene includes a run of six A's at nucleotides 123 to 128 (11).

In an attempt to determine whether the presence of mutations was an in vitro artifact, we also derived cDNA clones from RNA made by in vitro transcription. Plasmid DNA from a clone of nucleotides 531 to 917 of the G gene of isolate RSB89-6190 inserted into Bluescribe vector following *PstI* and *AvaI* digestion (3) was restricted with *PvuII* and *EcoRI*, and the appropriate fragment was gel purified with Geneclene (Bio 101). RNA transcripts were made in vitro with T3 RNA polymerase (Promega) as specified by the manufacturer. RQ1 DNase (Promega) was added, and the RNA was phenol-chloroform extracted and ethanol precipitated. cDNA synthesis and PCR were carried out as above with primers based on oligo(T) and nucleotides 531 to 549. To eliminate a band appearing in control PCR reactions set up with "cDNA" made without reverse transcription, it was found necessary to dilute the cDNA 1,000-fold before performing PCR. Although some of the DNA of the PCR product may have been derived directly from residual plasmid DNA, the vast majority could be assumed to be derived via RNA. These PCR products were digested with *AvaI* and *HaeIII*, gel purified, and ligated into Bluescribe digested with *AvaI* and *HincII*.

We examined 80 clones of the G gene (nucleotides 531 to 917) derived from mRNA from 18 isolates of subgroup A RSV. We completely sequenced 55 clones for the region between nucleotides 560 and 665. We initially screened 40 clones for the presence of frameshift mutations by single-nucleotide sequencing only; 25 of the clones that showed no frameshifts were not analyzed further, and the remaining 15 clones, which comprised some normal controls together with mutant clones, are included among the 55 that were completely sequenced. We found that 11 clones (14%) had

mutations in this region; 10 of the mutations were insertions or deletions of A's, of which 9 (11%) resulted in frameshifts and 1 was a G-to-A substitution. The nature and effect of these mutations are shown in Table 1, and the positions of the mutations are illustrated in Fig. 1.

The most common region for mutation (four clones) was a run of six A's ending at nucleotide 594: mutations here included both insertions and deletions. This run of six A's is present in all 18 strains examined, with the exception of RSB89-642, which has AAAAAG, and RSB90-8336 in which nucleotide 588 is also A, giving seven A's but no frameshift (3; unpublished data).

One clone showed an insertion and one showed a deletion of one A at a run of four A's finishing at nucleotide 607, and two clones showed one or two insertions of A's at nucleotide 630, where there are seven A's in all strains of subgroup A RSV isolates examined except RSB89-642, which has AAA GAAG (3). One clone showed a deletion of one A at 633. In contrast, however, no mutations were observed in the run of six A's starting at nucleotide 649 in any of the 80 clones.

The consequences of these mutations varied from strain to strain depending on the subsequent sequence, but they included immediate termination, changing of some amino acids and then termination, and changing of all subsequent amino acids. Some frameshifts resulted in a reading frame that showed no termination; if these strains are assumed to have the same intergenic sequence as the prototype sub-

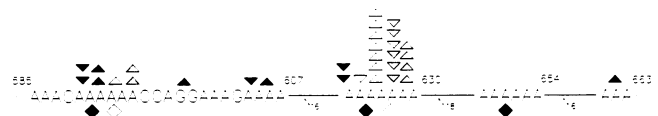


FIG. 1. Schematic diagram of the region of the RSV G gene (mRNA sense) between nucleotides 585 and 663, indicating oligo(A) tracts where frameshift mutations have been observed. The sequence shown is that of strain A2 (16); some other strains show minor variations (3). The number of insertion and deletion symbols indicates the number of clones detected with that mutation. Symbols: ▼, insertions observed in mRNA clones; ▽, insertions observed in genomic clones; ▲, deletions observed in mRNA clones; △, deletions observed in genomic clones; ◆, sites of insertions observed in clones from escape mutants (7); ▲, deletions observed in mRNA clones; △, deletions observed in genomic clones; ◇, sites of deletions observed in clones from escape mutants (7).

group A strain A2 (6), termination could occur on a polycistronic mRNA at termination codon UGA at fusion (F) protein gene nucleotides 100 to 102.

Nineteen clones derived from genomic RNA of RSB89-6190 were sequenced for A residues alone between nucleotides 566 and 665. Only six of these clones gave the consensus sequence obtained from clones derived from mRNA. One clone showed loss of one A in the run of six A's ending at nucleotide 594; five clones showed insertion of one A in the run of seven A's at nucleotide 630; one clone showed insertion of two A's in this region; three clones showed deletion of one A in this region; one clone showed deletion of one A at both nucleotide 594 and nucleotide 630; and point mutations resulting in the replacement of an A residue but no frameshift were seen in two clones, one at nucleotide 577 and one at 649. No frameshift mutations were observed in the run of six A's starting at nucleotide 649. The positions of these various mutations are shown in Fig. 1.

We examined 31 clones of part of the P gene of strain RSS2 between nucleotides 37 and 144 by sequencing one nucleotide only; no frameshift mutations were detected in the run of six A's between nucleotides 123 and 128. It would therefore seem that the presence of oligo(A) tracts alone does not account for the high error rate seen in the G gene.

We sequenced 18 clones derived from PCR of in vitro-transcribed RNA for one nucleotide only. Nine (50%) of these clones showed frameshift mutations: one clone showed deletion of one A at nucleotide 594, one showed a two-nucleotide deletion in the region of nucleotide 630, six showed a one-nucleotide deletion in the region of nucleotide 630, and one showed an insertion of one A at nucleotide 630. Again, no frameshift mutations were observed in the run of six A's starting at nucleotide 649.

Thus, higher levels of mutations were observed in clones derived from genomic RNA and T3 transcripts than in clones from mRNA. These experiments have in common that the reverse transcription of genomic RNA and the T3 transcription occur on the negative-sense template as opposed to the reverse transcription of mRNA, which is on the positive-sense template; also, the reverse transcriptase, T3, and 5' PCR primers were closer to the region in which errors occurred than in the clones derived from full-length G mRNA. It seems likely that many of the mutations observed are a consequence of error during T3 or reverse transcription or during PCR and are therefore probably artifacts. For example, human immunodeficiency virus reverse transcriptase is prone to produce single frameshift mutations in homopolymeric runs (1), although the avian myeloblastosis virus reverse transcriptase used in the experiments reported here is 10-fold more accurate than human immunodeficiency virus reverse transcriptase (1, 10). Nevertheless, undoubtedly viral genomes mutated in this region do exist, or it would not be possible to select monoclonal antibody escape mutants of this type. In addition, in the escape mutant experiments mRNA heterogeneity was shown by direct mRNA sequencing, so at least in those experiments the frameshift mutations observed were not PCR artifacts (7).

The only firm conclusion that can be drawn is that errors frequently arise during transcription of this part of the RSV genome where runs of A's occur in the mRNA sense. Use of reverse transcriptase or T3 polymerase followed by PCR is probably accentuating the susceptibility of this part of the genome to error during replication. These hot spots for mutation appear to depend on surrounding sequence in that errors are concentrated in two of the oligo(A) tracts (beginning at nucleotides 589 and 624) examined and are very rare

in a third (beginning at nucleotide 649). These error-prone regions appear to be conserved within subgroup A viruses, and runs of A's are also present in a similar region of subgroup B viruses.

Mechanisms for the introduction of nontemplated residues into mRNA by "stuttering" have been proposed for other nonsegmented negative-strand RNA viruses such as mumps virus, measles virus, simian virus 5, and Sendai virus (15). In these viruses the P gene gives rise to two mRNAs by insertion of nontemplated G residues in a specific run of three to six G's, allowing ribosomal access to a second reading frame. Two proteins with common N-terminal sequences but alternative C-terminal sequences of different lengths can thus be synthesized (5, 13, 14). This is a similar effect to that reported here for the G gene of RSV in that if the different mRNAs described here are generated naturally during transcription, G proteins with alternative C terminals would be produced. However, for RSV, it is not known whether such alternative gene products are synthesized during normal transcription rather than as a consequence of selection of mutants under antibody pressure during replication of the genome. Accumulations of polymerase errors have also been found in vesicular stomatitis virus, for which it is suggested that slippage takes place on a poly(U) tract in a noncoding part of the G gene in a manner analogous to the slippage that occurs during the polyadenylation of mRNA (2). However, the mechanism behind the polymerase errors described here may differ in that many of the mutations involve deletion of A's rather than insertion. The influence of adjacent sequences in determining the susceptibility of oligo(A) tracts to error will be investigated further.

Frameshifts changing the carboxy-terminal one-third of the RSV G protein were associated with loss of reaction with a wide selection of monoclonal antibodies and with a polyclonal serum raised against purified soluble G protein (7). It may be beneficial for the virus, therefore, to be able to evade neutralization by incorporating proteins of various antigenic specificities into its envelope during replication. It is a significant feature of the RSV G gene that alternative long open reading frames are present in the carboxy-terminal one-third of the gene. The conservation of these alternative frames closely following an area apparently extremely susceptible to polymerase error could have biological significance, particular in view of the results of Sullender et al. (12), who detected frameshifts in a natural isolate.

This work was supported by MRC program grant PG8322715.

#### REFERENCES

1. **Bebenek, K., J. Abbotts, J. D. Roberts, S. H. Wilson, and T. A. Kunkel.** 1989. Specificity and mechanism of error-prone replication by human immunodeficiency virus-1 reverse transcriptase. *J. Biol. Chem.* **264**:16948-16956.
2. **Bisel, P. A., and S. T. Nichol.** 1990. Polymerase errors accumulating during natural evolution of the glycoprotein gene of vesicular stomatitis virus Indiana serotype isolates. *J. Virol.* **64**:4873-4883.
3. **Cane, P. A., D. A. Matthews, and C. R. Pringle.** 1991. Identification of variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. *J. Gen. Virol.* **72**:2091-2096.
4. **Caravokyri, C., and C. R. Pringle.** 1992. Effect of changes in the nucleotide sequence of the P gene of respiratory syncytial virus on the electrophoretic mobility of the P protein. *Virus Genes* **6**:53-62.
5. **Cattaneo, R., K. Kaelin, K. Bacsko, and M. A. Billeter.** 1989. Measles virus editing provides an additional cysteine-rich protein. *Cell* **56**:759-764.

6. Collins, P. L., L. E. Dickens, A. Buckler-White, R. A. Olmsted, M. K. Spriggs, E. Carmargo, and K. V. W. Coelingh. 1986. Nucleotide sequences for the gene junctions of human respiratory syncytial virus reveal distinctive features of intergenic structure and gene order. *Proc. Natl. Acad. Sci. USA* **83**:4594-4598.
7. Garcia-Barreno, B., A. Portela, T. Delgado, J. A. Lopez, and J. A. Melero. 1990. Frame shift mutations as a novel mechanism for the generation of neutralization resistant mutants of human respiratory syncytial virus. *EMBO J.* **9**:4181-4187.
8. Johnson, P. R., M. K. Spriggs, R. A. Olmsted, and P. L. Collins. 1987. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. *Proc. Natl. Acad. Sci. USA* **84**:5625-5629.
9. Olmsted, R. A., B. R. Murphy, L. A. Lawrence, N. Elango, B. Moss, and P. L. Collins. 1989. Processing, surface expression, and immunogenicity of carboxy-terminally truncated mutants of G protein of human respiratory syncytial virus. *J. Virol.* **63**:411-420.
10. Roberts, J. D., B. D. Preston, L. A. Johnson, A. Soni, L. A. Loeb, and T. A. Kunkel. 1989. Fidelity of two retroviral reverse transcriptases during DNA-dependent DNA synthesis in vitro. *Mol. Cell. Biol.* **9**:469-476.
11. Satake, M., N. Elango, and S. Venkatesan. 1984. Sequence analysis of the respiratory syncytial virus phosphoprotein gene. *J. Virol.* **52**:991-994.
12. Sullender, W. M., M. A. Mufson, L. J. Anderson, and G. W. Wertz. 1991. Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. *J. Virol.* **65**:5425-5434.
13. Thomas, S. M., R. A. Lamb, and R. G. Paterson. 1988. Two mRNA's that differ by two non-templated nucleotides encode the amino co-terminal proteins P and V of the paramyxovirus SV5. *Cell* **54**:891-902.
14. Vidal, S., J. Curran, and D. Kolakofsky. 1990. Editing of the Sendai virus P/C mRNA by G insertion occurs during mRNA synthesis via a virus-coded activity. *J. Virol.* **64**:239-246.
15. Vidal, S., J. Curran, and D. Kolakofsky. 1990. A stuttering model for paramyxovirus P mRNA editing. *EMBO J.* **9**:2017-2022.
16. Wertz, G. W., P. L. Collins, Y. Huang, C. Gruber, S. Levine, and L. A. Ball. 1985. Nucleotide sequence of the G protein gene of human respiratory syncytial virus reveals an unusual type of viral membrane protein. *Proc. Natl. Acad. Sci. USA* **82**:4075-4079.