JOURNAL OF VIROLOGY, Mar. 1986, p. 1207–1209
0022-538X/86/031207-03$02.00/0

Vol. 57, No. 3

Heterogeneity in Base Sequence among Different DNA Clones Containing Equivalent Sequences of Rotavirus Double-Stranded RNA

CARLOS F. ARIAS,* SUSANA LÓPEZ, AND ROMILIO T. ESPEJO

Departamento de Biología Molecular, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, 04510, Mexico

Received 1 July 1985/Accepted 1 November 1985

The nucleotide sequences for several complementary DNA clones of the rotavirus genome were determined. When the sequences obtained from different clones for the same regions (16,000 bases) were compared, differences in eight base positions were observed. These discrepancies, approximately 1 in 2,000 bases, may be due to (i) differences in individual RNA genomes resulting from multiple passages; (ii) infidelity of DNA synthesis in the cloning procedure; or (iii) both factors. Whatever the cause, this frequency of base substitution found in sequences of complementary DNA obtained from the same isolate should be considered when comparing DNA sequences obtained from independent isolates. On the other hand, the frequency of base changes observed suggests that the rotavirus genome is very conserved since the virus used for cDNA synthesis has been continuously passaged for 6 years without plaque purification.

When the base sequence for a virus type is reported, it is customary to compare the sequence with those of different virus types, whether obtained by the same or different methods and investigators. For example, when two independently reported nucleotide sequences for segment 9 of the rotavirus SA11 are compared (1, 4), a total of 12 nucleotide differences can be observed.

Because of our concern over the meaning of the base substitutions observed among equivalent genes in rotavirus isolates (1, 4, 7), we sequenced the same regions of the double-stranded (ds) RNA genome of rotavirus several times to determine the possible cause(s) for these differences. Intrinsic differences may occur because, as originally proposed for QB RNA bacteriophage (6), the genome of each rotavirus isolate should not be considered as a unique, defined sequence, but rather as a weighted average of a large number of different individual sequences. Differences in the base sequence may also arise because of the infidelity of the DNA polymerases employed for cloning the RNA (2). The sequence obtained from the cloned cDNA therefore probably corresponds to one of the different individual sequences that make up the viral genomic population. This sequence may contain nucleotide substitutions that originated through incorrect base pairing during DNA synthesis. In this paper, we report that the sequences obtained from cDNA clones for equivalent regions of the genome of a multiply passaged population of rotavirus SA11 differed in about 1 position per 2,000 sequenced bases.

SA11 rotavirus was obtained form H. H. Malherbe in 1977 and has been repeatedly propagated for more than 6 years without being plaque purified. Total genomic dsRNA was prepared from this virus stock and, after denaturation, was used as the template for the synthesis of cDNA by the DNA polymerase from avian myeloblastosis virus (AMV) as previously reported (1). The cloning of this cDNA and the selection of bacterial clones containing hybrid plasmids with sequences of different dsRNA genes has been described previously (1). Several clones were obtained for each of the dsRNA segments. Most of the clones were partial copies of the corresponding gene. A total of 16 clones for segment 4, 19 clones for segment 6, 12 clones for segment 8, and 36 clones for segment 9 were partially or fully sequenced by the procedure of Maxam and Gilbert (15). On the average, 200 bases were read from each clone.

When a total of 16,600 sequenced bases, representing several regions of the SA11 genome (some regions were read more than twice), were compared, differences at eight positions were observed. From this value, a frequency of one nucleotide substitution per 2,070 bases sequenced was obtained. One of these differences in base sequence is illustrated in Fig. 1. Each difference observed is explained in detail in Table 1.

The 16,600 bases compared amounted to only 20% of the rotavirus genome, because these bases represented regions sequenced 2 or more times and up to 11 times in one case (in segment 9). The base substitution frequency obtained from this study is similar to that reported by Domingo et al. (6) for QB RNA using a completely different methodology. In two-dimensional polyacrylamide gel electrophoresis of the oligonucleotides resistant to RNase T_1, they observed that about 15% of the clones (plaque-purified phages) that were derived from a multiply passaged QB population showed fingerprint patterns which deviated from that of the RNA of the total population. Since the oligonucleotides which they examined represented only about 10% of the entire genome, they concluded that for each nucleotide change detected there must have been about nine not detected. Since four different types of variant (about 16%) were found among 25 clones, they concluded that the RNA sequence of each plaque-purified virus differed on the average by 1.6 positions from the composite sequence of the population from which it was derived. We used the same reasoning to analyze our data. Sequences of about 200 bases (representing about 1% of the rotavirus genome) were compared from each clone. Since among 83 clones eight different types of variant (about 10%) were found, each genome in the population differs in 10 positions. If correction is made for the genome size of 4,500 nucleotides in QB (9) and of 20,000 nucleotides in rotavirus

* Corresponding author.

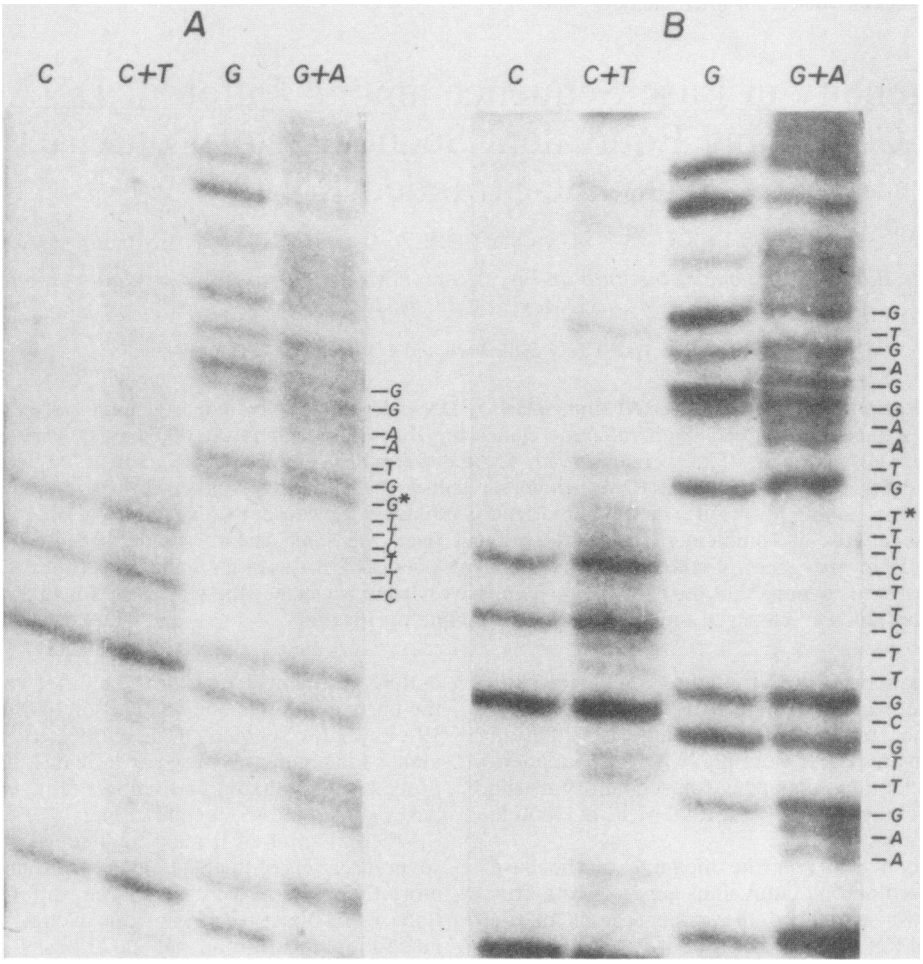

FIG. 1. Sequence analysis of two different cDNA clones derived from dsRNA segment 9 of rotavirus SA11. A portion of the sequencing gel autoradiogram that displayed a single nucleotide change between these two clones is shown. Nucleotide 267 (marked with an asterisk) is clearly shown to be altered from a G (clone A and the other clones of this region) to a T (in clone B, the variant clone). The sequence shown corresponds to a cDNA copy of the plus strand of the dsRNA gene; therefore, the change in the plus strand would be C → A. For each clone, four sequencing reactions described by Maxam and Gilbert (15) are shown.

(3), differences of 1 per 2,800 nucleotides for QB and 1 per 2,070 nucleotides for rotavirus are observed. However, when sequences obtained from cDNA are compared, an additional cause for the observed differences should be considered. In this case, nucleotide substitution may occur during DNA synthesis with reverse transcriptase (AMV DNA polymerase) or during filling in by DNA polymerase I of the partially ds molecules obtained after hybridization (1). The frequency of incorrect incorporation of paired bases by the latter enzyme (about 10^{-6}) is well below the differences observed in our comparison (12). However, AMV DNA polymerase, which lacks the proofreading 3′→5′ exonuclease activity, is more error prone. The frequency of error observed with this enzyme is approximately 1 in 600 when copying homopolymer templates and 1 in 6,000 when copying alternating copolymer templates (2). Frequencies from 1 in 329 to 1 in 17,000 have been observed when ϕX174 DNA is copied (5, 12). Since the values reported are variable and depend on many factors, such as conditions of synthesis, nature of the mismatch considered, and even the preparation of the DNA polymerase used (2, 5, 13), it is not possible to infer to what extent the AMV DNA polymerase contributes to the differences observed between clones. However, the observation that every difference found was silent or conservative (with the exception of the Gly to Glu change in gene 9; Table 1) indicates that these differences may be maintained in the population as viable variants that might be expected for a multiply passaged virus.

TABLE 1. Nucleotide changes found in a cDNA library of rotavirus SA11

Gene segment	Nucleotide no.[a]	Nucleotide change	Triplet change	Amino acid change
4		C → G	CTA → GTA	Leu → Val
6	305	T → C	AAT → AAC	Silent
6	320	G → A	GAG → GAA	Silent
6	385	C → G	GCC → GGC	Ala → Gly
9	267	C → A	CCA → ACA	Pro → Thr
9	311	T → C	GCT → GCC	Silent
9	385	G → A	GGA → GAA	Gly → Glu
9	527	C → T	AAC → AAT	Silent

[a] The nucleotide number was assigned by the methods of Estes et al. (segment 6) (8) and Arias et al. (segment 9) (1). For RNA segment 4, no position is shown since the sequence of this segment has not yet been completed (14).

The assumption that the differences are randomly distributed may not hold for the whole extent of the genome: in viable rotaviruses some dsRNA segments or some regions of these segments probably accept fewer nucleotide changes than do others. In fact, the existence of conserved sequences in the terminal regions of the rotavirus dsRNA segments is well documented (3, 11). However, in this work, the regions compared correspond in every case to the coding regions of the segment. When the differences found for each analyzed segment were considered, one difference was found in 16 of the clones analyzed for segment 4, three in 19 clones analyzed for segment 6; zero in 12 clones analyzed for segment 8, and four in 36 clones analyzed for segment 9. Even though the values seemed at first sight quite dissimilar, they are not statistically different from those expected if the differences follow a Poisson distribution. Values of P greater than 0.1 were obtained when the frequencies (number of different clones/total number of clones examined) obtained for each segment were compared with those expected to be found in a Poisson distribution.

Twelve nucleotides were found to differ when two sequences of dsRNA segment 9 of rotavirus SA11 were compared, independently reported by Both et al. (4) and Arias et al. (1). If the infidelity of the AMV DNA polymerase under the conditions employed by Both et al. is less than or the same as that under our conditions, the differences observed are probably due to the divergence between independently passaged virus populations because, according to the difference found between equivalent regions of about 1 in 2,000, only 0.5 difference owing either to heterogeneity in the genome or to miscopying would be expected in the 1,062-nucleotide-long segment 9 (1, 4).

A difference of about 1 per 1,000 nucleotides has been reported between cDNA clones of vesicular stomatitis virus, a single-stranded RNA virus (16). Considering that vesicular stomatitis virus was plaque purified before cDNA synthesis whereas the rotavirus used in this study was not, it appears that the rotavirus genome may be more conserved than the vesicular stomatitis virus genome.

The high mutation rate and the consequent rapid evolution and large variability observed in single-stranded RNA viruses has been explained by the lack of a proofreading mechanism for single-stranded RNA replication (10). The high conservation which was observed between cDNA clones obtained from a rotavirus which has been continuously passaged for 6 years without plaque purification suggests a very low mutation rate for dsRNA which, in turn, suggests the possible existence of a proofreading mechanism in its RNA polymerase. If the sequence of the genome segments of rotavirus is indeed highly conserved, the differences observed between independent isolates probably arise by reassortment of segments in mixed infections rather than through sequential point mutations.

LITERATURE CITED

1. **Arias, C. F., S. López, J. R. Bell, and J. H. Strauss.** 1984. Primary structure of the neutralization antigen of simian rotavirus SA11 as deduced from cDNA sequence. J. Virol. **50**:657–661.
2. **Battula, N., and L. A. Loeb.** 1974. The infidelity of avian myeloblastosis virus deoxyribonucleic acid polymerase in polynucleotide replication. J. Biol. Chem. **249**:4086–4093.
3. **Both, G. W., A. R. Bellamy, J. E. Street, and L. S. Siegman.** 1982. A general strategy for cloning double-stranded RNA: nucleotide sequence of the simian-11 rotavirus gene 8. Nucleic Acids Res. **10**:7075–7088.
4. **Both, G. W., J. S. Mattick, and A. R. Bellamy.** 1983. The serotype-specific glycoprotein of simian 11 rotavirus: coding assignment and gene sequence. Proc. Natl. Acad. Sci. USA **80**:3091–3095.
5. **Copinathan, K. P., L. A. Weymonth, T. A. Kunkel, and L. A. Loeb.** 1979. Mutagenesis in vitro by DNA polymerase from an RNA tumor virus. Nature (London) **278**:857–859.
6. **Domingo, E., D. Sabo, T. Taniguchi, and C. Weissman.** 1978. Nucleotide sequence heterogeneity of an RNA phage population. Cell **13**:735–744.
7. **Elleman, T. C., P. A. Hoyme, M. L. Dyall-Smith, I. H. Holmes, and A. A. Azad.** 1983. Nucleotide sequence of the gene encoding the serotype-specific glycoprotein of UK bovine rotavirus. Nucleic Acids Res. **11**:4689–4701.
8. **Estes, M. K., B. B. Mason, S. Crawford, and J. Cohen.** 1984. Cloning and nucleotide sequence of the simian rotavirus gene 6 that codes for the major inner capsid protein. Nucleic Acids Res. **12**:1875–1887.
9. **Fiers, W.** 1979. Structure and function of RNA bacteriophages. Compr. Virol. **13**:69–104.
10. **Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. Vande Pol.** 1982. Rapid evolution of RNA genomes. Science **215**:1577–1585.
11. **Imai, M., K. Akatani, N. Ikegami, and Y. Furuichi.** 1983. Capped and conserved terminal structures in human rotavirus genome double-stranded RNA segments. J. Virol. **47**:125–136.
12. **Kunkel, T. A., F. Eckstein, A. S. Mildvan, R. M. Koplitz, and L. A. Loeb.** 1981. Deoxynucleoside (1-thio) triphosphates prevent proofreading during *in vitro* DNA synthesis. Proc. Natl. Acad. Sci. USA **78**:6734–6738.
13. **Kunkel, T. A., and L. A. Loeb.** 1979. On the fidelity of DNA replication. Effect of divalent metal ion activators and deoxyribonucleoside triphosphate pools on *in vitro* mutagenesis. J. Biol. Chem. **254**:5718–5725.
14. **López, S., C. F. Arias, J. R. Bell, J. H. Strauss, and R. T. Espejo.** 1985. Primary structure of the cleavage site associated with trypsin enhancement of rotavirus SA11 infectivity. Virology **144**:11–19.
15. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
16. **Schubert, M., G. G. Harmison, and E. Meier.** 1984. Primary structure of the vesicular stomatitis virus polymerase (L) gene: evidence for a high frequency of mutations. J. Virol. **51**:505–514.