## Occurrence of partial deletion and substitution of the *src* gene in the RNA genome of avian sarcoma virus

Sager in 199

(deletion mutant/35S RNA/polyacrylamide gel electrophoresis/oligonucleotide mapping/heteroduplex)

MICHAEL M. C. LAI, SYLVIA S. F. HU, AND PETER K. VOGT

Department of Microbiology, University of Southern California, School of Medicine, Los Angeles, California 90033

Communicated by Norman Davidson, July 28, 1977

ABSTRACT The genome size of 20 transformation-defective (td) viruses derived from different strains of Rous sarcoma viruses [Prague (subgroups A and C), Schmidt-Ruppin (subgroups A and D) (SR-D), Bratislava 77, and Carr-Zilber (subgroup D)] was examined by polyacrylamide gel electrophoresis. All of the td viruses except td SR-D have 35S RNA of the same size—i.e., class b RNA. Two of five td SR-D viruses examined have a slightly larger RNA, corresponding to a td deletion that is about 25% smaller than that of class b RNA. However, the RNase  $T_1$ -oligonucleotide fingerprints of all the td SR-D viruses are identical, lacking two sarcoma-specific oligonucleotides. The fingerprints of these viruses also showed a minor oligonucleotide present at very low concentration. A study of heteroduplex molecules formed between genome-length cDNA made from wild-type SR-D and 35S RNA of td SR-D showed a deletion loop of 2.0 and 1.5 kilobases, respectively, at the map position of the src gene for these two classes of td SR-D viruses, confirming the results of polyacrylamide gel electrophoresis. In addition, some heteroduplex molecules with a substitution loop of 0.6-0.7 kilobase at the same site as the deletion loop were observed in all five of the td SR-D viruses. We conclude that some of the td SR-D viruses have a partially deleted src gene and that all of the td SR-D viruses have incorporated heterologous sequences of distinct length in some RNA molecules at the position of the src gene. The nature and origin of these heterologous sequences are discussed.

Transformation defective (td) viruses arise spontaneously in stocks of nondefective avian sarcoma viruses at a high frequency (1-3). These td viruses have a smaller 35S RNA genome, termed class b, than the parental sarcoma viruses which have size class a RNA (4-6). td viruses are generated by a deletion of *src*, the viral gene responsible for the induction and maintenance of oncogenic transformation in vitro and in vivo (6). The deleted sequence has been estimated to represent about 13–15% of the avian sarcoma virus genome and appears to span most of the temperature-sensitive lesions in the src gene (7-9). Most independently isolated td viruses have class b RNA of the same size, suggesting either that the src deletion of this size occurs at high frequency or that it is selected for in the course of virus growth. However, a few td viruses with a slightly larger genome than class b RNA have been described (10, 11). These appear to have retained part of src (S. Kawai and H. Hanafusa, personal communication).

The purpose of this study was to examine genomes of independently generated *td* viruses to see whether all of them have the same size deletion and a uniform structure. In accord with previous observations by Graf and coworkers (10) and by S. Kawai and H. Hanafusa (personal communication), we found that the Schmidt-Ruppin strain of Rous sarcoma virus (SR) contains, besides the common deletions of *src*, also smaller deletions with a genome that is larger than that of the more common *td* variants. We also show by heteroduplex mapping that both types of deletion mutants contain RNA molecules with short heterologous sequences of distinct length in the position of the *src* gene. These heterologous sequences may represent substitutions of cellular gene sequences.

## MATERIALS AND METHODS

Viruses. td viruses of Prague strain Rous sarcoma virus of subgroups A (PR-A), and C (PR-C), Carr-Zilber strain of subgroup D, SR strain of subgroup D (SR-D), the same virus passaged in mammalian cells, and avian sarcoma virus Bratislava 77 (B77) were used. td viruses were isolated by picking small areas of nontransformed fibroblasts with a capillary pipette from plates of about 100 transformed foci induced by the respective sarcoma virus. These nontransformed cells were then tested for the presence of td virus by standard interference techniques (12). No further cloning of the infected cells was performed. The culture media from these infected cells were harvested every 12 hr. Viruses were purified from pooled culture media and 35S RNA was extracted with phenol/sodium dodecyl sulfate according to published methods (6).

**Electrophoresis and Fingerprinting.** Polyacrylamide gel electrophoresis was performed as described (6). Oligonucleotide fingerprinting of <sup>32</sup>P-labeled RNA was carried out by twodimensional polyacrylamide gel electrophoresis according to published methods (13).

**Preparation of Genome-Length cDNA.** cDNA was prepared by a modification of the method developed by Junghans *et al.* (14). The virus was incubated at 41° for 18 hr in 8 ml of reaction mixture containing 1 mM each of dCTP, dGTP, and dATP, 0.2 mM TTP, 150  $\mu$ Ci of [<sup>3</sup>H]TTP (40–60 Ci/mmol) (New England Nuclear), 2 mM magnesium acetate, 30 mM dithiothreitol, 0.1 M Tris-HCl (pH 7.6), and 0.02–0.03% Triton X-100. The optimum concentration of Triton X-100 was determined beforehand for every virus preparation. The DNA product was treated with RNase A (50  $\mu$ g/ml) at 37° for 1 hr or 0.3 M KOH for 18 hr, and then analyzed by alkaline sucrose gradient. About 1  $\mu$ g of genome-length cDNA could be obtained from 4 mg of virus.

Heteroduplex Studies. Heteroduplex molecules were formed by hybridizing full-length cDNA with 35S RNA under the high formamide condition described by Casey and Davidson (15). Hybrids were spread with the urea–formamide method without glyoxal treatment. This spreading condition greatly facilitated the distinction between duplex hybrid regions and nonhy-

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: *td*, transformation defective; SR-A, SR-D, subgroups A and D of Schmidt–Ruppin strain of Rous sarcoma virus; PR-A, PR-B, PR-C, subgroups A, B, and C of Prague strain of Rous sarcoma virus; B77, avian sarcoma virus Bratislava 77; SV40, simian virus 40; kb, kilobase.



FIG. 1. Polyacrylamide gel electrophoresis of 35S RNA of td viruses.  $[^{32}P]$ - ( $\bullet$ ) or  $[^{3}H]$ uridine-labeled ( $\triangle$ ) 70S RNA of different viruses was heated at 100° for 45 sec and then electrophoresed on 2% polyacrylamide gels. The electrophoresis was run at 5 mA per gel for 6 hr (6). (A)  $\bullet$ , td B77;  $\triangle$ , td SR-D isolate 3. (B)  $\bullet$ , td B77;  $\triangle$ , td SR-D isolate 14. (C)  $\bullet$ , td SR-D isolate 21;  $\triangle$ , td SR-D isolate 25.

bridized single-stranded regions. In addition, there was no random denaturation as was observed after treatment with glyoxal (16). These conditions will be described in detail elsewhere (Y. Chien and N. Davidson). Poly(A) mapping was done by a modification of the previously described procedure (17). The modification consisted of adding tails of poly(BrdUrd) rather than of poly(dT) to the simian virus 40 (SV40) circular duplex DNA. Poly(BrdUrd) forms a more stable hybrid with poly(A) than does poly(dT), thus simplifying the spreading procedure (W. Bender and N. Davidson, personal communication).

## RESULTS

The Genome Size of Several Independently Isolated tdDeletion Mutants. In order to determine whether all td viruses have an *src* deletion of uniform size, we examined the 35S RNA of 20 separate isolates by polyacrylamide gel electrophoresis. Included were two isolates of td PR-A, one isolate of td PR-C, four isolates of td B77, four isolates derived from SR-D passaged in mammalian cells, four isolates of td Carr-Zilber strain of subgroup D, and five isolates of SR-D. All these viruses except those obtained from SR-D had class b RNA of the same size as that of td B77 (data not shown). Among the five isolates from td SR-D, we found two that had a 35S RNA slightly larger than that of td B77 (td SR-D isolates 3 and 14, Fig. 1 A and B)

whereas the other three (isolates 21, 22, and 25) had class b RNA of uniform size (Fig. 1C). Coelectrophoresis of the 35S RNAs from both types of td SR-D confirmed that td isolate 3 had an RNA genome larger than td isolate 21 (Fig. 1D). However, there was no size difference between isolates 3 and 14, both of which had larger RNA than that of td B77 (Fig. 1E); nor was there a detectable size difference between isolates 21 and 25, both of which had small RNA (Fig. 1F). The difference between the RNA of the two types of deletion mutants was estimated from electrophoretic mobility to be about 25-30% of the difference between class a and class b RNA-i.e., about 4% of the total genome (Fig. 2). Thus, the td SR-D viruses we examined can be classified into two groups: one with a more extensive deletion of the src gene, and the other with a smaller deletion, retaining possibly about 25% of src. However, it could not be ruled out that the latter group of viruses might have the same size deletion as the former but with an insertion of some other sequences.

Oligonucleotide Fingerprint Analysis. Nondefective sarcoma viruses have been shown to contain sarcoma-specific, RNase  $T_1$ -resistant oligonucleotides that are missing from the corresponding td derivatives (6, 18). Fig. 3 A and B shows oligonucleotide fingerprints of 70S RNA from SR-D (which also contained about 60% td segregants) and of td SR-D isolate 21, respectively. The sarcoma virus contained at least two oligo-



FIG. 2. Polyacrylamide gel electrophoresis of 35S RNA of  $[^{3}H]$ -uridine-labeled td SR-D isolate 3 ( $\Delta$ ) and  $^{32}P$ -labeled B77 and td B77 ( $\bullet$ ). a and b refer to RNA classes a and b, respectively.

nucleotides that were absent from td SR-D (arrows). If isolates 3 and 14 contain a partially deleted *src* gene, the oligonucleotide fingerprints of these two viruses may retain some of these sarcoma-specific oligonucleotides. However, as shown in Fig. 3C, the fingerprint of td SR-D isolate 3 RNA could not be distinguished from that of isolate 21. The same finding was made with isolate 14 (data not shown). We conclude that both sarcoma-specific oligonucleotides are deleted in both types of tdviruses.

The fingerprints of 70S RNA of SR-D and of all *td* SR-D viruses contained one minor oligonucleotide present at very low concentrations (the circled spots in Fig. 3). Although its concentration varied from virus preparation to preparation, it was always present in amounts that were substantially lower than those of the remaining "regular" nucleotides that occurred in equimolar quantities. This spot was also seen in the fingerprints of 35S RNA of these viruses (data not shown). However, it could not be determined whether it came from viral gene sequence or some cellular sequences.

Heteroduplex Studies. Because the oligonucleotide finger-



FIG. 4. Circular heteroduplex molecules between cDNA from SR-D and 35S RNA from td SR-D isolate 3. (A) Heteroduplex with a deletion loop (b) (1.5 kb). (B) Heteroduplex with a substitution loop (c) (0.6 kb). The 3' end of the 35S RNA is attached to a circular SV40 DNA.

prints of all td SR-D viruses were identical, we could not distinguish between the following two possibilities: either the src gene is deleted to a smaller extent in td SR-D isolates 3 and 14 or these viruses have an src deletion that is similar or identical in size to that in isolates 21 and 25 but have acquired some other sequences instead. Therefore, we studied the heteroduplex molecules formed between genome-length cDNA made from nondefective SR-D and the 35S RNA of the td SR-D viruses. About 25% of the heteroduplex molecules formed between SR-D DNA and the td RNA were circular as shown in Fig. 4 A and B. This observation suggests that the 35S RNA of the deletion mutants in the heteroduplex molecules were of full genome length (16). The homologous and nonhomologous regions were exactly the same in circular and linear heteroduplex molecules. Both the circular and linear heteroduplex molecules of td SR-D isolate 3 and td SR-D isolate 21 formed with the cDNA of nondefective SR-D had only one deletion loop along the entire genome (Figs. 4 and 5). The length of this loop was 2.0 kilobases (kb) and 1.5 kb for isolate 21 and isolate 3, respectively (Table 1). The location of the deletion loop was



FIG. 3. Oligonucleotide fingerprinting of SR-D and deletion mutants. The <sup>32</sup>P-labeled 70S RNAs of SR-D (A), td SR-D isolate 21 (B), and td SR-D isolate 3 (C) were digested with RNase  $T_1$  and then "fingerprinted" by two-dimensional polyacrylamide gel electrophoresis. First-dimension separation was performed in 6 M urea at pH 3.3; the second dimension was in Tris-citrate buffer at pH 8.0 (13). Spots marked by arrows in A are the oligonucleotides present only in SR-D. The circled spots in all three fingerprints are the oligonucleotides present at very low concentration (see *text*).



FIG. 5. Linear heteroduplex molecules formed between cDNA from SR-D and 35S RNA from various deletion mutants. (A) Heteroduplex between SR-D and td SR-D isolate 21, showing a deletion loop (b) close to the 3' end, which is attached to a circular SV40 DNA molecule. (B) Heteroduplex between SR-D and td SR-D isolate 3. This heteroduplex is very similar to that in A. However, the deletion loop (b) is smaller (see Table 1). (C) Heteroduplex between SR-D and td SR-D isolate 21, showing a substitution loop (c) (at the same site. (D) Heteroduplex between SR-D and td SR-D isolate 21 showing a substitution loop (c) (0.6 kb) (see Table 1). a represents the distance from circular SV40 DNA molecule (3' end) to the deletion loop, b represents the length of the deletion loop, and c represents the length of the substitution loop. The actual measurements for various clones are given in Table 1.

0.6-0.8 kb from the 3' end of the RNA molecule. This location was consistent with the site of the *src* gene estimated from oligonucleotide and heteroduplex mapping (16, 18). Because the deletion loop of *td* SR-D isolate 3 was smaller, less of the *src* gene must have been deleted in this virus than in *td* SR-D isolate 21. The size difference between the RNA genomes of isolates 3 and 21 is therefore not due to an insertion of some other sequences. The heteroduplex molecules of isolates 14 and 22 were similar to those of isolates 3 and 21, respectively. The *td* SR-D isolate 25 was not examined by heteroduplex studies. These data are in agreement with the results of polyacrylamide gel electrophoresis.

In both types of td SR-D viruses, one other kind of heteroduplex was observed in which, besides the *src* deletion loop of 2.0 or 1.5 kb, a substitution loop of 0.6 and 0.7 kb occurred at the site of the *src* gene (Fig. 5C). This substitution loop represented a completely heterologous gene sequence that could not form a duplex with the deletion loop representing *src* gene sequence for the following reasons: (i) The hybridization and spreading conditions used in this study did not cause any detectable partial denaturation of perfectly matched bases. (ii) Both the size and location of the substitution loops were re-

Table 1. Site and length of deletion loops and substitution loops

<i>td</i> virus isolate	a*†	b*	c*
3	$0.74 \pm 0.10$ (36)	$1.52 \pm 0.20$ (34)	$0.55 \pm 0.14$ (13)
14	$0.86 \pm 0.18$ (31)	$1.47 \pm 0.22$ (28)	$0.60 \pm 0.03$ (3)
21	$0.75 \pm 0.17$ (20)	$1.98 \pm 0.23$ (16)	$0.73 \pm 0.08$ (6)
22	$0.69 \pm 0.16$ (35)	$1.87 \pm 0.24$ (27)	$0.70 \pm 0.14$ (13)

The numbers in parentheses are the numbers of heteroduplex molecules observed and measured.

\* a represents the distance between 3' end [including poly(A)] and deletion loop; b represents the length of deletion loop; and c represents the length of substitution loop as described in Figs. 4 and 5.

<sup>†</sup> Circular SV40 DNA (1.76  $\mu$ m) was the internal standard used for length measurement. For single-stranded DNA or RNA, 1  $\mu$ m = 3.906 kb; for RNA·DNA hybrids, 1  $\mu$ m = 3.153 kb. producible. (iii) The two single-stranded arms of the substitution and deletion loops were of different lengths. The substitution loop was observed in all preparations of td SR-D RNA. The fraction of RNA molecules with this substitution loop varied from 5 to 30% in different RNA preparations. However, this number could only be taken as a rough estimate because only a fraction of the heteroduplex molecules was interpretable in electron micrographs. The observation of this substitution loop suggests that the src gene sequence was replaced by heterologous sequences in some of the td genomes. A third kind of heteroduplex molecule with a smaller deletion or insertion loop of 0.6-0.7 kb at the site of the src gene was also observed in all five isolates of td viruses (Figs. 4B and 5D). These molecules probably represented the heteroduplexes formed between td RNA with inserted heterologous sequences and cDNA made from td segregants present in the stocks of SR-D. This conclusion is supported by the following observations: (i) In length and location, the loop was similar to the substitution loop in the heteroduplex molecules that show both deletion and substitution loops. (ii) The entire length of these heteroduplex molecules was equal to the total length of td RNA.

## DISCUSSION

Two conclusions can be drawn from this study: (i) Two of five td isolates from SR-D have an src gene deleted to a smaller extent than most of the *td* viruses described in the literature. (ii) All of these td SR-D viruses contain a subpopulation of viruses that have incorporated heterologous gene sequences at the site of src in their RNA genome. Both polyacrylamide gel electrophoresis and heteroduplex analysis showed that td SR-D isolates 3 and 14 have retained about 25% of the src gene sequences. These residual src sequences remain to be characterized; they are not represented in the two src-specific oligonucleotides identified in fingerprints. The deletion loop of all td SR-D viruses starts at 0.6-0.8 kb from the 3' end. The variation of 0.1-0.2 kb among different td clones could be accounted for by the slippage in the hybrids between the poly(A)tract of RNA molecules and the BrdUrd tails of marker SV40 DNA molecules (16, 17). Thus, the deletions of src in both types of td SR-D viruses appear to start at the same distance from the 3' end of the genome, leading to the tentative conclusion that the residual *src* sequences in the *td* mutants with smaller deletions are derived from the 5' end of the *src* gene, possibly those sequences adjacent to the *env* gene (19, 20). *td* mutants with residual sequences will, of course, be useful in mapping point mutations in the *src* gene.

A substitution loop of 0.6-0.7 kb at the site of the src deletion occurs in some RNA molecules of all td SR-D viruses. The percentage of the RNA molecules with this substitution is apparently low so that no heterogeneity of RNA size could be detected in the polyacrylamide gel electrophoresis. Whether the heterologous sequences present in this loop correspond to the minor oligonucleotide of the fingerprints is not known. Preliminary characterization showed that the minor oligonucleotide has the same base sequences in all five td viruses studied. The lengths and locations of the heterologous sequences are also the same in all cases. However, it is not clear whether these heterologous regions in different td isolates are identical. The origin of the heterologous sequences is also unknown. They could have been acquired from the cellular genome, or they could represent redundant viral sequences. The latter possibility is less likely, because we did not detect any additional cDNA hybridized to the substitution loops in any heteroduplex molecules. However, the presence of these kinds of redundant sequences could not be completely ruled out by our techniques.

There are several possible mechanisms for the acquisition of the heterologous sequences in td genomes. They could have been present in the genome of the wild-type SR-D, or they could have been derived during or after deletion of the src gene. Because the SR-D we used had not been rigorously cloned recently, it is possible that some of the SR-D genomes might have incorporated heterologous sequences. A deletion in the src gene of such RNA would produce td viruses with heterologous sequences. This possibility has not been ruled out. Alternatively, the acquisition of heterologous sequences might be related to the generation of *td* viruses in one of the following three ways. First, the deletion of src could occur during reverse transcription and before integration of proviral DNA, and the heterologous sequences could be acquired by some of the td viral genomes after loss of src. This mechanism would require that the insertion of heterologous sequence occur at a high frequency, because it was found in all preparations of td SR-D virus. In the second mechanism, insertion of heterologous sequences could be coupled to the deletion of src. This could happen if we assume that the deletion of the src gene takes place when proviral DNA is integrated in the chromosome. The deletion of src and insertion of cellular sequence could result from a difference in the integration site. The third mechanism would involve first the insertion of heterologous sequences into a nondefective sarcoma viral genome by "illegitimate recombination" (21). The insertion of this heterologous DNA could then induce the deletion of the src gene. It has been previously reported that the presence of an inserted sequence (IS-element) in the Escherichia coli chromosome can induce the formation of deletions around the IS-element (22, 23). If this is the mechanism of src deletion, it would be expected that some of the RNA genomes of avian sarcoma viruses might contain inserted cellular sequences. Indeed, it has been reported that nondefective avian sarcoma viruses produced by clonal colonies of transformed cells occasionally contain an RNA genome that is larger than 35 S. The size of this RNA is genetically unstable; it may reflect cellular sequences incorporated in the viral genome and lost upon passage (5). Whether our SR-D stock also has incorporated heterologous sequences requires further study.

Partial deletions of the *src* gene have so far been found only in SR (10, 11) (S. Kawai and H. Hanafusa, personal communication). Despite extensive screening in this and other laboratories, no similar viruses have been discovered in PR, in B77, or in the Carr-Zilber strain of Rous sarcoma virus (P. H. Duesberg and P. K. Vogt, unpublished data). Whether SR uses a unique mechanism for the deletion of the *src* remains to be investigated.

We thank Dr. R. Junghans for his advice on the synthesis of cDNA, Dr. Y. H. Chien for sharing her unpublished procedures for heteroduplex studies, and Dr. W. Bender for the generous gift of SV40 DNA with BrdUrd tails. We also acknowledge the excellent technical assistance of Mrs. Cynthia Yung and Mrs. Jean Wang. This work was supported by U.S. Public Health Service Research Grant CA 16113 and American Cancer Society Grant VC-176 (to M.M.C.L.) and by U.S. Public Health Service Grant CA 13213 and Cancer Research Emphasis Grant CA 19725 to P.K.V.; S.H. is a postdoctoral fellow of the Leukemia Society of America.

- 1. Vogt, P. K. (1971) Virology 46, 939-946.
- 2. Kawai, S. & Hanafusa, H. (1972) Virology 49, 37-44.
- 3. Martin, G. S. & Duesberg, P. H. (1972) Virology 47, 494-497.
- Duesberg, P. H. & Vogt, P. K. (1970) Proc. Natl. Acad. Sci. USA 67, 1673–1680.
- 5. Duesberg, P. H. & Vogt, P. K. (1973) Virology 54, 207-219.
- Lai, M. M. C., Duesberg, P. H., Horst, J. & Vogt, P. K. (1973) Proc. Natl. Acad. Sci. USA 70, 2266–2270.
- 7. Duesberg, P. H. & Vogt, P. K. (1973) J. Virol. 12, 594-599.
- Neiman, P. E., Wright, S. E., McMillin, C. & MacDonnel, D. (1974) J. Virol. 13, 837-846.
- 9. Bernstein, A., MacCormick, R. & Martin, G. S. (1976) Virology 70, 206–209.
- Graf, T., Bauer, H., Gelderblom, H. & Bolognesi, D. P. (1971) Virology 43, 427-441.
- 11. Stone, M. P., Smith, R. E. & Joklik, W. K. (1974) Cold Spring Harbor Symp. Quant. Biol. 39, 859-868.
- 12. Vogt, P. K. & Ishizaki, R. (1965) Virology 26, 664-672.
- Lee, Y. G. & Wimmer, E. (1976) Nucleic Acid Res. 3, 1647– 1658.
- Junghans, R. P., Duesberg, P. H. & Knight, C. A. (1975) Proc. Natl. Acad. Sci. USA 72, 4895–4899.
- Casey, J. & Davidson, N. (1977) Nucleic Acid Res. 4, 1539– 1552.
- Junghans, R. P., Hu, S., Knight, C. A. & Davidson, N. (1977) Proc. Natl. Acad. Sci. USA 74, 477–481.
- 17. Bender, W. & Davidson, N. (1976) Cell 7, 595-607.
- Wang, L.-H., Duesberg, P., Beemon, K. & Vogt, P. K. (1975) J. Virol. 16, 1051–1070.
- Wang, L.-H., Duesberg, P. H., Kawai, S. & Hanafusa, H. (1976) Proc. Natl. Acad. Sci. USA 73, 447-451.
- Wang, L.-H., Duesberg, P. H., Mellon, P. & Vogt, P. K. (1976) Proc. Natl. Acad. Sci. USA 73, 1073–1077.
- Franklin, N. C. (1971) in *The Bacteriophage Lambda*, ed. Hershey, A. D. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 175–194.
- 22. Reif, H. J. & Saedler, H. (1975) Mol. Gen. Genet. 137, 17-28.
- Starlinger, P. & Saedler, H. (1976) Curr. Top. Microbiol. Immunol. 75, 111–152.