Restriction Endonuclease and Nucleotide Sequence Analyses of Molecularly Cloned Unintegrated Avian Tumor Virus DNA: Structure of Large Terminal Repeats in Circle Junctions

RICHARD A. KATZ,¹ CHARLES A. OMER,^{2†} JOHN H. WEIS,^{2‡} S. ALEX MITSIALIS,¹ ANTHONY J. FARAS,^{2*} AND RAMAREDDY V. GUNTAKA^{1*}

Department of Microbiology, College of Physicians and Surgeons, Columbia University, New York, New York 10032,¹ and Department of Microbiology, University of Minnesota School of Medicine, Minneapolis, Minnesota 55455²

Received 29 June 1981/Accepted 9 December 1981

Avian tumor virus supercoiled DNA was isolated from infected quail tumor cells and molecularly cloned in pBR322. Four different recombinant clones denoted pATV-6, pATV-7, pATV-8, and pATV-9 were characterized in detail by restriction endonuclease mapping and by DNA sequencing. The results of these studies indicate that (i) the two large terminal repeats (LTRs) present in pATV-6 are different sizes, (ii) pATV-8 and pATV-9 contain only one LTR, (iii) pATV-7 contains an inversion of 0.6 kilobase in the *env* gene and a deletion of the U₃ region and the *src* gene, and (iv) the *src* gene is deleted in pATV-6 and pATV-9. Circle formation from linear molecules was also examined in several of the clones by DNA sequencing through the circle joint. pATV-6 is an example of one class of circular molecules and contains a partially repeated LTR similar to that reported by Ju and Skalka (Cell 22:379-386, 1980). A second class of circles was exemplified by pATV-8 and pATV-9, which contain a single copy of the LTR with no base changes or deletions. This is in contrast to a class of circles containing a complete double LTR structure described by Swanstrom et al. (Proc. Natl. Acad. Sci. U.S.A. 78:124-128, 1981) and suggests that circles containing a single intact LTR may be formed by a homologous recombinational event in which an entire LTR or complementary regions from both LTRs are removed from the linear DNA molecule during circularization.

Replication of RNA tumor viruses occurs through a DNA intermediate which is covalently integrated into the host genome (1, 2). Prior to integration, the DNA exists in at least two different configurations, linear and circular (8, 23). The linear viral DNA molecules contain large terminal repeats (LTR) of about 300 base pairs (bp) (10, 19) representing unique sequences derived from the 5' end (U₃) and the 3' end (U₃) of the RNA genome interspersed by a short terminal repeat of 21 bp (R) of the RNA genome (2). Some molecules of circular viral DNA bear only a single copy of the LTR, whereas others contain two LTRs arranged in tandem (10, 19). Integrated proviruses contain the LTR sequences at their termini similar to linear forms of unintegrated viral DNA, suggesting that recombinational events leading to integration occur at specific sites in the LTR sequence (5, 11, 15, 18).

Linear viral DNA present in the cytoplasm of avian sarcoma virus (ASV)-infected cells has

been demonstrated to be the precursor of circular viral DNA located in the nucleus (20). Based upon this precursor-product relationship, it is reasonable to presume that the termini of linear viral DNA are joined by direct end-to-end ligation or by a recombinational event in forming circular viral DNA. To distinguish between these possibilities and delineate the mechanism by which linear viral DNA molecules are circularized, we and others (4, 9, 12, 21) molecularly cloned several circular viral DNA molecules and analyzed their circle junctions by restriction endonuclease mapping and DNA sequencing.

Molecular cloning of circular viral DNA (7) of the Prague C (PrC) strain of ASV was accomplished by partially digesting purified viral DNA with the restriction endonuclease *Hind*III and by ligating the cleaved DNA into the *Hind*III site of pBR322. This DNA was then used to transform *Escherichia coli* HB101, and transformants with viral DNA inserts were detected by colony hybridization with the use of ³²P-labeled viral cDNA (6). Since *Hind*III cleaves viral DNA at three sites and yields DNA fragments of 6.3, 3.2, and 0.1 kilobases (kb) in length (19), different

† Present address: Department of Genetics, Stanford University, Stanford, CA 94305.

‡ Present address: Molecular Genetics, Inc., Minnetonka, MN 55343.

regions of the viral genome could be present in different clones. Six recombinant clones with inserts were identified. Two of these clones contained the 3.2-kb viral DNA fragment, one (pATV-6) contained a 4.3-kb circle junction fragment derived from a transformation-defective genome, and three others (pATV7-9) contained most of the PrC genome, including the circle junction. Since we are mainly interested in characterizing the circle junction fragments, we describe here only those four recombinants (pATV-6, pATV-7, pATV-8, and pATV-9) that contain circle junctions.

To characterize the inserted viral DNA fragments, the DNAs were singly and doubly digested with several different restriction endonucleases and compared with known fragments derived from unintegrated viral DNA (19; data not shown). The restriction endonuclease maps

of pATV-6, pATV-7, pATV-8, and pATV-9 thus derived are presented in Fig. 1 and 2. pATV-6 appears to exhibit a deletion in the *src* gene and contains two copies of the LTR. One LTR in pATV-6, however, appears to be incomplete, since only a 190-bp *Eco*RI junction fragment appears instead of the expected 340-bp *Eco*RI fragment. pATV-8 contains the entire PrC genome and has only a single LTR at the circle junction. pATV-9 also contains a single LTR at the circle junction but, in addition, contains a deletion in the *src* region. pATV-7 appears to have an inversion of 0.6 kb of the *env* gene. From a detailed restriction analysis of this clone (Fig. 3), it appears that this inversion involves *Eco*RI, *Hind*III, *Xba*I, *Sac*I, and *Xho*I restriction endonuclease sites. pATV-7 is also missing the *src* gene and the U₃ region of the LTR. The latter conclusion is based upon the observation

FIG. 1. Restriction enzyme maps of molecularly cloned viral DNA inserts. The maps were constructed from analysis of fragments derived from single and multiple restriction enzyme digests of pATV-6, pATV-7, pATV-8, and pATV-9. The maps are aligned so that the gene order corresponds to that of viral RNA. A map of the genes of ASV is included for reference: *gag*, group-specific antigen; *pol*, reverse transcriptase; *env*, envelope glycoproteins; and *src*, transformation gene. The plasmid vector pBR322 is depicted as a loop and the relative positions of the *Eco*RI and *Bam*HI restriction endonuclease sites in the vector are shown to indicate the orientation of the viral DNA insert. Deletions are denoted by triangles. The boundaries of the inversion in pATV-7 are marked by arrows. The left and right ends of the restriction maps are linked in the uncleaved plasmids.

To determine the mechanism by which ASV linear DNA is circularized, the nucleotide sequence of the junction regions for pATV-6, pATV-8, and pATV-9 was determined (Fig. 4 and 5). Previous DNA sequence analysis of

molecularly cloned avian retrovirus circular DNAs has examined clones with either partially or completely repeated LTRs (13, 22) which constitute one class of circular viral DNAs found in infected cells (10, 19). pATV-6 is an

FIG. 4. DNA sequence of the circle junction regions of pATV-6, pATV-8, and pATV-9. The DNA sequences of the circle junction region were determined for pATV-6, pATV-8, and pATV-9 as shown in Fig. 5. Only the viral RNA polarity strand of DNA is shown. R refers to the sequence corresponding to the viral genomic RNA terminal redundancy. IR₃ and IR₅ refer to the 12- or 15-bp inverted repeats found at the ends of the LTR with the subscript 3 or 5 denoting whether the sequence is derived from copying 3' end or 5' end RNA sequences. U₃ and U₅ refer to 3' end or 5' end of the viral RNA-derived DNA sequences in the LTR which are unique to their respective ends. (-)PB refers to the tRNA^{ppv} primer binding site for (-)DNA synthesis.

FIG. 5. Strategy for sequencing pATV-6, pATV-8, and pATV-9. DNA sequencing was done by using the technique of Maxam and Gilbert (16). The 3' end-labeled fragments (14) (○) or 5' end-labeled fragments (16) (●) were used to determine the DNA sequence.

example of this type and contains a partially repeated LTR. From its nucleotide sequence, it appears very similar to the λ SRB-td7 clone of Ju and Skalka (13). pATV-8 and pATV-9 contain only one LTR and are examples of the second type of circular DNA found in ASV-infected cells. The LTR region of pATV-9 was sequenced and compared with that of pATV-6 in an attempt to understand the mechanism by which circles containing a single LTR are formed. In comparison to the complete LTR of pATV-6, pATV-9 contains a single complete copy of the LTR with no base changes or deletions. The regions on either side of the LTR, including the tRNA^{trp} primer binding site and the putative plus-strand DNA initiation site, are also intact in pATV-9. DNA sequence analysis of pATV-8 was performed for only one strand of the DNA and serves to confirm the results found for pATV-9, since an identical nucleotide sequence was found.

A comparison of the LTR sequence from the PrC strain of ASV to that previously determined for the LTRs of the Schmidt-Ruppin strain of ASV (22) and Rous-associated virus type 2 (RAV-2) (13) indicates that the LTR of PrC is 335 bp in length compared with 330 and 348 bp for Schmidt-Ruppin and RAV-2, respectively. These length differences are due to insertions and deletions entirely within the 160 nucleotides of the U₃ region immediately adjacent to the IR₃ sequence. Nucleotide comparisons of the LTRs of PrC, Schmidt-Ruppin ASV, and RAV-2 indicate a difference of 16 bp between the LTRs of

PrC and Schmidt-Ruppin ASV and a 21-bp difference between the LTRs of PrC and RAV-2. These changes are randomly distributed throughout the LTR, with the exception that no changes could be detected in the inverted repeats or in the 40 bp of the U₃ region immediately adjacent to R. The lack of changes in the U₃ region adjacent to R could reflect the requirement to retain the polyadenylate addition signal sequence (17) (AATAAA) and putative promoter sequence (24) located in this region.

The aforementioned data indicate that deletions in the molecularly cloned viral DNA exist in three of the four ATV clones examined and that one clone also contains an inversion. This inversion includes regions of DNA encompassing the right end of the *env* gene and the sequences to the left of the U₃ region. In the same clone, almost all of the *src* region is also deleted. The significance of these results is not clear. Whether deletion and inversion in pATV-7 represent the result of inaccurate reverse transcription or intramolecular recombination remains to be determined. Deletions in the *env* and *src* genes of avian retroviruses and recombinational events between ecotropic and xenotropic murine retroviruses are known to exist. Whether specific nucleotide sequences at the boundaries of, or within, the *env* gene contribute to the kinds of genetic exchange that are observed here remains to be explored. The possibility that these deletions and inversions occur in the bacterial host is also not ruled out. Additional recombinants will have to be isolated and characterized before the significance of these observations can be appreciated.

Restriction endonuclease and nucleotide sequence analysis of pATV-6 which contains a complete and partial LTR in tandem suggested that the recombination event leading to the generation of circular molecules must have occurred between the *EcoRI* and *PvuI* sites. Since the repeat is 0.19 kb long, circularization must have taken place at about 27 bp to the right of the *PvuI* site or 32 bp to the left of the *EcoRI* site. Thus, this clone is very similar to λ SRB-td7 described by Ju and Skalka (13). The significance of genomes with partial LTRs is not currently understood. Nucleotide sequence analysis of clones containing a single LTR indicated that sequence alterations were absent either within the LTR or regions surrounding it. Single LTR formation thus appears to occur by a mechanism which precisely removes one LTR, or complementary regions of both LTRs, from the linear provirus during circularization. This appears to be very reminiscent of a single homologous recombinational event. This is in contrast to the mechanism proposed for complete double LTR formation in which an end-to-end

ligation of the linear provirus appears to occur (13, 22) and suggests that these circular molecules are generated by distinct mechanisms.

This work was supported by Public Health Service grants CP 71055 and CA 28990 (to R.V.G.) and CA 20011 and CA 18303 (to A.J.F.). R.V.G. is a Research Career Development awardee from the National Cancer Institute. C.A.O. and J.H.W. were supported by predoctoral and postdoctoral traineeships, respectively, from Public Health Service training grant CA 09138.

We thank C. Kosman for typing this manuscript.

LITERATURE CITED

1. Bishop, J. M. 1978. Retroviruses. *Annu. Rev. Biochem.* 47:35-88.
2. Coffin, J. M. 1979. Structure, replication and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* 42:1-26.
3. Copeland, N. G., N. A. Jenkins, and G. M. Cooper. 1981. Integration of Rous sarcoma virus DNA during transfection. *Cell* 23:51-60.
4. DeLorbe, W. J., P. A. Luciw, H. M. Goodman, H. E. Varmus, and J. M. Bishop. 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. *J. Virol.* 36:50-61.
5. Dhar, R., W. L. McClements, L. W. Enquist, and G. F. Vande Woude. 1980. Nucleotide sequence of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. *Proc. Natl. Acad. Sci. U.S.A.* 77:3937-3941.
6. Grunstein, M., and D. S. Hogness. 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. U.S.A.* 72:3961-3965.
7. Guntaka, R. V. 1978. A simple method for rapid purification of avian sarcoma virus supercoiled DNA by selective precipitation of infected cell chromatin. *Anal. Biochem.* 90:256-261.
8. Guntaka, R. V., O. C. Richards, P. R. Shank, H. J. Kung, N. Davidson, E. Fritsch, J. M. Bishop, and H. E. Varmus. 1976. Covalently closed circular DNA of avian sarcoma virus: purification from nuclei of infected quail tumor cells and measurement by electron microscopy and gel electrophoresis. *J. Mol. Biol.* 106:337-357.
9. Highfield, P. E., L. F. Raffield, T. M. Gilmer, and J. T. Parsons. 1980. Molecular cloning of avian sarcoma virus closed circular DNA: structure and biological characterization of three recombinant clones. *J. Virol.* 36:271-279.
10. Hsu, W., J. L. Sabran, G. E. Mark, R. V. Guntaka, and J. M. Taylor. 1978. Analysis of unintegrated avian RNA tumor virus double-stranded DNA intermediates. *J. Virol.* 28:810-818.
11. Hughes, S. H., P. R. Shank, D. H. Spector, H. J. Kung, J. M. Bishop, H. E. Varmus, P. K. Vogt, and M. L. Briltman. 1978. Proviruses of avian sarcoma virus are terminally redundant, co-extensive with unintegrated linear DNA and integrated at many sites. *Cell* 15:1397-1410.
12. Ju, G., L. Boone, and A. M. Skalka. 1980. Isolation and characterization of recombinant DNA clones of avian retroviruses: size heterogeneity and instability of the direct repeat. *J. Virol.* 33:1026-1033.
13. Ju, G., and A. M. Skalka. 1980. Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. *Cell* 22:379-386.
14. Lacy, E., and T. Maniatis. 1980. The nucleotide sequence of a rabbit β -globin pseudogene. *Cell* 21:545-553.
15. Majors, J. E., and H. E. Varmus. 1981. Nucleotide sequences at host-proviral junctions for mouse mammary tumor virus. *Nature (London)* 289:253-258.
16. Maxam, A. M., and W. Gilbert. 1980. Sequencing end labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
17. Proudfoot, N. J., and G. G. Brownlee. 1976. 3'-Non-coding region-sequences in eukaryotic messenger RNA. *Nature (London)* 263:211-214.
18. Sabran, J. L., T. W. Hsu, C. Yeater, A. Kaji, W. S. Mason, and J. M. Taylor. 1979. Analysis of integrated avian RNA tumor virus DNA in transformed chicken, duck, and quail fibroblasts. *J. Virol.* 29:170-178.
19. Shank, P. R., S. H. Hughes, H. J. Kung, J. E. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus. 1978. Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. *Cell* 15:1383-1395.
20. Shank, P. R., and H. E. Varmus. 1978. Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closed circular viral DNA in the nucleus. *J. Virol.* 25:104-114.
21. Shoemaker, C., S. Goff, E. Gilboa, M. Paskind, S. W. Miha, and D. Baltimore. 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proc. Natl. Acad. Sci. U.S.A.* 77:3932-3936.
22. Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus. 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. *Proc. Natl. Acad. Sci. U.S.A.* 78:124-128.
23. Weinberg, R. A. 1977. Structure of the intermediates leading to the integrated provirus. *Biochim. Biophys. Acta* 473:39-55.
24. Yamamoto, T., B. deCrombrughe, and I. Pastan. 1980. Identification of a functional promoter in the long terminal repeat of Rous sarcoma virus. *Cell* 22:787-797.