5'-Terminal Nucleotide Noncoding Sequences of Retroviruses: Relatedness of Two Old World Primate Type C Viruses and Avian Spleen Necrosis Virus

GERALD G. LOVINGER,¹* GEORGE MARK,² GEORGE J. TODARO,² AND
GERALD SCHOCHETMAN¹

Biological Carcinogenesis Program, Frederick Cancer Research Center,¹ and Laboratory of Viral Carcinogenesis, National Cancer Institute, National Institutes of Health,² Frederick, Maryland 21701

Received 16 March 1981/Accepted 13 April 1981

Computer-assisted comparison of the 5'-terminal regions of mammalian type C viruses serves as a useful model of evolutionary divergence of noncoding nucleic acid sequences. It has led to the concept that regions of conserved nucleic acid sequences, the slowly divergent sequences, contain signals of translational, transcriptional, or integrative significance. Interspersed among the conserved regions are rapidly divergent sequences in which base changes, insertions, and deletions are especially prevalent. In the present study, CPC-1, a type C virus isolated from *Colobus polykomos*, was shown to be related to another Old World type C monkey virus, endogenous stump-tailed monkey virus, MAC-1, by analysis of their 5'-terminal nucleotide sequences. The 5'-terminal regions of CPC-1 and MAC-1 showed a 76% nucleotide correspondence and were of similar lengths, 132 and 127 nucleotides, respectively. Previous strong-stop analyses of other type C viruses have defined two subgroups: (i) Rauscher murine leukemia virus and gibbon ape leukemia virus and (ii) baboon endogenous virus and endogenous cat virus RD114. Based on the present sequence analysis of their 5'-terminal sequences, CPC-1 and MAC-1 formed a third subgroup. Computer-assisted comparison of the 5'-terminal sequences of CPC-1 and MAC-1 to the previously reported sequence of avian spleen necrosis virus (SNV) (Shimotohno et al., *Nature [London]* **285**:550-554, 1980) showed SNV to be a member of that subgroup of mammalian type C viruses. Consistent with the inclusion of SNV in this subgroup of mammalian type C viruses, SNV was distantly related to other mammalian type C viruses. Interestingly, the SNV 5'-terminal sequences showed no significant evolutionary relationship by these criteria to the avian leukemia and sarcoma viruses. CPC-1, MAC-1, and SNV contained conserved regulatory signals in similar positions in their 5'-terminal RNA sequences analogous to those observed in other mammalian type C retroviruses. These sequences included the canonical AAUAAA sequence, a palindrome, a putative ribosome binding site, and an integration site. Some of these highly conserved subsequences were common to 3'- and 5'-terminal noncoding sequences of nonviral eucaryotic mRNA's (Efstratiadis et al., *Cell* **21**:653-668, 1980). Thus, analysis and comparison of 5'-terminal nucleotide sequences have been useful in defining common functional signals and in extending the matrix of relationships among retroviruses.

Mammalian type C retroviruses have been isolated from such diverse host as rodents, apes, Mediterranean cats, and Old and New World monkeys (24). The 5'-terminal noncoding regions of a variety of these viruses have been sequenced, providing an evolutionary model for understanding how genomes diverge while preserving elements vital to their life cycles (13, 14). Comparison of 5'-terminal noncoding regions of retroviruses has shown a slowly divergent set of sequences with postulated replication, transcrip-

tion, translation, or integration functions (14). Interspersed among these sequences are sequences which have diverged more rapidly, that is, more subject to single-base changes, insertions, and deletions during the evolution of the viruses.

Based on the results of the above studies, two subgroups of mammalian type C viruses have been definitely established: (i) viruses of murine origin, which include Rauscher murine leukemia virus and gibbon ape leukemia virus, and (ii)

viruses of Old World monkey origin, which include baboon endogenous virus (BaEV) and endogenous cat virus RD114. A possible third subgroup consists of the endogenous stump-tailed monkey virus MAC-1 and endogenous rhesus monkey virus MMC-1 isolates from two Old World monkeys (20, 25), with identical 5'-terminal nucleotide sequences (14). The intragroup classifications are based on the statistically high number of correspondences in both the rapidly divergent and the slowly divergent sequences. The intergroup relationships derive from the sequence correspondences found only in the slowly divergent sequences.

An interesting outcome of these studies was that BaEV and MAC-1, both endogenous viruses of Old World monkeys, fell into separate subgroups. Recently, another endogenous virus of Old World monkeys, CPC-1, from *Colobus polykomos* (21), was isolated and was shown to be partially related to MAC-1 by sequence homology and antigenic cross-reactivity. In the present study, nucleic acid sequencing and computer-assisted analysis of the 5'-terminal region of CPC-1 showed that the rapidly and slowly divergent sequences of CPC-1 had a high degree of correspondence only with MAC-1 and MMC-1, thereby forming the third subgroup.

The reticuloendotheliosis viruses (exogenous avian retroviruses) have previously been shown to be related to type C mammalian retroviruses (2, 3, 7, 11). Recently, the long terminal repeat of avian spleen necrosis virus (SNV; a member of the reticuloendotheliosis viruses) has been sequenced (22), and the 5'-terminal sequence of SNV was deduced from this region. Computer-assisted analysis of the rapidly and slowly divergent sequences of SNV demonstrated a statistically significant degree of correspondence only with CPC-1, MAC-1, and MMC-1, thereby placing it in the third subgroup. In contrast, no evidence of relatedness of SNV to the avian leukemia and sarcoma viruses was observed. Based on these types of quantitative analyses, we have been able to elucidate interrelationships among the type C retroviruses.

MATERIALS AND METHODS

Isolation and growth of virus. A primary culture of normal kidney cells from *C. polykomos* was cocultivated with human carcinoma A549 cells as described previously (21). The cells were maintained in the Dulbecco modification of Eagle medium supplemented with 10% fetal calf serum. Supernatant fluids were harvested at 24-h intervals, and the virus was isolated from the fluids by centrifugation.

Reagents. [γ - 32 P]ATP (specific activity, 2,000 to 3,500 Ci/mmol) was obtained from Amersham Corp. (Arlington Heights, Ill.); unlabeled dATP, dGTP, and

dCTP were obtained from P-L Biochemicals, Inc. (Milwaukee, Wis.). Acrylamide and *N,N'*-methylenebisacrylamide were purchased from Schwarz/Mann (Orangeburg, N.Y.). Polynucleotide kinase purified from bacteriophage T₄-infected *Escherichia coli* cells was purchased from Boehringer Mannheim Biochemicals (Indianapolis, Ind.). Hydrazine was purchased from Eastman Organic Chemicals (Rochester, N.Y.). Dimethyl sulfate was purchased from Aldrich Chemical Co. (Milwaukee, Wis.).

Synthesis and sequence analysis of cDNA. cDNA of CPC-1 was synthesized in a reconstructed reaction, using 70S virion RNA and avian myeloblastosis virus reverse transcriptase. Approximately 10 μ g of purified 70S virion RNA was incubated at 43°C for 60 min in a 20- μ l reaction volume containing 40 mM Tris-hydrochloride (pH 8.1), 80 mM NaCl, 1 mM MnCl₂, 5 mM MgCl₂, 5 mM dithiothreitol, 1 mM each deoxynucleoside triphosphate and 20 U of avian myeloblastosis virus polymerase. After Sephadex G-50 column chromatography and dilute alkali treatment, the cDNA's were 5'-end-labeled with [γ - 32 P]ATP and polynucleotide kinase. The end-labeled cDNA transcripts were fractionated on 8% slab gels and eluted. One hundred picomoles of [32 P]cDNA was divided into four portions to be partially cleaved by four base-specific reactions (15). The end-labeled partial cleavage products were fractionated by electrophoresis on 8 and 16% thin slab gels (0.4 mm by 20 cm by 40 cm). After autoradiography, the sequences were interpreted.

Computer analysis of nucleotide sequences. The computer program ALIGN (8) was used in sequence comparisons of viral cDNA's (see Results) and in the statistical evaluations of the degree of their relatedness. Details of the scoring matrix have been described elsewhere (14). Intragroup comparisons of sequences used penalties of 15 for gaps, and intergroup comparisons of sequences used penalties of 5 for gaps. Further data on the sequences were elicited by using a comprehensive computer program (19).

RESULTS

Comparison of the 5'-terminal region of CPC-1 with those of other type C retroviruses. Retroviruses initiate the synthesis of a DNA copy on a tRNA primer located close to the 5' end of the genome. In vitro, the major DNA species is a product of less than 200 nucleotides terminating at the 5' end, designated cDNA^{5'}. To synthesize CPC-1 cDNA^{5'}, virions were disrupted with detergent in the presence of four deoxyribonucleotide precursors. The cDNA's were treated with dilute alkali followed by 5'-end labeling with [γ - 32 P]ATP. (For further details of the procedure, see Materials and Methods and reference 13.) Electrophoresis on 8% polyacrylamide slab gels resolved cDNA^{5'} from two shorter cDNA fragments. The cDNA's were then eluted from the gel and subjected to nucleotide sequence analysis (15).

The sequenced cDNA^{5'} of CPC-1 was con-

verted into its RNA genome complement, using the computer program ALIGN, and compared with previously sequenced 5'-terminal regions of type C viruses (Table 1). The statistical measure of relatedness was the *alignment score*, which was defined as the number of standard deviations a pair of sequences scored higher than the mean of many pairs of random alignments. Thus, the higher the positive alignment score, the closer the relationship between a pair of sequences. The alignment scores between CPC-1 and other type C viruses fell into three categories. (i) CPC-1 and MAC-1 were highly related, showing 76% sequence correspondence and an alignment score of 16.2 standard deviations. The probability that these viral sequences might be unrelated was less than 10^{-26} . (ii) CPC-1 and the remainder of the mammalian type C viruses examined (Rauscher murine leukemia virus, BaEV, and RD114) showed significant but lower relatedness, with sequence correspondences of 61 to 64% and alignment scores in the

range of 3.1 to 3.4 standard deviations. The probabilities of these pairs being unrelated ranged from 3×10^{-4} to 8×10^{-4} . (iii) CPC-1 and the avian leukosis and sarcoma viruses showed little or no relatedness, with alignment scores of less than 1.0 standard deviations.

An alignment of the sequences of the two endogenous Old World monkey viruses from category 1, CPC-1 and MAC-1, is shown in Fig. 1. To maximize sequence correspondence, only four single-base gaps had to be inserted by the computer program, and the matches were distributed throughout the entire 5' region. An example of a category 2 alignment, CPC-1 and another endogenous Old World monkey virus (BaEV), is shown in Fig. 2. To maximize the sequence correspondences in this alignment, 10 gaps were inserted, with the longest gap being five nucleotides. Regions where a high percentage of nucleotide matches clustered represented the slowly divergent sequences at the 5' terminus.

TABLE 1. Relationship of 5'-terminal sequence of CPC-1 to those of other type C retroviruses^a

Other type C virus	Sequence correspondence of other virus to CPC-1 (%)	Alignment score (standard deviations)	Probability of random score
MAC-1	76	16.2	$>10^{-26}$
BaEV	62	3.1	8×10^{-4}
RD114	61	3.1	8×10^{-4}
Rauscher murine leukemia virus	64	3.4	3×10^{-4}
Gibbon ape leukemia virus	63	3.3	4×10^{-4}
Avian sarcoma virus	52	0.4	0.4
Avian myeloblastosis virus	53	0.6	0.3

^a Percentages of correspondence and alignment scores were generated by the computer program ALIGN. The probabilities of random scores were calculated from tables in the *Atlas of Protein Sequence and Structure* (8).

	10	20	30	40	
CPC-1	G U C U C C C G C U U U A C U C C G C U C C U G G A G - G C G U G A G A G U U C C A G G U U C G A				
	* * * * *	* * * *	* * * * *	* * * * *	
MAC-1	- G C U C C U G C - - - - - C G U U U C A C G A G C A A G U G A G A G C U C C A G G U U C G A				
	50	60	70	80	90
CPC-1	A C C U G U A A U A A A G A U C C U U U G C A C U U - G C A U U G A C G C C U G A C U C U G G U G				
	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *
MAC-1	A C C U G U A A U A A A G A U C C U U G C U G C U U A G C U U U G A C U C U G G A C U C U G G U G				
	100	110	120	130	
CPC-1	A G C U U U U C U G G G G A C U A G A C G G A C U U G G C A C A A C A U				
	* * * *	* * * * *	* * * * *	* * * * *	
MAC-1	G U C U U C U U C G G G G A A U A A C G G U C U G G G C A U A A C A U				

FIG. 1. Nucleotide sequences of the 5' termini of the CPC-1 and MAC-1 genomes. The cDNA^{5'} of CPC-1 was sequenced in this report according to the method of Maxam and Gilbert (15); the cDNA^{5'} of MAC-1 had been sequenced previously (14). The cRNA's of these sequences were aligned, using a computer program (8, 14), to achieve the maximum number of nucleotide matches (indicated by asterisks) with the minimum number of gaps (indicated by dashes).

Analysis of slowly divergent putative regulatory sequences of CPC-1. The primary structure of the 5'-terminal region of CPC-1 (Fig. 3) showed no AUG triplet. AUG triplets have not been found in any of the mammalian type C viral regions between the 5' end and the tRNA primer, indicating that this region does not code for protein. Using a computer program (19), the 5'-terminal region could be arranged into several base-paired duplexes (Fig. 3). Two duplexes, residues 13 through 37 and 42 through 55, contained a large number of base-paired residues, and the following evidence indicates that they do exist in the genome RNA. Along with the synthesis of cDNA^{5'}, two shorter cDNA's were generated. These two cDNA's were sequenced and found to be subsets of cDNA^{5'}. The sequences ended at residues 38 and 56, which immediately preceded the two duplexes. It is known that reverse tran-

scriptase is sensitive to secondary structure (9), and, presumably, the enzyme has difficulty in opening the duplexes and proceeding to the 5' ends of the RNA genomes.

The clustered regions of nucleotide matches of CPC-1 contained putative regulatory signals in common with the other mammalian type C viruses. These signals in CPC-1 are shown in Fig. 3 and included the following: (i) an 55 AAUAAA polyadenylate addition signal sequence identical to those found in MAC-1, Rauscher murine leukemia virus, and gibbon ape leukemia virus but different from the BaEV sequence AAACGA (14) (the AAUAAA signal sequence is also found in many nonviral mRNA's [17]); (ii) a putative ribosome loading site of 92 UGGUGAG, similar to the MAC-1 sequence of

FIG. 2. Nucleotide sequences of the 5' termini of the CPC-1 and BaEV genomes. The cDNA^{5'} of CPC-1 was sequenced in this report, and the cDNA^{5'} of BaEV was sequenced previously (14). The best computer-assisted alignment of their RNAs is shown. See legend of Fig. 1 for further details.

FIG. 3. Proposed secondary structure and functional sequences for the 5' terminus of CPC-1. The 5'-terminal region of CPC-1 is shown, including the computer-proposed base-paired duplexes and cap structure. The shorter duplexes or duplexes with few G-C pairs have less probability of occurring in vivo. The numbering begins at the 5' terminus of the RNA.

UGGUGGU and the BaEV sequence of GGGUGGU (we have recently proposed that the long type C retrovirus leader contains a ribosome binding site in or around these GUG-containing sequences [14]); (iii) the palindrome

40
CAGGUUCGAACCUG, 14 nucleotides long, identical to the MAC-1 palindrome and related to the BaEV palindrome CUGGUGCACCAG; and (iv) the CPC-1 right-hand putative integra-

123
tion site GGCACAACAUU, containing the transposable universal element CA (23) and identical to the MAC-1 right-hand putative integration site except for position 127, where C is present instead of U. Thus, analysis of the slowly divergent sequences of CPC-1 confirm its close relationship to MAC-1 and its more distant relationship to BaEV and other type C mammalian viruses.

Comparison of the 5'-terminal region of SNV with other type C retroviruses. The long terminal repeat of SNV has recently been cloned and sequenced (22). The length of the cDNA^{5'} was calculated to be 180 nucleotides, somewhat longer than any other retroviral 5'-terminal sequence (Table 2). This sequence of SNV in its RNA genome form was aligned with each of the type C viruses previously sequenced, and each alignment was statistically evaluated by using the ALIGN program. As summarized in Table 2, SNV, a member of the avian reticuloendotheliosis viruses, showed negative alignment scores with two members of the avian leukosis-sarcoma virus group, denoting no significant relatedness. However, SNV showed distant relationships with BaEV, RD114, Rauscher murine leukemia virus, and gibbon ape leukemia

virus, with positive alignment scores in the range 2.4 to 3.4 standard deviations. Surprisingly, in contrast to its alignment with BaEV, SNV showed significantly higher scores with the other two endogenous Old World primate viruses, CPC-1 and MAC-1 (5.2 and 6.1, respectively). Thus, in SNV alignments, as with the CPC-1 alignments (Table 1), there were three categories of relationships.

A computer-assisted alignment between SNV and CPC-1 is shown in Fig. 4. The correspondence of SNV to CPC-1 (and also MAC-1) was particularly evident in two sequences (residues 41 through 62 and 163 through 179). These slowly divergent sequences in SNV contained the putative regulatory signals common to retroviruses and, as shown in Fig. 5, included the

57
following: (i) an AAUAAA canonical sequence which was identical to those of CPC-1 and MAC-1; (ii) a 14-nucleotide palindrome,

42
CAGAUUCGAAUCUG, which was identical to the palindrome in MAC-1 and CPC-1 except for A → G and U → C base changes, preserving the dyad symmetry; and (iii) a right-hand integration site identical to that of MAC-1 for 12 nucleotides. Interspersed between these conserved regions were regions where the majority of base differences and gaps occurred. SNV contained two large regions for which there was no correspondence in CPC-1 or MAC-1: (i) residues 64 through 76, a 13-nucleotide, UC-rich insert; and (ii) residues 125 through 148, a 24-nucleotide insert which was particularly GC rich. These two regions contributed to forming two stable hairpins.

TABLE 2. Relationship of 5'-terminal sequence of SNV to those of other type C retroviruses^a

Other virus	cDNA ^{5'} length (no. of nucleotides) ^b	Sequence correspondence of other virus to SNV (%)	Alignment score (standard deviations)	Probability of random score
MAC-1	127	68	6.1	4×10^{-5}
CPC-1	132	66	5.2	1×10^{-4}
BaEV	133	63	3.2	7×10^{-4}
RD114	127	60	2.4	10×10^{-4}
Rauscher murine leukemia virus	140	61	2.6	9×10^{-4}
Gibbon ape leukemia virus	144	64	3.4	4×10^{-4}
Avian sarcoma virus	101	51	-0.1	0.6
Avian myeloblastosis virus	101	50	-0.7	0.8

^a Percentages of correspondence and alignment scores were generated by the computer program ALIGN. The probabilities of random scores were calculated from tables in the *Atlas of Protein Sequence and Structure* (8).

^b SNV cDNA^{5'} is 180 nucleotides long.

FIG. 4. Nucleotide sequences of the 5' termini of the CPC-1 and SNV genomes. The cDNA⁸ of CPC-1 was sequenced in this report, and the cDNA⁸ of SNV was determined from the sequence of the published long terminal repeat (22). The best computer-assisted alignment of their RNAs is shown. Nucleotide matches are indicated by asterisks, and gaps are indicated by dashes. See legend of Fig. 1 for further details on alignment.

FIG. 5. Proposed secondary structure and functional sequences for the 5' terminus of SNV. The 5'-terminal region of SNV is shown, including the computer-proposed base-paired duplexes and cap structure. The shorter duplexes or duplexes with few G·C pairs have less probability of occurring *in vivo*. The numbering begins at the 5' terminus of the RNA.

DISCUSSION

The 5'-terminal genome regions of two Old World monkey retroviruses, CPC-1 and MAC-1, have been found to be more closely related to one another than to other retroviruses. Previous 5'-terminal sequence analysis had established two subgroups (13, 14) of mammalian type C viruses: (i) viruses of murine origin, including Rauscher murine leukemia virus and gibbon ape leukemia virus and (ii) viruses of Old World monkey origin, including BaEV and RD114. This classification is similar to that based on NH₂-terminal sequences of the major structural proteins of retroviruses, suggesting that 5'-ter-

minal sequence relationships are representative of the genome (16).

Based on the present results, CPC-1 and MAC-1 are approximately equally distinct from the viruses of the other two subgroups and constitute a third subgroup of viruses. These results indicate that there are two distinct lineages of viruses endogenous to Old World monkeys. This observation confirms previous results demonstrating multiple sets of endogenous type C viral genes in Old World monkey hosts (6). Because viral gene sequences related to this third subgroup (MAC-1 and CPC-1) are found in the cellular DNAs of various carnivores (5), this subgroup may represent an ancient interspecies

transmission of viruses between carnivores and Old World monkeys.

SNV, a member of the avian reticuloendotheliosis viruses, has been shown to be related to type C mammalian viruses (2, 3, 11). The 5'-terminal region of SNV has recently been sequenced (22), allowing us to compare it with each of the other type C virus sequences. Surprisingly, SNV aligns statistically to a higher degree with members of the third subgroup, CPC-1 and MAC-1, than with the other mammalian type C viruses. In contrast, there is no evidence of relatedness between SNV and the avian leukosis-sarcoma group of viruses. There are two highly conserved regions (residues 41 through 62 and 163 through 179) containing several functional signals which are identifiable with other mammalian type C viruses but closest to CPC-1 and MAC-1. These results are in accord with those which show considerable NH₂-terminal amino acid sequence homology in the major structural proteins of MMC-1 and avian reticuloendotheliosis virus (S. Oroszlan, M. Barbacid, T. Copeland, S. Aaronson, and R. V. Gilden, submitted for publication).

Between the 5' terminus of the genome and the tRNA primer site, SNV contains 180 nucleotides, which makes it at least 30 nucleotides longer than other type C viruses. Alignment of SNV with other type C viruses shows two long sequences (residues 64 through 76 and 125 through 148), one U rich and the other G rich, not found elsewhere and thus probably acquired recently. The proposed secondary structure of SNV (Fig. 5) shows that both nucleotide inserts are involved in forming stable hairpin loops. Since potential secondary structure may differentiate the 3' and 5' ends of viral RNA transcripts (4), these hairpin loops may be involved in regulating initiation or termination of transcription. This is critical to the life cycle of the virus, because proposed signals for the termination of transcription must be silent when present at the 5' end of the integrated genome but functional when present at the 3' end to ensure synthesis of complete viral transcripts.

Since SNV and other avian reticuloendotheliosis viruses have no detectable genetic homologies with the DNAs of their natural hosts, waterfowl and turkeys (12), an ancestor of the avian reticuloendotheliosis viruses was probably transmitted from a monkey to an avian host. Inasmuch as the proteins and nucleic acids of avian reticuloendotheliosis viruses show very close sequence homology to one another (12, 18), interspecies transmission was probably rather recent. Furthermore, the intermediate alignment score (5 to 6 standard deviations) between SNV and CPC-1-MAC-1 would suggest that the

transmission was not directly from CPC-1 or MAC-1, but rather from another monkey host. In this connection, the recent report of the relationship of avian reticuloendotheliosis viruses with two viruses of New World species, the endogenous owl monkey virus (OMC-1) and deer kidney virus (1), is interesting.

Computer-assisted comparison of the 5'-terminal regions of mammalian type C viruses serves as a useful model of evolutionary divergence of noncoding nucleic acid sequences, allowing a fairly precise evolutionary tree to be drawn (Fig. 6). Islands of conserved nucleic acid sequences (the slowly divergent sequences) contained signals of translational, transcriptional, or integrative significance. These slowly divergent sequences were especially useful in discerning distant or intergroup relationships among the viruses. Between the conserved regions are rapidly divergent sequences in which base changes, insertions, and deletions are especially prevalent. These rapidly divergent sequences are useful in examining closely related viruses. Similar analysis of nonviral gene families (10) has shown great value in revealing conserved regions which would not otherwise be evident. Thus, analysis and comparison of 5'-terminal nucleotide sequences have been useful in defining common functional signals and in extending the matrix of relationships among retroviruses.

ACKNOWLEDGMENTS

We thank Sharon Bladen, Leah Ripper, and JoAnn Comer for their excellent technical assistance. We also thank Karen McNitt for guidance in the application of the computer programs.

FIG. 6. Evolutionary tree of mammalian type C viruses, showing relationships among mammalian retroviruses based on 5'-terminal nucleotide sequence comparisons. The lengths of the lines are proportional to the percentage differences among the nucleotides of the cDNA's. From the observed numbers of nucleotide differences, the numbers that must have occurred were calculated (8). The evolutionary distances thus include superimposed and back mutations. R-MuLV, Rauscher murine leukemia virus; GaLV, gibbon ape leukemia virus.

This work was supported by Public Health Service contract N01-CO-75380 with the National Cancer Institute.

LITERATURE CITED

1. Barbacid, M., M. D. Daniel, and S. A. Aaronson. 1980. Immunological relationships of OMC-1, an endogenous virus of owl monkeys, with mammalian and avian type C viruses. *J. Virol.* **33**:561-566.
2. Barbacid, M., E. Hunter, and S. A. Aaronson. 1979. Avian reticuloendotheliosis viruses: evolutionary linkage with mammalian type C retroviruses. *J. Virol.* **30**:508-514.
3. Bauer, G., and H. M. Temin. 1980. Specific antigenic relationships between the RNA-dependent DNA polymerases of avian reticuloendotheliosis viruses and mammalian type C retroviruses. *J. Virol.* **34**:168-177.
4. Benz, E. W., R. M. Wydro, B. Nadal-Ginard, and D. Dina. 1980. Moloney murine sarcoma proviral DNA is a transcriptional unit. *Nature (London)* **288**:665-669.
5. Bonner, T. L., and G. J. Todaro. 1979. Carnivores have sequences in their cellular DNA distantly related to the primate endogenous virus, MAC-1. *Virology* **94**:224-227.
6. Bryant, M. L., C. J. Sherr, A. Sen, and G. J. Todaro. 1978. Molecular diversity among five different endogenous primate retroviruses. *J. Virol.* **28**:300-313.
7. Charman, H. P., R. V. Gilden, and S. Oroszlan. 1979. Reticuloendotheliosis virus: detection of immunological relationship to mammalian type C retroviruses. *J. Virol.* **29**:1221-1225.
8. Dayhoff, M. O. 1976. Atlas of protein sequence and structure, 5, suppl. 2, p. 1. National Biomedical Research Foundation, Washington, D.C.
9. Efstratiadis, A., T. Maniatis, F. C. Kafatos, A. Jeffrey, and J. N. Vournakis. 1975. Full length and discrete partial reverse transcripts of globin and chorion mRNAs. *Cell* **4**:367-378.
10. Efstratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Barelle, C. C. Shoulders, and N. J. Proudfoot. 1980. The structure and evolution of the human β -globin gene family. *Cell* **21**:653-668.
11. Hunter, E., A. S. Bhowan, and J. C. Bennett. 1978. Amino-terminal amino acid sequences of the major structural polypeptides of avian retroviruses: sequence homology between reticuloendotheliosis virus, p30 and p30s of mammalian retroviruses. *Proc. Natl. Acad. Sci. U.S.A.* **75**:2708-2712.
12. Kang, C.-Y., and H. M. Temin. 1974. Reticuloendotheliosis virus nucleic acid sequences in cellular DNA. *J. Virol.* **14**:1179-1188.
13. Lovinger, G. G., and G. Schochetman. 1979. 5'-Terminal nucleotide sequences of the Rauscher leukemia virus and gibbon ape leukemia virus genomes exhibit a high degree of correspondence. *J. Virol.* **32**:803-811.
14. Lovinger, G. G., and G. Schochetman. 1980. 5' terminal nucleotide sequences of type C retroviruses: features common to noncoding sequences of eucaryotic messenger RNAs. *Cell* **20**:441-449.
15. Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* **74**:560-564.
16. Oroszlan, S., and R. V. Gilden. 1980. Primary structure analysis of retrovirus proteins, p. 299-344. In J. R. Stephenson (ed.), *Molecular biology of RNA tumor viruses*. Academic Press, Inc., New York.
17. Proudfoot, N. J., and G. G. Brownlee. 1976. 3' noncoding region sequences in eukaryotic messenger RNA. *Nature (London)* **263**:211-214.
18. Purchase, H. G., and R. L. Witter. 1975. The reticuloendotheliosis viruses. *Curr. Top. Microbiol. Immunol.* **71**:103-124.
19. Queen, C. L., and J. L. Korn. 1980. Computer analysis of nucleic acids of proteins. *Methods Enzymol.* **65**:595-609.
20. Rabin, H., C. V. Benton, M. A. Tainsky, N. R. Rice, and R. V. Gilden. 1979. Isolation and characterization of an endogenous type C virus of Rhesus monkeys. *Science* **204**:841-842.
21. Sherwin, S. A., and G. J. Todaro. 1979. A new endogenous primate type C virus isolated from the Old World monkey *Colobus polykomos*. *Proc. Natl. Acad. Sci. U.S.A.* **76**:5041-5045.
22. Shimotohno, K., S. Mizutani, and H. Temin. 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. *Nature (London)* **285**:550-554.
23. Temin, H. M. 1980. Origin of retroviruses from cellular movable genetic elements. *Cell* **21**:599-600.
24. Todaro, G. J. 1980. Interspecies transmission of mammalian retroviruses, p. 47-76. In J. R. Stephenson (ed.), *Molecular biology of RNA tumor viruses*. Academic Press, Inc., New York.
25. Todaro, G. J., R. E. Benveniste, S. A. Sherwin, and C. J. Sherr. 1978. MAC-1, a new genetically transmitted type C virus of primates: "low frequency" activation from stump-tail monkey cell cultures. *Cell* **13**:775-782.