JOURNAL OF VIROLOGY, Sept. 1988, p. 3540–3543
0022-538X/88/093540-04$02.00/0

Vol. 62, No. 9

Harvey Sarcoma Virus Genome Contains No Extensive Sequences Unrelated to Those of Other Retroviruses except *ras*

KENNETH F. MANLY,* GARTH R. ANDERSON, AND DANIEL L. STOLER

Department of Cell and Molecular Biology, Roswell Park Memorial Institute, Buffalo, New York 14263

Received 8 February 1988/Accepted 13 May 1988

The Harvey murine sarcoma virus genome contains two rat-derived sets of genetic information recombined with the Moloney mouse leukemia virus. The rat sequences represent a *ras* oncogene and a rat VL30 element. The VL30 sequences have several discrete regions of similarity with retroviral sequences which were detected by searching a protein database for similarities with predicted polypeptide sequences from the VL30 regions. On the 5′ side, the most similar sequences were those of feline sarcoma viruses; on the 3′ side, murine leukemia viruses were the most similar. Some of the regions of similarity could also be detected directly by searching a nucleic acid sequence database with the viral DNA sequences. The most extensive region of similarity was that which corresponded to the endonuclease in the *pol* gene of a murine leukemia virus. The majority of the rat-derived sequences present in the Harvey sarcoma virus genome can now be attributed exclusively to *ras* or retrovirus- or retrotransposon-related sequences.

VL30 elements are a multigene family, including about 50 to 100 members per genome in rats and mice (2, 13). These elements, approximately 5.5 kilobases in length, have structures similar to those of retroviruses and retrotransposons, with long terminal repeat sequences and tRNA primer-binding sites (15, 22). Transcripts of VL30 elements are efficiently packaged as pseudotypes by type C retroviruses and are also efficiently copied by reverse transcriptase (14, 16). When introduced into cells, pseudotyped VL30 RNA can lead to the integration of DNA copies at new sites in the cell genome, suggesting that VL30 elements should be considered a class of transposable elements (6, 21).

The VL30 elements particularly resemble the retrotransposon class of transposable elements (for a review, see reference 30). These elements, including copia in *Drosophila melanogaster* and Ty in *Saccharomyces cerevisiae*, contain their own long terminal repeat sequences and are similar in size to VL30 elements (5, 27). They can be expressed as RNA, which then serves both as a message for translation into proteins and as a template for synthesis of additional DNA copies. These new copies can integrate into host cell DNA at new sites, sometimes affecting the expression of host genes near the new integration site.

VL30 elements were originally thought to be unrelated to typical retroviruses (3, 13, 24). More recent work has shown that certain fragments from both a cloned mouse VL30 element and a cloned rat VL30 element hybridize to mouse retroviral sequences (also cloned) under moderately stringent conditions (12). The related sequences appear to include more than one region in the VL30 sequences and to represent more than one region in the retroviral sequences. The rat VL30 sequences appear to be more closely related to mouse retroviral sequences than are the mouse VL30 sequences.

The Kirsten and Harvey murine sarcoma viruses are acute transforming retroviruses which arose independently following passage of mouse leukemia viruses in rats. Each incorporated two distinct genomic elements derived from rat cells: a 1-kilobase *ras* oncogene which encodes the well-studied p21 polypeptide, and approximately 4 kilobases of a rat VL30 element(s) (10, 29). Both *ras* and VL30 sequences appear to contribute to the oncogenic activity of these viruses (29). VL30 element transcription is strongly induced as a cellular response to anoxic stress (2a). We previously presented evidence suggesting that rat VL30 sequences incorporated in the Kirsten sarcoma virus genome might encode the major anoxic stress protein p34, lactate dehydrogenase k (1, 2a).

The complete sequence of Harvey sarcoma virus (HaSV) has been reported (26). The sequences of this virus which are believed to be derived from VL30 sequences constitute the most extensively reported sequence for a VL30 element. To characterize the VL30 domain of HaSV and to evaluate the similarity of VL30 sequences with retroviral sequences, we have compared translated HaSV sequences with the sequences of the Protein Identification Resource (11) available through BIONET (17, 25). We found discrete regions of the VL30-derived sequences of HaSV which were separately similar to regions of the *gag* and *pol* genes of feline or murine retroviruses, with closest similarity to the endonuclease domain of murine retroviral *pol*. No sequences related to any known dehydrogenases were found.

The HaSV sequence (26) was obtained in computer-readable format from R. O'Neill. This file was uploaded to BIONET and translated in three reading frames with the PEP program (25). Similar sequences were found by searching the sequences of the Protein Identification Resource (release 13.0, June 1987) (11) with the IFIND program (25) or the FASTP program (18), available on BIONET as XFASTP. Selected similarities were evaluated for significance with the RDF program, which compares the observed similarity score with a group of similarity scores obtained by randomizing one of the sequences many times (18). Finally, subsequences of the HaSV RNA sequence corresponding to regions of polypeptide similarity were compared with GenBank (4) viral nucleic acid sequences by using the search program for nucleic acid sequences XFASTN.

Terminator codons in the HaSV RNA sequence were converted to X's to allow them to be accepted by FASTP. The X character is treated by FASTP as an unknown residue and given an intermediate relatedness score (8) in comparison with any other amino acid.

* Corresponding author.

TABLE 1. High-scoring matches between sections of translated HaSV sequences and sequences of the Protein Identification Resource database

Region	Database sequence (source)[a]	Similarity score[b]	z score[c]
A	Residues 100 to 220 frame A		
	gag polyprotein (McDonough feline sarcoma virus)	95	9
	gag polyprotein (Gardner-Arnstein feline sarcoma virus)	93	
	gag polyprotein (Snyder-Theilen feline sarcoma virus)	81	
B	Residues 660 to 730, frames A and B		
	gag polyprotein (simian sarcoma virus)	155	
	gag polyprotein (McDonough feline sarcoma virus)	149	20
	gag polyprotein (baboon endogenous virus)	149	
C	Residues 750 to 830, frame C		
	gag polyprotein (simian sarcoma virus)	83	
	gag polyprotein (McDonough feline sarcoma virus)	71	5
	Histone H5 (goose)	57	
D	Residues 790 to 880, frame A		
	gag polyprotein (AKV murine leukemia virus)	131	
	gag polyprotein (Moloney murine leukemia virus)	131	19
	gag polyprotein (baboon endogenous virus)	112	
E	Residues 910 to 970, frame C		
	pol polyprotein (AKV murine leukemia virus)	174	
	pol polyprotein (Moloney murine leukemia virus)	174	26
	Alpha-galactosidase precursor (*Saccharomyces cerevisiae*)	64	
F	Residues 1020 to 1070, frame C		
	pol polyprotein (Moloney murine leukemia virus)	67	7
	pol polyprotein (AKV murine leukemia virus)	65	
	Gelsolin precursor (Human plasma)	49	
G	Residues 1050 to 1160, frame A		
	pol polyprotein (AKR murine leukemia virus [fragment])	94	
	pol polyprotein (AKV murine leukemia virus)	94	
	pol polyprotein (Moloney murine leukemia virus)	90	9
H	Residues 1160 to 1420, frames A, B, and C		
	pol polyprotein (Moloney murine leukemia virus)	427	42
	pol polyprotein (AKV murine leukemia virus)	428	
	pol polyprotein (AKR murine leukemia virus [fragment])	416	

[a] The residue numbers indicate the extent of the sequence submitted to FASTP to produce the similarity score shown. The actual region of similarity is somewhat smaller.

[b] The FASTP program was used to search the Protein Identification Resource database for sequences similar to the indicated HaSV polypeptide sequences. The three sequences with highest optimized similarity scores are shown with their scores.

[c] The z score is the alignment score for the indicated sequence expressed as the number of standard deviations above the mean of a set of scores from randomized sequences.

Discrepancies in HaSV sequences. The *ras* sequences of HaSV have been sequenced from different viral DNA clones by two groups (9, 26). The sequence analyzed here (26) differs from the other (9) by 22 base changes, 9 nucleotide additions, and 2 nucleotide deletions. Most of these discrepancies occur in the 230 nucleotides immediately 5′ of the *ras* coding sequences. These discrepancies suggest that terminator codons or frameshifts in the published sequences may not exist in the VL30 element from which the HaSV sequences were derived or, possibly, in the HaSV sequences themselves. We therefore analyzed these sequences in all reading frames without respect to termination codons.

Polypeptide similarities. The searches described above yielded eight major regions of sequence similarity (Table 1). These are referred to as regions A through H for discussion. Regions A through C showed greatest similarity with *gag* sequences of feline sarcoma virus. The remaining five regions showed greatest similarity with *gag* or *pol* regions of murine leukemia viruses, particularly Moloney or AKR-derived leukemia viruses. (Regions C and D could be considered a single region, but different sequences match with feline sarcoma virus and murine leukemia virus.) No regions showed significant similarity to retroviral *env* genes except those believed to be derived directly from Moloney leukemia virus (10, 29).

In two cases (B and H), the regions shown are composites of more than one reading frame. In these cases, our original searches showed immediately adjacent HaSV regions in different frames which matched with immediately adjacent regions of viral sequences. We interpreted this to mean that a mutation in the HaSV sequence (or a sequencing inaccuracy) had split the original coding sequence between two or more reading frames. To help evaluate the original extent of similarity, we combined the sequences from different frames and searched the database again, treating the combined sequences as one.

To locate the regions of similarity, the HaSV regions were aligned with sequences of one of two viruses, either the McDonough strain of feline sarcoma virus or the Moloney strain of murine leukemia virus. These viral sequences were among those which had matched best with several of the HaSV regions. The three most significant alignments are shown in Fig. 1, and the arrangement of all similar regions in the viral genomes is summarized in Fig. 2.

The region of VL30 showing the greatest similarity to retroviral sequences is region H (Fig. 2), spanning residues 1160 to 1420. This corresponds to the C-terminal region of the *pol* polyprotein, which is cleaved to yield an endonuclease. This region, illustrated in detail in Fig. 1, shows identity between VL30 and *pol* for 110 of 210 residues. Regions A, C,

B. Similarity with N-proximal p30 region of McDonough feline sarcoma virus

```
               ---------------- frame A ---------------|--- frame B ----
                                                       |
                    670       680       690       700  |   710       720
HASV       VXLNAICPLQWTSLLXIPFLFSHHPTXDACXXLLQIFFTTEERQ|RILLKARKNVPDMDGRLLTVDF
                :.. :.:.  ..: .:.::.:.:..::: ..:.::::|:.::.:::.::. :::
FOMVMD     NPPFSQDPVALTNLIE-SILVTHQPTWDDCQQLLQALLTAEERQ|RVLLEARKQVPGEDGRPTQLPN
                    310        320       330       340 |   350       360
```

E. Similarity with protease region of *pol* of Moloney leukemia virus

```
               ---------------------------------- frame C ---------------------
                     910       920       930       940       950       960       970
HASV       PNALCSYAPQWANIQRPXVQGATGNKQYLETARRTVDLGVGRVTHEYMVIPDCPYPLLMRNLLSRMACVG
                         ....:::::::.:.:  :. :.:.:..:.:::... .:::::::: :.::...
GNMV1M     QHSVLTQNPGPLSDKSAWVQGATGGKRYRWTTDRKVHLATGKVTHSFLHVPDCPYPLLGRDLLTKLKAQI
                     40        50        60        70        80        90       100
```

H. Similarity with endonuclease region of Moloney leukemia virus

```
             -------- frame B -------|---------------- frame A --------------|-------
                                     |                                       |
          1160      1170      1180   | 1190      1200      1210      1220    | 1230
HASV        TTEXIKAFLTKRETASTIIKKILEEIF|P-LGMPKVIWSDNGPTFVAKVSQGVAKYLEVDXKLHCIY|RPQSSGQ
             ...:.:: ::.:::....::.:::::|: .:::.:. .::::.::.::::.::. :..:.:::: :|:::::::
GNMV1M      FSGWIEAFPTKKETAKVVTKKLLEEIF|PRFGMPQVLGTDNGPAFVSKVSQTVADLLGIDWKLHCAY|RPQSSGQ
             940       950       960   |  970       980       990      1000    |

        frame B -|----- frame C ------|-------------- frame A --------------|-----
                 |                    |                                     |
            1240 |     1250           |          1270      1280      1290   | 130
HASV    VGXINKTLKR|PDPTKLIMETGT-DWVTLLP|PLALFRARNTPSRFSLTPFEILYGASVLTVLDDVTEP|NMSXC
        :...:.:.: | . ::::..::. ::: :: |::::.:::::::.. .:::.:::::::.  . : . ..:|.:..
GNMV1M  VERMNRTIK-|ETLTKLTLATGSRDWVLLL-|PLALYRARNTPGPHGLTPYEILYGAP--PPLVNFPDP|DMTRV
     1010        |       1030          1040      1050      1060       1070  |

          ---------------------------------- frame C ----------------------------------
        1300      1310      1320       1330      1340      1350      1360
HASV       HSNNDLCARLKDLQVIQKEICSELAAAY--ALGTPETSHQFQ--SETRL-HIMAPNPDTHMKGPYLVLLTTLTA
           .. .: :.:..: ..:.:. . :::::  .:. : ..:...  ..... .  . : ...::::: ::::: ::
GNMV1M     TNSPSLQAHLQALYLVQHEVWRPLAAAYQEQLDRPVVPHPYRVGDTVWVRRHQTKNLEPRWKGPYTVLLTTPTA
        1080      1090      1100       1110      1120      1130      1140      1150
```

FIG. 1. Alignment of high-scoring regions of HaSV VL30 sequences. HASV sequences are numbered by amino acid residues from the 5′ end of the genome. Murine leukemia virus *gag-pol* sequences (GNMV1M) and feline sarcoma virus *pol* sequences (FOMVMD) are numbered by residues from the amino-terminal end of the polyprotein.

F, and G show relatively weak similarity, and B, D, and E show intermediate levels. VL30 sequences in the HaSV genome do not represent all regions of the leukemia virus genomes (Fig. 2). However, the VL30 regions which resemble leukemia virus sequences are arranged in the same order as their corresponding leukemia virus sequences.

Significance of alignments. The XRDF program (18) was used to estimate the significance of the alignments shown in Table 1. For each alignment, the leukemia virus sequence was randomized 20 times, and with each random sequence an alignment score was calculated for the HaSV sequence. The z score shown in Table 1 is the difference between the original alignment score and the mean of the random scores divided by the standard deviation of the random scores. Scores greater than 3 are considered marginally significant; scores greater than 10 are highly significant (18). In two cases (regions B and H), the significance may be overestimated because these sequences were deliberately constructed to allow for hyprothetical frameshifts. However, in both cases, similar matches were found by searching for similar sequences with the original nucleic acid sequences.

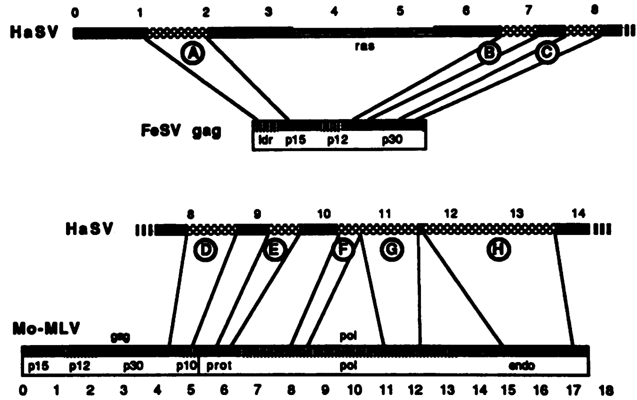

FIG. 2. Arrangement of similarities between HaSV VL30 sequences and mammalian leukemia virus sequences. HaSV and leukemia virus sequences are numbered in hundreds of amino acid residues; the HaSV sequences are drawn at twice the scale of the leukemia virus sequences. Matching regions from Table 1 are indicated by connecting lines and identified by the letters A through H. FeSV, Feline leukemia virus; Mo-MLV, Moloney murine leukemia virus.

Nucleic acid similarities. We sought confirmation of the protein sequence similarities by searching for nucleic acid similarities. Subsequences of the HaSV genome corresponding to the regions of protein similarity were constructed for searches of the viral nucleic acid sequences in the GenBank (4) database available on BIONET. In general, the results of these searches confirmed the protein sequence similarities. Region H matched especially well, achieving a 62% identity over a 625-base overlap with sequences of Moloney leukemia virus. In addition, the nucleic acid alignment showed base insertions or deletions corresponding in location to each of the frameshifts introduced into the polypeptide sequence shown in Table 1.

Conserved patterns. The similarity with p30 of Moloney leukemia virus includes an imperfect copy of the conserved $CX_2CX_4HX_4C$ motif (7) at residues 854 to 867 in region D. The similarity with the reverse transcriptase region (region F) includes only 10%, at most, of a pattern identified as being conserved among reverse transcriptases (28).

Other viruslike elements. Two other viruslike elements, not closely related to VL30 elements, have been sequenced: the murine retrovirus-related element (23) and the RTVL2-H2 (19). These elements resemble the VL30 sequences in having mosaic similarity with leukemia virus sequences, but the similarities occur in different regions (data not shown).

Structure of VL30 sequences. The size of the VL30 sequences of HaSV indicates that they constitute slightly over half of a complete VL30 element. Our observations confirm and extend previous hybridization and sequencing data which suggested a distant evolutionary relationship between rat VL30 sequences and murine leukemia virus sequences (12, 20). In addition, our observations suggest a relationship between some VL30 sequences and sequences of the retrovirus group which includes feline sarcoma virus and baboon endogenous retrovirus. The patchy nature of the sequence similarities suggests that recombination was an important process in the derivation of VL30 elements from the ancestor they share with the leukemia/sarcoma viruses.

We have previously suggested that an anoxic stress protein (LDH_k) with lactate dehydrogenase and nucleic-acid-binding activity might be encoded by the VL30 sequences incorporated into HaSV and Kirsten sarcoma virus (2a). Our sequence analysis makes this hypothesis less likely. The searches reported here failed to find any similarity with known dehydrogenase sequences. Furthermore, no large subsequence exists which does not relate to either *ras* or retroviral sequences. Either LDH_k activity would have to be encoded by sequences which resemble leukemia virus sequences or LDH_k would have to be encoded in a different frame from the leukemia virus sequences, overlapping them.

At first glance, the VL30 sequences would appear not to be coding sequences themselves; the reading frames are interrupted by numerous terminator codons and frameshifts. This does not necessarily reflect the state of these sequences in VL30 elements. Some of the interruptions may have been introduced during incorporation of these sequences into HaSV, during nonselective passage prior to cloning, or during cloning itself. Some may have been introduced by sequencing inaccuracies. Moreover, since VL30 elements are heterogenous, other elements may have retained more coding capacity than the element which was incorporated into HaSV. It seems reasonable, therefore, to let the HaSV VL30 sequences suggest what functions might actually be expressed by some VL30 elements. Since the similarity with endonuclease sequences is especially extensive, it may be prudent to consider that function first.

We thank Ray O'Neill for important conversations and for supplying the Harvey sequence data in computer-readable format.

This work was supported in part by grant 85-CRCRI-1900 to G.R.A. from the U.S. Department of Agriculture and by cooperative agreement 1 U41 RR01685 between the National Institutes of Health and Intelligenetics, Inc., in support of BIONET.

LITERATURE CITED

1. **Anderson, G. R., and B. K. Farkas.** 1988. The major anoxic stress response protein p34 is a distinct lactate dehydrogenase. Biochemistry **27**:2187–2193.
2. **Anderson, G. R., and K. C. Robbins.** 1976. Rat sequences of the Kirsten and Harvey murine sarcoma virus genomes: nature, origin, and expression in rat tumor RNA. J. Virol. **17**:335–351.

2a. **Anderson, G. R., D. L. Stoler, J. P. Scott, and B. K. Farkas.** 1988. Induction of VL30 element expression as a response to anoxic stress. Banbury Rep. **30**:265–274.

3. **Besmer, P. K., U. Olshevsky, D. Baltimore, D. Dolberg, and H. Fan.** 1979. Virus-like 30S RNA in mouse cells. J. Virol. **29**:1168–1176.
4. **Bilofsky, H. S., C. Burks, J. W. Fickett, W. B. Goad, F. I. Lewitter, W. P. Rindone, C. D. Swindell, and C. S. Tung.** 1986. The GenBank genetic sequence database. Nucleic Acids Res. **14**:1–4.
5. **Boeke, J. D., D. J. Garfinkel, C. A. Styles, and G. R. Fink.** 1985. Ty elements transpose through an RNA intermediate. Cell **40**: 491–500.
6. **Carter, A. T., J. D. Norton, Y. Gibson, and R. J. Avery.** 1986. Expression and transmission of a rodent retrovirus-like VL30 gene family. J. Mol. Biol. **188**:105–108.
7. **Covey, S. N.** 1986. Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucleic Acids Res. **14**:623–633.
8. **Dayhoff, M. O., R. V. Eck, and C. M. Park.** 1972. A model of evolutionary change in proteins. *In* M. O. Dayhoff (ed.), An atlas of protein sequence and structure. National Biomedical Research Foundation, Silver Spring, Md.

9 **Dhar, R., R. W. Ellis, T. Y. Shih, S. Oroszlan, B. Shapiro, J. Maizel, D. Lowy, and E. Scolnick.** 1982. Nucleotide sequence of the p21 transforming protein of Harvey murine sarcoma virus. Science **216**:934–937.

10. **Ellis, R. W., W. DeFeo, T. Y. Shih, M. A. Gonda, H. A. Young, N. Tsuchida, D. R. Lowy, and E. M. Scolnick.** 1981. The p21 *src* genes of Harvey and Kirsten sarcoma viruses originate from divergent members of a family of normal vertebrate genes. Nature (London) **292**:506–511.
11. **George, D. G., W. C. Barker, and L. T. Hunt.** 1986. The Protein Identification Resource (PIR). Nucleic Acids Res. **14**:11–15.
12. **Giri, C. P., C. P. Hodgson, P. K. Elder, M. G. Courtney, and J. J. Getz.** 1983. Discrete regions of sequence homology between cloned rodent VL30 genetic elements and Akv-related MuLV provirus genomes. Nucleic Acids Res. **11**:305–319.
13. **Howk, R. S., D. H. Troxler, D. Lowy, P. H. Duesberg, and E. M. Scolnick.** 1978. Identification of a 30S RNA with properties of a defective type C virus in murine cells. J. Virol. **25**:115–123.
14. **Itin, A., and E. Keshet.** 1985. Primer binding sites corresponding to several tRNA species are present in DNAs of different members of the same retrovirus-like gene family (VL30). J. Virol. **54**:236–239.
15. **Keshet, E., and Y. Shaul.** 1981. Terminal direct repeats in a retrovirus-like repeated mouse gene family. Nature (London) **289**:83–85.
16. **Keshet, E., Y. Shaul, J. Kaminchik, and H. Aviv.** 1980. Heterogeneity of virus-like genes encoding retrovirus-associated 30S RNA and their organization within the mouse genome. Cell **20**: 431–439.
17. **Kristofferson, D.** 1987. The BIONET electronic network. Nature (London) **325**:555–556.
18. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
19. **Mager, D. L., and J. D. Freeman.** 1987. Human endogenous retroviruslike genome with type C *pol* sequences and *gag* sequences related to human T-cell lymphotropic viruses. J. Virol. **61**:4060–4066.
20. **Pampano, C. L., and D. Meruelo.** 1986. Isolation of a retroviruslike sequence from the TL locus of the C57BL/10 murine histocompatibility complex. J. Virol. **58**:296–306.
21. **Rodland, K. D., A. Brown, and B. E. Magun.** 1986. Individual mouse VL30 elements transferred to rat cells by viral pseudotypes retain their responsiveness to activators of protein kinase C. Mol. Cell. Biol. **7**:2296–2298.
22. **Rotman, G., A. Itin, and E. Keshet.** 1986. Promoter and enhancer activities of long terminal repeats associated with cellular retrovirus-like (VL30) elements. Nucleic Acids Res. **14**:645–658.
23. **Schmidt, M., T. Wirth, K. Burkhard, and I. Horak.** 1985. Structure and genomic organization of a new family of murine retrovirus-related DNA sequences (MuRRS). Nucleic Acids Res. **13**:3461–3470.
24. **Sherwin, S. A., U. R. Rapp, R. E. Benveniste, A. Sen, and G. J. Todaro.** 1978. Rescue of endogenous 30S retroviral sequences from mouse cells by baboon type C virus. J. Virol. **26**:257–264.
25. **Smith, D. H., D. Brutlag, P. Friedland, and L. Kedes.** 1986. BIONET: national computer resource for molecular biology. Nucleic Acids Res. **14**:17–20.
26. **Soeda, E., and S. Yasuda.** 1985. Harvey murine sarcoma virus (Ha-MSV) genome, p. 928–939. *In* R. Weiss (ed.), RNA tumor viruses, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Strand, D. J., and J. F. McDonald.** 1985. Copia is transcriptionally responsive to environmental stress. Nucleic Acids Res. **13**: 4401–4410.
28. **Toh, H., H. Hayashida, and T. Miyata.** 1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature (London) **305**:827–829.
29. **Wei, C.-M., D. R. Lowy, and E. M. Scolnick.** 1980. Mapping of transforming region of the Harvey murine sarcoma virus genome by using insertion-deletion mutants constructed in vitro. Proc. Natl. Acad. Sci. USA **77**:4674–4678.
30. **Weinstein, I. B., J. F. McDonald, and M. E. Lambert.** 1988. Eukaryotic transposable elements as mutagenic agents. Banbury Rep. **30**:1–345.