Proc. Natl. Acad. Sci. USA
Vol. 77, No. 7, pp. 3937–3941, July 1980
Biochemistry

Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions

(retrovirus integration)

RAVI DHAR, WILLIAM L. MCCLEMENTS, LYNN W. ENQUIST, AND GEORGE F. VANDE WOUDE

Laboratory of Molecular Virology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

Communicated by David Baltimore, April 14, 1980

ABSTRACT **Integrated Moloney murine sarcoma provirus (MSV) has direct terminal repeat sequences (TRS). We determined the nucleotide sequence of both 588-base-pair TRS elements and the adjacent host and viral junctions of an integrated MSV cloned in bacteriophage λ. Sequences were identified corresponding to the $tRNA^{Pro}$ primer binding site in genomic RNA and the reverse-transcribed minus strong stop DNA. Each 588-base-pair repeat contains putative sites for promoting RNA synthesis and RNA polyadenylylation. The first and last 11 nucleotides of the TRS are inverted with respect to each other, and the same four-nucleotide host sequence is found bracketing integrated MSV. Some similarities of TRS and prokaryotic insertion sequence elements are discussed.**

The integrated and unintegrated proviruses of both avian and mammalian retroviruses have terminal repeat sequences (1–7). These sequences are derived from specific 5′ and 3′ regions of genomic RNA that are translocated to opposite termini during reverse transcription (8–10). For both Moloney leukemia virus (M-MuLV) (4) and Moloney sarcoma virus (MSV) (5–7), the length of the terminal repeat sequence is about 600 base pairs (8). Some consideration has been given to how terminal repeat sequences could function in provirus integration (ref. 3; unpublished data) and in the virus life cycle (3). We have demonstrated that these sequences are necessary for efficient transformation by the MSV *src* region (11).

The formation of the terminal repeat sequence is accounted for by two well-characterized intermediates of proviral DNA synthesis: minus strong stop DNA (ref. 12; see ref. 13 for review) and plus strong stop DNA (8–10, 14). The minus strong stop DNA for M-MuLV is a cDNA copy, 150 nucleotides in length, of the 5′ end of viral RNA (8, 12). An important feature of M-MuLV minus strong stop DNA is that it contains sequences complementary to the 3′ end of the viral RNA (15). This short terminal repeat sequence (trs) in the viral RNA enables the minus strong stop DNA to hybridize to the 3′ end of the RNA and to prime the synthesis of minus strand DNA (15–19). The plus strong stop DNA for M-MuLV and MSV is a complementary copy of the 5′ end of the minus strand DNA. It is approximately 600 nucleotides long (8–10) and contains sequences derived from the 3′ end of viral RNA and a complementary copy of the minus strong stop DNA, including the $tRNA^{Pro}$ primer binding site (8).

While elements of the terminal repeat sequences of unintegrated provirus may be identical to those of integrated provirus, the exact relationship is not known. We have cloned and characterized an integrated m1 MSV genome (7) that has terminal repeat sequences (TRS, to be distinguished from the terminal repeat sequences of unintegrated proviral DNA) apparently identical in size and having the same restriction endonuclease map as the terminal repeat sequence in *in vitro* synthesized (unintegrated) M-MuLV proviral DNA (refs. 4 and 6; unpublished data). Because MSV is derived from M-MuLV, it is highly likely that the structural and functional features of their terminal repeat sequences are the same. Although we use TRS to distinguish the long terminal repeat sequence of integrated provirus from its unintegrated form, the term LTR has been recently adopted for both forms.

One important step toward understanding the function(s) of TRS would be knowing its complete nucleotide sequence. We report here the nucleotide sequence of the 588-nucleotide MSV TRS elements and the adjacent host DNA and unique MSV DNA sequences.

MATERIALS AND METHODS

Source of DNA. An *Eco*RI fragment from transformed mink cells that contains a complete copy of integrated m1 MSV proviral DNA was cloned in phage λ (λ m1) as discussed (6, 7). The λm1 *Eco*RI fragment was subcloned in plasmid pBR322 in two portions by using the single *Bgl* II site and the two *Eco*RI sites indicated in Fig. 1*B*. Plasmid pm15 contains the 5′ end of the proviral DNA and pm13 contains the 3′ end. DNA fragments for sequence determinations were obtained from these plasmids by using restriction endonucleases and maps as described (ref. 6; unpublished data).

Isolation of Restriction Fragments, 5′-Terminus Labeling Procedures, and Methods of Sequence Determination. Restriction endonucleases were obtained from New England BioLabs or Bethesda Research Laboratories (Rockville, MD). Restriction fragments were first 5′-end labeled by using [γ-^{32}P]ATP (2000–5000 Ci/mmol; 1 Ci = 3.7×10^{10} becquerels; ICN) and T4 polynucleotide kinase as described by Maxam and Gilbert (20). The 5′-end labeled fragments were further digested with appropriate restriction endonucleases and isolated by elution from polyacrylamide gels after electrophoresis. The sequences of the fragments, labeled at only one 5′ end, were determined by a modification of the enzymatic method of Maat and Smith (21). Portions of the sequences were confirmed by the chemical method of Maxam and Gilbert (20).

RESULTS AND DISCUSSION

Organization of the λm1 *Eco*RI Fragment Containing m1 MSV Proviral DNA. A map of the *Eco*RI fragment of λm1 carrying one complete copy of integrated m1 MSV proviral DNA and flanking host sequences is shown in Fig. 1*B*. Viral RNA, with the specific $tRNA^{Pro}$ bound to the primer binding

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

Abbreviations: MSV, Moloney sarcoma virus; M-MuLV, Moloney leukemia virus; TRS, long terminal repeat sequence from integrated proviral DNA; trs, short terminal repeat sequence from viral RNA.

site (22), is shown in Fig. 1A. The fragment is oriented from left to right as 5′ to 3′ with respect to the viral RNA, with the first nucleotide of the sequenced portion near the left *Eco*RI site designated as 1 (Figs. 1C and 2). The TRS elements define the extent of integrated MSV information (6, 7). The notation trs in Fig. 1 *A* and *B* represents the terminal repeat sequences of M-MuLV RNA, estimated to be about 50–60 bases long (15).

Identification of TRS. The original identification and location of L- and R-TRS in λm1 were determined by mapping with five restriction enzymes that cleave within the repeat (ref. 6; unpublished data). The extent of each TRS element was obtained empirically by (*i*) determining the sequence of fragments that were predicted to span the provirus–host junctions as well as the TRS–unique MSV junctions and (*ii*) aligning the identical sequences from the left and right ends and identifying the points of divergence. This comparison revealed 588 base pairs of identical sequence that we define as TRS. In addition, it defined four junctions (Fig. 1*B*): external junctions of the host and provirus (a, d), and internal junctions of TRS and unique MSV (b, c). We will discuss the 588-base pair TRS elements first and then the junction sequences.

TRS of Integrated m1 MSV Provirus. We have determined the sequence of both TRS elements and of adjacent host and viral DNA. The particular stretches of sequence reported here are indicated in Figs. 1C and 2 (1–903, 5227–5872). Both TRS elements are identical. To facilitate their comparison, we have set the numbering of R-TRS to begin exactly 5000 base pairs from the beginning of L-TRS. When the L-TRS nucleotide coordinates are given, the corresponding position in R-TRS can be obtained by adding 5000. The features of the 1–903 and L-TRS nucleotide sequence described in detail below are illustrated in Fig. 1*D*. The complete nucleotide sequence is given in Fig. 2.

The identical 588-base-pair TRS elements (233–820 and 5233–5820) contain several notable features.

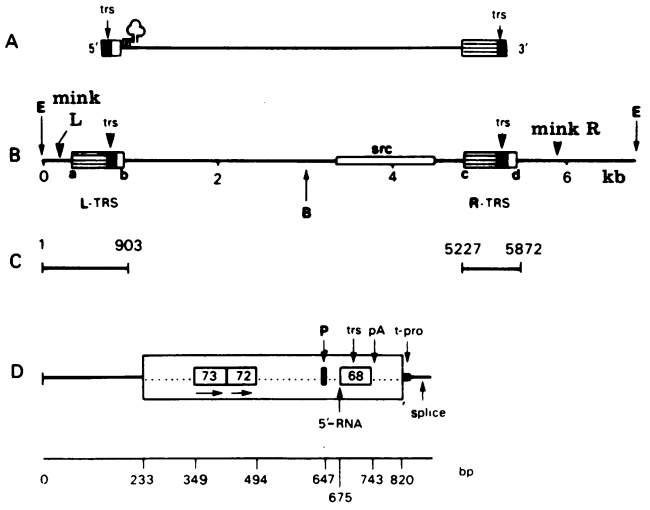

FIG. 1. Physical maps of m1 MSV. (*A*) MSV genomic RNA. The terminal repeat sequence in genomic RNA (15) (trs), $tRNA^{Pro}$ and binding site (cloverleaf structure), and the 5′ and 3′ genomic RNA contributions to TRS are indicated. (*B*) Physical map of λm1 cloned *Eco*RI fragment containing the integrated MSV provirus: mink L and R refer to the 5′ and 3′ mink DNA flanking sequences (refs. 6 and 7; unpublished data). The *Eco*RI (E) and *Bgl* II (B) sites were used for subcloning the fragment in pBR322 (11). The terminal repeat sequences are designated L- and R-TRS. Distance is indicated in kilobases (kb). (*C*) Regions of m1 MSV sequenced. (*D*) Identification of key locations within the L-TRS nucleotide sequence: P, possible promoter region; pA, poly(A) addition sequence; t-pro, $tRNA^{Pro}$ primer binding site; splice, possible splice-donor sequence. The locations of these features in the nucleotide sequence (Fig. 2) are indicated in base pairs (bp).

(*i*) *Inverted terminal sequences.* The first and last 11 bases are inverted repeat sequences (233–243; 810–820). To facilitate comparison of these sequences and junctions, we have drawn them in stem-and-loop configurations (Fig. 3 *A* and *B*). Because the TRS elements are direct repeats in the integrated provirus, the entire provirus has inverted terminal sequences (Fig. 3*C*).

(*ii*) *Sequence duplications.* An almost perfect duplication of 73 base pairs occurs in TRS (349–421; 423–494). The three nucleotides that are different have been underlined (two substitutions and one deletion). This duplication would be encoded near the 3′ end of genomic RNA. Duplications of similar size have been found in papovaviruses simian virus 40 and BKV (23), and possible functions have been discussed by Subramanian *et al.* (24). Computer analysis detected a much larger (≈200 base pairs) and highly diverged, partially overlapping duplication in TRS (480–680; 620–820). The significance of this duplication is unknown, but sequences from 620–820 include the putative promoter, the polyadenylylation sequence, trs, and the minus strong stop DNA.

(*iii*) *Location of the 5′ end of viral RNA within TRS.* A sequence within the TRS that is consistent with known elements of the 5′ end of the viral RNA genome was identified as follows. A 21-base sequence homologous with the 3′ sequence of $tRNA^{Pro}$ [5′ C-A-A-A-T-C-C-C-G-G-A-C-G-A-G-C-C-C-C-C-A 3′ (22)] was located at position 823–843. However, only 19 bases have been shown to interact (22). This defined the $tRNA^{Pro}$ binding site and locates the initial sequence of minus strong stop DNA. The nine-base sequence (5′ A-A-T-G-A-A-A-G-A 3′) immediately after the 5′ side of the tRNA binding site (814–822) is identical to the first nine bases incorporated into M-MuLV minus strong stop DNA (12). We used the reported length of the minus strong stop DNA, approximately 150 bases for M-MuLV (8, 12), to approximate the location of the 5′ end of the viral RNA sequence in TRS. It had been established that the methylated and "capped" sequence at the 5′ end of M-MuLV viral RNA is $7^{m}GpppG^{m}CG$ (25). Thus any G-C-G sequence in this region would be a likely candidate for the 5′ end of viral RNA. Three G-C-G triplets were considered. The first (698–700) was only 120 bases from the primer site and was thought to be too close to account for the estimated size of the minus strong stop DNA. The sequence G-C-G-C-G (675–679), containing two possible capping sites for the 5′ RNA terminus, is 148 or 146 bases from the primer binding site. The overlapping G-C-G sequences leave some ambiguity in assigning the 5′ end of the viral RNA. The next potential capping site (625–627) is too distant from the $tRNA^{Pro}$ binding site to be the 5′ end of viral RNA.

(*iv*) *Transcription initiation signals.* Putative eukaryote transcription initiation signals equivalent to the Pribnow box in *Escherichia coli* promoters (26), are A-T rich, favor a T-A-T-A-A-A-T-A sequence (generally called a Hogness box), and appear approximately 25 nucleotides preceding an RNA capping site. We find the sequence A-A-T-A-A-A-A (647–653) 22 bases preceding the G-C-G sequence at position 675 discussed in *iii*. In addition, six out of nine nucleotides in the model eukaryotic sequence (G-G-C-C-A-A-T-C-T; ref. 27) in the −70 to −80 position from the presumed sites for initiation of transcription, occur as part of a palindrome (592–603) 71 nucleotides before the G-C-G sequence. If this sequence were part of the promoter for viral RNA synthesis, it would be encoded in the 3′ end of the viral RNA and translocated to the 5′ side of the G-C-G sequences during proviral DNA synthesis. The generation of TRS would then represent a novel mechanism for assuring viral transcription after provirus integration. We note that another promoterlike sequence, 5′ T-A-T-A-A-A 3′

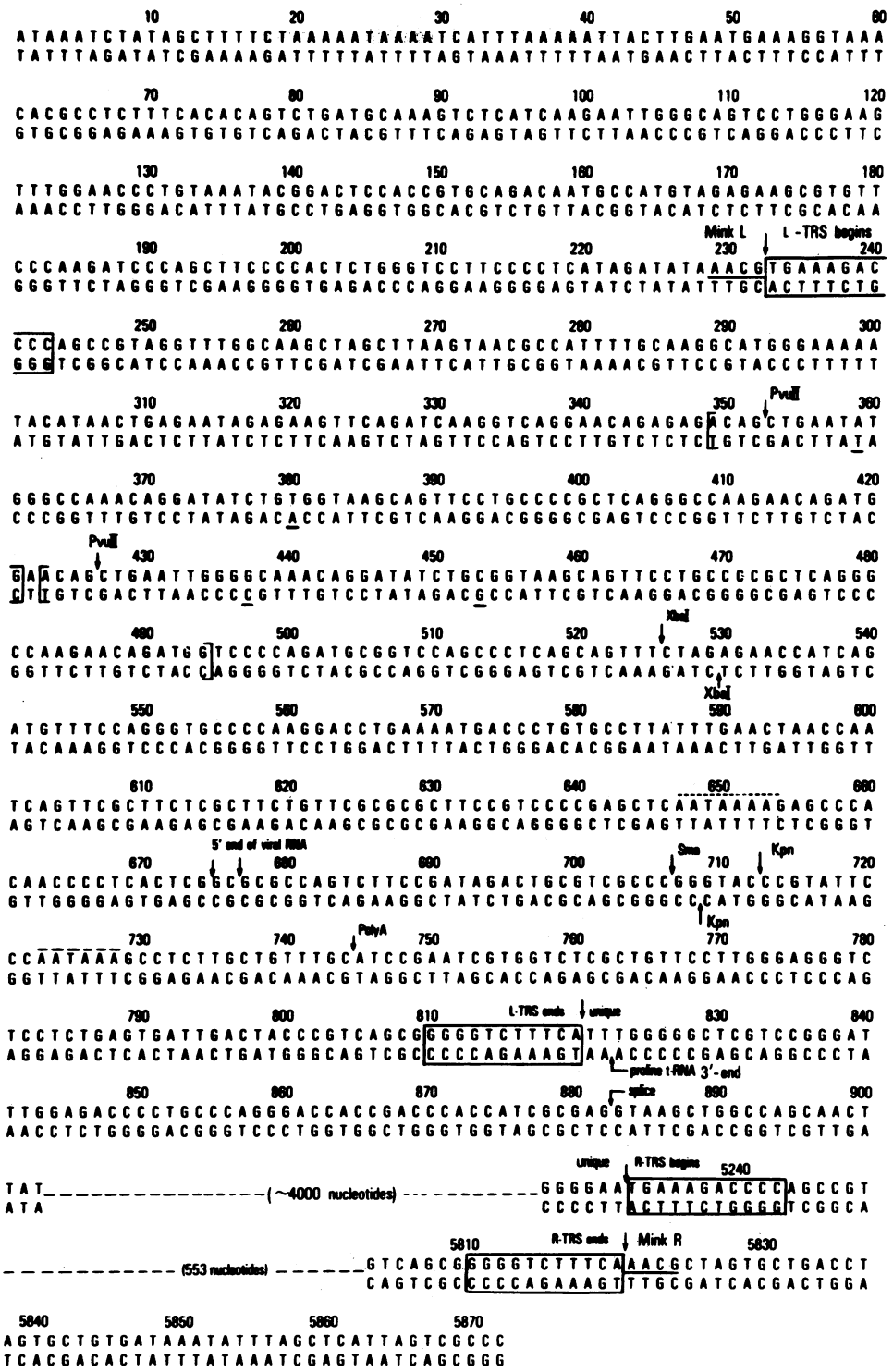

FIG. 2. The nucleotide sequence of m1 MSV TRS and its host and viral junctions. The nucleotide sequence includes the left mink flank (Mink L) (1–232), the 588-base-pair left TRS (L-TRS) (233–820), the unique 5′-MSV sequences (821–903), the unique 3′-MSV sequences (5227–5232), the 588-base-pair right TRS (R-TRS) (5233–5820), and the right mink flank sequences (Mink R) (5821–5872). From left to right (5′ to 3′ with respect to orientation of the viral RNA) the following sequences are identified: 11-base-pair inverted repeat (box) (compare with Fig. 3); 73-base-pair duplicated sequence (brackets) with the differences (two substitutions and one deletion) underlined; promoter-like sequence (overscored) 5′ end of viral RNA; possible polyadenylylation site (poly A); tRNA 3′ C-C-A end (proline tRNA 3′ end); and possible splice-donor site (splice). The restriction endonuclease sites, *Pvu* II, *Xba* I, *Sma* I, and *Kpn* I are indicated. The four-base duplicated host sequence is underlined (A-A-C-G; compare with Fig. 3*C*). For purposes of simplifying comparison, the first base of the R-TRS is synchronized with the first base of the L-TRS by adding 5000 base pairs. There are actually more than 5000 nucleotides between L- and R-TRS.

(225–230), is part of the duplicated host sequence at the site of MSV integration.

(*v*) *Polyadenylylation signals.* Polyadenylylation signals (5′ A-A-T-A-A-A 3′) are usually found 10–20 nucleotides preceding the dinucleotide C-A, the preferred site for polyadenylylation (28). We observed the sequence 5′ A-A-T-A-A-A 3′ (723–728) and a C-A dinucleotide 16 nucleotides downstream (744–745). This position is 68 or 70 bases from the two positions assigned

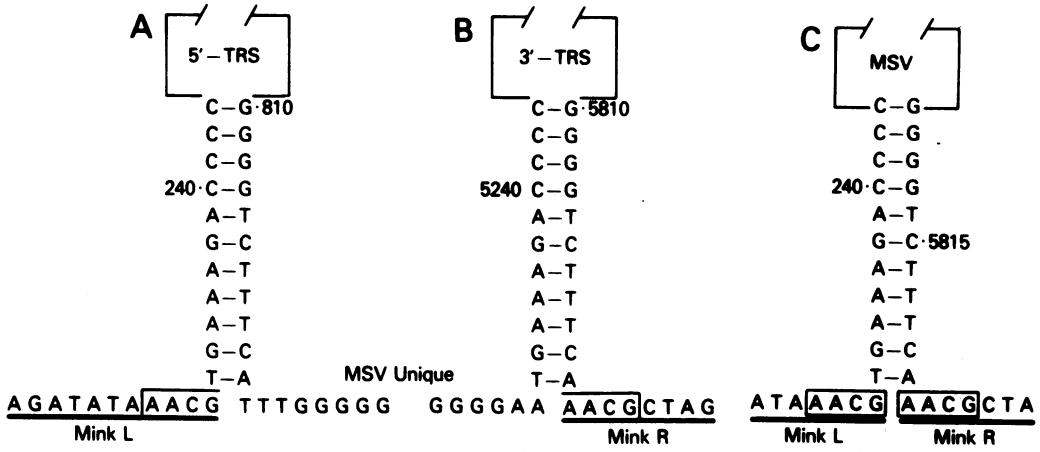

FIG. 3. Sequence arrangement of the host and viral TRS junctions. The 11-base-pair inverted repeat sequences of the TRS and specific mink or MSV junction sequences are indicated. (*A*) Mink L, L-TRS, and 5′ unique MSV junctions. (*B*) 3′ unique MSV, R-TRS, and Mink R junctions. (*C*) Mink L, MSV, and Mink R junctions. The identical four-base host sequences bracketing the integrated m1 MSV provirus are in boxes.

to the 5′ end of the viral RNA. The sequence between the putative 5′ end of the viral RNA and the polyadenylylation signal should delineate the trs. Its size is consistent with that estimated for M-MuLV trs (16). It is not known if this polyadenylylation signal functions in the L-TRS, but oligonucleotide mapping of the 3′ end of M-MuLV RNA (15) is consistent with termination of viral RNA synthesis occurring at the identical sequence in the R-TRS.

It is curious that the proposed promoter sequence (647–653) also satisfies the parameters of a polyadenylylation signal. The presence of two possible transcription termination signals in L-TRS may serve to terminate host transcription, so it does not interfere with provirus expression. In addition, the proposed polyadenylylation signal (723–728) has at the appropriate position (678–686) a sequence (5′ C-G-C-C-A-G-T-C-T 3′) similar to the −70 to −80 model eukaryotic sequence preceding the 5′ end of mRNA. If the proposed polyadenylylation signal functions as a promoter, its as yet unidentified transcript would be expected to begin at about position 753.

(*vi*) *Extent of TRS.* From the model of Gilboa *et al.* (8) for proviral DNA synthesis, we expected TRS to include *all* of the sequences of the minus strong stop DNA. The actual repeat sequences bracketing the integrated provirus, defined as TRS, lack the first two nucleotides of the minus strong stop DNA (compare A-A at positions 821 and 822 to TT at positions 5821 and 5822). While the products of proviral DNA synthesis may include these two nucleotides in TRS, they are not part of integrated TRS. This dinucleotide may be lost from the 3′ end of the R-TRS during proviral synthesis or upon integration. The critical sequences for integration could be the 11-nucleotide inverted repeat that excludes these two bases in an integration event analogous to prokaryote replicon fusion between a target site and an insertion sequence (IS) (29, 30).

Four TRS Junction Sequences. The external junctions (Fig. 1*B*, a and d) joining TRS with host sequences can best be visualized by pairing the TRS elements at their inverted repeat sequences (Fig. 3*C*). The striking feature of the external junctions is that the same four-base host sequence (A-A-C-G) at junction a (229–232) is found at junction d (5821–5824). A duplication at the site of integration is reminiscent of prokaryotic IS elements (31, 32), but until the site of integration in normal mink DNA is sequenced, we will not know if this duplication pre-existed or was generated by MSV integration.

The unique MSV sequences at the internal MSV–TRS junctions (Fig. 1*B*, b and c) have nearby runs of G-C base pairs (824–829 and 5226–5229). As described above, the G-C-rich sequence at position 824–829 is part of the tRNA binding site (22). The priming site for the plus strong stop DNA synthesis has not been identified. If priming occurs at the beginning of the contribution of 3′ RNA sequences to TRS (position c in minus strand DNA), we find no potential tRNA priming site. However, there is a C-C-A trinucleotide at the end of the inverted repeat region of TRS (5241–5243). It is in the proper orientation to be part of a tRNA primer for the plus strong stop DNA synthesis. Alternatively, the first 13 nucleotides of the minus strong stop DNA (810–822; including the 11-base inverted repeat) could serve to prime the plus strong stop DNA synthesis at position c. Products smaller than the minus strong stop DNA have been observed in M-MuLV provirus DNA synthesized *in vitro* (12).

Seif *et al.* (23) have proposed a general model for RNA splicing, in which the sequence 5′ A-G-G-T-A-A-G-T 3′ is a donor sequence for splicing to a receptor sequence located downstream in the primary RNA transcript. A sequence (5′ A-G-G-T-A-A-G-C 3′, 881–888) with 7 out of 8 nucleotides in common with the model occurs in the b junction sequence, 208 bases from the sequence encoding the 5′ end of viral RNA (see above). Three additional potential donor splice sites with 6 of 8 nucleotides in common with the model sequence can also be found in TRS preceding the 5′ end of viral RNA: A-G-G-T-C-A-G-G (332–339), T-G-G-T-A-A-G-C (380–387), and C-G-G-T-A-A-G-C (453–460). However, it is unknown whether any of these sites are involved in viral RNA maturation. No sequences related to the acceptor splice sequence defined by Seif *et al.* (23) were found in the ≈1500 nucleotides presented in Fig. 2.

We have found no translation initiation codons (A-T-G) in the first 227 bases of the 5′ end of the viral RNA genome. This would represent an unusually long leader sequence for eukaryotic genes, but it is consistent with the estimated placement of the amino terminal end of the defective m1 MSV polyprotein, pp60^{gag} (6).

Examination of the sequences of Mink L and Mink R (Fig. 2) reveals that both contain duplications. Beginning at positions 194 and 211, a seven-base sequence is duplicated (C-T-T-C-C-C-C), 15 bases from the mink–viral DNA junction. At positions 5825 and 5836, a nine-base sequence is duplicated (C-T-A-G-T-G-C-T-G). In both cases, the duplication is not tandem; the Mink-L duplication is separated by nine bases while the Mink-R duplication is separated by two bases.

Similarity of TRS Sequence Organization and Prokaryotic IS Elements. MSV TRS and prokaryotic IS elements have several parallel features. Both have inverted repeat sequences defining their ends. In MSV TRS this is 11 bases long, and if some mismatching is allowed, 19 more bases could be included. Since both the L- and R-TRS are directly repeated, the complete integrated m1 MSV proviral DNA has the 11-base-pair inverted repeat (Fig. 3C). Integrated m1 MSV resembles the Tn 9 transposon that confers chloramphenicol resistance in *E. coli* (30, 33) in that unique DNA is bracketed by directly repeated IS1 elements (34).

IS elements often have effects on the expression of adjacent sequences. Some are thought to contain promoter sequences that in the proper orientation cause transcription of adjacent DNA. Many IS elements also contain transcription terminator signals that block transcription passing through them (see ref. 34 for review). Since MSV TRS contains a retrovirus promoter as well as putative polyadenylylation signals, it may have transcription and termination control properties in eukaryotic cells.

The TRS sequence may respond in *E. coli* like certain IS elements. Several groups have reported sequence rearrangements involving TRS after growth of retrovirus–phage recombinants in *E. coli* (refs. 6, 7, and 35; unpublished data). A deletion similar to the "*cam*-cut out" of Tn 9 has been noted by us (unpublished data).

Perhaps TRS functions in eukaryotes like IS elements in prokaryotes. The putative duplication of host DNA at the site of m1 MSV integration is a characteristic of prokaryote IS element integration. Similarly, the MSV deletions with one end at, or near, TRS described by Tronick *et al.* (36) are like IS mediated deletions in *E. coli* (34). The transposition of sequences bracketed by repeated elements in yeast (37) and Drosophila (38) indicates that DNA rearrangements mediated by directly repeated sequences occur in eukaryotes. In addition, it has been proposed that inverted repeat sequences are involved in the joining of genetic elements that lead to antibody diversity (39). We have demonstrated that TRS is essential for the efficient transformation of cells by MSV *src* sequences (11). In this case, the TRS may provide elements of transcriptional control or facilitate integration. It remains to be seen if MSV integration uses TRS as an IS-like element and if rearrangements occur in eukaryotic cells mediated by TRS.

We thank M. Oskarsson for helpful discussions and review of this manuscript. We thank J. V. Maisel for performing computer analysis on the sequence which identified the 200-base-pair overlapping duplication in TRS and the RNA splice sites. We are grateful to W. A. Haseltine for providing us with a prepublication copy of the M-MuLV minus strong stop DNA sequence. We are indebted to L. Quinn, B. Nolte, and P. Yasem for their assistance in preparing this manuscript.

1. Hsu, T. W., Sabran, J. L., Mark, G. E., Guntaka, R. V. & Taylor, J. M. (1978) *J. Virol.* **28,** 810–818.
2. Shank, P. R., Hughes, S. H., Kung, H.-J., Majors, J. E., Quintrell, N., Guntaka, R. V., Bishop, J. M. & Varmus, H. E. (1978) *Cell* **15,** 1383–1395.
3. Hughes, S. H., Shank, P. R., Spector, D., Kung, H.-J., Bishop, J. M., Varmus, H. E., Vogt, P. K. & Breitman, M. L. (1978) *Cell* **15,** 1397–1410.
4. Gilboa, E., Goff, S., Shields, A., Yoshimura, F., Mita, S. & Baltimore, D. (1979) *Cell* **16,** 863–874.
5. Benz, E. W. & Dina, D. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3294–3298.
6. Vande Woude, G. F., Oskarsson, M. K., McClements, W. L., Enquist, L. W., Blair, D. G., Fischinger, P. J., Maizel, J. V. & Sullivan, M. (1979) *Cold Spring Harbor Symp. Quant. Biol.* **44,** in press.
7. Vande Woude, G. F., Oskarrson, M. K., Enquist, L. W., Nomura, S., Sullivan, M. & Fischinger, P. J. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4464–4468.
8. Gilboa, E., Mitra, S. W., Goff, S. & Baltimore, D. (1979) *Cell* **18,** 93–100.
9. Mitra, S. W., Goff, S., Gilboa, E., and Baltimore, D. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4355–4359.
10. Dina, D. & Benz, E. W. (1980) *J. Virol.* **33,** 377–389.
11. Blair, D. G., McClements, W. L., Oskarsson, M. K., Fischinger, P. J. & Vande Woude, G. F. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3504–3508.
12. Haseltine, W. A., Kleid, D. G., Panet, A., Rothenberg, E. & Baltimore, D. (1976) *J. Mol. Biol.* **106,** 109–131.
13. Taylor, J. M. (1977) *Biochim. Biophys. Acta* **473,** 57–71.
14. Varmus, H. E., Heasley, S., Kung, H.-J., Opperman, H., Smith, V. C., Bishop, J. M. & Shank, P. R. (1978) *J. Mol. Biol.* **120,** 55–82.
15. Coffin, J. M., Hagerman, T. C., Maxam, A. M. & Haseltine, W. A. (1978) *Cell* **13,** 761–773.
16. Haseltine, W. A., Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 989–993.
17. Schwartz, D. E., Zamecnik, P. C. & Weith, H. L. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 994–998.
18. Shine, J., Czernilovsky, A. P., Friedrich, R., Bishop, J. & Goodman, H. M. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 1473–1477.
19. Stoll, E., Billeter, M. A., Palmenberg, A. & Weissmann, C. (1977) *Cell* **12,** 57–72.
20. Maxam, A. H. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
21. Maat, J. & Smith, A. J. H. (1978) *Nucleic Acids Res.* **5,** 4537–4545.
22. Peters, G. & Dahlberg, J. E. (1979) *J. Virol.* **31,** 398–407.
23. Seif, I., Khoury, G. & Dhar, R. (1979) *Cell* **18,** 963–977.
24. Subramanian, K. H., Reddy, V. B. & Weissman, S. M. (1977) *Cell* **10,** 497–507.
25. Rose, J. K., Haseltine, W. A. & Baltimore, D. (1976) *J. Virol.* **20,** 324–329.
26. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 784–788.
27. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8,** 127–142.
28. Proudfoot, N. J. & Brownlee, G. G. (1974) *Nature (London)* **252,** 359–362.
29. Shapiro, J. A. & MacHattie, L. A. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43,** 1135–1142.
30. Shapiro, J. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1933–1937.
31. Calos, M. P., Johnsrud, L. & Miller, J. (1978) *Cell* **13,** 411–418.
32. Grindley, N. D. F. (1978) *Cell* **13,** 419–426.
33. MacHattie, L. A. & Jackowski, J. B. (1977) in *DNA Insertion Elements, Plasmids, and Episomes,* eds. Bukhari, A. I., Shapiro, J. A. & Adhya, S. L. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 219–228.
34. Bukhari, A. I., Shapiro, J. A. & Adhya, S., eds. (1977) *DNA Insertion Elements, Plasmids, and Episomes* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
35. Chan, H. W., Garon, C. F., Chang, E. H., Lowy, D. R., Hager, G. L., Scolnick, E. M., Repaske, R. & Martin, M. A. (1980) *J. Virol.* **33,** 845–855.
36. Tronick, S. R., Robbins, K. C., Canaani, E., Devare, S. G., Anderson, P. R. & Aaronson, S. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 6314–6318.
37. Cameron, J. R., Loh, E. Y. & Davis, R. W. (1979) *Cell* **16,** 739–751.
38. Potter, S. S., Brorein, W. J., Dunsmuir, P. & Rubin, G. M. (1979) *Cell* **17,** 415–427.
39. Seidman, J. G., Max, E. E. & Leder, P. (1979) *Nature (London)* **280,** 370–375.