Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions

H.R.Chen and W.C.Barker

National Biomedical Research Foundation, Georgetown University Medical Center, 3900 Reservoir Road, NW, Washington, DC 20007, USA

Received 7 November 1983; Accepted 16 January 1984

ABSTRACT

The nucleotide sequences of the LTRs and their adjacent regions from 19 type C and one type B retrovirus were compared. Salient features are: (a) The R regions in the genomes of most of the type C retroviruses begin with GC and end with CA. (b) The mammalian type C retroviruses have a polyadenylation signal "AATAAA" in the R region, and most have a "CAT" box and a "TATA" box in the U3 region. (c) The avian type C retroviruses have an AATAAA sequence, and some also have "CAT-like" and "TATA-like" boxes, in the U3 region. (d) As with many transposable elements, the IR regions of the proviruses begin with TG and end with CA, and the DR sequences in the host genomes flanking the proviruses are different from one another. Although SNV is an avian retrovirus, the nucleotide sequences in the R, U5, TBS, and PU region are more similar to the mammalian type C than to the avian type C retroviruses.

INTRODUCTION

The molecular structure of the retroviral genome has been well defined (1). Its genomic RNA has a short repeat (R) at the ends of the sequence (Fig. 1). The R at the 5' end is followed by a unique sequence called U5, a tRNA-binding site (TBS), and then a noncoding sequence. The middle portion of the genome consists of the coding regions, which may contain up to four genes: gag, pol, env, and onc. After the 3' end of the coding region is a noncoding region, a purine-rich sequence (PU), a unique sequence called U3, and the final R.

The genome of a retrovirus is replicated in the host cell by way of a DNA intermediate (2). The transcribed linear double-stranded DNA molecule is longer than the corresponding genomic RNA (Fig. 1) because of the addition of a U3 and a U5 at the 5' and the 3' ends of the DNA, respectively. The combination of U3, R, and U5 is called a long terminal repeat (LTR). The linear DNA becomes closed circular DNA with one or two copies of the LTR before it is integrated into the host genome. The integrated viral DNA is called a provirus. As with a transposable element (3), the sequence at each host-proviral junction consists of an inverted repeat (IR) at the end of the

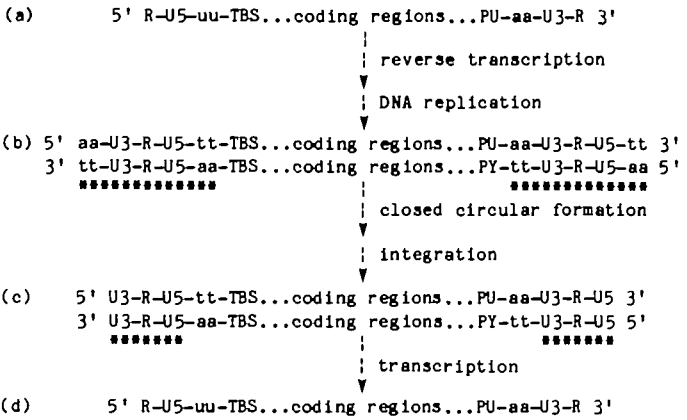

Fig. 1. Molecular structures of the retroviral genome during various stages of development. (a) The genome in the virion. The dinucleotides UU (uu) and AA (aa) are considered to be part of the U5 and the U3, respectively. The dotted lines flanking the coding regions represent noncoding regions. (b) The unintegrated double-stranded retroviral DNA, which becomes closed circular DNA with one or two copies of the LTR. (c) Proviral DNA in the host genome; notice that the dinucleotides TT (tt) and AA (aa) at the ends of the DNA are missing. Finally, (d) the viral genome is transcribed from the proviral sequence. Asterisks mark the LTR.

provirus and a direct repeat (DR) in the host genome flanking the provirus. Note that the dinucleotides, usually TT and AA, at the ends of the unintegrated double-stranded DNA are missing in the proviral sequence.

Retroviruses, which belong to the family Retroviridae and subfamily Oncovirinae (4), are classified into four types: A, B, C, and D. Many nucleotide sequences, mostly from type C retroviruses, have been determined (Table 1). Most of the sequenced avian type C retroviruses belong to the avian leukemia-sarcoma virus (ALSV) group, which includes RSV, Y73, FSV, AMV, and AMC29. The other avian retrovirus, SNV, is a reticuloendotheliosis virus and has been shown to be more closely related to the mammalian type C retroviruses than to the ALSV group (45). Among the mammalian type C are eight murine (MMLV, MMSV, AMLV, FBJ, AKR, FSFFV, BKMV, and RMLV), one feline (FESV), three primate (SSV, GALV, and BaEV), and one human retrovirus (ATLV). All of these except BaEV and ATLV show marked similarities in the sequences of their LTR. MMTV is the only type B retrovirus. This paper analyzes a collection of these nucleotide sequences and compares related sequences at the ends of the viral genomes and the host sequences at the host-viral junctions.

MATERIALS AND METHODS

The National Biomedical Research Foundation maintains a nucleic acid sequence database (5-9) that contained 1,123 entries with 1,593,561 bases as of July 20, 1983. The database contains nine completely sequenced retroviral genomes and fragments of many others (Table 1).

Computer programs ADDSEQ, ALIGN, and DISP were used to compare the sequences and to produce the alignments. Program ADDSEQ adds a new sequence to an existing alignment, inserting gaps where needed to align homologous regions of the sequences. Program ALIGN determines the best alignment of two sequences by computing the maximum match score (10), and program DISP displays alignments of related sequences. Only the homologous sequences are shown in this paper. The heterogeneous sequences, such as those in the LTR of SNV, BaEV, ATLV, and MMTV, are not shown here.

RESULTS AND DISCUSSION

The R Region

Avian Retroviruses R regions in the genomes of the ALSV group are 21 bases long (Table 1), and their sequences are highly conserved (Fig. 2a). SNV, on the other hand, has an R region 82 bases long and its sequence is more similar to those of the mammalian retroviruses discussed below. Like most of the mammalian type C retroviruses, the R region of the SNV contains the sequence "AATAAA" near the 3' end. However, no such polyadenylation signal is found in the R regions of the ALSV group.

Mammalian Retroviruses The R regions in the genomes of the mammalian type C retroviruses are longer than those of the ALSV group: 64-69 bases in the murine, the feline, and the primate retroviruses and 228 in the human retrovirus (ATLV) (Table 1). With the exception of BaEV and ATLV, portions of the sequences in the R regions of the mammalian type C retroviruses are conserved (Fig. 2b), and they contain an AATAAA sequence about 17 bases from the poly(A) site.

It should be noted that the R regions in the genomes of most avian and mammalian type C retroviruses begin with GC and end with CA.

Type B Retroviruses The R region of MMTV is only 11 bases long (Table 1), and its sequence does not contain the standard AATAAA sequence.

The U5 Region

Avian Retroviruses The U5 regions of the ALSV group are 80 bases long (Table 1), and their sequences are also highly conserved (Fig. 3a). Once

Table 1. Probable lengths of various regions in retroviral LTRs and probable lengths of their adjacent regions

	R	U5	U3	TBS	PU	IR	DR	Ref.
(1) Type C retroviruses								
Avian retroviruses								
RSV*	21	80	234	18	11	15	-	11
Y73*	21	80	215	18	11	15	-	12
FSV*	21	80	246	18	11	9	-	13
AMV	21	80	288	-	11	17	6	14-16
AMC29	21	80	216	18	-	11	-	17
SNV	82	97	420	18	13	9	5	18-20
Mammalian retroviruses								
MMLV*	68	77	449	23	13	13	-	21,22
MMSV*	69	76	444	23	13	13	4	23-25
AMLV*	68	77	449	23	12	13	4	26
FBJ*	68	76	475	23	13	13	4	27
AKR	68	76	482	-	14	13	4	28,29
FSFFV	69	76	373	23	13	13	4	30,31
BKMV	68	76	385	-	-	13	-	32
RMLV	64	76	-	-	-	-	-	33
FESV	68	76	342	18	-	14	-	34
SSV*	69	77	362	18	14	9	4	35,36
GALV	66	77	-	-	-	-	-	37
BaEV	66	67	421	23	-	-	-	38,39
ATLV*	228	176	355	18	12	4	6	40
(2) Type B retrovirus								
MMTV	11	124	1197	18	19	8	6	41-44

The lengths of U5, U3, and IR include the dinucleotide AA or TT at their ends, and the lengths of these regions at the ends of the proviruses should be two nucleotides shorter than that shown. IR, TBS, and PU may not be perfect.

Abbreviations: RSV, Rous sarcoma virus strain Prague C; Y73, avian sarcoma virus Y73; FSV, Fujinami sarcoma virus; AMV, avian myeloblastosis virus; AMC29, avian myelocytomatosis virus MC29; SNV, avian spleen necrosis virus; MMLV, Moloney murine leukemia virus; MMSV, Moloney murine sarcoma virus; AMLV, Abelson murine leukemia virus; FBJ, FBJ murine osteosarcoma virus; AKR, AKR murine leukemia virus; FSFFV, Friend spleen focus forming virus; BKMV, BL/Ka(B) murine nonleukemogenic virus; RMLV, Rauscher murine leukemia virus; FESV, feline sarcoma virus; SSV, simian sarcoma virus; GALV, gibbon ape leukemia virus strain GaLV(SF); BaEV, baboon endogenous virus; ATLV, human adult T-cell leukemia virus; MMTV, mouse mammary tumor virus.

*The entire genome has been completely sequenced.

again, the U5 of SNV is slightly longer (97 bases), and its sequence is very different from those of the ALSV group. It can be aligned with those of the mammalian type C retrovirus, but the homology is not significant when tested

```
(a) Avian retroviruses

RSV         GCCATTTTACCATTCACCACA
Y73         GCCATTTTACCTCCCACCACA
FSV         GCCATTTTACCATTCACCACA
AMV         GCCATTCTACCTCTCACCACA
AMC29       GCCATTTTACCATCCACCACA

Conserved   GCCATTTTACCATTCACCACA
            ''''''' ''''   '''''''

(b) Mammalian retroviruses

MMLV        GCGCCAGTCCTCCGATTGACTGAGTCG-CCCGGGTACCCGTGTATCCAATAAACCCTCTTGCAG-TTGCA
MMSV        GCGCCAGTCTTCCGATAGACTGCGTCG-CCCGGGTACCCGTATTCCCAATAAAGCCTCTTGCTGTTTGCA
AMLV        GCGCCAGTCCTCCGAGTGACTGAGTCG-CCCGGGTACCCGTGTATCCAATAAACCCTCTTGCAG-TTGCA
FBJ         GCGCCAGTCCTCCGATAGACTGAGTCG-CCCGGGTACCCGTGTATCCAATAAAGCCTTTTGCTG-TTGCA
AKR         GCGCCAGTCCTCCGATAGACTGAGTCG-CCCGGGTACCCGTGTATCCAATAAAGCCTTTTGCTG-TTGCA
FSFFV       GCGCCAGTCCTCCGACAGACTGAGTCG-CCCGGGTACCCGTGTTCCCAATAAAGCCTCTTGCTGATTGCA
BKMV        GCGCCAGTCCTCCGACAGACTGAGTCG-CCCGGGTACCCGTGTATCCAATAAAGCCTTTTGCTG-TTGCA
RMLV        GC-C--G-CCTCCGACAGCCTGAGTCG-CCCGGGTACCCGTGTATCCAATAAATCCTCTTGCTG-TTGCA
FESV        GCGCAAGTCTTTGCTGAGACTTGACCGCCCCGGGTACCCGTGTA-CGAATAAA-CCTCTTGCTGTTTGCA
SSV         GCGCCAGTCCTCCGAGAGACTGAGTCG-CCCGGGTACCTGTGTGTTCAATAAAACCTCTTGCTATTTGCA
GALV        CC---AGTCTTCCAAGAGACTGAGTCG-CCCGGGTACCCGTGTGATCAATAAAACCTCTTGCTACTTGCA

Conserved GCGCCAGTCCTCCGA AGACTGAGTCG CCCGGGTACCCGTGTATCCAATAAA CCTCTTGCTG TTGCA
          '   ' ' '     ' ''   '' '''''''''''' '' '   ''''''' ''' ''''  '''''
                                                    ---------
```

Fig. 2. Alignment of nucleotide sequences from the R regions of the LTR. Dashes indicate gaps inserted to align homologous regions of the sequences. The polyadenylation signal is underlined.

with program ALIGN.

Mammalian Retroviruses The U5 regions of most mammalian type C retroviruses are 76-77 bases long (Table 1), and portions of their sequences are conserved (Fig. 3b). BaEV and ATLV, on the other hand, have U5 regions 67 and 176 bases long, respectively, and their sequences are different.

Type B Retroviruses The U5 region in the genome of MMTV is 124 bases long (Table 1), and its sequence is different from that of the type C retroviruses.

Neither the promoter "TATA" box nor the polyadenylation signal AATAAA is found in the U5 region of any of the above retroviruses.

The U3 Region

Avian Retroviruses In contrast to the R and the U5 regions, the lengths of the U3 regions in the genomes of avian type C retroviruses are more variable (from 215 bases in Y73 to 420 bases in SNV) (Table 1). Length variations are also found in different strains of the SNV (361 to 420 bases). Despite the fact that these sequences are very heterogeneous, the U3 regions of some of the ALSV sequences contain "CAT-like" and "TATA-like" sequences. The AATAAA polyadenylation signal is found at the 3' end (Fig. 4a).

Mammalian Retroviruses The lengths of the U3 regions in the genomes of

```
(a) Avian retroviruses

RSV         TTGGTGTGCACCTGGGTTGATGGCCGGACCGTCGATTCCCTAACGATTGCGAACACCTGAATGAAGCAGAAGGCTTCA(TT)
Y73         TTGGTGTGCACCTAGGTTGATGGCCGGACCGTCGATTCCCTGACGACTACGAGCACCTGAATGAAGCAGAAGGCTTCA(TT)
FSV         TTGGTGTGCACCTGGGTAGATGGACAGACCGTTGAGTCCCTAACGATTGCGAACACCTGAATGAAGCGGAAGGCTTCA(TT)
AMV         TTGGTGTGCACCTGGGTTGATCGCCGGACCGTCGATTCCCTGACGACTGCGAACACCTGAATGAAGCTGAAGGCTTCA(TT)
AMC29       TTGGTCTGCACCTGGGTAGATGGACAGACCGTTGAGTCCCTAACGATTACGCGAACCTGAATGAAGCAGAAGGCTTCA(TT)

Conserved   TTGGTGTGCACCTGGGTTGATGGCCGGACCGTCGATTCCCTAACGATTGCGAACACCTGAATGAAGCAGAAGGCTTCA(TT)
            ***** ******* *** *** * * ****** ** ***** **** * **  ************** ********** **

(b) Mammalian retroviruses

MMLV        TCCGACT-TGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCGTCAGCGGGGGTCTTTCA(TT)
MMSV        TCCGAAT-CGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCACGA-CGGGGGTCTTTCA(TT)
AMLV        TCCGACT-TGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCGTCAGCGGGGGTCTTTCA(TT)
FBJ         TCCGAAT-CGTGGTCTCGCTGATCCTTGGGAGGGTCTCCTCAGAGTGATTGACTGCCCAGCC-TGGGGGTCTTTCA(TT)
AKR         TCCGAAT-CGTGGTCTCGCTGATCCTTGGGAGGGTCTCCTCAGAGTGATTGACTGCCCAGCC-TGGGGGTCTTTCA(TT)
FSFFV       TCCGAAT-CGTGGACTCGCTGATCCTTGGGAGGGTCTCCTCAGATTGATTGACTGCCCACCT-CGGGGGTCTTTCA(TT)
BKMV        TCCGAAT-CGTGGTCTCGCTGATCCTTGGGAGGGTCTCCTCAGAGTGATTGACTGCCCAGCT-TGGGGGTCTTTCA(TT)
RMLV        TCCGACT-TGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCAGAGTGATTGACTACCCGTCT-CGGGGGTCTTTCA(TT)
FESV        TCTGACT-CGTGGTCTCGGTGTTCCGTGGGTACGGGGTCTCATCGCCGAGGAAGACCTAATT-CGGGGGTCTTTCA(TT)
SSV         TCCGAAGCCGTGGTCTCGTTGTTCCTTGGGAGGGTCTCTCCTAACTGATTGACTGCCCACCT-CGGGGGTCTCTCA(TT)
GALV        TCCGAAGTCGTGGTCTCGCTGTTCCTTGGGAAGGTCTCCCCTAATTGATTGACCGCCCGGAC-TGGGGGTCTCTCA(TT)

Conserved   TCCGAAT CGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCAGAGTGATTGACTGCCCA C  CGGGGGTCTTTCA(TT)
            ** **    **** **** ** *** ****   *      *          **   **       ******** *** **
```

Fig. 3. Alignment of nucleotide sequences in the U5 regions of the LTR. The dinucleotide TT inside the parentheses at the 3' end is missing in the provirus. Dashes indicate gaps inserted to align homologous regions of the sequence.

the mammalian type C retroviruses vary from 342 bases in FESV to 482 bases in the AKR. As in the avian retroviruses, the sequences in the U3 regions of the mammalian type C retroviruses are heterogeneous. However, only a "CAT" box and a TATA box, but not the polyadenylation signal AATAAA, are found near the 3' end of the U3 regions of most mammalian type C retroviruses (Fig. 4b). In addition, the U3 regions of most murine retroviruses contain one or two copies of a 75-base repeat.

Type B Retroviruses The U3 region of the MMTV (1197 bases) is much longer than that of the type C retroviruses. Interestingly, its sequence contains a coding region for an unidentified protein varying in length from 198 to 324 amino acids in different strains of the virus.

The TBS Region

The sequence in the TBS region of the retroviral genome is complementary to the 3'-terminal sequence of a tRNA. It has been proposed that tRNA, upon binding to the TBS region by base pairing, serves as an initiation site for the minus DNA strand synthesis (46).

The TBS regions of the murine and BaEV retroviruses are 23 bases long; those of the other retroviruses, including even the type B retrovirus, are 18

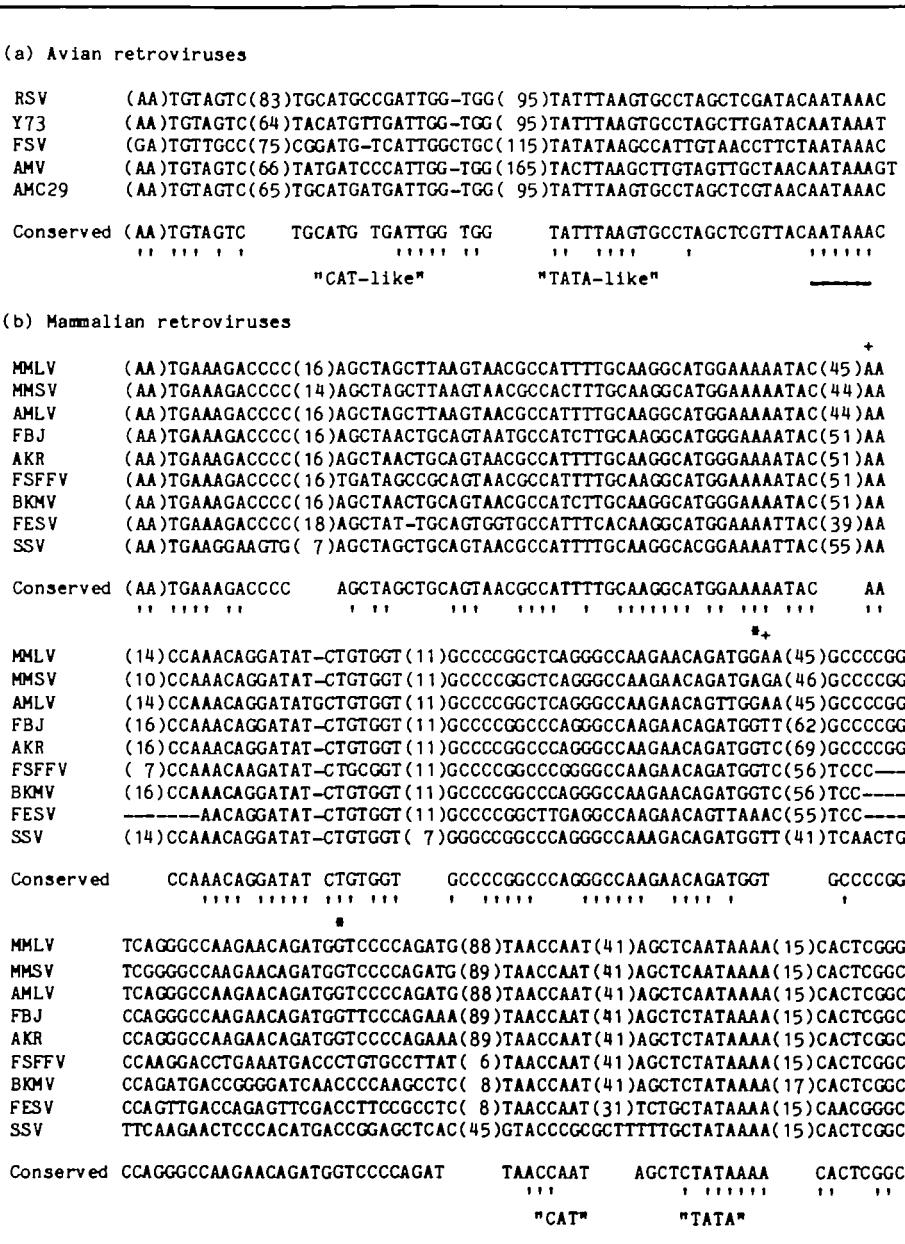

Fig. 4. Alignment of nucleotide sequences in the U3 regions of the LTR. Numbers in parentheses indicate the number of bases not shown. The dinucleotides in parentheses at the 5' end are missing in the provirus. Dashes indicate gaps inserted to align homologous regions of the sequences. The polyadenylation signal is underlined. There are two 75-base repeats, the beginnings and ends of which are shown by the symbols + and *, respectively.

```
(1) Type C retroviruses

 (a) Avian retroviruses          (b) Mammalian and related retroviruses

RSV          TGGTGACCCCGACGTGAT      MMLV       TGGGGGCTCGTCCGGGATCGGGA
Y73          TGGTGACCCCGACGTGAT      MMSV       TGGGGGCTCGTCCGGGATTTGGA
FSV          TGGTGACCCCGACGTGAT      AMLV       TGGGGGCTCGTCCGGGATCGGGA
AMC29        TGGTGACCCCGACGTGAT      FBJ        TGGGGGCTCGTCCGGGATTCGGA
                                     FSFFV      TGGGGGCTCGTCCGGGATTTGGA
                                     FESV       TGGGGGCTCGTCCGGGAT
Conserved    TGGTGACCCCGACGTGAT      BaEV       TGGGGGCTCGTCCGGGATTTGAG
             ::::::::::::::::::      SSV        TGGGGGCTCGTCCGGGAT
Trp tRNA 3' ACCACTGGGGCTGCACTA 5'    ATLV       TGGGGGCTCGTCCGGGAT
                                     SNV        TGGGGGCTCGTCCGGGAT

                                     Conserved TGGGGGCTCGTCCGGGAT  GGA
                                               ::::::::::::::::::  :::
                                  Pro tRNA 3' ACCCCCGAGCAGGCCCTAAACTT 5'


(2) Type B retrovirus

MMTV         TGGCGCCCGAACAGGGAC
             ::::::::::::::::::
Lys tRNA 3' ACCGCGGGCTTGTCCCTG 5'
```

Fig. 5. Alignment of nucleotide sequences in the TBS regions. Base pairings of tRNA to the TBS region are shown.

bases long (Table 1). The sequences in the TBS regions of the ALSV group are identical; each binds a Trp tRNA by base pairing (47) (Fig. 5). Sequence homology is also found in the TBS of five murine, one feline, two primate (SSV and BaEV), one human (ATLV), and one avian retrovirus (SNV), and the sequence binds a Pro tRNA (48). On the other hand, the sequence in the TBS region of the type B retrovirus (MMTV) is different from those of the type C, and it binds a Lys tRNA (49).

The PU Region

The PU region in the retroviral genome has been speculated to be an

```
(a) Avian retroviruses          (b) Mammalian and related
                                    retroviruses

RSV          AGGGAGGGGGA        MMLV       AGAAAAAGGGGGG
Y73          AGGGAGGGGGA        MMSV       AGAAAAAGGGGGG
FSV          AGGGAGGGGGA        AMLV       AGAAAAAGGGGGG
AMV          AGGGAGGGGGA        FBJ        AGAAAGAGGGGGG
                                AKR        AGAAAGAGGGGGGG
Conserved    AGGGAGGGGGA        FSFFV      AGAAAAAGGGGGG
             ''''''''''''       SSV        AAGAGAAATGGGGG
                                ATLV        GAAAAAGAGGCA
                                SNV        AAGAGCAGTGGGG

                                Conserved AGAAAAAGGGGGG
                                            '  '  ''
```

Fig. 6. Alignment of nucleotide sequences in the PU regions.

Fig. 7. Host-proviral junctions. DR regions in the host sequences are underlined; complementary nucleotides in IR regions of the proviruses are marked with arrowheads.

initiation site for the plus DNA strand synthesis (46). The PU regions in the genomes of the ALSV group are 11 bases long (Table 1), and their sequences are identical (Fig. 6a). On the other hand, SNV and mammalian type C retroviruses have a PU region varying from 12 to 14 bases (Table 1), and the sequences are partially conserved (Fig. 6b). MMTV has 19 bases in the PU region, and its sequence is different from those of the type C retroviruses.

Host-Viral Junctions

As mentioned above, the host-viral junction is characterized by two short sequences: IR in the provirus and DR in the host genome.

The lengths of the IR regions in the unintegrated retroviral DNA vary from 4 bases in the ATLV to 17 bases in the AMV (Table 1), and those in the integrated proviral DNA are two bases shorter (Fig. 1) because the terminal dinucleotide TT or AA is lost during viral integration. Some of the IR sequences, however, are not perfectly inverted (Fig. 7). As with many transposable elements (50), the IR of all integrated proviral DNA begins with TG and ends with CA.

Although all the DR regions are 4-6 bases long (Table 1), their sequences are different from one another (Fig. 7). Even for the same proviruses different DR sequences have been found in the host genome. The implication is that the site for viral integration may not be sequence-specific, and a retrovirus may be integrated into the host genome at many different locations (51).

ACKNOWLEDGMENTS

We thank Dr. L.T. Hunt and Bobby Baum for helpful comments and suggestions and Margaret C. Blomquist for editorial assistance. This work was supported by NIH grant GM08710.

REFERENCES

1. Temin, H.M. (1981) Cell 27, 1-3.
2. Varmus, H.E. (1982) Science 216, 812-820.
3. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.
4. Matthews, R.E.F. (1982) Classification and Nomenclature of Viruses, Karger, Basel, 199pp.
5. Dayhoff, M.O., Schwartz, R.M., Chen, H.R., Barker, W.C., Hunt, L.T. and Orcutt, B.C. (1981) DNA 1, 51-58.
6. Chen, H.R., Dayhoff, M.O., Barker, W.C., Hunt, L.T., Yeh, L.-S., George, D.G. and Orcutt, B.C. (1982) DNA 1, 103-108.
7. Chen, H.R., Dayhoff, M.O., Barker, W.C., Hunt, L.T., Yeh, L.-S., George, D.G. and Orcutt, B.C. (1982) DNA 1, 273-307.
8. Chen, H.R., Dayhoff, M.O., Barker, W.C., Hunt, L.T., Yeh, L.-S., George, D.G. and Orcutt, B.C. (1982) DNA 1, 365-374.
9. Chen, H.R., Dayhoff, M.O., Barker, W.C., Hunt, L.T., Yeh, L.-S., George,

D.G. and Orcutt, B.C. (1983) DNA 2, 273-278.
10. Dayhoff, M.O. (1979) in Atlas of Protein Sequence and Structure, Dayhoff, M.O., ed., Vol. 5, Suppl. 3, pp.1-8, National Biomedical Research Foundation, Washington, D.C.
11. Schwartz, D., Tizard, R. and Gilbert, W. (1983) Cell 32, 853-869.
12. Kitamura, N., Kitamura, A., Toyoshima, K., Hirayama, Y. and Yoshida M. (1982) Nature 297, 205-208.
13. Shibuya, M. and Hanafusa, H. (1982) Cell 30, 787-795.
14. Rushlow, K.E., Lautenberger, J.A., Papas, T.S., Baluda, M.A., Perbal, B., Chirikjian, J.G. and Reddy, E.P. (1982) Science 216, 1421-1423.
15. Rushlow, K.E., Lautenberger, J.A., Reddy, E.P., Souza, L.M., Baluda, M.A., Chirikjian, J.G., and Papas, T.S. (1982) J. Virol. 42, 840-846.
16. Klempnauer, K.-H., Gonda, T.J. and Bishop, J.M. (1982) Cell 31, 453-463.
17. Reddy, E.P., Reynolds, R.K., Watson, D.K., Schultz, R.A., Lautenberger, J. and Papas, T.S. (1983) Proc. Natl. Acad. Sci. USA 80, 2500-2504.
18. Shimotohno, K., Mizutani, S. and Temin, H.M. (1980) Nature 285, 550-554.
19. Shimotohno, K. and Temin, H.M. (1982) J. Virol. 41, 163-171.
20. O'Rear, J.J. and Temin, H.M. (1982) Proc. Natl. Acad. Sci. USA 79, 1230-1234.
21. Shinnick, T.M., Lerner, R.A. and Sutcliffe, J.G. (1981) Nature 293, 543-548.
22. Van Beveren, C., Goddard, J.G., Berns, A. and Verma, I.M. (1980) Proc. Natl. Acad. Sci. USA 77, 3307-3311.
23. Van Beveren, C., Van Straaten, F., Galleshaw, J.A. and Verma, I.M. (1981) Cell 27, 97-108.
24. Reddy, E.P., Smith, M.J. and Aaronson, S.A. (1981) Science 214, 445-450.
25. Dhar, R., McClements, W.L., Enquist, L.W. and Van de Woude, G.F. (1980) Proc. Natl. Acad. Sci. USA 77, 3937-3941.
26. Reddy, E.P., Smith, M.J. and Srinivasan, A. (1983) Proc. Natl. Acad. Sci. USA 80, 3623-3627.
27. Van Beveren, C., van Straaten, F., Curran, T., Muller, R. and Verma, I.M. (1983) Cell 32, 1241-1255.
28. Herr, W., Corbin, V. and Gilbert, W. (1982) Nucl. Acids Res. 10, 6931-6944.
29. Van Beveren, C., Rands, E., Chattopadhyay, S.K., Lowy, D.R. and Verma, I.M. (1982) J. Virol. 41, 542-556.
30. Clark, S.P. and Mak, T.W. (1982) Nucl. Acids Res. 10, 3315-3330.
31. Amanuma, H., Katori, A., Obata, M., Sagata, N. and Ikawa, Y. (1983) Proc. Natl. Acad. Sci. USA 80, 3913-3917.
32. Kim, J.P., Kaplan, H.S. and Fry, K.E. (1982) J. Virol. 44, 217-225.
33. Lovinger, G.G. and Schochetman, G. (1979) J. Virol. 32, 803-811.
34. Hampe, A., Gobet, M., Even, J., Sherr, C.J. and Galibert, F. (1983) J. Virol. 45, 466-472.
35. Devare, S.G., Reddy, E.P., Robbins, K.C., Andersen, P.R., Tronick, S.R. and Aaronson, S.A. (1982) Proc. Natl. Acad. Sci. USA 79, 3179-3182.
36. Devare, S.G., Reddy, E.P., Law, J.D., Robbins, K.C. and Aaronson, S.A. (1983) Proc. Natl. Acad. Sci. USA 80, 731-735.
37. Scott, M.L., McKereghan, K., Kaplan, H.S. and Fry, K.E. (1981) Proc. Natl. Acad. Sci. USA 78, 4213-4217.
38. Tamura, T.-A., Noda, M. and Takano, T. (1981) Nucl. Acids Res. 9, 6615-6626.
39. Tamura, T.-A. (1983) J. Virol. 47, 137-145.
40. Seiki, M., Hattori, S., Hirayama, Y. and Yoshida, M. (1983) Proc. Natl. Acad. Sci. USA 80, 3618-3622.
41. Kennedy, N., Knedlitschek, G., Groner, B., Hynes, N.E., Herrlich, P., Michalides, R. and van Ooyen, A.J.J. (1982) Nature 295, 622-624.
42. Donehower, L.A., Huang, A.L. and Hager, G.L. (1981) J. Virol. 37, 226-238.

43. Fasel, N., Pearson, K., Buetti, E. and Diggelmann, H. (1982) EMBO J. 1, 3-7.
44. Majors, J.E. and Varmus, H.E. (1981) Nature 289, 253-258.
45. Bauer, G. and Temin, H.M. (1980) J. Virol. 34, 168-177.
46. Gilboa, E., Mitra, S.W., Goff, S. and Baltimore, D. (1979) Cell 18, 93-100.
47. Cordell, B., Swanstrom, R., Goodman, H.M. and Bishop, J.M. (1979) J. Biol. Chem. 254, 1866-1874.
48. Harada, F., Peters, G.G. and Dahlberg, J.E. (1979) J. Biol. Chem. 21, 10979-10985.
49. Peters, G.G. and Glover, C. (1980) J. Virol. 35, 31-40.
50. Hishinuma, F., DeBona, P.J., Astrin, S. and Skalka, A.M. (1981) Cell 23, 155-164.
51. Fitts, R. and Temin, H.M. (1983) J. Gen. Virol. 64, 267-274.