JOURNAL OF VIROLOGY, Feb. 1986, p. 656–669
0022-538X/86/020656-14$02.00/0
Copyright © 1986, American Society for Microbiology
Vol. 57, No. 2

DNA Sequence of the Lymphotropic Variant of Minute Virus of Mice, MVM(i), and Comparison with the DNA Sequence of the Fibrotropic Prototype Strain

CAROLINE R. ASTELL,[1]* EDITH M. GARDINER,[2] AND PETER TATTERSALL[2,3]

Department of Biochemistry, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia V6T 1W5, Canada,[1] and Departments of Laboratory Medicine[3] and Human Genetics,[2] Yale University, New Haven, Connecticut 06510

Received 9 August 1985/Accepted 9 October 1985

The sequence of molecular clones of the genome of MVM(i), a lymphotropic variant of minute virus of mice, was determined and compared with that of MVM(p), the fibrotropic prototype strain. At the nucleotide level there are 163 base changes: 129 transitions and 34 transversions. Most nucleotide changes are silent, with only 27 amino acids changes predicted, of which 22 are conservative. Notable differences between the MVM(i) and MVM(p) genomes which may account for the cell specificities of these viruses occur within the 3′ nontranslated regions. The differences discussed include the absence of a 65-base-pair direct repeat in MVM(i), the presence of only two polyadenylation sites in MVM(i) compared with four in MVM(p), and sequences that bear a resemblance to enhancer sequences. Also included in this paper is an important correction to the MVM(p) sequence (C. R. Astell, M. Thomson, M. Merchlinsky, and D. C. Ward, Nucleic Acids Res. 11:999–1018, 1983).

Autonomous parvoviruses are small, icosahedral animal viruses that contain a linear, single-stranded DNA genome (57). Replication of these viruses requires a host factor expressed during the S phase of the cell cycle (43, 50, 53, 58). In addition, these viruses have a requirement for a host-cell differentiated state (32, 33, 54). For several parvoviruses, individual strains of the same serotype can show both different pathological spectra in vivo (25, 27) and different host-cell specificities in vitro (11, 19, 29, 54). While studies of the strain-specific tissue tropisms exhibited by viruses of several other groups have indicated that this specificity is often dictated by the presence or absence of a receptor on the host cell (20, 24, 35, 36, 51, 60, 61), this is not the case for at least two serologically identical variants of minute virus of mice, MVM(p), the prototype strain, and MVM(i), an immunosuppressive variant (52).

MVM(i) was first isolated from a murine T-cell lymphoma (8) and shown to inhibit several T-cell-mediated functions in vitro (17). This immunosuppressive property of MVM(i) is not associated with the prototype virus MVM(p), which grows productively in cells of fibroblast origin. In vitro studies on the interaction of both strains of MVM with differentiated cells revealed that while these viruses show a tropism for cells of lymphoid [MVM(i)] or fibroblast [MVM(p)] origin, both viruses bind to receptors on productive and restrictive cell types. Furthermore, competition experiments showed that both viruses recognize the same receptor. Because the genomes of MVM(i) and MVM(p) are converted to viral replicative forms (RFs) in either the productive or restrictive cell type, it is presumed that penetration and uncoating of the viruses occur regardless of the final outcome of the virus-host cell interaction (52). Development of a plaque assay for both MVM(p) and MVM(i) in 324K cells, a strain of simian virus 40-transformed human kidney cells which is permissive for both MVM strains, allowed a study of the interactions of both strains with host cells of lymphocyte or fibroblast origin (55). These results showed that the host range of each virus variant is specified by a viral genetic determinant which is stably inherited. In addition, the susceptible host-cell phenotype is dominant in T lymphocyte × fibroblast fusion hybrids, implying that different cell types express different developmentally regulated virus helper functions which can only be exploited by the virus variant which carries the appropriate strain-specific determinant (55).

The construction and characterization of an infectious clone of the prototype strain of MVM has been reported previously (30). As a first step in the process of mapping the lymphotropic and fibrotropic determinants of these viruses by constructing recombinant genomes, we decided to molecularly clone the MVM(i) genome and determine its complete sequence. This paper describes the construction of molecular clones of MVM(i) which were used to determine the sequences of the 3′ and 5′ ends of the viral genome. The sequence of most of the coding region was determined by direct cloning of restriction fragments from MVM(i) RF into M13 vectors. In addition, we determined two important corrections in our earlier sequence for MVM(p) (4). A preliminary report of some of these data has been presented previously (M. Blundell, M. Thomson, E. Gardiner, S. Cotmore, J. Bratton, P. Tattersall, and C. R. Astell, Can. Fed. Biol. Soc. 27:128, 1984).

While this manuscript was in preparation, a similar report on the sequence of MVM(i) has appeared (48), and we have included in our discussion a comparison of these results with those presented here.

MATERIALS AND METHODS

Material. Enzymes were purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.) and New England BioLabs, Inc. (Beverly, Mass.). Linkers were purchased from New England BioLabs. [α-^{32}P]dATP (2,000

* Corresponding author.

to 3,000 Ci/mmol) was purchased from Amersham Corp. (Arlington Heights, Ill.). Genetically engineered DNA polymerase (Klenow fragment) was a gift from C. Joyce.

Virus stocks and cell lines. The cloned stock of the prototype strain of MVM(p) was grown in A9 cells (53, 55) and used as the source of DNA for the sequence studies and the construction of molecular clones of this genome (4, 30). Derivation of the clonal isolate of the immunosuppressive strain, MVM(i), has been described previously (52) and is the same virus as that used in the studies of Sahli et al. (48). Stocks of MVM(i) are routinely grown in S49.1 TB2 cells, a T-cell-lymphoma line, as described previously (55). The viral DNA from which viral termini were cloned was prepared from MVM(i) particles grown in 324K cells, a simian virus 40-transformed newborn human kidney cell line.

Plasmids of MVM(p) genome. Molecular clones of MVM(p), pMM984 (30) and pABO (7), have been described previously.

Molecular cloning of MVM(i) genome. The 5′ end of the MVM(i) genome was cloned in a manner similar to that of the MVM(p) genome (30). Single-stranded DNA purified from MVM(i) particles was extended with DNA polymerase (Klenow fragment) plus all four deoxyribonucleotide triphosphates. *Bam*HI linkers were ligated to the right-hand end, and the *Eco*RI-*Bam*HI (*Eco*RI-B) fragment (map units 70 to 100) was cloned between the *Eco*RI and *Bam*HI sites of pAT153. This clone was designated pEG222. The left-hand end of the viral genome was cloned in a similar way, taking advantage of the fact that 1% of the packaged viral strands of MVM(p) are of the plus orientation (9). The *Eco*RI-*Bam*HI (*Eco*RI-C) fragment (map units 1 to 21) was cloned between the *Eco*RI and *Bam*HI sites of pAT153. This clone was designated pEG618.

Cloning of MVM(i) RF restriction fragments in M13 vectors. MVM(i) RF DNA was grown in S49.1 TB2 cells, and RF DNA was isolated as described previously (55). Restriction fragments from partial and complete digests with *Hae*III, *Alu*I, and *Sau*3A were cloned into appropriate sites of M13 mp8 and mp9 by using the M13 shotgun cloning approach (31). To obtain clones in specific regions of the genome, we hybridized clones containing the appropriate complementary region from MVM(p) DNA (4). One small region could not be read on any of our clones (~ map unit 92) and was sequenced on a longer clone with a synthetic oligonucleotide primer, 5′-TAAGACTAATAAAGA-3′. It should be noted that using the M13 vectors we had difficulty obtaining large clones of the MVM(i) genome (particularly the right half). We have no explanation for this other than that from the sequence it is clear there are many *Eco*PI sites in the DNA, and our host cell at the time, JM103, carries a P1 lysogen which encodes the restriction and modification system (62).

Determination of nucleotide sequence. Sequence of the M13 clones was determined by the dideoxy terminator method (49). Sequences were aligned with that of MVM(p) on a computer, using the DNA SEQNCE program of A. Delaney (14). This approach gave virtually all the sequences between map units 8 and 94 with the exception of ~ map unit 92. The synthetic 15-mer described above was used to sequence this region from map units 91 to 94. The sequence in both strands was obtained for at least 80% of the genome.

Sequence of 3′ and 5′ ends of the genome. Because clones of the 3′ and 5′ palindromes were not expected (and indeed were never obtained) with single-stranded vectors, we sequenced the ends of the genome by using plasmid clones and the chemical sequencing method (28). The sequence of the left-hand end (map units 0 to 10) was derived from clone pEG618. The palindromic sequence was resolved completely by running several different percent acrylamide gels (20, 12, and 8%) at 40 W per gel to ensure complete denaturation of the hairpin. The region between map units 87 and 100 including the 5′ palindrome was sequenced by the chemical method with plasmid pEG222 DNA. The full-length 5′ palindrome was propagated in *Escherichia coli* JC8111 (7).

RESULTS

Correction of MVM(p) sequence. From a comparison of the sequence of MVM(i) presented here with that of MVM(p) (4), we are now aware of two errors in the sequence of MVM(p). First, the region about the *Eco*RI site at map unit 70 was incorrect owing to an error at the computer level in joining the sequence of the *Eco*RI A fragment to the *Eco*RI B fragment from clone pABO. (NB: This clone contains both the A and B fragments with the A fragment inverted relative to B [7].) We have now confirmed this region on pMM984, the full-length infectious clone (30), and the correct sequence is given in Fig. 1. In comparing the sequence of MVM(i) with that of MVM(p), we were initially surprised to find that the sequences at ~ map unit 94 were identical (except for a few single base changes). However, in agreement with McMaster et al. (29), we observed a mobility difference of about 60 base pairs (bp) in the *Xba*I-*Msp*I fragment (map units ~86 to 99) of MVM(p) and MVM(i). We resequenced this region of the MVM(p) genome and discovered that there is a 65-bp direct repeat in this region (nucleotides 4720 to 4784 and 4785 to 4849; Fig. 1). This region in the MVM(p) genome was originally sequenced by the chemical method (28) with clone pABO (4). Because *Sau*3AI fragments were used to read this region, the 65-bp *Sau*3AI fragment contained within the direct repeat was missed. Sequencing leftward from the *Sau*96I site at nucleotide 4872 to get the overlap of the *Sau*3AI site should have detected the missing fragment. Unfortunately, at the time when these sequence experiments were done, commercial preparations of *Sau*96I were often not sufficiently pure to give sequences which could be read for the usual 150 to 200 nucleotides. Indeed, although this experiment was done, the overlap extended only ~25 to 30 nucleotides, a length which was considered adequate at the time. Clearly it was not, since a 65-bp repeat was missed. We subsequently sequenced this region with current preparations of *Sau*96I and read 198 nucleotides leftward from the *Sau*96I site, through both copies of the direct repeat. The possible significance of this direct repeat is discussed below.

Nucleotide sequence of MVM(i) genome. The sequence of MVM(i) is given in Fig. 1 (top line). This sequence is presented in comparison with that of the corrected sequence of MVM(p), using the DNA SEQNCE program (14). Nucleotide differences between the two sequences are displaced above [MVM(i)] and below [MVM(p)] the respective lines. The 65-bp repeat (nucleotides 4785 to 4849) in MVM(p) is further displaced to facilitate maintaining the homology lineup between the two genomes beyond this point. It should be noted here that the clone pEG222 is 21 nucleotides shorter than MVM(p) viral-strand DNA at the right-hand end of the genome owing to the DNA polymerase I (Klenow fragment) extension of viral DNA which stopped short of replicating the entire hairpin. The last 22 nucleotides of the MVM(i) right-end sequence shown here are inferred from the sequence of the clone between nucleotides 4880 and 4902.

It is clear from the sequence that with the exception of the

```
        10          20          30          40          50          60          70          80          90         100
                                                                  T        T
ATTTTTAGAACTGACCAACCATGTTCACGTAAGTGACGTGATGACGCGCGCT CGCGCGC GCCTTCGGACGTCACACGTCACTTACGTTTCACATGGTT
ATTTTTAGAACTGACCAACCATGTTCACGTAAGTGACGTGATGACGCGCGCT CGCGCGC GCCTTCGGACGTCACACGTCACTTACGTTTCACATGGTT
                                                                  G
        10          20          30          40          50          60          69          79          89          99

       110         120         130         140         150         160         170         180         190         200
                                A       G                                                               A G
GGTCAGTTCTAAAAATGATAAGCGGTTCAG GAGTTTA ACCAAGGCGCGAAAAGGAAGTGGGCGTGGTTTAAAGTATATAAGCAA T CTGAAGTCAGT
GGTCAGTTCTAAAAATGATAAGCGGTTCAG GAGTTTA ACCAAGGCGCGAAAAGGAAGTGGGCGTGGTTTAAAGTATATAAGCAA T CTGAAGTCAGT
                                G       A                                                               C A
       109         119         129         139         149         159         169         179         189         199

       210         220         230         240         250         260         270         280         290         300
  C                                G                                                                          A
TACTTATC TTTCTTTCATTCTGTGAGTCGAGACGC CAGAAAGAGAGTAACCAACTAACCATGGCTGGAAATGCTTACTCTGATGAAGTTTTGGGA CA
TACTTATC TTTCTTTCATTCTGTGAGTCGAGACGC CAGAAAGAGAGTAACCAACTAACCATGGCTGGAAATGCTTACTCTGATGAAGTTTTGGGA CA
  T                                A                                                                          G
       209         219         229         239         249         259         269         279         289         299

       310         320         330         340         350         360         370         380         390         400
                                                                 C   GG         A                             A
ACCAACTGGTTAAAGGAAAAAAGTAACCAGGAAGTGTTCTCATTTGTTTTTAAAA TGA  ATGTTCAACT AATGGAAAAGATATCGGATGGAATA TT
ACCAACTGGTTAAAGGAAAAAAGTAACCAGGAAGTGTTCTCATTTGTTTTTAAAA TGA  ATGTTCAACT AATGGAAAAGATATCGGATGGAATA TT
                                                                 A   AA         G                             G
       309         319         329         339         349         359         369         379         389         399

       410         420         430         440         450         460         470         480         490         500
      G                                                           C                                         T T
ACAAAAA GAGCTGCAGGAGGACGAGCTGAAATCTTTACAACGAGGAGCGGAAACTAC TGGGACCAAAGCGAGGACATGGAATGGGAA C ACAGTGGA
ACAAAAA GAGCTGCAGGAGGACGAGCTGAAATCTTTACAACGAGGAGCGGAAACTAC TGGGACCAAAGCGAGGACATGGAATGGGAA C ACAGTGGA
      A                                                           T                                         A C
       409         419         429         439         449         459         469         479         489         499

       510         520         530         540         550         560         570         580         590         600
                                A   C    A                G               G       A      GC      C          C
TGAAATGACCAAAAAGCAAGTATTCATTT TGA TCTTT GTTAAAAAAATGTTT TTTGAAGTGCTTA CACAAA AATATA  TCCTG TGATGTTA T
TGAAATGACCAAAAAGCAAGTATTCATTT TGA TCTTT GTTAAAAAAATGTTT TTTGAAGTGCTTA CACAAA AATATA  TCCTG TGATGTTA T
                                T   T    G                A               A       G      TT      G          A
       509         519         529         539         549         559         569         579         589         599

       610         620         630         640         650         660         670         680         690         700
          G          G                                                  C                            A
TGGTTTGTGCA CATGAATGGGG AAAGACCAAGGCTGGCACTGCCATGTACTAATTGGAGG AAGGACTTTAGTCAAGCTCAAGG AAATGGTGGAGAA
TGGTTTGTGCA CATGAATGGGG AAAGACCAAGGCTGGCACTGCCATGTACTAATTGGAGG AAGGACTTTAGTCAAGCTCAAGG AAATGGTGGAGAA
          A          A                                                  A                            G
       609         619         629         639         649         659         669         679         689         699

       710         720         730         740         750         760         770         780         790         800
   G                                                           G                                             G
GGCA CTAAATGTTTACTGGAGCAGATGGTTGGTAACAGCCTGTAATGTGCA CTAACACCAGCTGAAAGAATTAAACTAAGAGAAATAGCAGAAGACA
GGCA CTAAATGTTTACTGGAGCAGATGGTTGGTAACAGCCTGTAATGTGCA CTAACACCAGCTGAAAGAATTAAACTAAGAGAAATAGCAGAAGACA
   A                                                           A                                             A
       709         719         729         739         749         759         769         779         789         799

       810         820         830         840         850         860         870         880         890         900
             T    C         A                    G         T   A                  T                C          C
TGAGTGGGTTACT TACT ACTTATAA CATAAGCAAACCAAAAA GACTATAC AA TGTGTTCTTTTTGGAAA ATGATTGCTTACTA TTTTTAAC
TGAGTGGGTTACT TACT ACTTATAA CATAAGCAAACCAAAAA GACTATAC AA TGTGTTCTTTTTGGAAA ATGATTGCTTACTA TTTTTAAC
             C    T         G                    A         C   G                  C                T          T
       809         819         829         839         849         859         869         879         889         899

       910         920         930         940         950         960         970         980         990        1000
                  C     G     G                  A                                                  G     A
AAAAAGAAAATAAGCAC AGTCC CCAAG GACGGAGGCTATTTTCT AGCAGTGACTCTGGCTGGAAAACTAACTTTTTAAAAGA GGCGA CGCCATC
AAAAAGAAAATAAGCAC AGTCC CCAAG GACGGAGGCTATTTTCT AGCAGTGACTCTGGCTGGAAAACTAACTTTTTAAAAGA GGCGA CGCCATC
                  T     A     A                  T                                                  A     G
       909         919         929         939         949         959         969         979         989         999

      1010        1020        1030        1040        1050        1060        1070        1080        1090        1100
           T
TAGTGAGCAAA TATACACTGATGACATGCGGCCAGAAACGGTTGAAACCACAGTAACCACTGCGCAGGAAACTAAGCGCGGCAGAATTCAAACTAAAAA
TAGTGAGCAAA TATACACTGATGACATGCGGCCAGAAACGGTTGAAACCACAGTAACCACTGCGCAGGAAACTAAGCGCGGCAGAATTCAAACTAAAAA
           C
      1009        1019        1029        1039        1049        1059        1069        1079        1089        1099
```

```
      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
   G              C              A                          A
AGA GTTTCTATTAAAAC ACACTTAAAGAGCT GTGCATAAAAGAGTAACCTCACCAGA GACTGGATGATGATGCAGCCAGACAGTTACATTGAAATG
AGA GTTTCTATTAAAAC ACACTTAAAGAGCT GTGCATAAAAGAGTAACCTCACCAGA GACTGGATGATGATGCAGCCAGACAGTTACATTGAAATG
   A              T              G                          G
      1109      1119      1129      1139      1149      1159      1169      1179      1189      1199

      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
                                                     G                 A                 G
ATGGCTCAACCAGGTGGAGAAAACCTGCTGAAAAATACGCTAGAGATTTGTAC CTAACTCTAGCCAGAAC AAAACAGCATTTGACTT ATTTTAGAAA
ATGGCTCAACCAGGTGGAGAAAACCTGCTGAAAAATACGCTAGAGATTTGTAC CTAACTCTAGCCAGAAC AAAACAGCATTTGACTT ATTTTAGAAA
                                                     A                 C                 A
      1209      1219      1229      1239      1249      1259      1269      1279      1289      1299

                                                                1360      1370      1380      1390      1400
      1310      1320      1330      1340      1350      AG
A AGCTGAAACCAGCAAACTAACCAACTTTTCACTGCCTGACACAAGAACCTGCA  ATTTTTGCTTTTCATGGCTGGAACTATGTTAAAGTTTGCCATGC
A AGCTGAAACCAGCAAACTAACCAACTTTTCACTGCCTGACACAAGAACCTGCA  ATTTTTGCTTTTCATGGCTGGAACTATGTTAAAGTTTGCCATGC
                                                       GA
      1309      1319      1329      1339      1349      1359      1369      1379      1389      1399

      1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
             C                                        C           T
TATTTGCTGTGTT TAAACAGACAAGGAGGCAAAAGAAATACTGTTTTATTTCA GGACCAGCCAG ACAGGCAAATCTATTATTGCACAAGCCATAGCA
TATTTGCTGTGTT TAAACAGACAAGGAGGCAAAAGAAATACTGTTTTATTTCA GGACCAGCCAG ACAGGCAAATCTATTATTGCACAAGCCATAGCA
             T                                        T           C
      1409      1419      1429      1439      1449      1459      1469      1479      1489      1499

      1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
  G        T                       T     G
CA GCAGTTGG AATGTTGGTTGCTATAATGCAGC AATGT AACTTTCCATTTAATGACTGTACCAACAAGAACTTGATTTGGGTAGAAGAAGCTGGTA
CA GCAGTTGG AATGTTGGTTGCTATAATGCAGC AATGT AACTTTCCATTTAATGACTGTACCAACAAGAACTTGATTTGGGTAGAAGAAGCTGGTA
  A        C                       C     A
      1509      1519      1529      1539      1549      1559      1569      1579      1589      1599

      1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
                                                                                     A
ACTTTGGACAGCAAGTAAACCAGTTTAAAGCCATTTGCTCTGGTCAAACTATTCGCATTGATCAAAAAGGAAAAGGCAGCAAACA ATTGAACCAACACC
ACTTTGGACAGCAAGTAAACCAGTTTAAAGCCATTTGCTCTGGTCAAACTATTCGCATTGATCAAAAAGGAAAAGGCAGCAAACA ATTGAACCAACACC
                                                                                     G
      1609      1619      1629      1639      1649      1659      1669      1679      1689      1699

      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
                                                      G                       T              C
AGTCATCATGACCACAAATGAGAACATTACAGTGGTCAGAATAGGCTGCGAAGA AGACCAGAACACACTCAACCAAT AGAGACAGAATGCT AACATT
AGTCATCATGACCACAAATGAGAACATTACAGTGGTCAGAATAGGCTGCGAAGA AGACCAGAACACACTCAACCAAT AGAGACAGAATGCT AACATT
                                                      A                       C              T
      1709      1719      1729      1739      1749      1759      1769      1779      1789      1799

      1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
              A                             G
CATCTAACACATAC TTGCCTGGTGACTTTGGTTTGGTTGACAA AATGAATGGCCCATGATTTGTGCTTGGTTGGTAAAGAATGGTTACCAATCTACCA
CATCTAACACATAC TTGCCTGGTGACTTTGGTTTGGTTGACAA AATGAATGGCCCATGATTTGTGCTTGGTTGGTAAAGAATGGTTACCAATCTACCA
              C                             A
      1809      1819      1829      1839      1849      1859      1869      1879      1889      1899

      1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
             C                                                        G             C
TGGCAAGCTACTG GCTAAATGGGGCAAAGTTCCTGATTGGTCAGAAAACTGGGCGGAGCCAAAGGTGCC ACTCCTATAAATT ACTAGGTTCGGCACG
TGGCAAGCTACTG GCTAAATGGGGCAAAGTTCCTGATTGGTCAGAAAACTGGGCGGAGCCAAAGGTGCC ACTCCTATAAATT ACTAGGTTCGGCACG
             T                                                        A             T
      1909      1919      1929      1939      1949      1959      1969      1979      1989      1999

      2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
                                                 A
CTCACCATTCACGACACCGAAAAGTACGCCTCTCAGCCAGAACTATGCA TAACTCCACTTGCATCGGATCTCGAGGACCTGGCTTTAGAGCCTTGGAGC
CTCACCATTCACGACACCGAAAAGTACGCCTCTCAGCCAGAACTATGCA TAACTCCACTTGCATCGGATCTCGAGGACCTGGCTTTAGAGCCTTGGAGC
                                                 C
      2009      2019      2029      2039      2049      2059      2069      2079      2089      2099

      2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
ACACCAAATACTCCTGTTGCGGGCACTGCAGAAACCCAGAACACTGGGGAAGCTGGTTCCAAAGCCTGCCAAGATGGTCAACTGAGCCCAACTTGGTCAG
ACACCAAATACTCCTGTTGCGGGCACTGCAGAAACCCAGAACACTGGGGAAGCTGGTTCCAAAGCCTGCCAAGATGGTCAACTGAGCCCAACTTGGTCAG
      2109      2119      2129      2139      2149      2159      2169      2179      2189      2199
```

```
      2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
                                                G
AGATCGAGGAGGATTTGAGAGCGTGCTTCGGTGCGGAACCGTTGAAGA AGACTTCAGCGAGCCGCTGAACTTGGACTAAGGTACGATGGCGCCTCCAGC
AGATCGAGGAGGATTTGAGAGCGTGCTTCGGTGCGGAACCGTTGAAGA AGACTTCAGCGAGCCGCTGAACTTGGACTAAGGTACGATGGCGCCTCCAGC
                                                A
      2209      2219      2229      2239      2249      2259      2269      2279      2289      2299

      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400

TAAAAGAGCTAAAAGAGGTAAGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGTTTTACAGGCCTGAAATCACTTGGTTTTAGG
TAAAAGAGCTAAAAGAGGTAAGGGTTTAAGGGATGGTTGGTTGGTGGGGTATTAATGTTTAATTACCTGTTTTACAGGCCTGAAATCACTTGGTTTTAGG

      2309      2319      2329      2339      2349      2359      2369      2379      2389      2399

      2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
                  T
TTGGGTGCCTCCTGGCTA AAGTACCTGGGACCAGGGAACAGCCTTGACCAAGGAGAACCAACCAATCCATCTGACGCCGCTGCCAAAGAGCACGACGAG
TTGGGTGCCTCCTGGCTA AAGTACCTGGGACCAGGGAACAGCCTTGACCAAGGAGAACCAACCAATCCATCTGACGCCGCTGCCAAAGAGCACGACGAG
                  C
      2409      2419      2429      2439      2449      2459      2469      2479      2489      2499

      2510      2520      2530      2540      2550      2560      2570      2580      2590      2600

GCCTATGATCAATACATCAAATCTGGAAAAAATCCTTACCTGTACTTCTCTGCTGCTGATCAACGCTTTATTGACCAAACCAAGGACGCCAAAGACTGGG
GCCTATGATCAATACATCAAATCTGGAAAAAATCCTTACCTGTACTTCTCTGCTGCTGATCAACGCTTTATTGACCAAACCAAGGACGCCAAAGACTGGG

      2509      2519      2529      2539      2549      2559      2569      2579      2589      2599

      2610      2620      2630      2640      2650      2660      2670      2680      2690      2700

GAGGCAAGGTTGGTCACTACTTTTTTAGAACCAAGCGCGCTTTTGCACCTAAGCTTGCTACTGACTCTGAACCTGGAACTTCTGGTGTAAGCAGAGCTGG
GAGGCAAGGTTGGTCACTACTTTTTTAGAACCAAGCGCGCTTTTGCACCTAAGCTTGCTACTGACTCTGAACCTGGAACTTCTGGTGTAAGCAGAGCTGG

      2609      2619      2629      2639      2649      2659      2669      2679      2689      2699

      2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
                                   A
TAAACGCACTAGACCACCTGCTTACATTTTTAT AACCAAGCCAGAGCTAAAAAAAAACTTACTTCTTCTGCTGCACAGCAAAGCAGTCAAACCATGAGT
TAAACGCACTAGACCACCTGCTTACATTTTTAT AACCAAGCCAGAGCTAAAAAAAAACTTACTTCTTCTGCTGCACAGCAAAGCAGTCAAACCATGAGT
                                   T
      2709      2719      2729      2739      2749      2759      2769      2779      2789      2799

      2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
                       G
GATGGCACCAGCCAACCTGAC GCGGAAACGCTGTCCACTCAGCTGCAAGAGTTGAACGAGCAGCTGACGGCCCTGGAGGCTCTGGGGGTGGGGGCTCTG
GATGGCACCAGCCAACCTGAC GCGGAAACGCTGTCCACTCAGCTGCAAGAGTTGAACGAGCAGCTGACGGCCCTGGAGGCTCTGGGGGTGGGGGCTCTG
                       A
      2809      2819      2829      2839      2849      2859      2869      2879      2889      2899

      2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
                                             G
GCGGGGGTGGGGTTGGTGTTTCTACTGGGTCTTATGATAATCA ACGCATTATAGATTCTTGGGTGACGGCTGGGTAGAAATTACTGCACTAGCAACTAG
GCGGGGGTGGGGTTGGTGTTTCTACTGGGTCTTATGATAATCA ACGCATTATAGATTCTTGGGTGACGGCTGGGTAGAAATTACTGCACTAGCAACTAG
                                             A
      2909      2919      2929      2939      2949      2959      2969      2979      2989      2999

      3010      3020      3030      3040      3050      3060      3070      3080      3090      3100
                                               A          C          T
ACTAGTACATTTAAACATGCCTAAATCAGAAAACTATTGCAGAAT AGAGTTCACAA ACAACAGACAC TCAGTCAAAGGCAACATGGCAAAAGATGAT
ACTAGTACATTTAAACATGCCTAAATCAGAAAACTATTGCAGAAT AGAGTTCACAA ACAACAGACAC TCAGTCAAAGGCAACATGGCAAAAGATGAT
                                               C          T          A
      3009      3019      3029      3039      3049      3059      3069      3079      3089      3099

      3110      3120      3130      3140      3150      3160      3170      3180      3190      3200
                       G           A
GCTCATGAGCAAATTTGGAC CCATGGAGCTT GTGGATGCTAATGCTTGGGGAGTTTGGCTCCAGCCAAGTGACTGGCAATACATTTGCAACACCATGA
GCTCATGAGCAAATTTGGAC CCATGGAGCTT GTGGATGCTAATGCTTGGGGAGTTTGGCTCCAGCCAAGTGACTGGCAATACATTTGCAACACCATGA
                       A           G
      3109      3119      3129      3139      3149      3159      3169      3179      3189      3199

      3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
                  C                    T                                   C
GCCAGCTTAACTTGGT TCACTTGATCAAGAAATATT AATGTAGTGCTGAAAACTGTTACAGAGCAAGACT AGGAGGTCAAGCTATAAAAATATACAA
GCCAGCTTAACTTGGT TCACTTGATCAAGAAATATT AATGTAGTGCTGAAAACTGTTACAGAGCAAGACT AGGAGGTCAAGCTATAAAAATATACAA
                  A                    C                                   T
      3209      3219      3229      3239      3249      3259      3269      3279      3289      3299
```

```
     3310        3320        3330        3340        3350        3360        3370        3380        3390        3400
        C                                               C                                                         T
CAATGACCT ACAGCTTGCATGATGGTTGCAGTAGACTCAAACAACATT TGCCATACACACCTGCAGCAAACTCAATGGAAACACTTGGTTTCTACCC
CAATGACCT ACAGCTTGCATGATGGTTGCAGTAGACTCAAACAACATT TGCCATACACACCTGCAGCAAACTCAATGGAAACACTTGGTTTCTACCC
        T                                               T                                                         C
     3309        3319        3329        3339        3349        3359        3369        3379        3389        3399

     3410        3420        3430        3440        3450        3460        3470        3480        3490        3500
            T        G                   C                                   A     T                 A    G
TGGAAACCAAC ATAGCATC CCATACAGGTACTATTT TGCGTTGACAGAGATCTTTCAGT ACCTA GAAAATCAAGAAGGCACA TTGA CATAATG
TGGAAACCAAC ATAGCATC CCATACAGGTACTATTT TGCGTTGACAGAGATCTTTCAGT ACCTA GAAAATCAAGAAGGCACA TTGA CATAATG
            C        A                   T                                   G     C                 G    A
     3409        3419        3429        3439        3449        3459        3469        3479        3489        3499
     3510        3520        3530        3540        3550        3560        3570        3580        3590        3600
A                                                                                      T   T   T   G    T   T   A
T ATGGGAACACCAAAAGGAATGAATTCTCAATTTTTTACCATTGAGAACACACAACAAATCACATTGCTCAGAAC GG GA GA TTTGC AC GG AC
T ATGGGAACACCAAAAGGAATGAATTCTCAATTTTTTACCATTGAGAACACACAACAAATCACATTGCTCAGAAC GG GA GA TTTGC AC GG AC
G                                                                                      A   G   C   A    C   A   T
     3509        3519        3529        3539        3549        3559        3569        3579        3589        3599
     3610        3620        3630        3640        3650        3660        3670        3680        3690        3700
C                 CC          T          A          T
 TACTACTTTGACACAAA  CAGTTAAACT ACACACAC TGGCAAAC AACCGTCAACTTGGACAGCCTCCACTGCTGTCAACCTTTCCTGAAGCTGAC
 TACTACTTTGACACAAA  CAGTTAAACT ACACACAC TGGCAAAC AACCGTCAACTTGGACAGCCTCCACTGCTGTCAACCTTTCCTGAAGCTGAC
T                 TT          C          G          C
     3609        3619        3629        3639        3649        3659        3669        3679        3689        3699

     3710        3720        3730        3740        3750        3760        3770        3780        3790        3800
                                         G       G      A                             T
ACTGATGCAGGTACACTTACTGCTCAAGGGAGCAGACATGGA CAACACA ATGG GGTTAACTGGGTGAGTGAAGCAAT AGAACCAGACCTGCTCAAG
ACTGATGCAGGTACACTTACTGCTCAAGGGAGCAGACATGGA CAACACA ATGG GGTTAACTGGGTGAGTGAAGCAAT AGAACCAGACCTGCTCAAG
                                         A       A      G                             C
     3709        3719        3729        3739        3749        3759        3769        3779        3789        3799

     3810        3820        3830        3840        3850        3860        3870        3880        3890        3900
              G                                                T                  G                  G    G
TAGGATTTTGTCA CCACACAATGACTTTGAAGCCAGCAGAGCTGGACCATTTGCTGC CCAAAAGTTCCAGCAGAT TTACTCAAGGAGT GACA AGA
TAGGATTTTGTCA CCACACAATGACTTTGAAGCCAGCAGAGCTGGACCATTTGCTGC CCAAAAGTTCCAGCAGAT TTACTCAAGGAGT GACA AGA
              A                                                C                  A                  A    A
     3809        3819        3829        3839        3849        3859        3869        3879        3889        3899

     3910        3920        3930        3940        3950        3960        3970        3980        3990        4000
                                                            G    C                                            A
AGCCAATGGCAGTGTTAGATACAGTTATGGCAAACAGCATGGTGAAAATTGGGCT CACA GGACCAGCACCAGAGCGCTACACATGGGATGAAACAA C
AGCCAATGGCAGTGTTAGATACAGTTATGGCAAACAGCATGGTGAAAATTGGGCT CACA GGACCAGCACCAGAGCGCTACACATGGGATGAAACAA C
                                                            T    T                                            G
     3909        3919        3929        3939        3949        3959        3969        3979        3989        3999

     4010        4020        4030        4040        4050        4060        4070        4080        4090        4100
           A          G                      T                              G
TTTGGTTCAGG AGAGACACCA AGATGGTTTTATTCAATCAGCACC CTAGTTGTTCCACCACCACTAAATGG ATTCTTACAAATGCAAACCCTATTG
TTTGGTTCAGG AGAGACACCA AGATGGTTTTATTCAATCAGCACC CTAGTTGTTCCACCACCACTAAATGG ATTCTTACAAATGCAAACCCTATTG
           T          A                      A                              C
     4009        4019        4029        4039        4049        4059        4069        4079        4089        4099

     4110        4120        4130        4140        4150        4160        4170        4180        4190        4200
A
G ACTAAAAATGACATTCATTTTTCAAATGTTTTTAACAGCTATGGTCCACTAACTGCATTTTCACACCCAAGTCCTGTATACCCTCAAGGACAAATATG
G ACTAAAAATGACATTCATTTTTCAAATGTTTTTAACAGCTATGGTCCACTAACTGCATTTTCACACCCAAGTCCTGTATACCCTCAAGGACAAATATG
G
     4109        4119        4129        4139        4149        4159        4169        4179        4189        4199

     4210        4220        4230        4240        4250        4260        4270        4280        4290        4300
                                                           C
GGACAAAGAACTAGATCTTGAACACAAACCTAGACTTCACATAACTGCTCCATTTGT TGTAAAAACAATGCACCTGGACAAATGTTGGTTAGATTAGGA
GGACAAAGAACTAGATCTTGAACACAAACCTAGACTTCACATAACTGCTCCATTTGT TGTAAAAACAATGCACCTGGACAAATGTTGGTTAGATTAGGA
                                                           T
     4209        4219        4229        4239        4249        4259        4269        4279        4289        4299

     4310        4320        4330        4340        4350        4360        4370        4380        4390        4400
     T           G                                    G   T   T
CCAAA CTAACTGACCA TATGATCCAAACGGAGCCACACTTTCTAGAATTGT AC TA GGTACATTTTTCTGGAAAGGAAAACTAACCATGAGAGCAA
CCAAA CTAACTGACCA TATGATCCAAACGGAGCCACACTTTCTAGAATTGT AC TA GGTACATTTTTCTGGAAAGGAAAACTAACCATGAGAGCAA
     C           A                                    T   A   C
     4309        4319        4329        4339        4349        4359        4369        4379        4389        4399
```

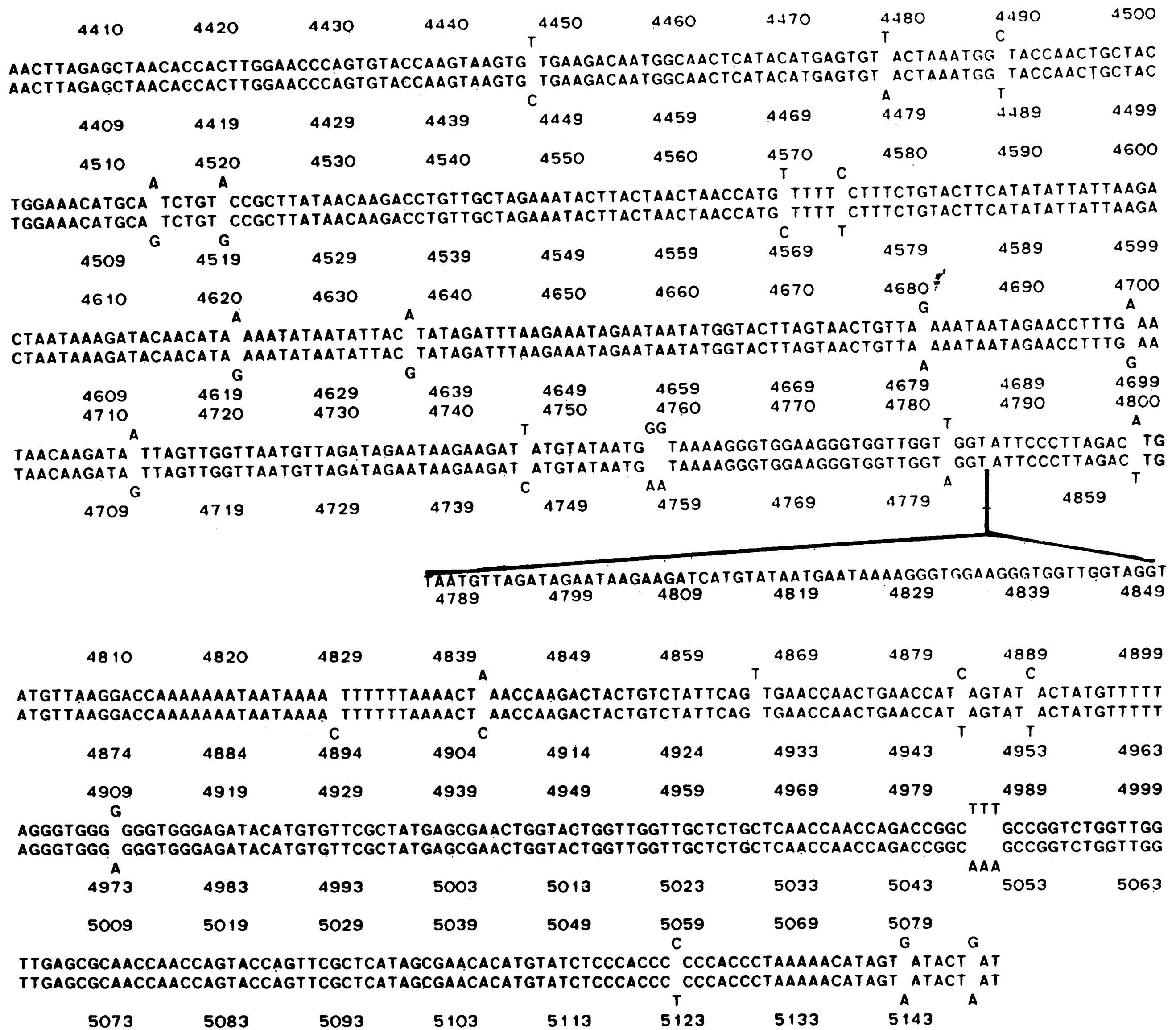


FIG. 1. Homology between MVM(i) (upper line) and MVM(p) (lower line). The two sequences were aligned by using the homology routine of the DNA sequence program (14). Mismatch nucleotides are displaced above and below the lines. The 65-bp tandem repeat in MVM(p) (nucleotides 4785 to 4849) is further displaced to maintain the lineup of the two sequences.

65-bp repeat, the sequences of MVM(p) and MVM(i) are very similar. In the MVM(i) genome three differences are an insertion of a T at position 61, a deletion of a C at position 4828, and a compensatory insertion of a T at position 4864 relative to MVM(p). None of these changes occurs within the coding region of MVM(p) (41) or, presumably, of MVM(i). Since the 5′ hairpin region of our sequence is derived from a clone, pEG222, it has only one orientation, in this case the flip orientation. It is interesting that the end of the loop in this flip orientation is 5′-TTT-3′ on the C strand, whereas in MVM(p) this sequence is the complementary 5′-AAA-3′. Of the 163 bp changes between MVM(i) and MVM(p), 129 are transitions and 34 are transversions. Distribution of these changes is nonrandom throughout the genome. In a genome size of ~5,000 nucleotides, 163 random base changes would average 3 changes per 100 nucleotides. The region between map units 40 and 60 is highly conserved (there are only six base changes in the sequence between nucleotides 2000 and 3000), whereas the regions between map units 8 to 20 and 70 to 80 have a much higher frequency of base changes (up to 8 to 10 changes per 100 nucleotides). This high degree of sequence conservation supports our previous suggestion that there may be multiple splice sites near map units 45 to 46 of MVM(p), with implied overlapping reading frames to generate different viral proteins (4). Another region (~ nucleotide 3580, map unit 70) seems to be a hot spot for base changes. There are eight changes within a distance of 25 nucleotides (Fig. 1). This region is within the coding region for the viral structural proteins; however, because all of these base changes occur in the third position of eight codons, the amino acid sequences between MVM(i) and MVM(p) in this region are identical (see below and Fig. 3).

Left-hand coding region. The sequence of MVM(i) is very similar to that of MVM(p) (4), with two major open reading frames which span almost the entire genome. The left-hand

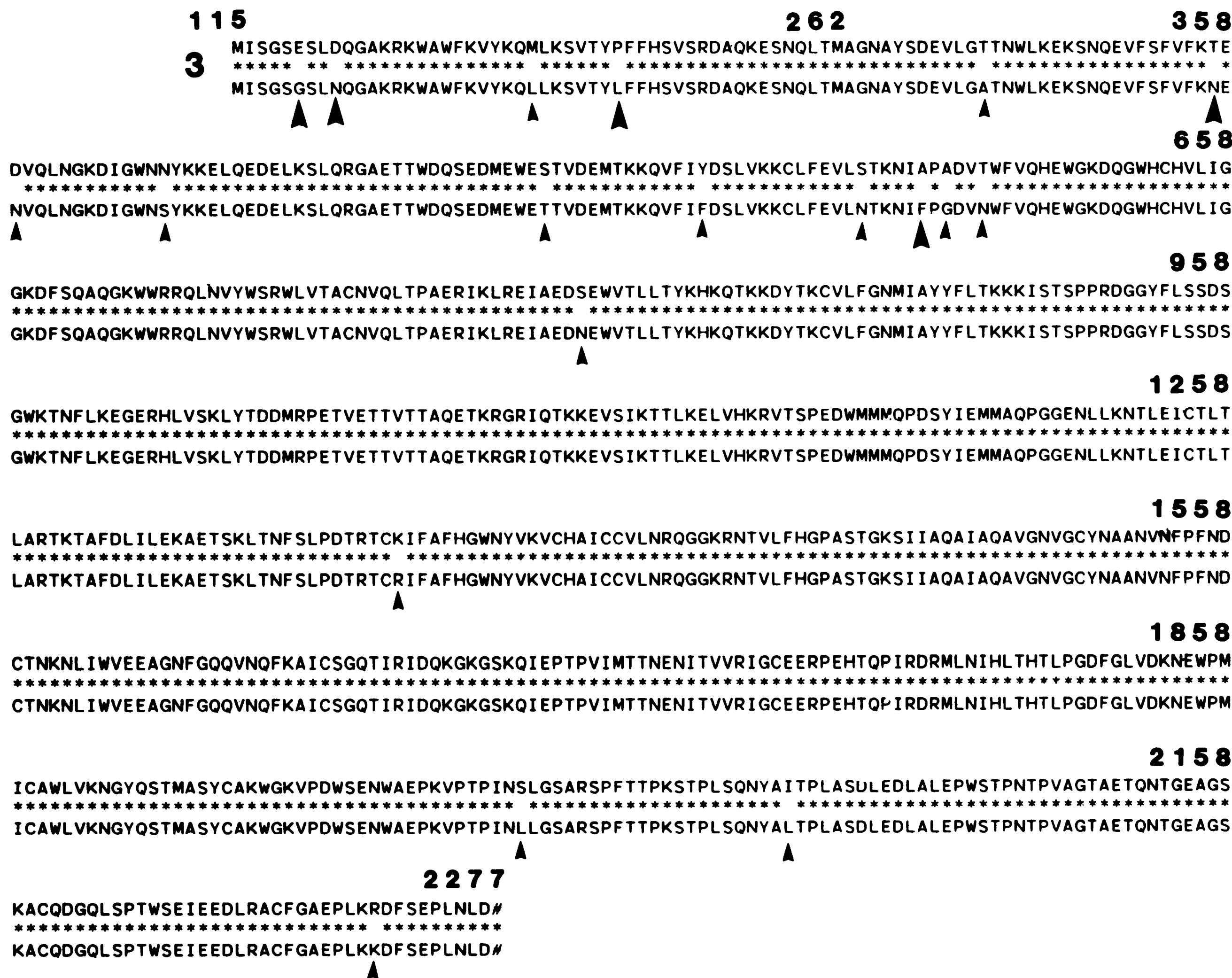


FIG. 2. Lineup of the left-hand open reading frame of MVM(i) (upper line) and MVM(p) (lower line). Identical amino acids are indicated by an asterisk. Conservative changes are noted by a small arrowhead below the amino acids. Radical changes are noted by a larger arrowhead. To facilitate comparison with Fig. 1, the numbers correspond to the nucleotide in the first position of the codon.

open reading frame (frame 3) extends from map units 2 to 45. The sequence of MVM(p) (4), transcription data (41), and subsequent studies on the mapping of the viral genome (13) indicate that NS-1, a nonstructural protein (83,000 daltons) of unknown function, maps to the left-hand coding region of the genome (13). A second nonstructural protein, NS-2 (24,000 daltons), has not yet been mapped precisely but is encoded in an mRNA transcribed from the left-hand (P4) promoter (13). Assuming that the protein products of the two strains, i and p, are similar, it is reasonable to compare the coding regions of the viruses to learn if there are significant differences at the amino acid level.

In Fig. 2, a lineup (14) of amino acids encoded within the potential left-hand open reading frames of MVM(i) (upper line) and MVM(p) (lower line) is given. Transcript R1 (4.8 kilobases) is a minor RNA component which encodes NS-1 (13). This transcript begins around nucleotide 201 (6), with the first methionine codon starting at nucleotide 262 of MVM(i) [261 of MVM(p)]. The protein (translated from reading frame 3) would begin with Met-Ala-Gly-Asn-Ala (MAGNA in Fig. 2). The first termination codon (TAA) in reading frame 3 occurs at position 2278. Beginning with the methionine at 262, there are a total of 15 amino acid changes between the MVM(i) and MVM(p) coding regions. Thirteen of these changes are conservative: Thr→Ala, Thr→Asn, Asn→Ser, Ser→Thr, Tyr→Phe, Ser→Asn, Ala→Gly, Thr→Asn, Ser→Asn, Lys→Arg, Ser→Leu, Ile→Leu, and Arg→Lys. There are two radical changes: Gly→Glu at position 358, and Ala→Phe at position 583. We noted earlier that for MVM(p), reading frame 3 is open upstream of the promoter region and the mapped mRNA start site. From the most upstream methionine codon, at position 115, in this extra open reading frame, the protein should begin with the amino acid sequence Met-Ile-Ser-Gly-Ser (MISGS) if this ATG is a functional start site. This region is also open in the MVM(i) sequence, with four additional amino acid differences from the polypeptide encoded between nucleotides 115 and 261 of MVM(p). Three of these changes are radical: Glu→Gly, Asp→Asn, and Pro→Leu. The fourth change, Met→Leu, is conservative. The possible significance of this region of frame 3 will be discussed below.

Right-hand coding region. The right-hand open reading frame of MVM (frame 1) extends from map units 46 to 89. Mapping studies have shown this region to encode the

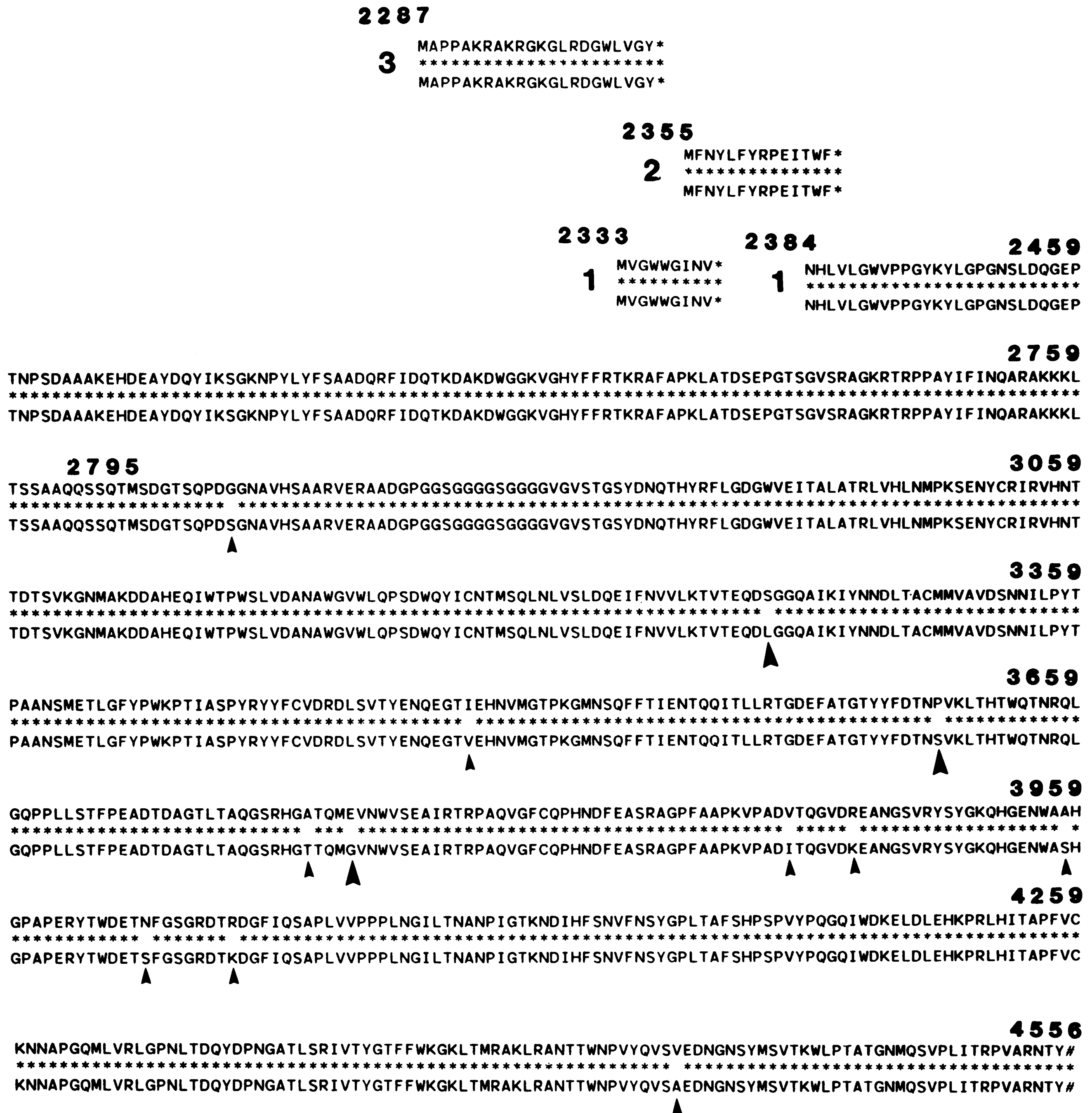


FIG. 3. Lineup of the right-hand open reading frame of MVM(i) (upper line) and MVM(p) (lower line). Identical amino acids are indicated by an asterisk. Conservative changes are emphasized by a small arrowhead below the amino acids. Radical changes are noted by a larger arrowhead. As discussed in the text and illustrated in Fig. 4, VP1 likely begins at the methionine at position 2287, and splicing of the transcript to frame 1 allows continued translation to the stop codon at nucleotide 4556. VP2 likely begins at 2795 and also terminates at the stop codon at nucleotide 4556 (see text for additional details). Numbers correspond to the first nucleotide of the codon.

overlapping structural polypeptides VP1 and VP2, which are both primary products of in vitro translation of viral mRNA (13, 46). The first methionine codon in this large right-hand open reading frame occurs at position 2795 (Fig. 3). The start of the mRNA(s) encoding these proteins maps to position 2005 ± 5 (6). We mapped this transcription start against the DNA sequence to position 2003, with two minor starts at 2006 and 2010 (C. Astell, unpublished data). Earlier, we reasoned that the AUG at 2795 was unlikely to be the functional start signal for VP2 (4). It would require ribosomes to traverse some 790 nucleotides of mRNA and to bypass at least three upstream AUGs (in closed reading frames) to initiate here the translation of the most abundant viral gene product. However, structural studies on the N terminus of VP2 in H1 have indicated (40) that this is indeed the likely start codon for VP2 translation and that the amino

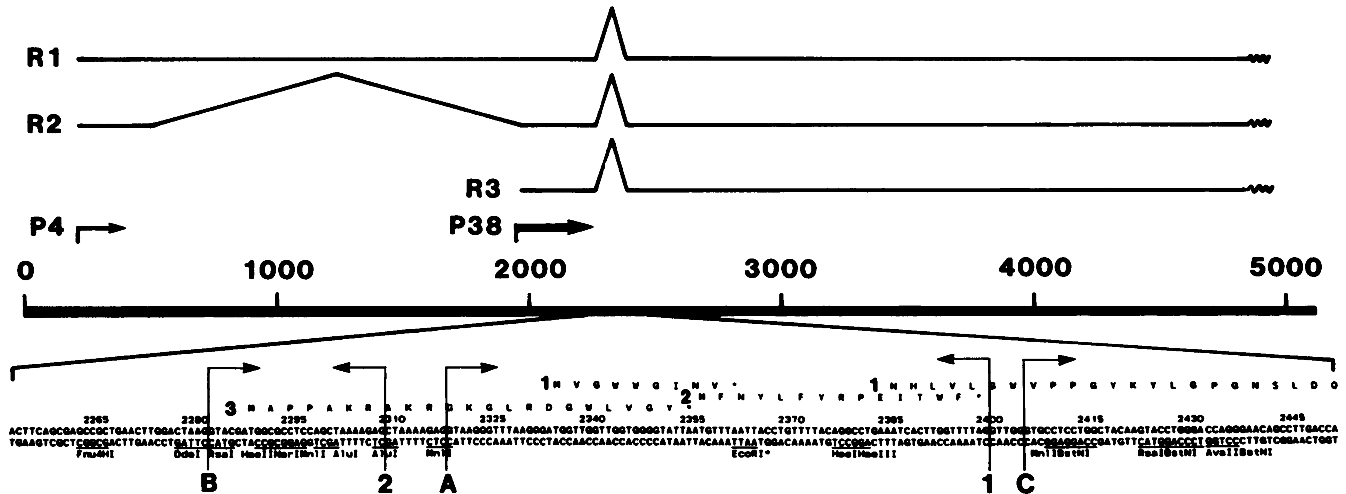


FIG. 4. Potential splice junctions at map unit ~45. The upper portion of the figure illustrates the RNA transcripts and splicing pattern for MVM(p) (41). As discussed in the text, a comparison of splice junctions at map unit ~45 for MVM(p), MVM(i), H1, and canine parvovirus indicate that a splice from A (5′ donor site) to 1 (3′ acceptor site) will generate the mRNA encoding VP1. Alternative splicing from B (5′ donor site) to 1 (3′ acceptor site) produces the probable mRNA encoding VP2 (see text for details). R1, R2, and R3 are the major spliced transcripts, and P4 and P38 are the two promoters (41).

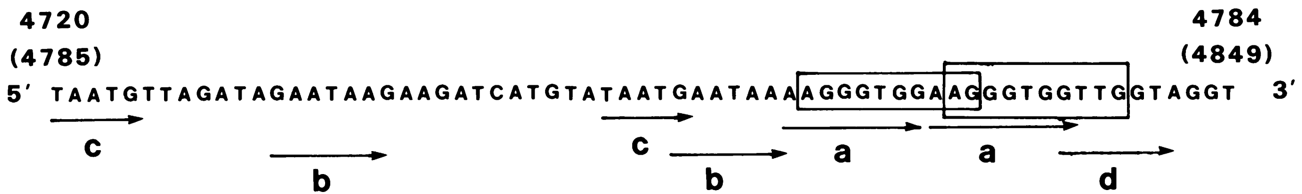
acid sequence of VP2 is identical to the C-terminal three-quarters of VP1.

Transcription studies on MVM (41) have identified a minor splice, common to all transcripts, lying between nucleotides 2250 and 2450. It is significant that through this region, and for several hundred nucleotides on either side of it, MVM(p) and MVM(i) are more strongly conserved than in all other regions of the genome, as though additional constraints on random genetic drift apply to this region and not to others. The reason for this may be the potential use of a second open reading frame in this region, perhaps encoding part of NS-2, since the regions showing diminished drift between MVM(p) and MVM(i) delineate extensive dual open reading frames on either side of the minor splice. These dual open reading frames are conserved between MVM(p) (4), MVM(i) (this paper), H1 (46), and to some extent canine parvovirus (44).

We have noted previously (4) that the DNA sequence of MVM(p) around the minor splice region offers several close matches to consensus splice donor and acceptor splice sites (38), and we used both the conservation and divergence of sequence between MVM(p), MVM(i), H1, and canine parvovirus in this region to explore which splicing options are most likely used. We matched sequences from nucleotides 2251 through 2450 in MVM(i) with equivalent regions from MVM(p), H1, and canine parvovirus, using the University of Wisconsin Genetics Computer Group GAP program. This region from each of the four genomes was then searched individually for the 10 best fits to both donor and acceptor consensus sequences (38), using the University of Wisconsin Genetics Computer Group FITCONSENSUS program. Potential splice sites common to all four were then ranked (A, B, C, etc. for donors and 1, 2, etc. for acceptors) by their average quality of fit and are displayed on the MVM(p) sequence in Fig. 4. Three potential donor splice sites and two potential acceptor sequences conserved between the four genomes were identified in this way. (Note that acceptor site 1 was not recognized earlier [4] because our search at that time was restricted to TNCAG.) We can discount donor site C since it is downstream of both potential acceptor sequences. The DNA sequence downstream of acceptor site 2 has translation terminators in all three reading frames and so would not allow translation of the VP1-specific region of the long right-hand open reading frame. While acceptor 1 is a 12-of-16-base fit to the consensus sequence, acceptor 2 is only a 9-of-16-base fit and has an AG dinucleotide at positions −4, −3, and −11, −10, from the A6 of the potential splice site which is never seen within 14 nucleotides of a functional acceptor site (38). On the basis of this analysis, we would predict that VP1 is synthesized starting with the methionine at nucleotide 2287 in MVM(i), with the sequence MAPPAKRA, and that this transcript is spliced from donor A to acceptor 1, joining a 10-amino acid N-terminal segment onto the right-hand long open reading frame through a fused glycine codon. An alternatively spliced transcript in which donor B is joined to acceptor 1 would circumvent the initiator methionine at nucleotide 2287 and allow the synthesis of VP2 to occur by delaying translation initiation until the ribosome had reached the methionine at nucleotide 2795 in MVM(i). For completeness, in Fig. 3 a lineup of the right-hand open reading frames (frame 1) of MVM(i) (upper line) and MVM(p) (lower line) includes potential leader regions (frames 1, 2, and 3) discussed earlier (4). However, assuming that VP1 begins in frame 3 at position 2287 as discussed above and is spliced to frame 1 with the 5′ donor site A and 3′ acceptor site 1 (Fig. 4), there are in the VP-1 coding region 12 amino acid changes between MVM(i) and MVM(p). Again, conservative changes predominate: Gly→Ser, Ile→Val, Ala→Thr, Val→Ile, Arg→Lys, Ala→Ser, Asn→Ser, Arg→Lys, and Val→Ala. There are three radical changes: Ser→Leu at position 3272, Pro→Ser at position 3619, and Glu→Gly at position 3755. It is noteworthy that the amino-terminal region unique to VP1 (i.e., not translated in VP2 if VP2 does begin at 2795) is invariant between MVM(i) and MVM(p).

Comparison of transcription signals between the two genomes. Two promoters with overlapping transcription units have been characterized for MVM(p) (41). The TATAAG of the left promoter (at position 177) is identical in i and p, as is the TATAAA of the second, much stronger promoter at position 1977. Whether the nucleotide changes in the vicinities of these consensus sequences are functionally significant is not known.

MVM(p) C-strand



H-1 C-strand

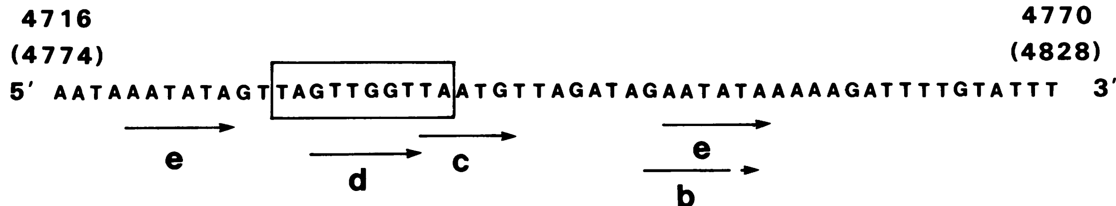


FIG. 5. Nucleotide sequences of the 65- and 55-bp repeats found in MVM(p) and H1. Arrows below the sequence correspond to short repeated sequences found within and between these tandem repeats from the MVM(p) and H1 genomes. The boxed regions are regions which share some homology with the consensus sequence for enhancers.

The most obvious difference between i and p in transcription signals involves potential polyadenylation sites, AATAAA. The second polyadenylation signal in MVM(p) at nucleotide 4755 (repeated at 4820) is mutated to GGTAAA in MVM(i) (nucleotide 4756). Because this is within the 65-bp repeat in MVM(p), there are four canonical polyadenylation signals in MVM(p), at nucleotides 4602, 4755, 4820, and 4885, compared with only two in MVM(i) at nucleotides 4603 and 4821.

DISCUSSION

We constructed clones of the genome of MVM(i), the immunosuppressive strain of MVM, and determined the complete nucleotide sequence of these clones. The sequence is virtually identical to that of MVM(p) (4) with the exception that the sequence of MVM(p) was found to contain a duplication of 65 bp which was missed in the original sequence (4). The presence of the 65-bp duplication in MVM(p) relative to MVM(i) explains the size difference between the *Xba*I-*Msp*I fragment (map units 86 to 99) observed by McMaster et al. (29) and confirmed by ourselves. It is noteworthy that a similar tandem repeat (of 55 bp) has been observed in another autonomous parvovirus genome (H1) as well as in defective interfering particles of H1 (45), though it is not present in the genome of a closely related virus, H3 (45). Both tandem repeats [MVM(p) and H1] (Fig. 5) contain a polyadenylation signal. The repeated sequence in H1 is 84% A+T in contrast with the repeat in MVM(p) which is only 66% A+T. When one compares the sequences of the 65-bp MVM(p) and 55-bp H1 repeats one of the most striking features is that on the complementary strand, the 65-bp repeat contains only one C residue while the 55-bp repeat does not contain a C. Within the MVM(p) 65-bp repeat there are several short direct repeats: AAG GGTGG (a), GAATAA (b), and TAATG (c). In H1 there is one short direct repeat, AATATA (e). Sequences b, c, and d (GTTGGT) are common to both the 65- and 55-bp repeat (Fig. 5).

The 65-bp direct repeats in the MVM(p) genome are of particular interest for three reasons. First, there is a polyadenylation signal within the repeat (hence two polyadenylation sites at 4755 and 4820; Fig. 1). In MVM(i), these sites are lacking due to a mutation of AATAAA → GGTAAA (nucleotide 4756; Fig. 1). Hence the MVM(i) genome has two fewer potential polyadenylation sites than does the MVM(p) genome. It is of possible interest that many genes possess multiple polyadenylation sites and that alternative processing of the 3′ ends of transcripts has been documented for several cellular genes as well as viral genes (1, 16, 23, 39, 47). While the factors that determine selection of a particular polyadenylation site are not well understood, it has been suggested that they probably reflect the growth state of the cell or the differentiated cell type (34). Second, Faust et al. (18) have recognized that the MVM genome contains 26 hexanucleotides, referred to as ψ, which conform to the consensus sequence $C_2A_{1-2}(C_{2-3}/T_2)$, and presented evidence that these sequences may be initiation sites for purified mouse DNA polymerase α primase in vitro. One of these sequences in the MVM(p) genome in the region of nucleotides 4762 to 4779 ($3'$-$C_3AC_2T_2C_3AC_2A_2C_2$-$5'$) is repeated at nucleotides 4827 to 4844. As yet, there is no evidence that ψ sequences are used by the DNA polymerase α primase in vivo (18). The third reason why the 65-bp repeats of MVM(p) are of interest is that they contain sequences related to transcriptional enhancer elements which have been identified in viral and cellular genes. Some of these enhancers show tissue- or cell-specificity (5, 12, 15, 21, 26, 37, 42, 62), although it has been reported recently that the simian virus 40 72-bp repeats do not alter the species specificity of polyomavirus (10). Hen et al. (22) have compared enhancer sequences from 10 genomes and have suggested a consensus sequence $GXTGTGG^{TTT}_{AAA}$. Weiher et al. (59) have suggested a core sequence of $(G)TGG^{TTT}_{AAA}(G)$ based on point mutations of the SV40 enhancer and comparison with other viral enhancers. Overlapping related sequences which occur within the MVM(p) 65-bp repeat are boxed in Fig. 5, as is a related sequence in the 55-bp repeat of H1. These MVM(p) sequences agree in 7 of 10 nucleotides with the former enhancer consensus sequence and in 7 of 8 nucleotides with the latter (viral) enhancer consensus sequence.

The structures of the 3′ and 5′ palindrome regions of the

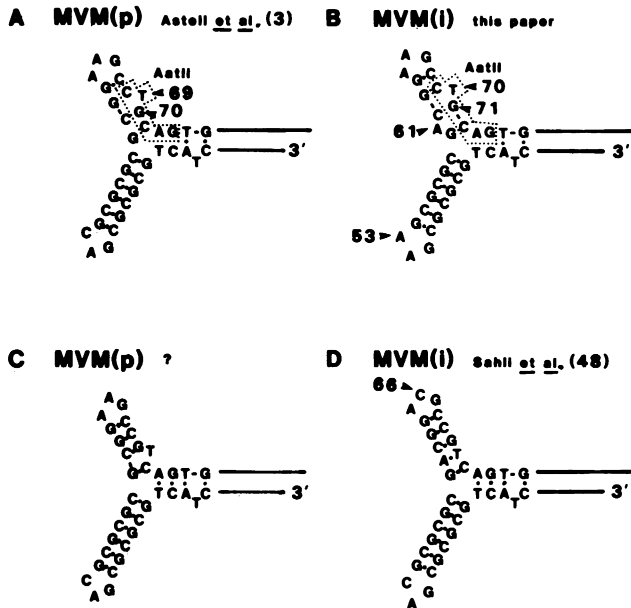


FIG. 6. Structure of the 3'-terminal hairpins from MVM(p) and MVM(i). The sequence determined for MVM(p) (3) is illustrated in Fig. 6A and that determined for MVM(i) (this paper) is given in Fig. 6B. Nucleotide differences between p and i are indicated by the arrowheads at positions 53 and 61. The sequence for MVM(i) determined by Sahli et al. (48) is illustrated in Fig. 6D and has an additional nucleotide change at position 66. Because we were not able to detect an *Aat*II site at nucleotide 69 in MVM(i) or 68 in MVM(p), we believe that the sequence of Sahli et al. (48) is more probably correct. The correct sequence for the 3' hairpin of MVM(p) is illustrated in Fig. 6C.

MVM(i) genome indicate that they are very similar to those of MVM(p). The 3' palindrome is identical, except for an additional T at position 61 and a G→T transversion at position 53 in MVM(i). This additional base (an A in viral-strand DNA) occurs in the short arm of the Y-shaped hairpin in viral-strand DNA (3) and thus may alter the stability of this arm. The G→T base change at position 53 (C→A change in viral-strand DNA) lies within the 3-nucleotide loop of the long arm of the Y-shaped hairpin and would not be expected to alter the stability of this structure (Fig. 6A and B). The 5' palindrome is also highly conserved. Assuming that the inferred sequence of the last 21 nucleotides is correct, there are six single base changes. Each is matched by a compensatory base change such that in the hairpin configuration of viral-strand DNA, the number of potential base pairs and hence stability of the hairpin would not be expected to be altered. The loop end of the hairpin is AAA (flip orientation) and TTT (flop orientation), which is opposite to that of the MVM(p) genome. It is interesting to speculate on how such a change might have occurred. Possible explanations are three single point mutations, a gene conversion event, a double crossover, a single crossover event at the level of viral-strand DNA, or a single crossover event between flip and flop turnaround forms of RF DNA, distal to the asymmetric unpaired region of the hairpin. Of the five we consider the last to be most likely.

Comparison of our sequence for MVM(i) with that of Sahli et al. (48). We compared our sequence for MVM(i) with that of Sahli et al. (48) and found 11 differences in base changes (Table 1). The AC versus CA at position 70 is most likely to be due to the problem of resolving the hairpin on gels, noted

TABLE 1. Differences in sequence between MVM(i) of Sahli et al. (48) and MVM(i) of this paper

Nucleotide position	MVM(i) of Astell et al.	MVM(i) of Sahli et al.	Coding changes
66	T	G	
70	A	C	
71	C	A	
90	T	C	
1566	C	G	
1572	G	A	
1962	A	G	
2050	A	C	Ile→Leu
2832	C	G	Ala→Gly
4390	C	A	
4698	A	G	

earlier for MVM(p) (3). At this point we are not sure of the correct order for these nucleotides. We observed both orders for MVM(p) viral-strand DNA (C. Astell and M. B. Chow, unpublished data). More frequently the order was that reported earlier (3) and illustrated here in Fig. 6A. Also, sequencing studies of the complementary strand of MVM(p) RF DNA agreed with this order (2). However, we consider the order reported by Sahli et al. (48) to be more probably correct, for the following reason. Repeated attempts to restrict both MVM(p) and MVM(i) clones at a predicted *Aat*II site at position 70, in this laboratory and in several other laboratories, have failed. Although the secondary structure in this region is a possible explanation for this problem, we must also consider the likelihood that the site simply does not exist, as is predicted by the Sahli sequence. This nucleotide reversal does not change the net number of base pairs in the short arm of the 3' hairpin of MVM(p), although it could possibly increase the stability of the structure. It increases by one the number of paired bases in the MVM(i) 3' hairpin, thereby significantly increasing its stability (Fig. 6B through D).

In addition to this difference, there are nine base change differences between the two sequences of MVM(i). Of the seven changes within the coding sequence, none would lead to radical amino acid differences. We found two changes which were not reported by Sahli et al. (48), and they found seven changes which we did not see.

A possible reason for the differences between our sequence of MVM(i) and that of Sahli et al. (48) is that we cloned from different MVM(i) stock preparations. Although both laboratories use stocks derived from the same genetically purified MVM(i) stock (52), our virus is routinely grown in S49.1 TB2 cells; in the case of the preparation from which the MVM(i) terminal plasmid clones were derived, the virus was grown in 324K cells. Sahli et al. (48) grew their preparation of MVM(i) in EL-4 lymphoma cells. Perhaps growth in these different host-cell types has allowed for some degree of genetic drift.

ACKNOWLEDGMENTS

We would like to thank M. Thomson for expert technical assistance.

The work was supported by grants from the BCHCRF and MRC (Canada) to C.R.A. and by Public Health Service grant CA29303 from the National Institutes of Health to P.T.

ADDENDUM

After submission of this report, Rhode (J. Virol. 55:886–889, 1985) reported that sequences in the H1 genome

including the 55-bp repeat do not enhance expression of chloramphenicol acetyltransferase activity.

LITERATURE CITED

1. **Alt, F. E., A. L. M. Bothwell, M. Knapp, E. Siden, E. Mather, M. Koshland, and D. Baltimore.** 1980. Synthesis of secreted and membrane-bound immunoglobulin mu heavy chains is directed by mRNAs that differ at their 3′ ends. Cell **20:**293–301.
2. **Astell, C. R., M. B. Chow, and D. C. Ward.** 1985. Sequence analysis of the termini of virion and replicative form of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus DNA replication. J. Virol. **54:**171–177.
3. **Astell, C. R., M. Smith, M. B. Chow, and D. C. Ward.** 1979. Structure of the 3′ hairpin termini of four rodent parvovirus genomes: nucleotide sequence homology at origins of DNA replication. Cell **17:**691–703.
4. **Astell, C. R., M. Thomson, M. Merchlinsky, and D. C. Ward.** 1983. The complete DNA sequence of minute virus of mice, an autonomous parvovirus. Nucleic Acids Res. **11:**999–1018.
5. **Banerji, J., L. Olson, and W. Schaffner.** 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell **33:**729–740.
6. **Ben Asher, E., and Y. Aloni.** 1984. Transcription of minute virus of mice, an autonomous parvovirus, may be regulated by attenuation. J. Virol. **52:**266–276.
7. **Boissy, R., and C. R. Astell.** 1985. *Escherichia coli recBC sbcB recF* hosts permit the deletion-resistant propagation of plasmid clones containing the 5′-terminal palindrome of minute virus of mice. Gene **35:**179–185.
8. **Bonnard, G. D., and R. B. Herberman.** 1975. Supression of lymphocyte proliferative responses by murine lymphoid cells, p. 817. *In* A. S. Rosenthal (ed.), Immune recognition. Academic Press, Inc., New York.
9. **Bourguignon, G. J., P. J. Tattersall, and D. C. Ward.** 1976. DNA of minute virus of mice: self-priming, nonpermuted, single-stranded genome with a 5′-terminal hairpin duplex. J. Virol. **20:**290–306.
10. **Campbell, B. A., and L. P. Villarreal.** 1985. Host species specificity of polyomavirus DNA replication is not altered by simian virus 40 72-base-pair repeats. Mol. Cell. Biol. **5:** 1534–1537.
11. **Campbell, D. A., Jr., S. P. Staal, E. K. Manders, D. G. Bonnard, R. K. Oldman, L. A. Salzman, and R. B. Herberman.** 1977. Inhibition of *in vivo* lymphoproliferative responses by *in vivo* passaged rat 13762 mammary adenocarcinoma cells. II. Evidence that Kilham rat virus is responsible for the inhibitory effect. Cell. Immunol. **33:**378–391.
12. **Chandler, V. L., B. A. Maler, and K. R. Yamamoto.** 1983. DNA sequence bound specifically by glucocorticoid receptor *in vitro* renders a heterologous promoter hormone responsive *in vivo*. Cell **33:**489–499.
13. **Cotmore, S. F., L. J. Sturzenbecker, and P. Tattersall.** 1983. The autonomous parvovirus MVM encodes two non-structural proteins in addition to its capsid polypeptides. Virology **129:** 333–343.
14. **Delaney, A.** 1982. A DNA sequence handling program. Nucleic Acids Res. **10:**61–67.
15. **DeVilliers, J. L., L. Olson, C. Tyndall, and W. Schaffner.** 1982. Transcriptional enhancers from SV40 and polyoma virus show a cell type preference. Nucleic Acids Res. **10:**7965–7976.
16. **Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall, and L. Hood.** 1980. Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell **20:**313–319.
17. **Engers, H. D., J. A. Louis, R. H. Zubler, and B. Hirt.** 1981. Inhibition of T cell-mediated functions by MVM(i), a parvovirus closely related to minute virus of mice. J. Immunol. **127:** 2280–2285.
18. **Faust, E. A., R. Nagy, and S. K. Davey.** 1984. Mouse DNA polymerase α-primase terminates and reinitiates DNA synthesis 2-14 nucleotides upstream of $C_2A_{1-2}(C_{2-3}/T_2)$ sequences on a minute virus of mice DNA template. Proc. Natl. Acad. Sci. USA **82:**4023–4027.
19. **Feunteun, J., and T. L. Benjamin.** 1982. Isolation of transformation-defective host-range mutants of polyoma virus on normal mouse cells. Virology **119:**310–316.
20. **Fields, B. N., and M. I. Greene.** 1982. Genetic and molecular mechanisms of viral pathogenesis: implications for prevention and treatment. Nature (London) **300:**19–23.
21. **Gillies, S. E., S. L. Morrison, V. T. Oi, and S. Tonegawa.** 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell **33:**717–728.
22. **Hen, R., E. Borrelli, P. Sassone-Corsi, and P. Chambon.** 1983. An enhancer element is located 340 base pairs upstream from the adenovirus-2 E1A capsite. Nucleic Acids Res. **11:** 8747–8760.
23. **Henikoff, S., J. S. Sloan, and J. D. Kelly.** 1983. A *Drosophila* metabolic gene transcript is alternatively processed. Cell **34:**405–414.
24. **Holland, J. J.** 1961. Receptor affinities as major determinants of enterovirus tissue tropisms in humans. Virology **12:**312–326.
25. **Kilham, L., and G. Margolis.** 1966. Spontaneous hepatitis and cerebellar hypoplasia in suckling rats due to congenital infections with rat virus. Am. J. Pathol. **49:**457–475.
26. **Kriegler, M., and M. Botchan.** 1983. Enhanced transformation by a simian virus recombinant virus containing a Harvey murine sarcoma virus long terminal repeat. Mol. Cell. Biol. **3:**325–339.
27. **Lum, G. S.** 1970. Serological studies of rat viruses in relation to tumors. Oncology **24:**335–343.
28. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74:**560–564.
29. **McMaster, G. K., P. Beard, H. D. Engers, and B. Hirt.** 1981. Characterization of an immunosuppressive parvovirus related to the minute virus of mice. J. Virol. **38:**317–326.
30. **Merchlinsky, M., P. J. Tattersall, J. J. Leary, S. F. Cotmore, E. M. Gardiner, and D. C. Ward.** 1983. Construction of an infectious molecular clone of the autonomous parvovirus MVM. J. Virol. **47:**227–232.
31. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9:**309–321.
32. **Miller, R. A., D. C. Ward, and F. H. Ruddle.** 1977. Embryonal carcinoma cells (and their somatic cell hybrids) are resistant to infection by the murine parvovirus MVM, which does infect other teratocarcinoma-derived cell lines. J. Cell. Physiol. **91:**393–402.
33. **Mohanty, S. B., and P. A. Bachman.** 1974. Susceptibility of fertilized mouse eggs to minute virus of mice. Infect. Immun. **9:**762–763.
34. **Moore, C. L., and P. A. Sharp.** 1984. Site-specific polyadenylation in a cell-free reaction. Cell **36:**581–591.
35. **Morishima, T., P. R. McClintock, G. S. Aulakh, L. C. Billups, and A. L. Notkins.** 1982. Genomic and receptor attachment differences between Mengovirus and encephalomyocarditis virus. Virology **122:**461–465.
36. **Morishima, T., P. R. McClintock, L. C. Billups, and A. L. Notkins.** 1982. Expression and modulation of virus receptors on lymphoid and myeloid cells: relationship to infectivity. Virology **116:**605–618.
37. **Mosthaf, L., M. Pawlita, and P. Gruss.** 1985. A viral enhancer element specifically active in human hematopoietic cells. Nature (London) **315:**597–600.
38. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10:**459–472.
39. **Nevins, J. R., and S. Chen-Kiang.** 1981. Processing of adenovirus nuclear RNA to mRNA. Adv. Virus Res. **26:**1–35.
40. **Paradiso, P. R., K. R. Williams, and R. L. Costantino.** 1984. Mapping of the amino terminus of the H-1 parvovirus major capsid protein. J. Virol. **52:**77–81.
41. **Pintel, D., D. Dadachanji, C. R. Astell, and D. C. Ward.** 1983. The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units. Nucleic Acids Res. **11:**1019–1038.
42. **Queen, C., and D. Baltimore.** 1983. Immunoglobulin gene tran-

scription is activated by downstream sequence elements. Cell **33**:741–748.

43. **Rhode, S. L., III.** 1973. Replication process of the parvovirus H-1. I. Kinetics in a parasychronous cell system. J. Virol. **11**:856–861.
44. **Rhode, S. L., III.** 1985. Nucleotide sequence of the coat protein gene of canine parvovirus. J. Virol. **54**:630–633.
45. **Rhode, S. L., III, and B. Klassen.** 1982. DNA sequence of the 5′ terminus containing the replication origin of parvovirus replicative form DNA. J. Virol. **41**:990–999.
46. **Rhode, S. L., III, and P. R. Paradiso.** 1983. Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. J. Virol. **45**:173–184.
47. **Rosenfeld, M. G., J.-J. Mermod, S. G. Amara, L. W. Swanson, P. E. Sawchenko, J. Rivier, W. W. Vale, and R. M. Evans.** 1983. Production of a novel neuropeptide encoded by the calcitonin gene via tissue-specific RNA processing. Nature (London) **304**:129–135.
48. **Sahli, R., G. K. McMaster, and B. Hirt.** 1985. DNA sequence comparison between two tissue-specific variants of the autonomous parvovirus, minute virus of mice. Nucleic Acids Res. **13**:3617–3632.
49. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
50. **Siegl, G., and M. Gautschi.** 1973. The multiplication of parvovirus LuIII in a synchronized culture system. I. Optimum conditions for virus replication. Arch. Gesamte Virusforsch. **40**:105–118.
51. **Smith, A. L., and G. H. Tignor.** 1980. Host cell receptors for two strains of Sindbis virus. Arch. Virol. **66**:11–26.
52. **Spalholz, B. A., and P. Tattersall.** 1983. Interaction of minute virus of mice with differentiated cells: strain-dependent target cell specificity is mediated by intracellular factors. J. Virol. **46**:937–943.
53. **Tattersall, P.** 1972. Replication of the parvovirus MVM. I. Dependence of virus multiplication and plaque formation on cell growth. J. Virol. **10**:586–590.
54. **Tattersall, P.** 1978. Susceptibility to minute virus of mice as a function of host-cell differentiation, p. 131–149. *In* D. C. Ward and P. Tattersall (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
55. **Tattersall, P., and J. Bratton.** 1983. Reciprocal productive and restrictive virus-cell interactions of immunosuppressive and prototype strains of minute virus of mice. J. Virol. **46**:944–955.
56. **Tattersall, P., A. J. Shatkin, and D. C. Ward.** 1977. Sequence homology between the structural polypeptides of minute virus of mice. J. Mol. Biol. **111**:375–394.
57. **Tattersall, P., and D. C. Ward.** 1978. The parvoviruses: an introduction, p. 3–12. *In* D. C. Ward and P. Tattersall (ed.), Replication of mammalian parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
58. **Tennant, R. W., K. R. Layman, and R. E. Hand, Jr.** 1969. Effect of cell physiological state on infection by rat virus. J. Virol. **4**:872–878.
59. **Weiher, H., M. König, and P. Gruss.** 1983. Multiple point mutations affecting the simian virus 40 enhancer. Science **219**:626–631.
60. **Weiner, H. L., K.A. Ault, and B. N. Fields.** 1980. Interaction of reovirus with cell surface receptor. I. Murine and human lymphocytes have a receptor for the hemagglutinin of reovirus type 3. J. Immunol. **124**:2143–2148.
61. **Weiner, H. L., D. Drayna, D. R. Averill, and B. N. Fields.** 1977. Molecular basis of reovirus virulence: role of the S1 gene. Proc. Natl. Acad. Sci. USA **74**:5744–5748.
62. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
63. **Yaniv, M.** 1982. Enhancing elements for activation of eukaryotic promotors. Nature (London) **297**:17–18.