JOURNAL OF VIROLOGY, Aug. 1988, p. 2903–2915
0022-538X/88/082903-13$02.00/0

Vol. 62, No. 8

# Nucleotide Sequence and Genomic Organization of Aleutian Mink Disease Parvovirus (ADV): Sequence Comparisons between a Nonpathogenic and a Pathogenic Strain of ADV

MARSHALL E. BLOOM,* SOREN ALEXANDERSEN, SYLVIA PERRYMAN, DAVID LECHNER, AND JAMES B. WOLFINBARGER

18 February 1988/Accepted 2 May 1988

**A DNA sequence of 4,592 nucleotides (nt) was derived for the nonpathogenic ADV-G strain of Aleutian mink disease parvovirus (ADV). The 3′ (left) end of the virion strand contained a 117-nt palindrome that could assume a Y-shaped configuration similar to, but less stable than, that of other parvoviruses. The sequence obtained for the 5′ end was incomplete and did not contain the 5′ (right) hairpin structure but ended just after a 25-nt A+T-rich direct repeat. Features of ADV genomic organization are (i) major left (622 amino acids) and right (702 amino acids) open reading frames (ORFs) in different translational frames of the plus-sense strand, (ii) two short mid-ORFs, (iii) eight potential promoter motifs (TATA boxes), including ones at 3 and 36 map units, and (iv) six potential polyadenylation sites, including three clustered near the termination of the right ORF. Although the overall homology to other parvoviruses is <50%, there are short conserved amino acid regions in both major ORFs. However, two regions in the right ORF allegedly conserved among the parvoviruses were not present in ADV. At the DNA level, ADV-G is 97.5% related to the pathogenic ADV-Utah 1. A total of 22 amino acid changes were found in the right ORF; changes were found in both hydrophilic and hydrophobic regions and generally did not affect the theoretical hydropathy. However, there is a short heterogeneous region at 64 to 65 map units in which 8 out of 11 residues have diverged; this hypervariable segment may be analogous to short amino acid regions in other parvoviruses that determine host range and pathogenicity. These findings suggested that this region may harbor some of the determinants responsible for the differences in pathogenicity of ADV-G and ADV-Utah 1.**

In recent years, the autonomous parvoviruses have become recognized as major etiological agents of a variety of animal and human diseases (6, 16, 31, 44, 69, 72). Most of the parvoviruses cause an acute disease picture the pathogenesis of which can be directly related to sites of viral replication, but the Aleutian mink disease parvovirus (ADV) is somewhat unique in that it causes both a severe acute pneumonia in newborn kits (1, 2, 4, 5) as well as a chronic disorder of the immune system in adult mink (3, 16, 20, 21, 38, 58, 60). Although every ADV strain tested induces pneumonia in mink kits (1, 5), strains differ markedly in pathogenicity or virulence for adult mink (16, 20, 21, 38, 58–60). The virulent ADV-Utah 1 strain (60) causes full-blown Aleutian disease in all genotypes of adult mink (16, 21, 22, 60), whereas the ADV-G strain, which was derived in cell culture from ADV-Utah 1, has lost pathogenicity for mink (20). Consequently, it has been recognized that structural differences at a genomic level may have an important role in determining the virulence of ADV. Previous work demonstrates differences between pathogenic and nonpathogenic strains of ADV (17, 18), but the relationship of these differences to virulence is uncertain.

In other parvovirus models, elucidation of the viral DNA sequence has provided an important framework for analyzing how viral genes or their products might affect the outcome of virus host interactions (31, 69, 70). The complete nucleotide sequence for several autonomous parvoviruses has been determined and indicates several common features (9, 26, 29, 52, 62, 65, 66, 68, 71). The linear single-stranded genomes are all approximately 5,000 nucleotides (nt) in length and contain palindromic sequences at both 5′ and 3′ termini. All the parvoviruses sequenced to date have two major open reading frames (ORFs). A left ORF that is governed by a promoter at ~4 map units (MU) (12, 23, 27, 30, 42, 55, 57, 65) specifies at least one nonstructural protein necessary for viral replication and gene regulation, and a right ORF, the promoter for which lies at ~38 to 40 MU, encodes the sequences for a set of overlapping structural or capsid proteins (12, 18, 21, 27, 31, 42, 46, 49, 53, 55, 57, 64, 65, 68).

Sequence data has also provided interesting information on the interrelationships among the various parvoviruses. For example, although minute virus of mice (MVM) (9, 52, 66), bovine parvovirus (BPV) (29), B19 parvovirus (68), and the dependovirus adeno-associated virus (AAV) (71) share little overall homology, several short stretches of amino acids in both structural and nonstructural proteins seem to be conserved. These findings suggest an important role for these regions, although their function is not as yet clear.

In this report, we describe the DNA sequence of the nonpathogenic ADV-G strain of ADV and the basic genomic organization that is evident. In common with the other autonomous parvoviruses, the ADV genome contains large left and right ORFs, and there are promoterlike sequence motifs located at 3 and 36 MU. Furthermore, although the overall levels of DNA and putative amino acid homology to several representative members of the parvovirus family are low, there are highly conserved regions in both the left and right ORFs. In addition, several features of the ADV genome seem to be significantly different from the other parvoviruses and these unusual features may be partly responsible for the unique pathogenic properties of ADV. We also report partial sequence data for the highly pathogenic ADV-Utah 1 strain, and although the two strains of ADV show very high homology, we found an 11-amino-acid area in the ORF for

* Corresponding author.

the capsid proteins that shows marked divergence between the two strains. This region may be partly responsible for the marked differences in pathogenicity seen between these two ADV strains.

## MATERIALS AND METHODS

**Materials.** All enzymes were bought from either Bethesda Research Laboratories, Inc. (BRL), Gaithersburg, Md., New England BioLabs, Inc., Beverly, Mass., or International Biotechnologies, Inc., New Haven, Conn., and were used according to the recommendations of the manufacturers. Vector DNAs were purchased from either BRL (M13mp18 and M13mp19) (74) or Promega-Biotec (pGEM3) (51), Madison, Wis. pEMBL8 DNA (34) was a gift from Francis Nano. $^{32}$P- and $^{35}$S-labeled deoxynucleotide triphosphates were bought from Du Pont-New England Nuclear Corp., Boston, Mass. Unlabeled deoxy- and dideoxynucleotide triphosphates were obtained from Pharmacia-P-L Biochemicals, Piscataway, N.J. Synthetic nucleotide primers were either purchased (BRL or Promega-Biotec) or were prepared on an oligonucleotide synthesizer (SAM-1; Biosearch, San Rafael, Calif.). These latter oligomers (generally 20 bases in length) were separated from failure sequences either by preparative electrophoresis in denaturing polyacrylamide gels or by high-performance liquid chromatography.

**Viruses, cells, and viral DNA.** The ADV-G strain (20) of ADV was propagated in Crandell feline kidney cells (CRFK) as previously described (4, 20) except that no antibiotics were used in the growth medium. Duplex monomer replicative-form (RF) DNA was isolated by a modified Hirt procedure (19); however, the denaturation step for the enrichment of covalently linked hairpin forms was omitted. Instead, the partially purified RF DNA was resolved on a low-melting-temperature agarose gel and isolated on NACS columns (BRL). Alternatively, the RF DNA was purified by two cycles of agarose gel electrophoresis and electroelution, using an Elutrap apparatus (Schleicher & Schuell, Inc., Keene, N. H.). Single-stranded virion DNA was prepared as previously noted (17, 19, 20).

**Molecular cloning techniques.** Unless specifically noted, the machinations employed for molecular cloning of DNA, agarose gel electrophoresis, restriction endonuclease digestions, and large-scale preparation of plasmid DNA were the same as previously used (2, 17, 22, 48, 49). Colony hybridization was performed by a modification of an alkaline blotting procedure (63); colony replicas were transferred to 82-mm (0.45 μm pore size) nylon filters (Hybond-N; Amersham Corp., Arlington Heights, Ill.) and were laid colony side up on Whatman 3MM filter paper saturated with 0.5 M NaOH. After 5 min, the filters were blotted briefly on dry 3MM paper and subsequently were replaced on the 0.5 M NaOH-saturated paper for an additional 5 min. After a brief rinse in 2× SSPE (48)–0.1% sodium dodecyl sulfate, the filters were reacted with the appropriate ADV DNA or RNA probes (17, 22).

The purified *Bam*HI-*Hin*dIII (15 to 88 MU) fragment of pADV-G (17) was subcloned into M13mp18 in *Escherichia coli* JM109 [*recA1 endA1 gyrA96 thi hsdR17 supE44 relA1* Ω− Δ(*lac-proAB*) F′*traD36 proAB ΔlacI*$^q$*ZM15*) (74); the sense of the M13 templates from these recombinant bacteriophages was plus in relation to the ADV sequences. A series of M13 clones processively deleted (32) from the *Hin*dIII site were derived from this recombinant bacteriophage by using a commercially available kit (Cyclone; International Biotechnologies). The data obtained from these templates provided the initial sequence information. Attempts to develop minus-sense single-stranded template molecules by cloning this same DNA segment in M13mp19 did not yield stable full-length templates.

In other experiments, ADV-G RF DNA was dC tailed by using terminal deoxytransferase (48), and it was annealed to pEMBL8 that had been previously linearized with *Pst*I and dG tailed (a generous gift from Bruce Chesebro). The hybrid molecules were transformed into strain JM109, and carbenicillin-resistant colonies were screened for the presence of ADV sequences by colony hybridization. The plasmid clone containing the largest ADV insert (pADVG IQ-6) produced deleted molecules on batch propagation in JM109 and was subsequently transferred into a *recA* derivative of *E. coli* MM294 [*recA* F$^-$ *endA1 hsdR17* (r− m+) *supE44*, *thi-1* Ω−; obtained from Francis Nano] (40). The generation of deleted plasmid molecules in this strain was greatly reduced, and pADVG IQ-6 DNA prepared from these bacteria served as template for most of the sequences reported in this article. Physical mapping indicated that with the exception of the right-hand end, this DNA was the same size as ADV-G RF DNA isolated by Hirt extraction (19). Subsequently, however, sequence comparisons with other clones indicated that 18 base pairs (bp) (nt 2469 to 2487) were missing from this clone. Therefore, three other clones derived from the tailing experiments were prepared for analysis of sequences bordering this deletion, and all three contained the 18 bp which were included in the presented sequence. DNA from the same three clones was used to confirm the 3′ (left) terminal sequence.

Additional attempts to develop clones containing the extreme 5′ (right) terminus were made by adding *Eco*RI linkers (48) to ADV-G virion DNA that had been made double stranded with the Klenow fragment of *E. coli* DNA polymerase (17, 19, 48). After several cycles of *Eco*RI digestion and microdialysis, the DNA was ligated into *Eco*RI-digested pGEM3 DNA and transformed into *E. coli* DB1256 (*recA recB21 recC22 sbcB15 hsdR* F$^-$ *proA2 his-4 thi-1 argE3 lacY1 galK2 ara-14 xyl-5 mtl-1 str-31 tsx-33*; obtained from Grant McFadden) (33). This strategy yielded clones extending from the ADV *Eco*RI site at 53 MU (17, 19, 49) rightward in the direction of 100 MU. No clones representing the portion from 0 to 53 MU were obtained. Selected clones were analyzed by restriction mapping and plasmid DNA sequencing.

The ADV-Utah 1 DNA used for sequence determination was the previously described recombinant plasmid containing the *Bam*HI-*Hin*dIII fragment (15 to 88 MU) of ADV-Utah 1 molecularly cloned into pUC18 (17).

**DNA sequencing techniques.** All sequencing was performed by the dideoxy-chain terminating method (67) with either $^{32}$P- or $^{35}$S- (14, 74) labeled deoxynucleotides as label. Templates for sequencing reactions were either single-stranded bacteriophage DNA isolated from M13 clones (74) or, in most cases, purified double-stranded plasmid DNA (28). All M13 bacteriophage-derived sequences were confirmed by sequencing plasmid DNA, and almost all regions were sequenced at least twice in both directions. Initially, the appropriate commercially available primers were used, but most of the reactions were performed with ADV-specific oligonucleotide primers, the sequences of which were based upon ADV DNA sequences that were already obtained. Utilization of these ADV-specific primers made it possible to obtain extended sequence information on a single template preparation. In some instances, sequencing reactions were

run using a deaza-nucleotide kit (American Bionetics, Inc., Emeryville, Calif.) to minimize secondary structure. For sequencing the extreme left (3') end, ADV-G RF DNA was included in a standard double-stranded dideoxy sequencing reaction, using an ADV-specific 20-nt primer complementary to nt 122 to 142 of the plus strand.

**DNA sequence analysis.** DNA sequences were analyzed on a personal computer (IBM XT; International Business Machines, Inc., Boca Raton, Fla.), using the Microgenie sequence analysis program package (Beckman Instruments, Inc., Palo Alto, Calif.). Homology comparisons of coding regions (ORFs) and terminal regions were performed by using the Alignment function of the Analysis mode of this program. The melting temperatures ($T_m$ of 3' hairpins were calculated by the method of Wetmur and Davidson (48, 73). Published sequences of BPV (29), feline panleukopenia virus (FPV) (26), and the human parvovirus B19 (68) were graciously provided on computer disks by Muriel Lederman, Jon Carlson, and Carol Astell, respectively. Other sequence data (MVM [9] and AAV [71] were contained in the data base of the Microgenie program.

## RESULTS

**Determination of ADV-G sequence.** The primary aim of these experiments was to determine the genomic nucleotide sequence of the ADV-G strain (20) of ADV. Using as templates the DNAs detailed in Materials and Methods, we obtained 4,592 bases of DNA sequence. We have adopted the convention that defines the 3' end of the minus-sense virion strand as the zero map position or left-hand end (8, 70); thus, nt 1 would be the first base at the 5' end of the plus-sense strand. The sequence is given in Fig. 1 for the plus strand.

The genomic termini of all parvoviruses contain self-complementary sequences capable of forming hairpin structures. The sequences at the left-hand end (3' end of the minus-sense virion strand) can be arranged in a Y-shaped configuration (31), and those at the right-hand end can be arranged in a simple U-shaped structure (31). The left-hand or 3' terminus of ADV-G contains a palindrome that can be arranged as shown in Fig. 2. The first 9 bases were deleted from the longest molecular clones, and that sequence was obtained by direct dideoxy sequencing of purified RF DNA. The most stable configuration of the 3' palindrome was compared with that described for MVM (Fig. 2) (9, 31, 66). Several features were noteworthy: (i) a 7-bp A+T-rich extreme 3' terminus, (ii) a duplex stem (29 bp in ADV-G and 44 bp in MVM), (iii) a short mismatched gap or bubble in the duplex stem (nt 20 to 21 with 95 to 97 in ADV-G; nt 25 to 26 with 88 to 90 in MVM), and (iv) one G+C-rich arm of 5 to 6 bp (nt 76 to 86 in ADV-G; nt 60 to 71 in MVM). The other arm in the observed ADV-G sequence was considerably longer, less highly base paired and G+C poorer than that for MVM. This same left-hand end region was sequenced in a total of four separate clones, and none gave a different sequence than that shown. Taken together, these data imply that the structure of the 3' terminus of ADV-G is very similar but not identical to that of MVM and other autonomous parvoviruses. Furthermore, the calculated $T_m$ (48, 73) of the palindrome observed for ADV-G (37.3°C) is significantly less than that for MVM (48.7°C). Thus, the 3' hairpin of ADV-G is less stable than that of MVM.

All ADV-G clones containing extreme right-hand end sequences were approximately 200 bp shorter than ADV-G RF DNA (data not shown). This truncation mapped to the right of the *Hin*dIII site at nt 4170. The same result was obtained with clones produced in DB1256 (33), a strain with a *recB recC sbc* genotype (24). Because we were unable to arrange our right-hand terminal sequence in a hairpin, we concluded that most if not all of the palindrome at that end was absent in the clones we have obtained to date. As a result, we were unable to determine whether the 5' and 3' termini are different, as in the case of MVM (9, 31, 66), or the same, as in the case of B19 parvovirus (68) and AAV (71). We observed an imperfect A+T-rich direct repeat of 25 nt that ended at nt 4591:

```
4469 ATAAACCTACATTCTATACTATCTA 4493
     *** **  **** ******** ***
4569 ATATACTAACAT-CTATACTA-CTA 4591
```

Similar repeats have been found near the right-hand terminus of several other autonomous parvoviruses (31).

Summing the 4,592 bases of DNA with the presumed 200 bases missing at the right end gives a size of ~4,792 nt, very close to previous estimates (19). This makes ADV-G several hundred bases shorter than any of the other nondefective parvoviruses [MVM(i), 5,085 nt] (9), but ~100 nt longer than the dependovirus AAV (4,681 nt) (11, 71). On the basis of these considerations, we used 4,800 nt as the denominator when expressing genome coordinates in map units.

**Identification of potential promoters and transcriptional control elements.** Analysis of the sequences of other autonomous parvoviruses suggests the existence of two functional promoters (10, 45), referred to as TATA boxes, at 4 and at 38 to 40 MU (12, 23, 27, 30, 31, 42, 46, 55, 57, 65). We identified eight potential TATA boxes (TATAA and TATTAA) in ADV-G (Fig. 1 and 3). Speculating that ADV-G had two functional promoters at ~4 MU and at ~38 to 40 MU, analogous to those of other parvoviruses, the nt 154 TATAA (3 MU) and the nt 1729 TATTAA(36 MU) are the most likely candidates.

Six repeats of the sequence motif AATAAA, which generally occurs 10 to 30 nt upstream from the start of the 3' poly(A) tracts in mRNA (15, 50), were identified (Fig. 1 and 3). The coordinates are nt 665, 818, 2546, 4136, 4394, and 4468. Several of these motifs are clustered at the right end of the genome, and one occurs near the middle of the genome, as described for other parvoviruses (11, 29, 31, 55, 71).

Sequence elements in addition to TATA boxes and AA TAAA motifs are essential in determining functional promoters and polyadenylation sites (10, 15, 43, 45, 50). We have presented a detailed analysis of this material in a manuscript dealing with the ADV transcription program (S. Alexandersen, M. E. Bloom, and S. Perryman, submitted for publication) and thus, have deferred detailed discussion of those points in this report. For similar reasons, we have not discussed potential splice donors and acceptors (29, 54).

**Definition of ORFs in the ADV-G genome.** The structural organization of parvovirus genomes reveals that all have two major ORFs. (By informal convention [31, 70], the capsid proteins have been designated from the largest to the smallest as VP1, VP2, VP3, and the nonstructural proteins are designated NS-1 and NS-2. For simplicity, we will conform to these designations.) The left ORF contains the sequences for the largest of the nonstructural proteins, NS-1 (26, 30, 31), and the right ORF codes for the bulk of the sequences contained in the structural or capsid proteins (31). In addition, several smaller ORFs seem to be a common feature and likely encode portions of additional NS proteins (30, 31). All of these coding regions occur on the strand complementary to the virion DNA, i.e., the plus-sense strand (31).

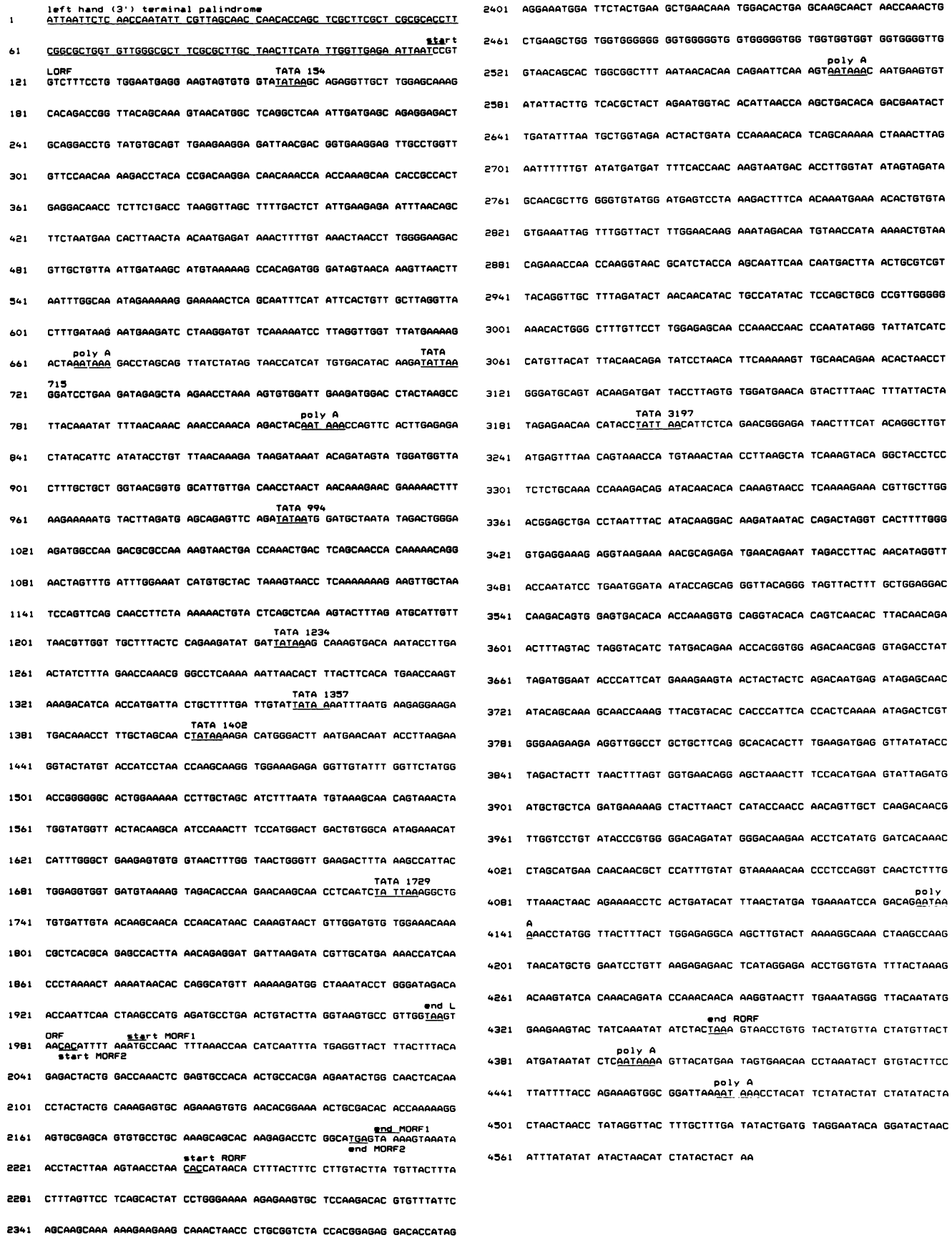


FIG. 1. Derived DNA sequence of the plus strand of ADV-G. Several features are indicated: the extent of the terminal palindrome at the left-hand end (i.e., 3′ end of virion strand); the positions of potential promoterlike TATA sequences; the locations of AATAAA polyadenylation signals; and the boundaries of the right, left, and the two mid- (MORF1 and MORF2) ORFs.

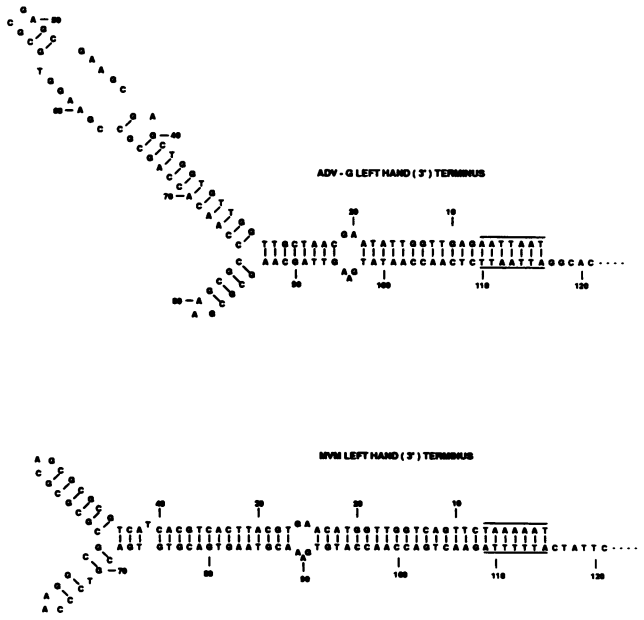


FIG. 2. Structures of the terminal palindrome sequences at the left or 3' end of the virion or minus strand of ADV-G and MVM. The sequences have been arranged to show maximal base pairing, that for MVM is based on that reported for the prototypic strain of MVM, MVM(p) (9, 31). The A+T-rich extreme 3' termini and their complements are indicated by the solid lines.

When the sequence of the ADV-G plus strand was translated in the three possible frames beginning with nt 1, 2, or 3, respectively, a pattern similar to that described above was obtained. A diagram depicting the location of the ORFs along with the respective ATGs and stop signals is depicted in Fig. 3. Those ATGs embedded in the most common context for initiation (ANNATGG) (45) are signified. The location of restriction enzymes with one or two recognition sites is also displayed in Fig. 3, and a partial list of enzymes without recognition sites in ADV-G is listed in the legend to that figure.

The left (left ORF), right (right ORF), and the 2 mid- (MORF1 and MORF2) ORFs were translated into amino acids; the protein sequences along with estimated molecular masses and nucleotide coordinates are shown in Fig. 4. In addition, the boundaries of the ORFs within the actual sequence are indicated in Fig. 1.

The left ORF is 1,859 nt in length and has a theoretical molecular mass of 70,927 daltons. This value is close to that reported for the 71,000-dalton mass of the ADV-G NS-1, p71 (5, 21). The right ORF is 2,105 nt long and could specify a protein of 79,970 daltons, a value intermediate to the two structural polypeptides of ADV-G (5, 20, 21). The 2 mid-ORFs could encode polypeptides less than 10,000 daltons in size.

**Relationship of ADV-G to other parvoviruses.** In order to investigate the relationship of ADV-G to several other members of the parvovirus family, we compared the homologies of the various ORFs at both the DNA and amino acid level. In addition, we also searched the first 200 nt to see whether significant homologies existed for the left-hand palindrome. None of these comparisons (AAV, B19 parvovirus, BPV, MVM, and FPV) yielded a relatedness of greater than 50% (data not shown), thus suggesting that ADV is not closely related to any of these parvoviruses.

The left ORF of the other parvoviruses contains a conserved domain at ~30 MU of approximately 60 amino acids referred to as the GKRN region (29, 31, 68). When the theoretical translations of the left ORFs of ADV-G and MVM are aligned, the presence of this GKRN element in the ADV sequence (31 MU) is evident (Fig. 5A). Several short stretches within this domain show extensive amino acid sequence conservation among ADV-G, MVM (Fig. 5A), and the other parvoviruses (data not shown), although the overall relatedness to MVM of the entire GKRN region is only 52%.

The right parvoviral ORFs have been alleged to contain a total of six conserved amino acid regions (29, 31, 68). In acronymic form, these have been designated NPYL, TPW, PIW, PGY, GGG, and YNN (29). Not all of these domains could be found in the ADV-G sequence we obtained. Regions that likely corresponded to the GGG, TPW, YNN, and the PIW elements were characterized and identified in an alignment of the ADV-G and MVM right ORFs (Fig. 5B). The amino acid homologies of these regions to the corresponding ones of MVM are: GGG, 73%; TPW, 78%; YNN, 76%; and PIW, 67%. Amino acid sequences corresponding to the NPYL or the PGY motifs could not be clearly identified in the right ORF of ADV-G. In fact, they could not be found in any of the three theoretical translations of the entire ADV-G sequence, suggesting that they are not features of ADV-G.

**Comparisons of ADV-G sequence to partial sequence of ADV-Utah 1.** The ADV-G strain is a cell culture derivative of the highly virulent ADV-Utah 1 strain that lost pathogenicity for adult mink after serial passage in CRFK cells (20). We previously cloned the 15- to 88-MU segment (*Bam*HI-*Hin*dIII fragment) of ADV-Utah 1 directly from DNA isolated from the tissues of infected mink (17). Using the ADV-specific primers developed for sequencing ADV-G, we also sequenced this segment of ADV-Utah 1 and compared it with the corresponding ADV-G segment (nt 721 to 4176). The ADV-Utah 1 *Bam*HI-*Hin*dIII segment contains 3,454 nt (1 less than in the corresponding segment of ADV-G), and the overall relatedness to ADV-G at the DNA level is very high (97.5%). Assuming the overall genomic lengths of both viruses are ~4,800 nt, this comparison encompasses approximately 76% of the genomes. A total of 73 nucleotide differences were observed between the two strains (Fig. 6). One short segment in the right ORF (64 to 65 MU) shows marked heterogeneity between ADV-G and ADV-Utah 1; this region, bounded by nt 3094 and 3130, contains 16 base changes, a single-base deletion at nt 3094, and a single-base addition at nt 3112 (ADV-G coordinates).

Because we had determined the ORFs of ADV-G, it was of interest to examine the ORFs of the two ADV strains. The ADV-Utah 1 sequence ends at the *Hin*dIII site at amino acid 645 of the right ORF; as a result, the extreme 3' portion of this ORF was not available for analysis. Nevertheless, we compared this sequence with the first 645 residues of the ADV-G right ORF; an analysis that encompasses 92% of the 702 residues in the ADV-G right ORF. The theoretical sizes for these truncated right ORF translation products are 73,388 and 73,516 daltons for ADV-G and ADV-Utah 1, respectively. The putative proteins are 95.8% related and have 15 isolated single-amino-acid changes, several of which are located in the conserved regions discussed above (Fig. 7). In addition, the region at 64-65 MU alluded to in the previous paragraph exhibits several singular features in the putative protein translations (Fig. 6 and 7). In a short cluster of 11 amino acids positioned just downstream from the YNN

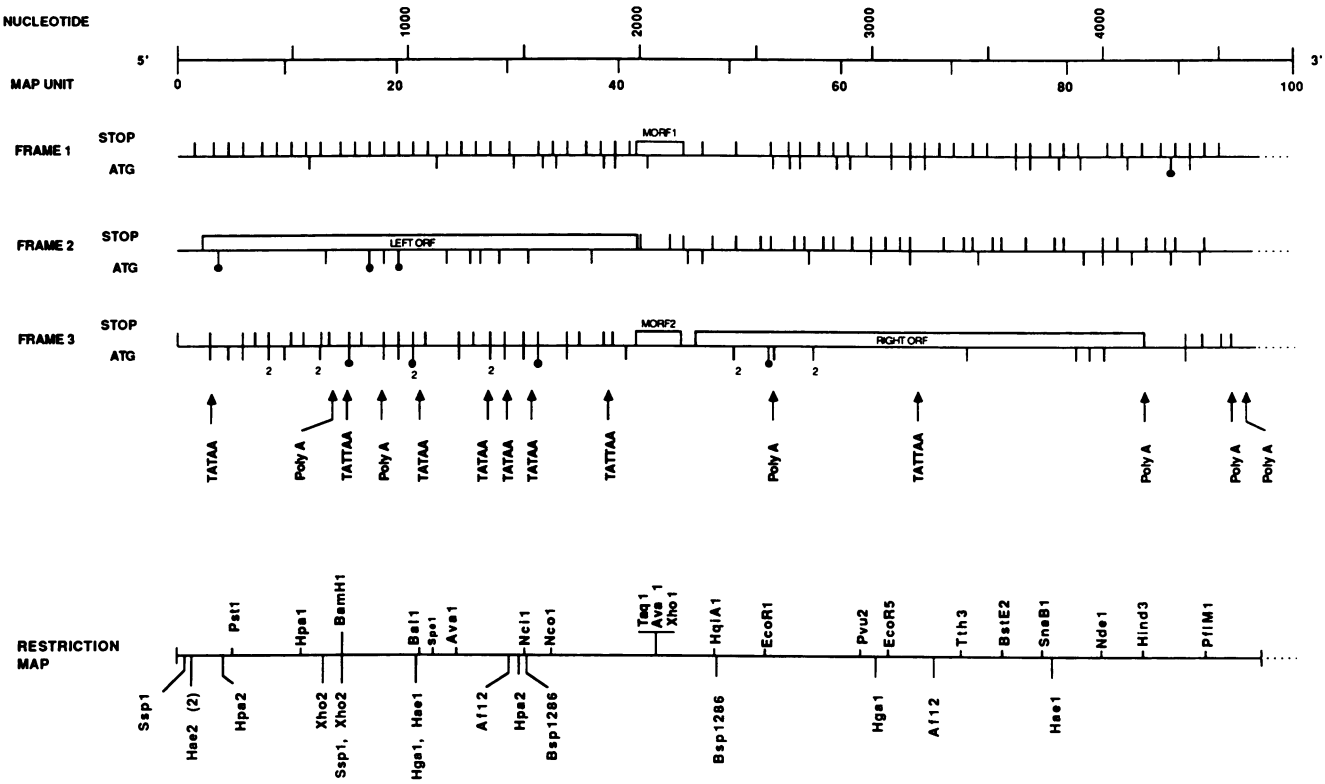


FIG. 3. Schematic representation of the genomic organization of ADV-G. A line diagram was prepared that displayed the major features of the ADV-G genome, including the ORFs and potential transcription-translation control signals in ADV-G. The location of termination codons (STOP) in the three potential reading frames is indicated by vertical marks above a horizontal line for each frame; some of these lines represent multiple STOPS too close to depict graphically. The boundaries of the ORFs are denoted by open boxes. Vertical lines below the horizontal lines indicate potential initiation (ATG) codons, and those ATGs in an optimal context for initiation (ANNATGG) (44) are marked ( ↓ ). The approximate position of promoterlike TATA boxes (TATAA and TATTAA) and polyadenylation signals (AATAAA) are denoted by arrows. The dotted portion of the horizontal lines represents the presumed unsequenced segment of the genome. The restriction map for those restriction endonucleases having one or two recognition sites in ADV-G is also shown. The following is a partial list of enzymes with no sites in ADV-G: *Aat*III, *Aha*II, *Asu*II, *Ban*I, *Bcl*I, *Bgl*I, *Bgl*II, *Cla*I, *Kpn*I, *Mlu*I, *Nar*I, *Not*I, *Pvu*I, *Sac*I, *Sac*II, *Sal*I, *Sma*I, *Sph*I, *Stu*I, *Xba*I, *Xma*III.

element, eight residues have diverged between the two ADV strains, suggesting that this might be a hypervariable region (Fig. 7, inset A).

The effect of these amino acid changes on the theoretical hydropathy (41) of the right ORFs was also examined (Fig. 8). As expected from the high level of relatedness, the general pattern of both strains is very similar, consisting of alternating hydrophobic and hydrophilic regions. Furthermore, the 15 amino acid changes not located in the hypervariable segment are found both in hydrophilic (9 of 15) and hydrophobic (6 of 15) areas. However, the amino acid changes within the hypervariable region produce a small, but obvious perturbation of this plot.

We also compared the available left ORF of ADV-Utah 1 with that of the ADV-G left ORF beginning at the *Bam*HI site at nt 721. The single-nucleotide deletion of an A corresponding to ADV-G nt 806 shifts the downstream portion of the left ORF to ADV-Utah 1 from this point into another translational reading frame. This deletion was observed on repeated sequencing of this region in both directions, but we could not determine whether the deletion was real or represented a cloning artifact, because only a single clone of ADV-Utah 1 was available for analysis (17). Assuming that a single base was deleted during cloning and that the left ORF of ADV-Utah 1 is single and continuous like that of ADV-G and all the other parvoviruses (31), there were four individual amino acid changes in this segment of the left ORF (Fig. 7), three of which were rated as conservative (data not shown). There was one change (nt 1646, F to L) which occurred within the GKRN region; however, this did not affect the hydropathy or the predicted protein structure (data not shown). MORF1 and MORF2 had one and three single-amino-acids changes, respectively.

## DISCUSSION

Two topics have been presented in this report. The first reports the nearly complete DNA sequence of the ADV-G strain of ADV and relationships to other parvoviruses. The second set of results is a comparison of the ADV-G sequence with a partial sequence of the virulent ADV-Utah 1 strain.

The first sections in this report point out structural similarities as well as differences among ADV-G and the other parvoviruses (9, 29–31, 65, 66, 68, 71). The overall levels of homology are low, confirming previous suggestions that ADV is not closely related to other parvoviruses (49, 58, 59). On the other hand, the gross organization of the ADV genome resembles in many but not all ways that recently summarized for MVM and other members of the group (31, 69, 70).

The terminal structures are known to play a pivotal role in parvoviral DNA replication, and the variation that can be

TRANSLATION OF ADV-G LEFT ORF

(NT 116-1975, FRAME 2, MOLECULAR WEIGHT=70,927 DALTONS)

```
                                                                    60
SVSFLWNEEVVWYISRGCLE  QSTDRLQQSNMAQAQIDEQR  RLQDLYVQLKKEINDGEGVA
                                                                   120
WLFQQKTYTDKDNKPTKATP  PLRTTSSDLRLAFDSIEENL  TASNEHLTNNEINFCKLTLG
                                                                   180
KTLLLIDKHVKSHRWDSNKV  NLIWQIEKGKTQQFHIHCCL  GYFDKNEDPKDVQKSLGWFM
                                                                   240
KRLNKDLAVIYSNHHCDIQD  IKDPEDRAKNLKVWIEDGPT  KPYKYFNKQTKQDYNKPVHL
                                                                   300
RDYTFIYLFNKDKINTDSMD  GYFAAGNGGIVDNLTNKERK  TLRKMYLDEQSSDIMDANID
                                                                   360
WEDGQDAPKVTDQTDSATTK  TGTSLIWKSCATKVTSKKEV  ANPVQQPSKKLYSAQSTLDA
                                                                   420
LFNVGCFTPEDMIIKQSDKY  LELSLEPNGPQKINTLLHMN  QVKTSTMITAFDCIIKFNEE
                                                                   480
EDDKPLLATIKDMGLNEQYL  KKVLCTILTKQGGKRGCIWF  YGPGGTGKTLLASLICKATV
                                                                   540
NYGMVTTSNPNFPWTDCGNR  NIIWAEECGNFGNWVEDFKA  ITGGGDVKVDTKNKQPQSIK
                                                                   600
GCVIVTSNTNITKVTVGCVE  TNAHAEPLKQRMIKIRCMKT  INPKTKITPGMLKRWLNTWD
RQPIQLSHEMPELYLGKCRW
```

COMPOSITION: 620 AMINO ACIDS

| | |
|---|---|
| ACIDIC (ASP + GLU) | 77(12.4%) |
| BASIC (ARG + LYS) | 84(13.5%) |
| AROMATIC (PHE + TRP + TYR) | 53(8.5%) |
| HYDROPHOBIC (AROMATIC + ILE + LEU + MET + VAL) | 188(30.3%) |

TRANSLATION OF ADV-G RIGHT ORF

(NT 2241-4346, FRAME 3, MOLECULAR WEIGHT=79,970 DALTONS)

```
                                                                    60
HHNTLLSLYLCYFTLVPQHY PGKKRSAPRHVFIQQAKKKK QTNPAVYHGEDTIEEMDSTE
                                                                   120
AEQMDTEQATNQTAEAGGGG GGGGGGGGGGGGVGNSTGGF NNTTEFKVINNEVYITCHAT
                                                                   180
RMVHINQADTDEYLIFNAGR TTDTKTHQQKLNLEFFVYDD FHQQVMTPWYIVDSNAWGVW
                                                                   240
MSPKDFQQMKTLCSEISLVT LEQEIDNVTIKTVTETNQGN ASTKQFNNDLTASLQVALDT
                                                                   300
NNILPYTPAAPLGETLGFVP WRATKPTQYRYYHPCVIYNR YPNIQKVATETLTWDAVQDD
                                                                   360
YLSVDEQYFNFITIENNIPI NILRTGDNFHTGLYEFNSKP CKLTLSYQSTRCLGLPPLCK
                                                                   420
PKTDTTHKVTSKENGADLIY IQGQDNTRLGHFWGEERGKK NAEMNRIRPYNIGYQYPEWI
                                                                   480
IPAGLQGSYFAGGPRQWSDT TKGAGTHSQHLQQNFSTRYI YDRNHGGDNEVDLLDGIPIH
                                                                   540
ERSNYYSDNEIEQHTAKQPK LRTPPIHHSKIDSWEEEGWP AASGTHFEDEVIYLDYFNFS
                                                                   600
GEQELNFPHEVLDDAAQMKK LLNSYQPTVAQDNVGPVYPW GQIWDKKPHMDHKPSMNNNA
                                                                   660
PFVCKNNPPGQLFVKLTENL TDTFNYDENPDRIKTYGYFT WRGKLVLKGKLSQVTCWNPV
KRELIGEPGVFTKDKYHKQI PNNKGNFEIGLQYGRSTIKY IY
```

COMPOSITION: 702 AMINO ACIDS

| | |
|---|---|
| ACIDIC (ASP + GLU) | 81(11.5%) |
| BASIC (ARG + LYS) | 66(9.4%) |
| AROMATIC (PHE + TRP + TYR) | 81(11.5%) |
| HYDROPHOBIC (AROMATIC + ILE + LEU + MET + VAL) | 207(29.4%) |

TRANSLATION OF ADV-G MID-ORF1

(NT 1993-2209, PHASE 1, MOLECULAR WEIGHT=7,918 DALTONS)

```
                                                                    60
MPTLNQHQFMRLLVFTETTG  PNSSATTATKNTGNSQPTTA  KSAESVNTENCDTPKRSASS
VPAKQHKRPRHE
```

COMPOSITION: 72 AMINO ACIDS

| | |
|---|---|
| ACIDIC (ASP + GLU) | 5(6.9%) |
| BASIC (ARG + LYS) | 9(12.5%) |
| AROMATIC (PHE + TRP + TYR) | 3(4.1%) |
| HYDROPHOBIC (AROMATIC + ILE + LEU + MET + VAL) | 10(13.9%) |

TRANSLATION OF ADV-G MID-ORF2

(NT 1983-2204, PHASE 3, MOLECULAR WEIGHT=8,922 DALTONS)

```
                                                                    60
HILNANFKPTSIYEVTLLYR  DYWTKLECHNCHEEYWQLTT  YYCKECRKCEHGKLRHTKKE
CEQCACKAAQETSA
```

COMPOSITION: 74 AMINO ACIDS

| | |
|---|---|
| ACIDIC (ASP + GLU) | 10(13.5%) |
| BASIC (ARG + LYS) | 11(14.9%) |
| AROMATIC (PHE + TRP + TYR) | 9(12.2%) |
| HYDROPHOBIC (AROMATIC + ILE + LEU + MET + VAL) | 18(24.3%) |

FIG. 4. Theoretical translations of the coding regions of ADV-G. The DNA sequence of the major ORFs was translated into amino acids, using the single-letter code. The nucleotide coordinates, translation frames, and calculated molecular masses in daltons are shown. In addition, the numbers and compositions of the residues are listed.

tolerated is of great interest (11, 31, 47). Unfortunately, these sequences have proven extremely difficult to molecularly clone and to maintain in procaryotic vectors (24, 52). The structure of the 3′ ADV-G terminus is very similar to that described for MVM (Fig. 3), but the calculated $T_m$ of the ADV-G 3′ terminus is 37.3°C, significantly below that of MVM (48.7°C), and the ADV-G hairpin is thus less stable. Because of the severe difficulties encountered by others (24, 29, 52, 62) in analyzing this structure and because our clones containing the 3′ end were not produced in *E. coli* strains tolerant of parvoviral palindromic structures (24), it is conceivable that the sequence we have presented may contain some errors, but the several clones we sequenced gave the same result. The host range of ADV is greatly restricted in vivo and in vitro (4, 20, 58), and we recently showed that intracellular levels of RF DNA vary markedly between permissive and nonpermissive systems (2–4, 22). Perhaps a relatively unstable 3′ terminus impairs the ability of ADV to replicate promiscuously and thus limits populations of susceptible host target cells. Interestingly, the two systems for highly permissive ADV replication are at reduced temperatures, i.e., propagation in CRFK cells at 32°C (20, 59) and growth in the lungs of poikilothermic newborn mink (1, 2, 4), and it is conceivable that one function of the reduced temperature is to help stabilize the left-hand hairpin and facilitate replication.

The sequence of ADV-G at the right-hand end does not contain a hairpin but terminates just downstream of a 25-bp A+T-rich direct repeat. Similar direct repeats occur immediately before the right-hand palindrome of MVM, H-1 virus, FPV, and canine parvovirus (CPV) (31), and we concluded that the ~200-bp deletion at this end of the ADV-G sequence likely contained the right-hand hairpin. Similar results were obtained when ADV DNA was molecularly cloned into *E. coli* DB1256, which has a *recB recC sbc* genotype and is tolerant of vaccinia virus palindromes (33). This might suggest that the ADV right-hand terminus will prove more difficult to clone than those of the other parvoviruses, nevertheless, additional cloning experiments with other recombination-deficient *E. coli* strains (JC8111) that are demonstrably tolerant of parvoviral palindromes (24) are in progress. Perhaps clones derived from these bacteria will provide definitive information on the terminal fine structure of ADV.

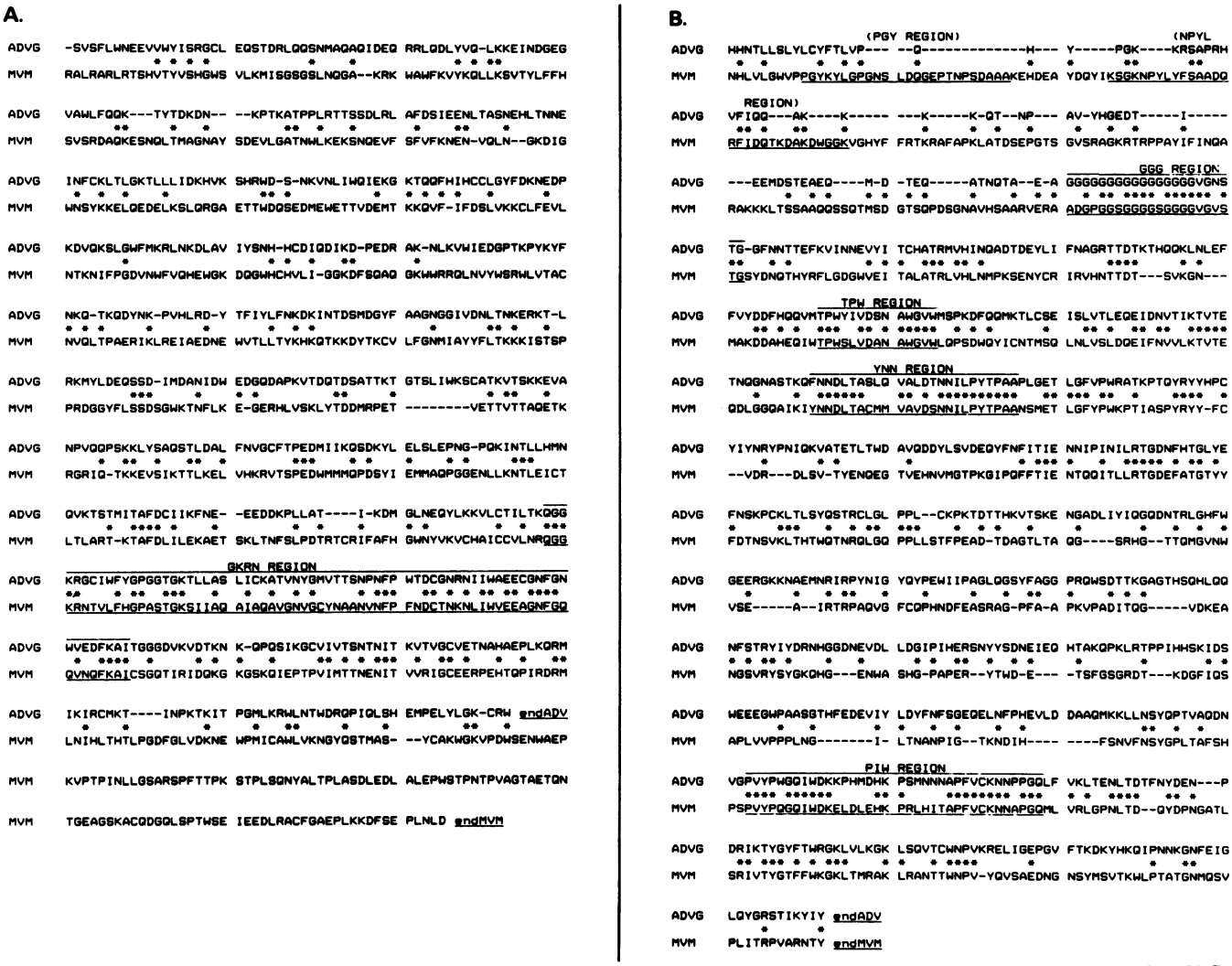


FIG. 5. Comparison of amino acid sequences of the left and right ORFs of ADV-G and MVM(p). The left (A) and right (B) ORFs of ADV-G and MVM(p) were translated into amino acids, using the single-letter code. The optimal overlap was determined by using the Alignment maneuver of the Comparison mode of the Microgenie sequence program. Identical residues are indicated (*). The locations of conserved regions (GKRN in the left ORFs; PGY, NPYL, GGG, TPW, YNN, and PIW in the right ORFs) (29, 31) are also defined.

Purified ADV-G virions contain two major capsid proteins with molecular masses of 85 (VP1) and 75 kilodaltons (VP2) (20, 21), and in addition, a nonstructural protein of 71 kilodaltons (NS-1) is found in ADV-infected cells (21). These proteins must be coded for by ORFs within the primary sequence, although noncontiguous coding regions are likely juxtaposed by RNA processing (27, 31, 53, 55). ADV-G, like the other parvoviruses, contains major left and right coding regions as well as several smaller ORFs located in the middle of the genome. Our results to date suggest that the genomic localization of these gene products in ADV grossly resembles that of the other parvoviruses.

NS-1 is the largest gene product of the left ORF (26, 30, 31). Its emerging role is that of a multifunctional protein involved both with DNA replication as well as regulation of viral and cellular promoters (12, 30, 31, 64). The conserved GKRN element (Fig. 5A) found in the available NS-1 sequences (29, 31, 68) resembles portions of several ATP- and GTP-binding proteins (7, 31) and has led others to speculate that NS-1 may be one of the presumed viral proteins directly involved with DNA replication (7, 31). A minimal amino acid consensus $[G:(X)_4:G:K:T/S:(X)_{5-6}:I/L/V]$ for the purine triphosphate-binding site has been proposed by Cotmore and Tattersall (31), and this consensus is preserved in ADV (Fig. 4 and 5A, left ORF amino acid 464 and following).

Additional NS proteins (30, 31) have not been formally identified in ADV-infected cells, but the presence of smaller mid-ORFs suggests that existence of such gene products, and, in fact, mRNA and cDNA analyses reinforce this suggestion (Alexandersen et al., submitted).

We previously showed that coding sequences for the ADV capsid proteins are found to the right of the *Eco*RI site (49) at nt 2553 and furthermore that all VP2 sequences are contained within VP1 (21). Therefore, it seemed reasonable to assume that the right ORF does in fact code for the bulk of the capsid proteins. The theoretical translation of this ORF (Fig. 5) yielded a protein with a mass (79,970 daltons) intermediate to those observed for VP1 (85,000 daltons) and VP2 (75,000 daltons). VP1 and VP2 are products of a spliced mRNA, the boundaries of which cannot be defined by simply examining the genomic sequence, and consequently, we have deferred discussing the arrangement of the ORFs into

| ADV-G | POSITION | UTAH 1 | ALTERATION | CODON POSITION |
|---|---|---|---|---|
| | | LEFT OPEN READING FRAME | | |
| A | 750 | G | I to V | 1 |
| A | 804 | | reading frame shift | |
| T | 820 | C | delete polyA site | 3 |
| G | 1182 | A | S to N | 2 |
| T | 1339 | G | I to M | 3 |
| T | 1646 | C | F to L | 1 |
| C | 1705 | T | | 3 |
| | | MID OPEN READING FRAMES | | |
| A | 2161 | G | MORF1 S to G | 2 |
| | | | MORF2 E to G | 3 |
| G | 2179 | C | MORF2 C to S | 2 |
| A | 2181 | G | MORF2 K to E | 1 |
| A | 2211 | G | | 3 |
| A | 2218 | G | non-coding region | |
| | | RIGHT OPEN READING FRAME | | |
| A | 2247 | G | N to D | 1 |
| C | 2260 | A | S to Y | 2 |
| G | 2333 | A | | 3 |
| C | 2378 | G | | 3 |
| G | 2390 | A | | 3 |
| C | 2396 | A | | 3 |
| A | 2415 | G | T to A | 1 |
| G | 2421 | C | A to P | 1 |
| A | 2429 | G | | 3 |
| T | 2471 | G | | 3 |
| G | 2477 | A | | 3 |
| G | 2487 | A | G to S | 1 |
| T | 2492 | G | | 3 |
| G | 2495 | T | | 3 |
| T | 2507 | G | | 3 |
| A | 2609 | G | | 3 |
| T | 2615 | C | | 3 |
| C | 2633 | T | | 3 |

| ADV-G | POSITION | UTAH 1 | ALTERATION | CODON POSITION |
|---|---|---|---|---|
| G | 2656 | A | G to D | 2 |
| C | 2679 | G | H to A | 1,2 |
| A | 2680 | C | | |
| G | 2684 | A | | 3 |
| C | 2685 | A | Q to K | 1 |
| A | 2749 | T | Y to F | 2 |
| A | 2777 | G | | 3 |
| A | 2825 | G | | |
| A | 2852 | G | | 3 |
| A | 2873 | G | | 3 |
| T | 2906 | C | | 3 |
| A | 3094 | | | |
| T | 3100 | C | | |
| C | 3103 | G | | |
| A | 3104 | G | | |
| A | 3105 | G | | |
| A | 3109 | | | |
| A | 3111 | G | | |
| | 3112 | A | | |
| C | 3114 | T | SEE FIGURE 7, INSET A | |
| A | 3117 | G | | |
| C | 3118 | A | | |
| C | 3119 | A | | |
| G | 3123 | A | | |
| A | 3124 | C | | |
| C | 3127 | G | | |
| A | 3128 | T | | |
| G | 3129 | A | | |
| T | 3130 | C | | |
| T | 3146 | G | | 3 |
| A | 3158 | G | | 3 |
| T | 3174 | C | | 3 |
| A | 3459 | G | I to V | 1 |
| C | 3590 | G | H to Q | 3 |
| A | 3705 | C | N to H | 1 |
| G | 3869 | A | | 3 |
| G | 3872 | A | | 3 |
| C | 3873 | T | | 1 |
| A | 3876 | G | N to E | 1,3 |
| C | 3878 | G | | |
| G | 3911 | A | | 3 |
| G | 3977 | A | | 3 |
| C | 4005 | G | H to D | 1 |

FIG. 6. Sequence differences between ADV-G and ADV-Utah 1. The DNA sequence of ADV-G from the 5′ end of the *Bam*HI site (nt 721) to the 3′ end of the *Hin*dIII site (nt 4176) was compared with the sequence of a *Bam*HI-*Hin*dIII molecular clone of ADV-Utah 1 (17). DNA sequence differences are listed, with the nucleotide coordinate of ADV-G as a reference point; where appropriate, the ADV-G nucleotide immediately 5′ to any insertion or deletion is denoted. The location of these changes with respect to ORFs or noncoding regions is given. In addition, the position of the nucleotide change within codons and any amino acid changes are also described. Adjacent nucleotide differences that altered single codons (nt 2679-2680 and nt 3876-3878) are grouped with a bracket. The region from nt 3094 to nt 3130 contains extensive changes, with a complex effect on the coding sequence; this set of differences is also grouped with a bracket, and a comparison of this region in the two ADV strains is depicted in detail in inset A of Fig. 7.

```
                                                    GGG REGION
ADVG  AEQMDTEQATNQTAEAGGGG GGGGGGGGGGGGVGNSTGGF NNTTEFKVINNEVYITCHAT
UTAH  PEQMDTEQATNQTAEAGGGG GGSGGGGGGGGGVGNSTGGF NNTTEFKVINNEVYITCHAT

                                                                TPW REGION
ADVG  RMVHINQADTDEYLIFNAGR TTDTKTHQQKLNLEFFVYDD FHQQVMTPWYIVDSNAWGVW
UTAH  RMVHINQADTDEYLIFNADR TTDTKTAQKKLNLEFFVYDD FHQQVMTPWFIVDSNAWGVW

                                                                YNN REGION
ADVG  MSPKDFQQMKTLCSEISLVT LEQEIDNVTIKTVTETNQGN ASTKQFNNDLTASLQVALDT
UTAH  MSPKDFQQMKTLCSEISLVT LEQEIDNVTIKTVTETNQGN ASTKQFNNDLTASLQVALDT

                                                          I--inset A--I
ADVG  NNILPYTPAAPLGETLGFVP WRATKPTQYRYYHPCYIYNR YPNIQKVATETLTWDAVQDD
UTAH  NNILPYTPAAPLGETLGFVP WRATKPTQYRYYHPCYIYNR YPNIQKLGQEQLEWTGTQDD

ADVG  YLSVDEQYFNFITIENNIPI NILRTGDNFHTGLYEFNSKP CKLTLSYQSTRCLGLPPLCK
UTAH  YLSVDEQYFNFITIENNIPI NILRTGDNFHTGLYEFNSKP CKLTLSYQSTRCLGLPPLCK

ADVG  PKTDTTHKVTSKENGADLIY IQGQDNTRLGHFWGEERGKK NAEMNRIRPYNIGYQYPEWI
UTAH  PKTDTTHKVTSKENGADLIY IQGQDNTRLGHFWGEERGKK NAEMNRVRPYNIGYQYPEWI

ADVG  IPAGLQGSYFAGGPRQWSDT TKGAGTHSQHLQQNFSTRYI YDRNHGGDNEVDLLDGIPIH
UTAH  IPAGLQGSYFAGGPRQWSDT TKGAGTHSQQLQQNFSTRYI YDRNHGGDNEVDLLDGIPIH

ADVG  ERSNYYSDNEIEQHTAKQPK LRTPPIHHSKIDSWEEEGWP AASGTHFEDEVIYLDYFNFS
UTAH  ERSNYYSDHEIEQHTAKQPK LRTPPIHHSKIDSWEEEGWP AASGTHFEDEVIYLDYFNFS

                                                          PIW REGION
ADVG  GEQELNFPHEVLDDAAQMKK LLNSYQPTVAQDNVGPVYPW GQIWDKKPHMDHKPSMNNNA
UTAH  GEQELEFPHEVLDDAAQMKK LLNSYQPTVAQDNVGPVYPW GQIWDKKPDMDHKPSMNNNA

ADVG  PFVCKNNPPGQLFVKLTENL TDTFNYDENPDRIKTYGYFT WRGKL
UTAH  PFVCKNNPPGQLFVKLTENL TDTFNYDENPDRIKTYGYFT WRGKL
```

```
inset A

          IleGlnLysValAlaThrGluTh rLeuThrTrpAspAlaValGln
ADVG  3090 ATTCAAAAAGTTGCAACAGAAAC-ACTAACCTGGGATGCAGTACAA 3134
UTAH       ATTC-AAAAGCTGGGGCAGGAGCAATTAGAATGGACTGGTACACAA
           IleG lnLysLeuGlyGlnGluGlnLeuGluTrpThrGlyThrGln
```

the probably viral polypeptides to a report dealing with ADV transcription (Alexandersen et al., submitted).

The amino terminus of VP1 for MVM occurs in a different ORF and, as noted (31, 42), is joined to the bulk of the capsid coding region by mRNA splicing. This VP1-specific segment contains for MVM, AAV, BPV, FPV, and B19 parvovirus two conserved sequences (31) denoted NPYL (~47 amino acids in size) and PGY (~30 amino acids), the functions of which are obscure (31). Analogs to these two amino acid elements could not be identified in ADV (Fig. 5B). Perhaps the absence of the PGY and NPYL regions is partly responsible for some of the unusual biological properties of ADV, such as the inability to spread well in culture (20, 59). The effect of mutations in this region would therefore be interesting to observe in MVM or other parvoviruses.

The final set of data that we present here is a comparison of the ADV-G sequence to a partial sequence of the virulent ADV-Utah 1 (Fig. 7 and 8). The two strains are highly related (>95% at both the DNA and amino acid levels). A deletion of a single base occurs in the sequence of ADV-Utah 1 in the position corresponding to nt 806 of ADV-G that shifts the left ORF of ADV-Utah 1 from translational frame

FIG. 7. Comparison of the right ORFs of ADV-G and ADV-Utah 1. The right ORF of ADV-G from the start at nt 2241 to the amino acid corresponding to the 3′ end of the *Hin*dIII site (nt 4176) was aligned with the theoretical translation of the analogous segment of ADV-Utah 1, using the single-letter amino acid code. Nonidentical or unmatched residues are emphasized (*). (Inset A) Detail of nucleotide and amino acid changes at the localized area of marked difference described in the text. To better illustrate the effects of nucleotide changes on amino acid sequence, the three-letter amino acid code was used.

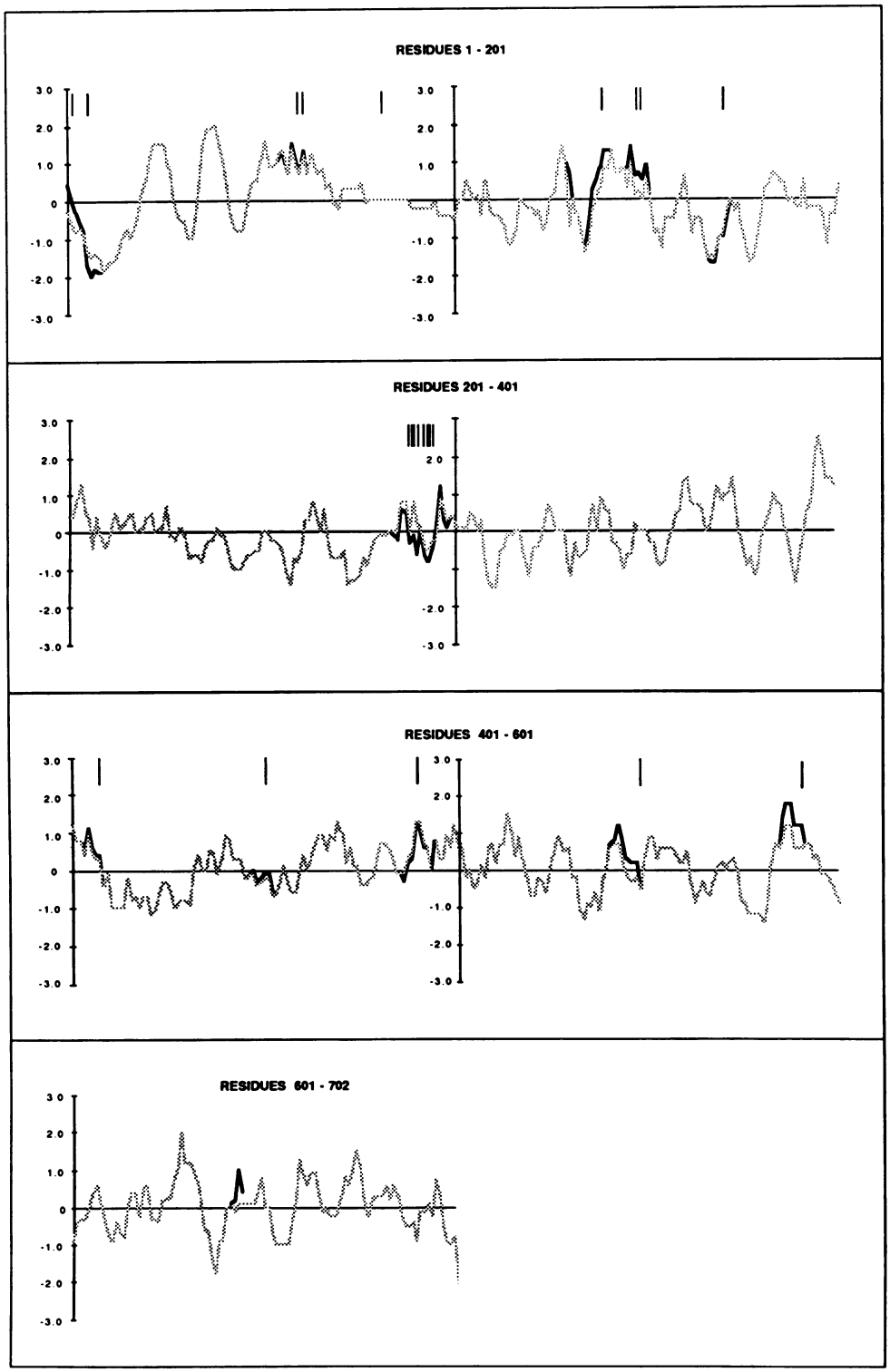


FIG. 8. Comparative hydropathy plot of the right ORFs of ADV-G and ADV-Utah 1. Hydropathy plots of the entire 702-amino-acid right ORF of ADV-G (nt 2241 to 4346) and the 645-amino-acid right ORF sequence available for ADV-Utah 1 were constructed with the Microgenie sequence program, using the procedure of Hopp and Woods (41). The plots were compared, and the regions unique to ADV-Utah 1 are signified with a solid line. The positions of the individual amino acid changes between ADV-G and ADV-Utah 1 are indicated by vertical marks above the plot. Those portions above the horizontal axis were hydrophilic, and those below were hydrophobic. The apparent divergence at the end of the ADV-Utah 1 plot was due to the fact that the available right ORF of ADV-Utah 1 ends at residue 645.

2 to frame 1. Although we did not have additional molecular clones of ADV-Utah 1 to compare, we speculated that this change was a cloning artifact because the left ORFs of ADV-G and all other parvoviruses are continuous. Additional support for this notion may be obtained from the observation that p71, the NS-1 product of the ADV left ORF (M. E. Bloom, S. Alexandersen, and J. B. Wolfinbarger, Second Parvovirus Workshop, 1988) is the same size in both ADV-G and ADV-Utah 1 (~71,000 daltons) (5, 61), close to the theoretical coding capacity of the total left ORF of ADV-G.

The major differences between ADV-G and ADV-Utah 1 relate to the ability of the former to grow well in CRFK but poorly in mink (20) and the capacity of the latter to grow and to cause fulminant disease in mink (20, 21, 38, 58, 60) but not to replicate productively in CRFK cells (20, 39). There are only 23 amino acid changes between the right ORFs of the two ADV strains; however, a short amino acid stretch at 64-65 MU shows a marked divergence between the two strains (8 of 11 amino acids) (Fig. 6, 7, and 8). The observation of this short divergent segment is particularly intriguing because hypervariable regions have been reported in the capsid genes for two strains of MVM [the immunosuppressive MVM(i) and the prototypic MVM(p)] (9, 66) and for CPV-FPV (56). In both instances, the regions map to a similar location in the right ORF. This hypervariability correlates with the viral host range (CPV-FPV) (56) and the ability of the virus to initiate a productive rather than restrictive infection [MVM(i)-MVM(p)] (44, 72; P. Tattersall, R. Moir, and E. Gardner, Second Parvovirus Workshop, 1987). Perhaps this region in ADV functions in an analogous fashion and is a major determinant of host range and pathogenicity.

It is possible to discriminate ADV strains antigenically by using monoclonal antibodies, although most monoclonal antibodies react with all ADV strains (61). Therefore, it is interesting to speculate that the sequence variation observed in the right ORFs of ADV-G and ADV-Utah 1 may relate to these minor differences in antigenicity. The amino acid changes in the right ORFs were found both in hydrophobic and hydrophilic regions (Fig. 8). Although single-amino-acid changes can greatly influence pathogenicity and antigenicity (25, 35, 36), it is difficult to predict from the primary nucleotide sequence the location of major epitopes (13, 37) or the effect of a single change (25, 33, 35, 37). Nevertheless, it would be interesting to synthesize peptides that spanned these variable regions and to determine whether they could induce virus strain-specific responses or modify the outcome of ADV infection.

Finally, when infectious ADV clones are constructed, it will be possible to construct ADV-G/ADV-Utah 1 chimeras and to assess the role of sequence differences as determinants of pathogenicity. These experiments are being actively pursued.

In summary, the work presented here reinforces the prevailing notion that all the parvoviruses, including ADV, exhibit a very similar genomic organization. At the same time, however, the apparent absence in ADV of two otherwise conserved regions in the right ORF (PGY and NPYL) and an unusual left-hand terminus tentatively suggest structural features that might account for the unusual nature of ADV infections (2–4, 16, 22, 58). Last, the comparisons of ADV-G and ADV-Utah 1 give preliminary evidence that the sequence differences between these two virus strains are slight. The marked difference in pathogenicity of the various ADV strains may be determined by the variation we have observed, but on the other hand, the unsequenced regions may also play a role in pathogenicity. Further molecular analysis will hopefully provide additional insight into these problems.

## ACKNOWLEDGMENTS

We acknowledge the assistance of Anders Cohn, Oskar Ruger-Kaaden, Sven Bergstrom, Camille Locht, Eric Huggans, Kenneth Robbins, and Danny Weidbrauk in the initial stages of this work. Other members of the staff of the Rocky Mountain Laboratories provided helpful discussion throughout the work. Robert Evans and Gary Hettrick aided in the production of the illustrations, and Irene Cook Rodriguez assisted with preparation of the manuscript.

This work was supported in part by the Danish Fur Breeders Association Research Foundation. Soren Alexandersen is an NIH Visiting Associate on leave from the Institute of Veterinary Pathology, Royal Veterinary and Agricultural University of Copenhagen, Denmark.

## LITERATURE CITED

1. **Alexandersen, S.** 1986. Acute interstitial pneumonia in mink kits: experimental reproduction of the disease. Vet. Pathol. **23:** 579–588.
2. **Alexandersen, S., and M. E. Bloom.** 1987. Studies on the sequential development of acute interstitial pneumonia caused by Aleutian disease virus in mink kits. J. Virol. **61:**81–86.
3. **Alexandersen, S., M. E. Bloom, and J. Wolfinbarger.** 1988. Evidence of restricted viral replication in adult mink infected with Aleutian disease of mink parvovirus. J. Virol. **62:**1495–1507.
4. **Alexandersen, S., M. E. Bloom, J. Wolfinbarger, and R. E. Race.** 1987. In situ molecular hybridization for detection of Aleutian mink disease parvovirus DNA by using strand-specific probes: identification of target cells for viral replication in cell cultures and in mink kits with virus-induced interstitial pneumonia. J. Virol. **61:**2407–2419.
5. **Alexandersen, S., A. Uttenthal-Jensen, and B. Aasted.** 1986. Demonstration of non-degraded Aleutian disease virus (ADV) proteins in lung tissue from experimentally infected mink kits. Arch. Virol. **87:**127–133.
6. **Anderson, M. J.** 1987. Human parvovirus infections. J. Virol. Methods **17:**175–181.
7. **Anton, I. A., and D. P. Lane.** 1986. Non-structural protein 1 of parvoviruses: homology to purine nucleotide using proteins and early proteins of papovaviruses. Nucleic Acids Res. **14:**7813.
8. **Armentrout, R., R. Bates, K. Berns, B. Carter, M. Chow, D. Dressler, K. Fife, W. Hauswirth, G. Hayward, G. Lavelle, S. Rhode, S. Strauss, P. Tattersall, and D. Ward.** 1978. A standard nomenclature for restriction endonuclease fragments, p. 523–526. *In* D. Ward and P. Tattersall (ed.), Replication of parvoviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Astell, C. R., E. M. Gardiner, and P. Tattersall.** 1986. DNA sequence of the lymphotropic variant of minute virus of mice, MVM(i), and comparison with the DNA sequence of the fibrotropic prototype strain. J. Virol. **57:**656–669.
10. **Bensimhon, M., J. Gabarro-Arpa, R. Ehrlich, and C. Reiss.** 1983. Physical characteristics in eucaryotic promoters. Nucleic Acids Res. **11:**4521–4540.
11. **Berns, K. I., and R. A. Bohenzky.** 1987. Adeno-associated viruses: an update. Adv. Virus Res. **32:**243–307.
12. **Berns, K. I., and M. A. Labow.** 1987. Parvovirus gene regulation. J. Gen. Virol. **68:**601–614.
13. **Berzofsky, J. A.** 1985. Instrinsic and extrinsic factors in protein antigenic structure. Science **229:**932–940.
14. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and $^{35}S$ label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80:**3963–3965.
15. **Birnstiel, M. L., M. Busslinger, and K. Strub.** Transcription termination and 3′ processing: the end is in site! Cell **41:**349–359.
16. **Bloom, M. E.** 1984. Parvovirus infections: features reminiscent

of AIDS. Ann. N. Y. Acad. Sci. **437**:110–120.
17. **Bloom, M. E., O.-R. Kaaden, E. Huggans, A. Cohn, and J. B. Wolfinbarger.** 1988. Molecular comparisons of in vivo- and in vitro-derived strains of Aleutian mink disease parvovirus. J. Virol. **62**:132–138.
18. **Bloom, M. E., D. Lechner, D. L. Wiedbrauk, and J. B. Wolfinbarger.** 1987. Analysis of molecularly cloned DNA reveals minor differences among three virus strains of Aleutian disease of mink parvovirus. Arch. Virol. **92**:175–181.
19. **Bloom, M. E., L. W. Mayer, and C. F. Garon.** 1983. Characterization of the Aleutian disease virus genome and its intracellular forms. J. Virol. **45**:977–984.
20. **Bloom, M. E., R. E. Race, and J. B. Wolfinbarger.** 1980. Characterization of Aleutian disease virus as a parvovirus. J. Virol. **35**:836–843.
21. **Bloom, M. E., R. E. Race, and J. B. Wolfinbarger.** 1982. Identification of a nonvirion protein of Aleutian disease virus: mink with Aleutian disease have antibody to both virion and nonvirion proteins. J. Virol. **43**:606–616.
22. **Bloom, M. E., R. E. Race, and J. B. Wolfinbarger.** 1987. Analysis of Aleutian disease of mink parvovirus infection using strand-specific hybridization probes. Intervirology **27**:102–111.
23. **Blundell, M. C., C. Beard, and C. R. Astell.** 1987. In vitro identification of a B19 parvovirus promoter. Virology **157**:534–538.
24. **Boissy, R., and C. Astell.** 1985. An *Escherichia coli recBCsbc BrecF* host permits the deletion-resistant propagation of plasmid clones containing the 5′-terminal palindrome of minute virus of mice. Gene **35**:179–185.
25. **Both, G. W., C. H. Shi, and E. D. Kilbourne.** 1983. Hemagglutination of swine influenza virus: a single amino acid change pleiotropically affects viral antigenicity and replication. Proc. Natl. Acad. Sci. USA **80**:6996–7000.
26. **Carlson, J. O., M.-K. Lynde-Maas, and S. Zheng-da.** 1987. A nonstructural protein of feline panleukopenia virus: expression in *Escherichia coli* and detection of multiple forms in infected cells. J. Virol. **61**:621–624.
27. **Carter, B. J., C. A. Laughlin, and C. J. Marcus-Sekura.** 1984. Parvovirus transcription. p. 209–258. *In* K. I. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
28. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**: 165–170.
29. **Chen, K. C., B. C. Shull, E. A. Moses, M. Lederman, E. R. Stout, and R. C. Bates.** 1986. Complete nucleotide sequence and genome organization of bovine parvovirus. J. Virol. **60**: 1085–1097.
30. **Cotmore, S. F., and P. Tattersall.** 1986. Organization of nonstructural genes of the autonomous parvovirus minute virus of mice. J. Virol. **58**:724–732.
31. **Cotmore, S. F., and P. Tattersall.** 1987. The autonomously replicating parvoviruses. Adv. Virus Res. **33**:91–174.
32. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid **13**:31–40.
33. **DeLanqe, A. M., M. Reddy, D. Scraba, C. Upton, and G. McFadden.** 1986. Replication and resolution of cloned poxvirus telomeres in vivo generates linear minichromosomes with intact viral hairpin termini. J. Virol. **59**:249–259.
34. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single stranded plasmids. Nucleic Acids Res. **11**:1645–1655.
35. **Diamond, D. C., B. A. Jameson, J. Bonn, M. Kohara, S. Abe, H. Itoh, T. Komatsu, M. Arita, S. Kuge, A. Nomoto, A. D. M. E. Osterhaus, R. Crainic, and E. Wimmer.** 1985. Antigenic variation and resistance to neutralization in poliovirus type 1. Science **229**:1090–1093.
36. **Dietzschold, B. W., W. H. Wunner, T. J. Wiktor, A. D. Lopes, M. Lafon, C. L. Smith, and H. Koprowski.** 1983. Characterization of an antigenic determinant of the glycoprotein that correlates with the pathogenicity of rabies virus. Proc. Natl. Acad. Sci. USA **80**:70–74.
37. **Fieser, T. M., J. A. Tainer, H. M. Geysen, R. A. Houghten, and R. A. Lerner.** 1987. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein α-helix. Proc. Natl. Acad. Sci. USA **84**:8568–8572.
38. **Hadlow, W. J., R. E. Race, and R. C. Kennedy.** 1983. Comparative pathogenicity of four strains of Aleutian disease virus for pastel and sapphire mink. Infect Immun. **41**:1016–1023.
39. **Hahn, E. C., L. Ramos, A. J. Kenyon.** 1977. Expression of Aleutian mink disease antigen in cell culture. Infect. Immun. **15**:204–211.
40. **Hanahan, D.** 1983. Studies on transformation of *E. coli* with plasmids. J. Mol. Biol. **166**:557–580.
41. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinant from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
42. **Jongeneel, C. V., R. Sahli, G. K. McMaster, and B. Hirt.** 1986. A precise map of splice junctions in the mRNAs of minute virus of mice, an autonomous parvovirus. J. Virol. **59**:564–573.
43. **Khoury, G., and P. Gruss.** 1983. Enhancer elements. Cell **33**: 313–314.
44. **Kimsey, P. B., H. D. Engers, B. Hirt, and C. V. Jongeneel.** 1986. Pathogenicity of fibroblastic- and lymphocyte-specific variants of minute virus of mice. J. Virol. **59**:8–13.
45. **Kozak, M.** 1986. Regulation of protein synthesis in virus-infected animal cells. Adv. Virus Res. **31**:229–292.
46. **Labieniec-Pintel, L., and D. Pintel.** 1986. The minute virus of mice $P_{39}$ promoter can encode both capsid proteins. J. Virol. **57**: 1163–1167.
47. **Lefebvre, R. B., S. Riva, and K. I. Berns.** 1984. Conformation takes precedence over sequence in adeno-associated virus DNA replication. Mol. Cell. Biol. **4**:1416–1419.
48. **Maniatis, T., E. F. Fritsch, J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
49. **Mayer, L. W., B. Aasted, C. F. Garon, and M. E. Bloom.** 1983. Molecular cloning of the Aleutian disease virus genome: expression of Aleutian disease virus antigens by a recombinant plasmid. J. Virol. **48**:573–579.
50. **McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins.** 1984. Requirement of downstream sequence for generation of a poly(A) addition site. Cell **37**:993–999.
51. **Melton, D. A., P. A. Krieg. M. R. Rebagliati, T. Maniatis. K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
52. **Merchlinksy, M. J., P. 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 minute virus of mice. J. Virol. **47**:227–232.
53. **Morgan, W. R., and D. C. Ward.** 1986. Three splicing patterns are used to excise the small intron common to all minute virus of mice RNAs. J. Virol. **60**:1170–1174.
54. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
55. **Ozawa, K., J. Ayub, H. Yu-Shu, G. Kurtzman, T. Shimada, and N. Young.** 1987. Novel transcription map for the B19 (human) parvovirus. J. Virol. **61**:2395–2406.
56. **Parrish, C. R., and L. E. Carmichael.** 1986. Characterization and recombination mapping of an antigenic and host range mutation of canine parvovirus. Virology **148**:121–132.
57. **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.
58. **Porter, D. D.** 1986. Aleutian disease: a persistent parvovirus infection of mink with a maximal but ineffective host immune response. Prog. Med. Virol. **33**:42–60.
59. **Porter, D. D., A. E. Larsen, N. A. Cook, H. G. Porter, and S. L. Suffin.** 1977. Isolation of Aleutian disease virus of mink in cell culture. Intervirology **8**:129–144.

60. **Porter, D. D., A. E. Larsen, and H. G. Porter.** 1969. The pathogenesis of Aleutian disease of mink. I. In vivo viral replication and the host antibody response to viral antigen. J. Exp. Med. **130**:575–589.
61. **Race, R. E., B. Chesebro, M. E. Bloom, B. Aasted, and J. Wolfinbarger.** 1986. Monoclonal antibodies against Aleutian disease virus distinguish virus strains and differentiate sites of virus replication from sites of viral antigen sequestration. J. Virol. **57**:285–293.
62. **Reed, A. P., E. V. Jones, and T. J. Miller.** 1988. Nucleotide sequence and genome organization of canine parvovirus. J. Virol. **62**:266–276.
63. **Reed, K. C. and D. A. Mann.** 1985. Rapid transfer of DNA from agarose gels to nylon membranes. Nucleic Acids Res. **13**:7207–7221.
64. **Rhode, S. L., III.** 1985. *trans*-Activation of parvovirus $P_{38}$ promoter by the 76K noncapsid protein. J. Virol. **55**:886–889.
65. **Rhode, S. L., III, and P. K. Paradiso.** 1983. Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. J. Virol. **45**:173–184.
66. **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.
67. **Sanger, F. S., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
68. **Shade, R. O., M. C. Blundell, S. F. Cotmore, P. Tattersall, and C. R. Astell.** 1986. Nucleotide sequence and genome organization of human parvovirus B19 isolated from the serum of a child during aplastic crisis. J. Virol. **58**:921–936.
69. **Siegl, G.** 1984. Biology and pathology of autonomous parvoviruses, p. 296–332. *In* K. I. Berns (ed.), The parvoviruses. Plenum Publishing Corp., New York.
70. **Siegl, G., R. C. Bates, K. I. Berns, B. C. Carter, D. C. Kelly, E. Kurstak, and P. Tattersall.** 1985. Characteristics and taxonomy of parvoviridae. Intervirology **23**:61–73.
71. **Srivasta, A., E. W. Lusby, and K. I. Berns.** 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J. Virol. **45**:555–564.
72. **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.
73. **Wetmur, R. and R. Davidson.** 1968. Kinetics of renaturation of DNA. J. Mol. Biol. **31**:349–370.
74. **Yanisch-Perron, C., J. Viera, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC 19 vectors. Gene **33**:103–119.