Nucleotide Sequence of an Immediate-Early Frog Virus 3 Gene

DAWN WILLIS,* DAVID FOGLESONG, AND ALLAN GRANOFF

St. Jude Children’s Research Hospital, Memphis, Tennessee 38101

Received 4 June 1984/Accepted 8 August 1984

We have used “gene walking” with synthetic oligonucleotides and M13-dideoxy nucleotide sequencing techniques to obtain the complete coding and flanking sequences of the gene encoding a major immediate-early RNA (molecular weight, 169,000) of frog virus 3. R-loop mapping of the cloned Xbal K fragment of frog virus 3 DNA with immediate-early RNA from infected cells showed that an RNA of approximately 500 to 600 nucleotides (the right size to code for the immediate-early viral 18-kilodalton protein of unknown function) hybridized to a region within 100 base pairs of one end of the Xbal K fragment; no evidence for splicing was observed in the electron microscope or by single-strand nuclease analysis. Further restriction mapping narrowed the location of the gene to the Xbal end of a 2-kilobase-pair Xbal-BglII fragment, which was bidirectionally subcloned into the bacteriophage pair mp10 and mp11 for sequencing. Mung bean nuclease mapping was used to identify both the 5’ and the 3’ ends of the mRNA. The 5’ end mapped within an AT-rich region 19 base pairs upstream from two in-phase AUG start codons that were immediately followed by an open reading frame of 157 amino acids. Another AT-rich sequence was found at the 3’ end of the mRNA start site; this sequence may function as a TATA box. The 3’ end of the message displayed considerable microheterogeneity, but clearly terminated within a third AT-rich region 50 to 60 base pairs from the translation stop codon. The eucaryotic polyadenyllic acid addition signal (AATAAA) was not present, a finding to be expected since frog virus 3 mRNA is not polyadenylated. Both the single-stranded mp10 clone of the Xbal-BglII fragment and a 15-base oligonucleotide complementary to the region flanking the two AUG translation start codons inhibited translation of the immediate-early 18-kilodalton protein in vitro, confirming the identity of the sequenced gene. As the regulatory sequences of this gene did not resemble those of known eucaryotic genes or of the cytoplasmic vaccinia virus, we conclude that frog virus 3 has evolved unique signals for the initiation and termination of transcription.

Specific DNA sequences that interact with RNA polymerase or other regulatory proteins serve to initiate and terminate transcription in both procaryotic and eucaryotic cells. DNA viruses that replicate in the nucleus have promoters (regulators for initiation of transcription) that are similar to those of eucaryotic cells. With few exceptions, nuclear eucaryotic genes have a Goldberg-Hogness or TATA box at approximately −30 base pairs (bp) from the start site of transcription, and terminate near the AAAAAA signal for polyadenyllic acid [poly(A)] addition (32). On the other hand, poxviruses, which replicate exclusively in the cytoplasm, code for their own RNA polymerase and poly(A) polymerase and have evolved a different set of regulatory sequences (42).

Frog virus 3 (FV3) is a large DNA virus that belongs to the genus Ranavirus of the family Iridoviridae. This virus cannot be categorized as either nuclear or cytoplasmic, but has features of each. The initial round of DNA synthesis takes place in the host cell nucleus (16), but later in infection DNA concatamer formation and assembly into mature virions occur in the cytoplasm (14). FV3 appears to have no virion-associated RNA polymerase; furthermore, Goorha (13) has demonstrated that host cell RNA polymerase II is required for the first round of viral RNA synthesis (immediate-early RNA), and that this synthesis takes place in the nucleus. In contrast, late viral mRNA synthesis seems to be a cytoplasmic function, perhaps under the direct control of a virus-coded enzyme (R. Goorha, personal communication). Somewhat surprisingly, in view of the requirement for host RNA polymerase II, purified viral DNA is not infectious. It can, however, be non-genetically reactivated by virion proteins (47). This observation suggests that a virion protein must interact with the host cell RNA polymerase II to transcribe the viral DNA. Infection with FV3 alters the α-amantin binding site of cellular RNA polymerase II (3, 9), but we do not know how this alteration relates to the turning on of FV3 transcription. One reason that an interaction of a virion protein with polymerase might be necessary is the high degree of methylation of FV3 DNA. Apparently every cytosine in the sequence CpG is methylated (46), and methylated DNA is thought to be resistant to transcription by eucaryotic RNA polymerase II (12). Another explanation for the requirement for a virion protein is that it is needed for the host RNA polymerase to recognize viral promoter sequences that may differ from those of cellular genes.

Because of the unique position of FV3 in viral taxonomy, we are sequencing representative FV3 genes and comparing the immediate-early, delayed-early, and late-regulatory sequences to those of established eucaryotic promoters. If early and late viral mRNAs turn out to be synthesized by different enzymes, we anticipate that the DNA sequences regulating these syntheses will reflect this difference. Here, we report the coding and flanking nucleotide sequences of the gene for a major immediate-early infected cell RNA (ICR-169) that encodes an infected-cell protein of approximate molecular weight 18,000, ICP-18 (48). As this paper documents the first frog virus gene to be cloned or sequenced, it necessarily includes a substantial amount of technical detail.

MATERIALS AND METHODS

Cells and virus. Fathead minnow (FHM) cells were propagated at 33°C as monolayers in roller bottles or 90-mm tissue culture dishes with Eagle minimal essential medium

* Corresponding author.
containing 5% fetal calf serum. A clonal isolate of FV3 was used to prepare virus stocks by infecting cells at a multiplicity of 1 PFU/cell. Virus was harvested and assayed as previously described (31).

**Purification of viral mRNA.** FHM cell monolayers in 90-mm dishes were infected at a multiplicity of 20 PFU/cell (1 h, 25°C). Unadsorbed virus was removed, and the infected cells were overlaid with Eagle’s minimum essential medium containing 2% fetal calf serum and incubated at 30°C. When RNA was labeled with 32P, the overlay was phosphate free and contained dialyzed fetal calf serum. Labeled viral RNA was obtained by adding carrier-free 32P at a concentration of 100 μCi/ml at appropriate times after infection. RNA was extracted from the cytoplasm as described by Wilis et al. (48), and viral RNA species were separated by electrophoresis on acid urea-agarose gels (26). Bands of labeled RNA were visualized by autoradiography, and in some instances, labeled RNA was eluted from the gels and purified by the procedure of Landridge et al. (24). For some experiments, RNA was electroblotted onto a nylon membrane (Gene Screen Plus) by the procedures recommended by the manufacturer (New England Nuclear Corp.).

**Bacterial strains and plasmids.** The plasmids used in this work were pKB111 (5) and MBV17. Both of these plasmids were propagated in Escherichia coli strain LE-392, which was provided by G. Kitchingman. Plasmid MBV17 was a gift of J. Corden and is a derivative of pBR322 containing the left-end, 1,345 bp of adenovirus 2 cloned into the EcoRI site of the plasmid after filling in all ends. This plasmid contains an XbaI site, which was needed to clone the FV3 XbaI K fragment. The single-stranded M13 bacteriophage pair mp10 and mp11 and the host strain JM103 were purchased from Pharmacia Fine Chemicals. Growth of bacteria and purification of plasmids and all recombinant DNA methodology were essentially as outlined by Maniatis et al. (27).

**Cloning of FV3 restriction fragments.** Viral DNA was isolated as previously described (14) from a DNA methyltransferase-negative, 5-azacytidine-resistant mutant of FV3. It was necessary to use the unmethylated mutant DNA because the methylated DNA of the wild-type virus was refractory to cloning (R. Goorha, personal communication). Restriction digests of 5-azacytidine-resistant mutant DNA are identical to those of the wild-type virus, except for those enzymes inhibited by cytosine methylation. A KpnI digest of a 5-azacytidine-resistant mutant was cloned into the KpnI site of the vector pKB111, and a plasmid (BD18) containing FV3 KpnI fragment C was isolated (G. Chinchar, personal communication). An XbaI digest of this plasmid released the two FV3 XbaI restriction fragments, XbaI-F and XbaI-K (25). XbaI-K (3.5 kilobase pairs [kb]) was then subcloned into the XbaI site of vector MBV17 to give plasmid XK15. The XK15 plasmid was subsequently digested with BglI, and this digest was cloned into the bidirectional bacteriophage pair mp10 and mp11 (30) in the XbaI and BamHI restriction sites (XbaI-K is cut by BglI to give a 2-kilobase pair [kb] and a 1.5-kbp fragment). The identity of the clones was confirmed by nick translation of the plasmids (35) and subsequent hybridization to Southern blots of restriction digests of FV3 DNA (40). The RNAs coded for by the cloned regions were identified by hybridization of the nick-translated plasmids to northern blots of RNA from FV3-infected cells (1). Individual restriction fragments were separated from vectors by sedimentation in sucrose-ethidium bromide gradients (37) or by electrophoresis on low-melting-temperature agarose gels, with subsequent purification over NACS Prepac columns (Bethesda Research Laboratories).

**DNA sequence analysis.** DNA sequence analysis was carried out by the dyeoxy chain termination method (36) with the 2-kbp XbaI-BglII fragment (XB2) cloned into the single-stranded bacteriophage pair mp10 and mp11. The initial 200 to 300 bases of sequence were obtained with the M13 universal primer. The remainder of the sequence was obtained by “gene walking” with synthetic oligonucleotides of 12 to 13 bases complementary to the 3′-terminal nucleotides of the already sequenced regions (22). The oligonucleotides were synthesized with the Applied Biosystems apparatus and subsequently purified by polyacrylamide gel electrophoresis and NACS Prepac chromatography. The second-strand DNA sequence was obtained in a similar manner. Maxam and Gilbert (29) chemical sequencing was used to clarify ambiguous regions.

**Computer analysis.** Hydrophilicity analysis was performed by the method of Hopp and Wood (21). Data compiling and sequence analysis were done with the programs of Staden (41) adapted for use on a Data General MV8000 computer.

**Single-strand nuclease analysis.** Mung bean nuclease (19) was used to determine both the 3′ and the 5′ ends of the RNA, and S1 nuclease was used for the less precise determination of the 5′-end of the RNA. The oligonucleotide probe was hybridized with cytoplasmic RNA from infected cells under the conditions described by Berk and Sharp (8). After 3 h at 52°C, the samples were diluted in a 10-fold volume of mung bean buffer (0.3 M NaCl, 0.1 M sodium acetate, 0.001 M ZnCl2, pH 4.5), and mung bean nuclease or S1 nuclease was added at a concentration of 500 U/ml. Samples were then incubated at 37°C for 30 min. The reaction was stopped by the addition of EDTA to 0.01 M, and 5 μg of carrier yeast tRNA was added. The samples were extracted with phenol-chloroform and precipitated with 0.3 M sodium acetate and 3 volumes of absolute ethanol. The precipitates were dissolved in 10 μl of formamide sample buffer, heated to 90°C, and subjected to electrophoresis on polyacrylamide sequencing gels in parallel with the sequence representing the total fragments.

**R-loop analysis.** Cytoplasmic RNA extracted from FV3-infected FHM cells was enriched for ICR-169 by hybridization to and elution from nitrocellulose filters bearing the complementary single-stranded region cloned in bacteriophage mp10 (10). This RNA was then hybridized to the purified cloned XbaI K fragment under the same conditions used for mung bean nuclease analysis. After 3 h of incubation at 52°C, the samples were quickly diluted and spread for the electron microscope by the technique of Chow et al. (11).

**In vitro translation of FV3 mRNA.** The conditions for translation of FV3 mRNA in micrococcal nuclease-treated reticulocyte lysates (Promega Biotec) were modified from the procedures of Raghoff and Granoff (34). Each reaction contained 18 μCi of [35S]methionine, 10 μg of total cytoplasmic RNA isolated from FV3-infected BHK cells 6 h postinfection, and 5 μl of reticulocyte lysate in a final volume of 30 μl in vitro translation buffer. Assays were incubated at 30°C for 60 min and stopped by the adding 30 μl of 10% sodium dodecyl sulfate and heating to 100°C for 2 min. Samples from each assay were subjected to electrophoresis on a 5 to 15% gradient sodium dodecyl sulfate-polyacrylamide gel (23). After electrophoresis, the gels were prepared for fluorography by soaking in Enhance (New England Nuclear Corp.) and exposing the dried gel to Kodak XAR-5 film at −70°C. Hybrid absorbance analysis was carried out by hybridizing (65°C, 10 min) FV3 mRNA to single-stranded cloned DNA or to synthetic oligonucleotides in 10 mM Tris (pH 7.5)–50 mM potassium acetate–5 mM MgCl2. The sam-
RESULTS

The gene for ICR-169 is located near the end of XbaI fragment K. Preliminary results obtained by hybridizing labeled infected-cell RNAs extracted from urea-agarose gels to Southern blots of FV3 DNA restriction digests indicated that the gene for ICR-169 was located on the 3.5-kbp XbaI fragment K (D. B. Willis, M. Lee, and A. Granoff, Fifth Intl. Congr. of Virology Abstracts, W37/02, p. 344, 1981). To determine precisely from where on the fragment the approximately 500-nucleotide RNA was transcribed, as well as to determine whether there was any processing of the transcript by splicing, we turned to R-loop analysis with the electron microscope (11). Figure 1 illustrates the results of R-loop mapping between immediate-early FV3 RNA, obtained from cells infected in the presence of cyclohexamide (45), and the XbaI K fragment. Only one R-loop was formed between the immediate-early FV3 RNA and this fragment. The loop began at 80 to 100 bp from the end of the fragment and extended for 500 to 600 nucleotides, the approximate size of the total ICR-169 molecule as measured on denaturing polyacrylamide gels (48). As no other R-loops were observed, we assumed that splicing, if present, could involve only a very small portion of the molecule.

The gene for ICR-169 is located on the 2-kbp XbaI-BgII fragment. Further localization of the gene for ICR-169 RNA was facilitated by the discovery that the enzyme Bgll cut the 3.5-kbp XbaI K fragment into a 2-kbp and 1.5-kbp fragment. Figure 2 shows the results of S1 nuclease analysis after hybridization of these two fragments to FV3 RNA. Only a portion of the 2-kbp XbaI-BgII fragment was protected from S1 nuclease degradation; the length of the protected fragment, here calculated at 500 bp, was long enough to encompass the entire RNA. The 2-kbp XbaI-BgII fragment, hereafter known as XB2, was subcloned into the single-stranded bacteriophage pair mp10 and mp11 as a prelude to dideoxynucleotide sequencing as well as to determine the orientation of the mRNA.

The 5' end of the gene for ICR-169 is near the XbaI end of fragment XB2. By inference from the R-loop mapping, the coding area of ICR-169 was near the XbaI end of fragment XB2. However, we did not know whether the 5' or 3' end of the message was at this end, there being no long tracts of poly(A) on FV3 messages to aid in orientation (44). In the experiment illustrated in Fig. 3, S1 nuclease analysis was carried out after hybridization of 32P-labeled FV3 RNA to single-stranded DNA from XB2 cloned in either mp10 or mp11; the bidirectional cloning ensured that opposite strands of the cloned fragment would be in each clone. The results illustrate that essentially all of the ICR-169 was protected by

FIG. 1. The gene for ICR-169 is near one end of the XbaI fragment K. Immediate-early RNA was extracted from the cytoplasm of FHM cells infected with FV3 in the presence of 50 μg of cyclohexamide per ml. The RNA was hybridized to 100 μg of denatured XbaI fragment K immobilized on a nitrocellulose filter as described in the text. RNA was eluted from the filters and then hybridized to XbaI-K in 0.4 M NaCl-80% formamide (52°C, 3 h) before diluting and spreading for the electron microscope. SS, Single stranded R-loop.

FIG. 2. The gene for ICR-169 is located on the 2-kbp XbaI-BgII fragment. XbaI fragment K was cut by Bgll to give two fragments of 2 and 1.5 kbp, and the two fragments were separated by electrophoresis on 1% low-melting-temperature agarose. The DNA was extracted from the appropriate band as described in the text, hybridized in formamide to immediate-early RNA as described in the legend to Fig. 1, and digested with 500 U of S1 nuclease after dilution with the appropriate buffer (30 min, 37°C). The protected hybrid was subjected to electrophoresis on a 2% agarose minigel for 1 h. Lanes: 1, φX174 HaeIII digest markers; 2, FV3 RNA plus the 2-kbp XbaI-BgII fragment; 3, FV3 RNA plus the 1.5-kbp XbaI-BgII fragment.
the mp10 XB2 clone (Fig. 3, lane 2). No FV3 RNA was observed to hybridize to the mp11 XB2 clone (Fig. 3, lane 1). As nucleotide polymerization proceeds from the XbaI site of mp10 toward the BamHI site, the direction of mRNA synthesis must therefore be from the XbaI site toward the BglII site of the cloned fragment. In addition, it appeared that the entire RNA molecule was protected by the cloned DNA, since the protected RNA molecule was the same size as ICR-169 extracted from uninfected cells (Fig. 3, lane 3).

Detailed restriction map of XB2. To construct suitable probes for determining the 5' and 3' ends of ICR-169, we used a more detailed restriction map of XB2. The XbaI K fragment was 5' end labeled with \[^{32}P\]ATP and cut with BglII, and the XB2 fragment (now labeled at one end only) was purified by electrophoresis on low-melting-temperature agarose. The location of restriction sites for HpaII, Sau3A1, and HinFl were determined by the partial digestion method of Smith and Bernstiel (39). The map thus obtained is illustrated in Fig. 4; the 3'- and 5'-labeled probes used later to determine the initiation and termination sites of transcription are also shown. Confirmation of the position of ICR-169 on this restriction map was obtained by hybridizing \[^{32}P\]-labeled ICR-169, extracted from urea-agarose gels, to Southern blots of the various digests (Fig. 4).

Sequence of gene coding for ICR-169. Over 1,000 bases of the XB2 fragment were sequenced as described above. The sequence was compiled and analyzed by a computer, and the following information was obtained (Fig. 5). (i) The coding strand contained only one open reading frame large enough to accommodate the 18-kilodalton (18K) protein believed to be coded by ICR-169. This reading frame began with an ATG start codon 98 nucleotides from the XbaI site (position +19 in Fig. 5). A second ATG start codon was found four nucleotides downstream from the first in the same phase. (ii) A TAG stop codon was located 472 nucleotides downstream from the translation start site (position +490 in Fig. 5).

Translation in this reading frame would result in a protein of 157 amino acids. The total molecular weight of the protein represented by this sequence was 17,640, which is in good agreement with the 18,000-molecular-weight estimate obtained by polyacrylamide gel electrophoresis (48).

Sequences at 5' end of ICR-169. Because FV3 requires host RNA polymerase II for synthesis of immediate-early RNA (13), we were particularly interested in comparing the promoter sequences of FV3 messages with those of other eucaryotic genes. The only restriction site in our map suitable for determining the 5' region by single-strand nuclease mapping was an HpaII site about 275 bp from the presumed start site of transcription (Fig. 4). The 355-bp XbaI-HpaII fragment was 5' labeled with \[^{32}P\]ATP by polynucleotide kinase, hybridized to RNA from FV3-infected cells, and digested with mung bean single-strand nuclease (43). We then synthesized an oligonucleotide identical to this 5'-labeled sequence that was the same sense as the coding strand and had the digested HpaII site at its 5' end (5' CGGCCGCCCC-TCCACC 3', which corresponds to the reverse of positions 261 through 277 in Fig. 5). This primer was used for dideoxy sequencing, and the sequencing reactions were run on the same gel as the mung bean nuclease-protected fragment (Fig. 6). The results show that the major protected fragment terminated within a highly AT-rich region; a number of smaller, minor bands were also observed at one-nucleotide intervals. These may represent true minor start sites, or they may be artifacts caused by RNA degradation or nuclease nibbling. A close inspection of the sequence surrounding this region (Fig. 6) revealed a
second AT-rich region at position -29 bases upstream from the major start site of transcription. Although slightly different from the canonical TATA box (6), this sequence is located in the correct position to serve the same function. Interestingly, except for one base, the two AT-rich regions represent 11-bp inverted repeats. This complementarity could lead to formation of a stem-and-loop structure such as is often found in or near transcriptional promoters (19).

Genome location of 3' ends of ICR-169. FV3 mRNAs lack poly(A) (44), so we did not expect to find an AATAAA polyadenylation signal near the 3' terminus of the gene. However, some specific recognition signals for termination of transcription or terminal processing must be located in that area. A 630-bp HpaII-Sau3A1 fragment, labeled at the 3' HpaII end (Fig. 4), was used to map the 3' terminus of ICR-169 by the mung bean nuclease method (43). The size of the protected fragments was determined by polyacrylamide gel electrophoresis in parallel with a Maxam and Gilbert sequence of the entire HpaI-Sau3A1 fragment (Fig. 7). The number of protected fragments observed suggests that the transcript terminated, either directly or via a processing mechanism, over a range of 25 nucleotides, with a minor stop at 32 bases and a major stop at 51 bases from the translational stop codon (TAG). The microheterogeneity of such sites could be due to a mung bean nuclease artifact or to RNA degradation, but it is clear that most of the ICR-169 molecules terminated within or near the sequence TATAAAA, which in eukaryotic cells is a classical Goldberg-Hogness box for promotion.

Hybrid arrested translation of ICP-18, the protein coded by ICR-169. Although we had shown by both Southern and

TCTAGATCITTTAAAGACTGGCATCAGCTCTCAGAAGGGAATGTAACGACTTACAGTATTTACG
-70  -60  -50  -40  -30  -20

M R M I O A Y L
CACATGCGCTACTTAAAAATCTACTTTAGCAGCTATCGCCATGAACTGACTCCTTCT
50  60  70  80  90  100

C O S V E C P Y T K G C L E I P P
GTCGACTACCTTGCTCCGAGAATGCTATGGCAGCCTGACACTGTCCCTGCTTATCACTACTCCCT

110  120  130  140  150  160

N N N F T I D L V N S V S T E P Q V K
TACAGAAATCTCATACAGATTTGTCCACCTGCTCTCTCTACTACAGATCCAGCTA

I T M T P H O L G T F V V P E P N V
GATCACTACCTTGCTCCGAGAATGCTATGGCAGCCTGACACTGTCCCTGCTTATCACTACTCCCT

170  180  190  200  210  220

S I K R A V K G D A A F K V E R A A G W
TTCTACATACAGAAGGCGATTTAAAGGGAAGGCAATTAGCTGGAAGCGGGGGGGGTTG

230  240  250  260  270  280

L P D T P Q V L T L F V V E R L N P V
GCTGCGCCGACCTCGGCAAGCTTGGCCGACCTGCTTCGCGGACCTGCTTATCACTACTCCCTG

290  300  310  320  330  340

W H S C M Y E N L E T D G G Y V I P
ATGCGACAGGAGGCTGCATGAAACCGAGATCGGCGGGGCGACCTGCTCCC

350  360  370  380  390  400

G E A T G Q R F G T A T E V P T M M L F
CGCGAGGCGCGGCGCGCAAGCTTGGCCGACCTGCTTCGCGGACCTGCTTATCACTACTCCCTG

410  420  430  440  450  460

K R M F V V K G V *
TAAAGAAGTGTCTTGGTAAAGGCTTAGACATTCAGGCTTACCAGCTGAG

FIG. 5. Complete nucleotide sequence of gene and deduced amino acid sequence of polypeptide coded for by ICR-169. The presumed start site of transcription is designated at +1. The two in-phase translational start codons are underlined, and the translational stop codon is indicated by an asterisk. The sequence, beginning at the XbaI site, is that of the noncoding strand, or the same sense as the mRNA.

FIG. 6. The start site of transcription is within an AT-rich area. The nuclease protected fragment (mung bean) was obtained after hybridizing the 5'-labeled XbaI-HpaII fragment (Fig. 4) to RNA extracted from FV3-infected cells and digesting with mung bean nuclease as described in the text. The deoxy sequencing reactions were primed by a synthetic oligonucleotide with the exact 5'-end sequence of the labeled, protected fragment. This primer was hybridized to a single-stranded phage clone containing the noncoding strand (mp11 XB2); therefore the coding strand sequence is displayed in the sequencing ladder. The reactions were run on an 80-cm 6% polyacrylamide urea gel at 1,500 V for 36 h. The wet gel was wrapped in Saran wrap and exposed to Kodak XAR-5 film at -70°C overnight. The 5' flanking sequence of the noncoding strand of the gene for ICR-169 is shown beginning at the XbaI site. An arrow points to the major start site for transcription, designated as position 1. The two AT-rich inverted repeats are boxed, and the translational start codons are underlined.

Northern blot analysis that our sequenced clone contained the region complementary to ICR-169, the previous evidence that this RNA coded for ICP-18 was based on size and kinetics of formation (48) and therefore circumstantial. To confirm the relationship between ICR-169 and ICP-18, we programmed a reticulocyte lysate with FV3 mRNA that had been hybridized to either single-stranded DNA corresponding to the whole coding strand (Fig. 8, lane 4), the whole noncoding strand (Fig. 8, lane 3), an oligonucleotide complementary to positions 189 through 201 of the message (Fig. 8, lane 5), or an oligonucleotide complementary to the two AUG translational start codons at positions 17 through 31 (Fig. 8, lane 6). The synthesis of ICP-18 was specifically eliminated by hybridizing the RNA with the coding strand of DNA, and it was significantly reduced by the oligonucleotide 5'GATCATGCGCATTG3', which is complementary to the translation start site (Fig. 6). Interestingly, the oligonucleotide complementary to the internal position of the message failed to inhibit translation of ICR-169 (Fig. 8, lane 5). Other oligonucleotides, complementary to various regions along the length of the gene, were equally ineffective in blocking translation of ICP-18 (results not shown). As the
FIG. 7 The 3' end of ICR-169 is microheterogeneous. The 1,100-bp fragment obtained from an HpaII digest of fragment XB2 (Fig. 4) was 3' labeled with [32P]PCTP and Klenow DNA polymerase and then digested with Sau3A1 to remove the irrelevant 3' end. The resulting 630-bp fragment was purified by electrophoresis on low-melting-temperature agarose and NACS column chromatography before chemical sequencing and mung bean nuclease mapping with FV3 RNA as described in the text. The sequence of the noncoding strand surrounding the 3' end of ICR-169, beginning with the TAG translational stop codon (double lines), is shown. The heterogeneous regions of termination are underlined, and the TATA-like region is boxed.

conditions for translation in vitro are not as stringent as those required for maintaining hydrogen bond formation between the RNA and the oligonucleotide, it seems likely that the impetus toward translation, once initiated, was sufficient to remove loosely bound oligonucleotides. The results leave no doubt that the gene we have sequenced is for the mRNA ICR-169 and that this RNA is the message for ICP-18, as we previously predicted (48).

DISCUSSION

We have sequenced the entire coding region of a major immediate-early FV3 gene and shown that it codes for the mRNA (ICR-169) for polypeptide ICP-18. In addition, we have determined the probable location of both the start site of transcription and the approximate position of the 3' terminus of the message. Without direct RNA sequence data, we hesitate to assign a unique 5' penultimate nucleotide as the transcriptional start site, but the apparent absence of splicing observed by R-loop analysis (Fig. 1) supports the site we have determined.

Since other data suggest that immediate-early FV3 RNAs are transcribed by host RNA polymerase II modified by a virion-associated protein (13, 47), we were particularly interested in investigating the 5' flanking (promoter) region of the gene and comparing it with promoter regions of other cellular and eucaryotic viral genes (38). The 5' end of ICR-169, determined by mung bean nuclease mapping, was found to lie within an AT-rich region that was preceded at position −29 by the sequence TATTITA; this sequence, although not a classic TATA box (6), was in the proper position to serve such a function. Moreover, the two AT-rich regions were part of an 11-bp inverted repeat that might form a stem-and-loop structure similar to those seen at other promoter regions (38). Several other faint bands at approximately one-nucleotide intervals were also noted by mung bean mapping. These may be artifacts, but mung bean nuclease is not supposed to cause any nibbling (19). These multiple bands could have resulted from degradation of RNA during extraction or may represent true minor start sites. Microheterogeneity of the 5' end of a message is often seen in the absence of a classical TATA box (32). Genes that lack TATA boxes are all viral and appear to initiate transcription within a microheterogeneous region—for example, the adenovirus E1A and E1B genes (4). Viruses in the simian virus 40-polyoma group also do not possess classic TATA boxes upstream from the late genes, and they initiate late transcription at a variety of sites spread over 100 bp (38).

Both adenovirus and simian virus 40-polyomavirus DNAs are transcribed by host cell RNA polymerase II, showing that TATA boxes are not essential for transcription. Vaccinia virus, which synthesizes its own DNA-dependent RNA polymerase to transcribe RNA in a cytoplasmic locale, has evolved unique signals for the initiation of transcription (42). Of the 60 nucleotides immediately upstream from the major start site of an early vaccinia virus mRNA, 88% were AT. The vaccinia virus gene contained no homologies to the CAAT box in the 70 to 90 region that has been observed in many other eucaryotic messages (7). We also found no CAAT-like region in the 5' flanking region of ICR-169.

Although we cannot eliminate the possibility of splicing of the ICR-169 transcript, three lines of evidence suggest that this message is not spliced. (i) [32P]-labeled RNA protected by the cloned DNA migrated on polyacrylamide gels at the same position as ICR-169 from infected cells (Fig. 3). (ii) The R-loop map did not reveal a free single-stranded 5' end (Fig. 1), although a very short one could not have been detected. (iii) The sequence around the presumed start site (Fig. 8) did not contain the AG 3' splicing signal (6).

The sequence of the 3' end of the ICR-169 was also determined and found to be microheterogeneous, terminating principally in a sequence that reads AATATAAA (Fig. 8).
As FV3 mRNAs do not contain poly(A) (44), we did not expect to find the poly(A) addition signal site AATAAA, which is typically found 10 to 30 nucleotides upstream from the poly(A) tail itself (33). However, the termination of ICR-169 within a TATA box sequence deserves some mention. One could speculate that the reason for the abrupt switch-off of host cell mRNA synthesis after FV3 infection (45, 48) results from normal eucaryotic promoters being recognized as terminators. However, because unknown and complex sequences surrounding the TATA box are equally important in controlling transcription in vivo (38), this hypothesis is highly speculative. Venkatesan et al. (42) also found heterogeneity at a vaccina virus early message terminus, and the typical poly(A) signal was missing as well. Although vaccina virus messages do contain poly(A), it is added by a virus-specific enzyme, and the signals for its addition are different from the cellular poly(A) polymerase.

The 3′ ends of all mRNAs that have been thoroughly analyzed—including the poly(A)-negative histone mRNAs—appear to be generated by simple cleavage of longer transcripts (33, 38). More experiments need to be carried out to determine whether the 3′ end of ICR-169 is also the result of processing or is generated by transcription termination, as we have suggested here.

Although we now know the DNA sequence of the coding and flanking regions of the gene ICR-169 for polypeptide ICP-18, we have no clue as to the function of the protein itself. It is unquestionably a very important molecule, for it is synthesized at a high rate throughout infection and is stringently regulated at the posttranscriptional level (15, 17). The hydrophobic nature of this polypeptide (data not shown) would tend to rule out any interaction with nucleic acids; in fact, Aubertin et al. (2) have clearly shown that ICP-18 is not a DNA-binding protein.

The presence of an AT-rich region in the same position as the TATA box suggests that this region functions as a promoter. However, we have failed to transcribe the cloned gene either in HeLa whole cell extracts in vitro (28) or in vivo after injection into Xenopus laevis oocytes (20). Perhaps this failure is not too surprising, for the absence of infectious DNA indicates that a viral protein is somehow required for transcription (47). Unfortunately, extracts from infected cells contain too much virus-induced nuclease to identify any definitive run-off transcripts, even if present. We are presently following three approaches to identify the regulatory sequences of immediate-early FV3 genes. One is to sequence additional immediate-early genes and to look for sequences homologous to those we have observed at the 5′ and 3′ ends of ICR-169. Another possibility for our failure to obtain transcription in vitro is that “enhancer” sequences not included in our clone are required for efficient expression of ICR-169 mRNA. Therefore, we are extending the 5′ end of our clone to include sequences in the adjoining fragment. The third and most promising approach is to insert our presumed promoter into a vector bearing the bacterial chloramphenicol acetyltransferase gene (18), to introduce this reconstructed vector via transfection into appropriate cells concomitant with FV3 infection, and to look for FV3-induced expression of the bacterial gene.

ACKNOWLEDGMENTS

We thank K. G. Murti for the electron micrograph and Clayton Naeve for the computer analysis and oligonucleotide synthesis. We also thank R. Staden for providing the computer programs and Ted Culp for adopting the DNA sequencing program to the Data General MV/8000. Mary Shuck, Evelyn Stigger, Janet Winkler, and Kathy Troughton provided excellent technical assistance.

This study was supported by Public Health Service research project grant CA 07055 and Cancer Center Support (CORE) grant CA 21765 from the National Cancer Institute and by American Lebanese-Syrian Associated Charities.

ADDENDUM IN PROOF

The ligation of nucleotides −79 to −2 of the ICR-169 gene (Fig. 5) in front of the coding sequences for chloramphenicol acetyltransferase produced a plasmid that expressed chloramphenicol acetyltransferase only in FV-3 infected cells. This result demonstrates that the promoter for ICR-169 is within this 78-bp region and that it is activated by a viral protein(s).

LITERATURE CITED

18. Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyl-


