

Three Splicing Patterns Are Used To Excise the Small Intron Common to all Minute Virus of Mice RNAs

WILLIAM R. MORGAN¹ AND DAVID C. WARD^{1,2*}

Departments of Human Genetics¹ and Molecular Biophysics and Biochemistry,² Yale University School of Medicine, New Haven, Connecticut 06510

Received 25 June 1986/Accepted 8 September 1986

We identified three splicing patterns used to excise the small intron common to all three transcripts encoded by minute virus of mice. Sequence analysis of minute virus of mice-specific cDNAs indicated that two donor and two acceptor splice sites were used: in pattern 1, the most frequent, nucleotide 2280 was spliced to nucleotide 2377; in pattern 2, nucleotides 2317 and 2399 were joined. Oligonucleotide probes, each specific for one of the four possible splice junction sequences, were synthesized and hybridized to viral mRNAs immobilized on nitrocellulose filters. The probes specific for splice patterns 1 and 2 hybridized to all three viral mRNAs, as did a third oligomer specific for a splicing pattern in which nucleotides 2280 and 2399 were joined. The fourth potential splicing pattern, linking nucleotides 2317 and 2377, was not detected. The presence of three splicing patterns in the transcripts designated R2 and R3 would allow the translation of five distinct polypeptides from these two mRNAs.

Minute virus of mice (MVM), an autonomous parvovirus, possesses a linear, single-stranded DNA genome 5,149 nucleotides long (1, 2). The transcriptional and translational organization of MVM DNA has been elucidated from extensive studies with the prototype strain, MVM(p) (Fig. 1). The three viral mRNAs, transcribed from two overlapping transcription units, are polyadenylated and spliced, and each includes a small intron between map units 46 and 48 (16). Although previous experiments have assigned specific viral polypeptide products to each mRNA (5, 6), neither the absolute number nor the amino acid sequence of these proteins has been established definitively because of uncertainty about the precise splice junctions and the translational initiation sites used.

The initial nuclease protection experiments used for RNA mapping indicated that the length of the leader exon from map units 40 to 46 was heterogeneous (16). Furthermore, analysis of the sequence around the small intron identified numerous potential donor and acceptor splice sites which could generate nine possible splicing patterns (2). Comparative sequence studies demonstrated that several autonomous parvoviruses had conserved two of the potential 5' and one of the potential 3' splice sites in this region of their genomes (1, 3). In MVM DNA, these sequences occur at nucleotides 2280, 2317, and 2399, respectively. It has been hypothesized that the alternative use of the two conserved donor sites with the conserved acceptor site could generate two forms of the major viral transcript R3, which would then code for capsid protein VP-1 or VP-2. Alternative splicing within the R3 transcript was further suggested by recent studies with an MVM-bovine papillomavirus chimera containing only the downstream (P-39) promoter and therefore presumably transcribing only the R3 mRNA species. Cells transformed by the chimera DNA expressed both capsid proteins in the same molar ratio as did cells lytically infected with MVM (8).

To establish unambiguously the number and location of splice sites within the MVM transcripts, we deduced viral RNA sequences from cDNAs synthesized by primer exten-

sion of a radiolabeled complementary oligonucleotide. The cDNAs were analyzed directly because the overlapping transcriptional organization complicates identification of the RNA origin of a cDNA and because this approach avoids potential artifacts that could arise during the construction or analysis of cDNA clones.

Poly(A)⁺ RNA was isolated from uninfected or MVM(p)-infected A9 cells 20 to 24 h postinfection by standard methods (11). Oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer and purified by polyacrylamide-urea gel electrophoresis and then on a Sep-Pak column (10). Oligonucleotides were end labeled with polynucleotide kinase (12). An oligonucleotide complementary to viral RNA just downstream of the small intron shared by all viral messages was used as a primer (Fig. 2B). This primer was hybridized to mRNA from MVM(p)-infected or uninfected A9 cells and extended with reverse transcriptase essentially as described by Ghosh et al. (7), except that the purification of primer mRNA hybrids by oligo(dT) cellulose chromatography was omitted. Several specific cDNAs were observed with mRNA from MVM(p)-infected cells (Fig. 3): two products slightly larger than 400 nucleotides (B₁ and B₂) and another doublet of approximately 0.7 kilobases (A₁ and A₂). The other minor bands observed were probably prematurely terminated cDNAs or derived from degraded RNA.

The cDNAs of approximately 400 nucleotides were individually eluted and subjected to Maxam-Gilbert sequence analysis (12). A different splice junction was identified in each cDNA. In the shorter, more abundant cDNA (B₂), nucleotide 2280 (D1 in Fig. 2A) was spliced to nucleotide 2377 (A1, pattern 1); nucleotides 2317 (D2 in Fig. 2A) and 2399 (A2, pattern 2) were joined in the less abundant, longer cDNA (B₁). Other cDNAs were not analyzed further because of insufficient radioactivity.

Based on previous nuclease mapping studies (16), full-length primer extension cDNAs derived from intact R3 transcripts were expected to be about 400 nucleotides long, R2 derivatives were expected to be approximately 700 nucleotides long, and R1 derivatives were expected to be more than 2 kilobases long. Although we expected that both 400-nucleotide-long cDNAs were derived from R3, they

* Corresponding author.

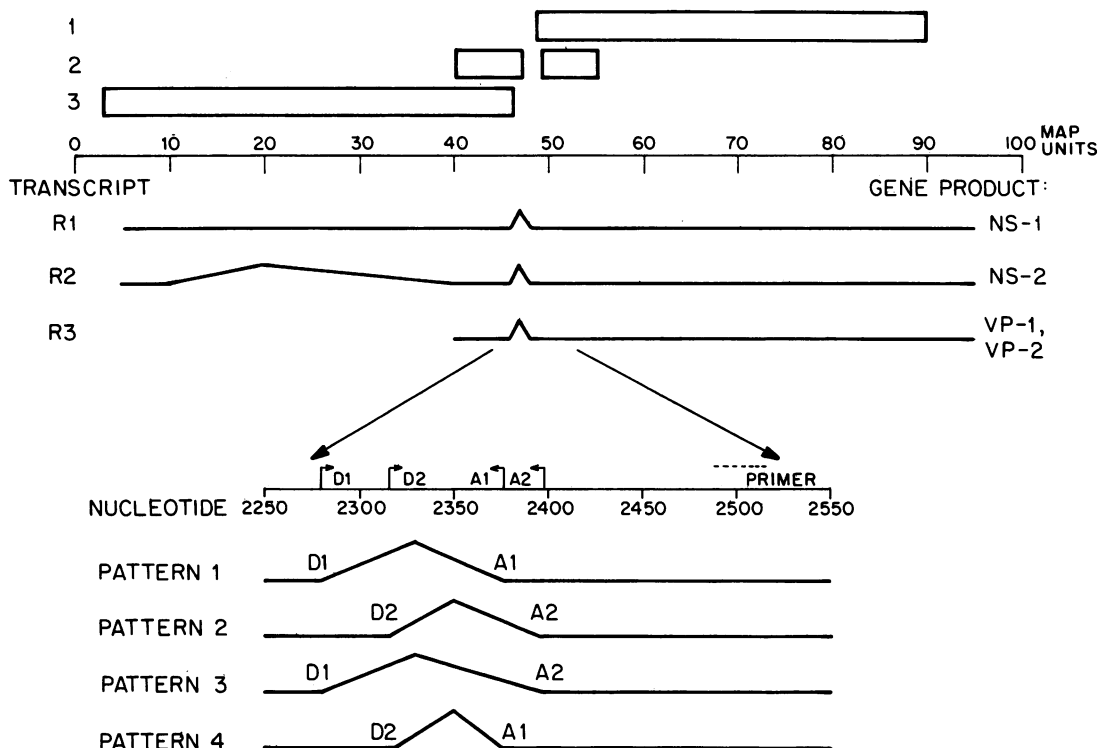


FIG. 1. Transcriptional and translational organization of the MVM genome. The map coordinates and general splicing patterns of the three viral transcripts, R1, R2, and R3, as determined by Pintel et al. (16) are shown in relationship to the two large and two small open reading frames previously identified (2). The translation products of each transcript were determined by hybrid selection experiments (5, 6) and by transfection studies of C127 cells with MVM-bovine papillomavirus chimeras (8). The location of the primer used for cDNA synthesis and the donor-acceptor sites shown here to generate the alternate splicing patterns for excision of the common intron between map units 46 and 48 are indicated at the bottom of the figure.

A

```

2251      D1                                D2      2350
ACTTCAGCGAGCCGCTGAACTTGGACTAAGCTACGATGGCGCCTCCAGCTAAAAGAGCTAAAAGAGCTAAAGGGTTAAGGGATGGTTGGTTGGTGGGGTA
TGAAGTCGCTCGGCGACTTGAACCTGATTTCATGCTACCGCGGAGGTGGATTTCTCGATTTCTCCATCCCAAAATCCCTACCAACCAACCCACCCCAT
XXXXXXXXXXXXX
MetAlaProProAlaLysArgAlaLysArgG

2351      A1                                A2      2450
TTAATGTTTAAATACCTGTTTTACAGCCCTCAAATCACTTGGTTTTGGTTGGCTGCCTCCGGCTACAAGTACCTGGGACCAGGGAACAGCCTTGACCA
AATTACAAATTAATGGACAAAATGTCCGGACTTTAGTGAACCAAATCTAAGCCGACCGAGACCGATGTTTCATGGACCCTGGTCCCTTGTCCGAACTGGT
XXXXXXXXXXXXX
1yTrpValProProGlyTyrLysTyrLeuGlyProGlyAsnSerLeuAspG1

2451      2550
AGGAGAACCACCAATCCATCTGACGCCGCTGCCAAAGAGCACGACGAGGCTATGATCAATACATCAAATCTGGAAAAATCCTTACCTGTACTTCTCT
TCCTCTTGGTTGGTTAGGTAGACTGCGGCGACGGTTTCTCGTCTGCTCCGGATACTAGTTATGTAGTTAGACCTTTTTAGGAATGGACATGAAGAGA
nGlyGluProThrAsnProSerAspAlaAlaAlaLysGluHisAspGluAlaTyrAspGlnTyrIleLysSerGlyLysAsnProTyrLeuTyrPheSer

```

B

OLIGONUCLEOTIDE	SEQUENCE	SPLICE JUNCTION PROBE FOR:
PRIMER	CTCGTGTCTGCCGATACTAGTTATGTAG	-----
OLIGO A	TGAACCTGATTCCGGACTTTAGTG XXXXXXXXXXXXX+++++	D1 TO A1
OLIGO B	TCTCGATTTTCTCCAACCCACGGAG XXXXXXXXXXXXX//////////	D2 TO A2
OLIGO C	TGAACCTGATTCCAACCCACGGAG XXXXXXXXXXXXX//////////	D1 TO A2
OLIGO D	TCTCGATTTTCTCCGGACTTTAGTG XXXXXXXXXXXXX+++++	D2 TO A1

FIG. 2. (A) Nucleotide sequence of MVM(p) near the region of small intron. The nucleotide sequence is revised from that of Astell et al. (1). D1 and A1 indicate the donor and acceptor splice sites of splicing pattern 1, respectively; while D2 and A2 similarly indicate the donor and acceptor splice sites of minor splicing pattern 2. Below the nucleotide sequence is the putative amino acid sequence for the amino terminus of VP-1 (see the text). Nucleotide components of the genome which are juxtaposed in the oligonucleotides used in these studies are indicated by X, *, +, /, or -. Note that sequences complementary to the splice junction-specific oligonucleotides are interrupted in the genomic DNA by intron sequences. (B) Sequences of synthetic oligonucleotide probes. All oligonucleotides are complementary to viral RNA. Sequences are written in the 3' to 5' direction for easy alignment with panel A. Symbols are as in panel A.

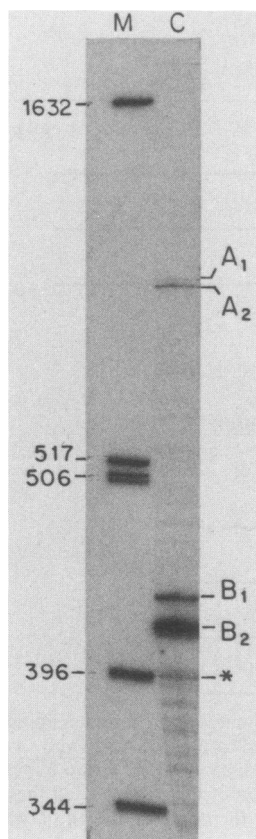


FIG. 3. MVM-specific cDNAs migrating as doublets. cDNAs were synthesized by using poly(A)⁺ RNA from MVM(p)-infected A9 cells. Samples were electrophoresed on a 5% polyacrylamide-urea gel until the xylene cyanol migrated about 60 cm (two gel lengths). Lane M, Marker DNA fragments with nucleotide lengths indicated at the left; lane C, cDNA products from MVM(p)-infected cells. A1 and A2, Minor cDNA bands of about 700 nucleotides; B1 and B2, major cDNA products slightly larger than 400 nucleotides. The asterisk indicates a cDNA band equivalent in size to that expected from RNA with splicing pattern 3 (see the text).

might also have been derived from R1 or R2, either being prematurely terminated products or degraded viral RNA fragments. To verify that R3 did indeed use both splicing patterns and to determine if these same splices were present as well in R1 or R2 transcripts, we probed Northern blots with oligonucleotides (Fig. 2B) specific for each splice junction. RNA and DNA samples were electrophoresed on agarose-formaldehyde gels and transferred to nitrocellulose after a mild alkaline treatment by standard procedures (11). The filters were then hybridized with ³²P-labeled MVM plasmid DNA (nick translated as described by Maniatis et al. [11]) or with oligonucleotide probes (5' end labeled with polynucleotide kinase). Posthybridization washes were done as described by Leary et al. (9), except as noted below. Previous studies have shown that sequences diverging by a single nucleotide can be distinguished by oligonucleotide probes (4). For maximum specificity, oligomers complementary to sequences symmetrically spanning each splice junction were used. Under high-stringency conditions, these oligonucleotide probes should hybridize uniquely to the appropriate splice junction and not to unspliced or other, alternatively spliced messages. As expected, the oligonucle-

otide primer used for cDNA synthesis hybridized to all three viral RNAs, as well as to an MVM genomic DNA clone (Fig. 4D, lanes 1 and 2). Similar results were obtained when plasmid pMM984, an infectious MVM clone constructed in pBR322 (14), was used as the probe (data not shown). Interestingly, both the oligomer specific for splicing pattern 1 (oligo A, Fig. 2) and the oligomer specific for splicing pattern 2 (oligo B, Fig. 2) also hybridized to all three viral RNAs with high specificity (Fig. 4A and B, lane 2). In contrast, no hybridization signal was detected with MVM genomic DNA (Fig. 4A and B, lane 1), a plus-strand genomic RNA synthesized by using SP6 RNA polymerase as described in Melton et al. (13), or poly(A)⁺ RNA from uninfected cells (Fig. 4A and B, lane 3) when either the oligo A or oligo B probe was used, thus confirming the specificity of the splice junction probes. In addition, the ratio of the signals detected by equivalently labeled splice junction oligo A and B probes was similar to that seen for the major and

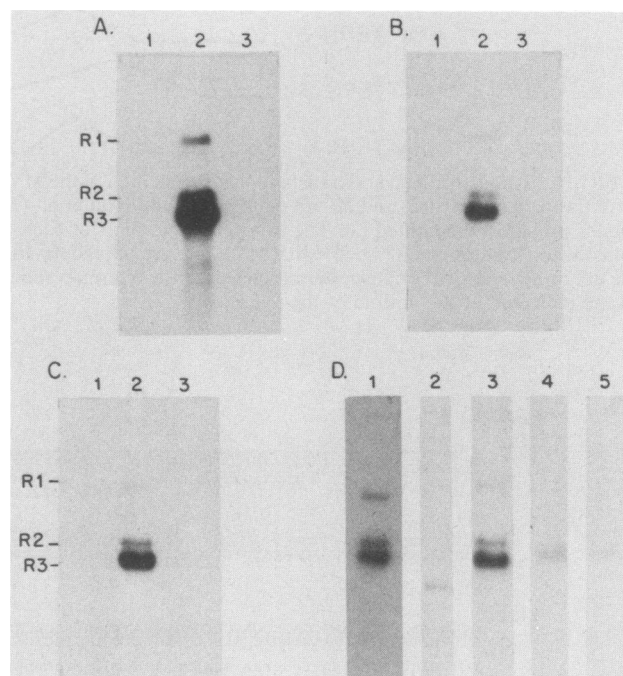


FIG. 4. Oligonucleotide probes specific for each splice junction hybridizing to all three viral RNAs. Samples were electrophoresed on an 0.8% formaldehyde-agarose gel, transferred to nitrocellulose filters, and probed with the oligonucleotides shown in Fig. 2B. In panels A, B, and C, oligo A, oligo B, and oligo C, respectively, were hybridized to the genomic DNA clone pMM984 digested with *Eco*RI (lanes 1), 1 μg of poly(A)⁺ RNA from MVM(p)-infected A9 cells (lanes 2), or 1 μg of poly(A)⁺ RNA from uninfected A9 cells (lanes 3). The specific activity of each oligonucleotide was $\sim 2.5 \times 10^8$ cpm/μg. Autoradiography was for 18 h at -70°C with an intensifying screen. In panel D, each lane was similarly exposed to demonstrate the relative abundance of each splicing pattern. Lanes 1, 3, 4, and 5 contain 1 μg of poly(A)⁺ RNA from MVM(p)-infected cells, and lane 2 contains pMM984 digested with *Eco*RI. Lanes 1 and 2 were hybridized with the primer (Fig. 2B), and lanes 3, 4, and 5 were hybridized with oligo A, oligo B, and oligo C, respectively. Oligonucleotides were similarly labeled ($\sim 2.5 \times 10^8$ cpm/μg), and autoradiography was for 18 h at room temperature with no intensifying screens. Hybridization was at 42°C for the primer and at room temperature for oligo A, oligo B, and oligo C. Stringent washes were done in $0.2\times$ SSC ($1\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 45°C for panels A, B, and D and at 40°C for panel C.

minor cDNAs approximately 400 nucleotides long. The signal from each viral RNA was proportional to the relative abundance of that RNA species in the infected cell (16), and the melting temperatures of the oligomer-RNA hybrids were equivalent for all three RNAs with each splice junction probe (data not shown). Furthermore, since both oligomers were radiolabeled to the same specific activity, it is apparent that each transcript possesses both alternative splicing patterns in approximately the same relative molar ratio.

Oligonucleotides specific for the other possible splicing patterns that use the identified donor and acceptor sites were also synthesized and used to probe Northern blots. Oligo C, which is specific for the D1-A2 splicing pattern (pattern 3), hybridized to all three viral RNAs (Fig. 4C), while no signal was detected with oligo D, which is specific for the D2-A1 splicing pattern (data not shown). The hybridization signal observed with oligo C was slightly lower than that seen with oligo B when probes of identical specific activity and film exposure times were used (Fig. 4D, lanes 4 and 5), and both gave a much weaker signal than did oligo A (Fig. 4D, lane 3). A cDNA with splicing pattern 3 was probably overlooked during our initial sequence analysis because of its relatively low abundance. Interestingly, reinspection of the primer extension data indicated that the band marked with an asterisk in Fig. 3 had the size expected for a full-length cDNA with a pattern 3 splice junction.

Based on the putative splice sites conserved among several autonomous parvoviruses, previous authors (1, 3) hypothesized that by alternatively joining two donor splice sites at nucleotides 2280 (D1) and 2317 (D2) to a common acceptor site at 2399 (A2), one could produce two forms of R3 to differentially code for VP-2 and VP-1, respectively. Our studies demonstrate that these two splicing patterns are indeed used; however, the most frequently used splicing pattern in MVM joins nucleotides 2280 (D1) and 2377 (A1), a splice site not previously identified to be conserved. A close reexamination of the sequences in this region reveals that feline panleukopenia virus (3), H-1 virus (18), canine parvovirus (17), and MVM(i) (the lymphotropic variant [1]) have all conserved an AG dinucleotide approximately 20 base pairs upstream of the conserved acceptor site previously identified. It appears then that in a subset of R3 transcripts which codes for VP-1, the removal of nucleotides 2318 to 2398 (D2-A2, pattern 2) allows the AUG at nucleotide 2286 to be spliced in-frame with the long right-hand open reading frame (Fig. 2A). However, in the majority of R3 transcripts the putative VP-1 initiation codon is removed by splicing nucleotide 2280 (D1) to nucleotide 2377 (A1, pattern 1) or 2399 (A2, pattern 3). As previously suggested (3), the removal of this AUG may be necessary for efficient translation of VP-2 initiating at nucleotide 2795 (8, 15).

Recent experiments indicate that the NS-2 protein is translated from the R2 transcript and that reading frame 2 is used in the second exon, which is just upstream of the common intron (6). Because translation of the second exon occurs in reading frame 2, the three splicing patterns could result in three NS-2 polypeptides with different carboxy termini. The predominant form, translated from messages with splicing pattern 1, would have a carboxy terminus of -Leu-Arg-Pro-Glu-Ile-Thr-Trp-Phe. The two minor species would terminate with -Leu-Arg-Tyr-Asp-Gly-Ala-Ser-Ser (from pattern 2 RNA) or -Leu-Arg-Leu-Gly-Ala-Ser-Trp-Leu-Gln-Val-Pro-Gly-Thr-Arg-Glu-Gln-Pro (from pattern 3 RNA). The role that three different carboxy peptides might play in the function of NS-2 is not known. Translation of the NS-1 protein from the R1 transcripts by use of the long

left-hand open reading frame 3 is not affected by the different splicing patterns, since this reading frame terminates upstream of both donor sites.

A fourth possible pattern, D2-A1, is not used to remove the common intron. Experiments with truncated introns of the rabbit β -globin gene (20) showed that efficient splicing occurred with an intron of 81 nucleotides but not with one of 69 nucleotides. Indeed, the vast majority of introns are larger than 75 nucleotides (although the splicing of shorter introns has been seen at low efficiency), suggesting that the minimum intron size for efficient splicing is about 80 nucleotides (19, 20). It is interesting therefore to note that three of the observed splicing patterns remove introns larger than 80 nucleotides (pattern 1, 96 nucleotides; pattern 2, 81 nucleotides; pattern 3, 118 nucleotides), while the fourth splicing pattern would remove introns of only 59 nucleotides.

This work was supported by Public Health Service grants CA-16038 and AI-19973 from the National Institutes of Health.

LITERATURE CITED

1. 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.
2. Astell, C. R., M. Thomas, M. Merchlinsky, and D. C. Ward. 1983. The complete DNA sequence of minute virus of mice, an autonomous parvovirus. *Nucleic Acids Res.* 11:999-1018.
3. Carlson, J., K. Rushlow, I. Maxwell, F. Maxwell, S. Winston, and W. Hahn. 1985. Cloning and sequencing of DNA encoding structural proteins of the autonomous parvovirus feline panleukopenia virus. *J. Virol.* 55:574-582.
4. Conner, B. J., A. A. Reyes, C. Morin, K. Itakura, R. L. Teplitz, and R. B. Wallace. 1983. Detection of sickle cell β^s -globin allele by hybridizations with synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA* 80:278-282.
5. Cotmore, S. F., L. J. Sturzenbecker, and P. Tattersall. 1983. The autonomous parvovirus MVM encodes two nonstructural proteins in addition to its capsid polypeptides. *Virology* 129:333-343.
6. Cotmore, S. F., and P. Tattersall. 1986. Organization of nonstructural genes of the autonomous parvovirus minute virus of mice. *J. Virol.* 58:724-732.
7. Ghosh, P. K., V. B. Reddy, M. Piatak, P. Lebowitz, and S. M. Weissman. 1980. Determination of RNA sequences by primer directed synthesis and sequencing of their cDNA transcripts. *Methods Enzymol.* 65:580-595.
8. Labieniec-Pintel, L., and D. Pintel. 1986. The minute virus of mice P₃₉ transcription unit can encode both capsid proteins. *J. Virol.* 57:1163-1167.
9. Leary, J. J., D. J. Brigati, and D. C. Ward. 1983. Rapid and sensitive colorimetric method for visualizing biotin-labeled DNA probes hybridized to DNA or RNA immobilized on nitrocellulose: bio-blotting. *Proc. Natl. Acad. Sci. USA* 80:4045-4049.
10. Lo, K.-M., S. S. Jones, N. R. Hackett, and H. G. Khorana. 1984. Specific amino acid substitutions in bacteriopsin: replacement of a restriction fragment in the structural gene by synthetic DNA fragments containing altered codons. *Proc. Natl. Acad. Sci. USA* 81:2285-2289.
11. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual, p. 109-112, 191-193, and 202-203. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
13. 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.
14. **Merchlinsky, M. J., P. J. Tattersall, J. J. Leary, S. F. Cotmore, E. M. Gardiner, and D. C. Ward.** 1983. Construction of an infectious molecular clone of the autonomous parvovirus minute virus of mice. *J. Virol.* 47:227-232.
 15. **Paradiso, P. R., K. R. Williams, and R. L. Constantino.** 1984. Mapping of the amino terminus of the H-1 parvovirus major capsid protein. *J. Virol.* 52:77-81.
 16. **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.
 17. **Rhode, S. L., III.** 1985. Nucleotide sequence of the coat protein gene of canine parvovirus. *J. Virol.* 54:630-633.
 18. **Rhode, S. L., III, and P. R. Paradiso.** 1983. Parvovirus genome: nucleotide sequence of H-1 and mapping of its genes by hybrid-arrested translation. *J. Virol.* 45:173-184.
 19. **Upholt, W. B., and L. J. Sandell.** 1986. Exon/intron organization of the chicken type II procollagen gene: intron size distribution suggests a minimal intron size. *Proc. Natl. Acad. Sci. USA* 83:2325-2329.
 20. **Wieringa, B., E. Hofer, and C. Weissmann.** 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β -globin intron. *Cell* 37:915-925.