

Gene Organization and Transcription of TED, a Lepidopteran Retrotransposon Integrated within the Baculovirus Genome

PAUL D. FRIESEN* AND MARK S. NISSEN†

Institute for Molecular Virology and Department of Biochemistry, The Graduate School and College of Agricultural and Life Sciences, University of Wisconsin—Madison, Madison, Wisconsin 53706

Received 27 December 1989/Accepted 13 March 1990

A single copy of the retrotransposon TED, from the moth *Trichoplusia ni* (a lepidopteran noctuid), was identified within the DNA genome of the baculovirus *Autographa californica* nuclear polyhedrosis virus. Determination of the complete nucleotide sequence (7,510 base pairs) of the integrated copy indicated that TED belongs to the family of retrotransposons that includes *Drosophila melanogaster* elements 17.6 and gypsy and thus represents the first nondipteran member of this invertebrate group to be identified. The internal portion of TED, flanked by long terminal repeats (LTRs), is composed of three long open reading frames comparable in size and location to the *gag*, *pol*, and *env* genes of the vertebrate retroviruses. Sequence similarity with the dipteran elements was the highest within individual domains of TED open reading frame 2 (*pol* region) that are also conserved among the retroviruses and encode protease, reverse transcriptase, and integrase functions, respectively. Mapping the 5' and 3' termini of TED RNAs indicated that the LTRs have a retroviral U3-R-U5 structural organization that is capable of directing the synthesis of transcripts that represent potential substrates for reverse transcription and intermediates in transposition. Abundant RNAs were also initiated from a site within the 5' LTR that matches the consensus motif for the promoter of late, hyperexpressed baculovirus genes. The presence of this viruslike promoter within TED and its subsequent activation only after integration within the viral genome suggest a possible symbiotic relationship with the baculovirus that could extend transposon host range.

The retrotransposons represent a class of eukaryotic transposable elements that bear a striking resemblance to the vertebrate retroviruses in genetic structure and function. Identified thus far in the plant, fungus, and animal kingdoms, these elements share a common mechanism for transposition involving reverse transcription of an RNA intermediate. Numerous families of retrotransposons containing long terminal repeats (LTRs) have been identified in invertebrates, including more than 10 such families in *Drosophila melanogaster* alone (for recent reviews, see references 5 to 7). On the basis of their size and overall gene organization, the *D. melanogaster* elements 17.6, 297, and gypsy most closely resemble the vertebrate retroviruses (for retrovirus reviews, see references 48 and 49). These retrotransposons possess three long open reading frames (ORFs) comparable to the retroviral *gag*, *pol*, and *env* genes (24, 28, 41); of these, the *pol* region is typically the most highly conserved among the retrotransposons and the retroviruses, a possible consequence of the importance of reverse transcription in the maintenance of these elements through evolution (25, 29, 47). The LTRs of the *D. melanogaster* elements also possess a U3-R-U5 structural organization that is used to progress through stages of reverse transcription that are analogous to those of the retroviruses (2). Finally, the *D. melanogaster* retrotransposons, like the retroviruses, represent an important class of insertion mutagens owing to their ability to integrate randomly into the host genome and alter gene expression (5, 7).

We have investigated the structure and gene organization of the retrotransposon TED, a lepidopteran element (7.5 kilobases [kb]) that, as we report here, is closely related to

the dipteran elements 17.6 and gypsy. TED represents a dispersed, middle repetitive sequence within the genome of the cabbage looper *Trichoplusia ni*, a nocturnal moth (31). Active transposition of this element within *T. ni* is evident from its amplification and alternate locations within the genome of cultured cells compared with that of larvae. Moreover, TED is capable of transposing to the DNA genome of the baculovirus *Autographa californica* nuclear polyhedrosis virus (AcMNPV) during viral replication in susceptible *T. ni* cells (31). Such large insertions of foreign DNA, including a variety of other host-derived transposable elements, can be accommodated by the double-stranded, circular genome (~128 kb) of AcMNPV without compromising replication functions (for baculovirus reviews, see references 17 and 32). TED was first identified as an insertion of host *T. ni* DNA within the genome (86.7 map units) of virus FP-D, an AcMNPV mutant distinguished from wild-type virus by the reduced number of occluded virus particles it produced in cell cultures (31, 38). Integration disrupted an early viral gene and resulted in alterations in the transcription of surrounding genes (18, 19).

Identification of a single copy of TED integrated within the AcMNPV genome has provided an opportunity to examine the structure, expression strategies, and mutagenic properties of an active retrotransposon in a context different from that of the cell in which it resides. In this study, we report the complete nucleotide sequence of TED as integrated within AcMNPV mutant FP-D. Since the LTRs play an essential role in the expression of transposon-encoded genes, element transposition, and alterations in the expression of nearby genes, we have also examined the transcriptional activities of the TED LTRs embedded within the baculovirus AcMNPV genome and the host *T. ni* genome. Transcriptional mapping indicated that the TED LTRs possess a U3-R-U5 structural organization analogous to that of

* Corresponding author.

† Present address: Department of Biochemistry, Washington State University, Pullman, WA 99164.

the retroviruses; moreover, full-length (genomic) RNAs were synthesized from the 5' LTR to the 3' LTR. These data, in addition to the presence of a gene with striking sequence similarities to the *pol* gene of the dipteran retrotransposons and the vertebrate retroviruses, support the contention that TED transposition occurs via reverse transcription of an RNA intermediate.

MATERIALS AND METHODS

Recombinant plasmids and DNA sequencing. Overlapping restriction fragments containing portions of element TED and flanking viral DNA at the site of integration (86.7 map units) of AcMNPV mutant FP-D were cloned into the Bluescript (KS) vector (Stratagene, La Jolla, Calif.) by standard methods (27). The unidirectional exonuclease III deletion procedure of Heinkoff (21) was used to generate progressive deletions of each plasmid in the directions indicated in Fig. 1B. Single-stranded phagemid DNA was isolated after infection of plasmid-containing *Escherichia coli* JM101 cells with VCS-M13 helper bacteriophage (Stratagene) and was sequenced with T7 DNA polymerase (U.S. Biochemicals, Cleveland, Ohio) in conjunction with the dideoxy chain termination method of Sanger et al. (42). Nucleotide sequences shown were determined for both DNA strands. Computer-assisted sequence comparisons and analyses were conducted with the sequence analysis software package of the University of Wisconsin—Madison Genetics Computer Group (14).

Cells and viruses. Established moth cell lines *T. ni* TN368 (cabbage looper; Lepidoptera: Noctuidae) (22) and *Spodoptera frugiperda* IPLB-SF21 (fall army worm; Lepidoptera: Noctuidae) (50) were propagated in TC100 growth medium (GIBCO Laboratories, Grand Island, N.Y.) supplemented with 10% heat-inactivated fetal bovine serum. AcMNPV insertion mutant FP-D was isolated after 25 serial passages of wild-type strain L-1 of AcMNPV through the *T. ni* TN368 cell line (31, 38). Virus FP-D is referred to as an FP (or few-polyhedron) mutant because of the reduced numbers of polyhedral occlusion bodies it produces in cell cultures relative to wild-type AcMNPV. FP-D represents a mixture of two virus genotypes, FP-DL and FP-DS, containing the full-length TED element or a single (solo) TED LTR, respectively. During infection, FP-DL spontaneously generates a mixture of FP-DL and FP-DS in an approximate ratio of 4:1, whereas the FP-DS genotype is stable (31).

RNA isolation and Northern (RNA) blot analysis. For isolation of RNA from uninfected or virus-infected cells, cell monolayers (10^7 cells per 100-mm plate) were mock infected

or inoculated with extracellular virus at a multiplicity of 20 PFU per cell. After a 1-h adsorption period, residual inoculum was removed and replaced with fresh growth medium. Cells were harvested 20 h later by dislodging the monolayer and washing the cells with phosphate-buffered saline (pH 6.2). Total RNA was extracted by the guanidine isothiocyanate-cesium chloride method (11), and poly(A)⁺ RNA was purified by oligo(dT)-cellulose chromatography (23).

For Northern blot analysis, poly(A)⁺ RNA was denatured by glyoxalation (30), subjected to electrophoresis on 1.2% agarose–10 mM sodium phosphate (pH 7.0)–1 mM EDTA gels, and transferred to nitrocellulose membranes (46). Blots were hybridized with RNA probes for 40 h at 65°C in 25 mM Tris hydrochloride (pH 7.5)–1× PE (0.1% sodium PP_i, 5 mM EDTA, 1% sodium dodecyl sulfate, 0.2% each polyvinylpyrrolidone, Ficoll, and bovine serum albumin)–5× SSC (1× SSC is 0.15 M NaCl plus 15 mM trisodium citrate)–50% formamide–150 µg of denatured salmon sperm DNA per ml. Strand-specific RNA probes were synthesized by *in vitro* transcription of plasmid DNA as described previously (3). Bluescript (KS) plasmids containing appropriate DNA fragments of the TED element juxtaposed with T3 or T7 bacteriophage promoters were linearized with various restriction endonucleases and transcribed with T3 or T7 RNA polymerases, respectively, in the presence of [α -³²P]UTP (800 Ci/mmol; Amersham Corp., Arlington Heights, Ill.). Greater than 90% of the resulting RNA transcripts were full length, as judged by denaturing agarose gel electrophoresis.

Primer extension and S1 nuclease analysis of TED transcripts. For mapping of the 5' ends of RNAs initiated from the 5' TED LTR, a 36-nucleotide primer (extending from nucleotide positions 303 to 268; Fig. 1) was radiolabeled exclusively as its 5' end (located within the internal portion of TED) with T4 polynucleotide kinase and [γ -³²P]ATP. The end-labeled primer was annealed to poly(A)⁺ RNA from infected or uninfected cells and extended with Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) as described previously (34).

The 3' ends of RNAs terminating within the 3' TED LTR were mapped by the S1 nuclease procedure of Weaver and Weissman (51). A 433-base-pair (bp) *Sau3A-EcoRI* probe (extending from nucleotide position 7222 within the internal portion of TED to the *EcoRI* site located 149 bp from the end of the 3' LTR within flanking viral sequences) was 3' end labeled exclusively at the *Sau3A* site with the Klenow fragment of *E. coli* DNA polymerase and [α -³²P]dGTP. The probe was denatured, annealed to poly(A)⁺ RNA under the

FIG. 1. Nucleotide sequence of TED as integrated within the genome of AcMNPV mutant FP-D. (A) The sense strand of TED is shown along with the predicted translation products of the three ORFs (ORFs 1, 2, and 3) as indicated by single-letter amino acid designations. The first methionine of ORF 1 is underlined, and termination codons are indicated by asterisks. Amino acids corresponding to putative protease (prt), reverse transcriptase (rt), RNase H (rnh), and integrase (int) domains of ORF 2 that are the most conserved among the retrotransposons and retroviruses are also underlined. Brackets mark the boundaries of the 5' and 3' LTRs, which are further subdivided into U3, R, and U5 regions. The U3-R and R-U5 junctions within the 5' and 3' LTRs, respectively, designate the initiation and polyadenylation sites of TED RNAs as synthesized in host *T. ni* cells (see text). The RNA start sites that are located within the baculoviruslike promoter (boxed sequences in 5' LTR) and that are utilized after viral integration (see text) are indicated by small arrows. Also underlined are the putative polyadenylation signal (PA), the tRNA PBS, the polypurine tract (PP), and the 4-bp duplication (dup) (AATG) of viral DNA at the site of integration. (B) Restriction map of TED (darkest area) illustrating the plasmids and the direction of exonuclease III deletions used to obtain the sequence. Also shown are the locations of RNA probes (arrows) used in Northern hybridizations (see text): probe P₂P₁ (0.98 kb) extends from the *PstI* site at position 1237 (P₂) to the *PstI* site at position 261 (P₁), probe SsS (1.13 kb) extends from the *SstI* to *Sall* sites (positions 2831 to 1705), probe BgX₁ (0.71 kb) extends from the *BglII* to *XbaI* (X₁) sites (positions 5414 to 4706), and probe X₂H (0.54 kb) extends from the *XbaI* (X₂) to *HindIII* sites (positions 6606 to 6070). The restriction sites at the boundary of each probe are indicated in the sequence. Restriction site abbreviations: B, *BamHI*; Bg, *BglII*; H, *HindIII*; K, *KpnI*; P, *PstI*; R, *EcoRI*; S, *Sall*; Ss, *SstI*; V, *EcoRV*; X, *XbaI*; Xh, *XhoI*; Xm, *XmaI*.

A ^{5'}LTR
1 TGTTAGTATGGAGCCTTAATGGATATCATCGACGCTGCATTTTCGTGTTATTGTCGCCAGCTGCATAGAAACTGTCTGAATGACGTAATCGTCATGAACCGCTGATGACGAAATTT 120
- US | R - PA - R | US -
121 GTAATTAGTATTAACTCAAAATGTATGCATTCCTATTCTAAATCGAGTAGGCTCCAGCATTAAATTTGTAAATCTTATAAGTAATAATAGCATTAAAAATCATTITGGTTT 240
PstI 5'LTR tRNA PBS
241 TTTTCTATCTCGCGCTCGCAGTATACGTAATTTGGCGACTCGGTAGGATACTCGGGGGCGGTCGGGAGCAATAGAGTCATCTCATATCGATATTCGGGTCGGGAGTCGCCGGGGC 360
361 GCTTCAAGCCACATCTGAGTATCCTGAATCTACGGGCAACTGCGGAGCCAGCGGGAATCCAGACATCGATTCCGAGAGAAAACCAATATGAAACTTATGTAAGTACACTTGA 480
ORF 1 -
481 TCGTTCCTGTTGCTTCCTTCTATCTAGTACTTTAATCTCTAAAAATTTTTCGTCACGAAAAGTTTTACACCCGGGAACCTGCATCACGCCAGGCTCGCCGAATATCATTGATA 600
S L K I F R H E K F Y T R E P A S R Q A R R N I I D I
601 TTAATAGAGACTTAGAATTCAGATTTAACTTAGGTTAGATAGATTATTTAGTATAAAAATGCCACAGCATCTGACGTAATTTCAAGGCTTTGGCACTTGTCCGGAATTTAATG 720
N R D L D N S D L N L G L D R L F S I K M P H D P D V I F K A L R L V G P E A N G
721 GCAACCCCAATATTTAACGAGATTATAAATATATGTGTAACCTAGTAGCAATATGCCAGCGCTGAGCGGGAAGTGAAGTAGGAAATTTATGTTGTAATGGCATCTTGAACA 840
N P M I L T R F I N I C D K L V E Q Y A S A E P G S E L G N L C L L N G I L N K
841 AGGCTACTGGAACGCTGCCTTACCATAAACGCAATGGCATTCTGAAACCTGGTAGGCATTAGATCATCTTTAATTAACAACCTTTTCAGACCGCGGATGAAACGGCTTTATATA 960
V T G T A A S T I N A N G I P E T W V G I R S S L I M N F S D Q R D E T A L Y M
961 ATGACCTCTCATTAGCTTACAAGGTAATAAGACTCTCAGGAGTCTACGAAACAATGCCAACCTTATTCAGTACCAATAGCAGTATGAACTGGCATGAGACTTTACCAACGACTA 1080
D L S L A S Q G N K T P Q E F Y E Q C Q T L F S T I M T Y V T L H E T L P T T I
1081 TCGAAGCTAAACGGGCACTTTACAAGAAAGTAACGTGTCAGGCTTTTGTGCGGGGACTAAAAGACCTTTAGGTTACGTATAAGGTGATGCGCCCGGAGACTATCGAAGAACCCCTG 1200
E A K R A L Y K K V T V Q A F V R G L K E P L G S R I R C M R P E T I E K A L E
PstI
1201 AATATGTGCAAGGAGCAATTAATGTAATATCTGCAAGCAAGCAATGAGTCTCGAGGCTCACAGCTCCTTAAATGCTTCTATACCGCAACAGCTCTCTGACCCCAATCAATA 1320
Y V Q E E L N V I Y L Q R W E S S R A H S S P K M L P I P Q Q S P V T P F N T
1321 CATTAGGTTATCACAGCCGGTACCTAATGCGGCTTCGATGGGCAACGTCGCAATCAACCCACCTCAACCCCTTAAATTTAATGTGCTAACCAATATCATAAATCGCAT 1440
L G I N R P P V P M W P V P M Q P P P P P P P F K F N V P N H G R M P
1441 CTACTAAAACCAACAGTCTTAGAGCACCGCCACCAATATCATCTCAGAGTAACGTTTTCCGCTTACCACCAGTAATCCACCACCAATCAAAATGTAACCCGATGAGTGGAG 1560
T K T Q Q M L R A P P P N Y H P Q S N V F R L P P R N P P P N Q I V K P M S G V
1561 TTCAACATTTTCCCAAAAACCTTACCTGTAATGACGGGACATGACTGGCGTAAATCCGGGAATCCGCCCAAAATAACTTCAAAAACCTCGCGAATTAACGTTAACGAATTTACT 1680
Q H F V P K T L P V M T G H D W R K S G N P P P N N Y F K T R E L N W E F Y S
SalI
1681 CGTCTGACGACTCATATAACTCAGTCGACTATTTCGGAACGGGTGGGACTATTACTGACTATTATAACAACCCCTAGACTATGACACAACGGGACTGTTACGACTTACCTT 1800
S D S Y N S V D Y Y S E P G C D Y Y T D Y N W P Y D Y D T N A T C Y D L P Y
ORF 2 - ORF 1
1801 ACGACGCTAGTGAGACAGAAGCTCAACCGCCCTAGCCATGTACATGAAAGTCAGGATTTTCACTGACCAAAACCTCAACCAAGCAAGGATAGATATAACCTACAGTACCAAGACA 1920
D A S E T E A Q P G P S H V H E S Q D F Q S T K P S N E Q G *
K S G F S I D Q T I K R T R I D I N L Q Y Q R Q
1921 ACTACCATATAGAATTTCTGATCCACCATTAAATTTCTGATTGACACTGGTCCCAATCAATCATTTATAGCCCAAGCCGTCACAAAATACTTTTCCAAATCTCGGTGAATTA 2040
L P Y I E F S D P P L K F L I D T G A N Q S F I S P Q A V Q K Y F S N Y S V N Y
prt
2041 CGACCCATTTGAAATAACGAACATACCGGTGTAGTGAATGAGCACTCAATTCATACCATGTTTCCAGGAGTTTAAACGAAACCAAGATATTAATTTATATATACCATTTTCA 2160
D P F E I T M I N G V S R N E H S I T L P C F Q E F N E T Q D I K L F I Y H F H
2161 TGATATTTTCGATGGATTAATAGGACTAGATTATTGTCTAAATGGGAGCCAGATAGATTTAAAGACTTTTTACTAATAACTAAATTTGCAACTAACCCGATTAAGCTATATAACTC 2280
D Y F D G L I G L D L L S K W E A K I D L K D F L L I T K F A T N R I K L Y M S
2281 TCGTAATGTCAACCTGTACGAGGATGATACCTGCAAGAAGTCTAAGTTAGTAAGGATACCTTAAACGCTACGGTACGGGAGTTTGTAGAGAAACAATGTTTGAACATGCAT 2400
R N V N L Y E D M I P A R S S K T A A G T R I P I N A T D G E V L V E E Q N C V C I
2401 TGTCATGAATCGGTTACCATGGTAAAAGTGGCGTGGATGTGAGAATTAGAGAATCCAAACCCCTAATGATGTAATTTTTACCTGGATCAGCCAGCTCTCGCGAATTTCAATAT 2520
V H E C V T M V K D G R G Y V E L E N P T P N D V I F Y L D O P A S A E L F N I
2521 CAAGTGCACCAAGTTGAACAGTCAACGTTGACGATGTTTATCAGGATTCGGTACAGACCATCTCAATGAGGAGAAAAGCTAACCTTTAAGACTTTGCTCGATTTACAGA 2640
K C T Q V E Q S Q R V D D V L S R L R T D H L N E E E * K A N L L R L C S R Y S D
2641 TGATTTTACATGACGGGGAAGCCCTTACATCACTAATAAAAATAAACACCGTATAAGAAACACGGGAGGATACCCGTACACCAAAAAGTTACCGGTACCCCTTATCCATCCGCA 2760
V D F T R E G E A L T F T N K I K H R I R T D E V P V Y T K S Y R Q C Y A N R T A
2761 GGAAGTTAGGCAACCAATCACGAAAATTTGGACCAAGGAATATAAGACCATCAGACTCTGCATGGAGCTCACCCATATGGTGTGCCAAGAAAATCGACGCTTCTGGAAACAAAA 2880
E V R D Q I T K M L D Q G I I R P S D S A W S S P I W V V P K K I D A S G K Q K
2881 GTGGCGTCTCGTAGTTGACTTCGGTAAGTTGAACGAGAAGACTATCGATGACAAAATACCCGATACCAACATAAGTGACGACTTGAACAAGTAGGTAAGTGGCAACTACTCACCACCTT 3000
W R L V D F R K L N E K T I D K A Y P I N I S D V L D K L G K Y F D T T L
3001 AGATTTGGCAAGTGGTTTTATCAGGTGGAGATGGACCTCAAGATATATCGAAAACCGGGTTTAACTGAGAACACGGGCAATTTGAATTCCTCGAATGCCTATGGGATAAAAAATC 3120
D L A S G F Y Q V E M D P Q D I S K T A F N V E H G H F E F L R M P M G L K M S
rt
3121 ACCATCTACTTTTCAAAGAGTTATGGCAATGTCTAAGAGGCTCCAAAAATACATCTGTCTCGTCTACCTTGACGATATTATGCTATAGTACTTCCCTACAGGAACACTGGAGAA 3240
P S T F Q R V M D N V L R G L O M N I C L V Y L D D I I V Y S T S L Q E H L E N
rt
3241 CCTGGAACGAGTTTCAAAGACTTAGAGAAAGTAATCTCAAAAATCAAATGGCAAGTCCGAATCTTGAAGCTCGAACTGCTTATCTTGGTCACATCATAAGCAGGACGGTATCAA 3360
L E R V Q R L R E S N F K I G A V L D K A S E F L K L E T A Y L G H I S R D G I K
3361 GCTAACCCCTGATAAGATTTCCGCTATTCAAAAATCTGATTCAAAAGACCCCTAAGAAAATAAAAACAATTTTTAGGCCCTCTCGGTTATACCAGAAAATTCATTCCAGATTTTGCACG 3480
P N P D K I S A I Q K Y L I P K T P K E I K Q F L G L L G Y Y R K F I P D F A R
3481 ACTCACAACCCCTTACACAGTGCCTAAAAAAGTAGTAAAGTAACCTTAGTCCGGAATGTAAGTCTTTGAACACTGTAACCACTGTTAACCAACGACCAATATTAACAATA 3600
L T K P L T Q C L K K G S K V T L S P E Y V N A F E H C K T L L T N D P I L Q Y
3601 CCCAGACTTTACCAGAGAAATTAACCTCACGACAGCGCTTCAACTTCGCTATGGAGCGGACTATCCCAAGGACCAATAGGATCCGCAACCCCGCTGTGTACGCTCTCGAACACT 3720
P G F T R E F N L T T D A S M N F A I G A V L S Q G P I G S D K P V C Y A N R T A
rrh
3721 CAATGAGAGGCAACTAAACTATAGCAATAGAGAAAGAAATTAAGTCTGCTTGGGCTACAAAATATTTAGACCCCTACTTATCGGTAGAAAAATTAAGATATTGACTGACCCAAA 3840
N E S E L N Y S T I E K E L L A I V W A T K Y F R P Y L F G R K F K I L T D H K
rrh
3841 ACCACTACAGTGGATGAACTTAAAAGACCAAACTCACGAATGACTAGTGGCAGTACGCAATGAAATGACTTCTCTGATGTACAAGAAAGGAAAGCTAATACCAACCGC 3960
P L Q W M N L K D P N S R M T R W R L R L S E Y D F S V V Y K K G Y A N T A
3961 TGATGCCCTTCTCGTGTGAGATCCATACCAGGAAATAGACGAAATGACTCAGTAATAGAAAACCTTAAAGAACTTAGCTCAATGATTAACCCCTCCGAGACATCTCGACCTCA 4080
D A L S R V E I H T T E I D E I D S V I E N I K E L S S M I N N P S E T S R P Q

4081 AAGCAAAACCGAAAACAGCAGCAAACTCAGACAGAACTCGACAACCGACACTGTACATACTAGTGATGAACACCCCTATTCTGGAAGTGCCCATACAAAAGCAACCACTCAACAGATTCA 4200
 R Q T E M T D E I R Q N S T T D T V H T S D E H P I L E V P I T M E P L N R F H

4201 TAGACAGATTTCATCTACCGTAGTAGGACATAAAACGACGACCTATTGTGCAAAAACCTTTGAAAGTCATACCCGAATGCAATCCAACATCAGAGTCAAACCTAGAACAAAGATG 4320
 R Q I H L T V V G D I K R R P I V T K P F E S H T R I A I Q L S E S N L E Q D V

4321 TATAAGTCCATCAAGAATACGTAAACCCAAAAGTCAAAACAGCCCTGATCATAAATCCGCCCTAAAGATGATTTCTATTATCCCTATTATACAAAAGACGCTCAGAAGTTCATCCCT 4440
 I S A I K E Y V N P K V K T A L I I N P P L K M Y S I I P I I Q K T F R S S L

4441 TAACCTAGTGTAAACCAAGTCGAACCTGAAAATGTCAAAGAGTATCTTAGACAACAAGCATTATACGACATACCATGACGGAAAGACAACCCCGGGAAATAAATGAATGCTACCT 4560
 N L V L T K V E L E N V K E Y L R Q Q D I I R H Y H D G K T M H R G I N E C Y L

4561 AGCACTCTCAAAAAGTATTACTGGCCAGGATGAAAGTCAAACTCACTAAATTTATCAATGAGTGTACTATCTGGTCAAGCAAATATGACAGGAACCAATACGACCTCAGTTAA 4680
 A L S K R Y Y W P R M K D Q I T K F I N E C T I C G Q A K Y D R N P I R P Q F N

4681 TATTGTACCCCGACGTACGAAACCTCTAGAGACCGTTCATATGGACCTATTACAGTTCAAATGAGAAATATAACGTTCAATGACGATTACGAAAGTCGGTCAGGACATACCACCT 4800
 I V P P A T T K P L E T V H M D L F T V Q W E K Y I T F I D V F T T Y K Y G Q A Y H

4801 ACGTGATGGCACCCTATTAGTATTTACAGGATTTTACAGTATTTGCACTCATCAGGATACCCATAACCAATAGTACTGATAACGGCACCAGAAATTTCCAACTCAATTTCTCTGA 4920
 R D G T A I S I L Q A L L R F C T H G L P I T I V T D N G T E F S M Q L F S E

4921 ATTCGTAGTATTATAAGATAATTCATCATAAGACCTTACCCACAGCCGAGCATAACGAAATATTGAACGTTCCATCCACAATTTCTGAGCATATCGAATTTCAAACCTACA 5040
 F V R I H K I I H H K T L P H S P S D N G N I E R F H S T I L E H I R I L K L Q

5041 ACATAAGGATGAACCAATTTGTAACCTTATGCCATACGCTATACAGGCTACAATAGTTCATACATAGTTTCAACAAATGACAGCCTTCGATCTACAAATGGACACTTTGATCCAA 5160
 H K D E P I V N L M P Y A I I G Y N S S I H S F T K C R P F D L L S N G H F D R

5161 GGACCCGCTTGACATAGACCTAACCGAACACATCTGACGCAATATGCTCAGAATCATCGTCAACAGATGAAACAAGTCTACGAAATATCAATGAAACATCTCTGCTAATCGTACGGC 5280
 D P L D I D L T E H I L Q Q Y A Q N H R Q Q M K Q V Y E I I M E T S L A N R T A

5281 TTTAATAGAAAGTAGAAAACAAACCGCGAATCTGAAGTAGAATACATCCCTCAGCAGCAAGTATTCATTAACCACTCTGCCACCGCTCAGAAGGTAGCACCACGCTATACCCAGGA 5400
 L I E S R M K T R E S E V E Y I P Q Q Q V F I K M P L A S R Q K V A P R Y T Q D

5401 TACGGCTTAGCAGATCTGCCATACACATCTATACGCTTAAAAACGTTGGGCGGTAGCTAAAGCGGCACTAAAACGTTTCTTAAAGGTAACACATGTTACAGGACTCTGCTGCTAC 5520
 T V L A D L P I H I Y T S K R G P V A K A R L K R V P K G N T L L Q D S A A T
 ~ ORF 2 N V F L K V T H C Y R T L L L L

5521 TGCAATACATGGCAGCCTCTCAAGAGATAAGACTTGAGACCTAGTGTGACCGGACTATTACCATAACAACCTGGGACCGACGACTAACCATACACTATCATCTCTTATAC 5640
 D N T C D A S S R D K T *
 T I H A T H P Q E I R L E T L A D G P G L L P Y K L G P T R L T I H Y N S F I Q

5641 AACCCATTGACCTCAACGATATTGAAAATAAAATCGACTCTGACAGACTCAACTTAATACATTTAGGACTAAACTCGATAACGAAACTTACCTACTCTATGAATCAAAATGATTATC 5760
 P I D L N D I E M K I D S V Q T Q L N T F R T K L D N E T Y L L Y E Y Q I D Y L

5761 TTACTAATAAGGTTGGCAATTAACCTTACTACACAAAATAAATCTTTAGAACCTGTTAGAGTAAAAGAGGCTTATAGATGGCTAGGGCTATAGTAAAAAGTGTCTGCGCACTTAGACT 5880
 T N K V P G K L L H Q I K S L E P V R V K R G L I D G L G S I V K S N V T G N F L D Y

5881 ACCAAGATGCCCTTAAATACGACGAGGCTCTTAAACCTTACAGACCAACGAAGGCAATTAACATCAGAATTTAATAGCCATCTGAGTCTTTGCAAAAGTGGATGCCCAACACAATA 6000
 Q D A L K Y D E A L K T L Q T N E G K L T S E F N S H L S L C K E W M S Q H N K

6001 AAGTGTAGAACCACTAACTTTAATCAAAATAGAGTAAATGCCACTTGAACCTACTATTACAAAAGAGCTTATAGGCACTATAGCTTAATTAATTTGCGAAATCGCAAAATCT 6120
 V L E Q L T L N Q I R V M A T L E L L L Q K E A Y R D Y S L I K F A K F A Q I L
 HindIII

6121 TAGGAATTAACAACCAACGATAGAGATTTAATGTTAGAATAATCAGATAGAAAATATGATGGCTTTTATACGGCATCTAGTACTCATTCATGATTGATAGAGGCTTGC 6240
 G I I T N N V E D L M L E I I R L E N M M A F I R A S S T H H S I G A E A L Q

6241 AGTCAATGATAGATAGATTAATCCCTTTATACTCCAAATCAAATCTAAATTTAGAACCTTAGGGAATATTACAGCTTAATCAAGCCAGGATCTTATTTCATCGATAAAGCTATAGTAA 6360
 S M I D R L K S L Y T P N Q I L N L E L R E Y Y S L I K P G S Y F I D K R I V I

6361 TAGTATATAATTTTCCAATGTTTCCAAGATACATATGACCTATACAACCTTCCATTTGACCCACAAAAGACAACCTTGCCTTATCTCTCTCTTATAGCAACAGATGAGA 6480
 V Y N F P I V S Q D T Y D L Y K L S I V P N K R Q L A L I P S S P Y I A T D E K

6481 AATCGTTCGTGATAGAGGCTGAATCCCGAAGTATAGCAGTCTTCTCTCGAAAAGAAAGACCGGACGAGATCCAGTCAAACTGATTGTATCAGAACTCATGCTCATC 6600
 S F V Y I E A E C P K Y S S T Y L C E K K T G Q Q I Q S K P D C I Q K L I V H Q

6601 AGAGTCTAGAGAATACTGTCAATTCACGAAGATATCTCTCATCAAGGAAGCAGTAGAAAATAGACGACCAACATTACGCTGTCTCTACCCGAACTACCAAGTTCAGTTGGCAT 6720
 S L E N T C Q F T K I S L I K E A V E K L D D Q H Y V L S L P E P T K V Q L A C

6721 GTGGGAAAGGACTTCAACACACTCAAGGAAGCTACCTCGTAACCATCCCTATGGGTTGATCTACAGACTCCAGAATTAACATAATAACAGTACAAACGGAATAAAGGCTCAAC 6840
 G R K D F N T L Q G S Y L V T I P M G C Y L Q T P E L T I I N D D N A I K G Q P

6841 CATTGAACTAGCAAGATACCATACGATGAAATGAATCTGACTCCGCTCTACACACATCAATTCAGCTCGATCGATCTGGAAGACTTACACAGCATCCAACCTAAATTCATGTGG 6960
 L K L A K I P Y D E N M L T A V S T H I N F S S I D L E D L H S I Q T K F M L G

6961 GAAAACCTATTGACATCGAGGAGATCAACCACTGCCCTGACACACAACCATCCCACTATACGTCATATTACTGGGCGCAATCCTATTTTCACTCTGAGATTAATCGCAATACA 7080
 K P I D I E E I Q P T A L Y H T T I P L Y V I L L G A I L F F T L R L I R K Y K

7081 AATGTTGGAGACTAAAATCAGAAGATAAAGAAAGCAGTCACTCTTGTAGATACATACATACGAAGACGTCAAGAAAATACCAGAAAACGAGATGACTTCCAGCAACATTTTCTCTTA 7200
 C W R L K S E D K E K Q S S L E I N T Y E D V K K N T R K R D D F P A T F S L M
 ~ ORF 3 PP ~ 3'LTR

7201 ATATGGTCAAAAATGTTGCTGATCTATGGGTTGAGGTTAGGTATGGAGCCTAATGAGTATCATCGACGCTGCATTTCTGTTATTGTCGCCGACGTCATAGAAACTGTCTGAATG 7320
 M V K N S C * ~ US | R ~

7321 ACGTAAATCGTCAAGCCGCTGATGACGAAATTTGTAATAGTATTAACCTCAAAATGATGATCTTCTAATATCGAGTAGGCTCACCGCAATTAATTTGTAATTTCTTA 7440
 PA ~ R | US ~ 3'LTR ~

7441 TAAATAAATATAGCATTAAATCATTTTGGTTTTTTCTATCTCGCGCTCGCAGTATACGTAATATG - 4 bp dup 7514

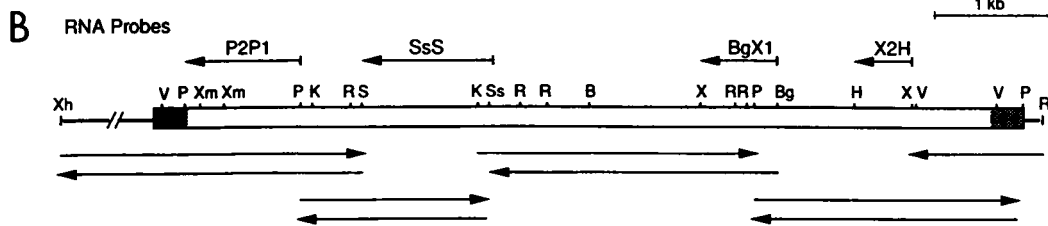


FIG. 1—Continued.

TED	268	IDGEALFTFNKIKHRIRTTDEVPVYKSYRYPFIHRQEVDRDQITKMLDQGIIRPSDSAUSPPIVVVPPK.IDASGKQKWRVVDVFRKLNEKTIDDKYPINISDVLDKLGKCY	380
17.6	185	hEGDKLFTFNqtKhtIntkhnLPLYSK.YsYPqayeQEVesQIQMLNQGttRTSNsPYnSPIVVVPPK.qDASGKQKFRIVIDYRKLNEiTVgDRHPiPNMDEILgKLGRCNY	297
Gypsy	159	ttnEALPFntaVtatIRTvDnePVYSRATptlmgvsdFvNEVekLkDGIIRPSrSPYnSptVVVKKgtdAFGnprnkRLVIDFRKLNEKTIpDRYPMPsIPmILanLgKakF	273
TED	381	FTTLDLASGFYQVEHDPQDISKTAfNVEHGHEFLRMPMGLKNSPSTFQRVMDNVLRLGNNICLVYLDIIIVYSTSLQEHLENLERVFORLRESNFKIQMDKSEFLKLETAYL	494
17.6	298	FTTIDLakGFhQIEMDPESVSKTAFstkhGHYEYLRMPfGLKNAPATFQRcMNDILRPLlnkhCLVYLDIIIVFSTSLDEHLQsLgLVFEKLakANIQLQDKCEFLKqETTFLL	411
Gypsy	274	FTTLDLkSGYhQIylAehDreKTSFsVNgGKYEFcRlPFLGRNASSiFQRALDDVLRReqigkICyVYVDDVIFSenesDhVrhIdtVlKclIDANmRVsqEKTffKesveYL	386
TED	495	GHIISRDGIKPNPDKISAIQKYLIPKTPKEIKQFLGLLGYRRKFIPIDFARLTKPLTQCLK.....KGSKVTLSPEYVNAFEHCKTLT.LNDPILQYPDFTRFNLTDD	596
17.6	412	GhVLTpDGIKPNPEKIEAIQKYPIPtKPKKIKaFLGLtGYRRKFIPIPNFAdIAKPMtKCLK.....knmkidTtnPEYdsAFkKlK.yLisEDPILkvPDFTKKfLTTD	514
Gypsy	387	GfIVSkDgtKSDPEKVAIQeYpePdcvykVRSFLGLaSYRvFIKDFaAIArPIITDILKgensvskhmsKkipVefnetqrNAFQRIRnILAsEdVlKvPDFKkFpDLTDD	500
TED	597	ASNFAIGAVLSQGPISGDKPVCYASRTLNESELNYSTIEKELLAIVWATKYFRPYLFG.RKFkILTdhKPLQMMNLKDPNSRMTRWRRLSEYDFSVVYKGGKSNMADALSR	710
17.6	515	ASDVALGAVLSQd....gHPLsYiSRTLNEHEINYSTIEKELLAIVWATktFRhYLIG.RHFeisSDHqPLsWlyrMKDPNSKLRWRVKLSEDFDikYiKGGeNcVADALSR	623
Gypsy	501	ASasGIGAVLSQe....gRPItmisRTLkQPEqNYATrERELLAIVWAlgkLqnlFYLGSReiniFDHqPLtFavAdRNTNAKIKRWKsyIdQhNakvYfKpGKeNfVADALSR	610

FIG. 2. Comparison of the putative amino acid sequence of the *pol* regions (ORF 2) of TED and *D. melanogaster* retrotransposons 17.6 and gypsy. Residues that are identical or chemically similar (44) among each of the three elements are indicated by double and single overlines, respectively. Individual residues chemically similar to those of TED are also capitalized. Numbers indicate the positions of residues from the start of ORF 2 for each element beginning with residue 268 (nucleotide position 2651) of TED. The amino acid sequences for 17.6 and gypsy were taken from references 41 and 28, respectively. Alignments were determined with the Bestfit program of the University of Wisconsin—Madison GCG sequence analysis package (14).

conditions described previously (34), and treated with S1 nuclease. S1 nuclease-resistant fragments and primer extension products were subjected to electrophoresis on 6% polyacrylamide–8 M urea–TBE (100 mM Tris borate [pH 8.3], 2 mM EDTA) gels followed by autoradiography.

RESULTS

Complete nucleotide sequence of element TED. The DNA sequence (7,510 bp) of TED as integrated within the genome of AcMNPV mutant FP-D is presented in Fig. 1A. Characteristic of a transposon integration event, four nucleotides of viral DNA (AATG) were duplicated at the site of insertion. This copy of TED originated from the *T. ni* (a lepidopteran noctuid) genome, where it represents a dispersed, middle repetitive sequence present in approximately 50 copies. Southern blot analysis of total genomic DNA indicated that only limited sequence polymorphism or partial copies of TED exist in cultured cells of this host (31). The restriction map (Fig. 1B) derived from the sequence determined here suggested that this particular copy of TED, found integrated within the AcMNPV genome, was representative of a majority of these elements within the host *T. ni* genome.

TED is flanked at both ends by LTRs, a feature characteristic of the copialike retrotransposons and retroviruses. Both LTRs (273 bp long) are identical in sequence (Fig. 1). The first (leftmost) and second (rightmost) LTRs were designated the 5' and 3' LTRs, respectively, on the basis of structural analogy with the retroviruses, the location of the major ORFs, and the predominant direction of transcription (left to right; see also below). Immediately adjacent to its 5' LTR, TED possesses a putative tRNA primer-binding site (tRNA PBS); the 18-bp sequence overlaps the 5' LTR by 1 nucleotide. In the case of the retroviruses, the first step in reverse transcription of genomic RNA is the initiation of first-strand DNA synthesis from a tRNA primer bound to the tRNA PBS (reviewed in references 48 and 49). Second-strand synthesis is then initiated from a polypurine tract located to the immediate left of the 3' LTR. The presumed polypurine tract for TED is 11 bp long (Fig. 1).

Translation of the DNA strand of TED shown in Fig. 1 revealed three long ORFs (ORFs 1, 2, and 3). Each ORF,

consisting of 457, 1,236, and 582 amino acids, respectively, overlaps the previous one. The longest ORF not shown (148 amino acids) is encoded by the opposite strand of DNA and is complementary to the N-terminal portion of ORF 1.

Predicted gene products of TED resemble those of other retrotransposons and the retroviruses. A computer search for sequence similarities revealed that the genetic and structural organizations of TED most closely resemble those of the *D. melanogaster* retrotransposons 17.6 (41), 297 (24), and gypsy (28), which in turn resemble the provirus form of avian leukosis virus (43), a vertebrate retrovirus. Flanked by LTRs, the interior portion of each element contains three overlapping ORFs with sizes and locations comparable to those of TED. The greatest sequence similarity is centered within ORF 2, corresponding to the *pol* domain of the *D. melanogaster* elements. Over a 440-amino-acid stretch, TED exhibits 64% amino acid identity (79% chemical similarity) and 45% amino acid identity (65% chemical similarity) with ORF 2 of elements 17.6 and gypsy, respectively (Fig. 2). Interestingly, the same stretch of TED ORF 2 is 39% identical (57% chemically similar) to the fungal retrotransposon Ty3 from *Saccharomyces cerevisiae* (20). In each case, these similarities extend over domains containing amino acid stretches that are highly conserved among the retroviruses (25, 29, 47) and encode reverse transcriptase and RNase H activities (Fig. 1). In addition, the N-terminal portion of TED ORF 2 possesses a hexapeptide, (hydrophobic residue)₂-D-T/S-G-A/S (Fig. 1), that is conserved among retroviruses and retrotransposons and thought to correspond to the active site of an aspartyl protease involved in the proteolytic processing of *gag-pol* polyproteins (1, 26, 47, 52). At the C-terminal region of TED ORF 2, several stretches of amino acids match highly conserved regions within the DNA endonuclease or integrase (Fig. 1) domain implicated in the retroviral integration process (15, 35). This region also possesses paired histidine (nucleotide positions 4517 and 4535) and cysteine (nucleotide positions 4625 and 4634) residues representing a potential site for coordinated binding of zinc ions, a feature found in many retroviral integrase domains (25). Thus, the inferred order of the conserved retroviruslike functional domains within TED ORF 2 (protease, reverse transcriptase, RNase H, and integrase) is

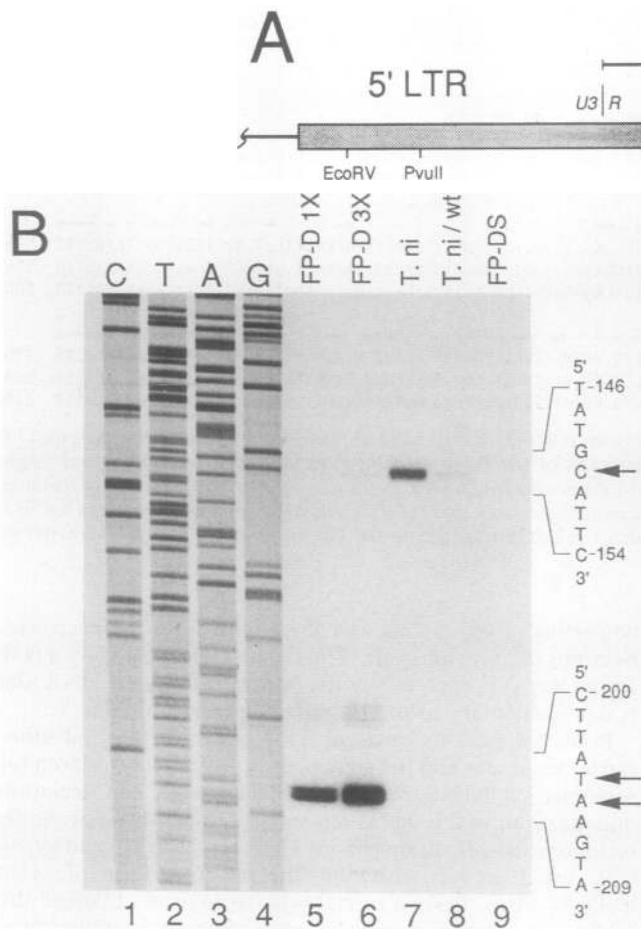


FIG. 3. Primer extension mapping of RNAs initiated from the 5' TED LTR. (A) A 5'-end-labeled primer (36 nucleotides [nt] long extending from position 303 within the body of TED to the *Sna*BI site at position 268 within the LTR) was annealed to poly(A)⁺ RNA and extended with reverse transcriptase; the asterisk denotes the position of the 5' end label of the primer. Shown are the relative locations (in nucleotides [nt] from the primer end) of the 5' ends of RNAs transcribed from the 5' LTR of TED within virus FP-D or host *T. ni* cells. Also shown are the U3-R and R-U5 junctions within the LTR (shaded box), the polyadenylation signal (AATAAA), the baculovirus late promoter consensus sequence (black box), and relevant restriction sites. (B) Extension products synthesized from RNA isolated from cultured *S. frugiperda* cells infected with AcMNPV mutants FP-D (2 μ g [lane 5] and 6 μ g [lane 6]) and FP-DS (15 μ g [lane 9]) and RNA isolated from *T. ni* cells either mock infected (15 μ g [lane 7]) or infected with wild-type (wt) AcMNPV (15 μ g [lane 8]) were subjected to electrophoresis along with a dideoxy sequencing ladder (lanes 1 to 4) generated with the same primer and single-stranded DNA from an M13mp19 clone of the 5' LTR. A threefold increase (3 \times) in RNA from FP-D-infected *S. frugiperda* cells was used to better visualize minor extension products (lane 6). Nucleotides at the initiation sites, numbered from the beginning of the LTR, are indicated on the right by arrows. We estimated that the resolution of these start sites was within 1 or 2 nucleotides.

identical to that within the *D. melanogaster* 17.6 and gypsy elements.

The sizes and locations of TED ORFs 1 and 3 are comparable to those of the retrovirus *gag* and *env* genes, respectively. Computer comparisons of ORFs 1 and 3, however, revealed no significant sequence similarities with other retrotransposons or vertebrate retroviruses. Nevertheless, the predicted translation product of TED ORF 1 resembles the *gag* products of Moloney murine leukemia virus (45) and the retrotransposon Ty1 (12) in exhibiting an unusually high content (10%) and uneven distribution of proline residues; for example, 29% of an 87-amino-acid stretch within ORF 1 is composed of proline residues. This property of the Ty1 *gag*-like proteins has been implicated in the formation of viruslike particles that contain both Ty RNA and reverse transcriptase activity (1). Like Ty1 and the *D. melanogaster* elements 17.6 and gypsy (12, 28, 41), TED ORF 1 lacks a putative nucleic acid-binding domain that is conserved among the retroviruses (13) and is also found in the *D. melanogaster* copia element (16, 33) and Ty3 (20). Suggestive that TED ORF 3 encodes an *env*-like membrane protein(s) is the presence of an uncharged, hydrophobic stretch of amino acids (nucleotide positions 7014 to 7055) within the predicted translation product at a C-terminal position analogous to the transmembrane portion of many retrovirus *env* proteins (48). Except for a similarly placed hydrophobic domain at the C terminus, it is noteworthy that ORF 3 of element 17.6 and ORF 3 of element gypsy are dissimilar despite residing within the same host, *D. melanogaster* (28, 41).

Identification of host- and virus-induced transcription initiation sites within the 5' TED LTR. The LTRs of the retrotransposons (and retroviruses) carry signals necessary for the promotion, initiation, and polyadenylation of RNA transcripts. To locate these signals within TED, we first mapped the major 5' and 3' ends of RNAs initiated and polyadenylated within the respective TED LTRs. The RNA start sites within the 5' LTR were determined by primer extension mapping (Fig. 3). The predominant extension product derived from poly(A)⁺ RNA isolated from uninfected *T. ni* cells (containing endogenous copies of TED) was a 154-nucleotide fragment (Fig. 3B, lane 7). Thus, when TED resided within the host *T. ni* genome, transcription was initiated from a single site (cytosine residue at nucleotide position 150) within the 5' TED LTR (Fig. 1). This initiation site represents the U3-R boundary that defines the end of the unique sequences at the 3' end (U3) and the beginning of the redundant (R) region of potential full-length RNAs (Fig. 1). Several A+T-rich regions that might function as TATA elements are present immediately upstream.

In contrast, the most prominent transcripts detected in *S. frugiperda* cells infected with virus FP-D (carrying a single copy of TED) were initiated further downstream within the LTR (positions 204 and 205), as demonstrated by the presence of 99- and 100-nucleotide extension products, respectively (Fig. 3B, lanes 5 and 6). Transcription from the upstream site (position 150) was greatly reduced relative to that from the downstream sites. Of particular interest was the observation that the abundant downstream RNAs were initiated 1 nucleotide apart within the sequence TTATAAG TAA (start sites indicated by underlining), which matches the conserved sequence comprising both the promoter and RNA start site for most, if not all, late AcMNPV genes (37, 39, 40). The absence of these RNAs in uninfected *T. ni* cells

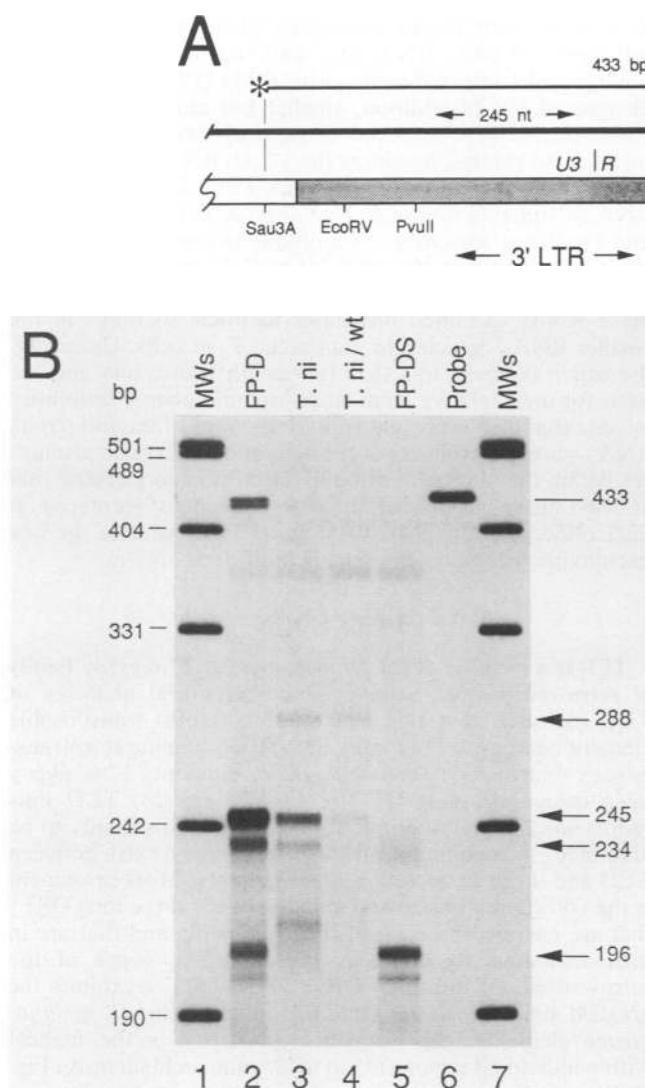


FIG. 4. Mapping the ends of poly(A)⁺ RNAs within the 3' LTR of TED. (A) A 433-bp DNA probe 3' end labeled exclusively at the *Sau3A* site (position 7222 within the internal portion of TED [asterisk]) and extending to the viral *EcoRI* site was annealed to poly(A)⁺ RNA and treated with S1 nuclease. The positions of the terminus of the major TED RNA (in nucleotides [nt] relative to the 3' end label) and the U3, R, and U5 domains of the 3' LTR (shaded box) are shown. (B) DNA fragments protected by RNA from *S. frugiperda* cells infected with virus FP-D (2.5 μ g [lane 2]) or FP-DS (2.5 μ g [lane 5]) and RNA from *T. ni* cells either mock infected (10 μ g [lane 3]) or infected with wild-type (wt) AcMNPV (10 μ g [lane 4]) were subjected to electrophoresis and autoradiography. Sizes in base pairs of DNA markers (lanes 1 and 7, MWs) are shown on the left, while prominent S1 nuclease-protected fragments and the 433-bp DNA probe alone (lane 6) are indicated on the right by the arrows.

(Fig. 3B, lane 7) suggested that transcription from the baculoviruslike LTR promoter was a consequence of the incorporation of TED into the genome of virus FP-D, possibly the result of *trans* activation by transcription factors specific for late viral promoters. To test whether viral infection was also capable of inducing transcription from the same site when TED resided within the host genome, we infected *T. ni* cells with wild-type AcMNPV and isolated RNA 24 h later. Primer extension revealed no evidence of transcription from the baculoviruslike promoter; only reduced levels from the upstream host promoter were observed (Fig. 3B, lane 8).

The lack of extension products synthesized in response to RNA from cells infected with virus FP-DS, carrying a solo TED LTR (Fig. 3B, lane 9), confirmed that the above-described prominent transcripts, initiated from LTR positions 204 and 205, were derived solely from the intact 5' LTR and not the solo LTR of virus FP-DS. Virus FP-DS, representing an element excision derivative, is generated spontaneously in the FP-D virus stocks used here (31).

Location of RNA termini within the 3' TED LTR. S1 nuclease protection analysis with a 3'-end-labeled DNA probe (Fig. 4A) was used to map the ends of transcripts

polyadenylated within the intact 3' LTR. RNA from FP-D-infected *S. frugiperda* cells (Fig. 4B, lane 2) protected three major fragments of 245, 234, and 196 bases, with the 245-base fragment predominating. The 245- and 234-base fragments were also predominant species protected by RNA from uninfected *T. ni* cells (Fig. 4B, lane 3). Thus, the major 3' ends of TED-specific RNAs mapped to nucleotide positions 7456 and 7466 within the 3' LTR (Fig. 1), corresponding to a pair of CA dinucleotides located 5 and 15 bases downstream from the single polyadenylation signal (AATAAA, positions 7446 through 7451) within the 3' TED LTR. Since the downstream (second) polyadenylation site (position 7466) was utilized most frequently by TED RNAs transcribed in both *T. ni* cells and FP-D-infected *S. frugiperda* cells, we have assigned the R-U5 boundary to this position. Thus, for potential full-length RNAs, this position marks the end of the redundant (R) region and the beginning of the U5 unique region (Fig. 1).

The less prominent 288-base fragment protected by *T. ni* RNA from both uninfected and wild-type AcMNPV-infected cells (Fig. 4B, lanes 3 and 4) corresponded to the distance from the labeled terminus of the probe to the end of the 3' LTR (288 bp). This result suggested that a significant number of RNAs extended through the 3' LTR without polyadenylation. A substantial number of transcripts that escaped polyadenylation were also detected in FP-D-infected cells, as demonstrated by the presence of a protected fragment identical in size (433 bp) to the full-length probe (Fig. 4B, lane 2). The solo TED LTR of virus FP-DS also exhibited termination leakiness in allowing viral transcripts initiated from outside the LTR to extend uninterrupted through the LTR (19). The lack of protection of the 234- and 245-base fragments by RNA from FP-DS-infected cells (Fig. 4B, lane 5) confirmed that these fragments were specific for termination within the intact 3' LTR and not the solo LTR. The 196-base fragment protected by both FP-D (Fig. 4B, lane 2) and FP-DS (Fig. 4B, lane 5) RNAs represented nonspecific protection of a portion of the probe, possibly by host *S. frugiperda* RNA.

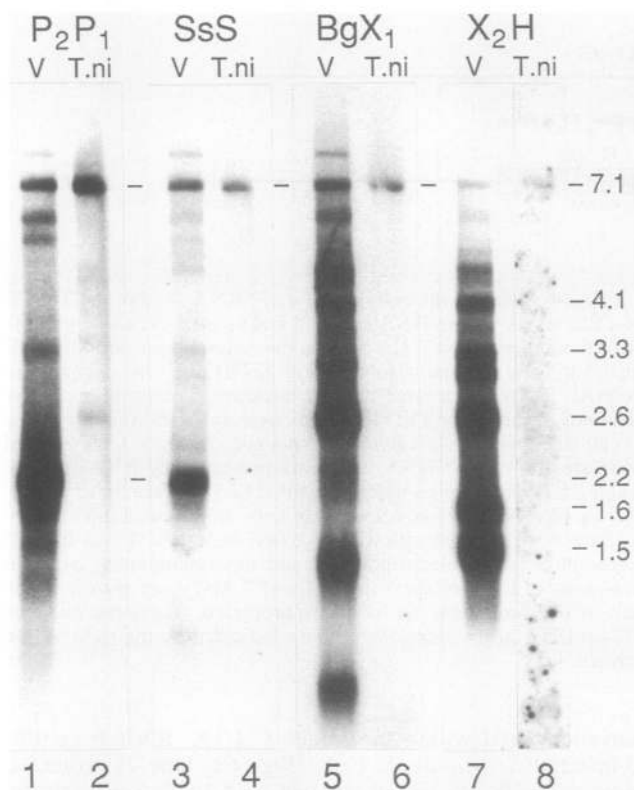


FIG. 5. Strand-specific Northern analysis of TED RNAs. Poly(A)⁺ RNAs isolated from *S. frugiperda* cells 24 h after infection with virus (V) FP-D (2.5 μ g per lane; lanes 1, 3, 5, and 7) and from uninfected *T. ni* cells (10 μ g per lane; lanes 2, 4, 6, and 8) were fractionated by agarose gel electrophoresis, blotted, and hybridized to the indicated strand-specific probes (P₂P₁, SsS, BgX₁, and X₂H). The location of each probe within the internal portion of TED is indicated in Fig. 1B. Longer exposures of autoradiograms for probes BgX₁ and X₂H (lanes 5 to 8) were used to better visualize the full-length RNAs. The sizes in kilobases of various RNAs are indicated on the right.

Transcription of full-length RNA copies of TED. The observation that RNA transcripts were both initiated and terminated within the 5' and 3' TED LTRs, respectively, suggested that the transcription of TED included the synthesis of a full-length (genomic) RNA extending from LTR to LTR. Such an RNA would be expected to play a role in the expression of TED-encoded gene products and in transposition. To examine transcription within the internal portions of TED, we hybridized Northern blots of poly(A)⁺ RNA to strand-specific RNA probes generated from four different regions of TED (Fig. 5). These probes (Fig. 1B) were designed to hybridize to transcripts extending from left to right (from the 5' LTR to the 3' LTR), the predominant direction of transcription, as determined in other hybridization experiments (data not shown).

The predominant TED-specific RNA identified in uninfected *T. ni* cells carrying endogenous copies of TED was approximately 7.1 kb in length (Fig. 5B, all lanes designated *T. ni*). That this RNA was sufficiently long (within the error of the empirical method used to determine RNA size) and hybridized to each of the four probes (P₂P₁, SsS, BgX₁, and X₂H) suggested that this transcript represented a full-length (genomic) RNA extending from the 5' LTR to the 3' LTR. Other smaller but less abundant RNAs were also detected in

T. ni cells with longer exposures (data not shown). The full-length (7.1-kb) RNA was also transcribed in *S. frugiperda* cells infected with virus FP-D (Fig. 5B, all lanes designated V). In addition, smaller but more predominant RNA species were observed. Several of these were detected by adjacent probes, including the 2.2-kb RNA hybridizing to probes P₂P₁ and SsS (Fig. 5B, lanes 1 and 3) and the 2.6-kb RNA hybridizing to probes BgX₁ and X₂H (Fig. 5B, lanes 5 and 7); others appeared to be unique to the region complementary to the specific probe (i.e., 1.5- and 1.6-kb RNAs hybridizing to probe X₂H [Fig. 5B, lane 7]). Few, if any, of these RNAs exhibited mobilities identical to those of the smaller RNAs detected in uninfected *T. ni* cells. Currently, the origin of these less-than-full-length transcripts and the basis for their relative abundance are unknown; possibilities include that they represent spliced versions of the full-length RNA or are the products of the activation of cryptic promoters within the internal portion of TED. It is noteworthy that at least three groups of internal sequences (centered at nucleotide positions 990, 5045, and 5770) resemble the late baculovirus promoter consensus motif (see above).

DISCUSSION

TED is a member of the *D. melanogaster* 17.6-gypsy family of retrotransposons. Sequence and structural analyses of TED indicated that this lepidopteran (moth) transposable element belongs to the family of LTR-containing retrotransposons that include *D. melanogaster* elements 17.6, gypsy (also designated mdg-4), 297, and Beagle (5). TED thus represents the first nondipteran member of this family to be identified. A number of striking similarities exist between TED and these *D. melanogaster* elements. Most prominent is the correlation in size and location of the three long ORFs that are encoded by each of these elements and that are in turn analogous to the *gag*, *pol*, and *env* genes of the retroviruses. Of the three ORFs, TED ORF 2 exhibits the greatest similarity to the same region within the *D. melanogaster* elements; this sequence similarity is the highest within individual regions (45 to 65% amino acid identity; Fig. 2) that also bear a high degree of resemblance to the domains within the retrovirus *pol* gene that encode protease, reverse transcriptase, RNase H, and integrase functions. As in the *D. melanogaster* elements, a frameshift in the -1 direction would be required for the expression of a putative TED *gag-pol* (ORF 1-ORF 2) fusion polypeptide.

As with gypsy expression in *D. melanogaster* (36), the most abundant TED RNA in host *T. ni* cells was a poly(A)⁺ transcript extending from the 5' LTR to the 3' LTR (Fig. 5). Mapping the ends of this RNA (Fig. 3 and 4) indicated that it possessed redundant termini (80 bases at each end), as initiation within the 5' LTR occurred upstream from the termination (polyadenylation) site within the corresponding 3' LTR. Thus, like the LTRs of gypsy (2), the LTRs of TED possess a retroviral U3-R-U5 organization (Fig. 1) that is capable of directing the synthesis of transcripts that correspond in structure to a full-length RNA copy (extending from the 5' LTR to the 3' LTR) of the integrated element (Fig. 6). Such transcripts represent potential intermediates for reverse transcriptase-mediated transposition (8). In turn, reverse transcription is expected to initiate from the putative tRNA PBS (Fig. 1) which is located adjacent to the 5' TED LTR and which is identical to that of 297 and closely related to that of 17.6 (24, 41); the 297 tRNA PBS is complementary to the 3' end of a *D. melanogaster* tRNA^{Ser}. Unlike their retroviral counterparts, the primer-binding sites of the *D.*

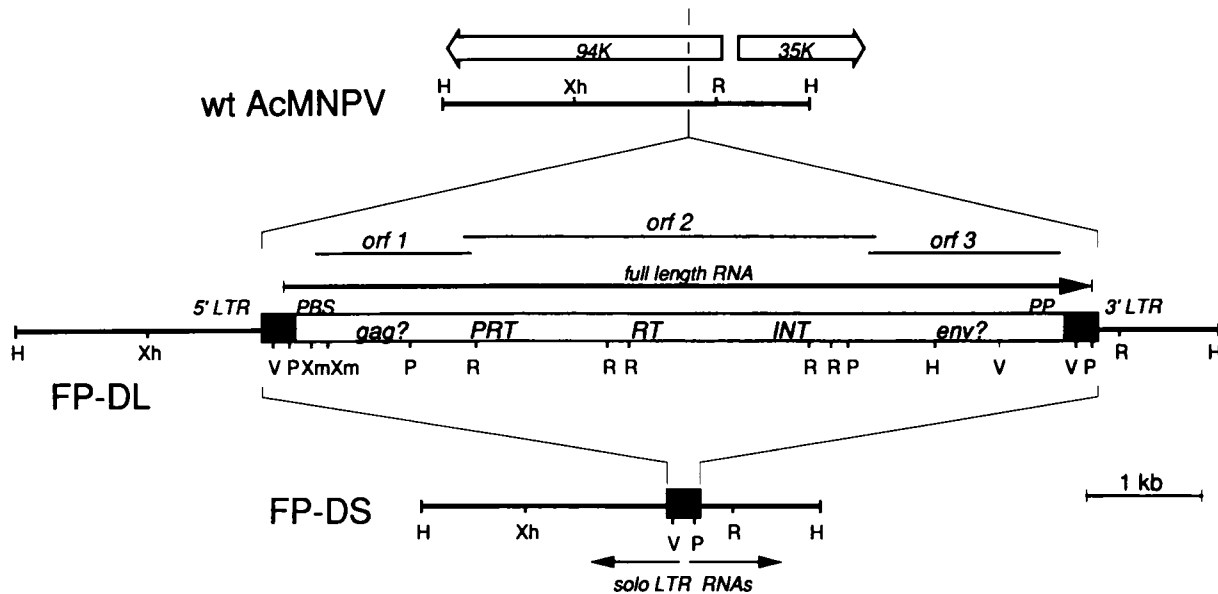


FIG. 6. Gene organization of retrotransposon TED as integrated within the AcMNPV genome. (Top) The site of TED insertion (86.7 map units) is within the N-terminal region of a viral gene encoding a 94,000-molecular-weight (94K) protein located on the *Hind*III K genome fragment of wild-type (wt) AcMNPV. (Middle) Virus FP-DL carries the intact TED element (7,510 bp) possessing identical 273-bp LTRs (darkly shaded regions). The putative tRNA PBS (PBS), polypurine tract (PP), and three ORFs (orf 1, orf 2, and orf 3) are located within the body of TED. The regions corresponding to the protease (prt), reverse transcriptase (rt), and integrase (int) functional domains conserved among the retrotransposons and retroviruses are indicated. The putative *gag* and *env* domains are also shown. The full-length TED RNA (long arrow) extends from the 5' LTR to the 3' LTR. (Bottom) Virus FP-DS, generated from virus FP-DL, carries a single (solo) LTR at the site of integration that directs the transcription of divergent RNAs from the baculoviruslike promoter into flanking viral sequences (19). Restriction site abbreviations are identical to those in Fig. 1.

melanogaster 17.6-gypsy retrotransposons, including that of TED, overlap the 5' LTR by 1 nucleotide; it has therefore been postulated that priming is accomplished by a tRNA lacking its 3'-terminal nucleotide (5). Finally, TED integration results in the duplication of 4 bp of target DNA, a characteristic feature of members of this family.

Despite their residence in invertebrates that are considered relatively distant phylogenetically (orders, Lepidoptera and Diptera, respectively), the marked similarities between TED and other members of the *D. melanogaster* 17.6-gypsy family support the contention that these elements were derived from a common ancestral transposon. The sequence similarity between the *pol* region of the fungal retrotransposon Ty3 from *S. cerevisiae* and those of 17.6 and TED further suggests that such a hypothetical ancestor existed prior to the divergence of the fungi from other eucaryotes (20) and continued to evolve as the Lepidoptera and Diptera diverged. On the other hand, available data cannot rule out the intriguing possibility that the observed similarity between these elements is due to a more recent horizontal transfer between different organisms (see also below) of an ancestral element that has diverged into the present-day elements.

Possible effects of the activation of baculoviruslike promoters on the transposition of TED. Integration of TED within the AcMNPV genome resulted in the activation of a promoter that directed the synthesis of new and abundant RNAs from the 5' LTR. Since these RNAs were initiated from a sequence that is identical to the consensus motif that constitutes the promoter and RNA start site for late AcMNPV genes (37, 39, 40), it is likely that transcription from this site is mediated by the viral transcriptional machinery specific for late, hyperexpressed AcMNPV genes. The same site

directs abundant transcription from the solo TED LTR of virus FP-DS during infection of both *S. frugiperda* and *T. ni* cells (19; J. Fleming and P. Friesen, unpublished results) and may account for the numerous RNAs transcribed from internal portions of TED within virus FP-D (Fig. 5). In contrast, transcripts initiated from the same viruslike promoter of TED LTRs residing within the *T. ni* genome were not detected even after infection with wild-type AcMNPV (Fig. 3). This result suggested that when embedded within the host genome, the baculoviruslike promoter of the LTR was unresponsive to the *trans*-acting factors that mediate late viral transcription. Thus, at least during the late period of infection, AcMNPV does not appear to stimulate host TED transcription. Earlier effects have not been ruled out.

Transcription from the baculoviruslike promoter within the 5' LTR of TED in virus FP-D produced RNAs that were nearly identical to the full-length RNAs of *T. ni* cells, differing only in the size of the redundant (R) region at each end (25 bases versus 80 bases for the full-length RNA). Assuming that this decrease in redundancy has no effect on nascent-strand transfer (reviewed in reference 48), the virus-induced transcripts could also serve as intermediates in transposition. We estimate that late in infection (24 h), *S. frugiperda* cells infected with virus FP-D (possessing a single copy of TED) contain levels of full-length or nearly full-length RNA that are 5 to 10 times higher than those in uninfected *T. ni* cells (possessing approximately 50 resident copies of TED). As is the case with transposition of the yeast Ty elements (8, 20), this increased synthesis of full-length TED RNAs may act to increase the frequency of TED transposition. Thus, a possible effect of the integration of TED within the viral genome and the resulting activation of the baculoviruslike promoter may be to increase the mobility

of this retrotransposon. Increased mobility late in infection during the period of active viral DNA replication could facilitate the spread of TED through the virus population, a feature expected of "selfish" DNA.

Our observations here, coupled with numerous examples of the insertion of host-derived DNA and/or transposable elements into the AcMNPV genome (4, 9, 10, 31), have also raised the interesting possibility that baculoviruses are capable of shuttling transposons (and therefore host DNA) between insect hosts. Such a transfer strategy (31) could serve to extend transposon host range. Successful shuttling by a virus would depend upon several factors, including the transpositional competence of the mobile element after viral integration and an ensuing nonlethal or abortive infection of the recipient host by the transposon-carrying virus. Studies intended to test the capacity of TED to transpose from the AcMNPV genome are currently under way.

Transposon-mediated evolution of the baculovirus genome. The frequent insertion of host-derived transposons at various sites within the AcMNPV genome also suggests that transposition is a general mechanism for the spontaneous introduction of foreign DNA into the baculovirus genome. If such mutagenic events provide a selective growth advantage, such as that provided by mutations resulting in the baculovirus FP (few-polyhedron) phenotype (reviewed in reference 32), then transposition may play a significant role in virus evolution.

Our studies here indicated that the integration of TED generated numerous mutational changes within the AcMNPV genome. Most conspicuous was the addition of several new retroviruslike genes (*gag*, *pol*, and *env*) to the viral genome (Fig. 6). Synthesis of poly(A)⁺ RNAs from the viral copy of TED suggested that at least some of the products of these genes are synthesized; indeed, FP-D-infected *S. frugiperda* cells contained significantly higher levels of reverse transcriptase activity than did wild-type virus-infected cells (R. Lerch and P. Friesen, unpublished results). Another mutagenic effect was the potential inactivation of the early 94K protein gene (located on the AcMNPV *Hind*III K genome fragment), since TED integration disrupted the coding region of this gene (Fig. 6). Finally, the incorporation into the viral genome of viruslike promoter sequences located within the TED LTRs resulted in the transcription of new and abundant RNAs having the potential to affect nearby viral genes. Most dramatic in this respect was the effect of the solo TED LTR within virus FP-DS (Fig. 6). Representing an element excision derivative, the solo LTR directs abundant transcription of divergent RNAs that initiate from the viruslike promoter and extend into flanking viral genes (19). Studies of the role of TED integration in altering AcMNPV gene expression are in progress.

ACKNOWLEDGMENTS

This work was supported in part by Public Health Service grant AI25557 from the National Institute of Allergy and Infectious Diseases and by institutional grant IN-119H from the American Cancer Society.

LITERATURE CITED

- Adams, S. E., J. Mellor, K. Gull, R. B. Sim, M. F. Tuite, S. M. Kingsman, and A. J. Kingsman. 1987. The function and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. *Cell* **49**:111-119.
- Arkhipova, I. R., A. M. Mazo, V. A. Cherkasova, T. V. Gorelova, N. G. Schuppe, and Y. V. Ilyin. 1986. The steps of reverse transcription of *Drosophila* mobile dispersed genetic elements and U3-R-U5 structure of their LTRs. *Cell* **44**:555-563.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 1988. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
- Beames, B., and M. D. Summers. 1988. Comparisons of host cell DNA insertions and altered transcription at the site of insertions in few polyhedra baculovirus mutants. *Virology* **162**:206-220.
- Bingham, P. M., and Z. Zuchar. 1989. Retrotransposons and the FB transposon from *Drosophila melanogaster*, p. 485-502. In D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
- Boeke, J. D. 1989. Transposable elements in *Saccharomyces cerevisiae*, p. 335-374. In D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
- Boeke, J. D., and V. G. Corces. 1989. Transcription and reverse transcription of retrotransposons. *Annu. Rev. Microbiol.* **43**:403-434.
- Boeke, J. D., D. J. Garfinkel, C. A. Styles, and G. R. Fink. 1985. Ty elements transpose through an RNA intermediate. *Cell* **40**:491-500.
- Carstens, E. B. 1987. Identification and nucleotide sequence of the regions of *Autographa californica* nuclear polyhedrosis virus genome carrying insertion elements derived from *Spodoptera frugiperda*. *Virology* **161**:8-17.
- Cary, L., M. Goebel, B. G. Corsaro, H. Wang, E. Rosen, and M. J. Fraser. 1989. Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **172**:156-169.
- Chirgwin, J. M., A. E. Przybyla, R. J. McDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**:5294-5299.
- Clare, J., and P. Farabaugh. 1985. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proc. Natl. Acad. Sci. USA* **82**:2829-2833.
- Copeland, T. D., S. Oroszlan, V. S. Kalyanraman, M. G. Sarnagadharan, and R. C. Gallo. 1983. Complete amino acid sequence of human T-cell leukemia virus structural protein p15. *FEBS Lett.* **162**:390-395.
- Devereux, J., P. Haeblerli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:378-395.
- Donehower, L. A., and H. E. Varmus. 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. *Proc. Natl. Acad. Sci. USA* **81**:6461-6465.
- Emori, Y., T. Shiba, S. Kanaya, S. Inouye, S. Yuki, and K. Saigo. 1985. The nucleotide sequences of copia and copia-related RNA in *Drosophila* virus-like particles. *Nature (London)* **315**:773-776.
- Faulkner, P., and E. B. Carstens. 1986. An overview of the structure and replication of baculoviruses. *Curr. Top. Microbiol. Immunol.* **131**:1-19.
- Friesen, P. D., and L. K. Miller. 1987. Divergent transcription of early 35- and 94-kilodalton protein genes encoded by the *Hind*III K genome fragment of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* **61**:2264-2272.
- Friesen, P. D., W. C. Rice, D. W. Miller, and L. K. Miller. 1986. Bidirectional transcription from a solo long terminal repeat of the retrotransposon TED: symmetrical RNA start sites. *Mol. Cell. Biol.* **6**:1599-1607.
- Hansen, L. J., D. L. Chalker, and S. B. Sandmeyer. 1988. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Mol. Cell. Biol.* **8**:5245-5256.
- Heinkoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**:351-359.
- Hink, W. F. 1970. Established insect cell line from the cabbage looper, *Trichoplusia ni*. *Nature (London)* **225**:466-467.
- Hruby, D. E., and W. K. Roberts. 1976. Encephalomyocarditis virus RNA: variations in polyadenylic acid content and biological activity. *J. Virol.* **19**:325-330.

24. Inouye, S., S. Yuki, and K. Saijo. 1986. Complete nucleotide sequence and genome organization of a *Drosophila* transposable genetic element, 297. *Eur. J. Biochem.* **154**:417-425.
25. Johnson, M. S., M. A. McClure, D. F. Feng, J. Gray, and R. F. Doolittle. 1986. Computer analysis of retrovirus *pol* genes: assignment of enzymatic function to specific sequences and homologies with nonviral enzymes. *Proc. Natl. Acad. Sci. USA* **83**:7648-7652.
26. Katoh, I., T. Yasunaga, Y. Ikawa, and Y. Yoshinaka. 1987. Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor. *Nature (London)* **329**:654-656.
27. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. Marlbor, R. L., S. M. Parkhurst, and V. G. Corces. 1986. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Mol. Cell. Biol.* **6**:1129-1134.
29. McClure, M. A., M. S. Johnson, D.-F. Feng, and R. F. Doolittle. 1988. Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. *Proc. Natl. Acad. Sci. USA* **85**:2469-2473.
30. McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc. Natl. Acad. Sci. USA* **74**:4835-4838.
31. Miller, D. W., and L. K. Miller. 1982. A virus mutant with an insertion of a copia-like transposable element. *Nature (London)* **299**:562-564.
32. Miller, L. K. 1986. The genetics of baculoviruses, p. 217-238. *In* R. R. Granados and B. A. Federici (ed.), *The molecular biology of the baculoviruses*. CRC Press, Inc., Boca Raton, Fla.
33. Mount, S. M., and G. M. Rubin. 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol. Cell. Biol.* **5**:1630-1638.
34. Nissen, M. S., and P. D. Friesen. 1989. Molecular analysis of the transcriptional regulatory region of an early baculovirus gene. *J. Virol.* **63**:493-503.
35. Panganiban, A. T., and H. M. Temin. 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of retrovirus *int* locus. *Proc. Natl. Acad. Sci. USA* **81**:7885-7889.
36. Parkhurst, S. M., and V. G. Corces. 1985. *forked*, gypsies and suppressors in *Drosophila*. *Cell* **41**:429-437.
37. Possee, R. D., and S. C. Howard. 1987. Analysis of the polyhedrin gene promoter of the *Autographa californica* nuclear polyhedrosis virus. *Nucleic Acids Res.* **15**:10233-10248.
38. Potter, K. N., and L. K. Miller. 1980. Correlating genetic mutations of a baculovirus with the physical map of the DNA genome, p. 71-80. *In* B. N. Fields, R. Jarnisch, and C. F. Cox (ed.), *Animal virus genetics*. Academic Press, Inc., New York.
39. Rankin, C., B. G. Ooi, and L. K. Miller. 1988. Eight base pairs at the transcriptional start point are the major determinant for baculovirus polyhedrin gene expression. *Gene* **70**:39-49.
40. Rohrmann, G. F. 1986. Polyhedrin structure. *J. Gen. Virol.* **67**:1499-1513.
41. Saigo, K., W. Kugimiya, Y. Mutasuo, S. Inouye, K. Yoshioka, and S. Yuki. 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. *Nature (London)* **312**:659-661.
42. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5465.
43. Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. *Cell* **32**:853-869.
44. Schwarz, R. M., and M. O. Dayhoff. 1978. Matrices for detecting distant relationships, p. 353-358. *In* M. O. Dayhoff (ed.), *Atlas for protein sequence and structure*, vol. 5. National Biomedical Research Foundation, Washington, D.C.
45. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukaemia virus. *Nature (London)* **293**:543-548.
46. Thomas, P. S. 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Natl. Acad. Sci. USA* **77**:5201-5205.
47. Toh, H., R. Kibuno, M. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo. 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and the *pol* gene product of Moloney murine leukaemia virus. *EMBO J.* **4**:1267-1272.
48. Varmus, H., and P. Brown. 1989. Retroviruses, p. 53-108. *In* D. E. Berg and M. M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.
49. Varmus, H. E. 1983. Retroviruses, p. 411-503. *In* J. A. Shapiro (ed.), *Mobile genetic elements*. Academic Press, Inc., New York.
50. Vaughn, J. L., R. H. Goodwin, G. L. Thompkins, and P. McCawley. 1977. Establishment of two insect cell lines from the insect *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *In Vitro (Rockville)* **13**:213-217.
51. Weaver, R. F., and C. Weissman. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S β -globin mRNA precursor and mature 10S β -globin mRNA have identical map coordinates. *Nucleic Acids Res.* **7**:1175-1193.
52. Youngren, S. D., J. D. Boeke, N. J. Sanders, and D. J. Garfinkel. 1988. Functional organization of the retrotransposon Ty from *Saccharomyces cerevisiae*: Ty protease is required for transposition. *Mol. Cell. Biol.* **8**:1421-1431.