MOLECULAR AND CELLULAR BIOLOGY, Feb. 1990, p. 615–624
0270-7306/90/020615-10$02.00/0

Vol. 10, No. 2

A Rapidly Rearranging Retrotransposon within the Miniexon Gene Locus of *Crithidia fasciculata*

ABRAM GABRIEL,[1,2]* TIM J. YEN,[1] DAVID C. SCHWARTZ,[3] CYNTHIA L. SMITH,[3] JEF D. BOEKE,[2] BARBARA SOLLNER-WEBB,[1] AND DON W. CLEVELAND[1]

Departments of Biological Chemistry[1] and Molecular Biology and Genetics,[2] The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, and Department of Embryology, Carnegie Institution of Washington, Baltimore, Maryland 21210[3]

Received 17 August 1989/Accepted 25 October 1989

The tandemly arrayed miniexon genes of the trypanosomatid *Crithidia fasciculata* are interrupted at specific sites by multiple copies of an inserted element. The element, termed *Crithidia* retrotransposable element 1 (CRE1), is flanked by 29-base-pair target site duplications and contains a long 3′-terminal poly(dA) stretch. A single 1,140-codon reading frame is similar in sequence to the integrase and reverse transcriptase regions of retroviral *pol* polyproteins. Cloned lines derived from a stock of *C. fasciculata* have unique arrangements of CRE1s. In different cloned lines, CRE1s, in association with miniexon genes, are located on multiple chromosomes. By examining the arrangement of CRE1s in subclones, we estimate that the element rearranges at a rate of ca. 1% per generation. These results indicate that the *C. fasciculata* miniexon locus is the target for a novel retrotransposon.

Over the past decade the *Trypanosomatidae* family of parasitic protozoa has provided important insights into the molecular biology of eucaryotic organisms. The discovery of sequence-specific bending of DNA in the mitochondrial minicircles of kinetoplastids (25), the realization that mRNA maturation in trypanosomatids involves *trans* splicing of discontinuously transcribed precursor molecules (29, 40), and, most recently, the detection of large-scale RNA editing of mitochondrial maxicircle transcripts (13, 38) have expanded current concepts of DNA structure and gene expression.

We have been studying transcription in the insect parasite *Crithidia fasciculata*, a species whose ease of cultivation, minimal nutritional requirements, and nonpathogenicity make it an ideal model trypanosomatid. As in *Trypanosoma brucei*, most (and probably all) translatable mRNAs in *C. fasciculata* possess an identical 39-nucleotide leader sequence at their 5′ termini (9, 15, 45). This sequence has been termed the miniexon (4) or spliced leader (31) sequence; the genes encoding this RNA are unlinked to protein coding genes. The miniexon sequence is also found at the 5′ terminus of an abundant, short, nonpolyadenylated RNA, termed the miniexon donor RNA (6). In *C. fasciculata*, miniexon donor RNA is approximately 90 nucleotides long and is transcribed from a family of multiple-copy, tandemly arrayed miniexon genes which have a unit length of 423 base pairs (bp) and a copy number of 200 to 500 per genome (15, 27).

During our investigation of the *C. fasciculata* miniexon gene locus, we detected multiple copies of a 3.5-kilobase (kb) insertion element which interrupt the tandem array at a specific site within the unit repeat. A number of other insertion elements have been identified in trypanosomes. RIME is a 511-bp repetitive element originally found interrupting a single copy of a rRNA gene in *T. brucei*; its ca. 200 copies are widely dispersed throughout the genome (17, 18). Ingi/TRS (22, 28) is a 5.2-kb dispersed repetitive element in *T. brucei*, flanked by the two halves of RIME. By sequence comparison, ingi/TRS is most closely related to mammalian LINE elements. It is not known to be associated with the miniexon gene cluster. SLACS/MAE (1, 7) is a multiple-copy 5.5- to 7.0-kb insert localized to the miniexon locus of *T. brucei* with features of a retroposon, i.e., target site duplications and a terminal poly(dA) tract. In this paper we report the complete 3,940-nucleotide sequence of a *Crithidia* insertion element and provide evidence that it retrotransposed into the miniexon gene locus. Furthermore, we demonstrate that the element actively rearranges in *C. fasciculata* at an estimated frequency of 1% per generation.

MATERIALS AND METHODS

***Crithidia* culturing.** The original stock of *C. fasciculata* was obtained from Paul Englund, John Hopkins University School of Medicine, and grown at 27°C as previously described (15). Clones were prepared by plating serial dilutions of the stock on nutrient agar plates containing 1% Bacto-Agar (Difco Laboratories), 37 g of brain heart infusion per liter, 1% penicillin-streptomycin, and 20 μg of hemin per ml and inoculating individual colonies 3 to 7 days later into fresh liquid medium, or by diluting log-phase *Crithidia* stocks to less than 1 organism per ml and dispensing 100-μl aliquots into microdilution wells. Wells were examined microscopically after 2 days, and those containing parasites were inoculated into fresh liquid medium.

Preparation of nucleic acids. *Crithidia* DNA was extracted from log-phase cultures by one of two methods: by phenol-chloroform extraction as described by Monteiro and Cox (26) (for Fig. 1, 4, and 5) or by the following modification of a standard yeast DNA extraction procedure (10) (for Fig. 7). Cultures (1 to 2 ml) were pelleted and suspended in 0.4 ml of 10 mM Tris (pH 8)–1 mM EDTA (TE). A solution of EDTA, Tris base, and sodium dodecyl sulfate was added to final concentrations of 56 mM, 88 mM, and 0.44%, respectively. After incubation at 65°C, a solution of unbuffered 5 M potassium acetate was added to a final concentration of 0.83 M, and the mixture was incubated on ice. After centrifugation, the supernatant was collected and precipitated with 95% ethanol at room temperature. The pellet was dried,

* Corresponding author.

suspended in 0.5 ml of TE containing 40 μg of heat-treated RNase A per ml, and incubated at 37°C. After centrifugation, the supernatant was mixed with an equal volume of isopropyl alcohol and the pellet was washed with 70% ethanol. The resulting pellet was dried and suspended in 50 μl of TE for Southern analysis.

Copy number was determined as previously described (15) and analyzed densitometrically.

Construction of plasmids. p4kb2.2 and additional independent copies of *Crithidia* retrotransposable element no. 1 (CRE1) were cloned from size-selected *Hin*dIII-digested *Crithidia* DNA ligated to *Hin*dIII-digested pUC18 and selected by hybridization to p400 probe, as previously described (15). pAUG-CRE was constructed as follows. The upstream ATG codon was created by polymerase chain reaction amplification of a p4kb2.2 template by using the oligonucleotide primers AG-13 (5′-GAGCGGCCGCCAT GACGGCATTCGGTCTAGTG-3′ [plus strand of the 4-kb insert from nucleotides 417 to 434]) and AG-14 (5′-GCCAGG CGTCGACAGGAATG-3′ [minus strand of the 4-kb insert for nucleotides 1016 to 1035]). AG-13 contains a *Not*I site and an ATG codon, followed by the first six codons found in CRE1. AG-14 overlaps the single *Sal*I site in CRE1. After 30 cycles of polymerase chain reaction amplification, the 632-bp product was purified, digested with *Not*I and *Sal*I, and ligated to the vector pBluescript KS+ digested with *Not*I and *Hin*dIII and a 3-kb *Sal*I-*Hin*dIII CRE1 fragment isolated from p4kb2.2. Once pAUG-CRE was obtained, the internal deletion plasmids pAUG-CREdelEcoRV and pAUG-CREdelBamHI were constructed by digestion of pAUG-CRE with *Eco*RV or *Bam*HI, respectively, followed by self-ligation of the digested plasmid in a dilute solution.

Hybridization and wash conditions. Nylon filters (GeneScreen Plus; Du pont Co.) were hybridized with ^{32}P-labeled probes at 42°C in 50% formamide–1 M NaCl–1% sodium dodecyl sulfate–10% dextran sulfate and then washed at 65°C in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate. The 3-kb and p400 probes (see Fig. 1B) were prepared from gel-purified fragments by the method of Feinberg and Vogelstein (14).

In vitro transcription and translation. In vitro transcription of pBluescript-derived plasmids was performed by using a kit from Stratagene Inc. Noncapping conditions resulted in fivefold-higher levels of translation (data not shown). In vitro translations were performed with treated rabbit reticulocyte lysate (Promega Biotec) and Tran^{35}S-label (ICN Pharmaceuticals Inc.).

Sequencing strategy. Subclones of p4kb2.2 were constructed in pUC18 by using the *Hin*dIII-*Sal*I, *Sal*I-*Pst*I, and *Pst*I-*Hin*dIII fragments. Vector was cleaved on either side of the insert, and deletion series of each subclone were constructed by timed exonuclease III digestion followed by trimming with S1 nuclease and secondary cutting on the opposite side of the insert. These deletions were then subcloned into M13mp18 and M13mp19 and sequenced by the dideoxy-sequencing method of Sanger et al. (34). The areas around the *Sal*I and *Pst*I sites were sequenced after preparation of pUC and M13 subclones by using restriction sites on either side of these restriction sites to obtain overlapping sequence information. All parts of the p4kb2.2 insert were sequenced completely, in both strands, at least twice.

For sequencing the 5′ insertion site and the first 80 codons from multiple clones, a *Sma*I-*Sma*I fragment (from nucleotides 150 to 628) from each clone was gel purified, subcloned into M13mp18 in both orientations, and then sequenced by using the oligonucleotides AG-3 (minus strand of p4kb2.2 from nucleotides 615 to 629) and AG4 (plus strand of p4kb2.2 from nucleotides 351 to 371) as primers. For sequencing the 3′ insertion sites, a 500-bp *Stu*I-*Hin*dIII fragment (from nucleotides 3440 to 3940) from each clone was subcloned into M13mp18 and then sequenced by using the M13 universal primer.

Pulsed-field gel electrophoresis. We prepared 0.5% agarose inserts containing log-phase *Crithidia* cells at a final concentration of 10^9/ml as described previously (35, 43). Electrophoresis was performed by a modification of the pulsed-field electrophoresis method of Schwartz et al. (36). Restriction digestion of *Crithidia* DNA in agarose inserts was performed as described previously (D. C. Schwartz, Ph.D. thesis, Columbia University, New York, N.Y., 1985).

Computer analysis. All sequence analyses were performed on a Digital Equipment Corp. Vax 8530 computer by using algorithms developed by Lipman and Pearson (24).

RESULTS

Cloning of a 4-kb element with homology to the miniexon gene. When *Crithidia* genomic DNA was digested to completion with restriction enzymes that cleave once within each miniexon gene and these digests were blotted and probed with a labeled copy of the miniexon gene (p400), the expected 423-bp fragment was observed (Fig. 1A, arrowhead). However, several additional bands were seen whose sizes differed from one DNA preparation to the next. The least variable of these was a 4-kb restriction fragment always present in genomic DNA digested with *Hin*dIII (Fig. 1A, arrow). The intensity of the 4-kb band suggested that it was present in multiple copies, a finding which would be unusual if it represented an "orphon," i.e., a gene copy dispersed from the tandem array (8, 30). This multiple-copy restriction fragment might encode a second class of miniexon containing repeats or might be a repeated sequence either flanking or interrupting the tandem array.

To determine its identity directly, we cloned the 4-kb fragment. The restriction map of one of the positive clones, termed p4kb2.2 (Fig. 1B), showed unique *Sal*I, *Eco*RI, and *Pst*I sites located 1,000, 1,175, and 2,900 nucleotides, respectively, from the left-hand *Hin*dIII site. The 1-kb *Hin*dIII-*Sal*I fragment hybridized strongly to the miniexon gene probe p400. The remaining 3-kb *Sal*I-*Hin*dIII fragment did not hybridize to p400 under the stringent conditions used. We therefore used the latter fragment as a specific probe for the 4-kb repeat (3-kb probe in Fig. 1B); we estimated the copy number of the 4-kb repeats to be approximately 10 per genome (data not shown).

The 4-kb element has features of a site-specific retroposon. To ascertain the degree of similarity between the cloned 4-kb fragment and the miniexon gene, we sequenced the p4kb2.2 insert. The salient structural features are depicted schematically in Fig. 1B; the sequence is given in Fig. 2. The left end of p4kb2.2 consisted of a nearly exact copy of the previously sequenced miniexon gene repeat, including the entire 39-bp miniexon sequence (Fig. 1B, boxes A and B; Fig. 2, nucleotides 377 to 415). However, just downstream of the 3′ end of the miniexon sequence, i.e., at the 5′ splice site, sequence identity diverged into an open reading frame (ORF) of 3,420 nucleotides (Fig. 2, nucleotides 417 to 3836) followed by a string of 27 adenosines (Fig. 2, nucleotides 3872 to 3898). The right-hand end of the clone consisted of 42 bases that were identical to sequences in the miniexon gene repeat; i.e., a duplication of the 3′ 29 nucleotides of miniexon sequence

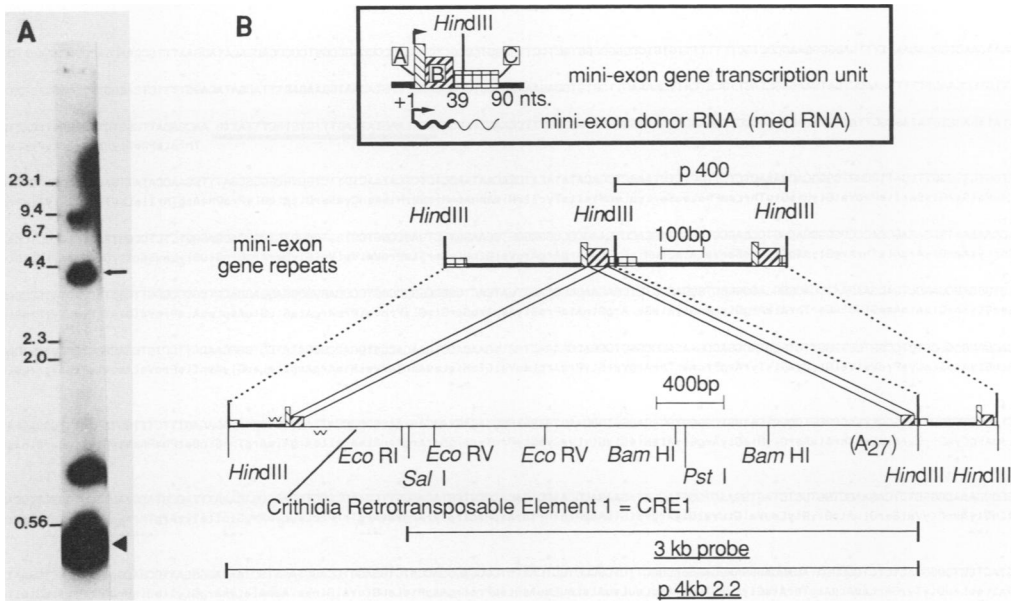

FIG. 1. Identification and structure of a 4-kb fragment which hybridizes to a miniexon gene probe. (A) Autoradiogram of total *Crithidia* DNA digested with *Hin*dIII, fractionated on 1% agarose, transferred to a nylon filter, and hybridized with a ^{32}P-labeled copy of the 423-bp miniexon gene (p400). In addition to the miniexon genes themselves (arrowhead), a band at 4 kb (arrow) is always observed. Other bands (e.g., at 8 kb) are found only in some DNA preparations. λ DNA digested with *Hin*dIII served as size markers. (B) Schematic of the structure of the cloned 4-kb fragment, designated p4kb2.2. As depicted, the clone consists of a copy of the miniexon gene interrupted by an inserted 3.5-kb element. Restriction sites used for sequencing the fragment are shown. p400 is a cloned copy of the miniexon gene used as a probe. The 3-kb probe is a *Sal*I-*Hin*dIII fragment used as an insertion-element-specific probe. Boxes A, B, and C denote the miniexon transcription unit. Box B is duplicated on either end of the insertion element. Wavy lines on the lower portion of the figure refer to oligonucleotides AG-3 and AG-4, used to sequence the insertion sites and the first 80 codons from multiple independent clones of p4kb.

(Fig. 1B, box B; Fig. 2, nucleotides 3899 to 3927), followed by the 13-nucleotide region downstream of the 5′ splice site (Fig. 1B, box C; Fig. 2, nucleotides 3928 to 3940).

The sequence shows that the 4-kb repeat consists of an insertion of a 3.5-kb element into the transcriptionally active region of an intact 423-bp miniexon gene repeat. The insertion event generated a 29-bp terminal duplication of the 3′ portion of the miniexon sequence. The fact that the sequence appears to be a DNA copy of a spliced and polyadenylated transcript strongly implies that p4kb2.2 represents an integrated reverse transcript. Such elements have been termed retroposons (33).

To examine the sequence specificity at the insertion site, we sequenced the 5′ and 3′ insertion sites of five additional cloned 4kb elements. Whereas three clones were identical to the sequence shown in Fig. 2 at their 5′ end, two other clones contained a single-base substitution (A to G) 1 base downstream of the 29-bp target site duplication (Fig. 2, nucleotide 416). Since G is also the nucleotide found in the miniexon gene at this position, these two clones appear to have 30-bp target site duplications and are concomitantly missing the first nucleotide of the inserted element. At the 3′ end the insertion site was identical in all clones. Surprisingly, however, the length of the poly(dA) tract just upstream of the insertion site varied from 16 to 57 bases, and in one clone, the three T residues within the poly(A) tract were replaced by eight T residues.

The insert contains a long ORF similar to retroviral *pol* polyproteins. The inferred amino acid sequence of the 1,140-amino-acid ORF is shown in Fig. 2. Remarkably, this ORF begins 1 nucleotide downstream of the miniexon splice junction and ends 20 nucleotides upstream of the 3′ poly(dA) tract. Because of its proximity and orientation, it could potentially be transcribed in continuity with the upstream miniexon gene.

The greatest similarities to the amino acid sequence of the *Crithidia* ORF found in the National Biomedical Research Foundation protein data base were to the *pol* polyproteins from equine infectious anemia virus and adult T-cell leukemia virus. These homologies extended on either side of the highly conserved tetrapeptide sequence (Tyr/Phe)-X-Asp-Asp (Y/F-X-D-D) (Fig. 2) from nucleotides 2382 to 2393. The *Crithidia* ORF contains 9 of the 10 invariant amino acid residues identified by Toh et al. in the polymerase and putative polymerase gene products from five different viruses (42).

To estimate the significance of these similarities, we compared a region of approximately 300 amino acids encoding the putative reverse transcriptase (RT) domain in the *Crithidia* ORF with the corresponding regions in the equine infectious anemia virus and adult T-cell leukemia virus *pol* protein sequences. Within this region the sequence identity is 17%; the *Crithidia* sequence was greater than 10 standard deviations more similar to each retroviral sequence than to 50 randomly permuted versions of either retroviral sequence, indicating significant evolutionary relationship between the *Crithidia* ORF and retroviral *pol* genes.

Since retroviral RT is only one domain within a larger polyprotein, which also encodes a viral protease (PR), integrase (IN), and RNase H, the *Crithidia* ORF was searched for additional domains. Using the consensus sequences derived by Doolittle and co-workers (11, 21), we identified a region containing two potential nucleic acid-binding domains beginning at nucleotide 921. Downstream of these potential nucleic acid-binding domains were five other amino acids invariant in retroviral INs. Although elements of

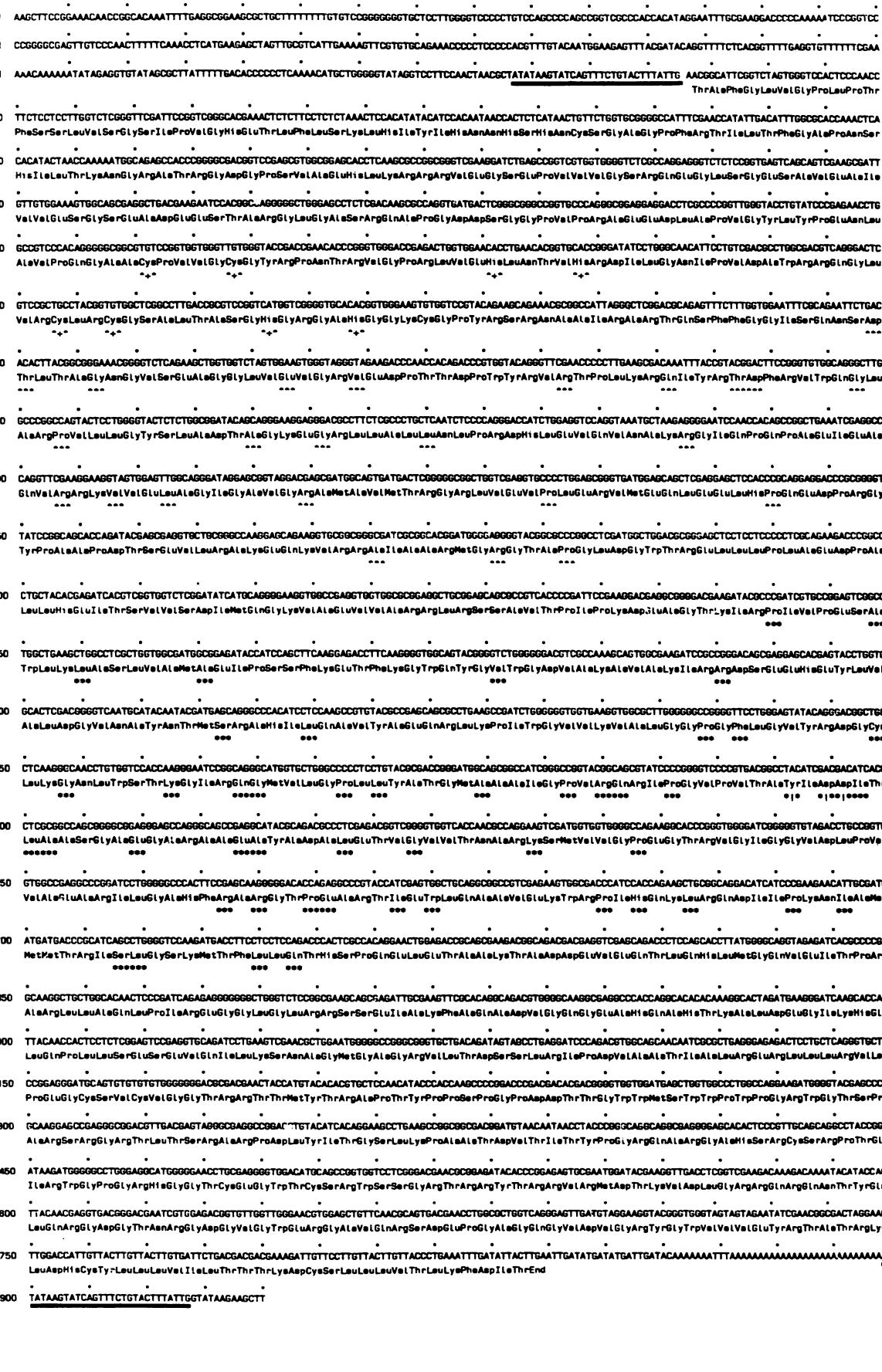

FIG. 2. Nucleotide and deduced amino acid sequence of p4kb2.2. The sequence of the entire 3.94-kb *Hin*dIII fragment is shown, along with the deduced 1,140-amino-acid ORF. The underlined 29-nucleotide regions are duplicated at either end of the ORF. The sequence from nucleotides 1 to 413 is identical to the previously determined miniexon gene sequence (15), except for the absence of a G residue at nucleotide 157 in p4kb2.2. In addition, a substitution of a G for an A is observed in some clones, 1 base downstream of the 5'-terminal duplication (nucleotide 416). Symbols: ^+^, consensus amino acids in potential metal-binding domains; ^^^, amino acid identities between the *Crithidia* ORF and the IN region of Moloney murine leukemia virus (39) after allowing for gaps; ***, amino acid identities between the *Crithidia* ORF and the putative RT of adult T-cell leukemia virus (37); *!*, sequence YXDD, which is the most highly conserved region in RT-like entities. The sequencing strategy is described in Materials and Methods.

retroviral PR and RNase H were also observed in the *Crithidia* ORF, the sequences were too divergent for unambiguous conclusions. The amino acid identities between the *Crithidia* ORF and both the Moloney murine leukemia virus IN and the adult T-cell leukemia virus RT domains are shown in Fig. 2. Independent analysis of the *Crithidia* ORF, using different alignment protocols and other protein data bases, confirmed the sequence similarity to retroviral RT and IN domains (R. F. Doolittle, personal communication).

It is noteworthy that the first AUG codon in the CRE1 sequence is not found until codon 380 (nucleotide 1554), although the IN domain begins at amino acid 169. Although it is possible that p4kb2.2 represents a pseudogene copy that has lost is initiation codon, the sequence of the first 80 codons in five additional independently cloned insertion elements are identical to the sequence in Fig. 2.

Thus, the *Crithidia* insertion element consists of a long ORF, which potentially encodes a polyprotein with domains similar to retroviral IN and RT. Since the structural and sequence features are most reminiscent of the class of elements designated non-long-terminal-repeat (LTR) retrotransposons or poly(A) retrotransposons (see Discussion) (3, 48), we have named the 3.5-kb insert *Crithidia* retrotransposable element 1 (CRE1).

The CRE1 ORF encodes a 140-kDa polypeptide. Since many non-LTR retrotransposon insertions represent pseudogene copies in which nonsense mutations have accumulated by neutral drift, we wanted to directly confirm the coding capacity of CRE1. We therefore constructed the plasmid pAUG-CRE, in which an AUG codon was created just upstream of the first codon of the CRE1 ORF and the entire potential coding region was placed downstream of a T7 promoter. In addition, we constructed two in-frame deletions of pAUG-CRE. After in vitro transcription and translation, a polypeptide of ca. 140 kilodaltons (kDa) is synthesized, as expected for a 1,141-amino-acid ORF beginning at the engineered start codon and ending at the CRE1 termination codon (Fig. 3A, lanes 2 and 6). In addition to the 140-kDa band, a strong band is present at 90 kDa which was most probably due to an internal initiation event, as demonstrated by translation of the in-frame deletion constructs (lanes 3 and 4). This result is consistent with the conclusion that the 90-kDa band is derived from the first endogenous AUG codon in CRE1. In lane 5, either a truncated transcript without a termination codon is poorly translated or the resulting protein is unstable.

So far, we have been unable to detect RT activity in either rabbit reticulocyte or wheat germ lysates after in vitro translation of either pAUG-CRE or the internal deletion constructs. However, it appears that the lysates themselves are inhibitory to RTs, since activity from either avian myeloblastosis virus RT or purified Ty virus-like particles is undetectable in the lysates (data not shown).

The genomic organization of CRE1 is surprisingly complex.

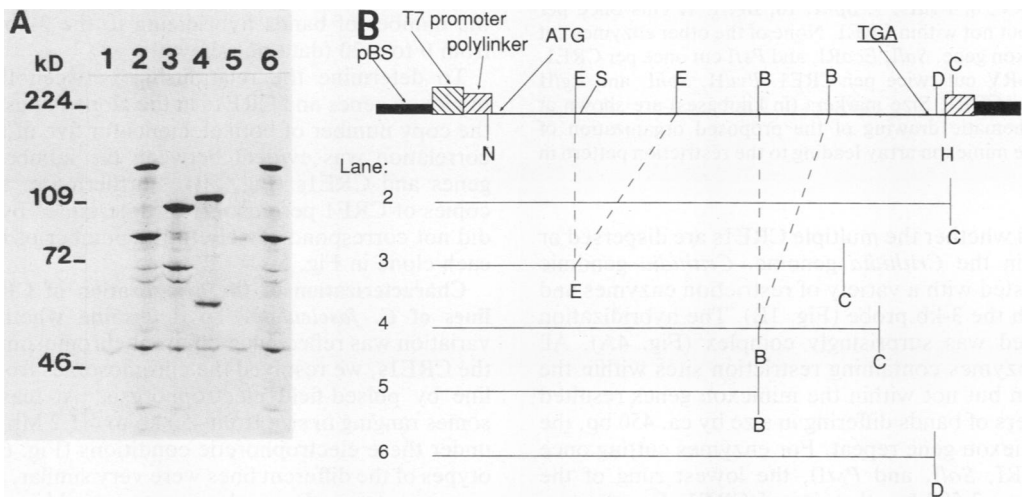

FIG. 3. In vitro transcription and translation of the CRE1 ORF. (A) Fluorograph of the in vitro translation products synthesized after in vitro transcription of the templates shown in Fig. 3B and resolved by electrophoresis through 7% polyacrylamide. The gel was fixed, treated with 1 M sodium salicylate, and dried before exposure. Lanes: 1, no RNA added; 2, pAUG-CRE cleaved at the *Cla*I site within the polylinker downstream of CRE1 sequences; 3, pAUG-CREdelEcoRV cleaved at the same site as in lane 2; 4, pAUG-CREdelBamHI, cleaved at the same site as in lane 2; 5, pAUG-CRE, cleaved at the *Bam*HI site within the CRE-coding region; 6, pAUG-CRE, cleaved at the *Dra*I site within the CRE1 3' noncoding region. (B) Important structural features of pAUG-CRE and structure of the templates used for in vitro transcription and translation. The thin line represents the sequence of CRE1 from nucleotides 417 to 3940 (Fig. 2) to which an upstream ATG codon has been added. The shorter hatched area is the pBluescript KS polylinker, and the taller hatched area is the bacteriophage T7 promoter. The bold line represents the vector pBluescript KS. Restriction sites: N, *Not*I; E, *Eco*RV; B, *Bam*HI; D, *Dra*I; C, *Cla*I.

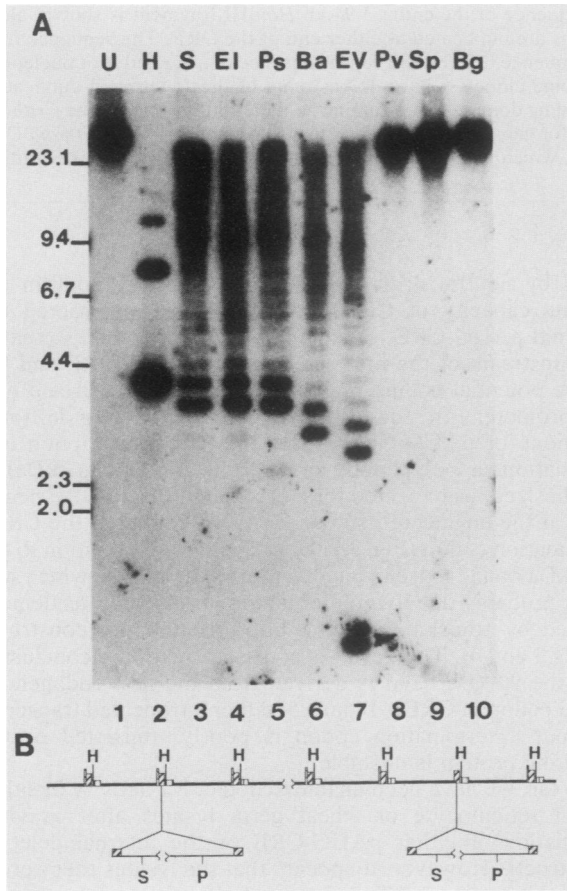

FIG. 4. A complex restriction enzyme pattern of *Crithidia* DNA hybridized to the 3-kb probe. (A) Autoradiogram of *Crithidia* DNA digested with a variety of restriction enzymes, fractionated through agarose, transferred to a nylon filter, and hybridized with the 3-kb probe. Lanes: 1, uncut; 2, *Hin*dIII; 3, *Sal*I; 4, *Eco*RI; 5, *Pst*I; 6, *Bam*HI; 7, *Eco*RV; 8, *Pvu*II; 9, *Sph*I; 10, *Bgl*II. H cuts once per miniexon gene, but not within CRE1. None of the other enzymes cut within the miniexon gene. *Sal*I, *Eco*RI, and *Pst*I cut once per CRE1. *Bam*HI and *Eco*RV cut twice per CRE1. *Pvu*II, *Sph*I, and *Bgl*II do not cut within CRE1. Size markers (in kilobases) are shown at the left. (B) Schematic drawing of the proposed organization of CRE1s within the miniexon array leading to the restriction pattern in panel A.

We determined whether the multiple CRE1s are dispersed or clustered within the *Crithidia* genome. *Crithidia* genomic DNA was digested with a variety of restriction enzymes and hybridized with the 3-kb probe (Fig. 1B). The hybridization pattern obtained was surprisingly complex (Fig. 4A). All digests with enzymes containing restriction sites within the retrotransposon but not within the miniexon genes resulted in regular ladders of bands differing in size by ca. 450 bp, the size of the miniexon gene repeat. For enzymes cutting once in CRE1 (*Eco*RI, *Sal*I, and *Pst*I), the lowest rung of the ladder was at ca. 3,500 bp, the size of CRE1. For the two enzymes which cut CRE1 twice (*Bam*HI and *Eco*RV), the size of the lowest rung was equal to 3,500 bp minus the length of the internal restriction fragment. For enzymes lacking sites within CRE1 or the miniexon gene (*Pvu*II, *Sph*I, and *Bgl*II), hybridization was to fragments of >23 kb.

The simplest interpretation of these data is that multiple copies of CRE1, present in the same orientation, are interspersed among the tandemly repeated miniexon genes (Fig 4B). Together, these form a superarray of >23 kb. However, any enzyme which cuts once per CRE1 (regardless of the position of the site within the element) will create multiple bands made up of the two halves of adjacent CRE1s plus a variable number of intervening miniexon genes. The fact that the smallest fragment in Fig. 4, lanes 2, 3, and 4, is 3.5 kb rather than 4 kb suggests that some CRE1 copies exist in tandemly duplicated arrays uninterrupted by miniexon genes, a prediction consistent with the 7.5-kb (dimer) and 11-kb (trimer) bands seen in the *Hin*dIII digest (Fig. 4, lane 2).

The CRE1 ladder results from heterogeneity within the population. Although the above explanation of the arrangement of CRE1 with regard to miniexon genes is consistent with the hybridization pattern shown in Fig. 4, there are more rungs on the ladder than expected, given our estimate of 10 copies of CRE1 per genome. Therefore, we reasoned that the ladder could represent the sum of multiple simpler patterns present within subpopulations of the culture. This hypothesis gained plausibility when we considered that the stock of *C. fasciculata* we had used had been cloned from a single cell in 1976 (12) and maintained by serial passage every 2 to 3 days since that time.

We took advantage of the ability of *C. fasciculata* to grow from single cells either in liquid media or on nutrient agar plates. We plated serial 10-fold dilutions from a log-phase culture. After visible colonies appeared, we picked random colonies from the highest-dilution plates and inoculated them into liquid media. DNA extracted from each of six freshly cloned cultures, as well as from the uncloned culture, was digested with *Eco*RI, fractionated on an agarose gel, transferred to nylon filters, and hybridized with the 3-kb probe. The resulting autoradiogram (Fig. 5) confirms our hypothesis dramatically. The originally cloned population of *C. fasciculata* has undergone a remarkable degree of rearrangement, such that randomly selected cells within the current population have unique genomic arrangements of CRE1. Examination of 54 additional colonies from the same stock culture revealed that no two clones had the same pattern and that the number of bands hybridizing to the 3-kb probe ranged from 0 to >20 (data not shown).

To determine the relationship between the number of miniexon genes and CRE1s in the cloned lines, we measured the copy number of both elements for five of the clones. No correlation was evident between the number of miniexon genes and CRE1s (Fig. 5B). Furthermore, the number of copies of CRE1 per genome as determined by densitometry did not correspond closely to the number of bands seen for each clone in Fig. 5A.

Characterization of the organization of CRE1s in cloned lines of *C. fasciculata*. To determine whether the clonal variation was reflected in different chromosomal locations of the CRE1s, we resolved the chromosomes from each cloned line by pulsed-field electrophoresis. At least 12 chromosomes ranging in size from 450 kb to >1.2 Mb were resolved under these electrophoretic conditions (Fig. 6A). The karyotypes of the different lines were very similar, although some apparent size polymorphisms were visible on the ethidium bromide-stained gel. The increased intensity of the chromosomes at the lower half of the gel suggests either that there are overlapping chromosomes of similar size in that region or, possibly, that the copy number of some of the chromosomes in this region is selectively amplified.

After transfer, we hybridized the filter with labeled 3-kb probe. The resulting autoradiogram showed that most copies of the retrotransposon are located on one or a few chromo-

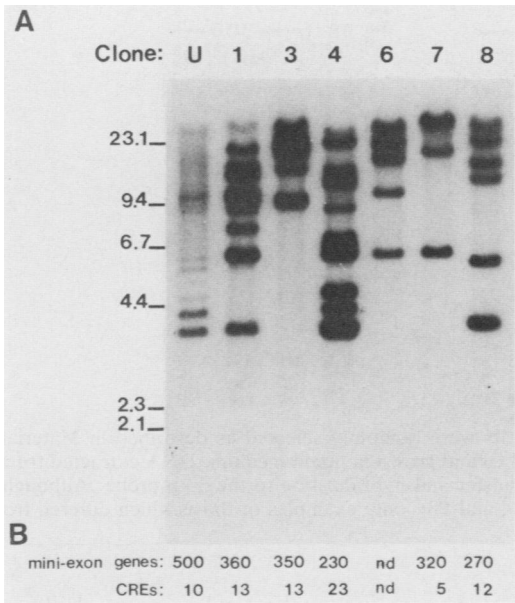

FIG. 5. Restriction enzyme pattern of DNA from cloned lines of *Crithidia* hybridized to the 3-kb probe. (A) Individual cells from the stock population were cloned. DNA was extracted from uncloned *C. fasciculata* (lane U) or six cloned lines (lanes 1, 3, 4, 6, 7, and 8), digested with *Eco*RI, fractionated on agarose, transferred to a nylon filter, and hybridized with the 3-kb probe. Size markers (in kilobases) are shown at the left. (B) Copy number determination of both miniexon genes and CRE1s per cell for uncloned *C. fasciculata* and five of the cloned lines. Abbreviation: nd, not determined.

somes, ranging from 600 to 900 kb (Fig. 6B). Whereas in clones 1, 3, 6, and 7 the retrotransposons are on at least one variably sized chromosome, in clones 4 and 8 they are found on three different-sized chromosomes. Thus, CRE1 is unlike most other retrotransposons in that it is not dispersed throughout the genome.

To determine whether CRE1s are always found on the same chromosomes as miniexon genes, we stripped the filter and rehybridized it with the p400 probe (Fig. 6C). A comparison of these two autoradiograms demonstrated that all of the chromosomes that contained CRE1s also contained miniexon genes. This was true even for the smallest hybridizing chromosome in clone 8 and the largest chromosome in clone 4, although that became apparent only after longer exposure times (data not shown). Conversely, clones 1 and 3 both had chromosomes which contained miniexon genes but no CRE1s.

Although Fig. 6A to C demonstrated that CRE1s reside on the same chromosomes as miniexon genes do, we sought to establish whether they were always physically adjacent within the DNA of each chromosome. Therefore, we digested whole chromosomes from each cloned line with the restriction enzyme *Pvu*II, for which there are no sites in either the miniexon genes or CRE1 (Fig. 4, lane Pv). After digestion, the products were resolved by pulsed-field gel electrophoresis. The ethidium bromide-stained pattern of the resulting gel is shown in Fig. 6D. Again, the digestion pattern for different clones was similar, except for specific bands >100 kb.

We again sequentially hybridized the filter with the 3-kb and then the p400 probe. The resulting patterns (Fig. 6E and F, respectively) complemented those obtained with whole chromosomes. The CRE1s were clustered together with

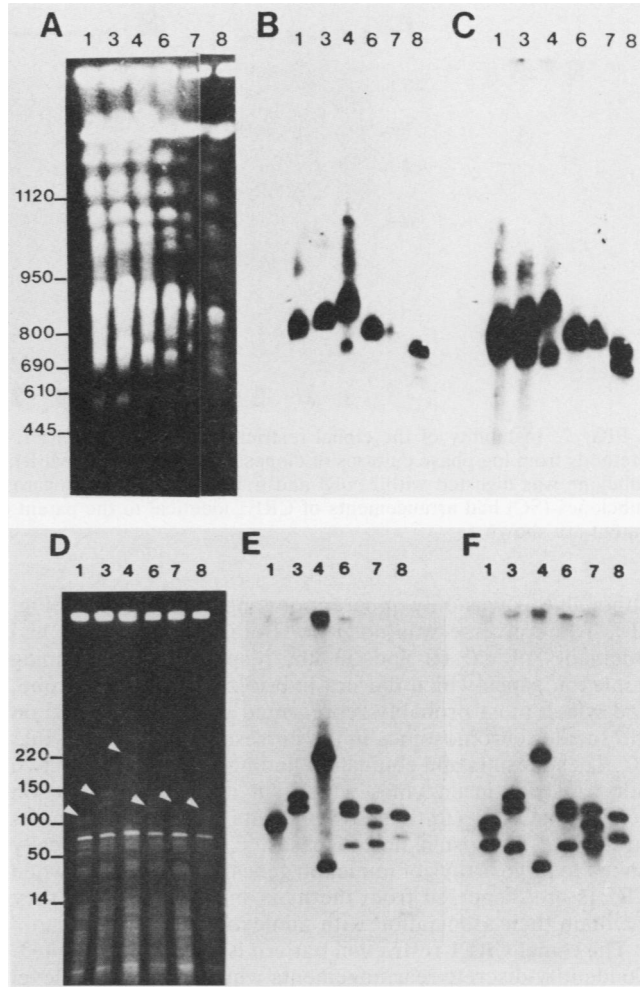

FIG. 6. CRE1s and miniexon genes are located on variable-sized chromosomes and chromosomal fragments in different cloned lines. (A) Ethidium bromide-stained pulsed-field gel of DNA from six cloned lines of *C. fasciculata*. Inserts of *C. fasciculata* in agarose were prepared from late-log-phase cultures of each cloned line at a final concentration of 10^9 cells per ml. Electrophoresis was through 1% HGT (FMC Corp.) agarose at 8 V/cm for 92 h, with stepped pulse intervals of 300, 200, 100, and 40 s at a ratio of 1:2:2:1. Size markers are based on *S. cerevisiae* chromosomes electrophoresed in an adjacent lane. (B) After transfer of the gel, the filter was hybridized with the 3-kb probe. (C) After the 3-kb probe was stripped from the nylon filter, the filter was rehybridized with p400. (D) Ethidium bromide-stained pulsed-field gel of the six *C. fasciculata* clones, each digested with *Pvu*II prior to electrophoresis through 1% agarose at 8 V/cm for 24 h at a 120-s pulse followed by 36 h at a 15-s pulse. Size markers used were a 1-kb DNA ladder (Bethesda Research Laboratories, Inc.), 50-kb λ concatemers, and *S. cerevisiae* chromosomes. Arrows mark large restriction fragments containing miniexon arrays. (E) After transfer of the gel, the filter was hybridized with the 3-kb probe. (F) After the 3-kb probe was stripped from the nylon filter, the filter was rehybridized with p400.

miniexon genes on multiple 50- to 250-kb restriction fragments. By comparison of these autoradiograms with the ethidium bromide staining pattern, it was apparent that the large polymorphic fragments seen by ethidium bromide staining corresponded to miniexon arrays (Fig. 6D, arrowheads). All fragments containing CRE1s also had associated miniexon genes (including the 10-kb fragment in clone 8,

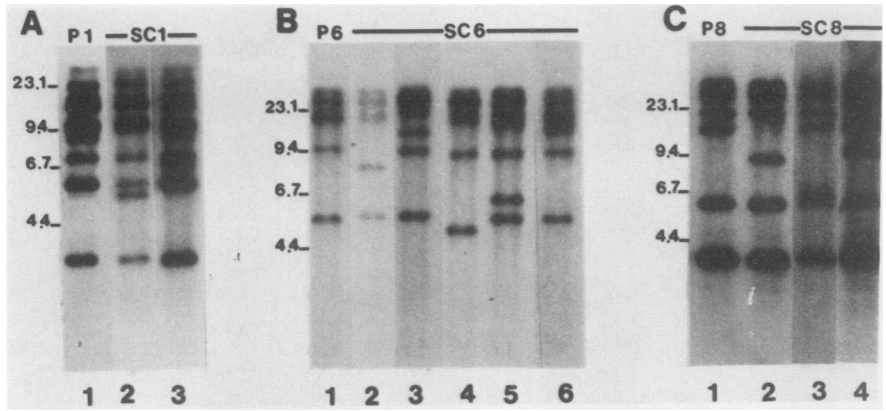

FIG. 7. Instability of the clonal restriction patterns for CRE1. Individual cells were randomly selected as described in Materials and Methods from log-phase cultures of clones 1 (panel A), 6 (panel B), and 8 (panel C) and grown in liquid medium. DNA extracted from each subclone was digested with *Eco*RI and fractionated through agarose prior to transfer and hybridization to the 3-kb probe. Although most subclones (SC) had arrangements of CRE1 identical to the parent clone (P1, P6, and P8), only examples of those which differed from the parent are shown.

although this was only apparent on longer exposures of Fig. 6F). The converse was not true. Both clones 1 and 3 had fragments of ca. 60 and 50 kb, respectively, containing miniexon genes which did not hybridize to the 3-kb probe, and which most probably represented the miniexon loci on the smaller chromosomes in the corresponding lanes in Fig. 6C. These results and equivalent findings generated with two other restriction enzymes which cut neither the miniexon genes nor CRE1 (*Bgl*II and *Ssp*I; data not shown) demonstrate that in most if not all cases, CRE1s are physically linked to large arrays of miniexon genes. Furthermore, when CRE1s are dispersed from the main miniexon arrays, they maintain their association with miniexon genes.

The clonal CRE1 restriction pattern is not stably inherited. To identify discrete rearrangements which occur at the level of individual cells, we prepared multiple subclones from three cloned lines (clones 1, 6, and 8). The subcloning took place some 30 generations after the cloning of the progenitor cells. By restriction analysis of the DNA from these progeny and hybridization with the 3-kb probe, we found that ca. 30% of the subclones had CRE1 patterns distinct from the parental pattern. Representative examples are shown in Fig. 7. These discrete changes included the concomitant appearance and disappearance of bands (+1, −1; Fig. 7A, lane 2; Fig. 7B, lanes 2 to 4; and Fig. 7C, lane 2), the appearance of one or more new bands without the apparent loss of any other bands (+1, +2; Fig. 7A, lane 3; Fig. 7B, lanes 5 and 6; and Fig. 7C, lanes 3 and 4), or the appearance and disappearance of multiple bands (data not shown). Interestingly, in more than 20 discrete rearrangements observed, none consisted simply of the deletion of a band.

Given that 30% of these progeny have discrete rearrangements and that we are observing events which have accumulated over 30 generations of growth, the total rate of rearrangement is estimated to be 1% per generation. This remarkably high rate of rearrangement is a minimum estimate, since not all rearrangements are detected by the agarose gel assay.

DISCUSSION

We have demonstrated that the *C. fasciculata* miniexon gene cluster is interrupted by multiple copies of a 3.5-kb element which encodes RT and IN domains and which rapidly rearranges within the tandem array. Unlike most other "retroid" (i.e., RT-encoding) entities, CRE1 has a low copy number and is associated with miniexon genes rather than being widely dispersed. It has a specific site of integration and rearranges at the remarkable rate of ca. 1% per generation. CRE1 is the first transposon described in *C. fasciculata*.

Does CRE1 actually encode an RT used in its transposition? Like retroviral *pol* genes, the CRE1 ORF does not contain an initiator methionine codon at its 5′ end. In many retroviral genomes, *pol* is transcribed and translated as part of a *gag-pol* fusion, alleviating the necessity for its own AUG initiation codon (44). Although we have no evidence to date for a *gag*-like gene fused to CRE1, we cannot exclude the novel possibility that a *gag* gene is located elsewhere in the *Crithidia* genome and that its mRNA is *trans* spliced to a CRE1 transcript prior to its translation. Other explanations for the absence of an initiator methionine in an otherwise open 1,140-amino-acid reading frame include the trivial possibility that the six independently sequenced 5′ ends of CRE1 are all pseudogenes, the possibility of altered rules for translation initiation in *C. fasciculata*, or the possibility that RNA editing is used to create an AUG posttranscriptionally (although this has not yet been observed outside of mitochondrial transcripts). Sequence analysis of CRE1 cDNAs may resolve these possibilities.

Of note, analysis of *Crithidia* RNA by filter hybridization does not reveal discrete RNA species which hybridize to CRE1-specific probes. This result suggests that CRE1 may be transcribed at very low levels, that its mRNA is rapidly degraded, or that its expression is regulated in an unknown way.

Relationship of CRE1 to other retrotransposons. Despite the sequence similarities with mammalian retroviruses, the order of domains in the CRE1 ORF (5′-IN-RT-3′) is opposite to that found in retroviruses. In fact, this arrangement has not previously been found outside of the Ty and copia group of LTR-containing retrotransposons, whose sequences are only very distantly related to CRE1. The organization and structure of the insertion element is most closely associated with a heterogeneous collection of RT-encoding elements categorized as either non-LTR retrotransposons (48) or poly(A) retrotransposons (3). None of these elements con-

tain LTRs, but all have short target site duplications and 3′ A-rich regions. Among this collection of retrotransposons, the *Bombyx mori* rDNA insertion elements R1Bm and R2Bm (5, 48) are organized most similarly to the *Crithidia* element in that they appear to integrate site specifically within tandemly arrayed genes. A phylogenetic tree of RT sequences from a variety of retroid elements generated to include the *Crithidia* ORF locates CRE1 as having branched from the progenitor of the non-LTR retrotransposons, despite its closer sequence similarity to retroviruses, suggesting rapid sequence evolution of this class of elements (Doolittle, personal communication). On the other hand, unlike CRE1, none of the established non-LTR retrotransposons have significant sequence homology to retroviral INs, although Xiong and Eickbush have demonstrated functional expression of a sequence-specific endonuclease from R2Bm (46).

How is CRE1 related to the miniexon insertion elements found in other trypanosomatids? Remarkably, the target sites for integration of CRE1, SLACS/MAE (1, 7), and the recently identified LINS 1, a 293-bp insert within the *Leptomonas seymouri* miniexon gene cluster (2), are all overlapping, occurring within 2 bases of one another within the highly conserved 39 nucleotides encoding the miniexon sequence. This is all the more striking considering the difference in length of the respective target site duplications and the lack of homology detected from the available sequence data on the other insertion elements. Collectively, these results suggest that the miniexon locus in trypanosomatids may serve as a sink for reverse-transcribed genes. It is a highly reiterated target, with each copy containing the same miniexon sequence found on all mRNAs. Furthermore, the conserved site within the miniexon genes may be the target for an endogenous endonuclease which facilitates the insertion of reverse transcripts.

Genetic mechanisms underlying rapid rearrangement. What genetic mechanisms account for the rapid rearrangement of this retrotransposon? The hypervariability noted in individual cells from the same population (Fig. 5) is reminiscent of the individual-specific "fingerprints" observed by Jeffreys et al. in human DNA probed with "minisatellites," i.e., short, dispersed tandem repeats (19, 20). The organization of CRE1s within long arrays of small tandem genes (Fig. 5) is analogous to the situation in transgenic mice which contain a single copy of a Moloney murine leukemia virus proviral genome integrated into highly repetitive tandem sequences on the sex chromosomes (16). In both of these examples, the high frequency of rearrangements is thought to be due to meiotic unequal crossing over within the tandem arrays.

Little, however, is known about the *Crithidia* life cycle. Trypanosomatid chromosomes do not condense at metaphase, and their ploidy has not been definitively established. Whether they even undergo meiosis remains controversial. Mitotic sister chromatid exchange within the miniexon array, on the other hand, is a plausible mechanism for generating polymorphisms at this locus. Szostak and Wu estimated the mitotic unequal crossover rate in the rDNA locus of *Saccharomyces cerevisiae* to be ca. 1% per generation, by extrapolating from the observed rate of loss of an inserted genetic marker (41). By using the same reasoning, the unequal crossover rate in the *Crithidia* miniexon locus would have to be substantially higher than our deduced 1%, since we are scoring only events which occur near copies of CRE1.

Thus, although it is conceivable that the observed rearrangements are due to unequal crossing over of miniexon genes, the following lines of evidence suggest that this is not the sole mechanism of rearrangement. (i) The variability within the 3′-terminal poly(dA) tract suggests multiple independent insertion events. A similar phenomenon has been noted by Pritchard et al. (32), who observed changes in the number of 3′-terminal TAA repeats during transposition of the *Drosophila* I element. (ii) CRE1 theoretically encodes a means for its own replication and transposition. If CRE1 were simply a hapless reverse transcript which integrated once, in the distant past, into the miniexon cluster, it should, over time, either be lost or at least accumulate nonsense mutations via neutral drift. The results shown in Fig. 3 demonstrate that the reading frame is intact. Furthermore, other morphologically distinct *Crithidia* species also contain 4-kb *Hin*dIII fragments which hybridize at high stringency with the 3-kb probe (A. Gabriel, unpublished observation). The maintenance of an intact ORF suggests that CRE1 encodes a functional element, while its appearance in other *Crithidia* species indicates that it is not an evolutionarily recent intruder into the *C. fasciculata* genome. (iii) Sister chromatid exchange cannot account for the presence of CRE1 on different chromosomes in different cloned lines. This would require transposition, translocation, or nonchromatid gene conversion events. However, in the absence of independent chromosome markers, we cannot categorically rule out the possibility that the multiple discrete chromosomes for clones 4 and 8 are the result of contraction and expansion of the long tandem arrays.

We speculate that reciprocal and/or nonreciprocal recombination (gene conversion), along with active transposition, may all occur concomitantly to maintain the variation within the *Crithidia* miniexon locus. In this regard, it is noteworthy that there is evidence for both site-specific recombination and retrotransposition mechanisms underlying the mobility of fungal mitochondrial introns, which encode RTs and/or INs and which appear to be evolutionarily related to non-LTR retrotransposons (23, 47). Current work is under way to express the functional domains of CRE1 and to unravel the specific genetic mechanisms underlying these rapid rearrangements.

ACKNOWLEDGMENTS

Special thanks to Russell Doolittle for independent analysis of our sequence data and to Paul Englund for providing us with *C. fasciculata*. We are grateful to Kim Arndt, Lawrence Brody, Carol Dowling, Karen Hofman, and Grant Mitchell for technical advice and to Andrew Fire, Hugh Rienhoff, and Sam Sisodia for critical review of the manuscript.

This work was supported in part by Public Health Service grants AI00803, GM34231, and CA16519 to A.G., D.W.C. and B.S.W., and J.D.B. respectively, from the National Institutes of Health, and by the John D. and Catherine T. MacArthur Foundation (to D.W.C., D.C.S., and B.S.W.). T.J.Y. was supported by an American Cancer Society Postdoctoral fellowship.

LITERATURE CITED

1. **Aksoy, S., T. M. Lalor, J. Martin, L. H. T. Van der Ploeg, and F. Richards.** 1987. Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. EMBO J. **6**:3819–3826.
2. **Bellofatto, V., R. Cooper, and G. A. M. Cross.** 1988. Discontinuous transcription in *Leptomonas seymouri*: presence of intact and interrupted mini-exon gene families. Nucleic Acids Res. **16**:7437–7456.
3. **Boeke, J. D., and V. G. Corces.** 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. **43**: 403–433.
4. **Boothroyd, J. C., and G. A. M. Cross.** 1982. Transcripts coding

for variant surface glycoproteins of *Trypanosoma brucei* have a short, identical exon at their 5′ end. Gene **20**:281–289.
5. **Burke, W. D., C. C. Calalang, and T. H. Eickbush.** 1987. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. Mol. Cell. Biol. **7**:2221–2230.
6. **Campbell, D. A., D. A. Thornton, and J. C. Boothroyd.** 1984. Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature (London) **311**:350–355.
7. **Carrington, M., I. Roditi, and R. O. Williams.** 1987. The structure and transcription of an element interspersed between tandem arrays of mini-exon donor RNA genes in *Trypanosoma brucei*. *Nucleic Acids Res.* **15**:10179–10198.
8. **Childs, G., R. Maxson, R. H. Cohn, and L. Kedes.** 1981. Orphons: dispersed genetic elements derived from tandem repetitive genes of eucaryotes. Cell **23**:651–663.
9. **Cornelissen, A. W. C. A., M. P. Verspieren, J. J. Toulme, B. W. Swinkels, and P. Borst.** 1986. The common 5′ terminal sequence on trypanosome mRNAs: a target for anti-messenger oligodeoxynucleotides. Nucleic Acids Res. **14**:5605–5614.
10. **Davis, R. W., M. Thomas, J. Cameron, T. P. St. John, S. Scherer, and R. A. Padgett.** 1980. Rapid DNA isolations for enzymatic and hybridization analysis. Methods Enzymol. **65**:404–411.
11. **Doolittle, R. F., D. F. Feng, M. S. Johnson, and M. A. McClure.** 1989. Origin and evolutionary relationships of retroviruses. Q. Rev. Biol. **64**:1–30.
12. **Englund, P. T.** 1978. The replication of kinetoplast DNA networks in *Crithidia fasciculata*. Cell **14**:157–168.
13. **Feagin, J. E., J. M. Abraham, and K. Stuart.** 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell **53**:413–422.
14. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. [Addendum.] Anal. Biochem. **137**:266–267.
15. **Gabriel, A., S. S. Sisodia, and D. W. Cleveland.** 1987. Evidence of discontinuous transcription in the trypanosomatid *Crithidia fasciculata*. J. Biol. Chem. **262**:16192–16199.
16. **Harbers, K., P. Soriano, U. Muller, and R. Jaenisch.** 1986. High frequency of unequal recombination in pseudoautosomal region shown by proviral insertion in transgenic mouse. Nature (London) **324**:682–685.
17. **Hasan, G., M. J. Turner, and J. S. Cordingley.** 1982. Ribosomal RNA genes of *Trypanosoma brucei*. Cloning of a rRNA gene containing a mobile element. Nucleic Acids Res. **10**:6747–6761.
18. **Hasan, G., M. J. Turner, and J. S. Cordingley.** 1984. Complete nucleotide sequence of an unusual mobile element from *Trypanosoma brucei*. Cell **37**:333–341.
19. **Jeffreys, A. J., N. J. Royle, V. Wilson, and Z. Wong.** 1988. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature (London) **332**:278–281.
20. **Jeffreys, A. J., V. Wilson, and S. L. Thein.** 1985. Individual-specific 'fingerprints' of human DNA. Nature (London) **316**: 76–79.
21. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc. Natl. Acad. Sci. USA **83**:7648–7652.
22. **Kimmel, B. E., O. K. Ole-Moiyoi, and J. R. Young.** 1987. Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. Mol. Cell. Biol. **7**:1465–1475.
23. **Lambowitz, A. M.** 1989. Infectious introns. Cell **56**:323–326.
24. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
25. **Marini, J. C., S. D. Levene, D. M. Crothers, and P. T. Englund.** 1982. Bent helical structure in kinetoplast DNA. Proc. Natl. Acad. Sci. USA **79**:7664–7668.
26. **Monteiro, M. M., and R. A. Cox.** 1986. Characterization of ard B and ard C actin gene loci of *Physarum polycephalum*. Mol. Gen. Genet. **204**:153–160.
27. **Muhich, M. L., D. E. Hughes, A. M. Simpson, and L. Simpson.** 1987. The monogenetic kinetoplastid protozoan, *Crithidia fasciculata* contains a transcriptionally active, multicopy mini-exon sequence. Nucleic Acids Res. **15**:3141–3153.
28. **Murphy, N. B., A. Pays, P. Tebabi, H. Coquelet, M. Guyaux, M. Steinert, and E. Pays.** 1987. *Trypanosoma brucei* repeated element with unusual structural and transcriptional properties. J. Mol. Biol. **195**:855–871.
29. **Murphy, W. J., K. P. Watkins, and N. Agabian.** 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA processing: evidence for *trans* splicing. Cell **47**:517–525.
30. **Parsons, M., R. G. Nelson, K. Stuart, and N. Agabian.** 1984. Variant antigen genes of *Trypanosoma brucei*: genomic alteration of a spliced leader orphon and retention of expression-linked copies during differentiation. Proc. Natl. Acad. Sci. USA **81**:684–688.
31. **Parsons, M., R. G. Nelson, K. P. Watkins, and N. Agabian.** 1984. Trypanosome mRNAs share a common 5′ spliced leader sequence. Cell **38**:309–316.
32. **Pritchard, M. A., J.-M. Dura, A. Pelisson, A. Bucheton, and D. J. Finnegan.** 1988. A cloned I-factor is fully functional in *Drosophila melanogaster*. Mol. Gen. Genet. **214**:533–540.
33. **Rogers, J.** 1983. Retroposons defined. Nature (London) **301**: 460–460.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell **37**:67–75.
36. **Schwartz, D. C., W. Saffran, J. Welsh, R. Haas, M. Goldenberg, and C. R. Cantor.** 1983. New techniques for purifying large DNA and studying their properties and packaging. Cold Spring Harbor Symp. Quant. Biol. **47**:189–195.
37. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80**:3618–3622.
38. **Shaw, J. M., J. E. Feagin, K. Stuart, and L. Simpson.** 1988. Editing of kinetoplastid mitochondrial mRNAs by uridine addition and deletion generates conversed amino acid sequences and AUG initiation codons. Cell **53**:401–411.
39. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–548.
40. **Sutton, R. E., and J. C. Boothroyd.** 1986. Evidence for *trans* splicing in trypanosomes. Cell **47**:527–535.
41. **Szostak, J. W., and R. Wu.** 1980. Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. Nature (London) **284**:426–430.
42. **Toh, H., H. Hayashida, and T. Miyata.** 1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature (London) **305**:827–829.
43. **Van der Ploeg, L. H. T., D. C. Schwartz, C. R. Cantor, and P. Borst.** 1984. Antigenic variation in *Trypanosoma brucei* analyzed by electrophoretic separation of chromosome-sized DNA molecules. Cell **37**:77–84.
44. **Varmus, H.** 1988. Retroviruses. Science **240**:1427–1434.
45. **Walder, J. A., P. S. Eder, D. M. Engman, S. T. Brentano, R. Y. Walder, D. S. Knutzon, D. M. Dorfman, and J. E. Donelson.** 1986. The 35-nucleotide spliced leader sequence is common to all trypanosome messenger RNAs. Science **233**:569–571.
46. **Xiong, Y., and T. H. Eickbush.** 1988. Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. Cell **55**:235–246.
47. **Xiong, Y., and T. H. Eickbush.** 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. **5**:675–690.
48. **Xiong, Y., and T. H. Eickbush.** 1988. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol. Cell. Biol. **8**:114–123.