Cell, Vol. 37, 333–341, May 1984, Copyright © 1984 by MIT 0092-8674/84/050333-09 $02.00/0

Complete Nucleotide Sequence of an Unusual Mobile Element from Trypanosoma brucei

Gaiti Hasan,* Mervyn J. Turner,* and John S. Cordingley†
*MRC Unit for Biochemical Parasitology
The Molteno Institute
Downing Street
Cambridge CB2 3EE, England
†Department of Pathology
University of Cambridge
Tennis Court Road
Cambridge CB2 1QP, England

Summary

The complete nucleotide sequence of a mobile element from Trypanosoma brucei is presented along with the sequence of its target site, which shows that the insertion has generated a 7 base pair direct repeat. The cloned copy of the element is a dimeric structure, one end of each monomer consisting of a stretch of 14 A residues preceded by a putative trypanosome polyadenylation signal. Six base pairs of DNA of unknown origin are found in the dimer between the two copies of the element. Evidence suggests that the element is present in the genome mainly as a monomer whose sequence is conserved across several species of trypanosome. The element contains an open reading frame encoding the same 160 amino acid protein in both sequenced copies and is extensively transcribed from both strands.

Introduction

Salivarian trypanosomes are parasitic protozoa that inhabit the bloodstream and tissue spaces of their mammalian hosts. They avoid elimination by the immune system by a process of antigenic variation that is the result of the sequential expression of a large repertoire of variant surface glycoprotein (VSG) genes (reviewed in Borst and Cross, 1982; Englund et al., 1982; Turner, 1982). The mechanism of antigenic variation has been the subject of intense investigation in recent years, and it is now well known that rearrangements of VSG genes are associated with expression. Recent sequence data (Michiels et al., 1983; Bernards et al., 1981) have shown that the rearrangement of VSG genes is not the result of a transposition event of the type described for certain well known eucaryotic mobile elements such as Tyl in yeast (Fink et al., 1981) or Copia in Drosophila (Rubin et al., 1981). Transposition of these eucaryotic mobile elements is in some ways similar to the transposition of procaryotic transposons (Kleckner, 1981); that is, they have defined ends and cause the formation of short flanking direct repeats, derived from sequences at their site of insertion.

This raises the obvious question whether the trypanosome genome contains, in addition to VSG genes, mobile elements analogous to other systems. We present the sequence of just such a mobile element, which we have isolated inserted into a ribosomal gene of Trypanosoma brucei MITat 1.6 (Hasan et al., 1982). The element has generated a short direct repeat from the target ribosomal gene sequence, and its sequence appears to be conserved across at least four species of trypanosomes. The cloned copy of the element has an unusual dimeric structure and is extensively transcribed from both strands into multiple polyadenylated RNAs. The sequence of the element suggests a hypothesis about the recent history of this particular copy of the element. Although this mobile element has features in common with other previously described mobile elements and processed pseudogenes in mammals, there are novel features of the DNA sequence that have not previously been seen. One obvious conclusion is that several different mechanisms for rearranging DNA sequences coexist in trypanosomes.

Results

Sequencing Strategy

We have previously described two ribosomal DNA clones from T. b. brucei, pGH330 and pGH331. These two clones both contain a Bgl II fragment encoding the 3′ 3 kb of the major ribosomal gene repeat, but additionally pGH330 has a 1.1 kb mobile element inserted into the 3 kb rDNA fragment. This ribosomal mobile element (RIME) was shown to be a member of a moderately repeated gene family and to be arranged differently in various trypanosome strains. RIME is present in a small fraction of the ribosomal genes of the MITAR 1 serodeme and absent from those of the ILTAR 1 serodeme (Hasan et al., 1982). A comparison of the restriction maps of pGH330 and pGH331 (Figure 1) showed that RIME was inserted into a 0.38 kb Alu I fragment of pGH331 (A 0.38). In pGH330, A 0.38 is lost and replaced by a 1.5 kb fragment (A 1.5) that must contain 1.1 kb of RIME in addition to 0.38 kb derived from pGH331.

A 0.38 and A 1.5 were purified from agarose gels, subcloned in both orientations into the single-stranded DNA phage M13mp8, and sequenced by the chain termination method of Sanger et al. (1977). Two additional strategies were employed. First, specific restriction fragments were subcloned. From the restriction map of pGH330 (Figure 1) all of the sequence of RIME and at least 275 bp of its flanking sequences are present in three Sau 3A fragments of 2.1 kb (S 2.1), 0.51 kb (S 0.51), and 0.36 kb (S 0.36). Both S 0.51 and S 0.36 were purified from gels and subcloned, while S 2.1 was gel-purified, digested with Alu I, and repaired with DNA polymerase I prior to subcloning into the Sma I site of M13mp8 by blunt-end ligation. Second, the A 1.5 fragment of pGH330 was partially digested with Bal 31 exonuclease, repaired, and subcloned into the Sma I site of M13mp8 in a similar fashion. The extent of DNA sequences read from these subclones and their orientation with respect to the direction

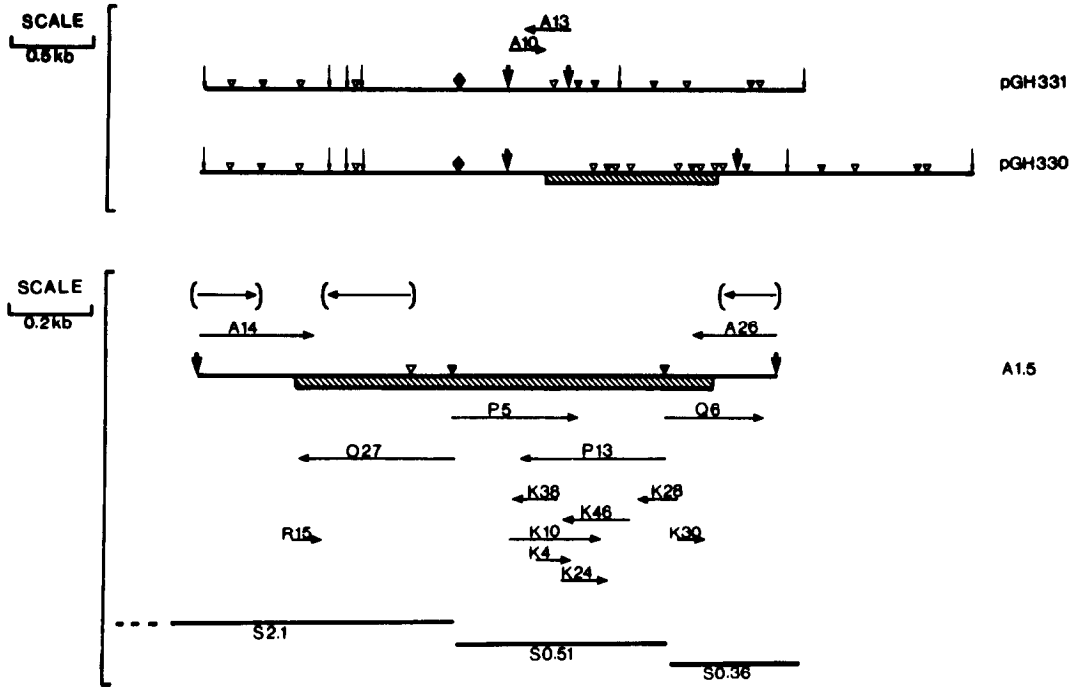

Figure 1. Restriction Maps of pGH331 and pGH330, Indicating the Strategy for Sequencing the A 0.38 Fragment of pGH331 and the A 1.5 Fragment of pGH330

The hatched region in pGH330 indicates the position of RIME. Symbols indicate restriction sites as follows: (♦) Hinc II; (↓) Alu I; (▽) Hpa II; (▼) Sau 3A. The thick vertical arrows (⬇) indicate corresponding Alu I sites that enclose the fragments A 0.38 in pGH331 and A 1.5 in pGH330. Both these fragments have been subcloned in phage M13mp8 and sequenced to completion. The extent and direction of sequence read from two subclones A10 and A13 of A 0.38 is indicated by horizontal arrows at the top. The extent and direction of sequence read from various subclones of A 1.5 is drawn on an enlarged scale in the lower half, and is also indicated by horizontal arrows. Subclones generated by Bal 31 digestion are indicated by the prefix "K." For the sake of clarity, not all of these subclones are shown. Only the restriction sites relevant for sequencing have been drawn on A 1.5. Regions that have also been sequenced chemically are indicated by horizontal arrows in parentheses. The positions of three Sau 3A fragments, S 2.1, S 0.51, and S 0.36, that have been used for subcloning and sequencing are indicated by horizontal lines at the bottom.

of rRNA transcription are shown in Figure 1. Parts of A 1.5 were also sequenced by the method of Maxam and Gilbert (1980), and are shown in parentheses in Figure 1.

A 7 Base Pair Direct Repeat Is Generated at the Target Site

Comparison of the complete sequences of A 0.38 and A 1.5 reveals that the sequences are identical for 234 bp from the 5′ ends relative to the direction of rRNA transcription. Similarly, the 3′ ends are identical for 132 bp. In A 1.5, a 7 bp sequence, found once in A 0.38, is present at both ends of the inserted sequence. The DNA sequences are such that the ends of the insert can be precisely defined, in particular the run of T residues at one end of RIME is striking. Figure 2 shows the complete sequences of A 0.38 and A 1.5; however, the sequences are presented in the opposite orientation to rRNA transcription for reasons explained below.

The Complete Sequence of RIME

Preliminary experiments demonstrated that the internal 0.51 kb Sau 3A fragment from RIME hybridized to the two Sau 3A fragments containing the ends of RIME (S 2.1 and S 0.36), suggesting that RIME is internally repetitive (data not shown). The completed DNA sequence revealed that RIME is a nearly perfect dimer of a 511 bp monomer (Figure 2). The dimeric nature of RIME presented difficulties in assembling the sequence, since many of the internal fragments could not be unambiguously placed. However, this problem was resolved from the sequences of clones O27, P5, P13, and Q6, which together cover the whole of RIME. Each fragment was sequenced from several independent subclones. P5 is a subclone of S 0.51 in one of the possible orientations and contains the junction of the two monomers. This shows 6 bp of DNA are present between the poly(A) tract defining the 3′ end of one monomer and the beginning of the second monomer at the sequence ACAC, also found at the 5′ end of the first monomer directly adjacent to the 7 bp direct repeat. P13 and O27 generate identical sequences; however, P13 is a subclone of S 0.51 in the opposite orientation to P5, and O27 is a subclone of S2.1. Consequently their sequences must be derived from the two different copies of the monomer. The orientation of S 0.51 was determined from the sequence of the subclone K28, which contains one of the Sau 3A sites and demonstrates that K28 comprises the ends of P13 and Q6 rather than those of P5 and Q6.

The total length of RIME is 1028 bases, counting from

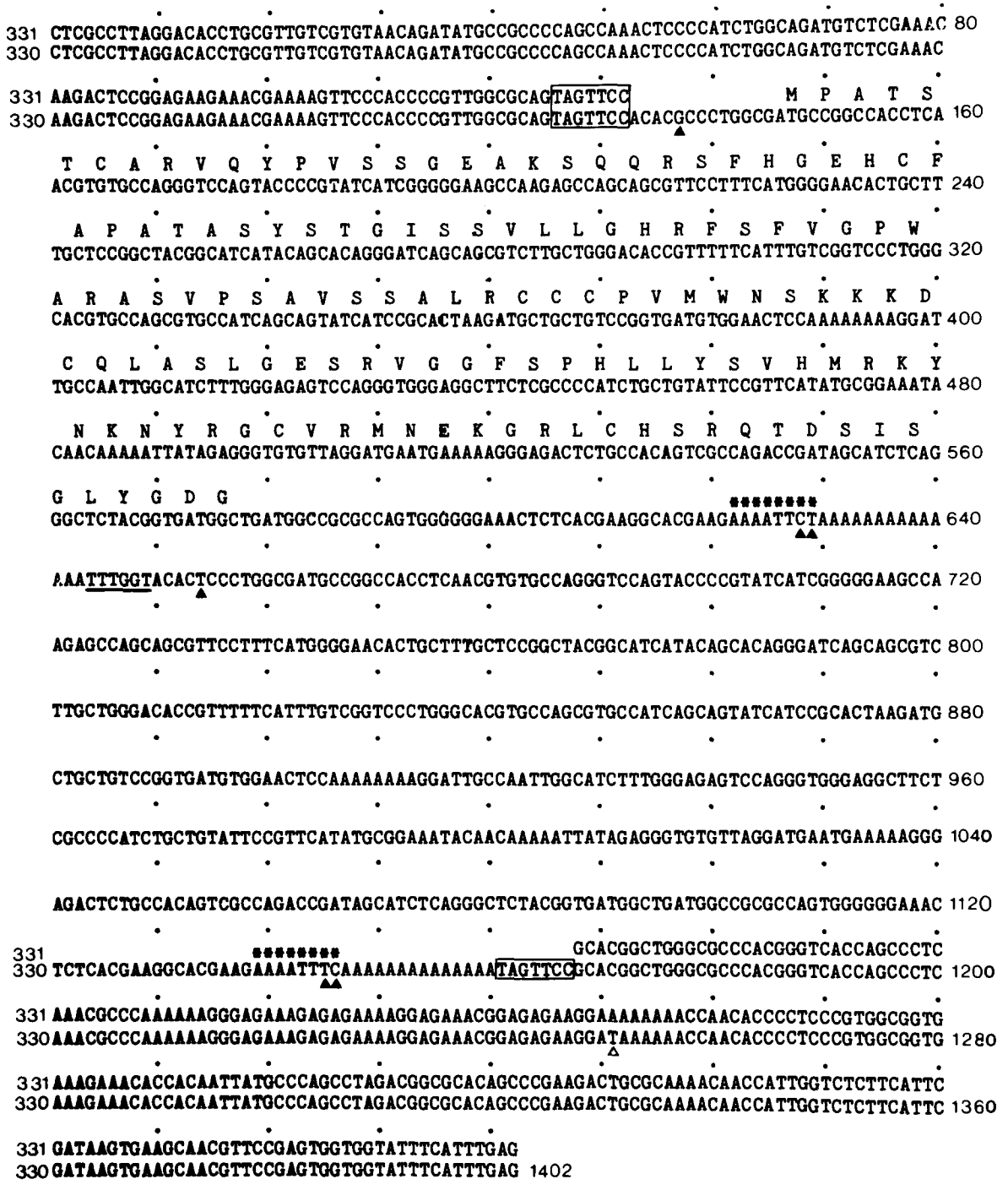

Figure 2. Complete Sequence of the A 0.38 Fragment of pGH331 and the A 1.5 Fragment of pGH330

The sequence of pGH331 is marked as 331. The corresponding sequence of pGH330 is given below the 331 sequence and is marked as 330. A single base difference between the sequence of 331 and 330 is marked by an open triangle (△). The sequence is given in the direction of the reading frame in RIME (and in the opposite orientation to rRNA transcription), which is shown by the single-letter code. The reading frame has been marked only for one monomer, though the same reading frame also exists in the second monomer. The 7 bp direct repeat at each end of RIME is boxed. Base differences between the two monomers are marked by solid triangles (▲). A possible trypanosome-specific polyadenylation signal is marked with asterisks. The 6 bp sequence of unknown origin, present between the two monomers, is underlined.

after the last nucleotide of the first direct repeat to before the first nucleotide of the second direct repeat. Of these, bases 1 to 511 (133 to 643 of A 1.5) match with bases 518 to 1028 (650 to 1161 of A 1.5), with the exceptions of positions 5, 496, 497, 522, 1013, and 1014. These bases are indicated by solid triangles in Figure 2. An open triangle

indicates a single base difference between A 1.5 and A 0.38 in the ribosomal gene sequence (1251 in A 1.5).

RIME Monomers Code for a Putative Protein of 160 Amino Acids

Each RIME monomer resembles a reverse transcript of an mRNA with a short poly(A) tail. Consequently both orientations of RIME were scanned for reading frames. A reading frame for 160 amino acids is present in one orientation. Both monomers have identical reading frames, since the three bases that are different in the two monomers lie outside the reading frame (Figure 2), which is in the opposite orientation to the direction of rRNA transcription. Certain sequenced eucaryotic pseudogenes have very similar structures to RIME monomers, though in their case the reading frame is frequently lost because of single base pair changes. A polyadenylation signal is often found in some of these processed pseudogenes 5′ to the poly(A) tail (Lemischka and Sharp, 1982). However, the polyadenylation signal AATAAA of higher eucaryotes has not been found in trypanosome VSG mRNAs (Bernards et al., 1981), suggesting that trypanosomes may have a different polyadenylation signal. It has not been possible to identify such a signal because the 3′ ends of VSG mRNAs show considerable homologies. Interestingly, when the 3′ ends of RIME monomers were compared with the 3′ ends of VSG mRNAs a short region of homology was detectable. Thus the sequence AAAATTCT, which is marked by asterisks in the first monomer, or the sequence AAAATTTT is present once or twice before the poly(A) site in all VSG mRNAs sequenced to date (Borst and Cross, 1982). It therefore seems likely that the sequence AAAATTPyT is the trypanosome polyadenylation signal. In the second monomer the sequence is AAAATTTC, and this may be another variation of the polyadenylation signal, though it is also possible that it is a mutation that has arisen in this monomer at a later stage.

Transcripts of RIME in Both Orientations Are Present in Polyadenylated RNA

To investigate the transcription of RIME sequences polyadenylated (PA+) and nonpolyadenylated (PA−) RNA from T. b. brucei were searched for transcripts of RIME by Northern blotting. Total RNA extracted from T. b. brucei was separated into PA+ and PA− fractions by chromatography on oligo(dT)-cellulose. Equal amounts of both RNAs were then separated on duplicate agarose–formaldehyde gels and blotted onto nitrocellulose filters. The filters were hybridized to single-stranded probes prepared by primed synthesis on the two M13 subclones, P5 and P13, of the internal Sau 3A fragment of RIME (S 0.51), which are in opposite orientations. As seen in Figure 3, transcripts of RIME are abundant in the PA+ fraction in both orientations. They are mostly longer than 1 kb. Hybridization is also visible to the PA− fraction, but it is fainter and to a subset of bands also seen in the PA+ fraction, suggesting it results from contamination of PA− RNA with PA+, and not vice versa. The bands of hybridization in PA− RNA are mainly of high molecular weight and may result from a small proportion of the transcripts from particular loci not being polyadenylated. Some abundant transcripts, however, are only detectable in PA+ RNA at the sensitivity of the blotting procedure. It is not unusual to find transcription in both orientations of various eucaryotic mobile elements (Georgiev et al., 1983). Presumably it is a consequence of their insertion, in either orientation, into genomic sites that are subsequently transcribed. The presence of Alu family members in RNA polymerase II transcription units of rats in both orientations has been described (Page et al., 1981). At the present level of detection no transcripts the size of RIME monomers (i.e. at approximately 0.5 kb) are visible. This is not merely a consequence of inefficient retention of small RNA molecules by nitrocellulose, since we have been able to detect RNA molecules of 0.2 kb using this methodology (data not shown).

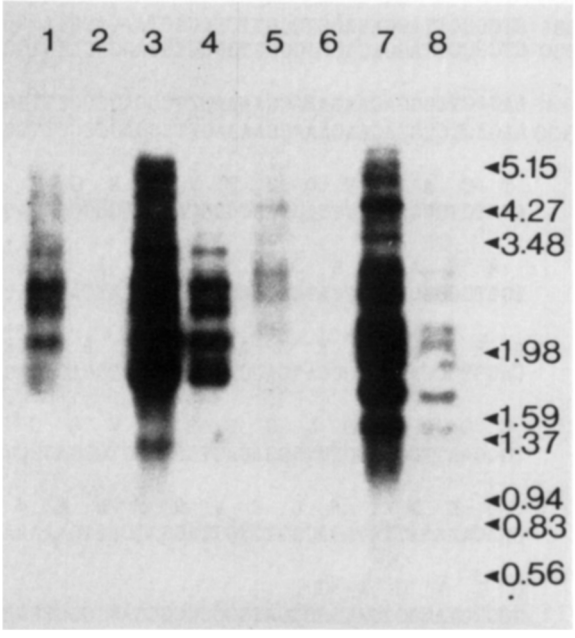

Figure 3. Northern Blotting Experiment to Investigate the Transcription of RIME Sequences in T. b. brucei

Nonpolyadenylated RNA (5 μg, lanes 1 and 5; 0.5 μg, lanes 2 and 6) and polyadenylated RNA (5 μg, lanes 3 and 7; 0.5 μg, lanes 4 and 8) were fractionated on a 1.2% agarose–formaldehyde gel that was subsequently blotted to a nitrocellulose filter. Lanes 1 to 4 were probed with a single-stranded probe prepared by primed synthesis on the M13 subclone, P5, of the S 0.51 fragment of RIME. Lanes 5 to 8 were probed with a single-stranded probe similarly prepared from the M13 clone P13, which contains the S 0.51 fragment of RIME, but in the opposite orientation to P5. The sizes of marker fragments are indicated in kilobases.

Arrangement of the RIME Family of Sequences in the Trypanosome Genome

If the coding sequence in RIME has a functional significance, its conservation within the trypanosome genome might be expected, and thus we would expect that the

many copies of RIME would have a conserved structure and sequence.

With the complete DNA sequence of RIME available we examined the sequence for restriction enzymes that cut the monomer twice, once close to each end, and Hae III was found to be suitable. If the sequence of the monomer is conserved and is found predominantly as single copies in unrelated sequences, then this internal Hae III fragment should be the major Hae III digestion product detected on Southern blots of genomic DNA. As a probe we used a subclone of S 0.51 that contains a complete monomer. Thus from the positions of the Hae III sites in the monomer only 85% of the hybridization will be to the central 0.43 kb fragment. A Southern blot with Hae III–digested DNA from pGH330, T. b. brucei (MITat 1.6 and ILTat 1.26), T. b. rhodesiense, T. b. gambiense, and T. evansii using this probe shows that in all the genomic DNA digests a predominant fragment at 0.43 kb is present and corresponds in size to the internal Hae III fragment from RIME seen in the pGH330 digest. Careful quantification of the T. b. brucei MITat 1.6 lane showed that 43% of the hybridization was to the 0.43 kb fragment. An additional 20% was to a 0.22 kb band present in all the genomic digests. Although it is not possible to be certain, it is likely that this band is derived from within RIME by a point mutation that has given rise to a Hae III site approximately midway between the existing Hae III sites. Several such possible mutation sites are present in RIME, and at least three of them are at third positions in the codons. Of these, two would cause no change in amino acid sequence while one would cause a mutation from cysteine to tryptophan (Figure 2, base number 370 of A 1.5). A series of other bands of hybridization can be seen in Figure 5, which together account for the remaining 36% of the hybridization. These bands are not conserved within all the genomic DNAs. From Figure 4 it is possible to estimate the copy number of RIME, since known amounts of genomic DNA and clone pGH330 were used. We estimate that 200 copies of RIME are present per haploid genome complement, assuming 2×10^7 base pairs as the genome size of T. b. brucei (Borst et al., 1980).

Sau 3A cuts the monomer only once, so for comparison with the Hae III digests we performed Sau 3A digests of genomic DNA from a number of trypanosome strains and species and analyzed them by Southern blotting in the same way as the Hae III digests. The results of such an experiment are shown in Figure 5, and they clearly show that S 0.51 is not the major digestion product of genomic RIME sequences. Therefore dimers such as that found in pGH330 are unlikely to be the predominant form of RIME in the genome, although they may be represented more than once. Alu I digests are shown in Figure 5 for comparison since there are no Alu I sites in the cloned copy of RIME. As expected, if RIME is mostly present as a monomer in the genome, many bands of hybridization are seen smaller than the size of the dimer (i.e., less than 1.1 kb). Some hybridization is apparent at or below 0.51 kb, suggesting that some copies of RIME have an Alu I site not present in our cloned copies.

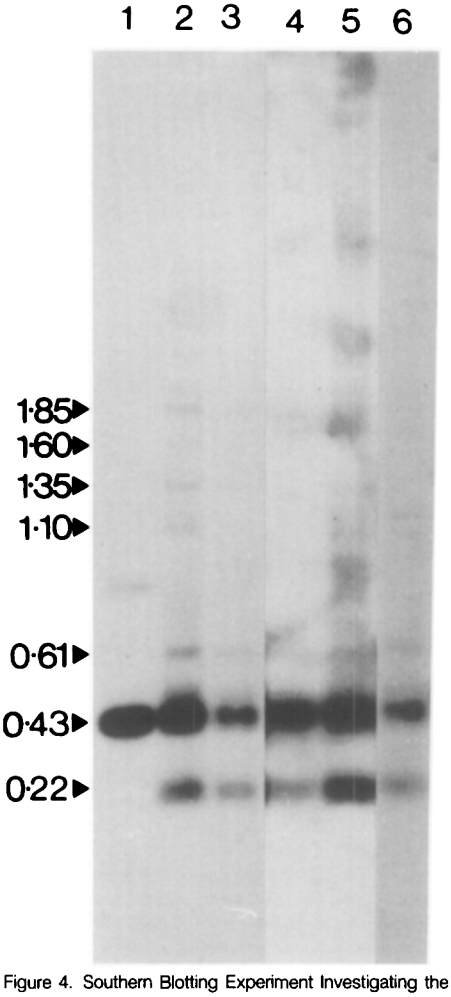

Figure 4. Southern Blotting Experiment Investigating the Conservation of Two Hae III Sites, That Are Present within the Reading Frame of Each RIME Monomer, in Genomic RIME Sequences

DNA from the following samples was digested with the restriction enzyme Hae III: lane 1, pGH330; lane 2, T. b. brucei MITat 1.6; lane 3, T. b. brucei ILTat 1.26; lane 4, T. b. rhodesiense; lane 5, T. b. gambiense; lane 6, T. evansii. DNA (2 ng) from the pGH330 digests and (1 μg) from each of the genomic DNA digests was fractionated on a 1% agarose gel, a Southern blot of which was probed with radiolabeled P5 (M13 subclone of the S 0.51 fragment of RIME). Sizes of relevant marker fragments are marked in kilobases.

Discussion

Transposition of RIME

Sequence analysis of RIME and a study of arrangement of the RIME family of sequences in the trypanosome genome have shown both similarities and dissimilarities with the structure of known eucaryotic mobile DNA sequences. The closest relationship appears to be to certain pseudogenes found in mammalian cells (Hollis et al., 1982; Vassin et al., 1980; Lemischka and Sharp, 1982; Wilde et al., 1982). Like RIME they too are found in the genome

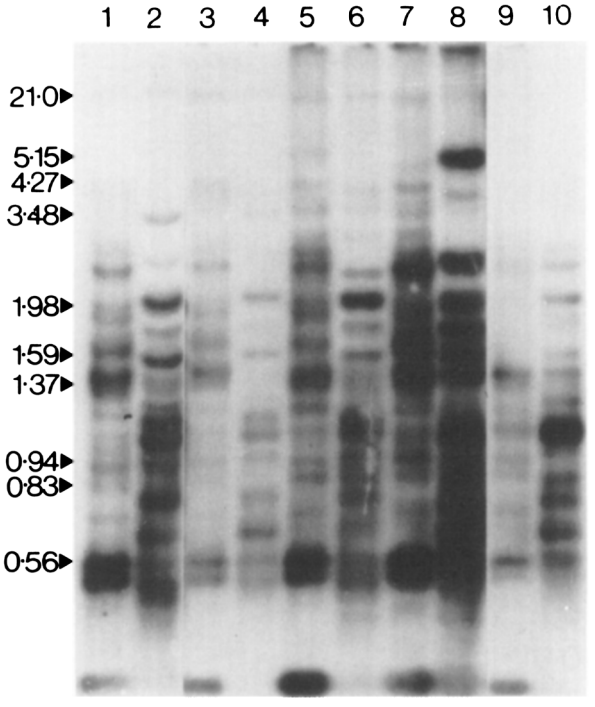

Figure 5. A Southern Blot of Genomic DNA Using Enzymes That Cut the Monomer of RIME Once (Sau 3A) or Not at All (Alu I)

Nuclear DNA from various members of the subgenus Trypanozoon was digested with Sau 3A and Alu I and run on a 1% agarose gel in the following order: lane 1, T. b. brucei MITat 1.6, Sau 3A; lane 2, T. b. brucei MITat 1.6, Alu I; lane 3, T. b. brucei ILTat 1.26, Sau 3A; lane 4, T. b. brucei ILTat 1.26, Alu I; lane 5, T. b. brucei ETat 1.8, Sau 3A; lane 6, T. b. brucei ETat 1.8, Alu I; lane 7, T. b. rhodesiense, Sau 3A; lane 8, T. b. rhodesiense, Alu I; lane 9, T. evansii, Sau 3A; lane 10, T. evansii, Alu I. Lanes 1 and 2 contain 1.5 μg of DNA, while all the other lanes contain 1 μg, with the exception of lanes 3 and 4 (0.5 μg). A Southern blot of the gel was probed with nick-translated 0.51 kb Sau 3A fragment of RIME. Sizes of marker fragments generated by digestion of lambda DNA with Eco RI and Hind III are indicated in kilobases.

with an oligo(A) stretch at one end, preceded by a polyadenylation signal, and flanked by direct repeats. Pseudogenes, however, need not be members of middle repetitive gene families. In this context, Drosophila is found to contain a middle repetitive family of sequences that are mobile, a member of which resembles a pseudogene (Dawid et al., 1981). But, unlike RIME, transcripts of this sequence are rare and only found in PA− RNA.

The proposed mechanism for the origin of these pseudogenes suggests reverse transcription of an RNA polymerase II transcript that has been spliced and polyadenylated (Lemischka and Sharp, 1982). The structure of RIME monomers suggests that RIME originated in a similar manner. As expected, no sequences resembling elements of eucaryote RNA polymerase II promoters are found in the sequenced copy of RIME, suggesting that its transcription is dependent on the presence of a promoter upstream from its site of insertion. This may be one reason that only transcripts of RIME larger than 0.5 kb are detected in Figure 3. Another deduction based on indirect evidence is that both monomers and dimers of RIME are transposed. The dimer of RIME found in pGH330 must have been inserted into the ribosomal gene by a single transposition event, since the whole structure is flanked by one pair of direct repeats. If two transposition events had been involved, one might expect 7 bp of the direct repeat to be present between the RIME monomers as well. This suggests a possible explanation for the extra 6 bp (of unknown origin) found between the two RIME monomers (see Figure 6). It is possible that these 6 bp are the remains of a direct repeat from an earlier target site of RIME, into which initially was inserted a monomer of RIME. This was followed by either insertion of a second monomer next to the 6 bp sequence or an unequal crossover between chromatids. The generation of 6 bp rather than a 7 bp duplication is not an unprecedented event since Alu repeats and F elements have also been observed with different duplication lengths. From the identical number of A residues in the two monomers, the second possibility seems more likely, since a difference in the number of A residues copied during reverse transcription of different transcripts might be expected. In either case a dimer of RIME would be formed that contained 6 extra bases in the middle derived from the target site. Subsequent transposition of the dimer would then occur with the extra 6 bp. A diagrammatic representation of the possible steps involved in transposition of what was originally a RIME monomer, to the final RIME dimer found in pGH330, is shown in Figure 6. An alternative explanation for the 6 bp is simply that these bases have been lost from one copy of the monomer in the dimer we have sequenced. The sequence of additional RIME monomers will clarify this point.

The presence of the putative internal polyadenylation site and the identical lengths of the two poly(A) lanes strongly suggest that the RIME dimer must have transposed to its position in the ribosomal gene without an RNA intermediate. Direct transposition and the involvement of an RNA intermediate have both been advanced as mechanisms of movement of Copia sequences in D. melanogaster (Flavell and Ish-Horowicz, 1981), but as in the case of RIME, the evidence is circumstantial. It remains to be seen if particular mobile elements in eucaryotes always favor the same mode of transposition.

Is There a Reason for RIME?

RIME occurs in the subgenus Trypanozoon, as an apparently conserved sequence family repeated approximately 200 times as judged by intensity of hybridization in Southern blots (Figure 5). Conservation of sequence in a middle repetitive family can imply either a recent evolutionary origin, which has not allowed enough time for significant divergence, or a high rate of gene conversion (Weiner and Denison, 1983). Divergence of Trypanozoons from the related subgenera like Nannomonas and Duttonella is thought to be a fairly recent occurrence in evolutionary terms (Baker, 1975). Nevertheless, from the presence of a reading frame within RIME, and also the presence of point

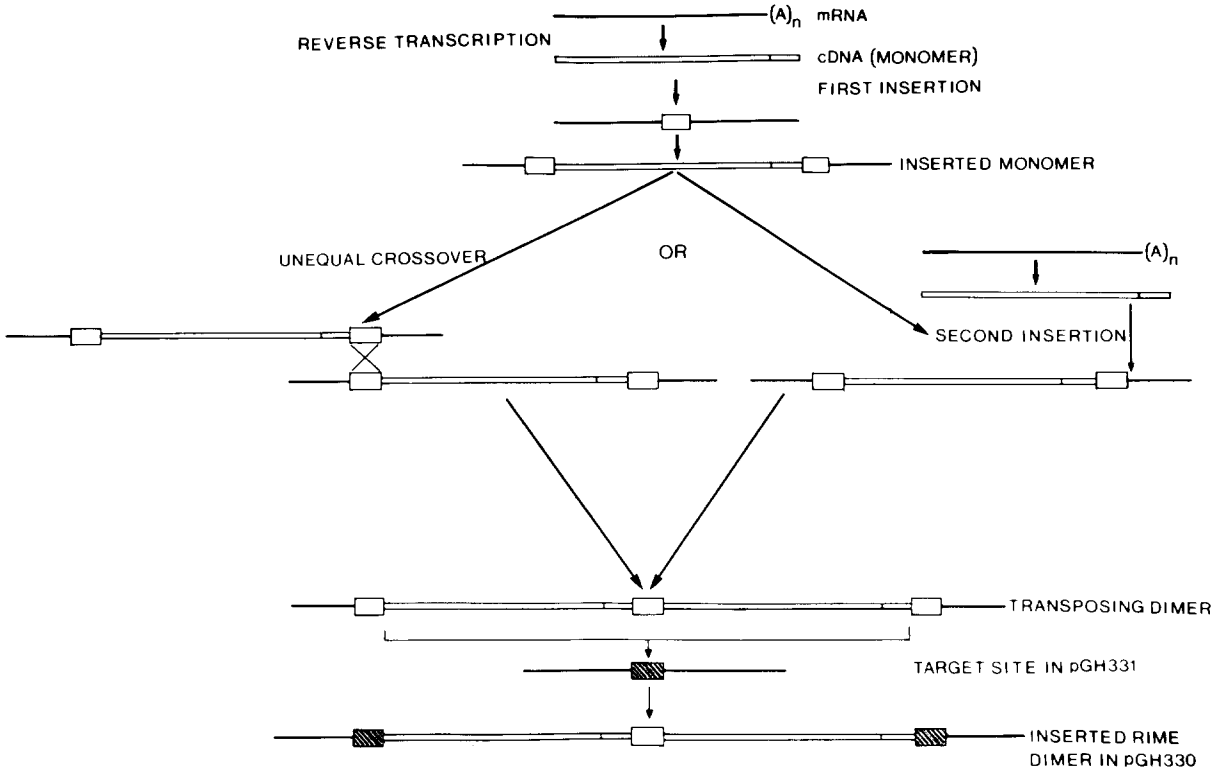

Figure 6. Model Showing the Possible Origin of RIME and the Subsequent Formation of the Dimeric Form Isolated in pGH330

Following reverse transcription of a mRNA sequence, the cDNA monomer is inserted into genomic DNA with the formation of a 6 bp direct repeat (open rectangle) generated from the target site. The start of the polyadenylated tail is indicated by the vertical line within the monomer. Unequal crossing-over between two monomers, or insertion of a second monomer sequence, could lead to the formation of a transposing dimer, which could insert into the ribosomal RNA gene cloned in pGH330, and so generating the 7 bp direct repeat (hatched rectangles).

mutations outside but not inside the reading frame, it is tempting to propose that sequence conservation of the RIME family is due to selection for this hypothetical functional significance, possibly by a gene conversion mechanism. Sequence conservation is also found in Copia and Copia-like families in Drosophila. Of these at least Copia sequences are transcribed in vivo and can be translated in vitro (Rubin et al., 1981), though as yet functions for Copia-encoded proteins are unknown. Not all middle repetitive families have conserved sequences. Specifically, the Alu family in humans shows sequence divergence from 4% to 8% (Jelinek and Schmid, 1982). On the other hand, an Alu-like family found in the intron of a rat growth hormone gene (Barta et al., 1981), and in an α-tubulin pseudogene (Lemischka and Sharp, 1982), is highly conserved. Interestingly, transcripts from this family are found to be brain specific (Sutcliffe et al., 1982). A correlation may therefore exist between functional significance of middle repetitive families and sequence conservation in them. Certainly at least one Alu-like sequence (not necessarily a member of a repeated family) has a highly functional role. Transcripts from this sequence have been found associated with, and essential for the activity of, the signal receptor particle (Walter and Blobel, 1982). A functional role for the putative protein encoded by RIME can be postulated in the actual transposition process of RIME, in a manner analogous to the transposase encoded by bacterial insertion sequences (Zupancic et al., 1983). However, until the presence of a RIME-encoded protein can actually be demonstrated, any suggestions regarding its function are mere conjecture.

A number of speculations have been made regarding the evolutionary advantage of mobile sequences to organisms (Young and Schwartz, 1981). Where species are still evolving and adapting to a particular niche, mobile sequences appear to confer a selective advantage by producing greater genetic variability (Choa et al., 1983). Since the subgenus Trypanozoon is thought to be currently undergoing a period of rapid evolution, the trend being toward developing a life cycle with a single vertebrate host (Baker, 1975), the RIME family of sequences may be advantageous to the subgenus Trypanozoon in a similar way.

Experimental Procedures

Trypanosomes

Trypanosomes were grown in rats infected intraperitoneally. At a parasitemia of 3×10^8–1×10^9 cells per milliliter, the rats were exsanguinated and the trypomastigotes purified from blood by the procedure of Lanham and Godfrey (1970). Trypanosome clones from the MITAR 1 serodeme

have been described by Cross (1975), and those from the ILTAR 1 serodeme, by Miller and Turner (1981). Nomenclature of the trypanosomes is in accordance with the published guidelines (Lumsden, 1978).

Subcloning and DNA Sequencing

Restriction fragments of 300 bp or more were purified from 1% agarose gels. The gels were run in 90 mM Tris, 90 mM boric acid, and 3 mM EDTA, pH 8.2 (TBE, pH 8.2), at 2 V/cm for 12–16 hr. Fragments of smaller sizes were purified from 6% acrylamide gels run in TBE. Partial digestion with Bal 31 of the A 1.5 fragment was according to Heitzeman et al. (1981). Subcloning of purified fragments was into the Sma I site of M13mp8 (Messing and Vieira, 1982). Approximately 200 ng of fragment DNA, the ends of which had been repaired in the presence of 50 μM dNTPs and 0.5 units of the Klenow fragment of DNA polymerase I, was used for ligation with 200–400 ng of M13mp8 DNA. Transformation of E. coli JM101 and selection of hybrid phages was as described by Messing and Vieira (1982). Dideoxy chain termination reactions for DNA sequencing were according to Sanger et al. (1977), with the modification that 5′-α-^{35}S-dATP was used instead of α-^{32}P-dATP for some of the reactions. Chemical sequencing of purified DNA fragments was as described by Maxam and Gilbert (1980).

Isolation of DNA and RNA from Trypanosomes

Isolation of DNA and RNA from 10^{10} T. b. brucei cells was by lysing the trypanosomes with 0.1% SDS. The SDS was heated to 90°C and added to a suspension of trypanosomes at 5×10^8–10^9 cells per milliliter in a buffer containing 100 mM sodium chloride, 10 mM Tris-HCl (pH 7.5), and 1 mM EDTA. DNA was spooled out from the lysate after the addition of ethanol to 70% and purified further. Total RNA was collected as an ethanol precipitate from the lysate and reprecipitated with 3 M LiCl. PA− and PA+ fractions were obtained by passage over an oligo(dT)-cellulose column. DNA from T. b. rhodesiense, T. b. gambiense, and T. evansii was a kind gift from Dr. R. O. Williams, ILRAD, Nairobi.

Blotting

Southern blotting was according to Wahl et al. (1979), and Northern blotting according to Hasan et al. (1982). All blots were washed three times at room temperature with 2× SSC (150 mM sodium chloride and 15 mM trisodium citrate) and 0.1% SDS for 5 min each. This was followed by two to three washes of 15 min each at 65°C with 0.1× SSC, 0.1% SDS. Radiolabeled probes were made by two methods. In one case purified S0.51 fragment was nick-translated (Rigby et al., 1977). In the other case single-stranded probes were made by using subclones of S0.51 fragment as templates for an "A" sequencing reaction without the addition of dideoxy A. The intensity of bands in Southern blots was measured using a gel scanner and a computer program for integrating the peak areas.

Acknowledgments

We are grateful to Judith Creighton for technical assistance in the preparation of trypanosomes and to Pippa Tuffnell for help with some of the dideoxy sequencing. We would also like to thank Mr. J. R. Young, Dr. I. Roditi and Dr. H. Haymerle for helpful discussions. One of us (G.H.) is in receipt of support from the Wellcome Trust (London), Churchill College (Cambridge), and the Inlaks Foundation (London).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 7, 1983; revised February 23, 1984

References

Baker, J. R. (1975). The evolutionary origin and speciation of the genus Trypanosoma. Symp. Soc. Gen. Microbiol. *24,* 343–366.

Barta, A., Richards, R. I., Baxter, J. D., and Shine, J. (1981). Primary structure and evolution of rat growth hormone gene. Proc. Nat. Acad. Sci. USA. *78,* 4867–4871.

Bernards, A., Van der Ploeg, L. H. T., Frasch, A. C. C., Borst, P., Boothroyd, J. C., Coleman, S., and Cross, G. A. M. (1981). Activation of Trypanosome surface glycoprotein genes involves a duplication–transposition leading to an altered 3′ end. Cell *27,* 497–505.

Borst, P., Fase-Fowler, F., Frasch, A. C. C., Hoejimakers, J. H. J., and Weijers, P. J. (1980). Characterisation of DNA from Trypanosome brucei and related trypanosomes by restriction endonuclease digestion. Mol. Biochem. Parasitol. *1,* 221–246.

Borst, P., and Cross, G. A. M. (1982). Molecular basis for trypanosome antigenic variation. Cell *29,* 291–303.

Choa, L., Vargas, C., and Spear, B. B. (1983). Transposable elements as mutator genes in evolution. Nature *303,* 633–635.

Cross, G. A. M. (1975). Identification, purification and properties of clone specific glycoprotein antigens constituting the surface coat of Trypanosoma brucei. Parasitology *71,* 393–417.

Dawid, I. B., Long, E. O., DiNocera, P. P., and Pardue, M. L. (1981). Ribosomal insertion-like elements in Drosophila melanogaster are interspersed with mobile sequences. Cell *25,* 399–408.

Englund, P. T., Hajduk, S. L., and Marini, J. C. (1982). The molecular biology of trypanosomes. Ann. Rev. Biochem. *51,* 695–726.

Fink, G., Farabaugh, P., Roeder, G., and Chaleff, D. (1981). Transposable elements (Ty) in yeast. Cold Spring Harbor Symp. Quant. Biol. *45,* 575–580.

Flavell, A. J., and Ish-Horowicz, D. (1981). Extrachromosomal circular copies of the eukaryotic transposable element copia in cultured Drosophila cells. Nature *292,* 591–595.

Georgiev, G. P., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., and Lukandin, E. M. (1983). Dispersed repetitive sequences in eukaryotic genomes and their possible biological significance. Cold Spring Harbor Symp. Quant. Biol. *47,* 1109–1121.

Hasan, G., Turner, M. J., and Cordingley, J. S. (1982). Ribosomal RNA genes of Trypanosoma brucei. Cloning of a rRNA gene containing a mobile element. Nucl. Acids Res. *10,* 6747–6761.

Heitzeman, R. A., Hagie, F. E., Levine, H. L., Goeddel, D. V., Ammerer, G., and Hall, B. D. (1981). Expression of a human gene for interferon in yeast. Nature *293,* 717–722.

Hollis, G. F., Heiter, P. A., McBride, O. W., Swan, D., and Leder, P. (1982). Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. Nature *296,* 321–325.

Jelinek, W. R., and Schmid, C. W. (1982). Repetitive sequences in eukaryotic DNA and their expression. Ann. Rev. Biochem. *51,* 813–844.

Kleckner, N. (1981). Transposable elements in prokaryotes. Ann. Rev. Genet. *51,* 341–404.

Lanham, S., and Godfrey, D. (1970). Isolation of salivarian trypanosomes from man and other mammals using DEAE cellulose. Exp. Parasitol. *28,* 521–535.

Lemischka, I., and Sharp, P. A. (1982). The sequences of an expressed rat alpha-tubulin gene and a pseudogene with an inserted repetitive element. Nature *300,* 330–335.

Lumsden, W. H. R. (1978). Proposals for the nomenclature of salivarian trypanosomes and for the maintenance of reference collections. Bull. World Health Organization *56,* 467–480.

Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labelled DNA with base-specific chemical cleavages. Meth. Enzymol. *65,* 499–560.

Messing, J., and Vieira, J. (1982). A new pair of M13 vectors for selecting either strand of double digested restriction fragments. Gene *19,* 269–276.

Michiels, F., Matthyssens, G., Kronenberger, P., Pays, E., Dero, B., Van Assel, S., Darville, M., Cravador, A., Steinert, M., and Hamers, R. (1983). Gene activation and re-expression of a Trypanosome brucei variant surface glycoprotein. EMBO J. *2,* 1185–1192.

Miller, E. N., and Turner, M. J. (1981). Analysis of antigenic types appearing in the first relapse populations of clones of Trypanosoma brucei. Parasitology *82,* 63–80.

Page, G. S., Smith, S., and Goodman, H. M. (1981). DNA sequence of the rat growth hormone gene: location of the 5′ terminus of the growth hormone mRNA and identification of an internal transposon-like element. Nucl. Acids Res. *9,* 2087–2104.

Rigby, P. W. J., Dieckman, M., Rhodes, C., and Berg, P. (1977). Labelling DNA to high specific activity in vitro by nick-translation with DNA polymerase I. J. Mol. Biol. *113,* 237–251.

Rubin, G. M., Brorein, W. J., Jr., Dunsmuir, P., Flavell, A. J., Levis, R., Strobel, E., Toole, J. J., and Young, E. (1981). Copia-like transposable elements in the Drosophila genome. Cold Spring Harbor Symp. Quant. Biol. *45,* 619–628.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain termination inhibitors. Proc. Nat. Acad. Sci. USA. *74,* 5463–5467.

Sutcliffe, J. G., Milner, R. J., Bloom, F. E., and Lerner, R. A. (1982). Common 82 nucleotide sequence unique to brain RNA. Proc. Nat. Acad. Sci. USA *79,* 4942–4946.

Turner, M. J. (1982). Biochemistry of the variant surface glycoproteins of salivarian trypanosomes. Adv. Parasitol. *21,* 69–153.

Vassin, E. F., Goldberg, G. I., Tucker, P. W., and Smithers, O. (1980). A mouse alpha-globin related pseudogene lacking intervening sequences. Nature *286,* 222–226.

Wahl, G. M., Stern, M., and Stark, G. R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridisation by using dextran sulphate. Proc. Nat. Acad. Sci. USA *76,* 3683–3687.

Walter, P., and Blobel, G. (1982). Signal recognition particle contains a 7S RNA essential for protein translocation across the endoplasmic reticulum. Nature *299,* 691–698.

Weiner, A. M., and Denison, R. A. (1983). Either gene amplification or gene conversion may maintain the homogeneity of the multigene family encoding human U1 small nuclear RNA. Cold Spring Harbor Symp. Quant. Biol. *47,* 1141–1149.

Wilde, C. D., Crowther, C. E., Cripe, T. P., Lee, M. G., and Cowan, N. J. (1982). Evidence that a human beta-tubulin pseudogene is derived from its corresponding mRNA. Nature *297,* 83–84.

Young, J. R., Shah, J. S., Matthyssens, G., and Williams, R. O. (1983). Relationship between multiple copies of a T. brucei variable surface glycoprotein gene whose expression is not controlled by duplication. Cell *32,* 1149–1159.

Young, M. W., and Schwartz, H. E. (1981). Nomadic gene families in Drosophila. Cold Spring Harbor Symp. Quant. Biol. *45,* 629–640.

Zupancic, T. J., Marvo, S. L., Chung, J. H., Peralta, E. G., and Jaskunas, S. R. (1983). RecA-independent recombination between direct repeats of IS50. Cell *33,* 629–637.