MOLECULAR AND CELLULAR BIOLOGY, Dec. 1993, p. 7766–7773
0270-7306/93/127766-08$02.00/0

Vol. 13, No. 12

A Trypanosomal CCHC-Type Zinc Finger Protein Which Binds the Conserved Universal Sequence of Kinetoplast DNA Minicircles: Isolation and Analysis of the Complete cDNA from *Crithidia fasciculata*

HAGAI ABELIOVICH, YEHUDA TZFATI, AND JOSEPH SHLOMAI*

Department of Parasitology and Department of Cellular Biochemistry, Hadassah Medical School, The Hebrew University, Jerusalem 91010, Israel

Received 11 June 1993/Returned for modification 26 July 1993/Accepted 2 September 1993

Replication of the kinetoplast DNA minicircle light strand initiates at a highly conserved 12-nucleotide sequence, termed the universal minicircle sequence. A *Crithidia fasciculata* single-stranded DNA-binding protein interacts specifically with the guanine-rich heavy strand of this origin-associated sequence (Y. Tzfati, H. Abeliovich, I. Kapeller, and J. Shlomai, Proc. Natl. Acad. Sci. USA 89:6891–6895, 1992). Using the universal minicircle sequence heavy-strand probe to screen a *C. fasciculata* cDNA expression library, we have isolated two overlapping cDNA clones encoding the trypanosomatid universal minicircle sequence-binding protein. The complete cDNA sequence defines an open reading frame encoding a 116-amino-acid polypeptide chain consisting of five repetitions of a CCHC zinc finger motif. A significant similarity is found between this universal minicircle sequence-binding protein and two other single-stranded DNA-binding proteins identified in humans and in *Leishmania major*. All three proteins bind specifically to single-stranded guanine-rich DNA ligands. Partial amino acid sequence of the endogenous protein, purified to homogeneity from *C. fasciculata*, was identical to that deduced from the cDNA nucleotide sequence. DNA-binding characteristics of the cDNA-encoded fusion protein expressed in bacteria were identical to those of the endogenous *C. fasciculata* protein. Hybridization analyses reveal that the gene encoding the minicircle origin-binding protein is nuclear and may occur in the *C. fasciculata* chromosome as a cluster of several structural genes.

Kinetoplast DNA (kDNA) is a unique extrachromosomal DNA network found in the single mitochondrion of parasitic flagellate protozoa of the family Trypanosomatidae. In *Crithidia fasciculata*, kDNA consists of about 5,000 duplex DNA minicircles (2.5 kb each) and about 50 DNA maxicircles (37 kb each) interlocked topologically to form a DNA network. Minicircles are virtually identical in size within a given kDNA network but show a species-dependent degree of sequence heterogeneity. Within a given species, minicircles contain conserved regions, which appear either as a single copy (e.g., in *Trypanosoma brucei*), two copies located 180° apart (e.g., in *C. fasciculata*), or four copies located 90° apart, as in *Trypanosoma cruzi* (reviewed in references 15, 37, 40, and 47). Overall, there are only few similarities among conserved sequences of different trypanosomatid species. Two short sequences are present, however, in the conserved regions of minicircles from almost every species thus far studied: the sequence GGGGTTGGT GTA in the heavy (H) strand and the sequence ACGCCC in the light (L) strand (the only exceptions are an AGGGTTG GTGTT found in the H strand of *Crithidia oncopelti* [34] and the sequence GAGCCC in the L strand of *Leishmania tarantolae* [23]).

The replication of kDNA networks occurs during the S phase of the cell cycle (10). On the basis of in vivo observations, Englund et al. have described the replication of kDNA minicircles (13, 14, 17) as a process in which individual minicircles are detached from the central zone of the disc-shaped network, replicated, and reattached to the periphery of the disc. The network increases in size until it doubles and then divides and segregates into two daughter networks. Extensive studies have been carried out on free minicircle replication intermediates of the trypanosomatids *Trypanosoma equiperdum* (31, 32, 38, 39, 41), *C. fasciculata* (4, 5, 13, 24, 25, 45), and *L. tarentolae* (46). These studies have suggested that synthesis of the minicircle L strand is initiated by RNA priming at a unique site and proceeds continuously. Synthesis of the H strand is discontinuous. The two daughter molecules resulting from the replication of an individual minicircle thus have distinctly different features. The daughter molecules containing a nascent L strand have a single gap of 6 to 10 nucleotides located at the newly synthesized L strand. This gap was mapped to the site of the suggested origin of L-strand replication and overlaps the conserved 12-nucleotide universal minicircle sequence (UMS) 5′-GGGGTTGGTGTA-3′ (5, 45). In contrast, minicircles containing a nascent H strand are highly gapped (4, 25). One of these gaps is located opposite the conserved hexameric sequence, which was suggested to be the replication origin site of the H strand (4, 5).

We have recently reported the specific recognition of the UMS element, conserved at the origin of replication of kDNA minicircles, by a unique UMS-specific single-stranded DNA-binding protein (UMSBP) from *C. fasciculata* (50). Here, we report on the isolation of a *C. fasciculata* cDNA encoding UMSBP. Analysis of its nucleotide sequence reveals a reading frame encoding a zinc finger protein of the CysXXCysXXXXHisXXXXCys class, highly similar to a human single-stranded DNA-binding protein specific for the sterol-responsive element (36), as well as to a *Leishmania major* single-stranded DNA-binding protein specific to

* Corresponding author. Electronic mail address: SHLOMAI@MD2.HUJI.AC.IL.

the hexameric repeats found in the 5′ region of genes encoding the major surface antigen of *Leishmania* species (52). The possible role of UMSBP in the initiation of kDNA minicircle replication is discussed.

MATERIALS AND METHODS

Nucleic acids and enzymes. kDNA and nuclear DNA were prepared from *C. fasciculata* by using centrifugation through a CsCl cushion as described previously (43). Synthetic oligonucleotides were prepared by using an Applied Biosystems oligonucleotide synthesizer at the Bletterman Laboratory of the Interdepartmental Division, Faculty of Medicine, Hebrew University of Jerusalem. Poly(dI-dC) · poly(dI-dC) and RNA molecular weight markers were purchased from Boehringer Mannheim. T4 polynucleotide kinase and restriction endonucleases were from New England Biolabs and Boehringer Mannheim. Radioactive nucleotides were from New England Nuclear.

Cells and culture conditions. *C. fasciculata* cultures were grown in brain heart infusion medium (Difco) containing hemin (20 μg/ml; Sigma), penicillin (100 U/ml), and streptomycin (100 μg/ml; Teva, Jerusalem, Israel) with agitation (200 rpm) in a rotary shaker at 28°C. Cells were harvested during logarithmic phase (1.5×10^7 cells per ml).

Screening of cDNA expression library. A λgt11 *Sfi-Not* vector (Promega) expression library was prepared, using poly(A)$^+$ RNA from *C. fasciculata* logarithmic-phase cultures. Clones (1.8×10^5) were screened by a procedure modified from those of Vinson et al. (51) and Singh (48). Briefly, 6×10^4 plaques in a 150-mm-diameter petri dish were induced and adsorbed onto isopropyl-β-D-thiogalactopyranoside (IPTG)-saturated nitrocellulose filters (Schleicher & Schuell). The filters were blocked for 30 min at 4°C in binding buffer (20 mM Tris-Cl [pH 7.5], 2 mM 2-mercaptoethanol, 0.5 mM EDTA, 5 mM $MgCl_2$) supplemented with 5% (wt/vol) nonfat dry milk (Difco). Probing was carried out for 4 to 8 h at 4°C in the presence of dodecamer UMS probe (5×10^8 cpm/μg, 1.2 ng/ml) and sheared, denatured calf thymus DNA (10 μg/ml; Sigma) in binding buffer supplemented with 0.25% (wt/vol) nonfat dry milk. Filters were washed six times in 60 ml of binding buffer as described above. Screening was carried out in duplicate.

UMSBP fusion protein expression in bacteria. λgt11 phage clone $\lambda_{1\text{-}5/6}$ was plaque purified and assayed for homogeneity by the UMS filter-binding assay. It was used to infect *Escherichia coli* Y1089 cells. Lysogens were selected (42), and two independent lysogenic clones plus a λgt11 lysogen and *E. coli* Y1089 control bacteria were induced by IPTG. β-Galactosidase induction kinetics was monitored by following the appearance of a new fusion protein band in denaturing, reducing polyacrylamide gels. Cell extracts prepared from induced cells (42) were frozen in liquid nitrogen and stored at −80°C until use. Protein was determined as described by Bradford (6).

Electrophoretic mobility shift analysis. Analyses were carried out as previously described (50) except that binding reaction mixtures were incubated at 0°C and electrophoresis was conducted at 4°C. Gels were dried and exposed to X-ray films (Agfa Curix RP2 or Kodak X-Omat AR), and quantitation of protein-DNA complexes was by densitometry of the autoradiograms, using a G300 scanning densitometer (Hoefer Scientific Instruments). Alternatively, dried gels were exposed to a type BAS-IIIs imaging plate (Fuji) and then scanned and analyzed in a model BAS1000 Bio-Imaging analyzer (Fuji).

Southern blot hybridization analysis. Southern blot hybridization of crithidial genomic and kinetoplast DNA was conducted on nitrocellulose filters as described previously (42). Blots were probed either with the entire UMSBP cDNA insert or with an *Eco*RI-*Ava*II fragment that spans the reading frame, as indicated.

Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis. Gel electrophoresis was conducted as described by Laemmli (27) in 8% acrylamide separation gels (1:30 bisacrylamide/acrylamide). Samples were electrophoresed alongside high-molecular-weight protein markers (Sigma).

DNA sequence analysis. DNA sequencing of double-stranded templates was carried out by the Sanger dideoxy method, using a Sequenase preparation (U.S. Biochemical) as instructed by the manufacturer. Sequences were further analyzed by using the University of Wisconsin Genetics Computer Group program package.

Partial sequencing of purified UMSBP. Partial sequence analysis by Edman degradation was carried out by using a fragment of UMSBP (lacking 1.1 kDa of the polypeptide N terminus) (50a), purified to apparent homogeneity from *C. fasciculata* cell extracts. A 4.5-μg sample of protein was mixed with loading buffer to final concentrations of 50 mM Tris-Cl (pH 6.85), 4% (wt/vol) SDS, 2% (vol/vol) 2-mercaptoethanol, 10% (vol/vol) glycerol, and 10 mM EDTA. The protein sample was incubated for 15 min at 37°C and loaded, alongside protein molecular weight markers (Rainbow prestained low molecular weight; Amersham) onto a polyacrylamide gel (16 by 18 by 0.075 cm; 16.5% T–3% C separating gel; 10% T–3% C spacer gel; 4% T–3% C stacking gel) as described by Schagger and Von Jagow (44). The gel was previously aged for 24 h at room temperature. Thioglycolic acid (0.1 mM) was added to the upper buffer chamber. Electrophoresis was performed as described by Schagger and Von Jagow (44), at constant 190 V and room temperature, for 16 h. The gel was washed in transfer buffer (12.5 mM Tris base, 96 mM glycine [pH 8.3], 20% [vol/vol] methanol) for 25 min and blotted onto a methanol-prewetted membrane (Problot; Applied Biosystems). Blotting was conducted in an ice-cooled electroblotter (Bio-Rad) at constant 75 V for 1 h. The membrane was washed three times for 5 min each time in distilled water, stained with Coomassie blue R-250 (0.025% [wt/vol] in 40% [vol/vol] methanol) for 5 min, and destained for 5 min in 50% (vol/vol) methanol. The UMSBP band was excised, washed three times for 5 min each time in distilled water, air dried, and kept desiccated at −20°C. Edman degradation sequencing was performed by the Bletterman Laboratory of the Interdepartmental Division, Faculty of Medicine, Hebrew University of Jerusalem.

RESULTS

Isolation of a cDNA clone encoding the *C. fasciculata* UMSBP. The specific binding of the UMS H strand by UMSBP was used as an assay in the screening of a *C. fasciculata* λgt11 expression library as described in Materials and Methods. A positive phage clone, $\lambda_{1\text{-}5/6}$, expressing a protein capable of specific UMS binding, was plaque purified for further analysis. *E. coli* Y1089($\lambda_{1\text{-}5/6}$) lysogens were found to express an IPTG-inducible β-galactosidase fusion protein (not shown). This clone contained a 1.2-kb cDNA insert (see Fig. 4) that did not contain the crithidial miniexon sequence, indicating that the $\lambda_{1\text{-}5/6}$ insert was not a complete cDNA. The original λgt11 library was rescreened by hybridization with the cDNA probes described below in order to identify an overlapping cDNA clone that does include a

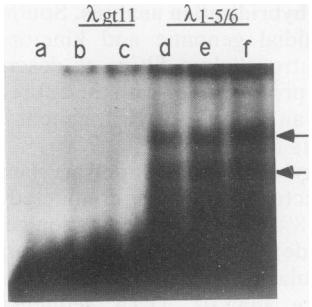

FIG. 1. Comparative UMS binding assays using $\lambda_{1\text{-}5/6}$ and λgt11 lysogen cell extracts. Bacterial extracts, prepared as described in Materials and Methods, were subjected to DNA binding reactions in the presence of 0.2 ng of ^{32}P-labeled UMS probe (10^9 cpm/μg) and then to electrophoretic mobility shift analysis as described in Materials and Methods. Lanes: a, free DNA; b and c, 1 and 2 μg, respectively, of protein of the λgt11 lysogen cell extract; d to f, $\lambda_{1\text{-}5/6}$ lysogen cell extract (0.56, 1.11, and 2.23 μg, respectively of protein). Arrows denote retarded bands that are specific for $\lambda_{1\text{-}5/6}$ lysogen extracts.

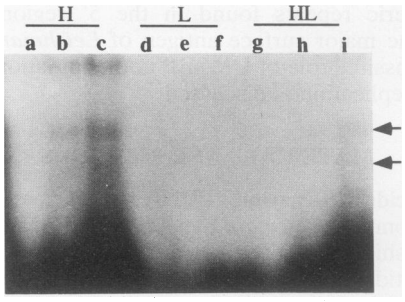

FIG. 2. Strand specificity of UMSBP fusion protein binding activity. Three micrograms of protein of $\lambda_{1\text{-}5/6}$ lysogen cell extracts, prepared as described in Materials and Methods, was incubated with the following ^{32}P-labeled UMS DNAs: 0.2, 0.6, and 2.0 ng of UMS H-strand probe (H, lanes a to c), 0.2, 0.6, and 2.0 ng of UMS L-strand probe (L, lanes d to f), or 0.4, 1.2, and 3.6 ng of double-stranded UMS probe (HL, lanes g to i). Electrophoretic mobility shift analysis was conducted as described in Materials and Methods. Arrows mark the positions of $\lambda_{1\text{-}5/6}$-specific protein-DNA complexes.

miniexon stretch. We used an *Eco*RI-*Ava*II restriction endonuclease fragment from the $\lambda_{1\text{-}5/6}$ insert, a sequence that does not include the 3′ untranslated region, as a probe in the isolation of 200 clones displaying positive hybridization signals in plaque hybridization assays of the original λgt11 cDNA library. These clones were pooled and reprobed in duplicate with the *Eco*RI-*Ava*II probe vis-à-vis a terminally labeled 21-base oligonucleotide corresponding to the 3′ terminus of the crithidial miniexon sequence (7). One such double-positive clone (N-2) was isolated and analyzed. It was shown to contain a 20-base 3′-miniexon sequence contiguous with a 5′ untranslated region that overlaps the 5′ end of the $\lambda_{1\text{-}5/6}$ insert. The identity of the clone was corroborated by restriction endonuclease analysis, sequencing of an internal segment, 3′-terminal sequencing, and Southern blot hybridization analysis (not shown).

Binding specificity of the UMSBP fusion expressed in bacteria conforms with that of *C. fasciculata* UMSBP. The specificity of DNA binding by the fusion protein, produced in $\lambda_{1\text{-}5/6}$-infected *E. coli* Y1089 lysogens, was monitored in soluble extracts by using the mobility shift electrophoresis analysis of protein-DNA complexes as an assay. As shown in Fig. 1, UMS DNA-binding activity could be detected only in extracts prepared from $\lambda_{1\text{-}5/6}$-infected lysogens, not in those derived from uninfected *E. coli* Y1089 or λgt11 lysogens lacking a cDNA insert. The additional protein-DNA complex, of a higher electrophoretic mobility, detected in this experiment (Fig. 1) probably represents the binding of a proteolytic degradation product of the fusion protein. The DNA-binding specificity displayed by the $\lambda_{1\text{-}5/6}$-encoded fusion protein is virtually identical to that of the endogenous UMSBP purified to homogeneity from *C. fasciculata* cell extracts (50). The fusion protein is a sequence-specific single-stranded DNA-binding protein, as demonstrated in the binding assays shown in Fig. 2. The bacterially expressed fusion protein bound the H-strand UMS sequence in a specific fashion. No binding was observed when the L strand was used as a probe, and only weak complex formation could be detected with the duplex (HL) probe. In binding competition experiments, a >100-fold molar excess of an unrelated oligomer could not compete with the UMS H strand for binding of the protein (Fig. 3). Furthermore, the binding activities of the cloned fusion protein and the endogenous UMSBP purified from *C. fasciculata* show the same sensitivity to mutations in the UMS sequence. Two UMS-derived mutated dodecamers were used in the binding competition analyses shown in Fig. 3. (i) UMS-mutG12, containing a transition mutation at the 3′-terminal adenine residue (position 12), has been previously shown to bind the endogenous crithidial UMSBP with an affinity identical to that of wild-type (wt) UMS and is, in fact, the only change in the UMS site that does not result in diminished binding (50). (ii) UMS-mutA10 contains a G-to-A transition mutation at position 10; the affinity of crithidial UMSBP for this oligonucleotide is significantly lower than its affinity for to wt UMS (50). As shown in Fig. 3, the relative affinities of the bacterially expressed fusion protein for wt UMS, UMS-

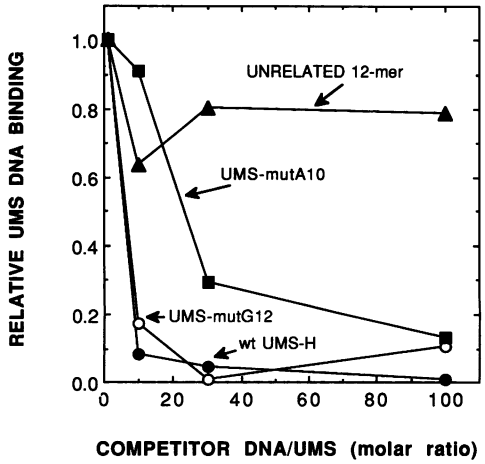

FIG. 3. Sequence specificity of the bacterially expressed UMSBP. Three micrograms of protein of $\lambda_{1\text{-}5/6}$ lysogen cell extracts, prepared as described in Materials and Methods, was incubated with 0.2 ng of UMS H-strand probe in the presence of various concentrations of the indicated constructs. The sequence of the unrelated dodecamer was 5′-AAGGGCAACCAG-3′. Electrophoretic mobility shift analysis and quantitation of protein-DNA complexes were carried out as described in Materials and Methods.

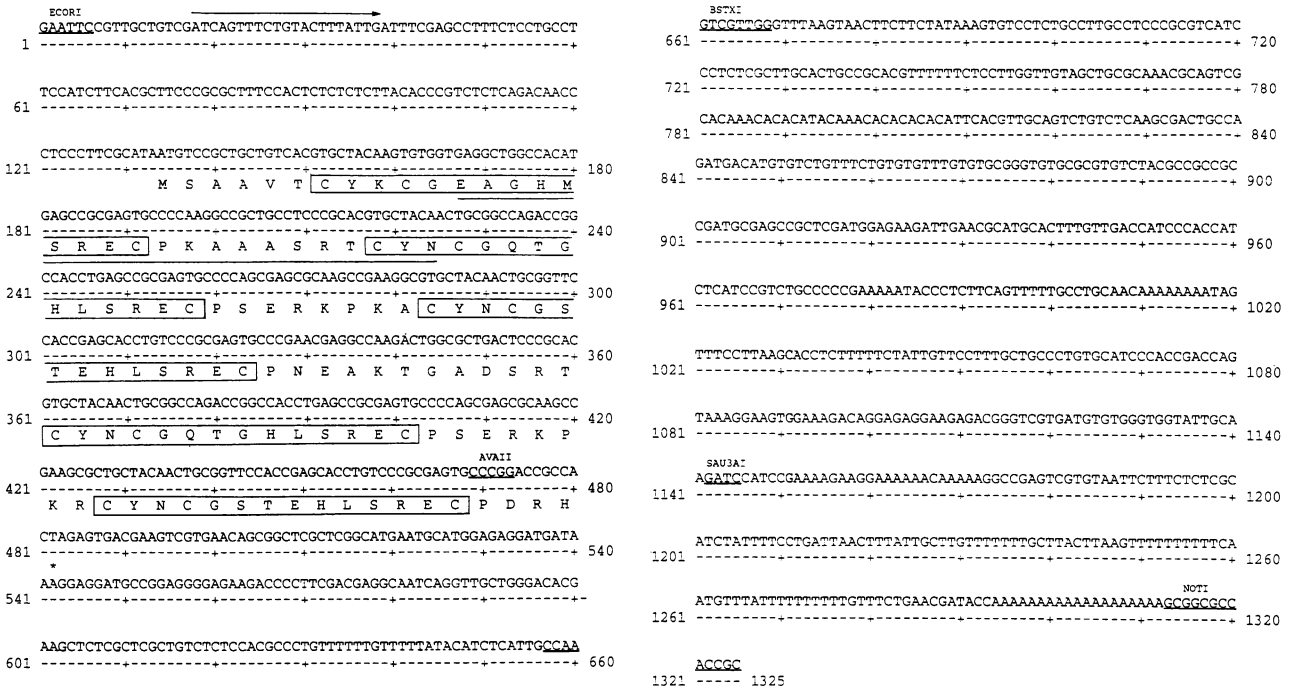

FIG. 4. Nucleotide sequence and derived amino acid sequence of *C. fasciculata* UMSBP cDNA. Nucleotides are numbered with position 1 corresponding to the first nucleotide of the *Eco*RI site of the cDNA insert. Nucleotides 1 to 17 are the synthetic *Eco*RI library linker. The arrow at the top marks the 21-bp miniexon sequence. Amino acids are numbered beginning at the initiator methionine. The asterisk represents the predicted stop codon. Zinc finger motif sequences are boxed. Underlined amino acid residues (12 to 32) have been cross-verified by Edman degradation sequencing of purified endogenous protein.

mutG12, and UMS-mutA10 are in complete concordance with those described for the endogenous UMSBP from *C. fasciculata* cell extracts. The fusion protein expressed in $\lambda_{1\text{-}5/6}$-infected *E. coli* lysogens displays similar binding affinities toward UMS-mutG12 and wt UMS and much lower affinity for UMS-mutA10 (a 10-fold molar excess of UMS-mutA10 displaces only 10% of the bound labeled wt UMS H-strand dodecamer, as opposed to 92% displacement by unlabeled UMS [Fig. 3]).

Primary sequence of full-length *C. fasciculata* UMSBP cDNA. The 1.2-kb insert found in the $\lambda_{1\text{-}5/6}$ clone was subcloned into a Bluescript pSK(−) plasmid (Stratagene), and the nucleotide sequences of both DNA strands were analyzed as described in Materials and Methods. Figure 4 shows the nucleotide sequence and the deduced amino acid sequence of the 1,295-bp full-length UMSBP cDNA, which includes the additional 96-nucleotide 5′ sequence contributed by the N-2 clone (see above). The nucleotide sequence of the λgt11 *Sfi*I-*Not*I polylinker was used to define the relevant reading frame expressed in the $\lambda_{1\text{-}5/6}$ clone. We presumed the first in-frame methionine codon downstream of the miniexon sequence to be the translation initiation site (Fig. 4). A termination codon at nucleotide 482 defines a 348-bp open reading frame encoding a putative 116-amino-acid protein with a deduced molecular mass of 12.8 kDa. The sequence also contains a 116-bp 5′ untranslated region (UTR) and a 831-bp 3′ UTR that includes a 19-nucleotide poly(A) stretch. Protein sequencing analysis by Edman degradation of 20 amino acid residues from a homogeneous fragment of endogenous UMSBP purified from *C. fasciculata* cell extracts is in complete agreement with the deduced protein sequence shown here (underlined amino acid residues 12 to 31; Fig. 4).

A striking structural characteristic of the UMSBP cDNA-encoded amino acid sequence is its potential to form five zinc finger structures (Fig. 5). The putative zinc finger motifs encompass amino acid residues 7 to 112. These C_2HC-type zinc fingers are composed of a general CysXXCysXXXX HisXXXXCys motif, typically found in retroviral Gag proteins (11). However, a much higher similarity is found between the UMSBP zinc fingers and those found in the sequences of *L. major* HEXBP (52) and the human sterol-responsive element-binding protein (designated CNBP [36]). All three proteins are DNA-binding proteins that bind specifically to G-rich, single-stranded sequences. A comparison of the crithidial UMSBP region containing the zinc fingers with those of the human CNBP and the leishmanial HEXBP is shown in Fig. 6. While a high degree of similarity is observed between these proteins' zinc finger motifs, UMSBP, HEXBP, and human CNBP show only a scant homology in the linker sequences between the zinc fingers and in the regions N terminal to the zinc finger stretch (Fig.

```
Finger2   C Y N C G Q T G H L   S R E C P S E R K P   K A . . . .
Finger4   C Y N C G Q T G H L   S R E C P S E R K P   K R . . . .
Finger3   C Y N C G S T E H L   S R E C P N E A K T   G A D S R T
Finger5   C Y N C G S T E H L   S R E C P D R H . .   . . . . . .
Finger1   C Y K C G E A G H M   S R E C P K A A A S   R T . . . .
Consensus C Y N C G - T G H L   S R E C P - E - K -   - - - - - -
```

FIG. 5. Alignment of the zinc finger repeats of the *C. fasciculata* UMSBP protein sequence. The amino acid positions of the repeats are as follows: zinc finger 1, 7 to 20; zinc finger 2, 29 to 42; zinc finger 3, 51 to 64; zinc finger 4, 77 to 90; and zinc finger 5, 99 to 112. The consensus sequence shows the characteristic features of a typical repeat unit. Shaded residues are common to at least three zinc fingers.

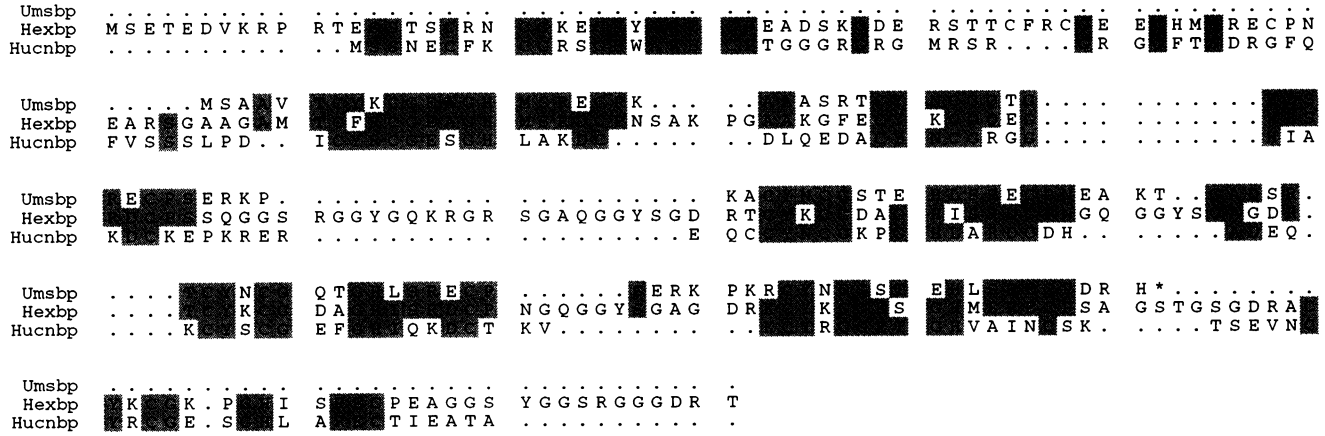

FIG. 6. Alignment and comparison of the amino acid sequences of *C. fasciculata* UMSBP, human CNBP (Hucnbp), and *L. major* HEXBP. Sequences of UMSBP, CNBP (36), and HEXBP (52) were aligned, and similarities were analyzed by using the University of Wisconsin Genetics Computer Group Prettybox program. Shaded residues are common to at least two of the proteins.

6). A unique attribute of UMSBP zinc fingers is the high degree of sequence homogeneity found between CCHC stretches in the protein, as demonstrated in Fig. 5. It is noteworthy that while zinc finger 1 is slightly divergent, fingers 2 and 4 are identical, as are fingers 3 and 5. This interdigitated structure is unique to UMSBP, and its mechanistic implications for the interactions of this protein with its binding sequence have yet to be explored.

Nuclear localization, genomic organization, and expression of the UMSBP gene in *C. fasciculata*. Southern blot hybridization analysis of *C. fasciculata* genomic and kinetoplast DNA, using an intact 1.2-kb cDNA insert (the $\lambda_{1\text{-}5/6}$ *Eco*RI-*Not*I fragment), shows hybridization of the probe with genomic DNA but not with kDNA preparations. This finding indicates a nuclear localization of the *C. fasciculata* UMSBP gene (Fig. 7). The general organization of the genomic sequence was probed by using six-cutter restriction endonucleases which do not cleave within the full-length cDNA sequence (*Eco*RI, *Bam*HI, and *Hin*dIII). A single DNA fragment, of approximately 21 kb, which hybridizes to the full-length $\lambda_{1\text{-}5/6}$ cDNA probe was detected in both *Eco*RI and *Bam*HI digests of *C. fasciculata* genomic DNA (not shown). With use of the same probe, three *Hin*dIII fragments of 7.5, 5.6, and 2.2 kb were observed, suggesting the existence of more than one copy of the UMSBP gene in the *C. fasciculata* genome (Fig. 7). Digestion of genomic DNA with *Sau*3AI, an enzyme with a single recognition site within the full-length cDNA, yielded two fragments, of 1.3 and 1.5 kb, that hybridized with the probe. More detailed analysis was performed in an effort to clarify the possible genomic organization of the UMSBP gene as suggested by these results. We compared the hybridization patterns obtained with use of the full-length *Eco*RI-*Not*I fragment versus the $\lambda_{1\text{-}5/6}$ *Eco*RI-*Ava*II fragment (which includes only the open reading frame sequence and 20 5′ UTR nucleotides; Fig. 4) as DNA probes. In these analyses, we used restriction endonucleases which cleave at unique sites within the cDNA but outside the *Eco*RI-*Ava*II fragment (e.g., *Sau*3AI and *Bst*XI). The only differences between hybridization with a complete insert compared with the shorter *Eco*RI-*Ava*II fragment is the addition of one 1.3-kb band in the *Bst*XI digest (not altered in the *Bst*XI-*Hin*dIII double digest) as well as a significant enhancement of the 1.3-kb *Sau*3AI signal. This result indicates that the three *Hin*dIII bands may represent linked loci, as otherwise one would expect three additional bands in the *Bst*XI digest when the large probe rather than the smaller one is used as well as additional bands in the *Sau*3AI digest. A plausible interpretation of these results is that the UMSBP gene occurs as a tandemly repeated cluster of at least three structural units (see below for further discussion). Partial *Hin*dIII digests of genomic DNA did not yield conclusive results (not shown), possibly because small gene clusters will give a low signal-to-noise ratio in such an analysis.

Northern (RNA) blot hybridization analyses of poly(A)$^+$ (Fig. 8) and total RNA (not shown) preparations yield identical results when probed with the UMSBP cDNA probe, using the intact *Eco*RI-*Not*I fragment. Three signals may be discerned: two major bands of approximately 3.4 and

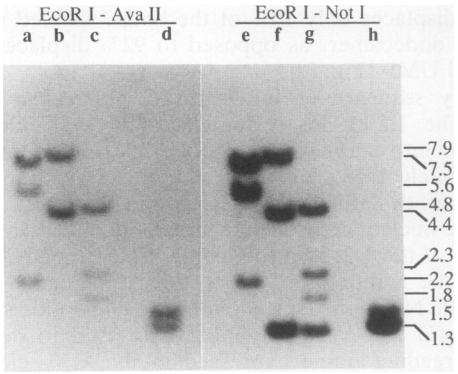

FIG. 7. Southern blot hybridization analysis of *C. fasciculata* genomic DNA. Hybridization analysis using ^{32}P-labeled fragments of the $\lambda_{1\text{-}5/6}$ insert as probes was conducted as previously described (42). Ten micrograms of *C. fasciculata* genomic DNA was electrophoresed in 0.7% agarose gels. A single DNA-blotted nitrocellulose filter was probed with the *Eco*RI-*Ava*II fragment (lanes a to d) and then washed and reprobed with the *Eco*RI-*Not*I fragment (lanes e to h). Restriction endonuclease digests of *C. fasciculata* genomic DNA were as follows: lanes a and e, *Hin*dIII digest; lanes b and f, *Bst*XI digest; lanes c and g, *Hin*dIII-*Bst*XI double digest; lanes d and h, *Sau*3AI digest. Radioactive DNA probes (~10^9 cpm/μg) were labeled by using a Multiprime labeling preparation (Amersham). All washes were with 0.1× SSC (15 mM NaCl, 1.5 mM sodium citrate) at 70°C. Values denote fragment sizes in kilobases.

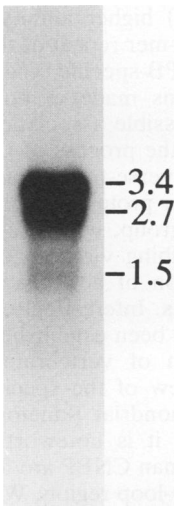

FIG. 8. Northern blot hybridization analysis of UMSBP. One microgram of *C. fasciculata* poly(A)$^+$ RNA was probed with the *Eco*RI-*Not*I fragment. Purification of poly(A)$^+$ RNA from a logarithmic 1-liter culture of *C. fasciculata* (2×10^7 cells per ml), using a CsCl cushion and oligo(dT) affinity chromatography, and Northern blot hybridization analysis were conducted as previously described (42). RNA size markers were from Boehringer Mannheim (size range, 0.3 to 7.4 kb). All washes were at high stringency (0.1× SSC, 70°C). Values represent RNA transcript sizes in kilobases.

2.7 kb, as judged by their electrophoretic mobilities, and a minor band of approximately 1.5 kb (Fig. 8). The same pattern was also observed with the shorter *Eco*RI-*Ava*II probe (not shown), indicating that 3′ UTR homologies with other transcripts do not interfere with the analysis.

DISCUSSION

Sequence-specific single-stranded DNA-binding proteins have been shown to play diverse roles in DNA metabolism: involvement in transcription regulation, recombination, and initiation of DNA replication have all been invoked. In more than one instance, a single protein was shown to partake in two or more of these activities (3, 9, 16, 28, 22, 35). We have recently described the recognition of the dodecameric nucleotide sequence GGGGTTGGTGTA (UMS), conserved at the origin of replication of kDNA minicircles, by a sequence-specific, single-stranded DNA-binding protein from the trypanosomatid protozoan *C. fasciculata*. Here, we report the isolation of two overlapping clones representing the full-length cDNA sequence of the *C. fasciculata* UMSBP. The presence of the 20-nucleotide miniexon-derived sequence at the very 5′ end of the projected cDNA sequence (Fig. 4) indicates that it coincides with the 5′ end of the mRNA molecule.

The identification of this cDNA as the sequence encoding the *C. fasciculata* UMSBP is supported by several lines of evidence. We find that the protein sequence of a fragment of the endogenous UMSBP purified from *C. faciculata* cell extracts is identical with the deduced amino acid sequence of the open reading frame encoded in the UMSBP cDNA described here over the 20-amino-acid stretch that was analyzed by Edman degradation (Fig. 4). Moreover, the DNA-binding specificity of the UMSBP–β-galactosidase fusion protein produced in *E. coli*($\lambda_{1\text{-}5/6}$) lysogens is virtually identical with that of the endogenous UMSBP purified to homogeneity from *C. fasciculata* cell extracts with respect to sensitivity to mutations in the UMS dodecamer and specificity for the H strand (Fig. 2 and 3).

The gene encoding UMSBP is nuclear, as *trans*-splicing of miniexon sequences is an attribute of trypanosomatid nuclear transcripts (29). Furthermore, using the UMSBP cDNA as a probe in Southern blot hybridization analysis, we detected a signal with genomic DNA but not with kDNA preparations. Considering the specific binding of UMSBP to a mitochondrial origin sequence, one expects, in line with findings for the majority of known nuclearly encoded, mitochondrially localized proteins, that the reading frame would encode a preprotein containing a cleavable leader sequence of 10 to 70 amino acid residues (2). As the first CCHC motif spans amino acid residues 7 to 20, a cleavable leader peptide longer than six residues would entail a sacrifice of the zinc finger. We find no evidence for the presence of a cleavable leader peptide sequence in UMSBP: the mobility of a recombinant UMSBP purified from *E. coli*, when analyzed on polyacrylamide gels under denaturing and reducing conditions, is identical to that of the endogenous UMSBP purified from *C. fasciculata* cell extracts (1). Nonetheless, the occurrence of a cleavable leader peptide is not mandatory for mitochondrial localization: several known nuclearly encoded proteins that are localized to the mitochondrial matrix lack a cleavable N-terminal leader (1a, 2, 30). Examples are also known in which the same gene codes for both a mitochondrially localized and a cytoplasmic protein, the destination being governed by differential promoter usage (2). In another example, the kinetoplast-associated DNA topoisomerase II of *C. fasciculata*, the predicted amino acid sequence does not reveal an obvious mitochondrial targeting signal. It was speculated in this case that specific internal sequences within the enzyme might be involved in directing the topoisomerase to the kinetoplast (33). We have found that extracts prepared from kinetoplasts that have been isolated from *Crithidia* cells by using nitrogen decompression and extensive purification through Percoll gradients as described by others (12a, 49) display the specific UMS-binding activity detected in *C. fasciculata* whole cell extracts (50a). However, further studies, addressed directly to the intracellular localization of the protein, are required to clarify UMSBP association with the trypanosomatid cell organelles.

As has been found with other trypanosomatid genes, the data described here support a tandem organization of several structural UMSBP genes in the crithidial chromosome. Southern blot hybridization analyses with three restriction endonucleases that do not cleave within the cDNA, *Eco*RI, *Bam*HI, and *Hin*dIII, gave two types of patterns: *Eco*RI and *Bam*HI endonuclease digests of *Crithidia* genomic DNA each yielded a unique large DNA fragment (of approximately 21 kb), possibly indicative of low sequence complexity within the genomic region carrying the UMSBP gene. Digestion with *Hin*dIII yielded three hybridizing bands. Two general interpretations can be considered; the first stipulates that the three *Hin*dIII bands represent three independent loci, and the other argues for some form of physical linkage between them, e.g., that all three reside within the single *Eco*RI fragment. A genomic cluster of three or more identical genes would be expected to yield only two bands in hybridization to the *Eco*RI-*Ava*II probe when digested with endonucleases that cleave once within the cDNA sequence and 3′ to the *Ava*II site, such as *Bst*XI and *Sau*3AI (Fig. 4). As shown in Fig. 7, this is indeed the case (lanes b and d). The appearance of three bands in the *Hin*dIII-*Bst*XI double

digest (Fig. 7, lane c) can be explained if one of the copies, or a portion of the intervening region, was inverted with respect to the others. Finally, a cluster model can accommodate the facts that when *Bst*XI digests and *Hin*dIII-*Bst*XI double digests are probed, only one additional band is observed when the probe is the intact *Eco*RI-*Not*I fragment (Fig. 7, lanes f and g) in comparison with results obtained with the short *Eco*RI-*Ava*II probe (Fig. 7, lanes b and c), and that only an enhancement of the 1.3-kb signal is observed with this probe in *Sau*3AI genomic digests (Fig. 7, lanes d and h).

From these results, we suggest that the UMSBP gene may occurs in the crithidial genome as a cluster of at least three copies. While an interpretation that incorporates restriction fragment length polymorphism cannot be ruled out completely at present, it fails to account for several of the observations described. Specifically, the occurrence of three *Hin*dIII genomic fragments that hybridize with the UMSBP probe implies, assuming an interpretation based on restriction fragment length polymorphism analysis, the presence of two UMSBP genomic loci. However, the observation of only one, high-molecular-weight genomic fragment that hybridizes with this probe in both *Eco*RI and *Bam*HI genomic digests is in support of a linkage of the three *Hin*dIII fragments.

Trypanosomatid gene clusters are known to be transcribed as polycistronic precursors that are later *trans* spliced to a short miniexon sequence in the generation of monocistronic mRNA (18, 19). Northern blot hybridization analysis of poly(A)$^+$ RNA reveals multiple transcripts that hybridize to both the intact *Eco*RI-*Not*I and the shorter *Eco*RI-*Ava*II probes (Fig. 8). Of the three bands discerned, two (at mobilities corresponding to approximately 3.5 and 2.7 kb) are too large to correspond to the UMSBP cDNA. The third, a minor band at about 1.5 kb, could represent the described cDNA, with a larger poly(A) tract. While the relationship between the larger transcripts and the 1.5-kb signal is not yet clear, an explanation in line with previous studies on transcription in trypanosomatids and with the results of our Southern blot hybridization analysis is that the locus is transcribed as a polycistronic pre-mRNA that is later trimmed by *trans* splicing. However, we cannot discount at present other possibilities, such as alternative promoter usage, alternative transcription termination sites, or the occurrence of some highly homologous transcripts for related protein products. Further analysis of the genomic cluster carrying the UMSBP gene is required to clarify its organization.

The deduced crithidial UMSBP protein sequence shows a high degree of similarity with those of human CNBP (36) and leishmanial HEXBP (52), all three being members of the CCHC-type zinc finger protein family (Fig. 6). These three nonviral CCHC motif proteins have similar cognate ligands: single-stranded, G-rich DNA molecules with recognition sites 6 to 12 nucleotides long. Such sequences have been found in regions involved in transcription regulation (12, 36), replication (3), recombination (9, 12, 26), and telomere management (21). This conservation of structure and biochemical function across evolutionary boundaries may underlie some general aspect of protein-DNA interactions in eukaryotes. The cross-reactivities of all the related CCHC-type zinc finger proteins for their cognate ligands has yet to be studied in detail. We have compared the binding affinities of the recombinant *C. fasciculata* UMSBP toward its UMS binding site and the *L. major* HEXBP ligand. Binding competition analyses reveal (1) that UMSBP exhibits a significantly (≥8-fold) higher affinity for the UMS dodecamer than for the 12-mer repeat of the GGCGAG hexamer reported as the HEXPB-specific binding sequence (52).

Several observations made in both viral and nonviral systems suggest a possible association of this type of zinc finger structure with the process of DNA replication initiation. A survey of reverse-transcribing viruses possessing CCHC-type zinc finger proteins (11) reveals that a common denominator of this group, which distinguishes them from other reverse-transcribing viruses not encoding such proteins, is the utilization of a tRNA primer in viral DNA minus-strand synthesis. Interestingly, tRNA and tRNA-like DNA sequences have been established as crucial *cis*-acting elements in initiation of vertebrate mitochondrial DNA replication (8). In view of the specificity of the crithidial UMSBP to a mitochondrial (kinetoplast) DNA origin of replication sequence, it is noteworthy that five potential binding motifs for human CNBP are found within the vertebrate mitochondrial D-loop region. While tRNA genes have not been found in trypanosomatid kDNA, other, as yet unidentified elements may play an analogous role in the minicircle molecule. A more detailed biochemical analysis of cell-free, soluble kDNA replication reactions will be required in order to further establish these possible analogies between kDNA minicircle replication and conventional mitochondrial replication systems. *C. fasciculata* UMSBP, being a minicircle origin-binding protein, will then lend itself to more function-related biochemical studies.

ACKNOWLEDGMENTS

This study was supported in part by grant 89-00190/1 from the United States-Israel Binational Science Foundation and grant DISNAT 132/GR 337 from the Israel Ministry of Science and Technology and the Society for Biotechnology and Research Braunschweig, Germany.

REFERENCES

1. **Abeliovich, H., and J. Shlomai.** Unpublished data.
1a. **Adrian, G. S., M. T. McCammon, D. L. Montgomery, and M. G. Douglas.** 1986. Sequences required for delivery and localization of the ADP/ATP translocator of the mitochondrial inner membrane. Mol. Cell. Biol. **6**:626–634.
2. **Attardi, G., and G. Schatz.** 1988. Biogenesis of mitochondria. Annu. Rev. Cell Biol. **4**:289–333.
3. **Bergmann, A. D., and E. M. Johnson.** 1992. The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. Mol. Cell. Biol. **12**:1257–1265.
4. **Birkenmeyer, L., and D. S. Ray.** 1986. Replication of kinetoplast DNA in isolated kinetoplasts from *C. fasciculata*. J. Biol. Chem. **261**:2362–2368.
5. **Birkenmeyer, L., H. Sugisaki, and D. S. Ray.** 1987. Structural characterization of site specific discontinuities associated with replication origins of minicircle DNA from *C. fasciculata*. J. Biol. Chem. **262**:2384–2392.
6. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye interaction. Anal. Biochem. **72**:248–254.
7. **Bruzik, J. P., K. Van Doren, D. Hirsch, and J. A. Steitz.** 1988. Trans-splicing involves a novel form of small nuclear riboprotein particles. Nature (London) **335**:559–562.
8. **Clayton, D.** 1991. Replication and transcription of vertebrate mitochondrial DNA. Annu. Rev. Cell Biol. **7**:453–478.
9. **Collick, A., M. G. Dunn, and A. J. Jeffreys.** 1991. Minisatellite binding protein MSBP-1 is a sequence specific single stranded DNA binding protein. Nucleic Acids Res. **19**:6399–6404.
10. **Cosgrove, W. B., and M. J. Skeen.** 1970. The cell cycle in *Crithidia fasciculata*. Temporal relationships between synthesis

of deoxyribonucleic acid in the nucleus and in the kinetoplast. J. Protozool. **17**:172–177.
11. **Covey, S. N.** 1986. Amino acid sequence homologies in the gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucleic Acids Res. **14**:623–633.
12. **Edelmann, W., B. Kroger, M. Goller, and I. Horak.** 1989. A recombinational hotspot in the LTR of a mouse retrotransposon identified in an in vitro system. Cell **57**:937–946.
12a. **Englund, P. T.** Personal communication.
13. **Englund, P. T.** 1979. Free minicircles of kinetoplast DNA in *Crithidia fasciculata*. J. Biol. Chem. **254**:4895–4900.
14. **Englund, P. T.** 1987. The replication of kinetoplast DNA networks in *Crithidia fasciculata*. Cell **14**:157–168.
15. **Englund, P. T., S. L. Hudjuk, and S. Marini.** 1982. The molecular biology of trypanosomes. Annu. Rev. Biochem. **1**:695–726.
16. **Erdile, L. F., W.-D. Heyer, R. Kolodner, and T. J. Kelly.** 1991. Characterization of a cDNA encoding the 70 kDa single stranded DNA binding subunit of human replication protein A and the role of the protein in DNA replication. J. Biol. Chem. **266**:12090–12098.
17. **Ferguson, M., A. F. Torri, D. C. Ward, and P. T. Englund.** 1992. In situ hybridization to the Crithidia fasciculata kinetoplast reveals two antipodal sites involved in kinetoplast DNA replication. Cell **70**:621–629.
18. **Gibson, W. C., B. W. Swinkels, and P. Borst.** 1988. Post transcriptional control of the differential expression of phosphoglycerate kinase genes in *Trypanosoma brucei*. J. Mol. Biol. **201**:315–325.
19. **Gonzales, A., T. J. Lerner, M. Huecas, B. Sosa-pineda, N. Nogueira, and P. Lizard.** 1985. Apparent generation of a segmented mRNA from two separate tandem gene families in *Trypanosoma cruzi*. Nucleic Acids Res. **13**:5789–5804.
20. **Gribscov, M., and R. R. Burgess.** 1986. Sigma factors from *E. coli*, *B. subtilis*, phage SP01 and phage T4 are homologous proteins. Nucleic Acids Res. **14**:6745–6763.
21. **Gualberto, A., R. M. Patrick, and K. Walsh.** 1992. Nucleic acid specificity of a vertebrate telomere binding protein: evidence for G-G base pair recognition at the core binding site. Genes Dev. **6**:815–824.
22. **Hoffman, J. F.-X., and S. M. Gasser.** 1991. Identification and purification of a protein that binds the ARS consensus sequence. Cell **64**:951–960.
23. **Kidane, G. Z., D. Hughes, and L. Simpson.** 1984. Sequence heterogeneity and anomalous electrophoretic mobility of kinetoplast minicircle DNA from *Leishmania tarantolae*. Gene **27**:265–277.
24. **Kitchin, P. A., V. A. Klein, and P. T. Englund.** 1985. Intermediates in the replication of kinetoplast DNA minicircles J. Biol. Chem. **260**:3844–3851.
25. **Kitchin, P. A., V. A. Klein, B. I. Fein, and P. T. Englund.** 1984. Gapped minicircles: a novel replication intermediate of kinetoplast DNA. J. Biol. Chem. **259**:15532–15539.
26. **Krowczynska, M. M., R. A. Rudders, and T. G. Krontiris.** 1990. The human minisatellite consensus at breakpoints of oncogene translocations. Nucleic Acids Res. **18**:1121–1127.
27. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
28. **Meyers, M. L., K. M. Keating, W. J. Roberts, K. R. Williams, J. W. Chase, and M. S. Horwitz.** 1990. Purification and functional characterization of adenovirus ts111A DNA binding protein. J. Biol. Chem. **265**:5875–5882.
29. **Muhich, M. L., D. E. Hughes, A. M. Simpson, and L. Simpson.** 1987. The monogenetic kinetoplastid protozoan, *Crithidia fasciculata*, contains a transcriptionally active, multi-copy miniexon sequence. Nucleic Acids Res. **15**:3141–3153.
30. **Murakami, K., S. Tanase, Y. Morino, and M. Mori.** 1992. Presequence binding factor dependent and independent import of proteins into mitochondria. J. Biol. Chem. **267**:13119–13122.
31. **Ntambi, J. M., and P. T. Englund.** 1985. A gap at a unique location in newly replicated kinetoplast DNA minicircles from *Trypanosoma equiperdium*. J. Biol. Chem. **260**:5574–5579.
32. **Ntambi, J. M., T. A. Shapiro, K. A. Ryan, and P. T. Englund.** 1986. Ribonucleotides associated with a gap in newly replicated kinetoplast DNA minicircles from *Trypanosoma equiperdium*. J. Biol. Chem. **261**:11890–11895.
33. **Pasion, S. G., J. C. Hines, R. Aebersold, and D. S. Ray.** 1992. Molecular cloning and expression of the gene encoding the kinetoplast-associated type II DNA topoisomerase of *Crithidia fasciculata*. Mol. Biochem. Parasitol. **50**:57–68.
34. **Pestov, G., L. A. Gladkaya, D. A. Maslov, and A. A. Kolesnikov.** 1990. Characterization of kinetoplast DNA in the lower trypanosomatid *Crithidia oncopelti*. Mol. Biochem. Parasitol. **41**:135–146.
35. **Radding, C.** 1982. Homologous pairing and strand exchange in genetic recombination. Annu. Rev. Genet. **16**:405–437.
36. **Rajavashisth, T. B., A. K. Taylor, A. Andalibi, K. L. Svenson, and A. J. Lusis.** 1989. Identification of a zinc finger protein that binds to the sterol regulatory element. Science **245**:640–643.
37. **Ray, D. S.** 1987. Kinetoplast DNA minicircles: high copy number mitochondrial plasmids. Plasmid **17**:177–190.
38. **Ryan, K. A., and P. T. Englund.** 1989. Replication of kinetoplast DNA in *Trypanosoma equiperdium*. J. Biol. Chem. **264**:823–830.
39. **Ryan, K. A., and P. T. Englund.** 1989. Synthesis and processing of kinetoplast DNA minicircles in *Trypanosoma equiperdium*. Mol. Cell. Biol. **9**:3212–3217.
40. **Ryan, K. A., A. T. Shapiro, C. A. Rauch, and P. T. Englund.** 1988. The replication of kinetoplast DNA in trypanosomes. Annu. Rev. Microbiol. **42**:334–358.
41. **Ryan, K. A., A. T. Shapiro, C. A. Rauch, J. D. Griffith, and P. T. Englund.** 1988. A knotted free minicircle in kinetoplast DNA. Proc. Natl. Acad. Sci. USA **85**:5844–5848.
42. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Saucier, J. M., J. Bernard, J. Da Silva, and G. Riou.** 1981. Occurrence of a kinetoplast DNA-protein complex in *Trypanosoma cruzi*. Biochem. Biophys. Res. Commun. **101**:988–994.
44. **Schagger, H., and G. Von Jagow.** 1987. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal. Biochem. **166**:368–379.
45. **Sheline, C., T. Melendy, and D. S. Ray.** 1989. Replication of kDNA minicircles isolated from *Crithidia fasciculata*: structure of nascent minicircles. Mol. Cell. Biol. **9**:169–176.
46. **Sheline, C., and D. S. Ray.** 1989. Specific discontinuities in *Leishmania tarantolae* minicircles map within universally conserved sequence blocks. Mol. Biochem. Parasitol. **37**:151–158.
47. **Simpson, L.** 1987. The mitochondrial genome of kinetoplastid protozoa: genetic organization, transcription, replication and evolution. Annu. Rev. Microbiol. **41**:363–382.
48. **Singh, H., J. M. LeBowitz, A. S. Baldwin, and P. A. Sharp.** 1988. Molecular cloning of an enhancer binding protein: isolation by screening of an expression library with a recognition site DNA. Cell **52**:415–423.
49. **Torri, A. F., and P. T. Englund.** 1992. Purification of a mitochondrial DNA polymerase from *Crithidia fasciculata*. J. Biol. Chem. **267**:4786–4792.
50. **Tzfati, Y., H. Abeliovich, I. Kapeller, and J. Shlomai.** 1992. A single stranded DNA binding protein from Crithidia fasciculata recognizes the nucleotide sequence at the origin of replication of kinetoplast DNA minicircles. Proc. Natl. Acad. Sci. USA **89**:6891–6895.
50a. **Tzfati, Y., H. Abeliovich, and J. Shlomai.** Unpublished data.
51. **Vinson, C. R., K. L. LaMarco, P. F. Johnson, W. H. Landschulz, and S. L. McKnight.** 1988. In situ detection of sequence specific DNA binding activity specified by a recombinant bacteriophage. Genes Dev. **2**:801–806.
52. **Webb, J. R., and W. R. McMaster.** 1993. Molecular cloning and expression of a *Leishmania major* gene encoding a single stranded DNA-binding protein containing nine "CCHC" zinc finger motifs. J. Biol. Chem. **268**:13994–14002.