Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 5534–5538, August 1986
Developmental Biology

Isolation of cDNA clones for differentially expressed genes of the human parasite *Schistosoma mansoni*

(cDNA expression library/λgt11/recombinant DNA/fusion proteins)

ALAN H. DAVIS*†, RONALD BLANTON*, FRITZ ROTTMAN†, RUSSELL MAURER†, AND ADEL MAHMOUD*†

*Division of Geographic Medicine, Department of Medicine, and †Department of Molecular Biology and Microbiology, Case Western Reserve University and University Hospitals, Cleveland, OH 44106

Communicated by Oscar D. Ratnoff, April 11, 1986

ABSTRACT **Little is known about the mechanisms that control transformations during the life cycle of *Schistosoma mansoni*. To enable isolation of DNA sequences encoding developmentally regulated antigens a cDNA expression library in the vector λgt11 amp3 was constructed from adult mRNA and immunologically screened with sera from infected individuals. We report here on the properties of three recombinant clones that derive from developmentally regulated genes. Clone 10-3 encoded a β-galactosidase fusion protein present in high abundance in infected *Escherichia coli*. Clones 7-2 and 8-2 also produced immunologically recognized proteins; however, the peptides did not appear to be β-galactosidase fusion proteins. The expression of mRNAs hybridizing to these cDNAs was examined in the different stages of the parasite life cycle. Messenger RNA corresponding to clone 10-3, ≈1000 bases in length, was present in higher abundance in male worms than in females but was not detected in schistosome eggs. A 900-base mRNA hybridizing to clone 7-2 was observed in adult worms and eggs. Both clone 10-3 and clone 7-2 hybridized to smaller mRNAs in cercariae and freshly transformed schistosomula than in adult worms. Clone 8-2 contained tandem cDNA inserts. One cDNA hybridized to a 1700-base mRNA present in all stages, while the second hybridized to an 800-base mRNA specific to adult female worms.**

Parasitic helminths are multicellular organisms that have developed unique biological features necessary for their specialized forms of life. Among these are their ability to infect more than one host, evade the host protective responses, and undergo remarkable morphologic and biochemical changes during their life cycle (1). *Schistosoma mansoni* is one of several blood flukes that infect humans and lead to considerable morbidity (2). The schistosomes are the only mammalian trematodes that are sexually dimorphic. Male and female *S. mansoni* worms parasitize the portal veins and survive in the blood stream for 5–10 years (3). Adult worms mate and produce eggs that pass to the outside environment with the stool. In fresh water, ova hatch, and the emerging miracidia infect specific intermediate snail hosts where they multiply asexually. After 4–6 weeks snails shed cercariae, which penetrate mammalian skin and immediately transform to intermediate larval forms, the schistosomula. These migrate through defined regions of the body until they reach maturity in the portal system.

The molecular mechanisms that underlie the complex developmental changes and expression of stage- and gender-specific strains in *S. mansoni* are reflected in the differential expression of several proteins (4, 5). We approached the investigation of schistosome biology by identifying antigens differentially expressed during development and sexual maturation of the parasite. An expression cDNA library for *S. mansoni* adults has been constructed in bacteriophage λgt11 amp 3 (6) and was screened using sera from infected humans. Twenty-one recombinants have been isolated; three were chosen for further study. One cDNA clone (10-3) corresponded to mRNA found in cercariae, schistosomula, and adult worms and was present in higher abundance in males than females. Egg RNA did not hybridize with this clone. A second clone (7-2) hybridized to mRNA in all stages examined. Both clones (10-3 and 7-2) hybridized to smaller messages in cercariae and schistosomula than in adult worms. A third recombinant (8-2) contained cDNAs arising from two different mRNAs, one of which was found only in adult female worms.

MATERIALS AND METHODS

Enzymes and Reagents. Restriction enzymes were purchased from Bethesda Research Laboratories. T4 DNA polymerase and *Eco*RI methylase were obtained from New England Biolabs. Alkaline phosphatase from calf intestine, T4 DNA ligase, and *E. coli* DNA polymerase I were purchased from Boehringer Mannheim. Reverse transcriptase was purchased from Life Sciences (St. Petersburg, FL) and RNase H was from Enzo Biochemicals (New York). Oligo(dT)-cellulose (type 3) and *Eco*RI linkers were obtained from Collaborative Research (Lexington, MA) and $(dT)_{12-18}$ from P-L Biochemicals. Nitrocellulose filters were purchased from Schleicher & Schuell. [α-^{32}P]dCTP (800 Ci/mmol; 1 Ci = 37 GBq) was obtained from New England Nuclear and [γ-^{32}P]ATP (3000 Ci/mmol) from Amersham.

Human sera used in this study were obtained from individuals infected with only one species of schistosomes by selecting individuals in the appropriate endemic area and then screening for infection with other species. Sera were obtained from *S. mansoni*-infected individuals in Egypt and Brazil, from *Schistosoma haematobium*-infected individuals in Kenya, and from *Schistosoma japonicum*-infected subjects in the Philippines. Goat anti-human immunoglobulin antiserum conjugated to horseradish peroxidase was purchased from Cappel Laboratories (Malvern, PA), as was rabbit anti-β-galactosidase. An IgG fraction was purified on a protein A column and generously provided by Charles McTiernan and Randy Huff. Protein A-horseradish peroxidase conjugate was purchased from Bio-Rad.

Organisms. Adult *S. mansoni* worms of a Puerto Rican strain were obtained by perfusion of hepatic portal systems of infected mice (7). Male and female worms were separated with fine forceps under a dissecting microscope. Cercariae were shed from infected *Biomphalaria glabrata* and concentrated at 2000 × *g*. Schistosomula were prepared by mechanical transformation of cercariae, and the tails were removed by Percoll gradient centrifugation (8). The organisms were

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

Abbreviations: bp, base pair(s); X-Gal, 5-bromo-4-chloro-3-indolyl β-D-galactoside; IPTG, isopropyl β-D-thiogalactoside.

then incubated at 37°C in Earle's lactalbumen in a 5% CO_2 atmosphere for 3 hr and were concentrated by centrifugation at 500 × *g*. Eggs were obtained from intestines of mice infected for 8 wk as described (9). All organisms were frozen in a dry ice/methanol bath and stored at −80°C until use.

Construction of *S. mansoni* cDNA Expression Library. Bacteriophage λgt11 amp3 was isolated from lysogens of bacterial strain W3110 (10) grown on superbroth (11) and thermally induced. Phage were purified on CsCl gradients, and DNA was isolated as described (12). The DNA was heated to 50°C, the cos ends were allowed to anneal at room temperature for 20 min in 66 mM Tris·HCl, pH 7.5/10 mM $MgCl_2$/5 mM dithiothreitol and were ligated at 14°C overnight with T4 DNA ligase. The DNA was precipitated with ethanol and suspended in TE buffer (10 mM Tris·HCl, pH 8.0/1 mM EDTA). Phage DNA was digested with *Eco*RI, ethanol-precipitated, and suspended in TE buffer. The DNA was then treated with calf intestinal alkaline phosphatase at 37°C for 45 min followed by phenol/chloroform and chloroform extractions and ethanol precipitation. The DNA was suspended in TE buffer. This sequence produced unit-length molecules with phosphatase-cleaved *Eco*RI ends.

Adult *S. mansoni* poly(A)$^+$ mRNA was purified from guanidinium-hot-phenol-extracted RNA by selection on oligo(dT)-cellulose (12). Six micrograms of poly(A)$^+$ RNA was copied into double-stranded cDNA using reverse transcriptase for first-strand synthesis and RNase H and DNA polymerase I for second-strand synthesis (13). The cDNA was treated with *Eco*RI methylase to protect internal *Eco*RI sites, made blunt-ended with T4 DNA polymerase, and was then ligated with kinased *Eco*RI linkers. Digestion with *Eco*RI was performed and the released linker fragments removed by precipitation of higher molecular weight DNA with ethanol and isopropanol. The cDNA was suspended in the TE buffer described earlier, and ≈200 ng was ligated with ≈2 μg of λgt11 amp3 DNA/25 μl using T4 DNA ligase. Ligation products were packaged *in vitro* (14). The size of the library and fraction of recombinant phage were assayed by titration of uncolored plaques on bacterial strain Y1090 in the presence of X-Gal (5-bromo-4-chloro-3-indolyl β-D-galactoside) and IPTG (isopropyl β-D-thiogalactoside) (15). In addition, randomly primed, ^{32}P-labeled first-strand cDNA (12) was hybridized to a known number of plaques lifted onto nitrocellulose. The library was amplified by growth on LB agar-containing plates (12) at 37°C using strain LE392 (16).

Antibody Screening of the cDNA Library. A pool consisting of two sera from *S. mansoni*-infected Brazilian individuals and one serum from an infected Egyptian was used. The choice of individual sera was based on the generation of numerous, discrete bands in gels of immunoprecipitated products of *S. mansoni* mRNA translated *in vitro* (data not shown). Human sera were diluted 1:10 in 20 mM Tris·HCl, pH 7.5/0.5 M NaCl containing 20% fetal calf serum. Goat anti-human immunoglobulin antiserum coupled to horseradish peroxidase was diluted 1:1000. All antisera were preabsorbed with lysates of *E. coli* bound to nitrocellulose (15). Lysates were made from strain Y1090 and lysogens of λgt11 amp3 in strain BTA282 (17). One hundred thousand phage were plated on LB agar in 150 mm-diameter Petri plates and allowed to produce plaques at 42°C. Proteins produced by the phage were screened as previously described (15). The bound immunoglobulin was visualized with horseradish peroxidase-coupled second antibody. Those plaques containing bacteriophage that generated positive signals on two replica filters were further purified and screened until pure lysates giving 100% positive plaques were obtained.

DNA Probe Preparation. Recombinant phage were prepared by thermal induction of lysogens in strain W3110. DNA was isolated and digested by *Eco*RI before gel electrophoresis on preparative 4.5 or 5% acrylamide gels (acrylamide/bisacrylamide, 30:1) (12) and staining with ethidium bromide. Gel slices containing the insert DNA were removed; the DNA was isolated by electroelution within dialysis membranes, concentrated by ethanol precipitation, and suspended in TE buffer.

RNA and Southern Blot Hybridization. *S. mansoni* poly(A)$^+$ mRNA was electrophoresed through agarose gels containing formaldehyde and then transferred to nitrocellulose (18). Some gels were made with preparative sample wells and the nitrocellulose was sliced into 0.5-cm-wide strips after transfer. *S. mansoni* DNA was isolated, digested with *Eco*RI or *Hin*dIII, electrophoresed, and transferred to nitrocellulose as previously described (19). Blots were probed with nick-translated (20), ^{32}P-labeled DNA prepared from the cDNA clones as indicated above. Specific activities of the DNA probes were approximately 1–3 × 10^8 cpm/μg of DNA. When blots were probed with clone 8-2, each of the three *Eco*RI fragments was labeled individually. Blots were hybridized for 12–24 hr at 65°C in 5× SSC (1× SSC = 0.15 M NaCl/0.015 M sodium citrate), 10× Denhardt's solution (1× Denhardt's solution = 0.02% Ficoll/0.02% polyvinylpyrrolidone/0.02% bovine serum albumin), 30 mM sodium phosphate (pH 7.0), 0.2% $NaDodSO_4$ containing sonicated, denatured salmon sperm DNA at 0.5 mg/ml. Blots were washed twice at room temperature in 3× SSC/30 mM sodium phosphate, pH 7.0/10× Denhardt's solution/0.5% $NaDodSO_4$ and three times for 20 min at 50°C in 0.2× SSC/0.1% $NaDodSO_4$. Some blots were washed in 0.1× SSC/0.1% $NaDodSO_4$ at 65°C. Autoradiographic exposures were made at −70°C using Kodak XAR5 film and intensifying screens. Before sequential probing with other clones, blots were washed in 0.1× SSC/0.1% $NaDodSO_4$ at 85°C for 30 min and autoradiographic exposures indicated no radioactivity.

Immunochemical Studies. *E. coli* strain BTA282 was lysogenized with selected phage. The lysogens were grown to mid exponential phase, induced at 42°C for 15 min, and incubated at 37°C for 2 hr. One-milliliter aliquots of cells were pelleted and lysed in 80 μl of $NaDodSO_4$ electrophoresis sample buffer (0.125 M Tris·HCl, pH 6.8/4% $NaDodSO_4$/20% glycerol/10% mercaptoethanol). Twenty-microliter aliquots were electrophoresed on gels containing $NaDodSO_4$ and were blotted to nitrocellulose (21). The antigens were identified with immune sera and peroxidase-conjugated protein A (diluted 1:1000). The blocking reagent in immunological blots was non-fat dry milk (22).

RESULTS

Construction of Schistosome cDNA Expression Library. From 6 μg of poly(A)$^+$ mRNA, approximately 200 ng of double-stranded cDNA was prepared. The average size of the cDNA molecules was approximately 1100 base pairs (bp) as determined by analysis of parallel, ^{32}P-labeled first- and second-strand reactions on alkaline agarose gels. The size of the resulting library before amplification was determined by two methods that distinguished recombinants from parental vector. The number of recombinant bacteriophage was assayed by the number of uncolored plaques in the presence of X-Gal and IPTG. In addition, a known number of plaques was lifted onto nitrocellulose and hybridized with randomly primed, ^{32}P-labeled first-strand cDNA. Both methods indicated that the library was at least 30% recombinant bacteriophage and 10^6 recombinants in size. Approximately 10^5 recombinants from an amplified portion of the library were immunologically screened with sera from infected humans. Twenty-one recombinants were selected for isolation when reproducibly positive signals occurred during each round of plaque purification. Three clones, 10-3, 7-2, and 8-2 were chosen for further characterization. The sizes of the

cDNA inserts were ≈330 bp for clone 10-3 and 580 bp for clone 7-2. Clone 8-2 contained three *Eco*RI fragments in the cDNA insert sized 1300 bp, 280 bp and 250 bp.

Proteins from induced lysogens of recombinant phage in *E. coli* strain BTA282 were examined by immunological blotting. *S. mansoni* antigens and β-galactosidase epitopes were visualized respectively with pooled immune human sera or anti-β-galactosidase antiserum (Fig. 1). Clone 10-3 encoded a β-galactosidase–*S. mansoni* fusion protein of approximately 130 kDa. Clone 7-2 produced an antigen approximately 21 kDa in size while clone 8-2 produced two antigens of approximately 83 kDa and 72 kDa. These were not visualized with the anti-β-galactosidase antiserum, although a number of peptides smaller than 116 kDa did react.

We then examined the species specificity of the proteins encoded by these clones. Recombinants were arrayed on a lawn of *E. coli* strain Y1090 for immunological comparisons. With pooled sera from *S. mansoni*-infected individuals, clone 10-3 gave a strong signal followed in intensity by 7-2 and 8-2 (data not shown). Normal human serum (from Cleveland donors) did not react. The three clones were also screened with pooled sera from individuals known to be infected only with *S. haematobium* or *S. japonicum*. None of the clones produced proteins recognized by these sera. Species specificity was also evaluated by RNA blot analyses using poly(A)$^+$ RNA that had been isolated from other platyhelminths, including *S. japonicum, Fasciola hepatica*, and *Taenia pisiformis*. Only *S. mansoni* RNA hybridized with each of the three cDNAs under the chosen experimental conditions.

Stage and Sex Comparisons of mRNA Expression. The expression of mRNAs corresponding to each of the three cDNA clones was examined by RNA blot hybridization. Data shown in Fig. 2 represent the same lanes from one RNA blot hybridized sequentially with each cDNA probe. Messenger RNA hybridizing to the insert from clone 10-3 was seen in RNA isolated from adult worms, separated male and female worms, and in cercariae. No mRNA hybridizing to this cDNA was seen in eggs obtained from mice following 8 wk of infection. Differences in mRNA sizes and quantity were noted with clone 10-3. In adult worms a 1000-base mRNA class appeared in greater abundance in male worms than in females, while in cercariae a 750-base mRNA was seen. All mRNAs were abundant, requiring only overnight autoradiographic exposures for the generation of intense signals. The most prominent hybridization, however, occurred to the 1000-base mRNA class in adult male worms.

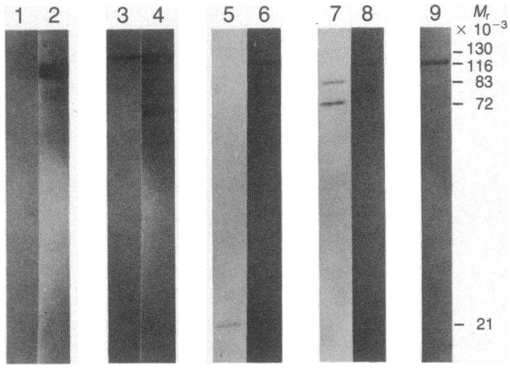


FIG. 1. Expression of *S. mansoni* antigens in *E. coli*. Immunological blots of proteins produced by *E. coli* strain BTA282 harboring recombinant clones were prepared. Lanes 1–8 contained proteins encoded by isolated clones and treated with sera from humans infected with *S. mansoni* (lanes 1, 3, 5, 7) or anti-β-galactosidase antibodies (lanes 2, 4, 6, 8). Lanes: 1 and 2, proteins produced from bacteria containing λgt11 amp 3; 3 and 4, clone 10-3; 5 and 6, clone 7-2; 7 and 8, clone 8-2; 9, molecular weight standards including β-galactosidase that were treated with anti-β-galactosidase antibodies.

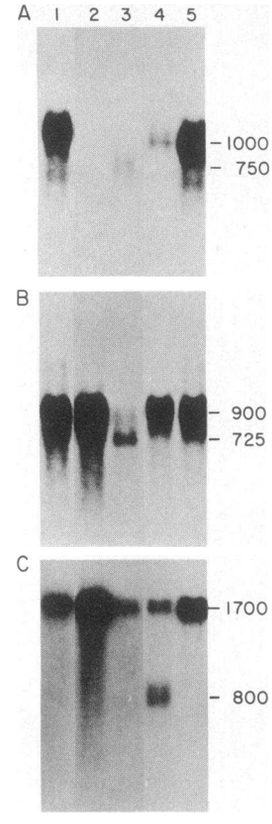


FIG. 2. Hybridization of cDNAs to *S. mansoni* mRNA. An RNA blot was sequentially probed with the inserts from the three cDNAs: *A*, clone 10-3; *B*, clone 7-2; *C*, clone 8-2. For each panel one autoradiographic exposure was chosen, and lanes were rearranged for ease of comparison. Each lane was loaded with 3 μg of poly(A)$^+$ RNA prepared from 1, adult worms; 2, eggs; 3, cercariae; 4, female worms; and 5, male worms. Sizes of the mRNAs (numbers in bases) were estimated using the ribosomal and tRNA bands in total HeLa cell RNA electrophoresed in adjacent lanes and stained with ethidium bromide. Longer autoradiographic exposures of *C*, lane 1, show a band at 800 bases.

The insert from clone 7-2 hybridized to abundant mRNA in all stages of the parasite examined: adults, eggs, cercariae, males and females. In adults and eggs, a broad band centered at ≈900 bases was seen, while RNA isolated from cercariae hybridized in two discrete bands—sized ≈900 and 725 bases. Superimposing the autoradiograms of 10-3 and 7-2 showed that the mRNAs did not comigrate, and Southern blot analyses indicated that these two cDNA clones hybridized to different genomic sequences. The hybridization patterns of clones 10-3 and 7-2 were unchanged when blots were washed under more stringent conditions.

Three *Eco*RI fragments were seen in the insert DNA in cDNA clone 8-2. To determine whether the clone arose from multiple mRNAs, poly(A)$^+$ RNA from adult worms was probed with each of the three *Eco*RI fragments (Fig. 3). The 1300-bp and 250-bp fragments both hybridized to a band approximately 1700 bases in length, while the 280-bp insert hybridized to mRNA ≈800 bases long. Thus, clone 8-2 contained two independent cDNAs. Hybridization patterns of RNA isolated from other stages were then examined using a mixture of all three labeled *Eco*RI fragments. The 1700-base mRNA was observed in adults of both sexes and in 8-wk eggs and cercariae, while the smaller mRNA was specific to adult female worms (Fig. 2).

mRNA Expression in Schistosomula. To determine the sizes of the corresponding mRNAs in freshly transformed

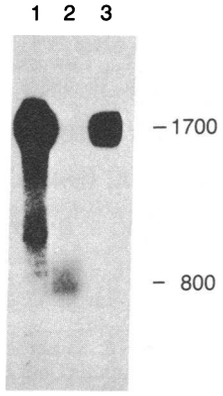


FIG. 3. Clone 8-2 contains two cDNAs. Adjacent strips of a nitrocellulose filter prepared from a preparative gel of adult poly(A)$^+$ RNA were probed with the following: Lanes 1, 1300-bp; 2, 280-bp; and 3, 250-bp *Eco*RI inserts from clone 8-2.

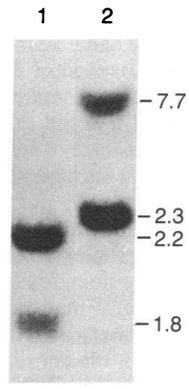


FIG. 5. Southern blot of *S. mansoni* DNA. Ten micrograms of *S. mansoni* DNA was digested with *Eco*RI (lane 1) or *Hin*dIII (lane 2), electrophoresed, transferred to nitrocellulose and hybridized with the 10-3 cDNA insert. Numbers refer to kilobase sizes as estimated using λ DNA digested with *Hin*dIII and φX174 DNA digested with *Hae* III as size standards. DNA isolated from separated male or female worms showed identical patterns.

schistosomula we used RNA blot analyses using RNA isolated from schistosomula that had been incubated in culture for 3 hr. Both clone 10-3 and clone 7-2 hybridized to smaller mRNAs in schistosomula compared with adult worms (Fig. 4). The sizes of the schistosomula mRNAs were similar to those in cercariae. In addition, the small, female-specific mRNA hybridizing with clone 8-2 was absent in schistosomula RNA. Thus, for these cDNAs, 3-hr-old schistosomula were similar to cercariae and did not appear to contain mRNAs characteristic of adult worms.

Southern Blot Hybridization. The hybridization of cDNA clones 10-3 and 7-2 to two sizes of mRNA might reflect use of multiple homologous genes. Southern blot analysis of *S. mansoni* DNA with clone 10-3 showed hybridization to only two fragments after *Eco*RI or *Hin*dIII digestion (Fig. 5). The inserts from clones 7-2 and 8-2 also hybridized to only two or three fragments of *S. mansoni* DNA after similar restriction digests (data not shown). It is therefore unlikely that any of the corresponding genes are part of multigene families.

DISCUSSION

The data presented here describe three cDNA clones encoding immunologically recognized polypeptides that correspond to genes differentially expressed during the life cycle of *S. mansoni*. Furthermore, two of these clones hybridized to different mRNA sizes in cercariae and schistosomula as compared to adult worms. Differential expression also has been described for other multicellular invertebrates such as collagen gene expression during the life cycle of *Caenorhabditis elegans* (23). The mechanisms regulating differential gene activity, particularly in multicellular parasites, are, however, unknown.

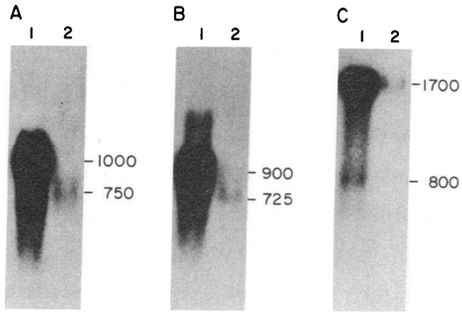


FIG. 4. Schistosomula RNA hybridization. One microgram of adult poly(A)$^+$ mRNA (lane 1) and approximately 25 μg of total schistosomula RNA (lane 2) were subjected to RNA blot analysis. The probes were *A*, clone 10-3; *B*, clone 7-2; and *C*, clone 8-2.

The schistosomes, like many other parasites, undergo developmental changes necessary for successful existence, not only in more than one host, but also in the external environment. At the protein level, several biochemical and immunologic determinants of schistosome differentiation have been described. In one study, five proteins specific to male *S. mansoni* and four for female worms were demonstrated (4). Furthermore, five proteins have been observed in adult worms but not in schistosomula (4). Immunologically, differentiation of parasite stages has been demonstrated using monoclonal antibodies developed against stage-specific proteins (24–27). Recently, immunoprecipitation of *in vitro* translation products of *S. mansoni* mRNA using polyclonal antisera suggested that there are several adult worm- and schistosomula-specific polypeptides (28–32). Previous reports concerning cloning of *S. mansoni* antigens (33, 34) did not describe RNA analyses which would have allowed comparisons to the cDNAs described in the present study.

We have approached the study of schistosome development by constructing a cDNA expression library and screening it with sera from infected humans. We then examined mRNA expression corresponding to isolated clones. After screening 10% of our library (10^5 recombinants), twenty-one clones were isolated, and three of the clones were selected for further characterization. Clone 10-3 hybridized most strongly to an adult male mRNA class of 1000 bases; less hybridization was observed in female worms, and none in eggs. In contrast, clone 7-2 hybridized to a 900-base mRNA class in eggs and adults. We also examined hybridization patterns in cercariae and schistosomula. Cercariae, the infective stage, are fresh water, free-living organisms that transform into the obligate parasitic stages, schistosomula and adults. Our results show that clones 10-3 and 7-2 hybridized to smaller mRNAs in cercariae and 3-hr schistosomula. Thus, although the free-living cercariae and the adult obligate parasite exhibit different mRNAs corresponding to clones 10-3 and 7-2, these differences were not correlated with the early events of schistosomula transformation.

The observed differences may reflect different sites of RNA initiation or termination, or they may result from differential processing of similar primary transcripts in different stages. A third alternative is that cDNA clones 10-3 and 7-2 correspond to genes that are part of multimember families. This is unlikely since the genomic Southern blot hybridization patterns were simple. Moreover, we have

isolated genomic clones corresponding to clone 10-3. All recombinants contained the same restriction fragments seen in *S. mansoni* DNA. The absence of *Eco*RI or *Hin*dIII sites in the cDNA clone and the presence of these sites in the genomic sequences suggest that the cDNA clone might span an intron. In addition, preliminary studies using the reported *S. mansoni* genome size (35) show that the gene corresponding to 10-3 is present in one copy per haploid genome.

Adult schistosomes are unusual platyhelminths in that they have separate sexes. Nevertheless, the two sexes are developmentally interdependent, since females in unisexual infection fail to develop to sexual maturity (36). Despite gross morphological differences between the sexes, only a few specific proteins have been observed (4). In our investigations a cDNA fragment of clone 8-2 hybridized to a female specific mRNA of 800 bases. Furthermore, we have demonstrated by Southern blot analyses of DNA prepared from isolated male and female worms that the corresponding gene was not sex-linked. The expression of this mRNA, therefore, must have resulted from differential gene activity.

Clone 10-3 encoded a fusion protein approximately 15 kDa larger than native β-galactosidase. This was approximately 4 kDa larger than expected based upon the size of the cDNA insert and the average size of an amino acid (110 daltons). The cDNA may encode a peptide with an unusual amino acid composition or one that causes anomalous migration of the fusion protein on gels. Clones 7-2 and 8-2 encoded antigens that did not appear to be β-galactosidase fusion proteins. Similar phenomena have been reported for other antigens cloned in λgt11 (37, 38) and may involve use of fortuitous signals in DNA sequences flanking the *Eco*RI site or contained within the cDNA. Since the manner in which bacteria synthesize these antigens is unknown, estimates of the expected sizes were not possible. Alternatively, fusion proteins encoded by clones 7-2 and 8-2 might be subject to proteolysis, and released peptides might not have reacted with the anti-β-galactosidase antiserum. The presence of peptides smaller than the 10-3 fusion protein and native β-galactosidase visualized with the anti-β-galactosidase antiserum (Fig. 1) suggested that proteolysis did occur.

Differentially expressed schistosome cDNAs provide essential probes to characterize in greater detail the mechanisms of schistosome morphogenesis.

We thank Earlene Moss and Pierre Peters for maintenance of the *S. mansoni* life cycle; Paul Klich, Sonji Gause, and Cathy DeHaven for technical help; Charles McTiernan and Randy Huff for the anti-β-galactosidase antiserum; and Pat Amato for typing the manuscript. This work was supported by the John D. and Catherine T. MacArthur Foundation and the Edna McConnell Clark Foundation.

1. Ellner, J. J. & Mahmoud, A. A. F. (1982) *Rev. Infect. Dis.* **4**, 698–714.
2. Mahmoud, A. A. F. (1985) in *Cecil Textbook of Medicine*, eds. Wyngaarden, J. B. & Smith, L. H., Jr. (Saunders, Philadelphia), 17th Ed., pp. 1809–1815.
3. Warren, K. S., Mahmoud, A. A. F., Cummings, P., Murphy, D. J. & Houser, H. B. (1974) *Am. J. Trop. Med. Hyg.* **23**, 900–902.
4. Atkinson, B. G. & Atkinson, K. H. (1982) *Exp. Parasitol.* **53**, 26–38.
5. Aronstein, W. S. & Strand, M. (1984) *J. Parasitol.* **70**, 545–557.
6. Kemp, D. J., Coppel, R. L., Cowman, A. F., Saint, R. B., Brown, G. V. & Anders, R. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3787–3791.
7. Duvall, R. H. & DeWitt, W. B. (1967) *Am. J. Trop. Med. Hyg.* **16**, 483–486.
8. Lazdins, J. K., Stein, M. J., David, J. R. & Sher, A. (1982) *Exp. Parasitol.* **53**, 39–44.
9. Boros, D. L. & Warren, K. S. (1971) *Nature (London)* **229**, 200–201.
10. Cox, E. & Yanofsky, C. (1969) *J. Bacteriol.* **100**, 390–397.
11. Botstein, D., Lew, K. K., Javik, V. & Swanson, C. A., Jr. (1975) *J. Mol. Biol.* **91**, 439–462.
12. Maniatis, T., Fritsch, E. F & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
13. Gubler, U. & Hoffman, B. M. (1983) *Gene* **25**, 263–269.
14. Hohn, B. (1979) *Methods Enzymol.* **68**, 299–309.
15. Young, R. A., Bloom, B. A., Grosskinsky, C. M., Ivanyi, J., Thomas, D. & Davis, R. W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2583–2587.
16. Murray, N. E., Brammer, W. J. & Murray, K. (1977) *Mol. Gen. Genet.* **150**, 53–61.
17. Bowtell, D. D. L., Saint, R. B., Rickard, M. D. & Mitchell, G. F. (1984) *Mol. Biochem. Parasitol.* **13**, 173–185.
18. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5206.
19. Davis, A. H., Blanton, R. & Klich, P. (1985) *Mol. Biochem. Parasitol.* **17**, 289–298.
20. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
21. Burnette, W. N. (1981) *Anal. Biochem.* **112**, 195–203.
22. Johnson, D. A., Gautsch, J. W., Sportsman, J. R. & Elder, J. H. (1984) *Gene Anal. Tech.* **1**, 3–8.
23. Cox, G. N. & Hirsh, D. (1985) *Mol. Cell. Biol.* **5**, 363–372.
24. Dissous, C. & Capron, A. (1983) *FEBS Lett.* **162**, 356–359.
25. Strand, M., McMillan, A. & Pan, X.-Q. (1982) *Exp. Parasitol.* **54**, 145–156.
26. Norden, A. P., Aronstein, W. S. & Strand, M. (1982) *Exp. Parasitol.* **54**, 432–442.
27. Harn, D. A., Mitsuyama, M. & David, J. R. (1984) *J. Exp. Med.* **159**, 1371–1387.
28. Bugra, K., Tanaka, R. D., Boyle, W. J. & MacInnis, A. J. (1983) *J. Parasitol.* **69**, 486–490.
29. Taylor, D. W., Cordingley, J. S. & Butterworth, A. E. (1983) *Mol. Biochem. Parasitol.* **10**, 305–318.
30. Tenniswood, M. P. R. & Simpson, A. J. G. (1984) *Parasitology* **84**, 253–261.
31. Knight, M., Simpson, A. J. G., Payares, G., Chaudri, M. & Smithers, S. R. (1984) *EMBO J.* **3**, 213–219.
32. Grausz, D., Dissous, C., Capron, A. & Roskam, W. (1983) *Mol. Biochem. Parasitol.* **7**, 293–301.
33. Cordingley, J. S., Taylor, D. W., Dunne, D. W. & Butterworth, A. E. (1983) *Gene* **20**, 25–29.
34. Lanar, D. E., Pearce, E. J. & Sher, A. (1985) *Mol. Biochem. Parasitol.* **17**, 45–60.
35. Simpson, A. J. G., Sher, A. & McCutchan, T. F. (1982) *Mol. Biochem. Parasitol.* **6**, 125–127.
36. Coles, G. C. (1984) *Parasitology* **89**, 603–637.
37. Dame, J. B., Williams, J. L., McCutchan, T. F., Weber, J. L., Wirtz, R. A., Hockmeyer, W. T., Maloy, W. L., Haynes, J. D., Schneider, I., Roberts, D., Sanders, C. S., Reddy, E. P., Diggs, C. L. & Miller, L. H. (1984) *Science* **225**, 593–599.
38. Goro, T. & Wang, J. C. (1984) *Cell* **36**, 1073–1080.

MOLECULAR AND CELLULAR BIOLOGY, Nov. 1988, p. 4745–4755
0270-7306/88/114745-11$02.00/0

Vol. 8, No. 11

Tandemly Repeated Exons Encode 81-Base Repeats in Multiple, Developmentally Regulated *Schistosoma mansoni* Transcripts

RICHARD E. DAVIS,[1]* ALAN H. DAVIS,[1,2] SIMON M. CARROLL,[1] ALEKSANDAR RAJKOVIC,[1] AND FRITZ M. ROTTMAN[1]

Department of Molecular Biology and Microbiology[1] and Division of Geographic Medicine,[2] Department of Medicine, School of Medicine and University Hospitals, Case Western Reserve University, Cleveland, Ohio 44106

Received 3 June 1988/Accepted 2 August 1988

The adult *Schistosoma mansoni* cDNA clone 10-3 encodes an antigen that is recognized by sera from infected humans. We characterized multiple developmentally regulated transcripts homologous to the 10-3 cDNA and portions of the complex genomic loci encoding those transcripts. Transcripts of approximately 950, 870, and 780 nucleotides were expressed in adults, whereas only the 780-nucleotide transcript was observed in the larval stage. These transcripts were highly similar, containing variable numbers of identical direct tandem repeats of 81 bases. Although the sequence of the repeating elements and sequences 3′ to them were identical in all the transcripts, sequences 5′ of the repeating elements exhibited variations, including a 27-base insertion, alternative start sites for transcription, and alternate 5′ exon usage. These transcripts appeared to be derived in part by the developmentally controlled alternative splicing of small exons and the use of alternative transcription initiation sites from the one or two complex loci of at least 40 kilobase pairs. Each 81-base repeat in the transcripts was encoded by three dispersed 27-base-pair exons. These 27-base-pair exons were contained within highly conserved, reiterated 3-kilobase-pair genomic tandem arrays.

Schistosoma mansoni is a blood fluke which infects humans. Considerable morbidity and mortality result from the infection of an estimated 200 million people worldwide by three species of schistosomes. These parasites undergo a complex life cycle that is initiated by the release of eggs from the human host; these eggs hatch on contact with fresh water and release free-swimming larval forms called miracidia. The miracidia penetrate certain species of snails and reproduce asexually to form the next infective larval form, cercariae, which are shed from the snails. On contact with the human host, cercariae penetrate the skin, transform to schistosomula, undergo a complex migration, and develop into sexually mature adults which produce eggs, thus completing the life cycle.

We have previously described the isolation of an adult *S. mansoni* cDNA clone, designated 10-3, that encodes an antigen that is recognized by sera from infected humans (10). Expression of transcripts homologous to 10-3 is developmentally regulated, and different-sized transcripts are expressed in the adult and cercarial stages. Little is known concerning gene structure and regulation in schistosomes. The present report describes our initial analysis of 10-3 gene structure, 10-3 transcripts, and the developmentally controlled alternative RNA processing from a 10-3 locus. We determined the sequence of several adult and cercarial cDNAs derived from 10-3 transcripts. These cDNAs contained variable numbers of identical 81-base-pair (bp) direct tandem repeats. Although sequences 3′ of these repeating elements were the same in all the cDNAs analyzed, 10-3 transcripts exhibited differences in sequence composition 5′ of the repeating elements. We suggest that there are one or two complex 10-3 loci in *S. mansoni* which are transcribed to yield multiple mRNAs that are developmentally regulated. These mRNAs are derived in part by alternative processing and different transcriptional initiation sites. The functions of these transcripts and their protein products are not known. Analysis of the genomic regions encoding the repeated elements in the schistosome transcripts revealed an unusual exon and intron organization in which three 27-bp exons that make up the 81-base repeat are interspersed within large, highly conserved 3-kilobase-pair (kb) genomic regions. These 3-kb regions are present as conserved tandem arrays.

* Corresponding author.

MATERIALS AND METHODS

Organisms. Adult *S. mansoni* of a Puerto Rican strain were obtained by perfusion of hepatic portal systems of infected mice (11). Cercariae were obtained after they were shed from infected snails (*Biomphalaria glabrata*). Schistosomula were prepared by mechanical transformation of cercariae, and the tails were removed by Percoll gradient centrifugation (23). Schistosomula were incubated in vitro as described previously (10). Additional frozen cercariae of the NMRI strain of *S. mansoni* were obtained from the Biomedical Research Center, Rockville, Md.

Construction of cDNA and genomic libraries. Adult and cercarial cDNAs were synthesized as described previously (24). The cDNA was methylated with *Eco*RI methylase, *Eco*RI linkers were ligated to the cDNA, and the cDNA was ligated to *Eco*RI-digested and phosphatase-treated lambda gt11 arms. The products of the ligation were packaged by using packaging extracts (Gigapack; Stratagene Cloning Systems, San Diego, Calif.). The adult cDNA was size fractionated into two pools for packaging: one that was less than 500 bp and another that was greater than 500 bp. The library containing cDNAs of greater than 500 bp was used in these studies. Recombinants were obtained at a frequency of $2.3 \times 10^6/\mu g$ of DNA for the cercarial library (1.2×10^6 total recombinants), $9.6 \times 10^5/\mu g$ of DNA for the >500-bp adult library (5.0×10^5 total recombinants), and $3.6 \times 10^6/\mu g$ of DNA for the <500-bp adult library (1.8×10^6 total recombinants). Five hundred thousand recombinants from each library were amplified (24).

An adult genomic DNA library was constructed from adult DNA that was partially digested with *Sau*3A. The partially

digested DNA was size fractionated, ligated into the *Bam*HI site of EMBL-3 (Stratagene Cloning Systems), and packaged into phage with packaging extracts (Gigapack); the library was then amplified. Recombinants were obtained at a frequency of $2 \times 10^6/\mu g$ of DNA, with a total of 2×10^6 recombinants.

Isolation of cDNA and genomic clones. Adult cDNA libraries were screened with the 10-3 clone (10) that was radiolabeled by nick translation (36) and with a 55-base 5'-end-labeled (24) synthetic oligonucleotide (oligonucleotide G; see Fig. 2) complementary to a repeating element in adult cDNAs. Cercarial cDNAs were isolated by screening with 5'-end-labeled oligonucleotide G. Adult genomic clones were isolated by screening the adult EMBL-3 library with clone 10-3 that was radiolabeled by nick translation (36). Hybridization conditions for the nick-translated probe were at 65°C in 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–10× Denhardt solution–30 mM sodium phosphate (pH 7.0)–0.2% sodium dodecyl sulfate–0.5 mg of sonicated, denatured salmon sperm DNA per ml. Blots were washed in 0.2× SSC–0.1% sodium dodecyl sulfate at 50°C. The hybridization conditions for screening with the 55-mer oligonucleotide included the same buffer plus 1% sodium dodecyl sulfate at 60°C, with washing in 2× SSC–0.1% sodium dodecyl sulfate at 55°C. Lambda phage DNA was isolated by CsCl equilibrium centrifugation (24). Genomic lambda phage restriction maps were obtained by cohesive lambda DNA termini oligonucleotide mapping (35) and standard mapping of plasmid subclones.

Subcloning and sequencing. cDNA inserts and restriction fragments of genomic clones were subcloned by standard methods into pBR322 or pBS(+) vector (Stratagene Cloning Systems). Subclones were sequenced by the dideoxynucleotide method on alkali-denatured supercoiled plasmids as described previously (16, 38) and by the chemical degradation method of Maxam and Gilbert (6, 25). Sequences produced by chemical degradation were resolved on 100-cm-long water-thermostatted gels, as described previously (41), allowing sequences to be read to 650 bases. cDNA fragments were sequenced in both orientations to ensure accuracy. Sequences were compiled and analyzed by using Microgenie software (Beckman Instruments, Inc., Fullerton, Calif.).

Nucleic acid preparation. Genomic DNA was isolated by using CsCl gradients essentially as described previously (40) or by dissolution of frozen worms, which were previously powdered on dry ice in 4 M guanidinium isothiocyanate, followed by phenol-chloroform extraction and ethanol precipitation. Adult and cercarial poly(A)$^+$ mRNAs were purified from guanidinium–hot phenol-extracted RNA and selection on oligo(dT)-cellulose (24).

Northern and Southern blot analyses. Poly(A)$^+$ mRNA was fractionated on formaldehyde-agarose gels and transferred to nitrocellulose in 20× SSC or to Zeta Probe (Bio-Rad Laboratories, Richmond, Calif.) in 25 mM $NaHPO_4$ (pH 6.5). In some experiments, the poly(A)$^+$ tails of mRNAs were removed by treatment with oligo(dT) and RNase H (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) prior to electrophoresis (30, 43). Molecular size markers included procaryotic and eucaryotic rRNAs and specific radiolabeled transcripts with known sizes (ranging from 700 to 1,100 bases) that were derived by T3 or T7 polymerase transcription of specifically truncated templates.

Genomic DNA was digested with restriction endonucleases, fractionated on agarose gels, and transferred to nitrocellulose in 20× SSC or Gene Screen (Dupont, NEN Research Products, Boston, Mass.) in 10× SSC. Following transfer of DNA to Gene Screen, the DNA was cross-linked to the nylon membrane with UV light.

Hybridization conditions for blots to nitrocellulose included the buffer described above for screening libraries. The Gene Screen hybridization buffer has been described by Church and Gilbert (7). The amount of $NaHPO_4$ was reduced in some experiments, to increase the stringency of hybridization. Hybridization conditions for oligonucleotides and restriction fragments were 15°C below the midpoint melting temperature (T_m), with the T_m determined from the equations described by Meinkoth and Wahl (26). Blots were washed at the same stringency in 2× SSC–0.1% sodium dodecyl sulfate. Hybridization probes were prepared by nick translation (36) or random primer labeling (13) of duplex DNA and end labeling of oligonucleotides (24). In some cases, blots were stripped in 0.1× SSC–0.1% sodium dodecyl sulfate at 90 to 100°C, autoradiographed to confirm the quantitative removal of probe, and rehybridized with different probes.

Primer extension analysis. Primer extension analysis was performed by annealing end-labeled oligonucleotides to 1 to 2 μg of poly(A)$^+$ mRNA and extending the primer with reverse transcriptase at 50°C in 16 mM Tris hydrochloride (pH 8.3)–8.4 mM $MgCl_2$–80 mM KCl–4.2 mM dithiothreitol–0.4 mM deoxynucleotides. Extended products were fractionated on denaturing polyacrylamide gels, and the gels were autoradiographed. Sequencing of mRNAs by primer extension was performed either by the dideoxynucleotide termination method (38) or by the chemical degradation of specific extension products isolated from gels, as described by Maxam and Gilbert (25). Both methods required 7 to 8 μg of poly(A)$^+$ mRNA. For chain-termination primer extension sequencing, deoxynucleotide triphosphate concentrations were 210 μM and dideoxynucleotide concentrations were 160 μM.

RESULTS

Sequence analysis of 10-3 cDNAs. We previously reported the isolation of a 313-bp adult cDNA clone, clone 10-3, from adult *S. mansoni* (10). Initial Northern blot hybridization of adult and cercarial mRNAs with the 10-3 cDNA used as a probe indicated the presence of two adult transcripts and a single cercarial transcript. Subsequent analysis, however, with poly(A)$^+$ mRNA treated with oligo(dT) and RNase H to remove poly(A)$^+$ tails resolved the adult transcripts into three sizes of approximately 950, 870, and 780 bases (Fig. 1). No increase in the number of cercarial transcripts was observed; the size of the cercarial transcript was approximately 780 bases.

To determine the coding capability and to characterize the differences among the adult and the cercarial mRNAs, we screened the adult and cercarial cDNA libraries for longer cDNAs homologous to the clone 10-3 cDNA. A total of 110 positive adult cDNA clones were observed in 160,000 recombinants, and 31 positive cercarial cDNA clones were observed in 50,000 recombinants. Six adult and three cercarial clones were chosen for further analysis (Fig. 2B). Sequence analysis of the original 10-3 cDNA and the nine additional cDNAs homologous to this clone indicated that these transcripts all contained variable numbers of exact, direct tandem repeats of 81 bases (Fig. 2A). The most 3' repeat element in all the cDNAs was truncated to 54 bp (Fig. 2A). The repeats were defined on the basis of their exon structure (see below) and the reiterated sequence observed in the cDNAs. All sequences 3' of the repeating elements in

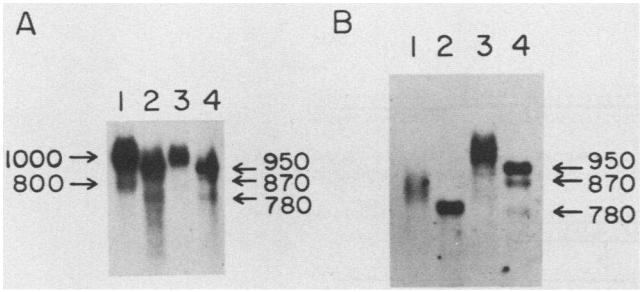


FIG. 1. Resolution of clone 10-3 mRNAs treated with oligo(dT)-RNase H prior to RNA blot analysis. Northern blots of adult and cercarial mRNAs untreated or treated with oligo(dT)-RNase H were hybridized with clone 10-3 that was radiolabeled by nick translation or an end-labeled oligonucleotide corresponding to the complement of a portion of the repeat (oligonucleotide G; see Fig. 2). All lanes contained 1 μg of poly(A)$^+$ mRNA. (A) Northern blots of adult mRNA hybridized with nick-translated clone 10-3 (lanes 1 and 2) or end-labeled oligonucleotide G (lanes 3 and 4). Lanes 1 and 3, Untreated adult mRNA; lanes 2 and 4, oligo(dT)-RNase H-treated adult mRNA. (B) Northern blot of adult and cercarial mRNAs hybridized with end-labeled oligonucleotide G. The increased resolution of these transcripts is due to separation on 1.7% agarose–formaldehyde gels that were electrophoresed for 16 h. Lane 1, Cercarial mRNA; lane 2, cercarial mRNA treated with oligo(dT)-RNase H; lane 3, adult mRNA; lane 4, adult mRNA treated with oligo(dT)-RNase H. Numbers are in bases.

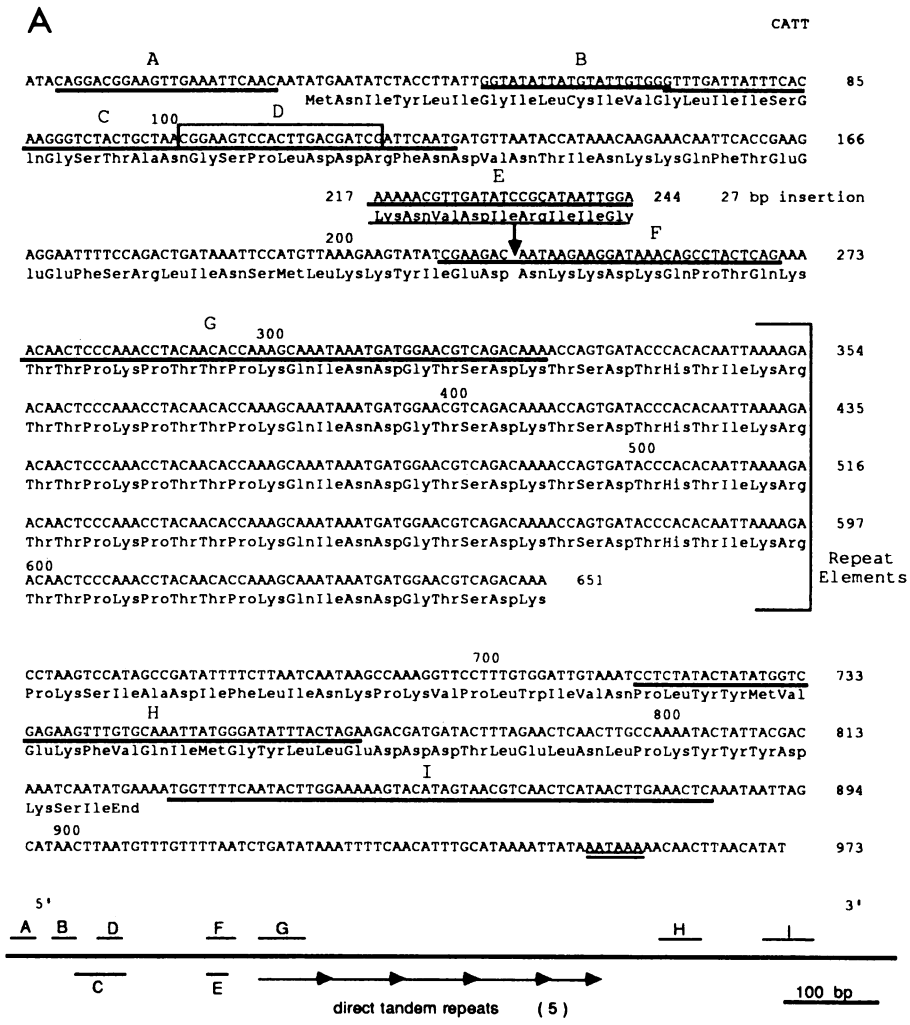
the cDNAs were identical, including the same apparent poly(A) addition site. All the transcripts contained 67% A and U. One adult cDNA (cDNA 2-4) contained four 81-bp repeats. Two adult cDNAs (cDNAs 30-2 and 29-1) contained five repeats, while two adult cDNAs (cDNAs 39-1 and 21-1) contained five repeats with a 27-bp insertion just 5′ of the repeats at bp 217 (Fig. 2). The three cercarial cDNAs had four repeat elements and did not contain the 27-bp insertion. The cercarial cDNA sequences were identical to the adult sequence, except for one cDNA (cDNA C-20), which differed by 2 bp substitutions 5′ of the repeats. The general organization and sequence of adult and cercarial cDNA clones were identical, except for variations in the number of repeat elements, the insertion of 27 bases 5′ of the repeat elements, and the extent to which the cDNAs extended toward the 5′ terminus of the transcript (Fig. 2B). Therefore, from the cDNA analysis, three highly similar transcripts containing 81-base repeats were found to be present in adults and cercaria: transcripts with four repeats, five repeats, and five repeats with a 27-base insertion.

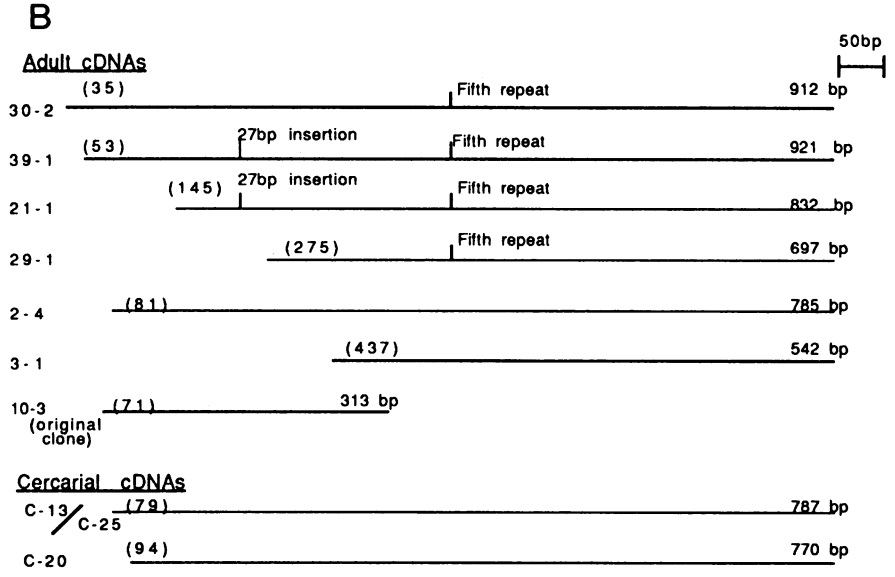
Characterization of 10-3 transcripts by primer extension and Northern blot hybridization. In order to correlate the cDNA clones with the transcripts observed on Northern blots, synthetic oligonucleotides complementary to a number of regions in the cDNAs were constructed (Fig. 2A) and used for Northern blot and primer extension analyses to further characterize adult and cercarial transcripts. Northern blot size estimates of the clone 10-3 cercarial transcript [with poly(A)$^+$ tails removed by RNase H-oligo(dT)] indicated a transcript of 780 bases, suggesting that the cercarial cDNAs were nearly full-length. It seemed likely that the longer adult cDNAs with five repeat elements (cDNAs 30-2 and 39-1) corresponded to the 950-base transcript observed on Northern blots. The transcript with only four repeat elements (cDNA 2-4) might then correspond to the 870-base transcript that was observed on Northern blots. Although all the 10-3 cDNAs had highly similar sequences, the extent of cDNA synthesis to the transcript 5′ terminus was variable (Fig. 2B). To characterize the 5′ ends and the sizes of the adult transcripts, we used primer extension analyses (Fig. 3). The dominant primer extension product observed with several primers (oligonucleotides B, D, E, and F; Fig. 2A) with adult mRNA was 34, 52, and 80 bases longer than the larger adult cDNA clones 30-2, 39-1, and 2-4, respectively (Fig. 3). The 5′ ends of these transcripts, as determined by primer extension analysis, were heterogeneous over 5 bases.

Nucleotide sequencing of these dominant primer extension products with two separate primers (oligonucleotides B and D) on adult mRNA demonstrated a sequence overlap with the 30-2 and 39-1 cDNA sequences. This indicates that the 5′ ends of cDNAs 30-2 and 39-1 were 34 and 52 bp, respectively, short of the 5′ ends of the transcripts. Nucleotides 1 to 34 constituted the primer extension sequence shown in Fig. 2A. This primer extension sequence was confirmed by sequence analysis of genomic clones corresponding to the 5′ end of the gene. These genomic clones contained the first two exons constituting bases 1 to 88 and 89 to 110; both exons exhibited standard splice junction sequences, and the exon sequence was identical to the cDNA sequence presented in Fig. 2A. This sequence analysis also indicated that the first 2 bases of the transcript, which were not discernible by primer extension sequencing, are CA. Appropriately positioned TATA and CAAT promoter elements were upstream of the transcription initiation site.

Nucleotides 8 to 30, which were determined by primer extension analysis, hybridized only to the adult 950- and 870-base transcripts but not to either the adult or the cercarial 780-base transcripts (Fig. 4, lanes 9 and 10). Hybridization of end-labeled synthetic oligonucleotides to Northern-blotted mRNA with poly(A)$^+$ tails removed indicated that all three adult transcripts and the cercarial transcript contained the repeated element and that sequences 3′ of these elements were similar in all the adult transcripts (Fig. 4, lanes 1 to 4). The 27-base insertion element hybridized to transcripts of all three sizes in adults and to the cercarial 780-base transcript (Fig. 4, lanes 5 and 6). The addition of the primer extension sequence to the length of clones 30-2 (five repeats), 39-1 (five repeats and the 27-bp insertion), and 2-4 (four repeats) yielded transcript sizes of 946, 974, and 866 bases, respectively. The hybridization and cDNA sequence data are consistent with the dominant primer extension product corresponding to the 5′ ends of the 950- and the 870-base adult transcripts. The 27-base difference between the longer transcripts was apparently not resolved on the Northern blots. From the organization of the adult cDNA clone sequences and hybridization data, the 870- and 950-base bands on Northern blots represent at least four transcripts: transcripts with four repeats, four repeats with the 27-base insertion, five repeats, and five repeats with the 27-base insertion.

Primer extension of adult mRNA with the complement of the 27-base insertion (oligonucleotide E) produced a product consistent with the 5′ ends of the 870- and 950-base transcripts. However, an additional product approximately 50 bases longer (ca. 300 bases) was also observed (Fig. 3B, lane 3). Although the cercarial cDNAs that were isolated did not contain the 27-bp insertion, Northern blot hybridizations indicated that a 780-base cercarial transcript contained this 27-base sequence (Fig. 4, lane 6). Primer extension analysis of cercarial mRNA with oligonucleotide E produced an extension product similar in size (ca. 300 bases) to the larger adult extension product observed with this primer (Fig. 3B, lane 4). It seems likely, therefore, that the extended 5′-terminal sequence observed with adult and cercarial mRNAs


A
CATT
A
B
ATACAGGACGGAAGTTGAAATTCAACAATATGAATATCTACCTTATTGGTATATTATGTATTGTGGGTTTGATTATTTCAC 85
MetAsnIleTyrLeuIleGlyIleLeuCysIleValGlyLeuIleIleSerG
C
100
D
AAGGGTCTACTGCTAACGGAAGTCCACTTGACGATCGATTCAATGATGTTAATACCATAAACAAGAAACAATTCACCGAAG 166
lnGlySerThrAlaAsnGlySerProLeuAspAspArgPheAsnAspValAsnThrIleAsnLysLysGlnPheThrGluG
E
217 AAAAACGTTGATATCCGCATAATTGGA 244 27 bp insertion
LysAsnValAspIleArgIleIleGly
200
F
AGGAATTTTCCAGACTGATAAATTCCATGTTAAAGAAGTATATCGAAGACAATAAGAAGGATAAACAGCCTACTCAGAAA 273
luGluPheSerArgLeuIleAsnSerMetLeuLysLysTyrIleGluAsp AsnLysLysAspLysGlnProThrGlnLys
G
300
ACAACTCCCAAACCTACAACACCAAAGCAAATAAATGATGGAACGTCAGACAAAACCAGTGATACCCACACAATTAAAAGA 354
ThrThrProLysProThrThrProLysGlnIleAsnAspGlyThrSerAspLysThrSerAspThrHisThrIleLysArg
400
ACAACTCCCAAACCTACAACACCAAAGCAAATAAATGATGGAACGTCAGACAAAACCAGTGATACCCACACAATTAAAAGA 435
ThrThrProLysProThrThrProLysGlnIleAsnAspGlyThrSerAspLysThrSerAspThrHisThrIleLysArg
500
ACAACTCCCAAACCTACAACACCAAAGCAAATAAATGATGGAACGTCAGACAAAACCAGTGATACCCACACAATTAAAAGA 516
ThrThrProLysProThrThrProLysGlnIleAsnAspGlyThrSerAspLysThrSerAspThrHisThrIleLysArg
ACAACTCCCAAACCTACAACACCAAAGCAAATAAATGATGGAACGTCAGACAAAACCAGTGATACCCACACAATTAAAAGA 597
ThrThrProLysProThrThrProLysGlnIleAsnAspGlyThrSerAspLysThrSerAspThrHisThrIleLysArg
600
ACAACTCCCAAACCTACAACACCAAAGCAAATAAATGATGGAACGTCAGACAAA 651
ThrThrProLysProThrThrProLysGlnIleAsnAspGlyThrSerAspLys
Repeat Elements
700
CCTAAGTCCATAGCCGATATTTTCTTAATCAATAAGCCAAAGGTTCCTTTGTGGATTGTAAATCCTCTATACTATATGGTC 733
ProLysSerIleAlaAspIlePheLeuIleAsnLysProLysValProLeuTrpIleValAsnProLeuTyrTyrMetVal
H
800
GAGAAGTTTGTGCAAATTATGGGATATTTACTAGAAGACGATGATACTTTAGAACTCAACTTGCCAAAATACTATTACGAC 813
GluLysPheValGlnIleMetGlyTyrLeuLeuGluAspAspAspThrLeuGluLeuAsnLeuProLysTyrTyrTyrAsp
I
AAATCAATATGAAAATGGTTTTCAATACTTGGAAAAAGTACATAGTAACGTCAACTCATAACTTGAAACTCAAATAATTAG 894
LysSerIleEnd
900
CATAACTTAATGTTTGTTTTAATCTGATATAAATTTTCAACATTTGCATAAAATTATAAATAAAAACAACTTAACATAT 973
5'
3'
A B D F G H I
C E
direct tandem repeats (5)
100 bp


B
50bp
Adult cDNAs
30-2
(35)
Fifth repeat
912 bp
39-1
(53)
27bp insertion
Fifth repeat
921 bp
21-1
(145)
27bp insertion
Fifth repeat
832 bp
29-1
(275)
Fifth repeat
697 bp
2-4
(81)
785 bp
3-1
(437)
542 bp
10-3 (original clone)
(71)
313 bp
Cercarial cDNAs
C-13/C-25
(79)
787 bp
C-20
(94)
770 bp

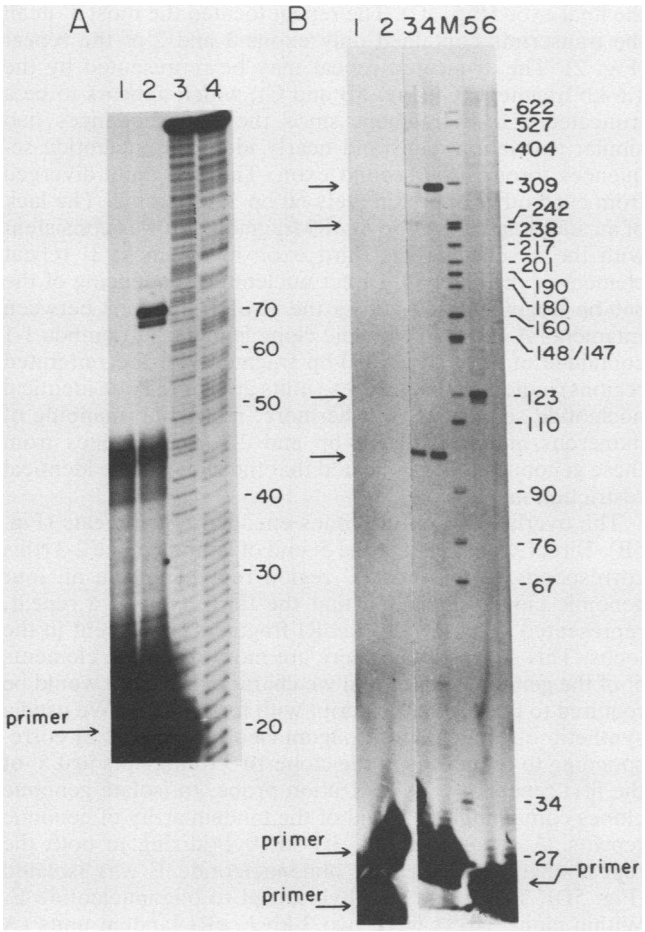


FIG. 3. Primer extension analysis of adult and cercarial mRNAs. End-labeled oligonucleotides were annealed to adult and cercarial mRNAs and extended with reverse transcriptase in the presence of deoxynucleotides. The extension products were analyzed on denaturing acrylamide gels and autoradiographed. (A) Adult mRNA primer extended with oligonucleotide B. Lanes: 1, end-labeled primer alone; 2, 1.5 μg of adult mRNA extended with end-labeled oligonucleotide B; 3 and 4, sequencing reactions (as described by Maxam and Gilbert [25]), with A and G (lane 3) and C and T (lane 4) used as size markers. Cercarial mRNA (up to 4 μg) did not demonstrate any significant extension products with oligonucleotide B. (B) Adult and cercarial mRNAs extended with oligonucleotides D and E. Lanes: 1, end-labeled oligonucleotide D alone; 2, end-labeled oligonucleotide E alone; 3, 1.5 μg of adult mRNA extended with oligonucleotide E; 4, 1.5 μg of cercarial mRNA extended with oligonucleotide E; 5, 1.5 μg of adult mRNA extended with oligonucleotide D; 6, 1.5 μg of cercarial mRNA extended with oligonucleotide D; M, end-labeled *Hpa*II pBR322 markers. The arrows designate the primer extension products described in the text. The numbers are in bases. The primer extension products corresponding to the 950- and 870-base mRNAs are 122 bases in length with primer D and 244 bases in length with primer E. RNA blot experiments suggest that the cercarial 780-base transcript lacking the 27-base insertion is present in a low abundance. The lack of extension products from oligonucleotides D in the size range predicted from the cercarial cDNAs is most likely because the conditions were not sensitive enough to detect these low-abundance products. The products observed in lanes 3 and 4 at 100 bases were not reproducibly observed in all mRNA preparations or primer extension reactions.

corresponds to the 780-base transcripts observed on Northern blots. Partial sequencing of the primer extension product of oligonucleotide E (27-base insertion) with cercarial mRNA indicated that portions of the extension product are different from any sequence observed 5′ of the 27-bp insertion in any of the adult cDNAs. Northern blot analysis of mRNA pretreated with oligonucleotide E and oligo(dT)-RNase H demonstrated that a cercarial mRNA contained both the repeat element and the 27-base insertion (Fig. 4, lanes 6 and 8). Together, these data suggest that the 780-base cercarial band on Northern blots represents at least two types of transcripts which contain the repeat element: a transcript with the 27-base element and a transcript without the 27-base element. The latter transcript was represented by the sequenced cercarial cDNAs. In addition, the 780-base transcripts in both adults and cercariae lacked the 5′-most 30 bp of the cDNA sequence presented in Fig. 1 (Fig. 4, lane 10). These data suggest that the 5′ termini of these transcripts are different from the 5′ termini of the larger adult transcripts. The shortened 5′ ends of the 780-base mRNAs and the variant sequence 5′ of the 27-base insertion in the cercarial mRNA were most likely the result of different transcriptional initiation sites and alternate 5′ exon usage, respectively.

FIG. 2. cDNA clones and sequence of clone 10-3 transcripts. (A) Nucleotide and predicted amino acid sequences of an adult 10-3 transcript and diagrammatic representation of a transcript with the position of the repeated elements and synthetic oligonucleotides used in this study. The sequence was derived by dideoxynucleotide termination (38) and chemical degradation (25) of the cDNA clones represented in panel B and previous extension sequencing of 10-3 transcripts as described in the text. The primer extension sequence constitutes nucleotides 1 to 34. The first 2 nucleotides were identified from the genomic sequence corresponding to the 5′ end of the transcript. The amino acid sequence of the predicted single long open reading frame is presented under the nucleotide sequence. The putative polyadenylation signal is doubly underlined. The cDNA sequence presented represents a transcript containing five repeating elements and a 27-base insertion (clones 39-1 and 21-1; see part B). Removal of the 27-base insertion represents the sequence of clone 30-2 (see part B). The removal of the fifth repeat element represents the sequence of clone 2-4. The sequences of all the adult cDNA clones are identical, except for the presence of the 27-bp insertion, an additional 81-base repeat, or the extent to which they extended in the 5′ direction. Cercarial cDNA C-20 differed from the other cercarial and adult cDNAs by 2 base substitutions at positions 202 (G) and 261 (C) of the sequence presented. Underlined and boxed residues indicate the oligomers used in this study. Oligonucleotide F was contiguous in those clones that lacked a 27-base insertion. The insertion of the 27 bases occurred within oligonucleotide F. (B) cDNA clones analyzed and sequenced in this study. The length of each of the clones is presented at the right, and the numbers in parentheses indicate the base in panel A to which each cDNA extended in the 5′ direction. All the cDNAs except clone 10-3 extended into the poly(A)$^+$ tail of the transcript. The points of insertion of the 27-bp element and the presence of a fifth repeat are noted. To aid in the alignment of the cDNAs, the extra 81 bases for the fifth repeat and the 27-base insertion are shown as extra elements and are not included in the linear length of the lines representing the cDNAs. The actual point of insertion of the fifth repeat is not known, since the repeats were direct and identical. Clone 3-1 extended only as far 5′ as three repeats, and clone 10-3 extended through 1 repeat and 56 bases of a second repeat.

Developmental transition to adult mRNA expression. During the transformation and development of cercariae to adults, the number and size of the 10-3 transcripts increased (Fig. 1 and 4). On penetrating the skin of the host, cercariae rapidly transform into schistosomula, with the loss of their cercarial tails. Artificially transformed schistosomula are capable of developing to maturity in the vertebrate host (1). To determine whether artificially transformed schistosomula incubated in vitro for 3, 12, and 24 h exhibited the cercarial or adult pattern of transcripts, Northern blots of RNase H-treated schistosomula mRNA were hybridized with the repeat oligonucleotide (oligonucleotide G). Schistosomula mRNA from all three time points exhibited the pattern of cercarial transcript hybridization, suggesting that the transition to the synthesis of adult-sized transcripts does not occur prior to 24 h of in vitro development of schistosomula to adults.

Characterization of genomic clones corresponding to the repeat elements in 10-3 cDNAs. To determine the genomic organization of the repeated elements, the clone 10-3 cDNA was used as a hybridization probe to isolate genomic clones encoding the repeating elements in the transcripts. Two hundred thousand recombinant bacteriophage (equal to approximately nine genomic equivalents) were screened. Seven positive clones were identified corresponding to four unique overlapping clones spanning 18.5 kb (Fig. 5A). These clones encoded the repeat elements in the cDNAs, as well as regions 3′ to at least bp 768 of the cDNA (Fig. 2). Restriction mapping, hybridization, and sequencing data derived from the genomic fragments encoding the repeating elements in the transcripts revealed that each 81-base repeat was derived from three 27-bp exons (Fig. 5B and C). Analysis of these 27-bp exons indicated that they each have consensus splice junctions (33) (Fig. 6). Furthermore, the exons which make up the repeated elements in the transcripts are located in highly conserved, reiterated 3-kb genomic regions spanning a total of at least 13 kb. This reiteration in the genomic organization is illustrated by fragments A and B in Fig. 5B. Each 81-base transcript repeat was derived from two adjacent *Eco*RI fragments: a 2.1-kb fragment (A) and an 860-bp fragment (B). The 2.1-kb fragment contained the first two exons of the repeat, whereas the 860-bp fragment contained the final exon (Fig. 5C). The repeat located the most 3′ in all the transcripts contained only exons 1 and 2 of the repeat (Fig. 2). This truncated repeat may be represented by the 1.6-kb fragment (C) (Fig. 5B and C), which appears to be a truncated 2.1-kb fragment since the two sequences had similar restriction maps and nearly identical nucleotide sequences through the second exon. The two units diverged from each other approximately 60 bp 3′ of exon 2. The lack of an additional 3′ 860-bp *Eco*RI fragment (B) was consistent with the absence of the third exon in the most 3′ repeat element of the cDNAs. Direct nucleotide sequencing of the 860-bp fragments and across the *Eco*RI junctions between fragments A and C in genomic clone lambda 1-1 (lambda 1-1 contained at least three 860-bp fragments in the reiterated regions) indicated that all these units appear to have identical nucleotide sequences. Furthermore, restriction mapping of numerous independent 860-bp and 2.1-kb subclones from these genomic clones indicated that the clones have identical restriction maps.

The overlapping set of clones encoded four repeats (Fig. 5B). Direct sequencing of the 5′ end of lambda clone 2-3 (this corresponds to the most 5′ region represented in all four genomic clones) indicated that the third exon of a repeat, represented by an 860-bp *Eco*RI fragment, is present in the locus. This suggests that there are more repeating elements 5′ of the genomic clones that we characterized that would be required to produce a transcript with five repeats. We used a synthetic oligonucleotide (oligonucleotide F; Fig. 2) corresponding to sequences in the clone 10-3 transcripts just 5′ of the first repeat as a hybridization probe, to isolate genomic clones containing the 5′ end of the tandem array of genomic repeats. A genomic clone, 3-1-34, hybridizing to both the repeat oligonucleotide and oligonucleotide F was isolated (Fig. 5D). This clone also hybridized to oligonucleotide E. Within clone 3-1-34 were two 3-kb *Eco*RI tandem units (A and B) and a 1.6-kb restriction fragment similar to fragment C. The 1.6-kb *Eco*RI fragment in the 3-1-34 clone, however, lacked the *Bam*HI site that was observed in the analogous fragment in the overlapping set of clones in Fig. 5C. Furthermore, the restriction map of regions 3′ of the 1.6-kb fragment in clone 3-1-34 was not the same as the regions 3′ of the tandem arrays in the genomic clones depicted in Fig.

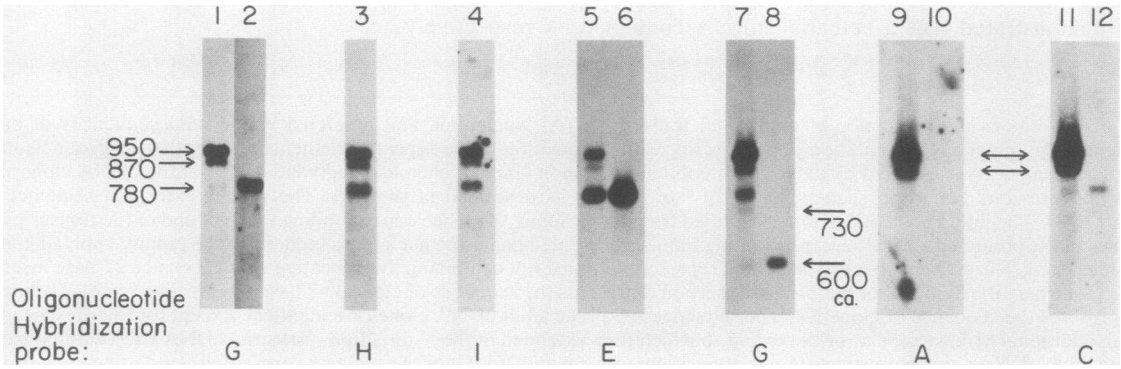


FIG. 4. Characterization of adult and cercarial transcripts by Northern blot hybridization analysis. Adult and cercarial mRNAs were treated with oligo(dT)-RNase H, electrophoresed on 1.7% agarose-formaldehyde gels, blotted to nitrocellulose, and hybridized to end-labeled oligonucleotides representing different regions of the clone 10-3 transcripts (see Fig. 2 for locations of oligonucleotides). All lanes contained 1 to 2 μg of mRNA. Lanes: 1, 3, 4, 5, 7, 9, and 11, adult mRNAs; 2, 6, 8, 10, and 12, cercarial mRNAs. mRNAs in lanes 7 and 8 were treated with oligo(dT)-RNase H in the presence of 150 ng of oligonucleotide E prior to electrophoresis. Longer exposures of lane 8 revealed residual repeat hybridization at 780 bases, consistent with the absence of oligonucleotide E in the cercarial cDNA clones. Lanes 9 to 12 were exposed at least 5 times longer than the other lanes were to demonstrate the absence of the 780-base hybridization in lanes 9 and 10 and to show the presence of hybridization of the 780-base transcripts in lanes 11 and 12. Exposure times for lanes 1 and 2 were 4 times less than those for lanes 3 to 6. Numbers are in bases.

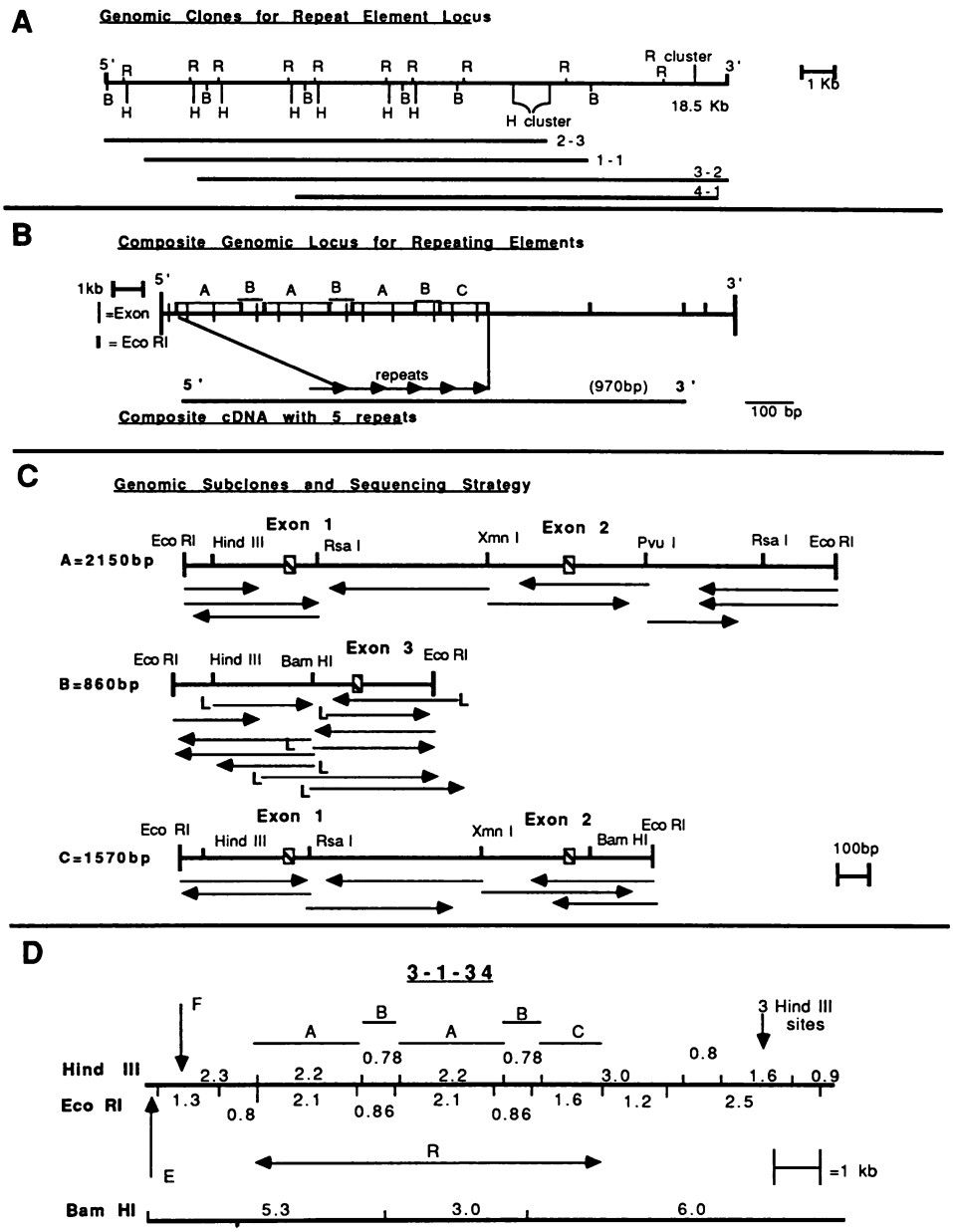


FIG. 5. Genomic clones corresponding to the repeat elements in clone 10-3 transcripts and the sequencing strategy used to sequence reiterated fragments encoding the repeats in 10-3 transcripts. (A) Overlapping genomic clones corresponding to the repeat element. These clones were isolated by screening with the 10-3 cDNA. Restriction site abbreviations: B, *Bam*HI; H, *Hin*dIII; R, *Eco*RI. (B) Composite of overlapping clones illustrating the repeating fragments and the area spliced to form the repeats in transcripts. The region correlated with the composite cDNA would encode four repeats, with the final repeat truncated to 54 bases. Regions A, B, and C are expanded in panel C. (C) Sequencing strategy for the fragments contributing to the reiterated genomic array encoding the 81-base repeats in clone 10-3 mRNAs. Sequencing was performed by the chemical degradation method of Maxam and Gilbert (25). The shaded boxes represent the location of the exons of the transcript repeats. Arrows with the letter L indicate fragments that were sequenced directly from the genomic clones. (D) Restriction map and hybridization pattern of genomic clone 3-1-34. Clone 3-1-34 was isolated by screening with an end-labeled oligonucleotide (oligonucleotide F) located just 5′ of the first transcript repeat. This clone contained regions homologous to the transcript repeats and to the 27-base insertion. Fragments A, B, and C corresponded by hybridization and restriction mapping to the same fragments shown in panels A to C. Arrows indicate the fragments and regions within the clone that hybridized with oligonucleotides E and F. R represents the *Eco*RI fragments which hybridized with a portion of clone 39-1 containing the repeats.

5B. These data could suggest that an interruption is present in the reiteration of the array in the schistosome genome that is perhaps produced by the duplicative events yielding the array. The sequence of selected 500-bp regions containing both exons and introns within the conserved arrays in clone 3-1-34 (portions of fragments A and B) indicated that there was greater than 98% identity with the same regions in the overlapping set of clones in Fig. 5A. Splicing of the two exons in the 1.6-kb region of clone 3-1-34 with the third exon located at the 5′ end of the set of four overlapping repeat clones would yield a typical 81-base repeat. Schistosomes are diploid, and laboratory strains are not clonal lines. Thus,

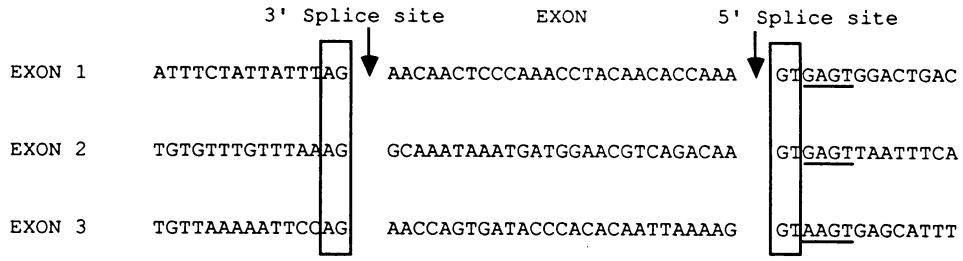


FIG. 6. Exon sequence and genomic intron sequence surrounding the splice junctions of the three exons making up the repeating element in clone 10-3 cDNAs. The typical pyrimidine stretch in the 3′ splice site may not be pronounced in *S. mansoni* because of the low G+C content observed in the gene. Underlined areas in the 5′ splice sites illustrate conservation of typical splice site sequences observed in the 10-3 exon-intron junctions.

this clone could also represent another allele or could be a pseudogene.

Four transcript repeats (consisting of three 81-base repeats and the typical fourth 54-base truncated repeat, which would be equivalent to 297 bases of a transcript) would be encoded in 11 27-bp exons interspersed in a contiguous conserved repetitive array of intron and exon sequences spanning at least 10 kb. mRNAs containing five repeats would be encoded over a larger region of genomic DNA. Both exons and introns within this locus are highly A and T rich (65%).

Southern blots with genomic DNA with hybridization probes containing the repeat elements produced a hybridization pattern consistent with the organization of the repeats in the genomic clones (Fig. 7). Southern blot titrations comparing cloned fragments from the 10-3 locus with restricted genomic DNA (oligonucleotides C and H used as probes) suggested that there are one or two copies of the gene. Moreover, genomic Southern blots, with oligonucleotides to regions in the cDNAs (oligonucleotides A, B, G, D, E, H, and I) and larger fragments of the cDNAs used as probes, suggest that the whole gene(s) is subdivided into many small exons. Characterization of additional genomic clones corresponding to the 5′ ends of the transcripts indicate that these regions are also composed of small exons and that an entire gene is at least 40 kb in length. The presence of many small A- and T-rich exons in the clone 10-3 locus, however, makes analysis of this locus difficult.

Putative proteins encoded by the 10-3 transcripts. Computer-assisted analysis of the translation of the cDNA presented in Fig. 2A displayed a long open reading frame from nucleotides 3 to 822. No other open reading frames of significant length were present. We designated the first AUG at base 34 as the translation initiation site based on its proximity to the 5′ end and sequence context (5, 21). Neither the 5′ 27-base insertion nor an additional fifth repeat changed the single long open reading frame. Thus, all the transcripts would use the same open reading frame, producing similar proteins with potential peptide insertions coded by the 27-base insertion or the number of repeat elements in the transcript. Use of the first AUG codon as the initiator indicates that the amino terminus has a hydrophobic region that could be a signal peptide. Alternatively, these hydrophobic residues could constitute a membrane domain in the protein. The amino acid residues of the putative protein depicted in Fig. 2A are predominantly hydrophilic and display relatively high frequencies of Thr (15.9%), Lys (12.9%), Ile (9.8%), Asp (9.5%), and Pro (8.3%). Glycosylation may occur at the serine at amino acid 25 (Asn-Gly-Ser), as well as the numerous Thr residues in the peptide repeats. Computer-assisted comparison of this putative protein with protein data banks did not reveal a significant identity with any other known proteins.

DISCUSSION

We have described *S. mansoni* transcripts containing repeat elements that are expressed differentially during the *S. mansoni* life cycle and the expression of different-sized transcripts in adults and cercariae. The variations in these clone 10-3 transcripts included the number of 81-base repeat elements, a 27-base insertion, differences in the 5′ ends of the transcripts, and alternate exons 5′ of the 27-base insertion in some transcripts. These differences could be the result of alternative processing and transcription initiation sites from a single locus. Alternatively, numerous highly similar transcripts might reflect the existence of multiple, highly similar loci. These conserved loci would have various numbers of repeat elements with the 27-bp insertion and a number of loci without the 27-bp insertion. This model would also require that some loci have additional 5′ differences to account for the clone 10-3 transcripts. Complicating these analyses is that schistosomes are diploid and laboratory strains are not clonal lines. However, the presence of multiple, highly conserved loci appears unlikely from the genomic titrations, Southern blot data, the genomic clones, and the frequency with which clones were isolated from the genomic libraries. Furthermore, the invariant nucleotide sequence of the repeating elements through the 3′ ends of nine adult and cercarial cDNAs strongly suggests that these mRNAs are derived from a single locus. Therefore, the data indicate that one or, perhaps, two loci produce developmentally regulated transcripts that are alternatively spliced and exhibit different transcriptional start sites, to yield transcripts of various sizes and compositions in adults and cercariae.

The nucleotide sequence of the cDNAs and primer extension and hybridization analyses with oligonucleotide probes permitted us to identify the sequence variations in the transcripts observed on Northern blots. The data are consistent with the fact that the largest band on Northern blots (950 bases) represents at least two transcripts which contain five tandem repeats. One of these transcripts contained a 27-base insertion 5′ of the repeats, while the other did not. The 870-base band on Northern blots included two transcripts containing four tandem repeats, with or without the 27-base insertion 5′ of the repeats. Although we did not isolate an adult cDNA clone with four repeats and the 27-base insertion, the data indicate that a 870-base transcript with these characteristics is present. The data from this study also suggest that the splicing of the repeat elements to yield transcripts with four or five repeats occurs in a coor-

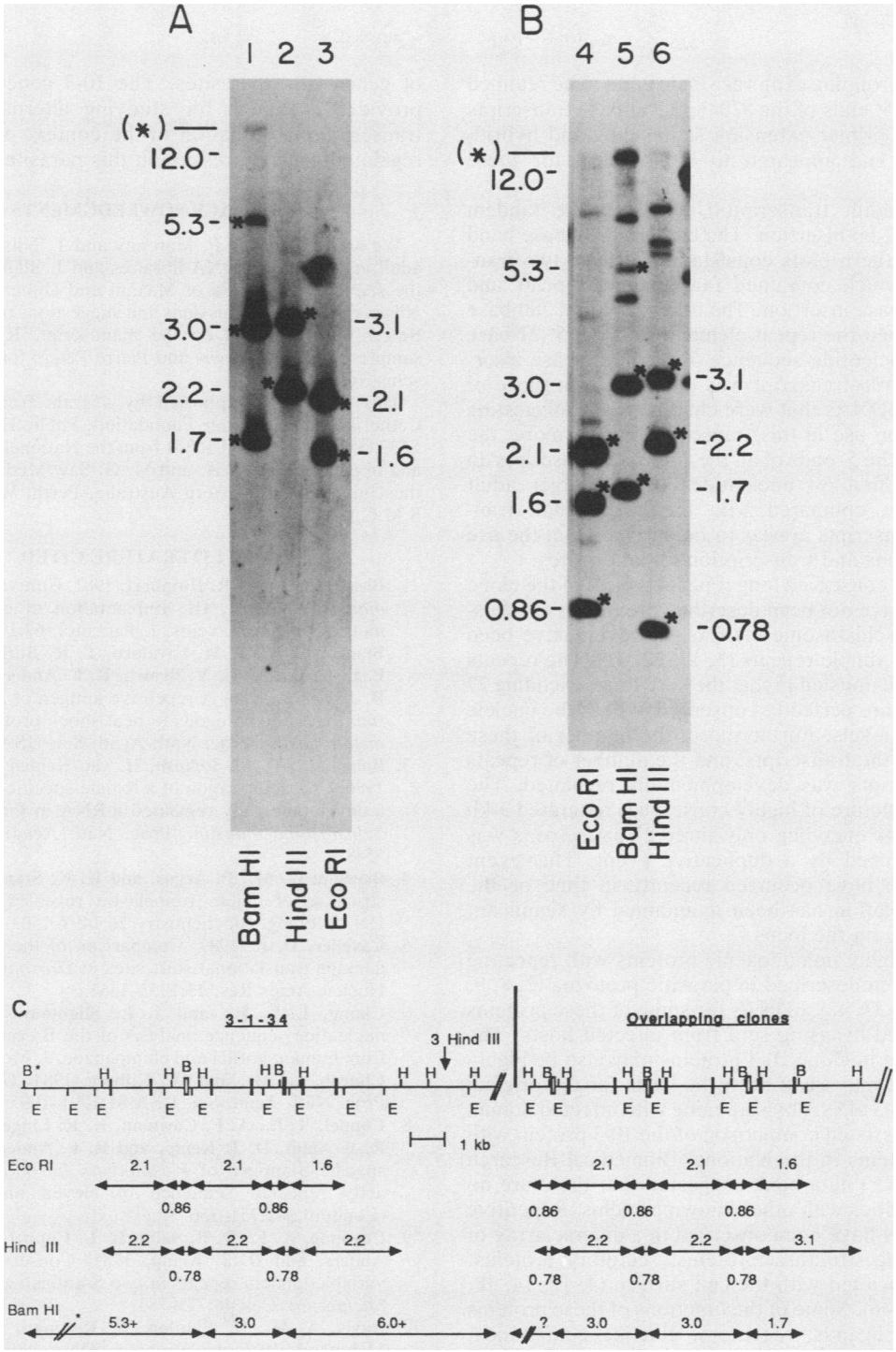


FIG. 7. Genomic adult DNA Southern blots probed with random primer-labeled clone 39-1 or end-labeled oligonucleotide G. Each lane contained 10 μg of genomic DNA digested with the respective restriction endonuclease, fractionated on 0.75% agarose gels, transferred to a Gene Screen nylon membrane, cross-linked to the membrane with UV light, and hybridized with radiolabeled probe. (A) End-labeled oligonucleotide G was used as the hybridization probe. (B) Random primer-labeled clone 39-1 was used as the hybridization probe. The multicopy bands observed with the 39-1 probe at 860 bp (*Eco*RI) and 780 bp (*Hin*dIII) represent exon 3, which was not present in the oligonucleotide G probe. The asterisks indicate fragments corresponding to restriction fragments containing the repeat elements predicted from the genomic clones depicted in Fig. 5 and in panel C. Southern blots with a restriction fragment of clone 39-1 containing just the five repeat elements produced the same pattern of bands in panel B (denoted by asterisks). The additional bands in panel B represent the other exons in clone 10-3 mRNAs. Shorter exposures showed differences in intensities of hybridization among bands within the repeat array, reflecting the number of copies observed in the array. For example, the 2.2-kb *Hin*dIII fragment was present in more copies than the 3.1-kb *Hin*dIII fragment. The 1.6-kb *Eco*RI band on lighter exposure indicated that two bands were present. Numbers indicate the size of hybridizing fragments (in kilobases). (C) Diagrammatic representation of fragments from the genomic clones predicted to hybridize with probes containing portions of the repeat elements. The 0.86-kb *Eco*RI and 0.78-kb *Hin*dIII fragments only hybridized to exon 3 of the repeat element, which was absent in the oligonucleotide G probe. Open boxes denote exon 3, and closed boxes denote exon 1 or 2. The discontinuous arrows indicate unknown fragment lengths for which genomic clone data were not available. It is not known whether the 5′-most *Bam*HI site in clone 3-1-34 is an endogenous *Bam*HI site or a *Bam*HI site created from the ligation of a *Sau*3A genomic site into the *Bam*HI site of the EMBL-3 arm (denoted by asterisks). Only two 3-kb genomic repeats are depicted from the overlapping clones.

dinated manner, grouping exons as single units to be retained or removed. The 5′ ends of the 870- and 950-base transcripts were mapped by primer extension sequencing and hybridization analyses and appeared to be identical for these transcripts.

The smallest adult transcript(s) exhibited the tandem repeats and the 27-bp insertion. The cercarial 780-base band observed on Northern blots consisted of at least two transcripts, one of which contained four tandem repeats and lacked the 5′ 27-base insertion. The other cercarial 780-base transcript contained the repeat elements and the 5′ 27-base insertion. The nucleotide sequence 5′ of the 27-base insertion in this cercarial transcript was different from those of any of the other cDNAs that were characterized, suggesting alternative 5′ exon use in this transcript. Furthermore, the data suggest that the 5′ ends of all the 780-base transcripts do not contain the first 30 nucleotides of the larger adult transcripts. Thus, compared with the larger adult transcripts, these transcripts appear to be derived from the use of alternative exons and transcription initiation sites.

Multiple highly conserved long repeats typical of the clone 10-3 transcripts have not been described previously in schistosomes. A few schistosome cDNAs, however, have been shown to contain simple repeats (3, 17, 22, 31). The repeats described here are unusual in that they are long, encoding 27 amino acids, and are perfectly conserved at both the nucleic acid and protein levels. Furthermore, the number of these repeats varied in the transcripts, and the number of repeats in certain transcripts was developmentally regulated. The complex gene structure of highly conserved, reiterated 3-kb genomic fragments encoding only three 27-base exons was most likely produced by a duplicative event. This event either appears to have occurred recently in time or the sequence conservation has been maintained by significant selective pressure on the locus.

A number of highly immunogenic proteins with repeating elements have been described in parasitic protozoa (2, 8, 9, 15, 20, 28, 32, 34, 37, 42). cDNAs for some of these proteins have been isolated by using sera from infected hosts. The repeating peptides in clone 10-3 proteins may also be highly immunogenic and may have been the basis for the original isolation of the 10-3 cDNA by screening with infected human sera. Computer-assisted comparison of the 10-3 protein with other known proteins in the National Biomedical Research Foundation (NBRF) data bank suggested that there are no significant similarities with other known proteins. Repetitive peptide sequences have been observed in a diverse array of proteins, including structural proteins, secretory proteins, and proteins associated with the cell surface (4, 12, 14, 18, 19, 27, 29, 39, 44–46). Some of the functions of these proteins include receptor ligands, receptors, gluelike compounds, connective tissue components, clotting proteins, and structural components. The number and length of the repeating elements observed in such proteins, as well as the sequence or residue conservation of the individual repeating units, is quite variable. The broad array of functions observed for proteins with repeating units precludes functional analogies from being drawn for the clone 10-3 protein. The presence of either a possible signal peptide or membrane domain could suggest that some 10-3 proteins are secreted or are membrane bound. The potential variance in amino acid composition and function of *S. mansoni* proteins derived from all of the 10-3 transcripts remains to be elucidated. However, the conserved repeated sequence organization of the 10-3 gene(s) and the multiple mRNAs encoded therein provide another example of the unique organization and expression of genes from parasites. The 10-3 gene(s) in *S. mansoni* provides a system for studying alternative splicing and transcriptional initiation in the context of developmentally regulated gene expression in this parasite.

ACKNOWLEDGMENTS

We are indebted to P. Maroney and T. Nilsen for providing the adult and cercarial cDNA libraries and J. Slightom for teaching us the sequencing method of Maxam and Gilbert (25). We thank T. Nilsen for helpful discussions and suggestions on the manuscript, D. Setzer for comments on the manuscript, R. Blanton for RNA samples, and Fred Lewis and Pierre Peters for cercariae and adult schistosomes.

This study was supported by a grant from the John D. and Catherine T. MacArthur Foundation, Public Health Service grants 5T32 GMO7250 and AI 15351 from the National Institutes of Health, and in part by the W. A. and M. G. Saw Medical Fellowship from the University of Western Australia, Perth, Western Australia (to S.M.C.).

LITERATURE CITED

1. **Basch, P. F., and R. Humbert.** 1981. Cultivation of *Schistosoma mansoni* in vitro. III. Implantation of cultured worms into mouse mesenteric veins. J. Parasitol. **67**:191–195.
2. **Bianco, A. E., J. M. Favalaro, T. R. Burkot, J. G. Culvenor, P. E. Crewther, G. V. Brown, R. F. Anders, R. L. Coppel, and D. J. Kemp.** 1986. A repetitive antigen of *Plasmodium falciparum* that is homologous to heat shock protein 70 of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **83**:8713–8717.
3. **Bobek, L., D. M. Rekosh, H. van Keulen, and P. T. Loverde.** 1986. Characterization of a female-specific cDNA derived from a developmentally regulated mRNA in the human blood fluke *Schistosoma mansoni*. Proc. Natl. Acad. Sci. USA **83**:5544–5548.
4. **Bressan, G. M., P. Argos, and K. K. Stanley.** 1987. Repeating structure of chick tropoelastin revealed by complementary DNA cloning. Biochemistry **26**:1497–1503.
5. **Cavener, D. R.** 1987. Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. Nucleic Acids Res. **15**:1353–1361.
6. **Chang, L.-Y. E., and J. L. Slightom.** 1984. Isolation and nucleotide sequence analysis of the B-type globin pseudogene from human, gorilla and chimpanzee. J. Mol. Biol. **180**:767–784.
7. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
8. **Coppel, R. L., A. F. Cowman, K. R. Lingelbach, G. V. Brown, R. B. Saint, D. J. Kemp, and R. F. Anders.** 1983. An isolate-specific S-antigen of *Plasmodium falciparum* contains an exactly repeated sequence of eleven amino acids. Nature (London) **302**:751–756.
9. **Cowman, A. F., R. B. Saint, R. L. Coppel, G. V. Brown, R. F. Anders, and D. J. Kemp.** 1985. Conserved sequences flank variable tandem repeats in two S-antigen genes of *Plasmodium falciparum*. Cell **40**:775–783.
10. **Davis, A. H., R. Blanton, F. Rottman, R. Maurer, and A. Mahmoud.** 1986. Isolation of cDNA clones for differentially expressed genes of the human parasite *Schistosoma mansoni*. Proc. Natl. Acad. Sci. USA **83**:5534–5538.
11. **Duvall, R. H., and W. B. DeWitt.** 1967. An improved perfusion technique for recovering adult schistosomes from laboratory animals. Am. J. Trop. Med. Hyg. **16**:483–486.
12. **Eckert, R. L., and H. Green.** 1986. Structure and evolution of the human involucrin gene. Cell **46**:583–589.
13. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
14. **Garfinkel, M. D., R. E. Pruitt, and E. M. Meyerowitz.** 1983. DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J. Mol. Biol. **168**:765–789.
15. **Guerin-Marchand, C., P. Druilhe, B. Galey, A. Londono, J. Patarapotikul, R. L. Beaudoin, C. Dubeaux, A. Tartar, O.**

Mercereau-Puijalon, and G. Langsley. 1987. A liver-stage-specific antigen of *Plasmodium falciparum* characterized by gene cloning. Nature (London) **329**:164–167.
16. **Hattori, M., and Y. Sakai.** 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152**:232–238.
17. **Johnson, K. S., D. W. Taylor, and J. S. Cordingley.** 1987. Possible eggshell protein gene from *Schistosoma mansoni*. Mol. Biochem. Parasitol. **22**:89–100.
18. **Kidd, S., M. R. Kelly and M. W. Young.** 1986. Sequence of the notch locus of *Drosophila melanogaster*: relationship of the encoded protein to mammalian clotting and growth factors. Mol. Cell. Biol. **6**:3094–3108.
19. **Klickstein, L. B., W. W. Wong, J. A. Smith, J. H. Weis, J. G. Wilson, and D. T. Fearon.** 1987. Human C3b/C4b receptor (CR1): demonstration of long homologous repeating domains that are composed of short consensus repeats characteristic of C3/C4 binding proteins. J. Exp. Med. **165**:1095–1112.
20. **Kochan, J., M. Perkins, and J. V. Ravetch.** 1986. A tandemly repeated sequence determines the binding domain for an erythrocyte receptor binding protein of *P. falciparum*. Cell **44**:689–696.
21. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
22. **Lanar, D. E., E. J. Pearce, S. L. James, and A. Sher.** 1986. Identification of paramyosin as schistosome antigen recognized by intradermally vaccinated mice. Science **234**:593–596.
23. **Lazdins, J. K., M. J. Stein, J. R. David, and A. Sher.** 1982. *Schistosoma mansoni*: rapid isolation and purification of schistosomula of different developmental stages by centrifugation on discontinuous density gradients of Percoll. Exp. Parasitol. **53**:39–44.
24. **Maniatis, T., E. F. Fritsch, and J. M Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
26. **Meinkoth, J., and G. Wahl.** 1984. Hybridization of nucleic acids immobilized on solid supports. Anal. Biochem. **138**:267–284.
27. **Moore, J. T., J. Hagstrom, D. J. McCormick, S. Harvey, B. Madden, E. Holicky, D. R. Stanford, and E. D. Wieben.** 1987. The major clotting factor protein from guinea pig seminal vesicle contains eight repeats of a 24-amino acid domain. Proc. Natl. Acad. Sci. USA **84**:6712–6714.
28. **Mowatt, M. R., and C. E. Clayton.** 1987. Developmental regulation of a novel repetitive protein of *Trypanosoma brucei*. Mol. Cell. Biol. **7**:2838–2844.
29. **Muskavitch, M. A. T., and D. S. Hogness.** 1982. An expandable gene that encodes a *Drosophila* glue protein is not expressed in variants lacking remote upstream sequences. Cell **29**:1041–1051.
30. **Narayan, P., D. F. Ayers, F. M. Rottman, P. A. Maroney, and T. W. Nilsen.** 1987. Unequal distribution of N^6-methyladenosine in influenza virus mRNAs. Mol. Cell. Biol. **7**:1572–1575.
31. **Newport, G. R., R. A. Harrison, J. McKerrow, P. Tarr, J. Kallestad, and N. Agabian.** 1987. Molecular cloning of *Schistosoma mansoni* myosin. Mol. Biochem. Parasitol. **26**:29–38.
32. **Ozaki, L. S., P. Svec, R. S. Nussenzweig, V. Nussenzwieg, and G. N. Godson.** 1983. Structure of the *Plasmodium knowlesi* gene coding for the circumsporozoite protein. Cell **34**:815–822.
33. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.
34. **Peterson, D. S., R. A. Wrightsman, and J. E. Manning.** 1986. Cloning of a major surface-antigen gene of *Trypanosoma cruzi* and identification of a nonapeptide repeat. Nature (London) **322**:566–568.
35. **Rackwitz, H.-K., G. Zehetner, A.-M. Frischauf, and H. Lehrach.** 1984. Rapid restriction mapping of DNA cloned in lambda phage vectors. Gene **30**:195–200.
36. **Rigby, P. W. J., M. Dieckmann, C. Rhodes and P. Berg.** 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**: 237–251.
37. **Roditi, I., M. Carrington, and M. Turner.** 1987. Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*. Nature (London) **325**:272–274.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
39. **Sharp, P. M., and W.-H. Li.** 1987. Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. J. Mol. Evol. **25**:58–64.
40. **Simpson, A. J. G., A. Sher, and T. F. McCutchan.** 1982. The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. Mol. Biochem. Parasitol. **6**:125–137.
41. **Slightom, J. L., T. W. Theisen, B. F. Koop, and M. Goodman.** 1987. Orangutan fetal globin genes: nucleotide sequences reveal multiple gene conversions during hominid phylogeny. J. Biol. Chem. **262**:7472–7483.
42. **Stahl, H.-D., P. E. Crewther, R. F. Anders, G. V. Brown, R. L. Coppel, A. E. Bianco, G. F. Mitchell, and D. F. Kemp.** 1985. Interspersed blocks of repetitive and charged amino acids in a dominant immunogen of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA **82**:543–547.
43. **Vournakis, J. N., A. Efstratiadis, and F. C. Kafatos.** 1975. Electrophoretic patterns of deadenylylated chorion and globing mRNAs. Proc. Natl. Acad. Sci. USA **72**:2959–2963.
44. **Yamada, Y., V. E. Avvedimento, M. Mudryi, H. Ohkubo, G. Vogeli, M. Irani, I. Pastan, and B. de Crombrugghe.** 1980. The collagen gene: evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54 bp. Cell **22**: 887–892.
45. **Yamamoto, T., C. G. Davis, M. S. Brown, W. J. Schneider, M. L. Casey, J. L. Goldstein, and D. W. Russell.** 1984. The human LDL receptor: a cysteine-rich protein with multiple Alu sequences in its mRNA. Cell **39**:27–38.
46. **Yang, C.-Y., S.-H. Chen, S. H. Gianturco, W. A. Bradley, J. T. Sparrow, M. Tanimura, W.-H. Li, D. A. Sparrow, H. DeLoof, M. Rosseneu, F.-S. Lee, Z.-W. Gu, A. M. Gotto, and L. Chan.** 1986. Sequence, structure, receptor binding domains and internal repeats of human apolippoprotein B-100. Nature (London) **323**:738–742.

Analyze and select for export

SEQ ID 1 BLASTP Show search parameters Select another search Update results Rerun

Filters Show available filters Reset

Add filter Select Field Add or Select saved filter

Alignment percent Identity:

Equals

100

Remove

Apply Save filter parameters

Seqs: 2 Redundant: 2 Docs: 2 Families: 1

Table View Graphical Representation

Add selected Add all Export results

Grouped by: Document Toggle visible columns

Showing 1 to 2 of 2 entries

DOCUMENT	PRIORITY	PUBLISHED	SCORE	ALIGN % ID	QUERY % ID	SUBJECT ID
AAA29855	Apr 26, 1993	Apr 26, 1993	205	100.00%	100.00%	100.00%	100.00%	100
BR102015027298	Oct 28, 2015	May 2, 2017	18	100.00%	100.00%	95.45%	8.00%	95.

First Previous 1 Next Last

Show 50 entries

AAA29855 GenBank®
Modification date: **1993-04-26**

Sequence number
AAA29855.1

Sequence key
SBNP001CLS6E

Locus
AAA29855

Division
INV

Definition
antigen 10-3 [Schistosoma mansoni].

Accession
AAA29855

Version
1

GI
160931

Title
Tandemly repeated exons encode 81-base repeats in multiple, developmentally regulated Schistosoma mansoni transcripts

Authority
Not provided

Author(s)
Davis, R.E.; Davis, A.H.; Carroll, S.M.; Rajkovic, A.; Rottman, F.M.

Modification date
1993-04-26

Molecule type
Protein

Molecule shape

linear

Organism (common name)

Schistosoma mansoni

Organism

Schistosoma mansoni

Taxonomy

Eukaryota
Metazoa
Spiralia
Lophotrochozoa
Platyhelminthes
Trematoda
Digenea
Strigeidida
Schistosomatoidea
Schistosomatidae
Schistosoma

Length

263

PubMed

3211127

References

Reference 1

authors	Davis, R.E.; Davis, A.H.; Carroll, S.M.; Rajkovic, A.; Rottman, F.M.
title	Tandemly repeated exons encode 81-base repeats in multiple, developmentally regulated Schistosoma mansoni transcripts
journal	Mol. Cell. Biol.
volume	8
issue	11
pages	4745-4755
year	1988
pubmed	3211127
embl_gb_record_number	1
sequence_position	residues 1-263

Sequence

>GenBank®|AAA29855.1|Protein
MNIYLIGILCIVGLIISQGSTANGSPLDDRFNDVNTINKKQFTEEEFSRLINSMLKKYIE
DKNVDIRIIGNKKDKQPTQKTTPKPTTPKQINDGTSDKTSDTHTIKRTTPKPTTPKQIND
GTSDKTSDTHTIKRTTPKPTTPKQINDGTSDKTSDTHTIKRTTPKPTTPKQINDGTSDKT
SDTHTIKRTTPKPTTPKQINDGTSDKPKSIADIFLINKPKVPLWIVNPLYYMVEKFVQIM
GYLLEDDDTLELNLPKYYYDKSI

Feature table

KEY	LOCATION	FEATURE
source	1..263	/organism="Schistosoma mansoni" /db_xref="taxon:6183"
Protein	1..263	/name="antigen 10-3"
CDS	1..263	/coded_by="M22346.1:34..825"

Comment

Draft entry and computer-readable sequence for [1] kindly provided by R.Davis, 26-JAN-1989.
The actual point of insertion of the fifth repeat is unknown, since the repeats were direct and identical. Clone 3-1 contained only three repeats, and clone 10-3 extended through only one repeat and 56 base pairs of a second one.
Method: conceptual translation.

Source

GENBANK; PROTEIN