BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **219,** 231–237 (1996)
ARTICLE NO. 210

Cloning and Characterization of a Ribosomal P Protein from *Taenia solium,* The Aetiological Agent of Human Cysticercosis

B. H. Kalinna,[1] and D. P. McManus

Molecular Parasitology Unit, Tropical Health Program and Australian Centre for International and Tropical Health and Nutrition, The Queensland Institute of Medical Research, The Bancroft Centre, P.O. Box Royal Brisbane Hospital, Brisbane, Queensland 4029, Australia

Received November 30, 1995

A cDNA encoding the complete open reading frame of the *Taenia solium* acidic ribosomal phosphoprotein P2 has been cloned. The 417 bp cDNA (arpP2) was isolated from a *T. solium* cDNA expression library by PCR using P protein specific pimers. The ORF encodes a protein of 121 amino acids exhibiting 40-55% identity to mammalian, fungal, insect, and parasite P2 proteins. The inferred molecular mass of the protein is 12,592 Daltons, similar to that reported for other ribosomal P2 proteins. After subcloning and expression, the recombinant protein was purified by affinity chromatography under non-denaturing conditions and was shown to have a molecular mass of 15 kDa which is equivalent to the expected size of the full length recombinant fusion protein, comprising the 12.6 kDa arpP2 plus an additional 2 kDa for the N-terminal fusion peptide incorporating the six histidine residues required for purification.

Dominant and evolutionarily conserved protein components of all ribosomes are the P (phospho)-proteins (also called the "A" acidic proteins). A recent overview of their location, function, structure and antigenicity is available in (1). The P proteins play an essential role in protein synthesis and are unique among the ribosome components due to their acidic p,*I* and to the fact that they are the only molecules in ribosomes which are present in multicopy. In general, eukaryotic ribosomal P proteins consist of three antigenically cross-reactive family members, P1, P2 and P0, which form a pentameric complex in which P1/P2 homodimers are attached to a single P0 protein via their NH_2-terminal end; they contain a variable region rich in alanine, proline and glycine adjacent to an acidic carboxy domain which is highly conserved both within and across species from humans to yeast. P0 is the largest protein of the family with an apparent molecular mass of approximately 38 kDa, while P1 and P2 migrate as a doublet in the 12-19 kDa range on SDS-polyacrylamide gels. This eucaryotic P protein system shows a marked degree of protein diversity, with at least three different ribosomal P protein genes in mammals (2) and five in yeast (3, 4). Furthermore, the acidic ribosomal proteins are phosphorylated in order to facilitate assembly into the ribosome (5). In higher eucaryotes the phosphorylation sites have been identified as C-terminal serine residues (5, 6).

The ribosomal P components may be important in the induction of autoimmune responses as, in the classical autoimmune disease systemic lupus erythematosus (SLE), anti-P antibodies are generated to the C-terminal region (residues SDDDMGFGLFD) of the human proteins (7). Latin American patients with chronic cardiac Chagas' disease, caused by the intracellular protozoan parasite, *Trypanosoma cruzi* (Tc), also present a strong humoral response against the same C-terminal region of the Tc acidic ribosomal proteins (TcP) (1). Furthermore, anti-TcP antibodies are cross-reactive with human P proteins, suggesting that TcP's may contribute to the development of autoreactive antibodies in Chagas' disease patients (8). Unlike most other known eukaryotic cells, *T. cruzi* possesses four types of ribosomal P proteins (TcPO, TcP1, TcP2a and TcP2b), the complete encoding cDNA sequences for which have been reported (1, 8). Sequences for ribosomal

[1] Corresponding author. Fax: 61 + 7-33620104. email:berndK@qimr.edu.au.

0006-291X/96 $18.00

proteins of other protozoan parasites, including *Leishmania* spp., *Trypanosoma brucei* and *Babesia bovis* are also available (1, 8-10) but, as far as we are aware, none have been reported for any multicellular parasite. Herein, we report the cloning, expression and characterization of a ribosomal P2 protein from the pig tapeworm (cestode), *Taenia solium*, whose larval stages (cysticerci), as well as using pigs as intermediate hosts, also develop in humans in the subcutaneous tissues, heart, skeletal muscle, brain, and eyes and cause cysticercosis, a serious and chronic disease. When cysticerci are present in the central nervous system, neurocysticercosis results, a particularly dangerous form of the disease.

MATERIALS AND METHODS

Bacterial strains, phage, and growth conditions. Escherichia coli Y1090 [F-△ *lacU169 Ion- 100 araD139 strA hflA150::Tn10* (pMC9; Tet^r Amp^r)] and XL1-Blue [F′::Tn*10 proA*$^+$*B*$^+$ *lacI*^{q} △*(lacZ)M15/recA1 endA1 gyrA96* (Nal^r)*thi hsdR17* (r_K^- m_K^-) *supE44 relA1 lac*] cells were cultured and maintained in LB medium (1% Bacto-Tryptone, 0.5% yeast extract, 1% NaCl) supplemented with ampicillin (100 μg/ml). Before use in phage propagation, Y1090 cells were cultured in LB medium supplemented with 0.2% maltose, 10 mM $MgSO_4$ (pH 7.4) at 37°C with shaking. The cells were pelleted and resuspended in 10 mM $MgSO_4$ to an optical density at 600 nm (OD_{600}) of 0.5. Aliquots of the packaged λgt11 cDNA library containing approximately 10,000 pfu were mixed with 600 μl of Y1090 cells (OD_{600} = 0.5) and incubated for 15 min at 37°C. Three ml melted top agarose (0.7% agarose in LB) was added to each aliquot, plated on 150mm LB plates, and incubated at 42°C for 3.5 h for immunoscreening. After the plates were overlaid with nitrocellulose filters impregnated with 10 mM isopropyl-β-D-thiogalactopyranoside (IPTG), incubation was continued for a further 3 h at 37°C.

Screening of a T. solium λgt11 cDNA library and DNA sequencing. A *T. solium* cDNA library was constructed in λgt11 using Poly $(A)^+$ RNA isolated from metacestodes (cysticerci) from pigs of Mexican origin. A total of 6×10^5 recombinant pfu of the unamplified library were screened with a pool of human infection sera (HIS) of patients with confirmed neurocysticercosis. HIS was adsorbed against *E. coli* proteins before use in immunoscreening. Plaques adsorbed onto IPTG-impregnated nitrocellulose filters were rinsed with TBST (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.05% Tween 20,) and then blocked with 3% skimmed milk in TBS (50 mM Tris-HCl (pH 7.5), 150 mM NaCl,). The filters were incubated with HIS (diluted 1:100 in TBST) at 37°C for 1h and then washed three times in TBST. After washing, the filters were incubated at 37°C for 1 h with goat anti-human IgG conjugated with horseradish peroxidase, washed three times in TBST and once in TBS, and developed with 4-chloro-1-naphthol. Positive plaques were identified, and each was extracted with a sterile pasteur pipette and plaque purified by another round of screening (at a lower concentration so that individual plaques could be isolated). For sequencing purposes the λ-DNA was isolated using a commercial kit (Wizard Lambda Preps DNA Purification System, Promega), restricted with *Eco*RI, and subcloned into the unique *Eco*RI site of the sequencing vector M13mp18. Manual sequencing was performed with a M13/pUC sequencing primer using the Sequenase Version 2 kit (United States Biochemical Corporation). Analysis of nucleotide and deduced aminoacid sequences was performed using the GCG software package, Version 7.3.1., 1993, (Genetics Computer Group), and the EGCG software (EMBL), and using the GenBank (release 87) and SwissProt (release 31) data bases.

Screening of a T. solium λgt10 cDNA library by PCR. Over 40 clones, putatively expressing *T. solium* antigens, were isolated from the λgt11 cDNA library and one clone (L4RTS1) with a cDNA insert of 150 bp, which on sequencing was shown to have more than 90% homology with the family of P ribosomal proteins, was selected for further study. A reverse primer (5′-CTTTATTCACAAATAATTAGTCAAAGAG-3′) matching the C-terminal sequence of the λgt11 clone was synthesised using the β-cyanoethyl phosphoramidite procedure on an ABI model 308B DNA synthesiser. This oligo was paired with a λgt10 forward primer (Promega) (5′-AGCAAGTTCAGCCTGGTTAAG-3′) and used in PCR to amplify a full length acidic ribosomal phosphoprotein P2 (arpP2) gene from a *T. solium* cDNA library constructed in λgt10 (library kindly provided by Dr. P. Laclette). The PCR involved denaturation for 4 min at 94°C, followed by 35 cycles at 94°C for 40 sec, 60°C for 1 min, and 72°C for 1.5 min, followed by a final extension at 72°C for 15 min, using a Perkin Elmer Cetus, model 480, thermal cycler.

cDNA cloning and analysis. PCR products were fractionated by electrophoresis through 2% low-melting point agarose (3:1, NuSieve GTG:Sea Kem ME, FMC Bioproducts) in Tris acetate-EDTA buffer and stained with ethidium bromide. Products were isolated from gels using Wizard DNA Clean-Up Columns (Promega), ligated into plasmid pGEM-T (Promega) using T4 DNA ligase (Amersham), and used to transform *E. coli* strain XL1-Blue cells by electroporation. Transformed cells were cultured on LB-agar supplemented with ampicillin, 5-bromo-4-chloro-3-indole-β-D-galactopyranoside, and IPTG. Maxipreps of recombinant pGEM-T plasmids were prepared from bacterial cultures using Qiagen-500 columns (Qiagen). Nucleotide sequences of both strands of plasmid DNA were determined with universal vector primers as sequencing primers by the dideoxy method using the ABI Taq DyeDeoxy Terminator Cycle Sequencing System and an automated DNA sequencer (ABI, model 373A). Analysis of nucleotide and deduced amino acid sequences were performed as described above.

Subcloning, expression and purification of recombinant arpP2 protein. The QIAexpress system (Qiagen) was used for

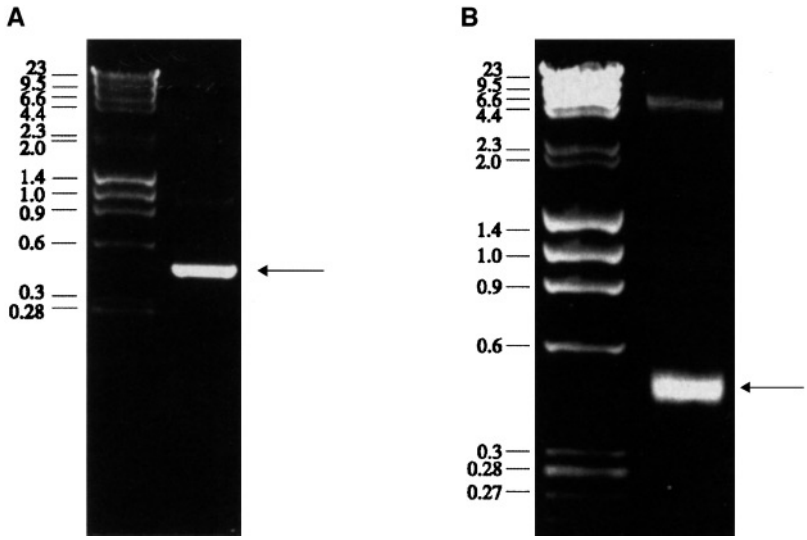

FIG. 1. Agarose gel electrophoresis of *T. solium* P protein cDNA. Molecular size markers are indicated in kilobases. (A) PCR product after amplifying a λgt10 cDNA library with a L4RTS1 specific reverse and a λgt10 forward primer. The arrow indicates the 450 bp amplicon. (B) Insert of the pGEM-T clone arpP2 after amplification with pGEM-T specific primers. The arrow indicates the 450 bp band consisting of the 417 bp arpP2 DNA plus additional 33 bases of λgt10 DNA.

high-level expression and purification of recombinant protein from *E. coli*. The pGEM-T plasmid DNA containing the full length insert was amplified by PCR using a forward primer that matched the N-terminal sequence of the arpP2 gene and incorporated a *BamHI* restriction site sequence (5′-TAATATGGATCCATGCGCTATCTCGCTGCTTATC-3′), paired with a reverse SP6 promoter primer. The *BamH*I and *Sal*I digested product was ligated into the *BamH*I and *Sal*I-digested pQE-30 vector (Qiagen) and used to transform *E. coli* XL1-Blue cells by electroporation. Clones were identified by restriction enzyme analysis and DNA sequencing. For expression, *E. coli* transformed with a recombinant clone was cultured at 37°C with vigourous shaking to an OD_{600} = 0.9, induced with 1 mM IPTG, and incubated for a further 4 hours. Cells were harvested by centrifugation and lysed by resuspending in 50 mM Tris-HCl (pH 8.0) supplemented with 1mg/ml lysozyme, incubating at room temperature for 1 h, freeze-thawing once, and sonicating at 20 kHz for 15 mins. After centrifugation the cleared lysate was mixed with Ni-NTA resin (Qiagen) and rotated for 1 h at room temperature. The resin was loaded into a 1-cm-diameter column and washed with 50 mM Tris-HCl (pH 8.0, flow rate 0.5 ml/min) until the OD_{280} was less than

```
  1  GCGGCCGCTGAACTACTATCGTGATTTCGTGAAA ATG CGC TAT CTC GCT GCT TAT CTT TTG GCC CAG
                                        Met Arg Tyr Leu Ala Ala Tyr Leu Leu Ala Gln

 68 CTT GGT GGC ACT GGT CGT CCT CAG GAG TCA GAC ATC AAG AAC ATC TTG AGC TCA GTT GGT
    Leu Gly Gly Thr Gly Arg Pro Gln Glu Ser Asp Ile Lys Asn Ile Leu Ser Ser Val Gly

128 GTC GAA TGC GAG CAG GAG CGA GCT AAG TTG GGG GTC GAT CAA TTG CAC GGC AAA AAT GTG
    Val Glu Cys Glu Gln Glu Arg Ala Lys Leu Gly Val Asp Gln Leu His Gly Lys Asn Val

188 CGT GAT TTG ATT AAC ACC GGC AAG GAG AAG ATG TCT GCA GTC TCT TTC GGT GCT ACT CCA
    Arg Asp Leu Ile Asn Thr Gly Lys Glu Lys Met Ser Ala Val Ser Phe Gly Ala Thr Pro

284 GTC GCT GTT CCA TCA GGT GGT GCT CCC GCT GCC GCT ACC GCT GCT GCT GAA GCT CCT AAA
    Val Ala Val Pro Ser Gly Gly Ala Pro Ala Ala Ala Thr Ala Ala Ala Glu Ala Pro Lys

308 GGT GGC GAC AAG GCT GCT GCT CCC CCC AAA GAG GAA AAG AAG GAG GAG TCT GAG GAA TCT
    Gly Gly Asp Lys Ala Ala Ala Pro Pro Lys Glu Glu Lys Lys Glu Glu Ser Glu Glu Ser

368 GAT GAC GAT ATG GGT TTC AGT CTC TTT GAC TAA TTATTTGTGAATAAAGA 417
    Asp Asp Asp Met Gly Phe Ser Leu Phe Asp *
```

FIG. 2. Nucleotide sequence and deduced amino acid sequence of clone arpP2. The stop codon is indicated by an asterisk.

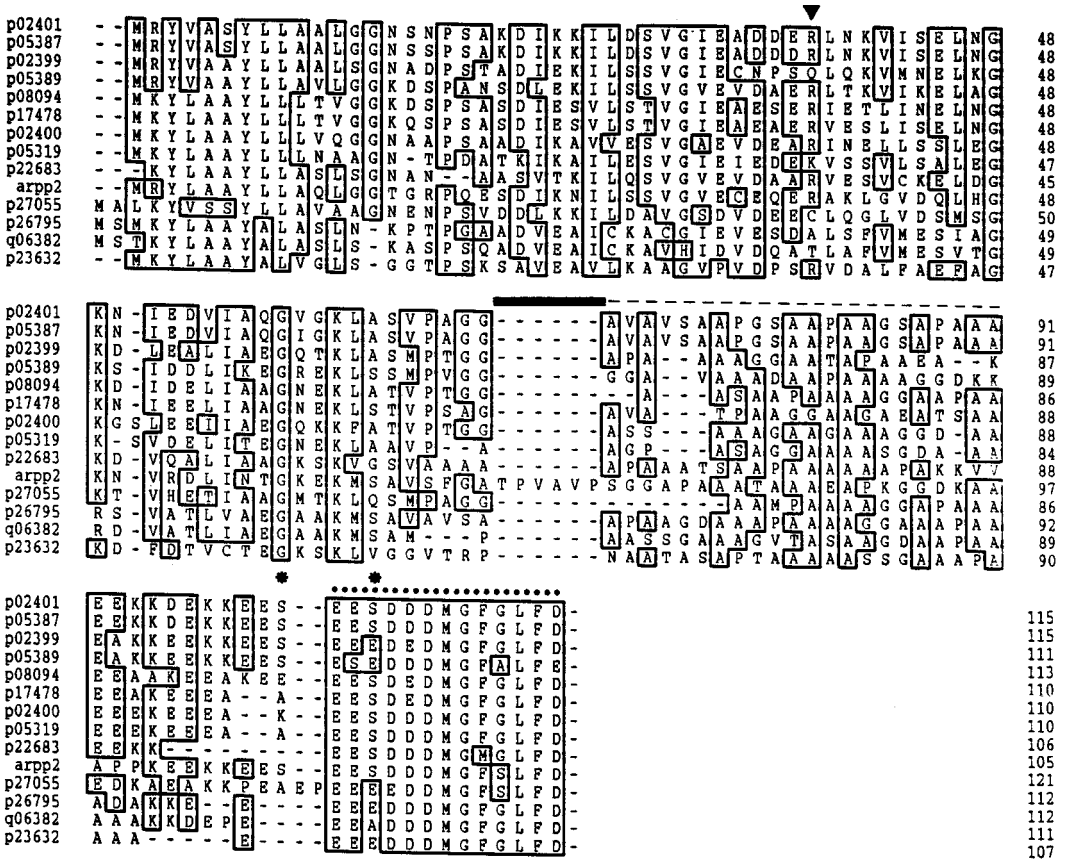

FIG. 3. Alignment of deduced amino acid sequences of ribosomal P proteins from *T. solium* (accession no. L39653), *Babesia bovis* (accession no. P27055), *T. cruzi* (accession no. P26795), *Leishmania infantum* (accession no. Q06382), *Dictyostelium discoideum* (accession no. P22683), *Saccharomyces cerevisiae* (accession no. P05319), *S. pombe* (accession no. P17478), *D. melanogaster* (accession no. P05389), *Artemia salina* (accession no. P02399), human (accession no. P05387), and rat (accession no. P02401). Blocks denote conserved residues and gaps have been introduced to maximise the alignment. Symbols on the first line of each alignment block designate the conserved arginine residue of ribosomal proteins P2 (solid triangle) and the phosphorylatable serine residues (asterisks). The 6 amino acid insertion of the *T. solium* arpP2 is indicated by a solid bar, the alanine-rich hinge region by a dashed line and the P consensus sequence by a dotted line.

0.01. The recombinant protein was eluted with a linear gradient to 0.25 M imidazole in 50 mM Tris-Hcl (pH 8.0). Protein was estimated using the Bradford assay (11).

RESULTS AND DISCUSSION

Cloning and DNA sequence analysis of the T. solium λgt11 L4RTS1 clone. Of the 6×10^5 plaques, 40 immunoreacted positively with HIS. All the positive plaques were picked and re-screened in a second round of immunoscreening to confirm positive signals and plaque-purify the clones. Eighteen clones were identified as definitely positive after the second screening. The λ DNA was isolated, subcloned into M13mp18 and sequenced from both ends. The resulting DNA sequences were used to search for homology in the GenBank database. One clone (L4RTS1) exhibited significant homology to the C-terminal end of the 60s acidic ribosomal phosphoprotein family of genes (P0, P1 and P2) from various species, in particular to the nucleotide sequence of chicken 60s acidic ribosomal phosphoprotein P1 (accession no. X13876). The alignment with this sequence showed that only a partial sequence of the *T. solium* P-protein had been cloned. In order to obtain a full-length cDNA clone, PCR was employed, using a *T. solium* λgt10 cDNA library as

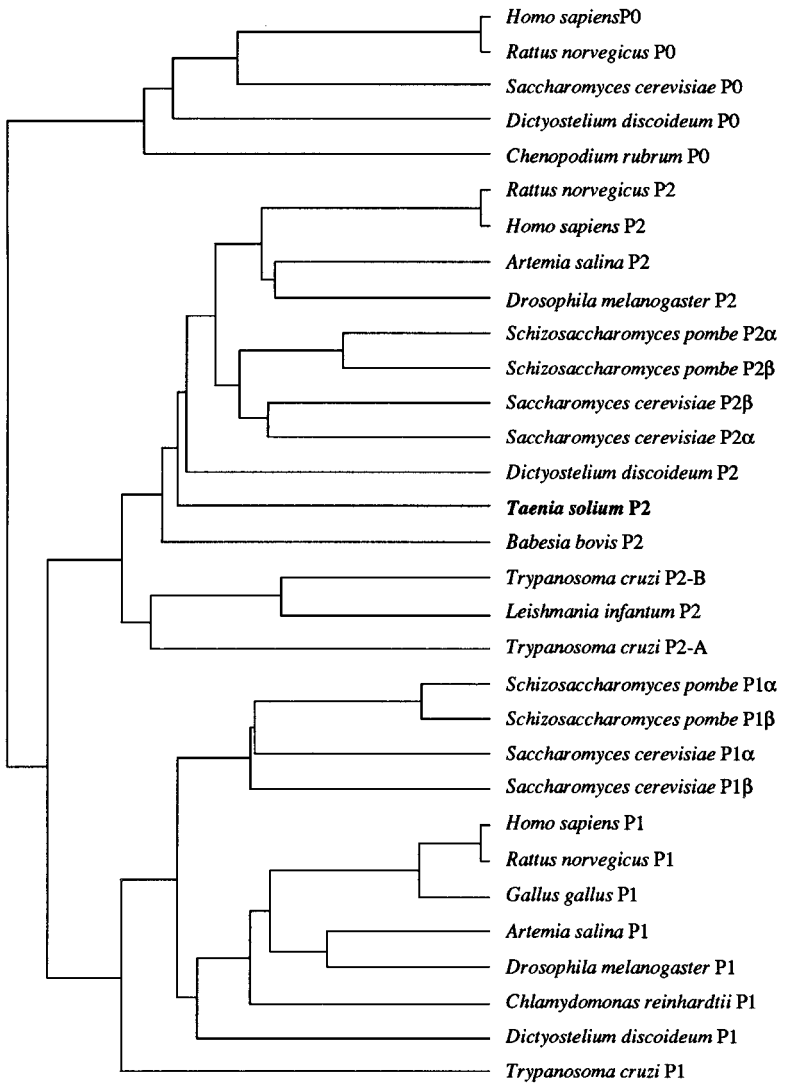

FIG. 4. Dendrogram of multiple pairwise alignments of deduced amino acid sequences of ribosomal P proteins from *T. solium* (highlighted) and other species producing clusters of similar sequences (GenBank accession nos. from the top: P05388, P19945, P05317, P22685, P29764, P02401, P05387, P02399, P05389, P08094, P17478, P02400, P05319, P22683, L39653, P27055, P26795, Q06382, P23632, P17476, P17477, P05318, P10622, P05386, P19944, P18660, P02402, P08570, P29763, P22684, P26643). The dendrogram predicts evolutionary relationships of arpP2 to other members of the P protein class (15).

template with a reverse primer that was designed to match the C-terminal sequence of the L4RTS1 clone in conjunction with a λgt10 forward primer. Figure 1A shows the size of the product, approximately 450 bp in size, amplified by using these primers. The cDNA was isolated from the gel and cloned into pGEM-T. Figure 1B shows the result of amplifying the insert from a selected recombinant plasmid (arpP2) in a PCR reaction with pGEM-T specific primers. The cDNA was sequenced from both ends and the resulting DNA sequences used to search for homology in the GenBank database. The clone exhibited, like the λgt11 partial clone L4RTS1, high homology to the 60s acidic ribosomal phosphoprotein family of genes (P0, P1 and P2). The full nucleotide sequence of arpP2 is shown in Fig. 2. The 417 bp cDNA contains an open reading frame beginning with the

initiation codon ATG at position 35, and ending with the termination codon TAA at position 400. When aligned with the cDNA sequences of rat P0 (accession no. X15096), mouse P0 (accession no. X15267), *Drosophila melanogaster* P2 (accession no. X05016), *Schizosaccharomyces pombe* P1 (accession no. M33137), and human P0 (accession no. H15026) proteins, identities of 67%, 70%, 67%, 69% and 65%, respectively were shown within the ORF.

Analysis of the protein structure. The ORF of arpP2 encodes a protein of 121 amino acids with a predicted molecular mass of 12,592 Da and a p*I* of 4.62, as inferred form the nucleotide sequence. This sequence was used to search for homology in the SwissProt database. In contrast to the search with the cDNA sequence, high homologies were found with protein sequences of P proteins of the P2 subclass. Alignment of the amino acid sequence with that of *Babesia bovis* (accession no. P27055), *T. cruzi* (accession no. P26795), *Leishmania infantum* (accession no. Q06382), *Dictyostelium discoideum* (accession no. P22683), *Saccharomyces cerevisiae* (accession no. P05319), *S. pombe* (accession no. P17478), *D. melanogaster* (accession no. P05389), *Artemia salina* (accession no. P02399), human (accession no. P05387), and rat (accession no. P02401) P2 proteins (Fig. 3) yielded identities of 38%, 42%, 36%, 43%, 47%, 48%, 55%, 47%, 47%, and 48%, respectively. The longest P2 proteins in the SwissProt database are 115 amino acids in length (human and rat), while the *T. solium* clone we have isolated is 6 residues longer. This is due to an insertion of 6 amino acids at positions 70-75 in the *T. solium* sequence (Fig. 3). This region comprises a highly variable fragment of the ribosomal P proteins (12, 13) and shows the lowest degree of homology in the alignment.

The overall structure of the *T. solium* P2 protein follows closely that described for eucaryotic ribosomal P proteins. It comprises a globular, N-terminal domain containing the central region of the protein, an alanine-rich hinge region, and a highly acidic, asparagine/glutamic acid-rich C-terminus, which contains the ribosomal P consensus sequence EESDDDMGFSLFD, but where the consensus glycine is exchanged for a serine (Fig 3). Like other P2 proteins the arpP2 contains a conserved arginine residue at position 38 of the alignment and two phosphorylatable serine residues in the C-terminal domain at positions 108 and 111 (Fig 3). The N-terminal region of the molecule is also highly conserved and shows only minor differences within the first 35 amino acids. This conservation may be related to the function of this domain of the protein which is responsible for the interaction with the body of the ribosomal major subunit (14). Since it has been considered that all ribosomal proteins P have evolved from a common ancestor (1), we constructed a dendrogram of P protein sequences in order to predict the phylogenetic relatedness of arpP2 to 30 other ribosomal P proteins (Fig. 4). Closest identities for arpP2 were with fungal P2 proteins (43%, 47% and 48%) and to the P2 protein of *B. bovis*, (38%) and *T. cruzi* (42%), and, accordingly, we consider the *T. solium* protein P2-like. By contrast, arpP2 showed less identity to P1 and P0 proteins from several species.

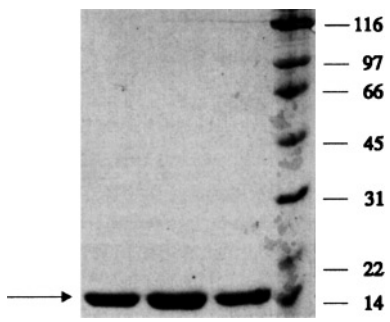

FIG. 5. SDS-PAGE analysis showing expression and purification of *T. solium* ribosomal protein P2. Molecular size markers are indicated in kilodaltons. Lanes 1–3 show consecutive fractions of the affinity purification.

Expression and purification of recombinant T. solium arpP2. After subcloning into pQE-30, and induction with IPTG, a protein of 15 kDa was highly expressed, that was not present in non-recombinant pQE-30 induced under the same conditions (data not shown). The 15 kDa protein is equivalent to the expected size of the full length recombinant fusion protein, comprising the 12.6 kDa arpP2 plus an additional 2 kDa for the N-terminal fusion peptide incorporating the six histidine residues required for purification. Purification by affinity chromatography under non-denaturing conditions yielded a single protein of 15 kDa (Fig. 5). The non-denaturing expression of arpP2 will facilitate studies on the *T. solium* recombinant acidic ribosomal phosphoprotein P2 to assess its potential as a diagnostic antigen and/or its potential role in auto immune disease, especially in patients with neurocysticercosis.

ACKNOWLEDGMENTS

We would like to thank Dr. Eduardo Garcia-Zepeda for identifying the L4RTS1 sequence in the λgt11 cDNA library. This work was supported by grants from the National Health and Medical Research Council of Australia and the Tropical Health Program and Australian Centre for International and Tropical Health and Nutrition. The DNA sequence of the *T. solium* arpP2 gene has been assigned GenBank accession number L39653.

REFERENCES

1. Levin, M. J., Vazquez, M., Kaplan, D., and Schijman, A. G. (1993) *Parasitology Today* **9,** 381–384.
2. Rich, B. E., and Steitz, J. A. (1987) *Mol. Cell. Biol.* **7,** 4065–4074.
3. Remacha, M., *et al.* (1988) *J. Biol. Chem.* **263,** 9094–9101.
4. Hunter-Newton, C., *et al.* (1990) *J. Bacteriol.* **172,** 579–588.
5. Naranda, T., and Ballesta, J. P. G. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 10563–10567.
6. Hasler, P., *et al.* (1991) *J. Biol. Chem.* **266,** 13815–13820.
7. Elkon, K., Bonfa, E., Llovet, R., Danho, W., Weissbach, H., and Brot, N. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5186–5189.
8. Skeiky, Y. A., Benson, D. R., Elwasila, M., Badaro, R., Burns, M. J., and Reed, S. G. (1994) *Infect. Immun.* **62,** 1643–1651.
9. Dalrymple, B. P., and Peters, J. M. (1992) *Nucleic Acid Res.* **20,** 2376.
10. Soto, M., Requena, J. M., Garcia, M., Gómez, L. C., Navarrete, I., and Alonso, C. (1993) *J. Biol. Chem.* **268,** 21835–21843.
11. Bradford, M. H. (1976) *Anal. Biochem.* **72,** 248–254.
12. Whittmann, H. G. (1982) *Annu. Rev. Biochem.* **51,** 155–183.
13. Ramirez, C., *et al.* (1989) *Can. J. Microbiol.* **35,** 234–244.
14. Uchiumi, T., Wahba, A. J., and Traut, R. R. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5580–5584.
15. Feng, D. F., and Doolittle, R. F. (1987) *Biochem J.* **290,** 205–218.