

MOLBIO 01965

## Characterization of the *Schistosoma mansoni* gene encoding the glycolytic enzyme, triosephosphate isomerase

Mitermayer G. dos Reis<sup>a,\*</sup>, Richard E. Davis<sup>b</sup>, Hardeep Singh<sup>b</sup>,  
Patrick J. Skelly<sup>a</sup> and Charles B. Shoemaker<sup>a</sup>

<sup>a</sup>Harvard School of Public Health, Department of Tropical Public Health, Boston, MA, USA; and

<sup>b</sup>Department of Biology, San Francisco State University, San Francisco, CA, USA

(Received 1 December 1992; accepted 9 February 1993)

The complete gene encoding *Schistosoma mansoni* triosephosphate isomerase (TPI) was isolated from a  $\lambda$  phage genomic library on 2 overlapping clones. These genomic clones have been characterized by restriction mapping and DNA sequencing of the 5' flanking region, the exons, the intron boundaries and the polyadenylation addition site. *S. mansoni* TPI is encoded by 6 exons spanning a region of about 12 kb. The 5 introns are located at positions precisely analogous to those of mammalian TPI genes but one of the 6 mammalian TPI introns is missing in *S. mansoni*. We find no evidence of spliced leader involvement in TPI gene expression. The gene is preceded by at least 4 tandem copies of a 2.5-kb repetitive sequence. While the 12-kb size for the *S. mansoni* TPI gene is much larger than the 3–4 kb typical of mammalian TPI genes, the 42-bp first intron is unusually short. The transcription initiation site for the *S. mansoni* TPI gene is heterogeneous. Genomic Southern blot analysis suggests that TPI is expressed from a single copy gene.

Key words: Schistosome; *Schistosoma mansoni*; TPI; Glycolysis; Intron; Gene

### Introduction

Because of its relative abundance, simple assay and central role in glycolysis, triose phosphate isomerase (TPI) has been extensively studied in terms of its biochemistry and its crystal structure [1]. The conserved nature of

its amino acid sequence has been established through characterization of TPI cDNAs from distant evolutionary sources. As genes encoding TPI were characterized from different organisms spanning a broad evolutionary range, it was noted that introns generally interrupt coding sequence between regions encoding distinct secondary structural features [2]. The intron positions are often identical, or near identical, even when comparing plant, fungal and animal species [3–5]. Based largely on TPI gene structure analysis, exons have been proposed to be protein coding modules that were joined together early in gene evolution during an 'exon shuffling' process [6].

The structure of genes encoding 'house-keeping' functions and the nature of their promoter elements within the animal phylum *Platyhelminthes* have not been extensively researched, yet several interesting features

*Correspondence address:* Charles B. Shoemaker, Harvard School of Public Health, Department of Tropical Public Health, 665 Huntington Avenue, Boston, MA 02115 USA. Tel.: 617-432-1339; Fax: 617-738-4914.

\**Present address:* Centrote Pesquisas Goncalo Moniz, FIOCRUZ/UFBA, Rua Valdemar Falcao, 121-Brotas, Salvador-Bahia-40000, Brazil.

*Note:* Nucleotide sequence data reported in this paper have been submitted to the GenBank™ data base with the accession numbers L06636, L07283, L07284, L07285 and L07286.

*Abbreviations:* TPI, triosephosphate isomerase; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends.

have already emerged. The single-copy hypoxanthine-guanine phosphoribosyltransferase gene of the parasitic blood fluke *Schistosoma mansoni* was shown to contain multiple introns, some of which are among the smallest yet characterized, and others many kilobases long [7]. An epidermal growth factor receptor homologue of *S. mansoni* also has introns ranging from extremely short to very large [8]. One exon of the *S. mansoni* 3-hydroxy-3-methylglutaryl-CoA gene can be spliced by either conventional cis-splicing or through the process of *trans*-splicing to a spliced leader [9]. In all cases, the intron/exon boundaries have sequences that fit the consensus established for eukaryotic splice junctions [10].

Since no *Platyhelminth* is, as yet, amenable to gene transfer studies, promoter analysis studies have been limited to determination of transcription initiation sites and searching upstream sequences for elements that resemble transcriptional promoters from other organisms. Through such analysis, several putative promoter elements have been reported [7,11–15].

In this paper, we report the characterization of the *S. mansoni* TPI gene. We include the DNA sequence upstream of the transcription start site, downstream of the polyadenylation signal and through all splice junctions. We find one intron of the 'mini-intron' variety and no evidence of spliced leader involvement. All introns are in coding DNA and are in the exact locations seen for mammalian and avian TPI introns. Interestingly, the one mammalian intron that is not found in the *S. mansoni* TPI gene is the only intron that is present within the coding region of the *Drosophila melanogaster* TPI gene [16].

## Materials and Methods

**Cloning and characterizing the *S. mansoni* TPI gene.** An *S. mansoni* genomic library in bacteriophage  $\lambda$  [8] was plated and screened with an *S. mansoni* TPI probe made from the previously cloned full-length cDNA [17] by a random primer labeling procedure using the

Multiprime kit (Amersham Corp., Arlington Heights, IL). Based on restriction mapping analysis and Southern blotting with synthetic oligonucleotide probes [18], appropriate DNA fragments were subcloned into M13 or pUC plasmid vectors prior to sequence analysis [19]. Sequence was frequently obtained using TPI oligonucleotide primers homologous to sequences within the exons.

**S1 protection studies.** An oligonucleotide primer (TPIS1-1; 5'-CGAGATCCAGACA-TTTCC-3') was prepared that anneals to *S. mansoni* TPI sense strand DNA surrounding the translation initiation codon. The primer was end-labeled and used to prime DNA synthesis from a single-stranded M13 DNA template (QZ3-1) containing the first two exons of the *S. mansoni* TPI gene and about 2 kb upstream. The reaction products were re-cut with *Hind*III and the 0.5 kb end-labeled S1 protection probe was purified from an alkaline gel. Probe preparation and S1 protection studies were performed essentially as described by Shoemaker and Mitsock [20] using 80  $\mu$ g of total *S. mansoni* adult RNA. Protected fragments were resolved on a sequencing gel alongside sequencing reactions in which the same primer and template were used as for the S1 probe preparation.

**Primer extension analysis.** Primer extension studies were performed as previously described [15] by annealing end-labeled oligonucleotides to adult total or poly (A)<sup>+</sup> RNA, extending the primer with Superscript II reverse transcriptase (Gibco BRL, Gaithersburg, MD), and resolving the extended products on denaturing polyacrylamide gels. Sequencing reactions employing the TPIS1-1 oligo and QZ3-1 template were used to define the length and termini of the extension products. Two TPI oligonucleotide primers were used in primer extension studies, TPIS1-1 and TPIS1-3 (5'-GAGCTTCTGAGAGTAAC-3', corresponding to a region 80 nucleotides further 3' in the mRNA).

**Rapid amplification of cDNA ends.** 5' RACE

was performed using two methods. Firstly, we used the 5' RACE System from Gibco BRL (Gaithersburg, MD) in which we followed the manufacturer's instructions. Secondly, we employed the method of Frohman et al. [21] as modified by Loh et al. [22]. Both methods produced similar results. cDNA was synthesized at 50°C from adult poly(A)<sup>+</sup> RNA with either random hexamer or oligo-dT primers and Superscript reverse transcriptase using the buffer and conditions recommended by the manufacturer (Gibco BRL, Gaithersburg, MD). The cDNA/RNA hybrids were hydrolyzed by treatment with 0.15 M NaOH at 65°C for 1 h before the sample was neutralized and extracted with phenol/chloroform. The primers were removed by 4 centrifugations in Centricon-100 microconcentrators (Amicon, Beverly, MA) and the cDNA was further purified and concentrated by ethanol precipitation. The cDNA was tailed for 1 h at 37°C using 1 mM dGTP and 12 units of deoxy-terminal nucleotide transferase (Promega, Madison, WI).

PCR was performed on the G-tailed cDNA using the anchor primer, 5'-TCTAGAAGTGTGGATCCCCCCCCCCCC-3' and the antisense TPI-specific oligonucleotide, 5'-CACGCTCTGATAATG-3' (TPI-E4A) that is homologous to TPI cDNA several hundred bases downstream from the translation initiation site. The PCR product was agarose gel purified (Magic PCR Prep, Promega, Madison, WI) and directly sequenced or cloned into a pT7Blue T-vector [23] (Novagen, Madison, WI) for sequence analysis. The purified 5' RACE PCR product was directly sequenced using the end-labeled TPIS1-3 primer and the fmol DNA Sequencing System (Promega, Madison, WI).

To validate the use of direct cycle sequencing of 5' RACE PCR products to characterize the 5' ends of *S. mansoni* mRNAs, we analyzed 3 mRNAs whose 5' ends had been previously characterized by primer extension sequencing [9,15,24]. Direct cycle sequencing of 5' RACE products identified the same 5' terminus and produced the identical sequence as previously defined by primer extension sequencing (R. Davis, unpublished).

*Analysis of S. mansoni TPI genomic DNA.* *S. mansoni* genomic DNA preparation, Southern blotting, and RNA preparation were all performed as previously described [8]. The probes were prepared by radiolabeling (as above) DNA fragments obtained by PCR amplification of *S. mansoni* TPI cDNA with specific oligonucleotide primers. To generate the DNA probe fragment for the genomic Southern, one PCR oligonucleotide matched sequence near the initiation codon (TPIPR; 5'-TCTGGATCTCGCAAATTT-3') and the opposing oligonucleotide was homologous to sequence near the 3' end of the third exon (TPI-E3B; 5'-CTACGCTCAGAATGTC-3').

To look for the presence of TPI mRNA having a spliced leader, PCR was performed using an adult cDNA template with a spliced leader specific primer (5'-AACCGTCACGG-TTTTACTC-3') and the TPI-E4A primer under the conditions previously reported [8]. To validate the negative results, several positive controls were employed. One positive control was the same PCR as above but with the TPIPR primer replacing the spliced leader primer. This control produced the TPI cDNA product of the expected size. Another positive control used the same cDNA template and the spliced leader oligonucleotide but employed a second oligonucleotide from within a different mRNA that is known to have a spliced leader (D. Davis, unpublished results). This reaction also produced an amplified cDNA product of the expected size.

## Results and Discussion

*Cloning and characterization of the S. mansoni TPI gene.* Approximately 200 000 recombinant phage from an *S. mansoni* genomic library were screened with a hybridization probe made from a full-length *S. mansoni* TPI cDNA [17]. Eight positively hybridizing phage plaques were identified from which 6 were purified for further characterization. Restriction mapping and Southern blotting were performed on the clones using oligonucleotide probes complementary to sequences at

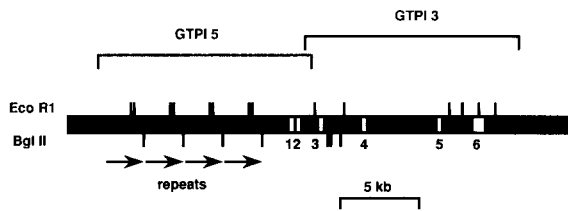


Fig. 1. Structure of the *S. mansoni* TPI gene. Exons, numbered 1–6, are shown as white blocks. *Eco*RI and *Bgl*II cleavage sites are indicated above and below the gene map, respectively. The approximate positions of the repeat element are represented by the arrows. The portions of the *S. mansoni* TPI gene encompassed by the two genomic clones, GTPI3 and GTPI5, are indicated by brackets.

various positions in TPI cDNA. Based on these studies, no single genomic clone contained the entire gene but 2 overlapping independent clones, GTPI5 and GTPI3 (Fig. 1), were shown to contain all of the exons of the *S. mansoni* TPI gene. With additional TPI oligonucleotide probes and Southern blotting, the location of each exon was determined within the restriction map of the two genomic clones. Restriction fragments containing exons were subcloned and the complete exon sequence was obtained, including the intron boundaries. Additional sequence was obtained for the regions upstream of the putative transcription start (cap) site (see below) and downstream of the polyadenylation signal. The results, shown in Fig. 2, demonstrate that the *S. mansoni* TPI gene contains 6 exons and 5 introns.

*S. mansoni* TPI gene structure analysis. Several observations can be made upon examination of the *S. mansoni* TPI gene structure. All of the intron/exon boundaries conform to the basic consensus for eukaryotic splice donor and acceptor sites [10]. The gene, at about 12 kb, is much larger than TPI genes reported from plants, fungi and animals [2–4,16] which are all less than 4 kb. The only other fully characterized *S. mansoni* gene for which homologous gene structures from other organisms are available encodes hypoxanthine-guanine phosphoribosyltransferase (HGPRT) [7]. In that case the *S. mansoni* gene is substantially smaller than its murine homo-

logue [25].

Another unusual feature of the *S. mansoni* TPI gene, and other *S. mansoni* genes, is the broad range of intron sizes. Although 4 of the 5 TPI gene introns are between 1 and 4 kb, the first intron is only 42 bp. This small intron is less than the minimum 80 bp reported to be required for intron splicing in mammals [26]. In contrast, the vast majority of introns in *Caenorhabditis elegans* are 40–60 bp in size [27] and are thought to employ fewer and/or smaller splicing components [28]. *S. mansoni* genes appear to have features of both mammalian and *C. elegans* genes. Of the *S. mansoni* genes characterized to date, most contain introns larger than 80 bp and many are substantially larger than 1 kb. But very small introns have been observed in both the *S. mansoni* HGPRT [7] and epidermal growth factor receptor homologue (SER) genes [8] and other *S. mansoni* genes (R. Davis, unpublished results). In TPI as well as the previously reported instances, the small introns are always the first, or first several introns in the gene. Thus, *S. mansoni* appears to be an organism that typically contains both large and small introns.

The positions of the introns within the *S. mansoni* TPI coding region are, in every case, identical to intron locations seen within mammalian, avian (chicken) and plant (maize) TPI genes [2,3,29]. One intron (intron 5) present in mammals and chickens is not present within *S. mansoni* or maize. Interestingly, this intron, lacking in the *S. mansoni* and maize genes, is the only intron that is present in the coding region of the *Drosophila* TPI gene [16]. *Aspergillus nidulans* has 5 introns within the TPI gene, but only the one corresponding to the third *S. mansoni* TPI intron is common to both species. Both *Aspergillus* and maize have several introns located at sites not found in the *S. mansoni* TPI gene but all at sites near the ends of secondary structure units, consistent with the model that an antecedent, primordial TPI gene was constructed of multiple protein coding modules separated by introns [2].

About 1 kb upstream of the TPI coding

```

CCAGACCCTTCAGATTATCTATATCACATTACAACCTAACGTATCAGCATTACCTAGTTCCAGAGAGTAATCA
ATAAATTTTCTTAAATCATGTCCATTAGTCTATGATGCTCATAGTCTATATCAACTGTTACTGTAGGAACTTCA
GACAAAATAGCACGGGCTTGCTATCAGGCAAGAATTATTTATCAACACGAGTTACCCAATTTGTTATCATATGAGT
AATTGGAGCATCTAACATAAGCGGTTGAGTATTAAGAAATTAGGTGAAATGATGATCTCTCTATACCAGCTACTT
GATTTTCGTCGGAAATAAACTACTAGGGTAATGGTTACTAACAGGAGATCAAATGTAATATTATTACTTATG
AACAACTTTGGTTACAATCAACTCAACGATATACTGTAACCTGAAGATTGTTTTAGTACTACTGAACTACTGAA
|---CAP region---|
CTTCTTCCTTAGAATATTAAGCGGTTGATCAAAGAGTCCCTCCAGTTAATAATCATATAGATTCCGTTACCAA
MetSerGlySerArgLysPhePheValGlyGlyAsnTrpLysMetAsnGlySerArgAspAspAsnAspLy
GGAAATGCTGGATCTCGCAAATTTTGTGGGGGAACTGGAAAATGAACGGAAAGCAGAGATGACAACGATAA
sLeuLeuLysLeuLeuSerGluAlaHisPheAspAspAsnThrG-----intron 1-----
CTTACTTAAGTTACTCTCAGAAGCTCATTTTGATGATAACACGGgtttgctcatgaagtcacgtacgaaaaatgct
-----luValLeuIleAlaProProSerValPheLeuHisGluIleArgLysSerLeuLysLysGluIle
ccccatttagAAGTTTTAATGTCTCCACCTTCAGTCTTTTTGCACGAAATTCGAAAAGTTTGAAGAAGAAATA
HisValAlaAlaGlnAsnCysTyrLysValSerLysGlyAlaPheThrGlyGluIleSe-----
CACGTGGCCGCTCAAAATTTGCTATAAAGTATCAAAGGGTGCATTCCACGGGAAATCAGgtaactgttcccatcc
-----intron 2-----rProAlaMetIleArgAspIleGlyCy
actgcct-----/ 1.5 kb /-----ttttcaacaatottgttttagCCAGCAATGATAAGAGATATTGGTTG
sAspTrpValIleLeuGlyHisSerGluArgArgAsnIlePheGlyGluSerAspGlu-----
TGATTGGGTCATACTTGGACATTTCTGAGCGTAGAAAACATTTTGGTGAATCTGATGAAgtaagtttttaactgt
-----intron 3-----LeuIleAlaGluLysValGlnHisAlaL
accatc-----/ 3 kb /-----ccatattttttataacaagCTTATTGCTGAAAAGTTCAACATGCAC
euAlaGluGlyLeuSerValIleAlaCysIleGlyGluThrLeuSerGluArgGluSerAsnLysThrGluGluV
TTGCTGAAGGCTAAGCGTTATTGCATGTATTGGTGAACATTTATCAGACCGTGAATCTAATAAAAACAGAGGAAG
alCysValArgGlnLeuLysAlaIleAlaA-----intron 4-----
TATCGCTTAGACAGTTAAAAGCTATTGCAAagtaagcttgcatgcctgcaggtc-----/ 4 kb /-----aa
-----snLysIleLysSerAlaAspGluTrpLysArgValValAlaTyrGluProVal
taatccatttccatttagATAAGATTAATCAGCTGATGAATGGAACGAGTAGTCGTAGCATATGAACCGATT
TrpAlaIleGlyThrGlyLysValAlaThrProJnGlnAlaGlnGluValHisAsnPheLeuArgLysTrpPhe
TGGGCTATTGGAACAGGTAAAGTTGCTACACCACAACAGCTCAAGAAGTTCATAATTCCTTCGTAATGGTTT
LysThrAsnAlaProAsnGlyValAspGluLysIleArgIleIleTyrGlyG-----
AAAACGAATGCACCAAAATGGAGTTGATGAAAAAATACGTATTATCTATGGTgtaagtcataatatttgattt
----intron 5-----lySerValThrAlaAlaAsnCysLysGluLeuAl
----/ 2 kb /-----ttttattactacacgcaagGATCTGTAACCTGCTGCCAATTTGAAAGAAATAGC
aGlnGlnHisAspValAspGlyPheLeuValGlyGlyAlaSerLeuLysProGluPheThrGluIleCysLysAl
TCAACAACATGATGTCGATGGATTTTGGTTGGTGGAGCTTCATTAACCAGGAATTCCTGAGATATGTAAGC
aArgGlnArg***
CAGACAACGTTGAACGTTTTTGTGGTCTTCTTCTTCTTACGCTTCTAAATTTTGTCTTCTCATGTCACTTACT
TGCTTTTTTGTCTTGTTTTTTGTTTTTGACAACCAAGCATATTATAATTAGTAGTATCTTAATGTCTAATGTA
AAAAAAAATATCACCACTAGTTCGAACATTTTAATTAATCTATCAGTCAAAAAAATAATATTGTCAATTTAGTTCA
TTGTTTCATTATCGTCCCTTCTAATTTAAATTTCTTATGGCCTAAATATACCTTGAATTAATAATTTAATAAGTT
TTGTAATAATTTTCACTGGATCTTACTTAATATAAAATGTACAAAGCTT

```

Fig. 2. DNA sequence of the *S. mansoni* TPI gene with coding region translation. Only the coding strand is shown and includes regions upstream of the cap site (indicated) and downstream of the polyadenylation site (indicated). Translated sequences are shown in three letter code above the coding DNA. Intron sequences are displayed in lower case. Where the entire intron sequence is not shown, the approximate size of the intron is given.

exons are multiple, identically spaced, *EcoRI* and *BglIII* restriction sites (Fig. 1) and *XbaI* sites (data not shown). The existence of these sites is consistent with the presence of an element of about 2.5 kb that is tandemly repeated at least four times upstream of the TPI gene. Detection of partial restriction digestion products in genomic Southern blots

supports the existence of the repeated element and suggests it exists as four tandem copies (data not shown). The significance of this element and its association with TPI gene function, if any, are not known.

The poly(A) signal for the *S. mansoni* TPI gene is not typical of poly(A) signals from other eukaryotes [30] as we do not find the

canonical AATAAA sequence 20–30 bp before the poly(A) addition site. The AATATA between 10–20 bp prior to poly(A) in *S. mansoni* may be an analogous signal although, due to the A/T rich nature of DNA in this organism, it could be random chance homology. While most *S. mansoni* cDNAs have been reported to have an AATAAA signal upstream of the poly(A) site, exceptions to the rule are common [8,31,32].

*Heterogeneous sites of TPI gene transcription initiation.* Several approaches were undertaken in an effort to determine the cap site for the *S. mansoni* TPI gene. S1 protection analysis (Fig. 3) and primer extension analysis (data not shown) both produced evidence for multiple initiation sites, most prominently in the region bracketed in Fig. 2, but with other weaker sites located nearby. For additional corroboration, the technique of rapid amplification of cDNA ends (RACE) [21,22] was employed. Direct sequencing of 5' RACE products (data not shown) confirmed that the cap site was heterogeneous with prominent sites located within the bracketed region shown in Fig. 2.

Recently it was shown that a subset of mRNAs from *S. mansoni* contain a 5' leader that is *trans*-spliced subsequent to transcription [9]. Sequence of the RACE product from *S. mansoni* TPI produced no evidence of a spliced leader on the TPI mRNA. To test whether a small subset of the TPI mRNA might contain the spliced leader, we used the PCR. The reaction employed an adult cDNA template with a sense strand oligonucleotide corresponding to spliced leader sequence and a TPI-specific anti-sense primer (see Materials and Methods). In several controlled attempts, we were unable to generate an amplified fragment by this reaction suggesting that the reported *S. mansoni* spliced leader is not joined to TPI mRNA.

Analysis of the sequence immediately upstream of the cap site, the region in which the promoter is predicted, reveals no clear evidence of a canonical TATA box [10]. Perhaps this is not surprising since the TPI promoter might be

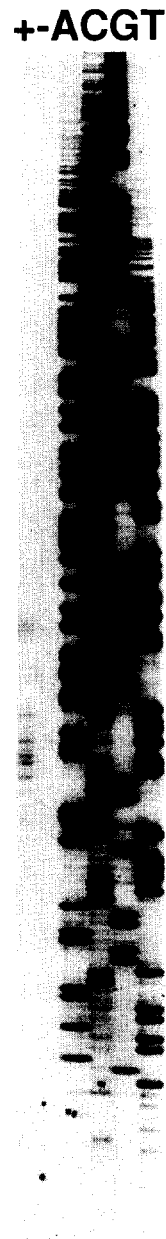


Fig. 3. S1 protection analysis of the *S. mansoni* TPI mRNA cap site. An end-labeled single-stranded DNA probe was prepared using an oligonucleotide primer (TPIS1-1) that is homologous to the sense strand near the translation initiation site (see Materials and Methods). The probe was annealed to total adult RNA, digested with S1 nuclease and the products were electrophoresed through a DNA sequencing gel (+). A control S1 digest without added RNA (-) and DNA sequencing reactions in which the TPIS1-1 primer was used with a template containing TPI genomic DNA (A,C,G,T) were also loaded onto the sequencing gel.

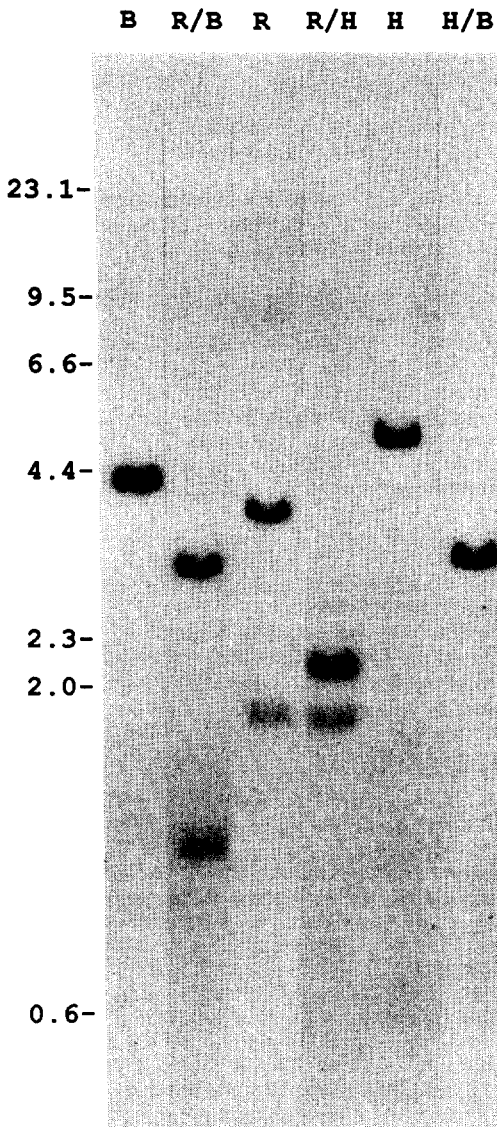


Fig. 4. Genomic Southern blot of exons 1–3 of the TPI gene. Genomic DNA, prepared from adult *S. mansoni* worms, was digested with various restriction enzymes (B, *Bgl*II; R, *Eco*RI; H, *Hind*III) and electrophoresed through an agarose gel. The resolved DNA fragments were transferred to a Nytran filter and hybridized to a radiolabeled cDNA probe specific for the first three exons of *S. mansoni* TPI. The positions of DNA size markers are indicated.

expected to have more in common with housekeeping gene promoters which are frequently characterized by a lack of the TATA box and the presence of a poorly defined cap site.

*S. mansoni* TPI is expressed from a single copy gene. Southern blot analysis was performed on restriction enzyme digests of *S. mansoni* genomic DNA to determine whether the TPI gene represented by the two genomic clones in Fig. 1 is a unique, single copy gene. As seen in Fig. 4, restriction fragments that hybridize to a TPI probe of exons 1–3 match the sizes expected based on the restriction map of the *S. mansoni* TPI genomic clones. The lack of additional hybridizing species, even when the filter is subjected to prolonged exposure (data not shown), strongly suggest that TPI is expressed from a single copy gene.

In summary, characterization of the *S. mansoni* TPI gene provides further evidence that schistosomes frequently employ an unusual combination of very small and large introns. All intron positions are identical to a subset of the introns found in both mammalian and plant TPI genes. This essential gene appears to be expressed from a single gene copy employing a heterogeneous cap site.

#### Acknowledgements

The authors wish to thank Tomas Sanchez for DNA sequencing assistance. These studies were funded, in part, by National Institutes of Health Grants 5RO1AI28499-02, United Nations Development Programme/World Bank/WHO Special Program for Research and Training in Tropical Diseases, the John and Catherine MacArthur Foundation and the TDR/Rockefeller Research Strengthening Program (C. S.). M. R. was supported by a fellowship from the Rockefeller Foundation. Additional support came from the National Institutes of Health 1R15A1327709-01 and the Office of Research and Sponsored Programs at San Francisco State University (R. D.).

#### References

- 1 Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G. (1975) Structure of chicken muscle triose

- phosphate isomerase determined crystallographically at 2.5 Å resolution using amino acid sequence data. *Nature* 255, 609–614.
- 2 Straus, D. and Gilbert, W. (1985) Genetic engineering in the precambrian: structure of the chicken triosephosphate isomerase gene. *Mol. Cell. Biol.* 5, 3497–3506.
  - 3 Brown, J.R., Daar, I.O., Krug, J.R. and Maquat, L.E. (1985) Characterization of the functional gene and several processed pseudogenes in the human triosephosphate isomerase gene family. *Mol. Cell. Biol.* 5, 1694–1706.
  - 4 Marchionni, M. and Gilbert, W. (1986) The triosephosphate isomerase gene from maize introns antedate the plant-animal divergence. *Cell* 46, 133–141.
  - 5 McKnight, G.L., O'Hara, P.J. and Parker, M.L. (1986) Nucleotide sequence of the triosephosphate isomerase gene from *Aspergillus nidulans*: implications for a differential loss of introns. *Cell* 46, 143–147.
  - 6 Dorit, R.L., Schoenbach, L. and Gilbert, W. (1990) How big is the universe of exons? *Science* 250, 1377–1382.
  - 7 Craig III, S.P., Muralidhar, M.G., McKerrow, J.H. and Wang, C.C. (1989) Evidence for a class of very small introns in the gene for hypoxanthine-guanine phosphoribosyl-transferase in *Schistosoma mansoni*. *Nucleic Acids Res.* 17, 1635–1647.
  - 8 Shoemaker, C.B., Ramachandran, H., Landa, A., Dos Reis, M.G. and Stein, L. (1992) Alternate splicing of the *Schistosoma mansoni* gene encoding a homologue of epidermal growth factor receptor. *Mol. Biochem. Parasitol.* 53, 17–32.
  - 9 Rajkovic, A., Davis, R.E., Simonsen, J.N. and Rottman, F.M. (1990) A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc. Natl. Acad. Sci. USA* 87, 8879–8883.
  - 10 Breathnach, R. and Chambon, P. (1981) Organization and expression of eukaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50, 349–383.
  - 11 Simurda, M.C., van Keulen, H., Rekosh, D.M. and LoVerde, P.T. (1988) *Schistosoma mansoni*: identification and analysis of an mRNA and a gene encoding superoxide dismutase (Cu/Zn). *Exp. Parasitol.* 67, 73–84.
  - 12 El-Sherbeini, M., Ramadan, N., Bostian, K.A. and Knopf, P.M. (1991) Cloning and sequence analysis of the *Schistosoma mansoni* membrane glycoprotein antigen gene GP22. *Mol. Biochem. Parasitol.* 49, 83–98.
  - 13 Ram, D., Grossman, Z., Markovics, A., Avivi, A., Ziv, E., Lantner, F. and Schechter, I. (1989) Rapid changes in the expression of a gene encoding a calcium-binding protein in *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 34, 167–176.
  - 14 Bobek, L.A., Rekosh, D.M. and LoVerde, P.T. (1988) Small gene family encoding an eggshell (chorion) protein of the human parasite *Schistosoma mansoni*. *Mol. Cell. Biol.* 8, 3008–3016.
  - 15 Davis, R.E., Davis, A.H., Carroll, S.M., Rajkovic, A. and Rottman, F.M. (1988) Tandemly repeated exons encode 81-base repeats in multiple, developmentally regulated *Schistosoma mansoni* transcripts. *Mol. Cell. Biol.* 8, 4745–4755.
  - 16 Shaw-Lee, R.L., Lissemore, J.L. and Sullivan, D.T. (1991) Structure and expression of the triose phosphate isomerase (Tpi) gene of *Drosophila melanogaster*. *Mol. Gen. Genet.* 230, 225–229.
  - 17 Shoemaker, C.B., Gross, A., Gebremichael, A. and Harn, D. (1992b) cDNA cloning and functional expression of the *Schistosoma mansoni* protective antigen, triose phosphate isomerase. *Proc. Natl. Acad. Sci. USA* 89, 1842–1846.
  - 18 Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning. A Laboratory Manual*. 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
  - 19 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
  - 20 Shoemaker, C.B. and Mitscock, L.D. (1986) Murine erythropoietin gene: Cloning, expression and human gene homology. *Mol. Cell. Biol.* 6, 849–858.
  - 21 Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* 85, 8998–9002.
  - 22 Loh, E.Y., Elliott, J.F., Cwirla, S., Lanier, L.L. and Davis, M.M. (1989) Polymerase chain reaction with single-sided specificity: analysis of T cell receptor chain. *Science* 243, 217–220.
  - 23 Marchuk, D., Drumm, M., Saulino, A. and Collins, F.S. (1990) Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.* 19, 1154.
  - 24 Meanaway, M.A., Aji, T., Phillips, N.F.B., Davis, R.E., Salata, R.A., Malhotra, I., McClain, D., Aikawa, M. and Davis, A.H. (1990) Definition of the complete *Schistosoma mansoni* hemoglobinase mRNA sequence and gene expression in developing parasites. *Am. J. Trop. Med Hyg.* 43, 67–78.
  - 25 Melton, D.W., Konecki, D.S., Brennand, J. and Caskey, C.T. (1984) Structure, expression, and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proc. Natl. Acad. Sci. USA* 81, 2147–2151.
  - 26 Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* 55, 1119–1150.
  - 27 Blumenthal, T. and Thomas, J. (1988) *Cis* and *trans* splicing in *C. elegans*. *Trends Genet.* 20, 305–307.
  - 28 Ogg, S.C., Anderson, P. and Wickens, M.P. (1990) Splicing of a *C. elegans* myosin pre-mRNA in a human nuclear extract. *Nucleic Acids Res.* 18, 143–149.
  - 29 Old, S.E. and Mohrenweiser, H.W. (1988) Nucleotide sequence of the triosephosphate isomerase gene from *Macaca mulatta*. *Nucleic Acids Res.* 16, 9055.
  - 30 Proudfoot, N. (1991) Poly(A) signals. *Cell* 64, 671–674.
  - 31 Laclette, J.P., Landa, A., Arcos, L., Willms, K., Davis, E. and Shoemaker, C.B. (1991) Paramyosin is the *Schistosoma mansoni* (trematoda) homologue of antigen B from *Taenia solium* (cestoda). *Mol. Biochem. Parasitol.* 44, 287–296.
  - 32 Wright, M.D., Henkle, K.J. and Mitchell, G.F. (1990) An immunogenic M<sub>r</sub> 3,000 integral membrane protein of *Schistosoma mansoni* worms that closely resembles a human tumor-associated antigen. *J. Immunol.* 144, 3195–3200.