Analysis of cDNA encoding the hypoxanthine-guanine phosphoribosyltransferase (HGPRTase) of *Schistosoma mansoni*; a putative target for chemotherapy

Sydney P.Craig III, James H.McKerrow[1], George R.Newport and Ching C.Wang

Department of Pharmaceutical Chemistry and [1]Department of Pathology, University of California, San Francisco, CA 94143, USA

ABSTRACT

Because of the lack of *de novo* purine biosynthesis, hypoxanthine-guanine phosphoribosyltransferase (HGPRTase) is a critical enzyme in the purine metabolic pathway of the human parasite, *Schistosoma mansoni*. Using a cDNA clone encoding mouse HGPRTase and subsequently a synthetic oligonucleotide derived from sequencing a clone of genomic DNA, two clones were isolated from an adult schistosome cDNA library. One clone is 1.374 Kilobases (Kb) long and has an open reading frame of 693 bases. The deduced 231 amino acid sequence has 47.9% identity in a 217 amino acid overlap with human HGPRTase. Northern blot analysis indicates that the full length of mRNA for the *S. mansoni* HGPRTase is 1.45-1.6 Kb. Analysis of the primary structures of the putative active site for human and parasite enzymes reveal specific differences which may eventually be exploitable in the design of drugs for the treatment of schistosomiasis.

INTRODUCTION

Many parasites are known to lack enzymes for *de novo* biosynthesis of purines (1-8). For this reason, enzymes involved in the salvage of purine bases and nucleo-sides have been proposed to be potential targets for antiparasitic chemotherapy (9-16). HGPRTase is of particular interest because it plays a major role in replenishing purine nucleotides for a number of parasites (6-8,17). Furthermore, the mammalian HGPRTase has been extensively studied because of its implication in gout (18), Lesch-Nyhan syndrome (19) and to its usefulness in the selection of human-mouse cell hybridomas (20). The primary structure for the HGPRTase of humans (21,22) as well as the deduced amino acid sequence (from cDNA sequence analysis) of the HGPRTase of mice and hamsters (23) have been reported. Hershey and Taylor (24) have aligned the deduced amino acid sequences of the mammalian HGPRTases with those deduced for the adenine phosphoribosyltransferase (APRTase) of *E. coli*, the APRTase of mice (25), and with bacterial glutamine phosphoribosyltransferase (26), xanthine-guanine phosphoribosyltransferase (27,28), and orotate phosphoribosyl-transferase (29). These alignments plus the results of genetic studies (30,31) permitted Hershey and Taylor to postulate specific sequences as the likely phospho-ribosylpyrophosphate (PRPP) and purine binding sites of HGPRTase.

Schistosoma mansoni is one of the three major species of trematode parasites that cause schistosomiasis (Bilharzia). This disease is endemic throughout most of the tropical world and affects over 200 million people, making it one of the world's greatest health problems (32). *S. mansoni* has long been known to lack the capability of *de novo* purine synthesis (2). Our recent studies on the schistosomula of *S. mansoni* indicate that the organisms depend primarily on salvaging purine bases to fulfill their requirement for purine nucleotides (8,15). Due to the limited interconversion between adenine and guanine nucleotides, the HGPRTase of *S. mansoni* provides the major source of guanine nucleotides for the parasite. An effective inhibition of this enzyme is anticipated to have a lethal effect on the parasite. This enzyme has been purified to homogeneity from adult worms of *S. mansoni* and demonstrates several properties which distinguish it from the enzymes of mammals (15). These observations and the availability of structural information from mouse and human clones encouraged us to attempt to clone and sequence the cDNA's encoding parasite HGPRTases with the aim of comparing the putative PRPP and purine binding sites of the host and parasite enzymes. A significant difference in the primary structures of these binding sites would justify further characterizations of the parasite enzyme for specific inhibitor design.

MATERIALS AND METHODS
Screening of *S. mansoni* cDNA libraries
Lambda gt11 *S. mansoni* cDNA libraries, prepared from adult worm poly (A) mRNA were prepared as previously described (33) the only modification being that after digestion with Eco R1, fragments larger than 1.5 Kb were enriched for by agarose gel electrophoresis. For screening, phage were plated at a density of 5-10 X 10^5 plaque forming units (pfu) per 150 mm plate on a lawn of *Escherichia coli*, strain Y-1090. Double nitrocellulose lifts were prepared according to standard methods (34,35). The filters were probed with a 691 base pair (bp) DdeI restriction fragment of the pHPT5 plasmid (23; kindly provided by Dr. Tom Caskey of Baylor College of Medicine in Houston, TX). This fragment contains the entire coding sequence for mouse HGPRTase plus 37 bp upstream of the start codon. The probe was radio-labeled using a nick translation kit from Bethesda Research Laboratories (BRL) and α-^{32}P-dCTP from Amersham. Specific radioactivity of the probe usually exceeded 10^7 cpm/μg DNA. Prehybridization and hybridization solutions contained 30% formamide (deionized), 5 X standard sodium citrate (SSC, 0.15 M NaCl, 0.015 M Na-citrate, pH 7.0), 0.1% sodium dodecyl sulfate (SDS), 1 mM EDTA, 10 mM Tris, pH 7.5, 1 X Denhardts' and 100 μg/ml heparin. Filters were prehybridized at least 1 hr before adding the freshly denatured nick translated probe to a concentration of more than 10^6 cpm/ml. Hybridization was allowed to proceed overnight at 37° C. The filters were

washed with 5 X SSC and 0.2% SDS at 42^0, blotted dry and autoradiographed with an intensifying screen at -70^0. Positive plaques were isolated and the phage were repeatedly plated and screened until 100% of the plaques on a plate hybridized with the probe.

Additional cDNA clones were isolated by using 20 base long oligonucleotides synthesized by the Biomolecular Resource Center at UCSF (see Fig. 1). These oligo-nucleotides were based on sequences identified during partial sequencing of a genomic clone (36). One hundred nanograms of the synthetic oligonucleotide was radiolabelled using 2 units of T_4 polynucleotide kinase from BRL with γ-^{32}P-ATP (3,000 Ci/mmole) from Amersham. Oligonucleotides thus labeled were separated from γ-^{32}P-ATP by centrifugation at 2,000 x g for 4 min at 4^0 through a G-25 Sephadex spin column from Boehringer/Mannheim. Hybridization with the labeled oligonucleotide was for 18 hours at 47^0 at a concentration of 10^6-10^7 cpm/ml in 6 X SSC, 10 X Denhardt's, 0.1% SDS, 0.1 mg/ml heparin, and 1 mM EDTA. The filters were washed with 5 changes of 6 X SSC, 0.1% SDS at 47^0 and autoradiographed.

<u>Purification of lambda phage DNA</u>

Lambda phage DNA was prepared by plating the phage at 10^5 pfu per plate on two 150 mm plates. Confluent lysis was achieved at 42^0 after approximately 12 hours. The phage were recovered by rotating the plates at 4^0 after the addition of 8-10 ml of SM buffer (35; 0.1 M NaCl, 8 mM $MgSO_4$, 0.2% gelatin, 50 mM Tris, pH 7.5). The buffer was collected after 12-16 hours and centrifuged at 2,000 x g for 5 minutes at 4^0. The pellets were discarded and DNase I (Sigma) and T_1 RNase (Boehringer/ Mannheim) were added to final concentrations of 0.3 µg/ml and 33 units/ml, respec-tively. After incubation at room temperature for 15 min, the sample was recentrifuged at 2,000 x g for 10 min at 4^0, and the supernatant was extracted with an equal volume of chloroform and recentrifuged at 2,000 x g for 15 min at 4^0. The supernatant was distributed equally among three SW 41 tubes (from Beckman), each containing 3.3 ml of 10% glycerol in TM buffer (35; 10 mM $MgSO_4$, 50 mM Tris, pH 7.8) overlaying a 3 ml cushion of 40% glycerol in TM buffer. After centrifugation for 1 hr at 35,000 rpm at 4^0 in a SW 41 rotor, the supernatants were removed and the pellets resuspended in a total volume of 300 µl of TM. DNase I and T_1 RNase were added to final concentra-tions of 5 µg/ml and 330 units/ml, respectively, and incubated at 37^0 for 30 min. EDTA (20 mM, pH 8), proteinase K (100 µg/ml), and SDS (0.5%) were added and followed by incubation at 65^0 for another hr. The sample was extracted successively with equal volumes of phenol, phenol/chloroform (1:1) and chloroform with 1 min microcentrifu-gations between extractions to separate phases. The DNA was precipitated from the aqueous phase by the addition of 0.5 volume of 7.5 M NH_4-acetate followed by an equal volume of isopropanol. The pellet was collected after 10 min at -20^0 by micro-

centrifugation and washed with 95% ethanol. Yields of phage DNA usually ranged from 20 to 200 μg.

Subcloning and sequencing cDNA inserts

The cDNA inserts were removed from the purified, cloned phage DNA by digestion with EcoR1 and were ligated into the dephosphorylated EcoR1 site of "Bluescript" from Stratagene. The plasmid DNA was used to transform *Escherichia coli*, strain dg98, and was purified from 500 ml liquid cultures of the transformed bacteria. The bacteria were harvested (2,000 x g for 10 min at 4°) and resuspended in 20 ml of the STET buffer (35; 0.1 M NaCl, 0.1 mM EDTA, 0.5% Triton X-100, and 10 mM Tris, pH 8.0) containing lysozyme at 0.5 mg/ml. The sample was rapidly brought to boiling over a Bunsen burner and maintained at 100° with gentle mixing for 2 additional min prior to cooling and centrifuging at approximately 30,000 x g for 30 min at 4°. Subsequent treatments with proteinase K, extraction with phenol, phenol/chloroform and chloroform, and banding in CsCl after intercalation with ethidium bromide were done according to standard methods (35). Yields of the plasmid DNA were usually about 1 μg/ml of the original bacterial culture. For sequencing the cDNA insert, the purified plasmid was annealed with SK and KS oligonucleotide primers (from Stratagene), and sequenced by the dideoxy chain termination method (37) using a "Sequenase" kit and protocols from United States Biochemicals. For these reactions, α-^{35}S-dATP (Amersham) was used for isotopic labeling and inosine triphosphate was substituted for GTP in the elongation reaction in order to reduce the effects of secondary structure on electrophoretic mobility (38-40). Complete sequencing of both strands of cDNA inserts was facilitated by the use of ten different 20 base long oligonucleotides prepared by the Biomolecular Resource Center at UCSF (see Fig. 1).

Probing of Northern blots

Total RNA was extracted from the hepatopancreas of *S. mansoni* infected *Biomphalaria glabrata* snails (the intermediate host for schistosomes) as previously described (41). RNA from uninfected snails was included as a control. The RNAs were resolved in MOPS-formaldehyde-agarose gels and transferred to nitrocellulose filters according to standard procedures (35). The blots were probed with the nick translated cDNA insert which had a specific radioactivity exceeding 10[7] cpm/μg DNA. Hybridizations were performed as described above except that the hybridization temperature was 55° and the filters were washed with 1 X SSC, 0.2% SDS at 65°.

Computer analysis of deduced protein structure

Predictions for the secondary structure of protein were performed according to Chou and Fassman (42,43) while hydrophobicity plots were generated using parameters established by Kyte and Doolittle (44).

RESULTS

The cDNA encoding *S. mansoni* HGPRTase

The use of a 691 bp cDNA fragment from pHPT5, encoding mouse HGPRTase, to probe for complementary sequences within a lambda gt11 *S. mansoni* cDNA library resulted in the identification of a single clone, designated Smc1, among approximately 200,000 phage plaques. The clone was purified to homogeneity, and the cDNA insert subcloned and sequenced (Fig. 1). Smc1 was then used to reprobe the same cDNA library under more stringent hybridization conditions (at 55°), yielding several homologous clones with partial cDNA inserts at a frequency of 1/50,000 phage plaques. Eventually, a 1374 bp cDNA clone, Smc2, was identified by screening the cDNA library with a 20 base synthetic oligonucleotide that is homologous to an upstream coding sequence (P in Fig. 1), deduced from sequencing a clone of genomic DNA (36). The two cDNA clones, Smc1 and Smc2, at 854 and 1374 bp in respective

AflIII 8
BsmI 106
TthIII2 109
HincII 178   XhoII 387
HaeI 191   AccI 486
ClaI 207   ClaI 581
AhaIII 241   MaeIII 626   DdeI 1044
HgiAI 287   BspMI 629   BclI 1121
XhoI 290   AhaIII 701   BclI 1278
SspI 811   SspI 1316

P   ORF-1   ORF-2

500   1,000
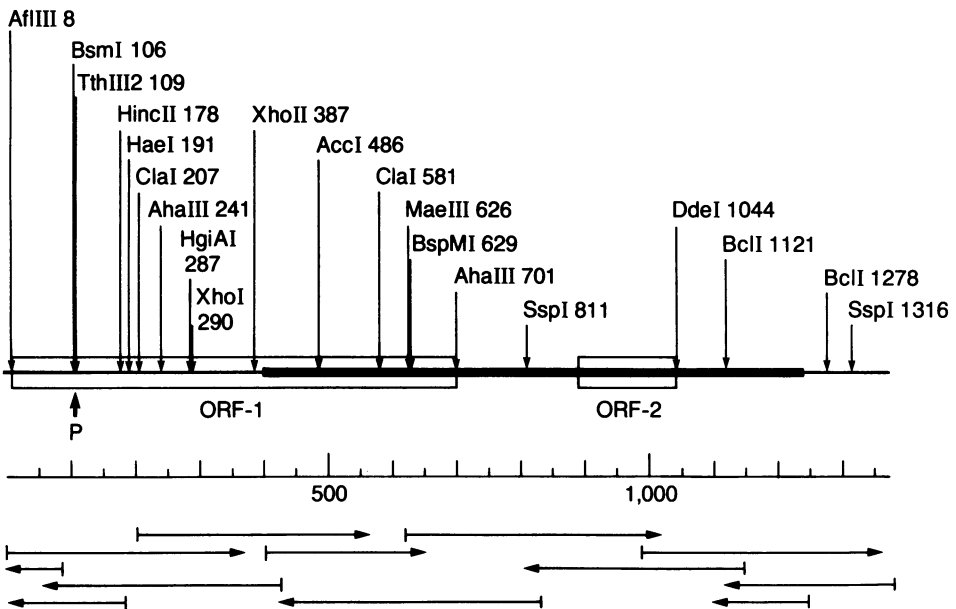
Figure 1. Map of two cDNA clones for schistosomal HGPRTase. The solid black line represents the full length of clone Smc2 at 1374 bp. The thicker black line shows the location of the 854 bp clone, Smc1, which starts at 400 bp from the 5' end of Smc2. ORF-1 codes for 231 amino acids (see Figs. 2 & 3) while ORF-2 may code for a smaller peptide of 51 amino acids. The P shows the position for the oligonucleotide that was used to probe the cDNA library leading to the isolation of Smc2. The arrows below the map show length and direction for sequences read.

```
  1-              Met Ser Ser Asn Met Ile Lys Ala Asp Cys Val Val
  1-   ATCCGCGAC  ATG TCT AGT AAC ATG ATA AAA GCT GAC TGT GTT GTG

 13-   Ile Glu Asp Ser Phe Arg Gly Phe Pro Thr Glu Tyr Phe Cys Thr
 46-   ATA GAA GAC AGT TTT CGA GGA TTT CCT ACG GAG TAT TTC TGC ACA

 28-   Ser Pro Arg Tyr Asp Glu Cys Leu Asp Tyr Val Leu Ile Pro Asn
 91-   TCT CCT CGG TAT GAC GAA TGC TTG GAT TAT GTT CTC ATA CCA AAT

 43-   Gly Met Ile Lys Asp Arg Leu Glu Lys Met Ser Met Asp Ile Val
136-   GGT ATG ATA AAA GAT AGG CTT GAA AAA ATG TCA ATG GAT ATT GTT

 58-   Asp Tyr Tyr Glu Ala Cys Asn Ala Thr Ser Ile Thr Leu Met Cys
181-   GAC TAT TAC GAG GCC TGT AAT GCG ACA TCG ATC ACA CTT ATG TGT

 73-   Val Leu Lys Gly Gly Phe Lys Phe Leu Ala Asp Leu Val Asp Gly
226-   GTC CTC AAA GGT GGA TTT AAA TTC CTT GCT GAT CTT GTT GAT GGG

 88-   Leu Glu Arg Thr Val Arg Ala Arg Gly Ile Val Leu Pro Met Ser
271-   CTT GAA CGC ACT GTC CGT GCT CGA GGT ATC GTC CTA CCA ATG TCC

103-   Val Glu Phe Val Arg Val Lys Ser Tyr Val Asn Asp Val Ser Ile
316-   GTT GAG TTT GTT CGT GTC AAG AGT TAT GTT AAT GAT GTC AGT ATT

118-   His Glu Pro Ile Leu Thr Gly Leu Gly Asp Pro Ser Glu Tyr Lys
361-   CAT GAA CCT ATA TTA ACT GGT TTA GGA GAT CCT TCG GAA TAC AAA

133-   Asp Lys Asn Val Leu Val Val Glu Asp Ile Ile Asp Thr Gly Lys
406-   GAT AAG AAT GTT CTT GTG GTC GAA GAT ATA ATT GAC ACA GGA AAA

148-   Thr Ile Thr Lys Leu Ile Ser His Leu Asp Ser Leu Ser Thr Lys
451-   ACA ATA ACG AAG CTC ATA AGC CAT TTG GAT AGT TTG TCT ACG AAA

163-   Ser Val Lys Val Ala Ser Leu Leu Val Lys Arg Thr Ser Pro Arg
496-   AGT GTT AAA GTC GCA AGC CTC CTC GTC AAG CGA ACA TCG CCC AGA

178-   Asn Asp Tyr Arg Pro Asp Phe Val Gly Phe Glu Val Pro Asn Arg
541-   AAT GAT TAC CGA CCA GAC TTT GTT GGT TTT GAA GTT CCA AAT CGA

193-   Phe Val Val Gly Tyr Ala Leu Asp Tyr Asn Asp Asn Phe Arg Asp
586-   TTT GTC GTT GGC TAT GCT TTA GAC TAT AAT GAT AAT TTC CGT GAC

208-   Leu His His Ile Cys Val Ile Asn Glu Val Gly Gln Lys Lys Phe
631-   CTG CAC CAT ATT TGC GTG ATT AAT GAA GTG GGT CAA AAA AAA TTC

223-   Ser Val Pro Cys Thr Ser Lys Pro Val OC
676-   TCT GTA CCC TGT ACA TCA AAA CCT GTT TAA ATGTTTTTTAGTTGAGATAAT

727-   AACAAGAAACTAAATAAATACAGAAACAGAATTTCCTTCAAAATGGTCATTCTGTAAATA
787-   ATTAATCAATATCACAGCCAAAAAAATATTTAACACTTCGTTGTTCTCTTCCAACTTCTT
847-   CTATACTTTTGTTATCCTATCTATACTATCTATACTACATATACTATGCTATCTATCCTA
907-   TCTATATACTATACTATCGTTAGTCTATACTTGTATAATCAACATATCTTCAATAAGATA
967-   TGTATTATTATCAACGAAACTGATTTTTCCATATCTATATATATTCTTAACTTATTTCTT
1027-  TTTTTTATATTTTATTTCTAAGATCATATTTGGCACACAAATGTTTTTCTTTCTTTTTTC
1087-  TCTTTTTTTTCTTGTTTTTACTCCTTTCATTTCCGTTGATCAATTGATTAATCAATTTAATT
1147-  GATTATTATTACACTTCTTATACTTCCTTCCCCTCCCCCAACTTCTATCATCAAAATAAA
1207-  ATGTTTCTTCATAAGAAAGATTCTAATGTTGTATGATGGACCAAAAAACAAAAAAAAACA
1267-  TTTTTCAATAATGATCATTAATTTCTAGTGTTATTTATTACTTATTAAAAATATTTTTTA
1327-  GTATTATTGTTATTAATAAAACTAAAATCGAAGAAAAAAAAAAAAAAA
```

```
                  Translated Mol. Weight = 26062.24
                  Codon usage:

                     7/UUU/phe    4/UCU/ser    7/UAU/tyr    4/UGU/cys
                     4/UUC/phe    1/UCC/ser    3/UAC/tyr    3/UGC/cys
                     3/UUA/leu    2/UCA/ser    1/UAA/OC     0/UGA/OP
                     3/UUG/leu    3/UCG/ser    0/UAG/AM     0/UGG/trp
                  --------------------------------------------------
                     6/CUU/leu    5/CCU/pro    3/CAU/his    3/CGU/arg
                     5/CUC/leu    2/CCC/pro    1/CAC/his    1/CGC/arg
                     1/CUA/leu    4/CCA/pro    1/CAA/gln    5/CGA/arg
                     1/CUG/leu    0/CCG/pro    0/CAG/gln    1/CGG/arg
                  --------------------------------------------------
                     5/AUU/ile    2/ACU/thr    9/AAU/asn    6/AGU/ser
                     2/AUC/ile    0/ACC/thr    1/AAC/asn    2/AGC/ser
                     8/AUA/ile    7/ACA/thr   12/AAA/lys    1/AGA/arg
                     7/AUG/met    3/ACG/thr    4/AAG/lys    1/AGG/arg
                  --------------------------------------------------
                    13/GUU/val    4/GCU/ala   12/GAU/asp    6/GGU/gly
                     9/GUC/val    1/GCC/ala    8/GAC/asp    1/GGC/gly
                     1/GUA/val    1/GCA/ala    9/GAA/glu    4/GGA/gly
                     4/GUG/val    1/GCG/ala    3/GAG/glu    1/GGG/gly
                  --------------------------------------------------
```

Figure 2. Translation and codon usage for the major open reading frame in clone Smc2. Possible poly (A) addition signals (AATAAA) are boxed and the repetitive sequences, TATCCTATCTATA, TATACTAT, and TCTATACT, are underlined.

length, were sequenced and the results are presented in Figs. 1 and 2. Since the entire sequence of Smc1 is duplicated within Smc2 (Fig. 1), much of the following data refers only to clone Smc2. Smc2 has a long open reading frame (ORF-1) starting with the 5'end, coding for a protein with 231 amino acid residues adding up to a molecular weight of 26,000 daltons (Fig. 2) and which possesses 47.9% identity in a 217 amino acid overlap with human HGPRTase and 36.9% identity in a 222 amino acid overlap with *Plasmodium falciparum* HGPRTase (see Fig. 4). When chemically similar amino acid substitutions are counted as no substitution, the amino acid sequence for the schistosomal HGPRTase has 67.3% and 59.4% agreement with the sequences for human and malarial HGPRTase, respectively. Codon usage appears to be non-random for some of the amino acids such as leucine, valine, proline, threonine, asparagine, arginine, and glycine. With the exception of valine, most of the codons seem to prefer the usage of A and T which is reflected in an AT content of 60.8% in the ORF-1 of Smc2 (Fig. 2).

Within the 3' tail of Smc2 there is a second open reading frame (ORF-2) beginning with a methionine and running for 51 amino acids (Fig. 1). The start codon for this open reading frame is in a favorable environment for initiation of translation as there is an A in the -3 and a C in the +4 position relative to the AUG codon (see 45,46). Thus, this open reading frame could theoretically be translated. However, the predicted amino acid sequence does not align significantly with any other proteins in the Dayhoff database using the Lipman & Pearson "dfastp" program (47). Also within the 3' tail, there are three AATAAA poly (A) addition signals (Fig. 2), two 13 base sequences, TATCCTATCTATA, and two 8 base sequences, TATACTAT and TCTATACT appearing 4 times each in a region 140-240 bases downstream from the stop codon. The significance of these repeated sequences is not known.

Hybridization of nick translated Smc1 with the Northern blots indicated that there may be a single species of mRNA encoding *S. mansoni* HGPRTase (Fig. 3). The size of the mRNA is estimated to be 1.45 to 1.60 Kb. Conservatively assuming that the mRNA may have a poly (A) tail of about 100 bases (16), whereas Smc2 has a size of 1374 bases and a poly (A) 3'-end of 15 bases, Smc2 may miss about 100 bases representing the 5'-end of the mRNA.

Analysis of the protein structure deduced from the cDNA

Alignments of predicted amino acid sequence (Fig. 4) for the coding region of Smc2 with sequences for the HGPRTases of humans (22) and the malarial parasite, *P. falciparum* (16), allow identification of putative purine and phosphoribosylpyro-phosphate binding sites in the parasite enzyme (24). Between amino acid positions 40 and 78, there is a predicted β–α–β structure in all three enzymes (Fig. 4). Many nucleotide binding proteins are known to possess this structure, which is characterized as a mononucleotide fold presumably involved in binding to PRPP among the
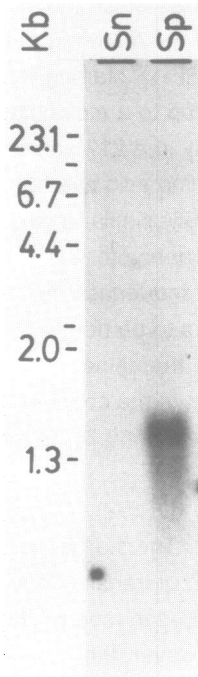
Figure 3. Northern blot of total RNA extracted from the hepatopancreas of infected (Sp) and non-infected (Sn) snails. Molecular weight markers include the 2.0-23.1 Kb Hind III digested lambda DNA fragments in addition to the 1.3 Kb mRNA for a schistosomal protease, the cercarial elastase (41).

phosphoribosyltransferases (48). Interestingly, the position and length of the predicted β–α–β structures is nearly identical for the human and schistosomal enzymes, even though primary structure has diverged extensively, with less than 11% identity within the amino acid sequences from position 54 to 76. Two amino acids from the end of the second β structure (at aa # 77) begins a 13 amino acid long sequence proposed to be a part of the purine binding site of mammalian enzymes (24). Within this region the schistosomal sequence differs from that of the human by only one nonconservative and one conservative substitution. On the other hand, the enzyme from *Plasmodium* differs substantially, particularly with the presence of the serine-arginine-glycine residues (position 82-84). A serine-arginine-glycine sequence also appears in the putative purine binding site of bacterial XGPRTase and murine APRTase (Fig. 5). All three HGPRTases appear to have a β turn at position 79 and the second half of the putative purine binding site is predicted to possess at least some α helical structure. The tyrosine and serine residues, also apparently important for purine binding (24,31) are found in all of the purine phosphoribosyl transferases at position 118-119 (Fig. 5) and are located just to the carboxyl side of a predicted β turn in the human, schistosomal, and malarial enzymes (Fig. 4). Although the serine residue at position

```
                                          <β>+<β><---- α --
H -           MATRSPGVVISDDEPGYDLDLFCIPNHYAEDLERVFIPHGLIMDR---TE
              ~  ~  ~|||  |    |    ~  ||    ~|  |  |~  |~||  |~|  ||
                                          <--β->+β><----α-----
S -     MSSNMIKADCVVIEDSFRGFPTEYFCTSPRYDECLDYVLIPNGMIKDRLEKMS
              ~  |  ~|~    |    ~~|~  ~~|    |  ||~|||~||  |    ~
                                          <---β--->+<-β--><- α --
P - MPIPNNPGAGENAFDPVFVKDDDGYDLDSFMIPAHYKKYLTKVLVPNGVIKNR---IE
    '          '         '          '          '          '
    1         10        20        30        40        50


  --α----->   <---β--->++  <------ α ----->  + +++   <---β--->++β>
  RLARDVMKEMGGHHIVALCVLKGGYKFFADL---LDYIKALNRNSDRSIPMTVDFIRLKSYC
    ~  |     ~  ~   |  ~~|||||  ||~|||   ~|  ~    |    ~||~|~|~|~|~|||~
  --α-----><----β---->+   <------ α ----->   <------β----->+<-β-
  MDIVDYYEACNATSITLMCVLKGGFKFLADL---VDGLERTVRARGIVLPMSVEFVRVKSYV
    ~  |    ~  |    ~  ~~|~|||   |~  |    ~   ~   ~   ~ ~   ~~~     |||||~
  ->          <--β->+ ++   <-α->  <β->+β><---α----> <-β->++β>+
  KLAYDIKKVYNNEEFHILCLLKGSRGFFTALLKHLSRIHNYSAVEMSKPLFGEHYVRVKSYC
    '          '          '          '          '          '          '
    60        70        80        90        100       110       120


  NDQSTGDIKVIGGDDLSTLTGKNVLIVEDIIDTGKTMQTLLSLVRQYNPKMVKVASLLVKRT
  ||  |   ~~  ~  |~  |~|    ||||~|||||||||||~~  |~|  ~  ~  ~  |  ||||||||||
  NDVSIHEPILTGLGDPSEYKDKNVLVVEDIIDTGKTITKLISHLDSLSTKSVKVASLLVKRT
  ||  |     ~   ~   |~|   |  |   ||~||||||||||~  |~~   |   ~    |~|  ~|  |~~|||
  NDQSTG-TLEIVSEDLSCLKGKHVLIVEDIIDTGKTLVKFCEYLKKFEIKTVAIACLFIKRT
    '          '          '          '          '          '
    130       140       150       160       170       180


  PRSVGYKPDFVGFEIPDKFVVGYALDYNEYFRDLNHVCVISETGKAKYKA
    |~||||||||~|  ~||||||||||~~||||  |~|||~|  |    |    ~
  SPRNDYRPDFVGFEVPNRFVVGYALDYNDNFRDLHHICVINEVGQKKFSVPCTSKPV
    ~  |   ~~|||||  ~|  ~|||||  ||||~  ||||  |~|~~|~  |  ||  ~
  PLWNGFKADFVGFSIPDHFVVGYSLDYNEIFRDLDHCCLVNDEGKKKYKATSL
    '          '          '          '          '          '
    190       200       210       220       230       239
```
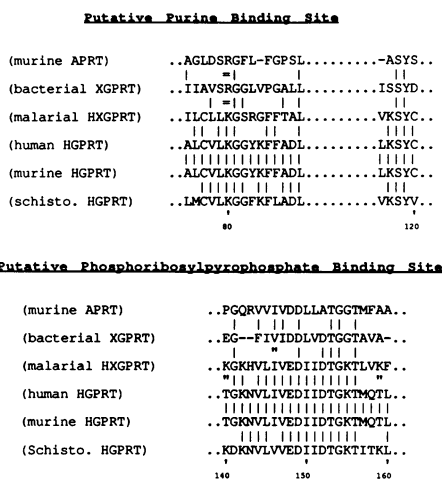
Figure 4. Alignment of amino acid sequences for the HGPRTases of humans (H, see 22), schistosomes (S), and *Plasmodium falciparum* (P, the etiologic agent for malaria, see 16). Vertical bars indicate where the human or malarial enzymes possess identical amino acids at that position in the sequence. The wavey lines indicate where the human and malarial enzymes possess amino acids with similar chemical properties. Regions with the symbol β represent areas of predicted beta secondary structure; those with α are predicted to be alpha helical; and + symbols show where there is a strong prediction for a beta turn (42,43).

124 of the HGPRTases is not found in other purine PRTases (after 24), it is probably involved with a β turn that can affect affinity for purine substrates (30; see Discussion).

A 15 amino acid segment at position 142-156 of the schistosomal enzyme aligns well with part of the putative phosphoribosylpyrophosphate binding site (24), and differs from the sequence in the human enzyme by only a single neutral amino acid substitution of valine for isoleucine (Fig. 4 & 5). In this same region the schistosomal enzyme differs from that of *Plasmodium* by the same neutral substitution plus a substitution of asparagine for histidine at position 143. Chou and Fasman analysis (42,43) predicts that the human, schistosomal and malarial enzymes would be in β conformation in this region with β turns just upstream at position 139, and with a β turn at position 152.

Putative Purine Binding Site

```
(murine APRT)      ..AGLDSRGFL-FGPSL.........-ASYS..
                     |   =|                   ||
(bacterial XGPRT)  ..IIAVSRGGLVPGALL.........ISSYD..
                     | =||    | |            ||
(malarial HXGPRT)  ..ILCLLKGSRGFFTAL.........VKSYC..
                     || |||  || |            ||||
(human HGPRT)      ..ALCVLKGGYKFFADL.........LKSYC..
                     |||||||||||||||         |||||
(murine HGPRT)     ..ALCVLKGGYKFFADL.........LKSYC..
                     |||||| || |||            |||
(schisto. HGPRT)   ..LMCVLKGGFKFLADL.........VKSYV..
                     '                        '
                     80                       120
```

Putative Phosphoribosylpyrophosphate Binding Site

```
(murine APRT)      ..PGQRVVIVDDLLATGGTMFAA..
                     |  | ||| |   ||| |
(bacterial XGPRT)  ..EG--FIVIDDLVDTGGTAVA-..
                     |   "  |  ||||| |
(malarial HXGPRT)  ..KGKHVLIVEDIIDTGKTLVKF..
                     "|| ||||||||||||||  "
(human HGPRT)      ..TGKNVLIVEDIIDTGKTMQTL..
                     |||||||||||||||||||||
(murine HGPRT)     ..TGKNVLIVEDIIDTGKTMQTL..
                     |||| |||||||||||   |
(Schisto. HGPRT)   ..KDKNVLVVEDIIDTGKTITKL..
                     '         '         '
                     140       150       160
```

Figure 5. Alignments of putative active sites for schistosomal HGPRTase with analogous regions of mouse APRTase (25), *E. coli* XGPRTase (27, 28), malarial HXGPRTase (16, 49), human HGPRTase (22), and mouse HGPRTase (23). Vertical bars show where the amino acids are identical with human HGPRTase. Quotation marks show where an amino acid differs with that for human HGPRTase but is identical with that for schistosomal HGPRTase. Equal sign denotes where positively charged arginine has been substituted for positively charged lysine.

There is an additional, well conserved region in the human, schistosomal and malarial enzymes near the carboxyl end (position 190-216). Within this region there is 81% identity in amino acid sequence for the human and schistosomal proteins and 70% identity for those of schistosomes and *Plasmodium*. Also, within this region there is only one nonconservative substitution between the sequence for the human and schistosomal enzymes while there are 4 between the sequence for schistosomes and *Plasmodium*. These areas have not formerly been identified as evolutionarily conserved. However, substitutions of neutral amino acids for the aspartates at positions 208 and 215 have been shown to alter the affinity of HGPRTase for both PRPP and purines (50,51).

Hydrophobicity analysis of the amino acid sequences for the enzymes of humans, schistosomes, and malaria indicate striking similarities with minor differences (Fig. 6). The most notable differences between the human and schistosomal enzymes are in positions 85-93 and 106-117 which are more hydrophillic for the human enzyme and positions 121-133 and 170-180 which are more hydrophillic for the schistosomal enzyme.
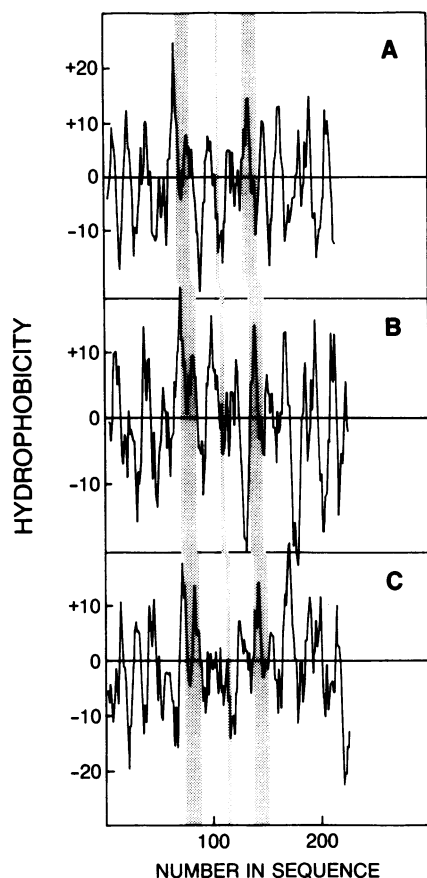
Figure 6. Hydrophobicity plots derived according to parameters established by Kyte and Doolittle (44) for human (A), schistosomal (B), and malarial (C) HGPRTases. The grey areas show the relative positions for the putative active sites shown in Fig. 5.

DISCUSSION

We have isolated and sequenced cDNA clones coding for the schistosomal HGPRTase. This allows predictions of the amino acid sequence for this pivotal enzyme in schistosome purine metabolism. One issue regarding the sequence is which one of the methionines at positions 6, 10 or any other upstream or downstream positions of Figure 4 is the actual amino terminus of the enzyme. Using DNA from a genomic clone of schistosomal HGPRTase (36), we have sequenced more than 450 bases upstream from where Smc2 ends. Assuming that this sequence does not contain any introns and will be found in the 5' end of a full length mRNA for HGPRTase, we have focused on the 160 bases immediately upstream from the putative start codon for the schistosomal HGPRTase identified in Fig. 3. Within this sequence there are two more AUG codons, both in the same reading frame with the sequence reported here

for the HGPRTase. The first is 54 and the second 18 amino acids upstream from the methionine at position 6. However, since the first AUG codon has a T in the +4 position and both of these codons have T's in the -3 position, they are considered to be in unfavorable environments to function as start codons, (45,46). Actually, the T in the +4 position relative to the methionine at position 6 is not considered to provide the most favorable possible environment for the initiation of translation. However, the methionine at position 10 is in a more favorable environment with A in both the -3 and +4 positions. Furthermore, sequences found within 50 bases of the start codon at position 6 (TCCT, TCCTT, and TCCGC) are complementary to sequences found at the 3' end of the 18s rRNA of a number of eucaryotes (52). That such sequences could play a role in the initiation of translation has formerly been proposed (52). Thus, although there remains the possibility that ribosomes would occasionally initiate translation at the two upstream methionines as well as the methionine at positions 6, methionine number 10 is more likely to be found at the amino terminus of the enzyme. The next methionine residue downstream is at position 49. We consider it unlikely to be the N-terminus because it is in the middle of a peptide sequence common to human, malarial and schistosomal HGPRTases.

Northern blot analysis indicates that the mRNA for *S. mansoni* HGPRTase is 1.45-1.6 Kb. Since Smc2 is 1,374 bp long with a poly (A) tail of only 15 bases, there could be about 100 bases of the mRNA sequence missing from the 5'-end of Smc2 assuming a poly (A) tail of about 100 bases in the mRNA. This evidence suggests that the methionine which appears in genomic sequence 54 amino acids upstream from the methionine at position 6, may be too far upstream to be found in the mRNA. However, the only way to determine with certainty the amino terminus for the schistosomal HGPRTase is to purify enough enzyme to permit sequencing of the amino teminus. This effort is currently underway in our laboratory.

Dovey *et al.* (15) reported that the subunit molecular weight for the HGPRTase of *S. mansoni* was 64,000 daltons by SDS-PAGE. The molecular weight deduced from cDNA sequencing is slightly over 26,000 daltons. Clearly the length for the mRNA as determined in Northern blots, which agrees well with the size for the mRNAs of mammals (23) and *P. falciparum* (16), does not allow for an enzyme subunit that could be much over 30,000 daltons. There are two possible explanations for this discrepancy; (1) that *S. mansoni* HGPRTase was present as a dimer form in SDS-PAGE; or (2) that the enzyme subunit was bound to another protein during SDS-PAGE. Multimerization of HGPRTase was in fact seen by activity staining of a native gradient gel. The native parasite enzyme had a molecular weight of 105,000 daltons (15), consistent with a tetramer of 26,000 dalton subunits. Multimerization also occurs with the native human HGPRTase (53,54).

Direct comparison of amino acid sequences for the HGPRTases of *Plasmodium* and humans yields 48% amino acid identity overall (16). Although this level of homology is identical to that for the schistosomal and human enzymes, divergence within specific regions of the sequences are dissimilar for *Plasmodium* and *Schistosoma*. The most intriguing area of dissimilarity is within the putative purine binding site (at amino acid positions 77-89 of Fig. 4). Within this region, the malarial enzyme differs significantly with the enzymes of both schistosomes and humans. In particular, the insertion of polar and charged serine and arginine residues at positions 82 and 83 constitute a major alteration in the environment. The significance of this dissimilarity may reside with the recent discovery that the malarial enzyme differs from the human and schistosomal HGPRTases by also recognizing xanthine as substrate, indicating that this enzyme is in fact an HXGPRTase (49).

Current efforts are focused on purifying more HGPRTase from adult worms in an effort to determine its amino terminal sequence, and to identify and clone the full-length cDNA for expression within bacteria or yeast. The enzyme thus generated will facilitate purification of the HGPRTase in large quantities and tests of drug sensitivity. Expression will also allow us to plan site directed mutagenesis for the purpose of studying substrate specificity of the enzyme. Also, we hope to crystallize the enzyme for X-ray analysis. These efforts should lead to a clearer picture of the tertiary structure of the active sites of parasite and human HGPRTases, thereby permitting the rational design of drugs to specifically inhibit them.

ACKNOWLEDGEMENTS

REFERENCES

1. Senft, A. W., Miech, R. P., Brown, P. R., and Senft, D. G. (1972) Int. J. Parasit. **2**, 249-260.
2. Marr, J. J., Berens, R. L. and Nelson, D. J. (1978) Biochim. Biophys. Acta **544**, 360-371.
3. Gutteridge, W. E. and Gaborek, M. (1979) Int. J. Biochem. **10**, 415-422.
4. Wang, C. C. and Simashkevich, P. M. (1981) Proc. Natl. Acad. Sci. USA **78**, 6618-6622.
5. Heyworth, P. G., Gutteridge, W. E. and Ginger, C. D. (1982) FEBS Lett. **141**, 106-110.

6. Wang, C. C., Verham, R., Rice, A., and Tzeng, S.-F. (1983) Mol. Biochem. Parasitol. **8**, 325-337.
7. Wang, C. C. and Aldritt, S. (1983) J. Exp. Med. **158**, 1703-1712.
8. Dovey, H. F., McKerrow, J. H., and Wang, C. C. (1984) Mol. Biochem. Parasitol. **11**, 157-167.
9. Marr, J. J., Berens, R. L. and Nelson, D. J. (1978) Science **201**, 1018-1020.
10. Nelson, D. J., LaFon, S. W., Tuttle, J. V., Miller, W. H., Miller, R. L., Krenitzky, T. A., Elion, G. B., Berens, R. L and Marr, J. J. (1979) J. Biol. Chem. **254**, 11544-11549.
11. Carson, D. A. and Chang, K.-P. (1981) Biochem. Biophys. Res. Commun. **100**, 1377-1383.
12. Rainey, P. and Santi, D. V. (1983) Proc. Natl. Acad. Sci. USA **80**, 288-292.
13. Rainey, P., Garrett, C. E. and Santi, D. V. (1983) Biochem. Pharmacol. **32**, 749-756.
14. Wang, C. C., Verham, R., Cheng, H.-W., Rice, A. and Wang, A. L. (1984) Biochem. Pharmacol. **33**, 1323-1329.
15. Dovey, H. F., McKerrow, J. H., Aldritt, S. M., and Wang, C. C. (1986) J. Biol. Chem. **261**, 944-948.
16. King, A. and Melton, D. W. (1987) Nucl. Acids Res. **15**, 10469-10481.
17. Reyes, P., Rathod, P. K., Sanches, D. J., Murema, J. E. F., Reikmann, K. H., and Heidrich, H. G. (1982) Mol. Biochem. Parasitol. **5**, 275-290.
18. Kelley, W. N., Greene, M. L., Rosenbloom, F. M., Henderson, J. F., and Seegmiller, J. E. (1969) Ann. Intern. Med. **70**, 155-206.
19. Seegmiller, J. E., Rosenbloom, F. M., and Kelley, W. N. (1967) Science **155**, 1682-1684.
20. Davidson, R. L. ed. (1974) Somatic Cell Hybridization, Raven Press, New York.
21. Wilson, J. M., Tarr, G. E., Mahoney, W. C., and Kelley, W. N. (1982) J. Biol. Chem. **257**, 10978-10985.
22. Jolly, D. J., Okayama, H., Berg, P., Esty, A. C., Filpula, D., Bohlen, P., Johnson, G. G., Shively, J. E., Hunkapillar, T., and Friedmann, T. (1983) Proc. Natl. Acad. Sci. U. S. A. **80**, 477-481.
23. Konecki, D. S., Brennand, J., Fuscoe, J. C., Caskey, C. T., and Chinault, A. C. (1982) Nucl. Acids Res. **10**, 6763-6775.
24. Hershey, H. V. and Taylor, M. W. (1986) Gene **43**, 287-293.
25. Dush, M. K., Sikela, J M., Sohaib, A. K., Tischfield, J. A., and Stambrook, P. J. (1985) Proc. Natl. Acad. Sci. USA **82**, 2731-2735.
26. Tso, J. Y., Zalkin, H., VanCleemput, M., Yanofsky, C., and Smith, J. M. (1982) J. Biol. Chem. **257**, 3525-3531.
27. Richardson, K. K., Fostel, J., and Skopek, T. R. (1983) Nuc. Acids Res. **11**, 8809-8816.
28. Pratt, D. and Subramani, S. (1983) Nucl. Acids Res. **11**, 8817-8823.
29. Poulsen, P., Jensen, K. F., Valentin-Hansen, P., Carlsson, P., and Lundberg, L. G. (1983) Eur. J. Biochem. **135**, 223-229.
30. Wilson, J. M., Tarr, G. E. and Kelley, W. N. (1983) Proc. Natl. Acad. Sci. USA **80**, 870-873.
31. Wilson, J. M. and Kelley, W. N. (1984) J. Biol. Chem. **259**, 27-30.
32. Mott, K. E. (1984) Schistosomiasis: New goals. World Health, World Health Organization, Geneva.
33. Newport, G. R., Hedstrom, R. C., Kallestad, J., Tarr, P., Klebanoff, S., and Agabian, N. (in press) Am. J. Trop. Med. Hyg. **38**,
34. Benton, W. D. and Davis, R. W. (1977) Science **196**, 180-182.
35. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning, A Laboratory Manual, Cold Spring Harbour Laboratory Press, New York.

36. Craig, S. P., Muralidhar, M. G., McKerrow, J. H. and Wang, C. C. (manuscript in preparation).
37. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
38. Barnes, W. M., Bevan, M. and Son, P. H. (1983) Methods in Enzymology **101**, 98-122.
39. Gough, J. A. and Murray, N. E. (1983) J. Mol. Biol. **166**, 1-19.
40. Mizusawa, S., Nishimura, S. and Seela, F. (1986) Nucleic Acids Research **14**, 1319-1324.
41. Newport, G. R., McKerrow, J. H., Hedstrom, R., Petitt, M., McGarrigle, L., Barr, P. J. and Agabian, N. (in press) J. Biol. Chem.
42. Chou, P. Y. and Fassman, G. D. (1974) Biochem. **13**, 211-221.
43. Chou, P. Y. and Fassman, G. D. (1974) Biochem. **13**, 222-245.
44. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105-132.
45. Kozak, M. (1986) Cell **44**, 283-292.
46. Kozak, M. (1986) Cell **47**, 481-483.
47. Lipman, D. J. and Pearson, W. R. (1985) Science, **22**, 1435-1441.
48. Argos, P., Hanei, M., Wilson, J. M. and Kelley, W. N. (1983) J. Biol. Chem. **258**, 6450-6457.
49. Queen, S. A., Jagt, D. V. and Reyes, P. (in press) Mol. Biochem. Parasitol..
50. Wilson, J. M. and Kelley, W. N. (1983) J. Clin. Invest. **71**, 1331-1335.
51. Melton, D. W., Konecki, D. S., Brennand, J. and Caskey, C. T. (1984) Proc. Natl. Acad. Sci. USA **81**, 2147-2151.
52. Hagenbuchle, O., Santer, M., Steitz, J. A. and Mans, R. J. (1978) Cell **13**, 551-563.
53. Holden, J. A., and Kelley, W. N. (1978) J. Biol. Chem. **253**, 4459-4463.
54. Johnson, G. G., Eisenberg, L. R., and Migeon, B. R. (1979) Science **203**, 174-176.