Short communication

Cloning and characterization of *Leishmania tarentolae* adenine phosphoribosyltransferase

Otavio H. Thiemann a, Juan D. Alfonzo b, Larry Simpson a,b,c,*

a Department of Molecular, Cell and Developmental Biology, Howard Hughes Medical Institute, UCLA School of Medicine, UCLA, Los Angeles, CA 90095-1662, USA
b Howard Hughes Medical Institute, UCLA School of Medicine, UCLA, Los Angeles, CA 90095-1662, USA
c Department of Medical Microbiology, Immunology and Molecular Genetics, Howard Hughes Medical Institute, UCLA School of Medicine, UCLA, Los Angeles, CA 90095-1662, USA

Received 15 June 1998; accepted 17 June 1998

* Corresponding author. Tel.: +1 310 8254215; fax: +1 310 2068967; e-mail: simpson@hhmi.ucla.edu

Keywords: *Leishmania tarentolae*; APRT; Adenine phosphoribosyltransferase; Kinetoplastid protozoa

The biosynthesis of purine, pyridine and pyrimidine nucleotides is catalyzed by enzymes of the phosphoribosyltransferase (PRTase) family. Most organisms synthesize adenine nucleotides by both the de novo and the salvage pathways. In contrast, all protozoan parasites examined to date are strict purine nucleotide auxotrophs because of the absence of a purine de novo biosynthetic pathway [1]. Kinetoplastid protozoa from the genus *Leishmania* possess three enzymes involved in the recycling of purine nucleotides by the salvage pathway, adenine PRTase (APRT) (EC 2.4.2.7), hypoxanthine-guanine PRTase (HGPRT) (EC 2.4.2.8) and xanthine PRTase (XPRT) (EC 2.4.2.22) [2]. APRT is responsible for catalyzing the conversion of adenine and D-5-phosphoribosyl 1-pyrophosphate (PRPP) into adenosine-5-monophosphate (AMP) and pyrophosphate (PPi) by the anomeric inversion of the ribofuranose ring [3]. The only aprt gene that has been cloned and characterized from a kinetoplastid protozoan is that from *Leishmania donovani* [4].

*Leishmania tarentolae*, a parasite of the gecko, has been exploited in our laboratory as a model trypanosomatid for a variety of molecular, biochemical and evolutionary studies. In this communication we describe the cloning and expression of the aprt gene from this organism. It has been recently shown that the lizard *Leishmania* forms a monophyletic clade branching between the *Viannia* and *Leishmania* subgenera [5,6]. These data confirm the inclusion of *L. tarentolae* in the *Leishmania* genus, which was previously
proposed by kinetoplast DNA and lipid analyses [7]. As in the case of other trypanosomatids, Leishmania tarentolae is a purine nucleotide auxotroph [8]. The ease of cell culture and genetic analysis of L. tarentolae should facilitate its use for site-directed mutagenesis of the aprt gene as well as for functional complementation and testing of inhibitory substrates. In view of the close phylogenetic relationship, the results should be of general significance for the pathogenic Leishmania species.

A 230 bp fragment of L. tarentolae aprt gene was PCR-amplified from genomic DNA with degenerate oligodeoxynucleotides for two highly conserved regions (DARGFLFG and VVLID- DVL) of known conserved regions (DARGFLFG and VVLID-). The DNA was PCR-amplified from genomic DNA with degenerate oligodeoxynucleotides for two highly conserved regions (DARGFLFG and VVLID-DVL) of known conserved regions (DARGFLFG and VVLID-DVL) of known aprt genes, S-2181 (5'-ACCGGAAATTCGATGCTCGCGGCTTCCTCTTG-CC-3') and S-2182 (5'-TGCTAAGCTTCCCACC- CGTGTGTTGCCAGGACGTCATCGATTAGCA-CCAC-3') (the underlines indicate added EcoRI and HindIII sites). The amplified DNA contained a deduced open reading frame of 76 amino acids with sequence identity to other APRT polypeptides. The full-length aprt gene was isolated from a L. tarentolae UC strain λZAP express BamHI-SalI 3A genomic library by hybridization screening with the PCR amplified L. tarentolae-specific probe. A 3 kbp genomic fragment was sequenced and found to contain a 708 nucleotide open reading frame that encodes a protein of 236 amino acids with a molecular weight of 25 kDa.

The 5' RACE of the aprt transcript using the 3' primer S-2488 (5'-CAGTAAAGAGTTGGTCAGAAAAGCTCCGACGGTCA-CCG-3') which is antisense to the aprt mRNA, and the 5' primer, S-2273 (5'-AACTAACGCTATATAAGTGATCAGTTTCTGTACTTTATTG-3'), which is specific to the L. tarentolae spliced leader RNA, indicated the presence of one splice acceptor site located at position -155 (SAS in Fig. 1) from the adenylate residue of the predicted methionine initiation codon (underlined in Fig. 1). The splice acceptor site is preceded by a polypyrimidine track of 28 pyrimidines at positions -238—-210. The aprt gene is preceded by an in-frame stop codon at position -294.

The mapping of the 3' UTR of the aprt transcript was performed by 3' RACE. cDNA synthesis from polyA+ RNA was performed with an oligo-dT primer. The cDNA was PCR amplified with the primers, S-2272 (5'-TTGAATTCTCGAGTACCTGC-3') and S-2181. An aliquot of the first PCR reaction was subjected to a second PCR amplification with the primers S-2272 and S-2184 (5'-ACCGGAAATTCGATGCTCGCGGCTTCCTCTTG-CCAC-3'). The amplified product was cloned into the pCR 2.1-TOPO vector (Invitrogen) and transformed into Escherichia coli DH5α. Sequencing identified a single polyadenylation site at position 856, which is 148 nucleotides from the TAG stop codon. These results would indicate a mature mRNA transcript of 1050 nt.

The L. donovani and L. tarentolae sequences share 87% amino acid identity with sequence divergence being mostly in the N- and C-terminus regions. Inspection of the sequence alignments in Fig. 2 reveals the conserved purine and PRPP binding domains [9] flanked by regions without significant sequence identity. The prediction of protein folding motifs from the individual polypeptides and the aligned sequences indicate a conservation of α-helical (H) and β-sheet (B) structural motifs (Fig. 2). The Leishmania sequences are ≈50 amino acids longer than the other APRT sequences, and the L. tarentolae sequence is 9 amino acid residues shorter at the C-terminus than the L. donovani homologue [4]. The function of these extended C-terminal sequences is not understood. A pairwise alignment of the different sequences revealed a 29–36% sequence identity with the Leishmania sequences (data not shown). The measurements of distances between the sequences showed that the APRT genes from Leishmania (L. tarentolae and L. donovani) together with the APRT of the methanogenic bacterium, Methanococcus janaschii, are the most rapidly evolving sequences within the APRT family (data not shown).

Southern analysis of genomic DNA indicated that the L. tarentolae aprt gene is single copy (data not shown). Contour-clamped homogeneous field electrophoresis analysis (CHEF), run in two different pulse and field strength conditions, was used to localize aprt to a 1.2 Mbp chromosome (data not shown).
Fig. 1. Nucleotide and predicted amino acid sequences of the *L. tarentolae* APRT gene. A 1246 bp sequence of the 3 kbp genomic fragment is shown. The nucleotides within the protein coding region and the predicted amino acid residues are enumerated from the methionine start codon (bold face underlined ATG). A termination codon (TAG) is indicated in bold at position 708. Nucleotides within the 5% untranslated region are numbered negatively starting from the start codon. The G residue corresponding to the sRNA splice-acceptor site is indicated (SAS). The tentative polypyrimidine tracks identified 5% to each splice-acceptor site are boxed, and the polyadenylation site identified in the transcript is indicated at position +856.

Northern analysis revealed two low abundance *aprt* transcripts of \( \approx 1.1 \) and 7.5 kb in length (data not shown). The 1.1 kb band is consistent with the mature *aprt* mRNA which was deduced from the sequence and from 5' and 3' RACE experiments. We speculate that this band represents an accumulation of pre-mRNA transcript, possibly previous to trans-splicing and/or polyadenylation.
Fig. 2. Multiple alignment of representative APRT sequences. The amino acid sequences of several APRT proteins are shown aligned with the *L. tarentolae* and *L. dono* sequences. The amino acid positions are indicated at the top of the alignment. The consensus of the alignment is indicated. * Indicates amino acid identities, : and . indicate conserved substitutions. The predicted secondary-structure fold, which was utilized as a guide to refine the sequence alignment, is shown: H, α-helices; L, loops; and B, β-sheets. The shaded boxes indicate the predicted purine binding domain and the boxed sequence indicates the predicted PRPP binding domain. The sequences are as follows: *Drosophila melanogaster* (APRT10, S34831), *Homo sapiens* (APRT14, P07741), *Saccharomyces cerevisiae* APT1 (APRT9, S49755) and APT2 (APRT16, L14434), *Rattus norvegicus* (APRT13, P36972), *Arabidopsis thaliana* APT1 (APRT15, P31166) and APT2 (APRT11, Q42563), *Triticum aestivum* (APRT2, U22442), *Leishmania dono* (APRT7, AF060886), *Escherichia coli* (APRT12, M14040, M25902), *Haemophilus influenzae* (APRT4, U32748, L42023), *Streptomyces coelicolor* (APRT3, X87267), *Cricetulus griseus* (APRT8, S36334), *Mastomys hildibrantii* (APRT1, U28722) and *L. tarentolae* (APRT6, AF060886). The sequences were aligned by the CLUSTAL-X program [10] with further manual refinements, taking into consideration the predicted secondary structures. The predict protein program (EBI at http://www.ebi.ac.uk/searches/searches.html) and the PHDsec program [11–13] were used to deduce the secondary structure predictions.
Fig. 3. Expression of *L. tarentolae* aprt in *E. coli* and purification of the recombinant protein. The *L. tarentolae* aprt gene was subcloned into the pQE-30 expression vector and overexpressed in *E. coli*. The proteins were separated in a SDS-PAGE gel. The 0 and 4 lanes represent lysates of non-induced and 4 h IPTG-induced *E. coli*, respectively. Sup., the supernatant of the 4 h-induced crude lysate. The 0–500 lanes represent fractions eluted from the Ni$^{2+}$-NTA column with increasing concentrations of imidazole (0–500 mM). M, molecular size markers in kDa.

The aprt open reading frame was amplified by a PCR reaction with oligodeoxynucleotides S-2284 (5'-CGTGGGATCCATGTCCCTCAAGGAAATCGGACCCAC-3') and S-2286 (5'-GCTATATTAACTTCTGGATCAGTAGATCGCGC-3') (added BamHI and HindIII restriction sites are underlined). The 728 bp PCR fragment was then cloned into the pQE-30 overexpression vector (Qiagen). The IPTG-induced transformed *E. coli* cells produced high levels of soluble, enzymatically active *L. tarentolae* APRT, which was purified to homogeneity on a Ni$^{2+}$-NTA column by elution with a step gradient of imidazole (Fig. 3). The expressed APRT eluted in the 100–150 mM imidazole fractions, together with enzymatic activity. The recombinant APRT protein with an N-terminal hexahistidine tag migrates as a 26 kDa protein in 10% SDS-PAGE. The recombinant APRT retained enzymatic activity and substrate specificity (data not shown). The purification of recombinant APRT protein from *L. donovani* has been previously reported [4] by a method which utilized the affinity of APRT for the AMP product and which required an enzymatically active protein. The ability to overexpress and purify the recombinant *L. tarentolae* protein in a single chromatographic step, independent of the active binding of the APRT to the substrate, should allow the purification of APRT mutants with different affinities for the substrate. This is a prerequisite for a structural-functional study of the enzyme.

Size exclusion chromatography of the purified *L. tarentolae* APRT indicates that the protein migrates as a dimer of 52 kDa in the absence of PRPP (data not shown). This is an interesting difference from the *L. donovani* homologue, which has been shown to elute as a monomer in the presence or absence of PRPP [4]. It is possible that the few amino acid differences seen in these two APRT proteins, principally at the N- and C-terminus sequences, may play a role in the formation of higher order structure.

Acknowledgements

The work was supported in part by a research grant AI-09102 to L.S. from the National Institutes of Health. We would like to thank the members of the Simpson laboratory for helpful discussions in the course of this work. We would like to express our gratitude to Dr James A. Lake for helpful discussion on the phylogeny of the APRT genes and to Dr Beatriz D. Lima for the *L. tarentolae* iZAP express genomic library.

References


