

Developmentally regulated gene from *Leishmania* encodes a putative membrane transport protein

(parasitic protozoan/gene expression/glucose transporter)

BRADLEY R. CAIRNS*[†], MICHAEL W. COLLARD[‡], AND SCOTT M. LANDFEAR*^{‡§}

*Department of Microbiology and Immunology and [†]Vollum Institute, The Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Road, Portland, OR 97201

Communicated by Harvey F. Lodish, June 29, 1989

ABSTRACT We have cloned a developmentally regulated gene from the parasitic protozoan *Leishmania enriettii*. The mRNA from this gene accumulates to a much higher level in the promastigote stage of the parasite life cycle that lives in the gut of the insect vector than in the amastigote stage of the parasite that lives inside the macrophages of the mammalian host. The predicted protein encoded by this gene is homologous to the human erythrocyte glucose transporter and to several sugar-transport proteins from *Escherichia coli*. These structural similarities strongly suggest that the cloned gene encodes a membrane transport protein that is developmentally induced when the parasite enters its insect vector. Regulated membrane transporters may be required for the parasite to adapt to the environment of the insect gut.

Leishmania are parasitic protozoa with a digenetic life cycle (1). Amastigotes are forms of the parasite that live within the macrophages of the mammalian host and are specifically adapted for growth and survival inside the macrophage lysosomes. These amastigotes are oval-shaped, nonmotile, obligate intracellular organisms. When a sandfly bites an infected host and takes a blood meal, it ingests the parasite-laden macrophages. Once inside the insect gut, the amastigotes are released from the macrophages and transform to the other stage of the life cycle, the promastigote. These promastigotes are elongated, flagellated, motile organisms that are specialized for colonization of the insect alimentary tract. The life cycle is completed when an infected sandfly bites another host and injects the promastigotes into the skin, where they invade macrophages and transform back into amastigotes.

Amastigotes and promastigotes are adapted to very different physiological environments. The macrophage lysosomes that host the amastigotes are acidic and are maintained at the temperature of the warm-blooded host. In contrast the insect gut has a pH close to neutrality, is lower in temperature than the mammalian host, and may contain different nutrient resources than the macrophage lysosome. A plausible hypothesis is that the parasite will possess certain developmentally regulated genes that are expressed exclusively or preferentially in one stage of the life cycle and that are involved in the adaptation of the microorganism to its changing milieu.

We report here the cloning of a developmentally regulated gene that is expressed much more abundantly in the promastigotes than in the amastigotes. From the sequence of this cloned gene[¶] we predict a protein product with significant structural similarity to several known membrane transport proteins, including the human erythrocyte glucose transporter and two bacterial sugar-transport proteins. These similarities suggest that the *Leishmania* gene also encodes a transporter. The ligand for this putative transporter, although

presently unknown, may be utilized preferentially by the parasite when it resides in the sandfly alimentary tract. This cloned gene may be useful for investigating the molecular mechanisms that control gene expression during the parasite life cycle.

MATERIALS AND METHODS

Growth of Organisms and Nucleic Acid Preparation. Promastigotes of *Leishmania enriettii* were cultured in Schneider's *Drosophila* medium containing 10% heat-inactivated fetal bovine serum, and amastigotes were grown in male Hartley guinea pigs, as described (2, 3). Genomic DNA and both total and polyadenylated RNA were isolated by published methods (2, 3).

Construction and Screening of cDNA and Genomic Libraries. A cDNA library was constructed in the vector λ gt11 by the protocol of Huyhn *et al.* (4). Aliquots of this library were plated and transferred to four replica nitrocellulose filters (5). Two identical replicas were hybridized with radiolabeled cDNA probe (6) templated from promastigote polyadenylated RNA, and the other two replicas were hybridized to labeled cDNA probe templated from amastigote polyadenylated RNA. Autoradiograms of the filters were inspected for plaques that hybridized to the promastigote cDNA probe but not to the amastigote probe. Inserts from these cDNA clones were subsequently subcloned (5) into the plasmid vector pBluescript SK(+) (Stratagene). A size-selected genomic library was generated by digesting promastigote genomic DNA to completion with *Cla* I, isolating a size fraction ranging from approximately 3.4 to 3.8 kilobases (kb) from an agarose gel (7), and cloning this fraction into *Cla* I-digested pBluescript SK(+). This library was screened (5) with the radiolabeled (8) insert from the Pro-c1 cDNA clone (see text).

DNA Sequencing. DNA sequencing was performed by the enzymatic method of Sanger *et al.* (9), using [α -³⁵S]thio-dATP (New England Nuclear) and the Sequenase kit (United States Biochemical). Deletions were performed with the Erase-a-Base kit (Promega) as recommended by the manufacturer. The entire 3592-base-pair (bp) insert of the Pro-g1 genomic clone was sequenced from both strands, using double-stranded plasmid DNA as template (10).

Computer-Assisted Data Analysis. The homology search of the National Biomedical Research Foundation (NBRF) data base (October 1988) employed the FASTP algorithm (11), and the RELATE algorithm (12) was performed on a Micro VAX II using the software provided by the Wisconsin Genetic Computer Group (13). For alignment of the protein sequences

Abbreviation: LTP, predicted *Leishmania* transport protein.

[†]Present address: Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305.

[§]To whom reprint requests should be addressed.

[¶]The sequence reported in this paper has been deposited in the GenBank data base (accession no. M26229).

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

(12), we used the Micro Genie software package and a General Electric personal computer.

RESULTS

Isolation of a Developmentally Regulated cDNA Clone. In our initial experiments, we isolated developmentally regulated clones by differentially screening (6) a promastigote cDNA library with probes made from promastigote and amastigote polyadenylated RNA. Aliquots of a λ gt11 cDNA library (4), constructed from promastigote polyadenylated RNA, were transferred to replica filters and hybridized in duplicate to radiolabeled (6) cDNA probes templated separately from promastigote and amastigote polyadenylated RNA. Plaques that hybridized intensely to the promastigote probe but not to the amastigote probe were selected as potential developmentally regulated clones. One such cDNA clone, Pro-c1, contained a 985-bp insert that hybridized to a single 3-kb band on a Northern blot of *Leishmania* RNA (Fig. 1A, lane p). Hybridization of the insert to equal amounts of promastigote and amastigote RNA confirmed that this mRNA is strongly developmentally regulated. The amastigote hybridization signal was barely detectable (Fig. 1A, lane a) under conditions that gave rise to an intense signal in the promastigote lane. Rehybridization of this Northern blot to a *Leishmania* rRNA clone (14) confirmed that both lanes of the blot contained equal amounts of total RNA (Fig. 1B).

Isolation and Sequencing of a Genomic Clone Encoding the Developmentally Regulated mRNA. Since the cDNA clone was not full-length, a direct route to obtaining the structure of the gene was to isolate a genomic clone homologous to the developmentally regulated cDNA. Because no intervening sequences have been found in any genes cloned from *Leishmania* or other related kinetoplastid protozoa, a genomic clone would be likely to contain a contiguous sequence for the protein-coding region of this mRNA.

Genomic Southern blots using the Pro-c1 cDNA insert as probe revealed that the sequence encoding this mRNA is arranged in tandem repeats containing at least seven copies of a 3.6-kb repeating unit (data not shown). This tandemly repeated arrangement of genes is common among the kinetoplastid protozoa and has been observed for several other gene families in both *Leishmania* and the related trypanosomes (15). To isolate a genomic clone representing one unit

of this repeat, we digested *L. enriettii* genomic DNA with the enzyme *Cla* I, which cleaves at a single site within the tandem repeat (data not shown), prepared a size-selected plasmid library, and screened this library with the Pro-c1 cDNA insert. One genomic clone, Pro-g1, was studied in detail.

The sequence of the Pro-g1 insert was determined in both strands (Fig. 2). Translation of this sequence in all reading frames predicts a single long open reading frame encoding a 567-amino acid polypeptide of 61,447 Da (Fig. 3). This predicted protein sequence is contained entirely within the region of the genomic clone that hybridizes contiguously to the mRNA, as determined by S1 nuclease mapping (20), and is encoded by the strand that does not hybridize to the 3-kb mRNA (data not shown).

Homology of Predicted Protein Sequence to Human Erythrocyte Glucose Transporter and to Bacterial Transport Proteins. An initial search of the NBRF protein data base (11) revealed that the predicted protein sequence in Fig. 3 bore

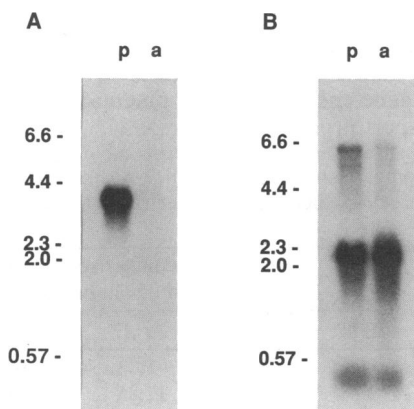


FIG. 1. Northern blot of promastigote (lane p) and amastigote (lane a) RNA probed with the insert from the Pro-c1 cDNA clone (A) and with a *Leishmania* rRNA clone (14) (B). Each lane contained 5 μ g of total RNA that had been resolved in a 1% agarose/formaldehyde gel (5). Numbers at left indicate the molecular sizes in kilobases of *Hind*III-digested phage λ DNA markers. The Pro-c1 probe from the initial hybridization was removed by boiling the blot in distilled water, and the filter was then rehybridized with the rRNA probe.

| | | | | | | |
|----------|------------|------------|------------|------------|------------|-------------|
| 1 | AGTACTGTCT | CTGCGTTGCC | GGGTATGAGC | GCGGATGAG | GATGCTGAGC | TGTAGAGTTG |
| 61 | CGTCCCGCT | CACCTCTCC | GGCATCCAT | GACCTGCCG | CTTCATTGTT | CCCTCCGGTT |
| 121 | CTCTTTGAC | TGCCCCCAC | CCACACGGC | GTTAACACTT | TCACTGGCCT | GCAGCTCTCT |
| 181 | CAGCAAGTAA | AACAAAACA | AGACCCCTC | TTTCAAGCTT | TGTTAGCCGT | TTGTGGCGA |
| 241 | CGCCCCACTG | CCCAACGCTG | TGCCCATCC | CCGCTAGTCT | GCCTTTCTTG | CATCAAAACC |
| 301 | ACTGAGAGCA | TTCACTAGAA | TCAATGAGCA | CAGGTCGAG | GTGAACGAGC | GCCGCTCGGA |
| M | | | | | | |
| 361 | CTCTGTATCG | GAGAANGAGC | CGGCCCGGA | CGATGCCAGG | AAGGAGCTCA | CGGACGACCA |
| 421 | AGAGACGGC | CCGCCGTTC | TGACGGCGA | TGACGGCGTA | GTCATCGGTA | TGCAGGCGAT |
| 481 | CGGTGCGAGC | CTGAACGGCT | ACTCGATCGG | CTTTGTCCGC | GTGTACTCAA | CGCTGTTTGG |
| 541 | CTACAGTACG | AACCTGCGCA | GCTTCTCCCA | GGAGAACAGC | TGCACGACGG | TGCCCAAGCG |
| 601 | TGACTGCAAG | TGGTTTGTGA | GCCCGACTGC | CAGCAGCTAC | TGTGGTGGC | CCGAGGTCC |
| 661 | GTGCCGCAAG | GAGTATGCTT | ACTCCAGTCC | TGCGGAGATG | CCAGGTGGCC | TTGCCGGTGG |
| 721 | CGAGGCAGAG | TGCGGTGCA | GGTGTGCTGA | CAGCAGCAGG | GAGTGCCAGA | ACCCGTCCGG |
| 781 | TACTCATCG | TGGAAAGCG | GATCTTTTGC | TGCGTCGATG | ATTGCCGGCT | GCCTGATCGG |
| 841 | CTCCGCTTTT | GCTGGTCCG | TTGCGTGAA | GATCGGTGCG | AGGCTCTCGT | CTCTGCTCGT |
| 901 | TGGTCTCGT | GCTGTTGGG | CGTCCGTGAT | GTACCACGGC | TGTTGCCGGG | CGGACGAGTT |
| 961 | TTGGTACTG | ATCGTGGGCC | GCTTCTGATC | TGCTCTGTTT | TTGGGGGTGA | TCGTGGTTGC |
| 1021 | GTGTCTGTG | TACACTGATC | AGAACGGCGA | CCCGAAGTGG | AAGCGCACCA | TTGGCGTATG |
| 1081 | GTTCACGTA | TTCACGACGT | TGGGCTATT | GCTCGCTGCG | CTGATGGGCG | TTGCCCTTGG |
| 1141 | CCAGAGCATC | CGGTTTGATC | ACGACGGAGA | CCAGAAGGTT | ATGGCCGCGA | TGACGGGCTC |
| 1201 | GTGCGTGTTC | TGCACCTGCT | TTAGTCTTCT | GACGTTGTTT | CTTGGGATTT | TGACGAGGGA |
| 1261 | ATCGCGCGCA | AAGTTCGACG | GGGTTGAGGA | GGGCCCGGCT | GAGCTGAACC | CGAGCGAGTA |
| 1321 | GCGTACTGTC | GAGATGATCC | CGGACTGCT | GATGGGCTGC | GTGATGGGCG | CCAGCTGCA |
| 1381 | GCTGACCGCC | ATCAACGCTG | TGATGAACTA | CGGCCGACG | ATCATGGGCA | GCCTTGGACT |
| 1441 | GGCACCCTG | GTGGCAACT | TGTTGTGAT | GCTGTGGAAC | TCGTGAGGCA | TCCTGTATC |
| 1501 | GATTCGCTT | TGTTATGTT | TCAAGATGG | CCACGTGTTT | CTGTTCCGGT | CGATCTTAC |
| 1561 | GTGCGTATG | TGCTTGTTC | TGTGGCGCAT | TGTTGGTATC | CTGTGGTGA | CAGAAGAGCT |
| 1621 | GGAGCGGAG | AACGCGGTG | CCATCACCGC | CATCTCCGTC | TTCATCTCG | GCTTTCGAGT |
| 1681 | GTGGTGAAG | CCGTGCTACT | ACGTTGCTAC | GAAGACATG | TTCCCGCCGC | TTCCCGCCGC |
| 1741 | CGGCGCGCC | TGTTTACGCG | AGTGGCTCA | GTTTCACTTC | AACCTGATCA | TTAAGTGTG |
| 1801 | CTATCCGAT | GGCAGGAAA | GCATCTCCG | CCGCCCTGCT | GGCAACGAC | GGCAACGCA |
| 1861 | GGCTGTGCC | TTCATCTTCT | TGTTGGCTC | TGGCTTCACT | TGCTTGGTTA | TCACGCTTT |
| 1921 | CTTCTGGAC | CCGTGGGAG | AGGACGCTG | AGGACGCTG | GTGGTCTCG | GTGGTCTCG |
| 1981 | GAAAGAGG | CTGAGCGAG | AGTCCATCG | GAACAGGCT | GACTGGCCG | AATGAGAGG |
| O | | | | | | |
| 2041 | AGGGTCACT | GCTCTCTCG | GCTGCTCC | CGTCCCGATC | CAAGGGTCTC | GTGTCGTGA |
| 2101 | CAAGACGGC | CTTGGGCGT | GACGGTGGT | GTGACTCCG | TGACTTCGCC | CGCGATCGG |
| 2161 | GGCAGCGCG | CATTTGCGC | CGCTCTTTT | TGTTGTCTT | GCCTTCCATG | AGACGGCTGG |
| 2221 | CGCGGTTGC | AATGCAACGA | CTGTGTGGC | TAGACAATG | TATACATAGT | CGTGTCTGC |
| 2281 | TGCGCCCGT | CGCTCTCTC | AGCTCAGATG | TAGACAACAG | GGTTTCCAGT | TTGTGAGAG |
| 2341 | GGACAGGGT | GTTGCTGCT | TCTTGGATG | GATACTCTC | TGCTTTCTG | CAGCGTTCAT |
| 2401 | CCCTCCGCC | CGCTTCTGT | GTCCTCTCT | TCCATAATG | CGCTCCGCG | GCCTGACTATG |
| 2461 | CGCGTCCG | CGCTCCGCA | GCCCGATGA | CTGTGATAG | CGGTAGGGG | TCCGCCACG |
| 2521 | CCCGTCCG | ATGCGCTTC | TTCACGCGC | TTCGTCCTT | TTACCCCTTT | ATCTTTAACA |
| 2581 | CACCAAAAA | ACGCTCTTT | CGGTAATGAT | ATTTTTTATG | TGGTAACGTC | CCCGGGGGC |
| 2641 | CTCTTCTCC | CCTCCGCC | CTGCGGCC | TTTCTGTAGT | GTCCGCTCGT | GTGTGTGTGT |
| 2701 | GCCACATGC | CCCAATCGG | TGTCTTTGC | CCTTCTTGG | CTCCGCGCA | CAGAGTGTCT |
| 2761 | TGCGATGCT | TCTGTCTTT | GCTCTGCTC | TCTTGTGGC | GTTGGGCCG | CCCTCACCC |
| 2821 | TCTTCTCAA | CCGCCCTCC | TGCGGGGCC | CGCTCAGAC | GCTTGTGCTC | TTTGCATAG |
| 2881 | GTGGAGCAG | GGCTGACGT | TGTGTTACT | GGGTCCGTT | CTGGGGTGC | TGCGGGTGT |
| 2941 | GGGATGTGA | GTCGCGGT | GAGTGGCGA | GGGGACGCA | ACGAGCGCA | GGGGCGGAG |
| 3001 | CATTCAAAG | ATGAAAAAT | CGACCACTT | CTCTCTTTT | TGCGGTGGG | TCCGAGATG |
| 3061 | CAGCGCCTT | TGCGACGAG | GCGATCCCG | TGCGAGCGG | CCATTTCCG | TGCACCGGG |
| 3121 | CCCGGTGCC | CCTTGACCG | GGTGGAGG | AGTGTGTGG | TAAAGATG | ATGTTGGCT |
| 3181 | TGTTTCACT | TGCCAAAAG | AAGTATAGC | GCATATTTCT | CTCTTCTGTT | TACTTCCCTC |
| 3241 | GCGCCTTCT | CAGGTGACTT | GCGGTGCCA | CGTACGAGC | CTCCGAAACA | CCGACAGGG |
| 3301 | TGGGGCTGC | GCCCTCTCAT | TTCAGCTGTG | TGAATGATG | GATGAAATG | CTTTCCCGT |
| 3361 | ACGCTTCTT | TCTTCTCCC | CTTCAGTGTG | CAGTGTACC | ATGCCACTG | AACGGTGTCT |
| 3421 | GTGAGGGGA | AGGACACATC | GCTACTCTG | CTGATTGTC | CTAGACTTC | CTGAGTGTCT |
| 3481 | GATGATGCT | TGATGAGCAA | AGACGCTCC | CCTCTCTTT | ATGCGTAGT | TTGAGGGAG |
| 3541 | GCCATGACA | ATCTGCCGT | GTGCCGGTT | TGACAGTCG | GTGCT | |

FIG. 2. DNA sequence for the insert of the Pro-g1 genomic clone. Nucleotides are numbered at left. The long open reading frame predicted from this sequence begins at the ATG codon [underlined methionine (M) codon] starting at nucleotide 323 and ends with the TGA opal (O) termination codon (underlined) that begins at nucleotide 2024. The complete amino acid sequence appears in Fig. 3. The ATG at position 323 was chosen as the initiation codon because it is the first ATG within the mRNA coding region and because the surrounding sequence fits most closely the eukaryotic consensus initiation sequence (16) CCRCCATG, where A occurs at position 3 more frequently than G. However, we cannot rule out the possibility that one of the methionine codons at nucleotide 440 or 464 is the true initiator.

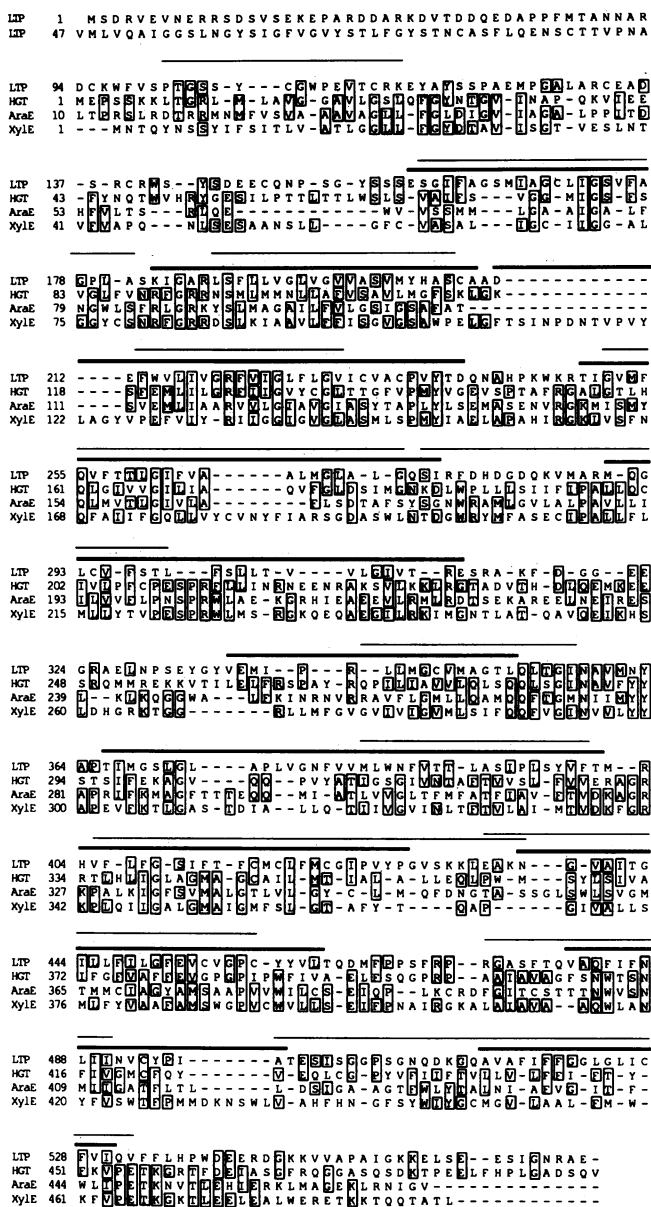


FIG. 3. Sequence homology among the predicted *Leishmania* transport protein (LTP), the human erythrocyte glucose transporter (HGT) (17), the *Escherichia coli* arabinose- H^+ (AraE) and xylose- H^+ (XylE) transporters (18). The LTP sequence (presented in single-letter amino acid code) was aligned to maximize identity with HGT. The other alignments are those reported in ref. 18. Boxes represent identities between any two sequences. Dashes represent gaps introduced to maximize alignment. The heavy lines indicate predicted transmembrane domains of LTP and the lighter lines indicate the predicted transmembrane regions of HGT, as determined by the algorithm of Klein *et al.* (19). The domain between residues 366 and 400 of LTP represents two overlapping segments, residues 366–384 (integral membrane segment) and 384–400 (possible integral membrane segment). Amino acids are numbered at left.

significant homology to the human erythrocyte glucose transporter (17), a protein that mediates the facilitated diffusion of glucose across the plasma membrane. A subsequent maximal alignment of the Pro-g1 protein sequence (hereafter referred to as LTP, for *Leishmania* transport protein) with the sequence of the erythrocyte glucose transporter revealed 21.7% identity between the two proteins over the 492-amino acid area of match (Fig. 3). When conservative amino acid substitutions were considered, a level of similarity of 44.4% was obtained, equivalent to the similarity of the glucose trans-

porter and several bacterial sugar-transport proteins (18), which are believed to have similar secondary and tertiary structures. Further, when the sequence of LTP was compared with two of the bacterial sugar-transport proteins (the arabinose- H^+ and xylose- H^+ transporters from *E. coli*, products of the *araE* and *xylE* genes, respectively; alignments are those reported in ref. 18), numerous identities arose between LTP and each *E. coli* protein (Fig. 3). In addition, there are 22 amino acids that are conserved in identity and relative position in all four of these proteins, consistent with the notion that these residues serve critical and related functions in all of these proteins (18).

A number of specific regions of the LTP sequence are noteworthy. Two blocks of sequence that are strongly conserved between LTP and the erythrocyte glucose transporter are Gly-Arg-Phe-Val-Ile-Gly and Gln-Leu-Thr-Gly-Ile-Asn-Ala-Val (residues 219–224 and 353–360 of LTP). Similarly, the (Arg/Lys)-Xaa-Gly-(Arg/Lys)₂ sequence (residues 89–93 and 330–334 of the erythrocyte glucose transporter) that is conserved and duplicated in the three known transporters (18) is moderately well conserved at a single location in LTP (residues 183–187 of LTP) as Lys-Ile-Gly-Ala-Arg. However, several regions that are highly conserved among the three known transporters are not conserved in LTP, most notably Pro-Glu-Ser-Pro-Arg and Pro-Glu-Thr-Lys-Gly (residues 208–212 and 454–458 of the erythrocyte glucose transporter). The N-linked glycosylation site that has been assigned to Asn-45 of the human glucose transporter is not conserved in the LTP sequence.

A statistical analysis of the degree of similarity between LTP and these known transporters, performed using the RELATE algorithm of Dayhoff *et al.* (12), revealed a high degree of relatedness among the sequences. Table 1 shows the segment comparison scores for pairwise comparisons of the LTP sequence with various known transporters. The value of ≈ 9 standard deviation units away from the mean for the comparison of LTP with the erythrocyte glucose transporter is well above the value of 4.75 that is required to group two sequences within the same superfamily; this value also predicts a probability of less than 10^{-19} that this relatedness occurs by chance. By the same criterion, LTP is in the same superfamily as the AraE and XylE proteins, but not the LacY permease, which is known to be structurally different from the AraE and XylE transporters (18). However, the sequences of the erythrocyte transporter, AraE, and XylE are more closely related to each other than they are to LTP (Table 1).

The indicated similarity between the LTP sequence and these three known transport proteins strongly suggests that the Pro-g1 gene encodes a developmentally regulated membrane transport protein.

Table 1. Comparison of LTP sequence with sequences of human erythrocyte glucose transporter (HGT) and *E. coli* transport proteins

| Sequence | Segment comparison score, standard deviation units | | | | |
|----------|--|-------|-------|-------|------|
| | LTP | HGT | AraE | XylE | LacY |
| LTP | — | 9.08 | 5.33 | 7.03 | 4.24 |
| HGT | 9.08 | — | 12.85 | 14.44 | 1.06 |
| AraE | 5.33 | 12.85 | — | 21.21 | 2.04 |
| XylE | 7.03 | 14.44 | 21.21 | — | 1.54 |
| LacY | 4.24 | 1.06 | 2.04 | 1.54 | — |

The sequences of the four known transporter proteins were compared to LTP and to one another by using the RELATE algorithm (12). The fragment length compared was 30 amino acids, and the number of random runs per comparison was 100. Two proteins are considered to be in the same superfamily if their segment comparison score is >4.75 (12).

Hydropathy Profile of the Predicted Protein Sequence and Membrane Orientation. Mueckler *et al.* (17) have presented a model for the orientation of the erythrocyte glucose transporter in the plasma membrane based primarily upon the hydropathy profile (21) of the protein sequence. In this model, 12 hydrophobic stretches of 21 amino acids are predicted to form membrane-spanning α -helices that pass back and forth through the membrane and are connected to each other by alternating hydrophilic regions. The amino- and carboxyl-terminal regions are hydrophilic domains that are believed to reside on the cytoplasmic side of the membrane (17). There are two other long hydrophilic stretches, one between hydrophobic helices 1 and 2, which is predicted to be extracellular, and the other between helices 6 and 7 on the cytoplasmic side of the membrane. The hydropathy profiles of the bacterial transport proteins (18), as well as those of other glucose-transport proteins whose genes have been cloned recently (22–26), all conform to this general model.

The hydropathy plot of the LTP sequence (Fig. 4), determined by the method of Kyte and Doolittle (21), is similar to those for the glucose transporter (17) and for the bacterial transport proteins (18); there are 12 distinct peaks of hydrophobicity (numbered 1–12 in Fig. 4) that alternate with hydrophilic regions. With the exception of peak 1, which is less hydrophobic than the other domains, these peaks correspond to regions of the sequence (heavy lines in Fig. 3) that are predicted to be membrane-spanning α -helices by the algorithm of Klein *et al.* (19). By analogy to the other known transporters, we propose that peak 1 is probably also a transmembrane domain, but this remains to be demonstrated. There are also extended hydrophilic regions between hydrophobic peaks 1 and 2 and between peaks 6 and 7, which correspond to the extracellular and intracellular loops, respectively, in the model for the glucose transporter (17). Like the mammalian and bacterial transport proteins, the LTP sequence lacks an apparent amino-terminal signal sequence. In contrast to the erythrocyte transporter, LTP contains more extended hydrophilic regions at the amino terminus and between several of the hydrophobic peaks. These larger hydrophilic domains account for the larger size of the LTP sequence (567 amino acids) compared to the 492-residue erythrocyte transporter.

In general the regions of the erythrocyte transporter that are predicted (19) to be transmembrane α -helices (narrow lines in Fig. 3) overlap the predicted transmembrane regions of LTP. The major exception is transmembrane segment 1 of the erythrocyte transporter, which is not predicted by this algorithm in LTP, as discussed above. In addition, the

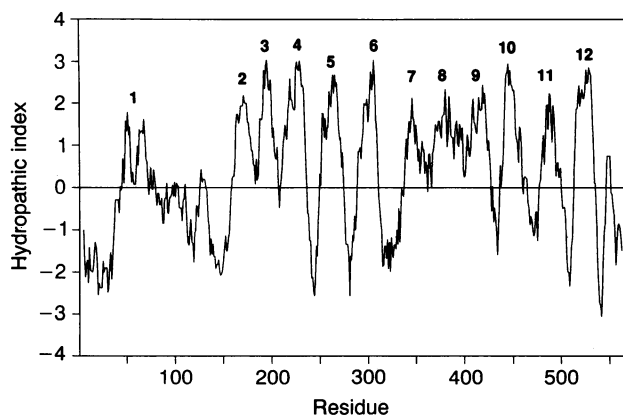


FIG. 4. Hydropathy plot for LTP was constructed using the algorithm of Kyte and Doolittle (21) and a window size of nine residues. Values above the zero baseline are hydrophobic, while those below are hydrophilic. The numbers above hydrophilic peaks indicate putative membrane-spanning domains (17).

precise location of the predicted helix termini differ between the two proteins.

DISCUSSION

Parasitic protozoa typically possess multiple morphologically distinct life-cycle forms that are adapted to a particular biological niche. There must exist developmentally regulated genes whose expression is modulated during the transitions between the life-cycle stages and whose gene products confer upon the parasite its unique phenotype in each stage. To understand these developmental transformations at the molecular level, it will be necessary to identify specific developmentally regulated genes and to study the mechanisms of their regulation. We have cloned a gene from *L. enriettii* whose mRNA accumulates much more abundantly in the promastigote (insect) stage of the parasite than in the amastigote stage that resides in the mammalian host.

The DNA sequence of this developmentally regulated gene has been conceptually translated and generates a single extended translational reading frame. Although we have no direct evidence that this predicted protein is synthesized by the parasite, a number of striking structural similarities between this polypeptide and the sequences of several known transport proteins strongly suggest that this is the correct gene product. These similarities also imply a functional role for the predicted protein as a membrane transporter.

The suggestion that this gene encodes a membrane transport protein is supported by several observations: (i) amino acid sequence similarity between the predicted protein and three known transport proteins (Fig. 3) that extends over the length of the maximally aligned sequences, (ii) the presence of 22 amino acid residues (those in Fig. 3 where the box covers all four rows) that are identical in all four of the maximally aligned sequences, (iii) a highly statistical significant relationship between the sequence of the predicted protein and each of the three transporters, as judged by the RELATE algorithm of Dayhoff *et al.* (12), and (iv) striking similarities among the hydropathy profiles of all four proteins, including the prediction of 12 alternating transmembrane segments interspersed with several extended hydrophilic loops.

We have identified several differences between LTP and the known transporters, including the absence in LTP of the conserved peptide sequences Pro-Glu-Ser-Pro-Arg and Pro-Glu-Thr-Lys-Gly, the absence of a predicted glycosylation site, and the absence of one of the otherwise duplicated (Arg/Lys)-Xaa-Gly-(Arg/Lys)₂ sequences. Since we do not know the function of these sequences in the known transporters, we cannot determine the significance of their absence in LTP. However, it is notable that the *E. coli* citrate transporter (18), which is structurally related to the erythrocyte transporter, AraE, and XylE, is also missing the two conserved peptides.

Ultimately, a functional test will be required to prove the hypothesis that this *Leishmania* gene encodes a membrane transport protein. The recent demonstration (27) that the human erythrocyte glucose transporter can be expressed functionally in *E. coli* and can complement a deficiency in glucose transport provides encouragement that a genetic test can be accomplished. A successful genetic assay would also provide proof for the identity of the ligand.

Several *Leishmania* transport systems have been characterized biochemically and genetically, including those for glucose (28), ribose (29), folate (30, 31), proline (32), and nucleosides (33), but none of the genes for these transporters has yet been cloned. The genes for two closely related cation-transporting ATPases have been cloned from *Leishmania donovani* (34); the RNA from one of these two genes accumulates to a higher level in amastigotes than in promastigotes. Hence the genes for at least two *Leishmania* trans-

port proteins are regulated differentially during the life cycle, one being expressed preferentially in amastigotes (a cation ATPase) and the other in promastigotes (LTP).

The biology of this parasite within the sandfly vector suggests a reason for the existence of developmentally regulated transport proteins. Although parasites are initially delivered to the female sandfly via a blood meal, the vertebrate blood is soon digested, and the newly transformed promastigotes are exposed to the environment of the insect alimentary tract. The sandflies subsequently ingest sugar meals from plants (35); several investigators have demonstrated that these sugar meals are required by the parasite for its further development to an infectious stage within the sandfly (36). Sugar meals enhance the ability of laboratory infected sandflies to successfully transmit the infection to a mammalian host upon taking a second blood meal (37), and the addition of sugars to cultures of promastigotes grown *in vitro* induces the formation of various morphologically distinguishable subforms that normally populate the insect (38). Thus the parasite is exposed to a different nutrient milieu in the sandfly compared to the macrophage lysosome and may require new transport proteins to utilize these different metabolites and to proceed along its developmental program.

Recently a promastigote-specific gene that encodes a reductase protein was cloned from *Leishmania major* (39, 40). The genes for this reductase and for the putative transporter described here may serve as valuable probes for investigating the molecular basis of developmentally regulated gene expression in *Leishmania*.

We thank Dr. Will Gilbert, Mr. Chen-Ming Fan, and Mr. Dan Chasman for generous help with the computer-assisted data analysis. This investigation received financial support from Grant AI25920 from the National Institutes of Health, from the United Nations Development Programme/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases, from the Oregon Affiliate of the American Heart Association, and from the Medical Research Foundation of Oregon.

- Zuckerman, A. & Lainson, R. (1979) in *Parasitic Protozoa*, ed. Krier, J. P. (Academic, New York), Vol. 1, pp. 57–133.
- Landfear, S. M., McMahon-Pratt, D. & Wirth, D. F. (1983) *Mol. Cell. Biol.* **3**, 1070–1076.
- Landfear, S. M. & Wirth, D. F. (1984) *Nature (London)* **309**, 716–717.
- Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning Techniques: A Practical Approach*, ed. Glover, D. (IRL, Oxford), Vol. 1, pp. 49–78.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- St. John, T. P. & Davis, R. W. (1979) *Cell* **16**, 443–452.
- Vogelstein, B. & Gillespie, D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 615–619.
- Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Zagursky, R., Baumeister, N., Lomax, N. & Bergman, M. (1985) *Gene Anal. Tech.* **2**, 89–94.
- Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
- Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
- Comeau, A. M., Miller, S. I. & Wirth, D. F. (1986) *Mol. Biochem. Parasitol.* **21**, 161–169.
- Muhich, M. L. & Boothroyd, J. C. (1988) *Mol. Cell. Biol.* **8**, 3837–3846.
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
- Mueckler, M., Caruso, C., Baldwin, S. A., Panico, M., Blench, I., Morris, H. R., Allard, W. J., Lienhard, G. E. & Lodish, H. F. (1985) *Science* **229**, 941–945.
- Maiden, M. C. J., Davis, E. O., Baldwin, S. A., Moore, D. C. M. & Henderson, P. J. F. (1987) *Nature (London)* **325**, 641–643.
- Klein, P., Kanehisa, M. & DeLisi, C. (1985) *Biochim. Biophys. Acta* **815**, 468–476.
- Berk, A. J. & Sharp, P. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1275–1278.
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
- Thorens, B., Sarkar, H. M., Kaback, H. R. & Lodish, H. F. (1988) *Cell* **55**, 281–290.
- Fukumoto, H., Seino, S., Imura, H., Seino, Y., Eddy, R. L., Fukushima, Y., Byers, M. G., Shows, T. B. & Bell, G. I. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5434–5438.
- Kayano, T., Fukumoto, H., Eddy, R. L., Fan, Y.-S., Byers, M. G., Shows, T. B. & Bell, G. I. (1988) *J. Biol. Chem.* **263**, 15245–15248.
- Celenza, J. L., Marshall-Carlson, L. & Carlson, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2130–2134.
- James, D. E., Strube, M. & Mueckler, M. (1989) *Nature (London)* **338**, 83–87.
- Sarkar, H. K., Thorens, B., Lodish, H. F. & Kaback, H. R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5463–5467.
- Zilberstein, D. & Dwyer, D. M. (1984) *Mol. Biochem. Parasitol.* **12**, 327–336.
- Pastakia, K. B. & Dwyer, D. M. (1987) *Mol. Biochem. Parasitol.* **26**, 175–182.
- Ellenberger, T. & Beverly, S. M. (1987) *J. Biol. Chem.* **262**, 10053–10058.
- Kaur, K., Coons, T., Emmett, K. & Ullman, B. (1988) *J. Biol. Chem.* **263**, 7020–7028.
- Bonay, P. & Cohen, B. E. (1983) *Biochim. Biophys. Acta* **731**, 222–228.
- Aronow, B., Kaur, K., McCartan, K. & Ullman, B. (1987) *Mol. Biochem. Parasitol.* **22**, 29–37.
- Meade, J. C., Hudson, K. M., Stringer, S. L. & Stringer, J. R. (1989) *Mol. Biochem. Parasitol.* **33**, 81–92.
- Young, C. J., Turner, D. P., Killick-Kendrick, R., Rioux, J. A. & Leaney, A. J. (1980) *Trans. R. Soc. Trop. Med. Hyg.* **74**, 363–366.
- Schlein, Y. (1986) *Parasitol. Today* **2**, 175–177.
- Warburg, A. & Schlein, Y. (1986) *Am. J. Trop. Med. Hyg.* **35**, 926–930.
- Schlein, Y., Borut, S. & Greenblatt, C. L. (1987) *J. Parasitol.* **73**, 797–805.
- Kidane, G. Z., Samaras, N. & Spithill, T. W. (1989) *J. Biol. Chem.* **264**, 4244–4250.
- Samaras, N. & Spithill, T. W. (1989) *J. Biol. Chem.* **264**, 4251–4254.