ELSEVIER

Biochimica et Biophysica Acta 128 (1995) 94–98

BBA Biochimica et Biophysica Acta

Short Sequence-Paper

The V-ATPase *A* subunit gene (*vma-1*) from *Giardia lamblia* ☆

Elena Hilario, Johann Peter Gogarten *

Department of Molecular and Cell Biology, University of Connecticut, 75 North Eagleville Rd., Storrs, CT 06269-3044, USA

Received 27 March 1995; accepted 12 May 1995

Abstract

The sequence of the gene encoding the *A* subunit of the vacuolar type ATPase from *Giardia lamblia* is reported. Comparison of the encoded protein with the homologous subunits of eukaryotic and archaebacterial ATPases reveals high levels of similarity throughout the sequence (e.g., overall 49.1 and 44.6% identity to the homologous subunit from carrot and *Halobacterium*, respectively). An exception are three regions which are unique to the *Giardia* subunit. The largest of these regions contains motifs characteristic for eukaryotic spliceosomal introns; however, comparison to the cDNA shows that this region is also present in the mRNA.

Keywords: V-ATPase; Intron; Promoter; (*G. lamblia*)

The diplomonad *Giardia lamblia* is a strict anaerobic parasite with two equivalent nuclei, four pairs of flagella (9 + 2 type), numerous acidic vacuoles, no mitochondria, a high G + C content in its ribosomal DNA (72–75%) but lower in genomic DNA (42–48%) [1]. *Giardia* has no structurally evident Golgi apparatus, however, a regulated secretory pathway has been characterized during encystation [2]. Phylogenetic analysis using 16S rRNAs as molecular markers suggests that *Giardia* constitutes an early branch within the eukaryotic lineage and that in some respects this organism might be regarded as a missing link between eukaryotes and prokaryotes [3].

Proton pumping ATPases have been divided into three groups: P-, V-, and F-type ATPases [4]. P-type ATPases are characterized by a phosphorylated intermediate present in the catalytic cycle. Many cation pumping ATPases located in the plasma membranes of eubacteria and eukaryotes belong to this ATPase type. The Ca^{2+} pumping ATPase located in the eukaryotic endoplasmic reticulum also is a P-type ATPase. A Ca^{2+} translocating P-ATPase has been described in a microsomal fraction from *Giardia* [5].

F-ATPases or coupling factor ATPases are present in the plasma membrane of Eubacteria, the inner mitochondrial membranes and in the thylakoid membranes of chloroplasts. These enzymes couple the electrochemical proton gradient produced by the electron transport chain to the synthesizes of ATP. In eubacteria, and under experimental conditions, F-ATPases can also function as ATP driven proton pumps. So far, no F-type ATPases have been described in *Giardia*.

Vacuolar type ATPases (V-ATPases) are present in the endomembrane systems of eukaryotes: vacuoles, lysosomes, Golgi apparatus, coated vesicles, etc. In some instances V-ATPases were also found to energize the plasma membranes of specialized cells in transport epithelia in animals. F- and V-ATPases consist out of two major portions: a hydrophilic and a membrane-bound sector, named F_1 or V_1 and F_0 or V_0, respectively. Sequences of the major subunits revealed that these two ATPase types have evolved from the same ancestral ATPase. Vacuolar type ATPases generate proton gradient across membranes. These gradients are used to energize the transport of solutes, to maintain a constant intracellular pH, and to provide a signal for intravesicular processes (for reviews see Ref. [6]). The archaebacterial counterpart of the V-ATPase has been called A-ATPase [7,8]. In contrast to the eukaryotic V-ATPases, A-ATPase can synthesize ATP consuming the H^+ gradient [8]. The primary structure of the A-ATPase A and B subunits is very similar to the V-ATPase, but functionally this enzyme appears more similar to the F-ATPase [7–9].

☆ The nucleotide sequence data reported in this paper have been submitted to the GenBank/EMBL Data Libraries under the accession number U18938.

* Corresponding author. E-mail: gogarten@uconnvm.uconn.edu. Fax: +1 (203) 4861784.

SSDI 0005-2736(95)00130-1

In this paper we report the nucleotide and deduced amino acid sequence of the gene encoding the catalytic subunit A of the vacuolar H^+-ATPase of the diplomonad *Giardia lamblia*.

Genomic DNA from *Giardia lamblia* grown in axenic cultures (ATCC 30888, strain Portland 1) was extracted according to Ref. [10]. A short segment of the coding region of the V-ATPase A subunit from *Giardia* was amplified by PCR using genomic DNA and a pair of primers based on previous work done by Starke and Gogarten [11]. The nucleotide sequences for the left and right primers are TCTACGTTGGTTGTGGGGARAGNG-GNAAYGARATG and GCCTCGCGAGCAGC-CACAGGCATRTTNGANGTRTT, respectively. A PCR product of 191 bases (corresponding to positions 831 to 1023) was obtained. This PCR product was subcloned in pBluescript, sequenced, and a non radioactive probe was prepared using digoxigenin dUTP (Boehringer) as a label (probe 1, see Fig. 1). Two samples of genomic DNA were digested with *Kpn*I and *Eco*RI, respectively. Only one band hybridized with the homologous probe 1 in a Southern blot (Fig. 2).

A genomic library *Giardia lamblia* strain WB constructed in Lambda Zap II was kindly provided by Dr. Frances Gillin (University of California, San Diego, CA) and screened with the homologous probe. 21 positive clones were picked after the third round of screening. Seven of them were excised in vivo and the pBluescript clones obtained were used to transform *Escherichia coli* XL1 Blue. One clone which contains the whole gene was selected and sequenced.

The open reading frame that encodes the V-ATPase *A* subunit of *Giardia lamblia* is 1962 bp long. The predicted protein sequence contains 654 amino acids and a predicted molecular mass of 71 880 Da. The G + C content of the coding region is lower (49%) than in other protein encoding *Giardia* genes (an average of 57% was calculated by [1]); however, the *A* subunit gene shows the same preference for G/C in the third codon position (50–60%) that has been reported for other *Giardia* genes [12].

```
                                                GATGTATCCACAGGCCTGGACACACGGTCTCCCGTATACAATGGGCCTCCTCTAGGCTGTTAAGCTCATCATCTAGGTTAGTCTCTGCACAGTCTC
CGAAACTCCGTGAGTTCATAGAATGCCGTACCTACTCCTCTGTCGTACGCCCAACGCCGTGGATGTTATGTCTCTAAATAAATCTGTTTCAAAAAACGCCGTTTTCTTTTAAAATTTGAA

ATGGTTACAAGCGAGACGACATCTAATGGACTGATCCAGTCCATTGCAGGTCCGGTCGTCACAGCAACTAATATGACCGGGTGCTTCATGTTCGAGGTTTGCTATGTGGGAAAAGCCCGT   120
 M  V  T  S  E  T  T  S  N  G  L  I  Q  S  I  A  G  P  V  V  T  A  T  N  M  T  G  C  F  M  F  E  V  C  Y  V  G  K  A  R

TTGATTGGGGAAATCATCCAACTCAAGGGGGATTCGGCTGTTATCCAGGTCTATGAAGATACCTCTGGGCTAGAAGTTAACGACTGCGTCTACAAATCCGGACGCCTACTTTCTGTCCAT   240
 L  I  G  E  I  I  Q  L  K  G  D  S  A  V  I  Q  V  Y  E  D  T  S  G  L  E  V  N  D  C  V  Y  K  S  G  R  L  L  S  V  H

CTTGGTCCCGGTTTACTTTCGTCTATTTACGATGGAATCCAGCGCCCCTTAGAGAAGATCGCACAGATTACTAACTCCCACTTCATCCCGCGAGGAATTTCTGCGCCAGCCTTGGATCTA   360
 L  G  P  G  L  L  S  S  I  Y  D  G  I  Q  R  P  L  E  K  I  A  Q  I  T  N  S  H  F  I  P  R  G  I  S  A  P  A  L  D  L

GAGAGGAGATGGACGTTCAGACCGCTTGTCAAGTTGGGGGATCTCCTATCTGTGGGTGACATCTTTGGCATTGTTCCAGAGAACGATCTTCTGGAGTGCAAGATAATGCTACAGAGTATG   480
 E  R  R  W  T  F  R  P  L  V  K  L  G  D  L  L  S  V  G  D  I  F  G  I  V  P  E  N  D  L  L  E  C  K  I  M  L  Q  S  M

                Fragment A
GATTCAACGAGGCTACAAATGAGAAAGCGAACATTCCAGGCGGAAAGGTTGTTAAAATACGTGCGGCCTACCGACGCATACACAGTCAAAGACGAAGAAAAGATCTTGGTCCTGGAATAC   600
 D  S  T  R  L  Q  M  R  K  R  T  F  Q  A  E  R  L  L  K  Y  V  R  P  T  D  A  Y  T  V  K  D  E  E  K  I  L  V  L  E  Y

GCCGGACAAAGGTATTACTACGGCCTGGCCCACTACCATCCTGTTCGTTCGCCTAAGAAGGTCTTTAAGAAGCTGCCGTGCATAAATCCCTTGATAACAGGCCAGAGAACGTTGGATGGA   720
 A  G  Q  R  Y  Y  Y  G  L  A  H  Y  H  P  V  R  S  P  K  K  V  F  K  K  L  P  C  I  N  P  L  I  T  G  Q  R  T  L  D  G

                                                                                                        Probe 1
CTCTTTCCCCTTGCAATAGGAGCCACCGCAGCGATCCCAGGAGGTTTTGGGTGTGGGAAGACTGTCGTGAGTCAATCCATCAGCAAGTATGGTAACACAGATGTTATTGTCTACATTGGC   840
 L  F  P  L  A  I  G  A  T  A  A  I  P  G  G  F  G  C  G  K  T  V  V  S  Q  S  I  S  K  Y  G  N  T  D  V  I  V  Y  I  G

                                                Fragment B
TGTGGAGAGCGCGGTAATGAAATGGCTGAGATCCTCACAGATTTCCCCGAAATGACATTCGAGGCTAAGAAGCGCGTTATCGGGCCATCTGGACAGCAAGAGATTAAAACGGTTGTATCC   960
 C  G  E  R  G  N  E  M  A  E  I  L  T  D  F  P  E  M  T  F  E  A  K  K  R  V  I  G  P  S  G  Q  Q  E  I  K  T  V  V  S

                                                                                                Probe 2
GACATTTTCAGCCGCACTGTGCTTGTCGCGAATACGTCTAATATGCCCGTTGCAGCCCGTGAGGCTAGCATCTACACAGGAATCACGATTTCTGAGTTCTTCCGTGACCAAGGGTACAAC  1080
 D  I  F  S  R  T  V  L  V  A  N  T  S  N  M  P  V  A  A  R  E  A  S  I  Y  T  G  I  T  I  S  E  F  F  R  D  Q  G  Y  N

GTCACTCTTCTGGCCGACTCTACTTCCCGCTGGGCTGAGGCACTGCGAGAGATCAGCGGGCGCTTGGGAGGTATTCCGGGAGAAGGCGGCTATCCTGCTGATCTCACCTCTAAGCTTAGC  1200
 V  T  L  L  A  D  S  T  S  R  W  A  E  A  L  R  E  I  S  G  R  L  G  G  I  P  G  E  G  G  Y  P  A  D  L  T  S  K  L  S

                                Fragment C
CACTTCTATGAGCGTGCTGGACGAGTCATTTGTCTGGGGAATTCCAATGGTGCGTGGAGCCCAGAGAATATCCTTGATCCTACGTCTGGAGCTATCTCGTCAACCGAACAAGGTCTCCTG  1320
 H  F  Y  E  R  A  G  R  V  I  C  L  G  N  S  N  G  A  W  S  P  E  N  I  L  D  P  T  S  G  A  I  S  S  T  E  Q  G  L  L

CAACCCGGCCAGCGTATCGGCTCTGTGTCGGTGATAGGTGCTGTTTCCCCGGCTGCAGGAGATCTTTCTGATGCAGTCGCCGTCAGTACACTGGCTATAGTCCAAGTTTTCTGGGGACTC  1440
 Q  P  G  Q  R  I  G  S  V  S  V  I  G  A  V  S  P  A  A  G  D  L  S  D  A  V  A  V  S  T  L  A  I  V  Q  V  F  W  G  L

AGCAAGAACCTGGCGAACGGCAAGCACTTTCCATCCGTCGATTGGCTCATTAGCTATTCCAGATGTACTGAGACTCTTGCGCCGTGGTACAACGCTAGAGACATCAGCTTCTTGACCAAT  1560
 S  K  N  L  A  N  G  K  H  F  P  S  V  D  W  L  I  S  Y  S  R  C  T  E  T  L  A  P  W  Y  N  A  R  D  I  S  F  L  T  N

AAGGAAAAGGCGAAGACTCTTCTGCAAAACGAGGTGGCCATCCAGGAAACCGCGCAGCTCGTTGGCTATGATAGCTTGGATGATAGCGAAAAGCTTATTCTCGACGTCTGTAACCTAATT  1680
 K  E  K  A  K  T  L  L  Q  N  E  V  A  I  Q  E  T  A  Q  L  V  G  Y  D  S  L  D  D  S  E  K  L  I  L  D  V  C  N  L  I

CAAGAGGGCTTTTTGCAGCAGAACTCCTACACGGTGTACGATAAGTTCTGCCCGTTTATAAAGACTGATCTGATGCTTCGTAACATTGTTCATTACTACGAGTCCTGTAAGACGGCTCTG  1800
 Q  E  G  F  L  Q  Q  N  S  Y  T  V  Y  D  K  F  C  P  F  I  K  T  D  L  M  L  R  N  I  V  H  Y  Y  E  S  C  K  T  A  L

GGGCAGGGACACAGCTATCAAGAGCTTAAGAACAGCCTGGAAAAGCTCATTACAGCTCTTTATAGGATGAAGTTTATAGACACTAATACTAACGGTATCGAGTACGGCGTCACCGAGTTA  1920
 G  Q  G  H  S  Y  Q  E  L  K  N  S  L  E  K  L  I  T  A  L  Y  R  M  K  F  I  D  T  N  T  N  G  I  E  Y  G  V  T  E  L

                                                Putative Polyadenylation Signal
GACAAACTCCACGCCCAGATCACTGCAGCATTTGACGCGCTGTAAGCTCTTTCCAGTGAATTTACTAAGAGTTCGGGTATCGGTAAATCCATATGGTCACTTCGGCCCTCCATACTTACT  2040
 D  K  L  H  A  Q  I  T  A  A  F  D  A  L  -

ATTGTTGTTACTATTGTTATTGTTGTTGTTATGATGGCGTCCGTTGCCGTGACCGCTGCCACGGTCAAAGCTCGATCTATTGAATCGGTCGCCGGTAGCCTGTCTAT
```

Fig. 1. Nucleotide sequence and deduced protein sequence of the V-ATPase A subunit gene (*vma-1*) from *Giardia lamblia*. The translated product was obtained using the universal genetic code. The open reading frame was identified by sequence alignment of several V-ATPase genes (see Fig. 4) and by comparing the AT region upstream from the putative start codon with several *Giardia* genes (see Fig. 3). Lines above the nucleotide sequence denote fragments A, B and C. Lines below the nucleotide sequence denote probes 1 and 2. Double underlined sequence mark the palindrome.

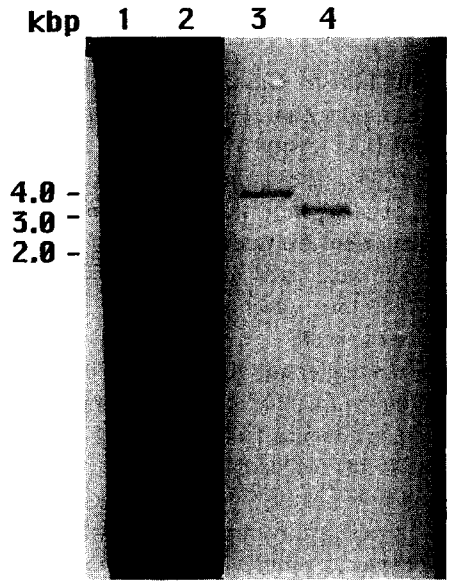

Fig. 2. Southern blot analysis of genomic DNA from *Giardia lamblia* digested with *Eco*RI (lanes 1 and 3) and *Kpn*I (lanes 2 and 4). The digested DNA was transferred to a nylon membrane and hybridized at 68°C for 12 h, using a digoxigenin dUTP labeled probe. Lane 1 and 2 are after staining with ethidium bromide, lane 3 and 4 show the crossreactivity to probe 1 (compare Fig. 3). With either endonuclease only a single band of 3 to 4 kbp hybridized to this probe.

An AT rich region of about 80 bases precedes the initiation codon (A + T 68%). A comparison of this region with the 5′ upstream regions of other *Giardia* genes is shown in Fig. 3. Some common motifs were contained within these regions; however, their physiological role has not been determined. A TA octamer is located between −13 and +1. From primer extension and S1 analyses of several genes, it has been determined that the 5′ untranslated region of mRNAs consists of 3–6 nucleotides only, starting within the TA octamer (see references in Fig. 3, except HSP 70). The TA octamer is preceded by a CG rich region located at position −40 to −10 (an additional CG region at −60 in V-ATPase A). A motif found within most of the CG region is CGCC. Another common motif is a CAAAAAA box at position −60 to −20. Some variations contain either T or G, but the A content is predominant in most cases. A CAAAT motif is found in many *Giardia* genes within the first 70 bases upstream from the initiation codon; however, this motif was not found within the 217 nucleotides preceding the putative initiation codon of the V-ATPase A-subunit gene. From these observations a general picture of the promoter features can be drawn:

CAAAAAA----------CGCC---------TA octamer **ATG**

-60 to -20 -40 to -10 -13 to +1 +1

Another potential TA octamer can be found at position −39, preceded by a CGCC motif; however, no start codon was found adjacent to this TA region. Potential TATA boxes have been identified in some *Giardia* genes (see Ref. [1] for review). Comparison with upstream regions of other *Giardia* genes suggests the putative initiation codon of the V-ATPase A gene to be localized 7 nucleotides downstream from the TA octamer. Another ATG is located 72 nucleotides downstream from the putative start codon (see Fig. 1); however, this second ATG is not preceded by the features typically associated with start codons in *Giardia* (compare Fig. 3). Therefore, methionine at position 72 is an improbable start codon.

The 3′ end of the gene contains the polyadenylation signal AGTRAA that was proposed for *Giardia lamblia* [13]. In the V-ATPase *A* subunit gene this putative polyadenylation signal is AGTGAA and is positioned ten nucleotides downstream from the stop codon. Interestingly, a palindrome flanks the polyadenylation signal (see Fig. 1). A stem loop structure is often part of the polyadenyla-

5′ region in:	-70 *	-60 *	-50 *	-40 *	-30 *	-20 *	-10 *	+1 *
α-2 giardin[12]	GGATAAGCGAACTCATGATGGAAATT*caaat*TAC*CTTAAAA*TATATTTTCTGAgcgcTTACATTTAGAAA**ATG**							
α-1 giardin[13]	ATTA*caat*TCAAAGTTGGT*caatt*TGAGCTT*CAAAAAA*TAc**cgcc***AAAAAAT*gg*caattc*ggTTTTAAAA**ATG**							
NADP-GDH[14]	GGGGAGATGGCTGTTCTTGAAAAGCTGACCA*CAAATAA**cgcc***TTTAATTACAgg**cgcc**ccAGATTTTAAA**ATG**							
α-tubulin[15]	TCTTCTGGAGCAGAAAA*CAATTTA*GAATT*caaat*CAG*caaat*TCCAGAGTCTGGAcgggcgg*AAATAAAA***ATG**							
β-tubulin II[16]	GATTGC*CATTTAA*ATTG*caat*CTGAACTTggc**cgccgcc***aaat*CCAGATTC**cgcc**gg**cgcc**g*ATTTAAAA***ATG**							
surface protein 11[17]	ATCTTAGGCTTGGGTGCTTGAGATTATTAAGAAGATTCTTAcccgTATAAGTGAGCTGAGAAGGATTTTA**ATG**							
ADP ribosyl. factor[18]	CTGCAGATCCTTTGGCATCCTGTACAAACTATTATGATTT*CAAAAAA*TAGAAggcTggccATTAAAAAGA**ATG**							
HSP 70 kD[19]	TGTTGCCTTCC*gaaat**CAAAAAA***ATCA*caat**CAAAAGA*TCTGACGATCAGGA**cgcc**GGGAAAATAAAATC**ATG**							
V-ATPase A	CCAA**cgcc**GTGGATGTTATGTCTCTAAATAAATCTGTTT*CAAAAA**cgcc***GTTTTCTTTTAAAATTTGAA**ATG**							

Fig. 3. Comparison of the first 70 nucleotides upstream the initiation codon of several genes from *Giardia lamblia*. Common motifs are underlined and highlighted. See text for details. See Refs. [12–19].

tion signal in higher eukaryotes [19]; however, none of the genes included in Fig. 3 contain a similar palindrome in this position (data not shown). Its function during transcription (if any) is unknown.

A multiple amino acid sequence alignment of various V-ATPase A-subunit (Fig. 4) reveals high levels of similarity throughout the sequence. The aligned *Giardia* sequence shows 52.1, 49.1, 50.2, 49.9, 44.4, and 44.6% identity to the sequences from *Entamoeba, Daucus, Manduca, Neurospora, Sulfolobus* and *Halobacterium*, respectively. However, three regions in the *Giardia* A-subunit cannot be aligned to the homologous sequences. These fragments are located at positions: 496 to 534 (fragment A); 898 to 960 (fragment B); and 1246 to 1332 (fragment C). In the homologous β-subunit of the bovine mitochondrial F-ATPase these regions correspond to position 112, 200–213, and 299. All of these positions are located on the surface of the F-ATPase [20]. At the nucleotide level, fragment C contains some features corresponding to a typical eukaryotic intron: it is flanked by GGT at its 5′ end (position 1249) and AG at the 3′ end (position 1332). No spliceosomal introns have been reported in *Giardia* genes. To establish the nature of fragment C, we used PCR amplification with a set of primers flanking this region using a cDNA library in λgt11 from *Giardia lamblia* (strain WB; this library was kindly provided by Dr. Michael Mowatt, Laboratory of Parasitic Diseases, NIAID, NIH, Bethesda, MD). The primers are located at positions: 1065 to 1984, and 1340 to 1359, corresponding to a fragment of 294 bp from the genomic clone. As shown in Fig. 5, no

```
Giardia              MVTSETTSNGLIQSIAGPVVTATNMTGCFMFEVCYVGKARLIGEIIQLKGDSAVIQVYEDTSGLEVNDCVYKSGRLLSVHLGPGLLS
Entamoeba          MNFDTDKKEKEF.KVY.VS....I.E..L.AA.N.LVR..SRG.M....R.E.TT.T.....E.A..QLG.M.ERTMKP...E....IMT
Daucus     MPSVYGDRLTTFEDSEKESEY.YVRKVS....V.DG.G.AA.Y.LVR..HDN......R.E....T.....E.A..M...P.LRTHKP...E....I.G
Manduca      MASKGGLKTIANEENEERF.YVFAVS.....EK.S.SA.Y.LVR..YNE.V....R.E..M.T.....E...VT.G.P.LRT.KP...E....I.G
Neurosp.         MAPQQNGAEVDGIHT.K.Y.VS....V.ED.I.VA.Y.LVK..HDQ.V..V.RIN..Q.T.....E.A.VM.G.P.LRT.KP...E.....N
Sulfol.                MV.E.RVVRVN..L.I.DG.REAQ....V..SDLK.V...TRIE..R.F.....S.D.VKPG.K..R..AP...E.....IG
Halobact.            MSQAEAI.DT.E.E.VS.......GL DAQ.ND.V...DEG.M..V.EIE..VTT.....E...IGPGQP.DNT.EP.T.D....M.D

Giardia    SIYDGIQRPLEKIAQITNSHFIPRGISAPALDLERRWTFRPL VKLGDLLSVGDIFGIVPENDLLEC KIMLQSMDSTRLQMRKRTFQAERLLKYVRPT
Entamoeba ..F.......VS..EKSG.I.......VAS..HQ.E.E.T.. ..K..HV.G...I.T...SA.VV H..LVPPTVMGTV              TW.AEA
Daucus     N.F.......KT..KRSGDVY....V.V....KDTL.E.Q.KKIGE....TG..LYAT.F..S.MQ HHVA.PPDAMGKI              T..A.A
Manduca    ..F.......KD.NEL.Q.IY..K.VNV.S.AR.VD.E.N..N..V.SHITG..LY...H..T.VK H.MLMPPRAKGTV              T.IA.A
Neurosp.   N.............EAS..IY.....AT....RKKK.E.T. TM.V..HIAG..VW.T.Y..SFISVH..L.PPRARGTI              TRIAEK
Sulfol.    K....L....DS..KVS..P.VA..V.I....RQTK.H.V. K..S..KVGP...I.V.Q.T..I. HR.LIPPNVHGT.              .ELARE
Halobact.  .....V....DVLEDEMGA .LD..VD..GI..DTD.E.E. T.EA..EVAA..VV.T.D.TVSI. H.VLVPPRSDGGE              VVAVES

Giardia    DAYTVKDEEKILVLEYAGQRYYYGLAHYHPVRSPKKVFKKLPCINPLITGQRTLDGLFPLAIGATAAIPGGFGCGKTVVSQSISKYGNTDVIVYIGCGE
Entamoeba GN..L  DD.VIGI.FN.KTEELSM..HW...K.RPTAE.ITSTT..V....I..S...CIQ.G.C....A.......I..AL...S.S...I.V....
Daucus     GQ.SL  KDTV.E..FQ.VKKQFTMLQTW...T.RP.AS:.AADT..L....V..A...SVL.G.C....A.......I..AL...S.S.TV..V....
Manduca    GN.K.  TDVV.ET.FD.EKAQ.TMLQVW...Q.RP.TE...ANH..L....V..S...CVQ.G.T....A.......I..AL...S.S...I.V....
Neurosp.   GE...  .....EV.FD.KKTE.PMMQTW...V.RPAAE.HSANQ.FLV...V..A...SVQ.G.V....A.......I...V..FS.S.....V....
Sulfol.    GD...  .DVVA.VDMN.DEIPVKMYQKW...I.RPYKE..EPVE..L..I.V..TV..I.K.G......P..S....TL..LA.WSAAK.VI.V....
Halobact.  GTF..  DDTVVE.DTG  EEIQMHQEW...RQRPTVD.QTPTE..VS...I......I.K.G......P..S....TQ..LA.FADA.IV.......

Giardia    RGNEMAEILTDFPEMTFEAKKRVIGPSGQQEIKTVVSDIFSRTVLVANTSNMPVAAREASIYTGITISEFFRDQGYNVTLLADSTSRWAEALREISGRL
Entamoeba .......V.R...ALS       IKVGD KE.       S.MT..A.....................L..YY..M....AMM.................
Daucus     .......V.M...QL.       MTL.D.RE.       SVMK..T.....................A.Y...M....SMM.................
Manduca    .....S.V.R....L.       .EI E.VT.       S.MK..A.....................L..Y...M....SMM.................
Neurosp.   .......V.K....LS       IEV D.RK.       P.MK..T.I...................VA.Y.....M..AMM...S.............
Sulfol.    .....TDE.RS..KLK       DPWTG   K       PLLL..I.............PS...V.V.MA.Y......D.L.V............DLG..M
Halobact.  .....T.VIE....LP       DPQTG   N       PLMA..T.I.........G..SC.......A.YY..M..D.A.M..........M....S..

Giardia    GGIPGEGGYPADLTSKLSHFYERAGRVICLGNSNGAWSPENILDPTSGAISSTEQGLLQPGQRIGSVSVIGAVSPAAGDLSDAVAVSTLAIVQVFWGLS
Entamoeba AEM.ADS....Y.AAR.AS......M.E...SP                                  K......IV.....PG..F..P.TT...N........D
Daucus     AEM.ADS....Y.AAR.AS......K.K...GP                                  E.N...TIV.....PG..F..P.TSA..S........D
Manduca    AEM.ADS....Y.GAR.AS........K....P                                  D.E....IV.....PG..F..P.TAA..G........D
Neurosp.   .EM.ADQ.F..Y.GA..AS......K.QA..SP                                  P.E....IV.....PG..F..P.TSA..G........D
Sulfol.    EEM.A.E.F.SY.P.R.AEY........A...P                                  E.Y...TIAS....PG..FTEP.TSN..RF.R...P.D
Halobact.  EEM...E....Y.AAR..E......YFENFNGT                                  E  ..I........PG..F.EP.TQN..R..KT..A.D

Giardia    KNLANGKHFPSVDWLISYSRCTETLAPWYNAR DISFLTNKEKAKTLLQNEVAIQETAQLVGYDSLDDSEKLILDVCNLIQEGFLQQNSYTVYDKFCPF
Entamoeba .K..QR....A.N.N..F.KYIKS.DSY..SK .EE.VPLRD.I.EI..M.EGLLQIV....Q...AETD..T.EIARV.KDD........P..FS...
Daucus     .K..QR......N......KYSTA.ESF.EKF .SD.IDIRT..REV..R.DDLN.IV....K.A.AETD.IT.ETAK.LR.DY.A..AF.P.......
Manduca    .K..QR.....IN......KYMRA.DDF.EKN YPE.VPLRT.V.EI..E.EDLS.IV....KA..AETD.IT.E.AK.LKDD.......SS..R....
Neurosp.   .K..QR.....INTSV...KYLTI.DK..ERE YPD.PRLRDRIRQ..SDSEELDQVV....KSA.S.PD.IT..MAT..K.D.....G.SD..Q...I
Sulfol.    VS..QAR.Y.AIN.IQGF.AYVDLV.Q.WHKNV.PNWKEMRDTMMKV.IR.DELRQIVR...PEE.AEKD..V.EAAK..KDA..KE.A.DDI.A.SSP
Halobact.  SD..ERR...AIN.DE...LYKDQ.D..FTDNVVDDWAEQRQS.VDI.DE.SELE.IV....K.A.PEDQQ.T.E.ARY.R.AW....ALHDV.RY..P

Giardia    IKTDLMLRNIVHYYESCKTALGQGHSYQE        LKNSLEKLITALYRMKFIDTNTNGIEYGVTELDKLHAQITAAFDAL
Entamoeba Y..CGII...I.F.NEAFQ..  SVD.EDHKITWATI.SAMSD.LVRIS...YEEPS Q.EQVINEKYGE.YRD..TR.AT.LE
Daucus     Y.SVW.M...I.F.NLANQ.VER.AGMDGQKISYTLI.HR.GD.FYR.VSQ..E.PA E.EDVL.GKFK...DDL.SG.RN.EDETR
Manduca    Y..VG..K..ISF.DMSRH.VE STAQSDNKVTWNVIRDAMGNVLYQ.SS...K.PVKD.EAKIKADF.Q.LEDMS...RN.ED
Neurosp.   W..EW.MKLMMGFHDEAQK.IA..QN        WNKVREATQD.QAQ.KSL..EVPS E.Q.KICKKYEAIQQ.MLDK.ASVIDE
Sulfol.    Q.QVRIM.L.YIF.NQSQDLISK.        VPLKKILDKVGPIEPEII.I.YT  IK.DEL   NKI.EIENKLK.T..S.LKEVS
Halobact.  E..YAI.SG.KTLH.ESFE..DA.VPVE.  ITSIDAAPR.NR.V         TTPDDEHEAEVAEIKQQITE.LRELY
```

Fig. 4. Sequence alignment of the *Giardia lamblia* V-ATPase A subunit with others V-ATPase A subunits. The following sequences were used (accession numbers for the GenPept databank are given in parenthesis): *Entamoeba hystolytica* (U04849), *Daucus carota* (J03769), *Manduca sexta* (X64233), *Neurospora crassa* (J03955), *Sulfolobus acidocaldarius* (J0321), *Halobacterium salinarium* (X70294). The multiple alignment was calculated using the Clustal V Multiple Sequence Alignment program [21] with the following parameter settings: fixed gap penalty = 10; floating gap penalty = 10; and Dayhoff PAM 250 matrix. Residues identical to the *Giardia* sequence are denoted by (.).

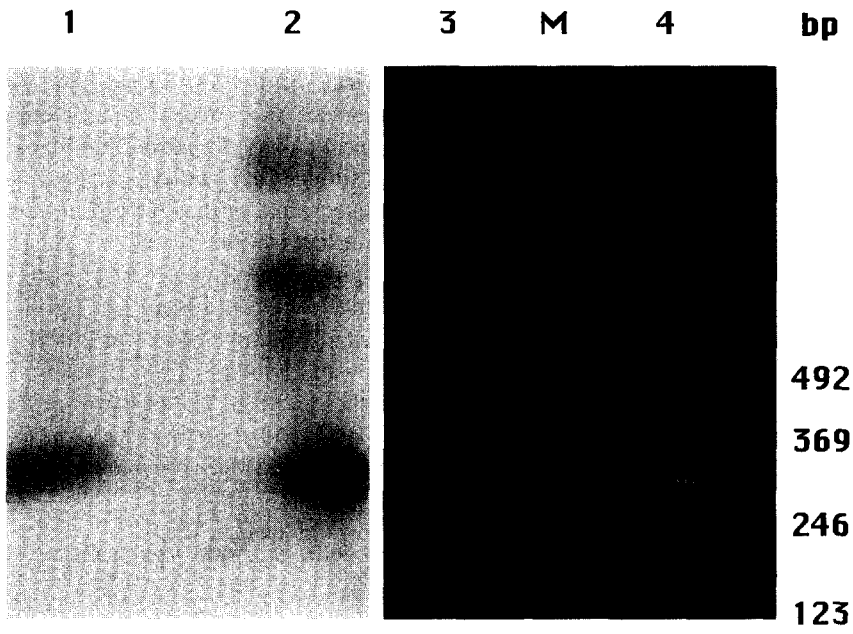

Fig. 5. PCR amplification and Southern blot analysis of fragment C of the V-ATPase A subunit encoding gene from *Giardia lamblia* (compare Fig. 2). Lanes 1 and 3 show the amplification products from a cDNA library ($1 \cdot 10^5$ phages); lanes 2 and 4 (control) show the products obtained from clone 5c α which contains a genomic sequence (10^5 copies). A set of primers flanking fragment C were used for the amplification. The PCR products were transferred to a nylon membrane and hybridized at 68°C for 2.5 h, with non-radioactive probe 2 (see Fig. 2). A PCR product of about 300 bases was amplified using either the genomic clone (lane 4) or the cDNA library (lane 3) as template (expected product: 294 bp). Both PCR products hybridized with the homologous probe 2 (lanes 1 and 2).

difference in the size of the PCR products was observed when either the genomic DNA clone or the cDNA library were used as substrate for the reaction. This result suggests that fragment C is part of the mRNA and not an intron. A Southern blot using the non-radioactive probe 2 (compare Fig. 1), which binds to 186 bases of fragment C, confirms that both PCR products encode the V-ATPase *A* subunit (Fig. 5).

The research was funded by NSF grant BSR-9020868, E.H. was supported in part by a fellowship from the DGAPA-Universidad Nacional Autónoma de México. We thank Frances Gillin and Michael Mowatt for providing us with DNA libraries, and Mark Brancieri for critically reading the manuscript.

References

[1] Adam, R.D. (1991) Microbiol. Rev. 55, 706–732.
[2] Reiner, D.S., McCaffery, M. and Gillin, F.D. (1990) Eur. J. Cell Biol. 53, 142–153.
[3] Sogin, M.L., Gunderson, J.H., Elwood, H.J., Alonso, R.A. and Peattie, D.A. (1989) Science 243, 75–77.
[4] Pedersen, P.L. and Carafoli, E. (1987) Trends Biochem. Sci. 12, 146–150.
[5] Muñoz, M.L., Claggett, C.E. and Weinbach, E.C. (1988) Comp. Biochem. Physiol. 91B, 137–142.
[6] Harvey, W.R. and Nelson, N. (1992) J. Exp. Biol. 172.
[7] Schäfer, G. and Meyering-Vos, M. (1992) Biochim. Biophys. Acta 1101, 232–235
[8] Mukohata, Y., Sugiyama, Y. and Ihara, K. (1992) J. Bioenerg. Biomembr. 24, 547–553.
[9] Kibak, H., Taiz, L., Starke, T., Bernasconi, P. and Gogarten, J.P. (1992) J. Bioenerg. Biomembr. 24, 415–424.
[10] Richards, E., Reichardt, M. and Rogers, S. (1994) in Current Protocols in Molecular Biology (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., eds.), John Wiley and Sons, USA.
[11] Starke, T. and Gogarten, J.P. (1993) Fed. Eur. Biochem. Soc. Lett. 315, 252–258.
[12] Alonso, R.A. and Peattie, D.A. (1992) Mol. Biochem. Parasitol. 50, 95–104.
[13] Peattie, D.A., Alonso, R.A., Hein, A. and Caulfield, J.P. (1989) J. Cell Biol. 109, 2323–2335.
[14] Yee, J. and Dennis, P.P. (1992) J. Biol. Chem. 267, 7539–7544.
[15] Kirk-Mason, K.E., Turner, M.J. and Chakraborty, P.R. (1989) Mol. Biochem. Parasitol. 36, 87–100.
[16] Ey, P.L., Khanna, K., Manning, P.A. and Mayrhofer, G. (1993) Parasitol. 58, 247–258.
[17] Murtagh, J.J., Mowatt, M.R., Lee, C.-M., Scott Lee, F.-J., Mishima, K., Nash, T.E., Moss, J. and Vaughan, M. (1992) J. Biol. Chem. 267, 9654–9662.
[18] Gupta, R.S., Aitken, K., Falah, M. and Singh, B. (1994) Proc. Natl. Acad. Sci. USA 91, 2895–2899.
[19] Heidecker, G. and Messing, J. (1986) Annu. Rev. Plant Physiol. 37, 439–466.
[20] Abrahams, J.P., Leslie, A.G.W. Lutter, R. and Walker, J.E. (1994) Nature 370, 621–628.
[21] Higgins, D.G. and Sharp, P.M. (1988) Gene 73, 237–244.