INFECTION AND IMMUNITY, Aug. 1994, p. 3572–3575
0019-9567/94/$04.00+0

Vol. 62, No. 8

Cloning of *Entamoeba* Genes Encoding Proteolipids of Putative Vacuolar Proton-Translocating ATPases

STEVEN DESCOTEAUX,[1] YI YU,[1,2] AND JOHN SAMUELSON[1,2]*

Department of Tropical Public Health, Harvard School of Public Health,[1] and Department of Pathology, New England Deaconess Hospital,[2] Boston, Massachusetts 02115

Received 11 March 1994/Returned for modification 12 April 1994/Accepted 23 May 1994

Molecular cloning techniques were used to identify genes encoding the proteolipids of putative vacuolar proton-transporting ATPases (V-ATPases; EC 3.6.1.35) of *Entamoeba histolytica* (Ehvma3) and *Entamoeba dispar*. The Ehvma3 gene encoded a 177-amino-acid peptide, with an M_r of 18,110, which showed extensive positional identities with peptides of *E. dispar* (92%), *Schizosaccharomyces pombe* (58%), and humans (56%).

The vacuolar proton-transporting ATPase (V-ATPase) of *Entamoeba histolytica*, the protozoan parasite that causes amebic dysentery, is of great interest because phagolysosomal acidification is necessary for contact-mediated cytolysis of host epithelial cells by these parasites (15). By analogy to macrophages (9), the *E. histolytica* V-ATPase is likely critical for killing of bacteria, which the parasites ingest within the colonic lumen (12). By analogy to osteoclasts, the *E. histolytica* V-ATPase may be present on the plasma membrane and so be involved in maintenance of cytosolic pH and secretion of acid (4, 9). Evidence for the presence of an *E. histolytica* V-ATPase comes from inhibition of acidification of pinocytotic vesicles by bafilomycin A, a specific inhibitor of V-ATPases (5, 11), and cloning of an *E. histolytica* gene (Ehvma1) encoding the catalytic peptide A of putative V-ATPase (2, 20).

Acidification of intracellular compartments including endosomes, lysosomes, secretory granules, and the lumen of the endoplasmic reticulum is caused by ATP-driven, proton-transporting pumps (V-ATPases) encoded by vacuolar membrane ATPase (*vma*) genes (2). These pumps, which are related in structure to the proton transporters of bacteria, mitochondria, and chloroplasts (F-ATPases or F_0F_1 ATPases [8]), are composed of at least 10 different peptides. The major component of the transmembrane segment of the V-ATPase is a 17-kDa peptide which is often referred to as the proteolipid because it partitions with the organic phase of a chloroform-methanol extraction (2). Indeed, the 17-kDa proteolipid contains no lipid but is composed of four very hydrophobic domains, which form the proton channel (13). A conserved Glu residue within the fourth hydrophobic domain of the *Saccharomyces cerevisiae* proteolipid (*vma3* gene product) is covalently labeled by *N,N'*-dicyclohexylcarbodiimide (DCCD), an irreversible inhibitor of V-ATPases (3, 14). A second specific inhibitor of V-ATPases is bafilomycin A, which produces the same phenotype in *S. cerevisiae* as disabling mutations of the V-ATPase genes do (5).

With the goal of better characterizing the amebic V-ATPase, genes encoding putative V-ATPase proteolipids were cloned from *E. histolytica* (called Ehvma3 after the *S. cerevisiae* gene) and from *Entamoeba dispar* (Edvma3), which is genetically, biochemically, and immunologically different from *E. histolytica* and frequently colonizes the human colon but is not associated with dysentery (6). Briefly, a 129-bp PCR product (Ehvma3-PCR) was made from *E. histolytica* HM-1:IMSS genomic DNA by using a sense primer [GC(AT)(AT)(AT)(AT)GG(AT)AC(AT)GC(AT)AA] to the hexapeptide Ala-Tyr/Ile-Gly-Thr-Ala-Lys and an antisense primer [(AT)(GC)C(AT)AC(AT)ATTA(AG)(AT)CC(AG)TA] to the hexapeptide Tyr-Gly-Leu-Ile-Val-Gly/Ala conserved in V-ATPase proteolipids (Fig. 1) (2, 13, 17). Next, Ehvma3-PCR was radiolabeled and hybridized to *E. histolytica* HM-1:IMSS genomic DNA and cDNA libraries and to an *E. dispar* SAW 142 cDNA library. A genomic clone from the *E. histolytica* DNA library (Ehvma3-gDNA1), which was sequenced completely on both strands by using vector and synthesized primers and the dideoxy method (17), contained the entire Ehvma3 open reading frame (ORF) (GenBank accession number U01057). An *E. histolytica* cDNA clone (Ehvma3-cDNA) encoded the entire ORF less the first 2 amino acids (aa) and included a 70-bp 3′ noncoding sequence containing a 27-poly(A) tail. An *E. dispar* cDNA clone (Edvma3-cDNA; GenBank accession number U01055), which was sequenced completely on both strands, encoded the entire ORF, a 13-bp 5′ noncoding sequence, and a 63-bp 3′ noncoding sequence including a 20-poly(A) tail.

The Ehvma3 gene had a 177-aa ORF with an M_r of 18,110 and a net charge of +1 at pH 7 (Fig. 1). Because there was an in-frame stop codon in the 5′ noncoding region of the Ehvma3 gene (data not shown), the Ehvma3 ORF could not be longer. However, the start of Ehvma3 translation may be Met-3, which was the start of the Edvma3 ORF (Fig. 1). The Ehvma3 coding region showed 93% nucleotide and 92% amino acid identities with that of Edvma3, which had a 176-aa ORF and an M_r of 18,082. Eighteen of the 33 base pair differences between Ehvma3 and Edvma3 coding regions were silent, while 9 of 13 amino acid changes were conservative. The 3′ noncoding sequences of the Ehvma3- and Edvma3-cDNAs were also remarkably similar, including a shared start of the poly(A) tails 44 bp after the stop codon (data not shown). Like most ameba genes, the Ehvma3 and Edvma3 ORFs contained no introns, and 88 and 86% of the third positions of the codons of Ehvma3 and Edvma3 ORFs, respectively, were A or T (not including Met and Trp).

When *E. histolytica* HM-1:IMSS DNA was digested with *Eco*RI, *Rsa*I, and *Taq*I, each of which cut once in the Ehvma3 ORF, and probed with a ^{32}P-labeled 173-bp segment of the Ehvma3 gene (Ehvma3 probe) beginning at Ser-121 of the ORF and ending at the stop codon, there was a single band in each lane of the Southern blots (data not shown). This result suggested that the Ehvma3 gene was single copy since it is very

* Corresponding author. Phone: (617) 432-4670. Fax: (617) 738-4914. Electronic mail address: descot@mbcrr.harvard.edu.

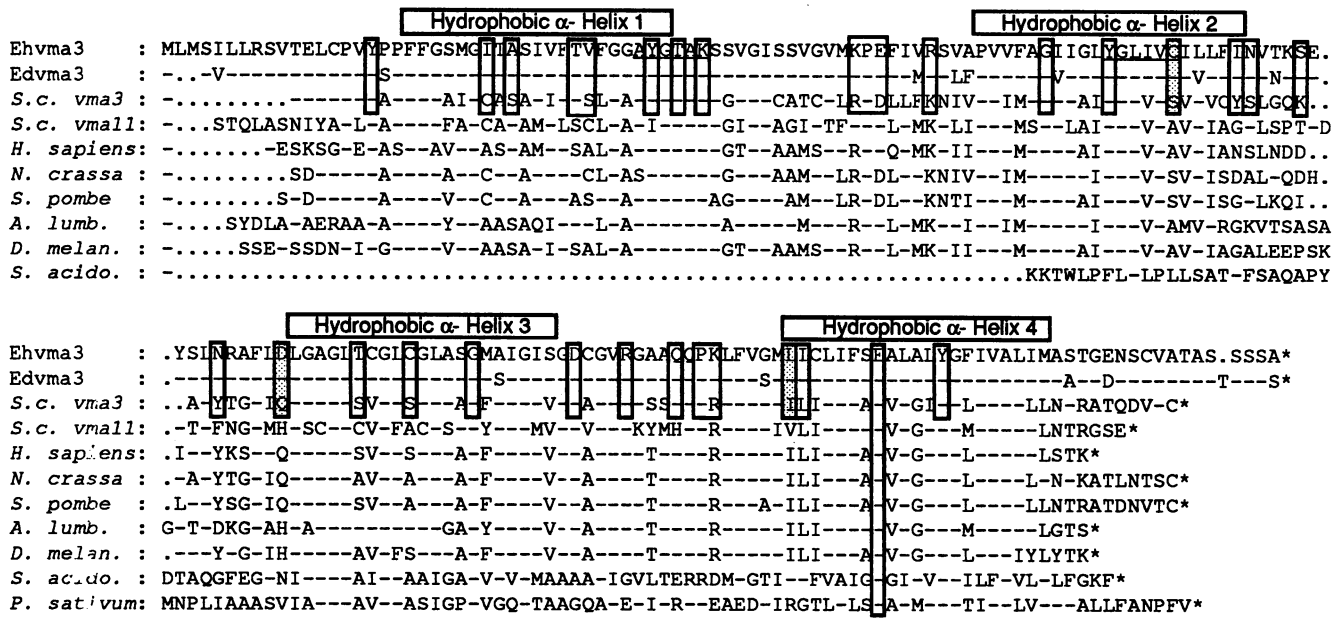

FIG. 1. Alignment of the putative *E. histolytica* V-ATPase proteolipid (Ehvma3) in single-letter code with those of *E. dispar* (Edvma3), *S. cerevisiae vma3* (GenBank accession number X15155 [13]) and *vma11* (D10486 [19]), *Homo sapiens* (polycystic kidney disease gene [M62762]), *N. crassa* (L07105), *S. pombe* (X59947), *Ascaris lumbricoides* (M33757), *Drosophila melanogaster* (X55979), and *S. acidocaldarius* (J04740), identified from the database at the National Center for Biotechnology Information by using the Blast network service (1). Also aligned is the F-ATPase proteolipid of garden pea chloroplasts (*Pisum sativum* [M57711]). Underlined amino acids in the Ehvma3 ORF mark the location of the sense and the antisense primers used to amplify the 129-bp Ehvma3 PCR product from *E. histolytica* genomic DNA. Dashes are present where the sequences are identical to the Ehvma3 ORF, periods mark gaps, and asterisks mark the ends of the ORFs. Ehvma3 hydrophobic α-helices are numbered from the amino terminus to the carboxyl terminus and enclosed in horizontal boxes (7). Residues studied by mutagenesis in *S. cerevisiae vma3* (14) are enclosed in small vertical boxes. These boxes are open if the Ehvma3 and Edvma3 ORFs showed positional identities or conservative substitutions (14) and shaded if the ameba proteolipids appear to have incompatible substitutions (14). The long vertical box marks the Glu-147 residue, which in other proteolipids has been shown to bind DCCD (3, 14).

unlikely that multiple Ehvma3 genes contained three identical restriction enzyme sites in their 3′ noncoding regions. In addition, the sequences of overlapping segments of the Ehvma3 PCR products and genomic DNA and cDNA clones were all identical. However, we cannot rule out the possibility that there exists another Ehvma3 gene, which did not cross-hybridize with the Ehvma3 probe used for the Southern blot or the Ehvma3 PCR product used to screen the cDNA and genomic DNA libraries. On Northern (RNA) blots of *E. histolytica*, the Ehvma3 probe identified a ~600-base band (data not shown).

The extreme hydrophobicity of the predicted Ehvma3 proteolipid was reflected in its overall amino acid composition (66% hydrophobic [GAVLIMFW or P], 23% neutral [STNQ or C], and 8% hydrophilic [DEKHR or Y]). The Ehvma3 ORF encoded neither Trp nor His residues, as is the case with *S. cerevisiae vma3* proteolipid (13). The hydropathy plot of the Ehvma3 ORF (10), which included four hydrophobic α-helices that form the putative transmembrane proton channel (7), was nearly superimposable onto that of *S. cerevisiae vma3* (Fig. 2) (13). The putative ectoplasmic loop between hydrophobic α-helices 1 and 2 (Fig. 2, asterisk) was hydrophobic in the ORFs of both Ehvma3 and *S. cerevisiae vma3*.

The 177-aa Ehvma3 ORF was 16 to 22 aa longer than the ORFs of *S. cerevisiae vma3* (160 aa) (13), *Schizosaccharomyces pombe* (161 aa), and human PDK1 (polycystic kidney disease gene; 155 aa) (Fig. 1) (1). Conserved in the Ehvma3 ORF was Glu-147, which was located in the fourth hydrophobic α-helix and has been shown in other V-ATPase proteolipids to covalently bind the irreversible inhibitor DCCD (3, 14). Of 33 aa present in *S. cerevisiae vma3* that have been tested with site-directed mutagenesis (14), 17 were identical in the Ehvma3 ORF and 13 showed conservative changes or nonconservative changes that did not inhibit function in the yeast mutants (Fig. 1, open boxes). Three amino acid substitutions in the Ehvma3 ORF were nonconservative and, in the yeast experiments, inhibited function (Fig. 1, shaded boxes).

The carboxyl half of the Ehvma3 ORF showed 24 to 25% positional identities, including conservation of the DCCD-binding Glu-147 residue, with the ~8-kDa proteolipids present in the V-ATPase of *Sulfolobus acidocaldarius*, an archaebacterium, and the F-ATPase of the garden pea chloroplast (Fig. 1) (8). Although it has been argued that the genes encoding the eukaryotic 17-kDa V-ATPase proteolipids result from duplication of a progenitor gene encoding an 8-kDa proteolipid (8, 13), the carboxyl half of the Ehvma3 ORF showed only 14% positional identity with the overlapping regions of the amino half of the Ehvma3 ORF. As is the case with other V-ATPases, the DCCD-binding Glu-147 present in the fourth hydrophobic α-helix of the Ehvma3 ORF was replaced by Gly in the second Ehvma3 hydrophobic α-helix (2, 13). F-ATPases and the archaebacterial V-ATPases, which both contain 8-kDa proteolipids, also use the proton motive force to make ATP (i.e., they act as an ATP synthase) (8, 13). In contrast, eukaryotic V-ATPases cannot function as ATP synthases, suggesting that this activity was lost after gene duplication and formation of the 17-kDA proteolipid (8, 13).

In overlapping regions, the Ehvma3 ORF showed extensive positional identities with *S. cerevisiae vma3* (53%), *S. pombe* (58%), and human PDK1 (56%) (Fig. 1). These positional

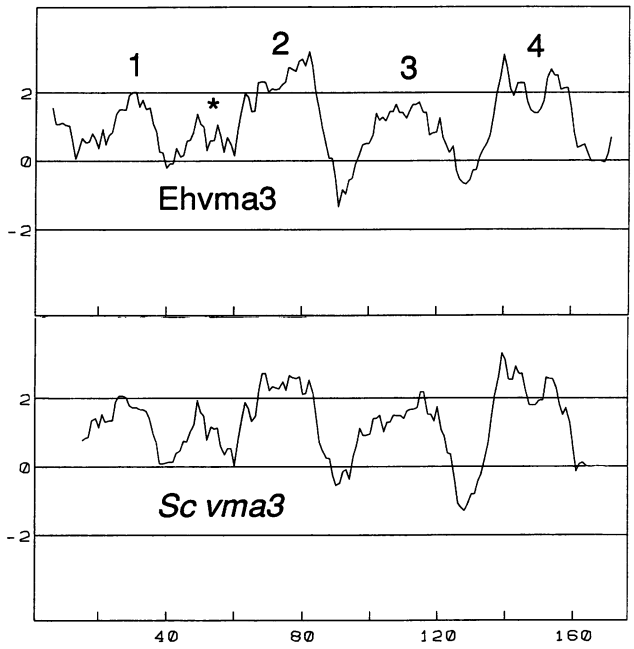

FIG. 2. Hydropathy plots of the ORFs of Ehvma3 and *S. cerevisiae vma3* (13) made with the algorithm of Kyte and Doolittle (10) by using a window of 11 aa. Hydrophobic α-helices, representing transmembrane segments of the Ehvma3 proteolipid that compose the putative channel, are numbered from the amino terminus to the carboxyl terminus (7). An asterisk marks the ectoplasmic segment between α-helices 1 and 2.

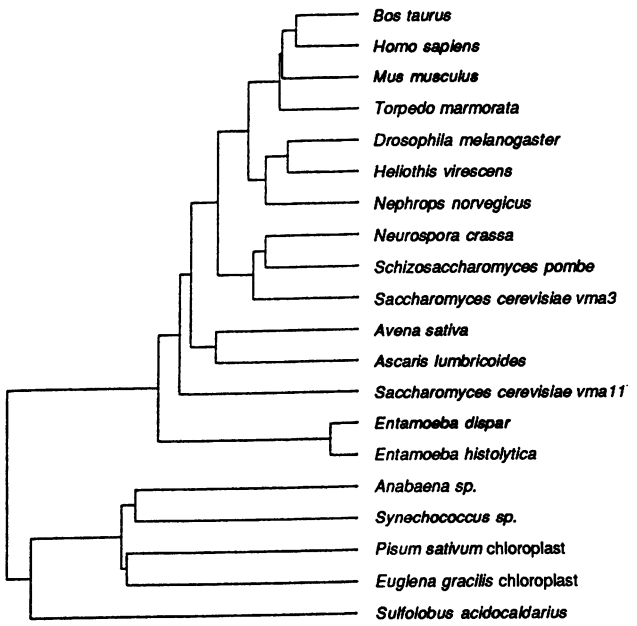

FIG. 3. Phylogenetic tree drawn with a maximal-alignment rule of 16 V-ATPase proteolipids and four F-ATPase proteolipids. The V-ATPase proteolipids included those aligned in Fig. 1 as well as those of *Mus musculus* (mouse [GenBank sequence accession number M64298]), *Nephrops norvegicus* (lobster [S40059]), *Bos taurus* (cow chromaffin granules [J03835]), *Torpedo marmorata* (marbled electric ray [X52002]), *Heliothis virescens* (moth [L16884]), and *Avena sativa* (oat [M73232]). The F-ATPase proteolipids included those of chloroplasts (garden pea [Fig. 1] and *Euglena gracilis* [M16844]) and cyanobacteria (*Anabaena* sp. [M21659] and *Synechococcus* sp. [X05302]).

identities were considerably less than those shared by the human and yeast proteolipids (69 to 77%). These positional identities were also somewhat less than those shared by the ameba V-ATPase catalytic peptide A with human PDK1 (66%) and *S. pombe* (63%) (20). The Ehvma3 ORF showed a 44% positional identity with the ORF of the *S. cerevisiae vma11* gene, which itself shows a 58% positional identity with *S. cerevisiae vma3* (19). The *S. cerevisiae vma11* gene is necessary for *S. cerevisiae* V-ATPase function even though the encoded proteolipid has not been identified in the yeast V-ATPase (19).

A phylogenetic tree containing 20 V- and F-ATPase proteolipid sequences from 19 species was drawn by means of a pattern-induced multisequence alignment program using a maximal-linkage rule (Fig. 3) (18). In this tree, 16 eukaryotes formed one large group or clade, at the edge of which were *E. histolytica* and *E. dispar*. A second clade included cyanobacteria, chloroplasts, and the archaebacterium, which all contain 8-kDa proteolipids. In a phylogenetic tree of V-ATPase catalytic peptides A, amebae are closer to humans than are *S. pombe*, *Neurospora crassa*, and *Plasmodium falciparum* (20), and archaebacteria are present in the eukaryotic V-ATPase clade rather than the F-ATPase clade (2).

These are the first V-ATPase proteolipid genes identified from a protozoan parasite, although V-ATPase catalytic peptide A genes from *E. histolytica*, *P. falciparum*, and *Trypanosoma congolense* have been identified. The 7% amino acid difference between the Ehvma3 and Edvma3 ORFs was consistent with previous reports showing 5 to 20% positional differences in the ORFs of homologous genes of *E. histolytica* and *E. dispar* parasites (6). It is not known whether (i) these amino acid substitutions, of which only four were nonconservative, can explain marked differences in the invasiveness of *E. histolytica* and *E. dispar* (6), (ii) phagocytosis- and virulence-deficient mutants of *E. histolytica* have defects in vacuolar acidification (16), (iii) the Ehvma3 gene will complement *S. cerevisiae vma3* or *vma11* mutants (13, 14, 19), (iv) bafilomycin A will inhibit amebic killing of bacteria, or (v) antibodies directed at the Ehvma3 ectoplasmic loop between α-helices 1 and 2 will bind to the parasite surface and/or produce protective immunity to the amebae in animal models.

This work was supported by NIH grant AI28395 and by the John D. and Catherine T. MacArthur Foundation. Y. Yu was supported in part by a Pan American Health Organization Fellowship.

We thank Egbert Tannich for the cDNA libraries and Tom Graf and Temple Smith for help with the sequence analyses.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.
2. **Anraku, Y., R. Hirata, Y. Wada, and Y. Ohya.** 1992. Molecular genetics of the yeast vacuolar H^+-ATPase. J. Exp. Biol. **172**:67–81.
3. **Arai, H., M. Berne, and M. Forgac.** 1987. Inhibition of the coated vesicle proton pump and labeling of a 17,000-dalton polypeptide by *N,N'*-dicyclohexylcarbodiimide. J. Biol. Chem. **262**:11006–11011.
4. **Bakker-Grunwald, T.** 1992. Ion transport in parasitic protozoa. J. Exp. Biol. **172**:311–322.
5. **Bowman, E. J., A. Siebers, and K. Altendorf.** 1988. Bafilomycins: a class of inhibitors of membrane ATPases from microorganisms, animal cells and plant cells. Proc. Natl. Acad. Sci. USA **85**:7972–7976.
6. **Diamond, L. S., and C. G. Clarke.** 1993. A redescription of *Entamoeba histolytica* Schaudinn, 1903 (emended Walker, 1911)

separating it from *Entamoeba dispar* Brumpt, 1925. J. Eukaryot. Microbiol. **40:**340–344.
7. **Eisenberg, D., E. Schwartz, M. Komaromy, and R. Wall.** 1984. Analysis of membrane and surface protein sequences with a hydrophobic moment plot. J. Mol. Biol. **179:**125–142.
8. **Futai, M., T. Noumi, and M. Maeda.** 1989. ATP synthase (H^+-ATPase): results by combined biochemical and molecular biological approaches. Annu. Rev. Biochem. **58:**111–136.
9. **Grinstein, S., A. Nanda, G. Lukacs, and O. Rotstein.** 1992. V-ATPases in phagocytic cells. J. Exp. Biol. **172:**179–192.
10. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
11. **Lohden-Bendinger, U., and T. Bakker-Grunwald.** 1990. Evidence for a vacuolar-type proton ATPase in *Entamoeba histolytica*. Z. Naturforsch. Sect. C **45:**229–232.
12. **Mirelman, D.** 1987. Ameba-bacterium relationship in amebiasis. Microbiol. Rev. **51:**272–284.
13. **Nelson, H., and N. Nelson.** 1989. The progenitor of ATP synthases was closely related to the current vacuolar H^+-ATPase. FEBS Lett. **247:**147–153.
14. **Nuomi, T., C. Beltran, H. Nelson, and N. Nelson.** 1991. Mutational analysis of yeast vacuolar H^+-ATPase. Proc. Natl. Acad. Sci. USA **88:**1938–1942.
15. **Ravdin, J. I., P. H. Schlesinger, C. F. Murphy, I. Y. Gluzman, and D. J. Krogstad.** 1986. Acid intracellular vesicles and cytolysis of mammalian target cells by *Entamoeba histolytica* trophozoites. J. Protozool. **33:**478–486.
16. **Rodriguez, M. A., and E. Orozco.** 1986. Isolation and characterization of phagocytosis- and virulence-deficient mutants of *Entamoeba histolytica*. J. Infect. Dis. **154:**27–32.
17. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Smith, R. F., and T. F. Smith.** 1992. Pattern-induced multi-sequence alignment (PIMA) algorithm employing secondary structure-dependent gap penalties for use in comparative protein modelling. Protein Eng. **5:**35–41.
19. **Umemoto, N., Y. Ohya, and Y. Anraku.** 1991. *VMA11*, a novel gene that encodes a putative proteolipid, is indispensable for expression of yeast vacuolar membrane H^+-ATPase activity. J. Biol. Chem. **266:**24526–24532.
20. **Yu, Y., and J. Samuelson.** Mol. Biochem. Parasitol., in press.