An Alignment of 17 Deduced Protein Sequences from Plant, Fungi, and Ciliate H$^+$-ATPase Genes

Achim Wach,[1] Alain Schlesser,[1] and André Goffeau[1]

Seventeen protein sequences of H$^+$-ATPases from plants (*Arabidopsis thaliana, Nicotiana plumbaginifolia, Lycopersicum esculentum*), fungi (*Saccharomyces cerevisiae, Schizosaccharomyces pombe, Zygosaccharomyces rouxii, Neuropora crassa, Candida albicans*), and a parasitic ciliate (*Leishmania donovani*) have been aligned. Twenty sequence fragments were identified which were conserved in H$^+$-, Na$^+$/K$^+$-, and Ca^{++} plasma membrane-ATPases. In addition, a total of 118 residues not located in these fragments were found to be conserved in all H$^+$-ATPases. Among those, 38 amino acid residues were screened out as being priority targets for site–directed mutagenesis experiments aimed at the identification of the amino acid residues specifically involved in cation specificity.

INTRODUCTION

That the yeast plasma membrane H$^+$-ATPase, initially identified by physiological studies (see review by Goffeau and Slayman, 1981), is structurally similar to the mammalian Ca^{2+}- and Na$^+$/K$^+$-ATPases has been inferred from the observation of a predominant 100-kDa polypeptide in an ATPase solubilized and purified from the plasma membranes of the fission yeast *Schizosaccharomyces pombe* (Dufour and Goffeau, 1978). A similar subunit composition has been reported for the transport ATPase enzyme isolated from the mold *Neurospora crassa* (Bowman *et al.*, 1981; Addison and Scarborough, 1981) and from the budding yeast *Saccharomyces cerevisiae* (Malpartida and Serrano, 1980; Foury *et al.*, 1981; Koland and Hammes, 1986; Perlin *et al.*, 1989; Wach *et al.*, 1990). The biochemical characterization of the three fungal enzymes (see review by Goffeau and Slayman, 1981) has established the electrogenic H$^+$- pumping capacity of the reconstituted purified enzyme and the similarity in

mechanisms with the mammalian Ca^{2+}- and Na$^+$/K$^+$-ATPases. In particular, the presence of an aspartyl-phosphate catalytic intermediate and the sensitivity of the ATPase activity to vandate were established for the three fungal enzymes. A similar class of plant H$^+$-ATPase was reported more recently (see review by Serrano, 1989).

Since 1984, at least 16 reviews reporting further properties of H$^+$-ATPases from fungi and plant plasma membranes have been published (Serrano, 1984; Scarborough, 1985; Bowman and Bowman, 1986; Goffeau and Boutry, 1986; Goffeau, 1987; Slayman, 1987; Goffeau, 1988; Serrano, 1988; Goffeau *et al.*, 1989; Nakamoto and Slayman, 1989; Serrano, 1989; Goffeau and Green, 1990; Serrano, 1991; Gaber, 1992; Scarborough, 1992).

Among this wealth of largely redundant information, at least two alignments of fungal and plant H$^+$-ATPase protein sequences have been previously proposed (Serrano, 1989; Goffeau and Green, 1990). However, during these last three years several new genes encoding presumed H$^+$-ATPases from plant and fungi have been reported but not yet listed and compared to each other. Therefore in the present

[1] Unité de Biochimie Physiologique, Université Catholique de Louvain, Place Croix du Sud 2-20, B-1348 Louvain-la-Neuve, Belgium.

```
                 10        20        30        40        50        60        70        80         90       100
                  |         |         |         |         |         |         |         |          |         |
H-PM,  Yeast    1 MTDTSSSSSSSSASSVSAHQPTQEKPAKTYDDAASESSDDDDIDALIEELQSNHGVDDEDSDNDGPVAAGEARLVPEEYLQTDPSY GLT SDEVLKRR--KKY GLN
NaK-PM,Sheep    1 --------------------------MGKGVGRDKYEPAAVSEHGDKKKAKKERDMDELKKEVSMDDHKLSLDELHRKYGTDLNR       GLT TARAAEIL--ARD GPN
Ca-PM, Rat      1 MGDMANNSVAYSGVKNSLKEANHDGDFGITLAELRALMELRSTDALRKIQESYGDVYGICTKLK---------------TSPNE       GLS GNPADLERREAVF GKN
                                                                                                 xx                x x
                                                                                               a (96%)          b (80%)


                 110       120       130                                              140       150       160       170       180
                  |         |         |                                                |         |         |         |         |
H-PM,  Yeast  104 QM-ADEKESLVVKFVMFFVGPIQF VMEAAAILAAGL -------------------------SDWVDFGVICGLLMLNAGVGFVQEFQAGSIVDELKKTLANTAV
NaK-PM,Sheep   77 ALTPPPTTPEWVKFCRQLFGGFSM LLWIGAVLCFLA YGIQAATEEEPQN--------------DNLYLGVVLSAVVIITGCFSYYQEAKSSKIMESFKNMVPQQAL
Ca-PM, Rat     89 FIPPKKPKTFLQLVWEALQDVTLI ILEIAAIVSLGL SFYQPPEGDNALCGEVSVGEEEGEGETGWIEGAAILLSVVCVVLVTAFNDWSKEKQFRGLQSRIEQEQKFT
                                           x
                                       c (71%)


                 190       200       210       220       230       240       250       260       270
                  |         |         |         |         |         |         |         |         |
H-PM,  Yeast  182 VIRDGQ LVEIPAN EVVPGDI LQLED GTIIPTD GRIVTEDCFLQ IDQSAITGESL AVDK--HYGDQT--FSSSTVKRGEGF MVVTATGDNTFVG RAAAL
NaK-PM,Sheep  169 VIRNGE KMSINAE EVVVGDL VEVKG GDRIPAD LRIISANG-CK VDNSSLTGESE PQTRSPDFTNENPLETRNIAFFSTNC VEGTARGIVVYTG DRTVM
Ca-PM, Rat    196 VIRGGQ VIQIPVA DITVGDI AQVKY GDLLPAD GILIQGND-LK IDESSLTGESD HVKKS--LDKDPLLLSGTHVMEGSGR MVVTAVGVNSQTG IIFTL
                  xxx x          xx          x   x x            x x  xxxx                          xx x      x
                  d (91%)       e (86%)     f (83%)            g (87%)                           h (73%)


                 280
                  |
H-PM,  Yeast  276 VNKAA------------------------------------------- GGQG -------------------------
NaK-PM,Sheep  266 GRIATLASGL-------------------------------------E GGQT PIAAEIE--------------------
Ca-PM, Rat    291 LGAGGEEEEKKDEKKKEKKNKKQDGAIENRNKAKAQDGAAMEMQPLKSEE GGDG DEKDKKKANLPKKEKSVLQGKLTKLAVQ
                                                                   xx
                                                                 i (83%)


                 290       300       310       320       330       340       350       360       370
                  |         |         |         |         |         |         |         |         |
H-PM,  Yeast  285 HFTEVLNGIGIILLVLVIATLLLVWTACFYRTNGIVR---------ILRYTLGITIIG VPVGLPAVVTTT MAVG AAYLAKKQAIVQKLSAIESLA GVEI
NaK-PM,Sheep  288 HFIHIITGVAVFLGVSFFILSLILE----------------YTWLEAVIFLIGIIVAN VPEGLLATVTVC LTLT AKRMARKNCLVKNLEAVETLG STST
Ca-PM, Rat    373 IGKAGLLMSAITVIILVLYFVIDTFWVQKRPWLAECTPIYIQYFVKFFIIGVTVLVVA VPEGLPLAVTIS LAYS VKKMMKDNNLVRHLDACETMG NATA
                                                                   xx xx  xx                x  x  x
                                                                 j (83%)                  k (76%)
```
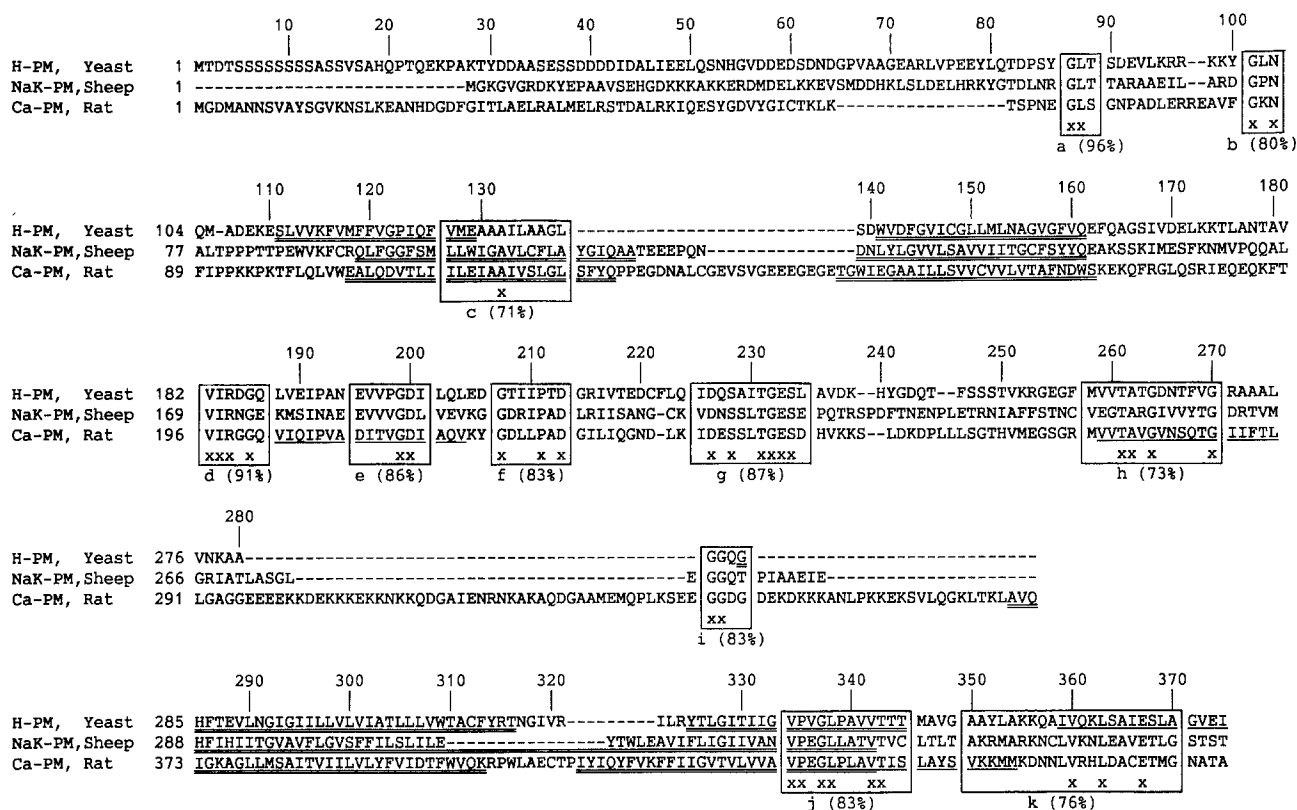
Fig. 1. Alignment of P-type ATPase sequences. Amino acid sequences are from *Saccharomyces cerevisiae* H⁺-ATPase (H-PM, Yeast), Na⁺/K⁺-ATPase from sheep (NaK-PM, Sheep), and Ca²⁺-ATPase from rat plasma membranes (Ca-PM, Rat). See text for references. The alignment was initially performed by using the program CLUSTAL (Higgins and Sharp, 1988) and then refined manually. The numbers on top of each block are positions in ScPMA1. Overall homologies were calculated by pairwise comparison of the sequences in each cluster using the structure-genetic matrix from Feng, *et al.* (1985). The clusters with mean homology scores above 70% are boxed and labeled *a* to *s*. Underlined sequences indicate predicted hydrophobic regions by the method of Klein *et al.* (1985): membrane integral sequences are underlined twice, and possibly membrane integral sequences are underlined once.

mini-review we will focus on the alignment of the 17 deduced H⁺-ATPase complete protein sequences reported, to our knowledge, up to January 1992. This will provide up-to-date information which might be useful for decisions concerning future site-directed mutagenesis or domain-shuffling experiments.

IDENTIFICATION OF 20 CONSERVED REGIONS IN THE Ca²⁺, Na⁺/K⁺, AND H⁺-ATPases

Green (1989) has identified about 30 amino acid sequence segments which share more than 15% mean identity in pairwise comparisons between skeletal Ca²⁺-ATPase versus sheep Na⁺/K⁺-ATPase, of sheep Na⁺/K⁺-ATPase versus yeast H⁺-ATPase, and of yeast H⁺-ATPase versus *E. coli* K⁺-ATPase (see also Goffeau and Green, 1990 and Van Dyck *et al.*, 1990).

Figure 1 represents an alignment, first performed by the CLUSTAL program (Higgins and Sharp, 1988) and subsequently refined manually, of three different eukaryotic plasma membrane ATPases pumping either Ca²⁺ (Schull and Greeb, 1988), Na⁺/K⁺ (Schull *et al.*, 1985), or H⁺ (Serrano *et al.*, 1986, as corrected by Van Dyck *et al.*, 1990). The 20 sequence fragments showing a mean homology higher than 70% as calculated by pairwise comparison of the amino acid clusters quantified by the structure-genetic matrix from Feng *et al.* (1985) are boxed and marked by the letters *a* to *s*. The length of the homologous segments *a* to *s* varies from 3 to 39 amino residues, and the mean percent homology varies from 71 to 96%. Since these 20 segments *a* to *s* are well conserved in all three pumps, they are not likely to be involved in cation specificity but rather in the mechanism of ATP hydrolysis or in the mechanism of energy transduction which is considered to be basically similar for all three transport-ATPases.
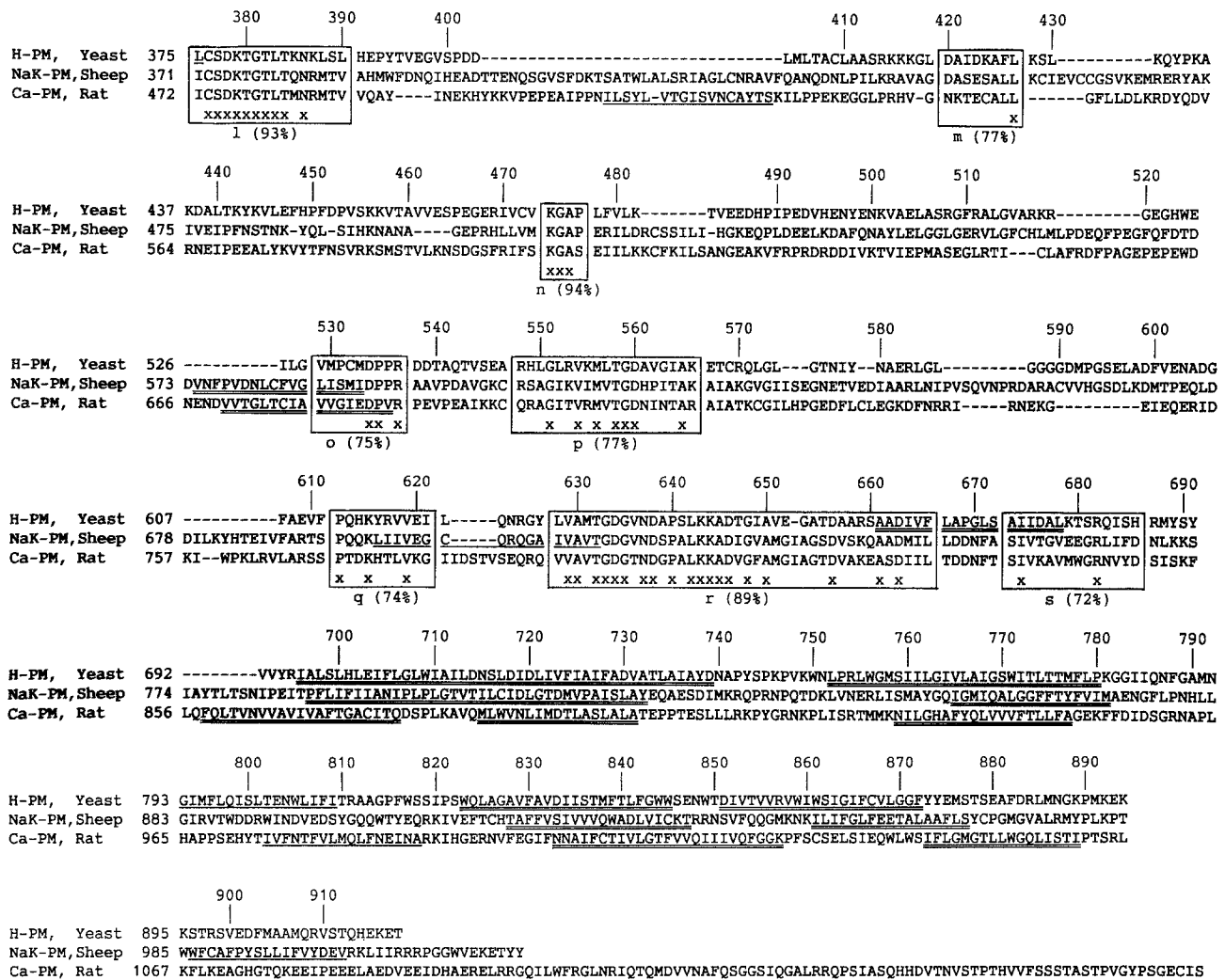
```
                 380       390       400                                                              410       420       430
                 |         |         |                                                                |         |         |
H-PM,  Yeast  375 LCSDKTGTLTKNKLSL HEPYTVEGVSPDD--------------------------------LMLTACLAASRKKKGL DAIDKAFL KSL---------KQYPKA
NaK-PM,Sheep  371 ICSDKTGTLTQNRMTV AHMWFDNQIHEADTTENQSGVSFDKTSATWLALSRIAGLCNRAVFQANQDNLPILKRAVAG DASESALL KCIEVCCGSVKEMRERYAK
Ca-PM, Rat    472 ICSDKTGTLTMNRMTV VQAY----INEKHYKKVPEPEAIPPNILSYL-VTGISVNCAYTSKILPPEKEGGLPRHV-G NKTECALL ------GFLLDLKRDYQDV
                 xxxxxxxxx x                                                                                  x
                  l (93%)                                                                             m (77%)

                 440       450       460       470       480       490       500       510       520
                 |         |         |         |         |         |         |         |         |
H-PM,  Yeast  437 KDALTKYKVLEFHPFDPVSKKVTAVVESPEGERIVCV KGAP LFVLK-------TVEEDHPIPEDVHENYENKVAELASRGFRALGVARKR---------GEGHWE
NaK-PM,Sheep  475 IVEIPFNSTNK-YQL-SIHKNANA----GEPRHLLVM KGAP ERILDRCSSILI-HGKEQPLDEELKDAFQNAYLELGGLGERVLGFCHLMLPDEQFPEGFQFDTD
Ca-PM, Rat    564 RNEIPEEALYKVYTFNSVRKSMSTVLKNSDGSFRIFS KGAS EIILKKCFKILSANGEAKVFRPRDRDDIVKTVIEPMASEGLRTI---CLAFRDFPAGEPEPEWD
                                                       xxx
                                                    n (94%)

                 530       540       550       560       570       580       590       600
                 |         |         |         |         |         |         |         |
H-PM,  Yeast  526 ----------ILG VMPCMDPPR DDTAQTVSEA RHLGLRVKMLTGDAVGIAK ETCRQLGL---GTNIY--NAERLGL---------GGGGDMPGSELADFVENADG
NaK-PM,Sheep  573 DVNFPVDNLCFVG LISMIDPPR AAVPDAVGKC RSAGIKVIMVTGDHPITAK AIAKGVGIISEGNETVEDIAARLNIPVSQVNPRDARACVVHGSDLKDMTPEQLD
Ca-PM, Rat    666 NENDVVTGLTCIA VVGIEDPVR PEVPEAIKKC QRAGITVRMVTGDNINTAR AIATKCGILHPGEDFLCLEGKDFNRRI-----RNEKG---------EIEQERID
                                xx x                     x   x x xxx     x
                   o (75%)                              p (77%)

                 610       620       630       640       650       660       670       680       690
                 |         |         |         |         |         |         |         |         |
H-PM,  Yeast  607 ----------FAEVF PQHKYRVVEI L-----QNRGY LVAMTGDGVNDAPSLKKADTGIAVE-GATDAARSAADIVF LAPGLS AIIDALKTSRQISH RMYSY
NaK-PM,Sheep  678 DILKYHTEIVFARTS PQQKLIIVEG C-----QRQGA IVAVTGDGVNDSPALKKADIGVAMGIAGSDVSKQAADMIL LDDNFA SIVTGVEEGRLIFD NLKKS
Ca-PM, Rat    757 KI--WPKLRVLARSS PTDKHTLVKG IIDSTVSEQRQ VVAVTGDGTNDGPALKKADVGFAMGIAGTDVAKEASDIIL TDDNFT SIVKAVMWGRNVYD SISKF
                                 x   x    x                 xx xxxx xx x xxxxx x x        x       x x              x    x
                     q (74%)                               r (89%)                               s (72%)

                 700       710       720       730       740       750       760       770       780       790
                 |         |         |         |         |         |         |         |         |         |
H-PM,  Yeast  692 --------VVYRIALSLHLEIFLGLWIAILDNSLDIDLIVFIAIFADVATLAIAYDNAPYSPKPVKWNLPRLWGMSIILGIVLAIGSWITLTTMFLPKGGIIQNFGAMN
NaK-PM,Sheep  774 IAYTLTSNIPEITPFLIFIIANIPLPLGTVTILCIDLGTDMVPAISLAYEQAESDIMKRQPRNPQTDKLVNERLISMAYGQIGMIQALGGFFTYFVIMAENGFLPNHLL
Ca-PM, Rat    856 LQFQLTVNVVAVIVAFTGACITQDSPLKAVQMLWVNLIMDTLASLALATEPPTESLLLRKPYGRNKPLISRTMMKNILGHAFYQLVVVFTLLFAGEKFFDIDSGRNAPL

                 800       810       820       830       840       850       860       870       880       890
                 |         |         |         |         |         |         |         |         |
H-PM,  Yeast  793 GIMFLQISLTENWLIFITRAAGPFWSSIPSWQLAGAVFAVDIISTMFTLFGWWSENWTDIVTVVRVWIWSIGIFCVLGGFYYEMSTSEAFDRLMNGKPMKEK
NaK-PM,Sheep  883 GIRVTWDDRWINDVEDSYGQQWTYEQRKIVEFTCHTAFFVSIVVVQWADLVICKTRRNSVFQQGMKNKILIFGLFEETALAAFLSYCPGMGVALRMYPLKPT
Ca-PM, Rat    965 HAPPSEHYTIVFNTFVLMQLFNEINARKIHGERNVFEGIFNNAIFCTIVLGTFVVQIIIVQFGGKPFSCSELSIEQWLWSIFLGMGTLLWGQLISTIPTSRL

                 900       910
                 |         |
H-PM,  Yeast  895 KSTRSVEDFMAAMQRVSTQHEKET
NaK-PM,Sheep  985 WWFCAFPYSLLIFVYDEVRKLIIRRRPGGWVEKETYY
Ca-PM, Rat   1067 KFLKEAGHGTQKEEIPEEELAEDVEEIDHAERELRRGQILWFRGLNRIQTQMDVVNAFQSGGSIQGALRRQPSIASQHHDVTNVSTPTHVVFSSSTASTPVGYPSGECIS
```

Fig. 1. Continued.

ALIGNMENT OF 17 AMINO ACID SEQUENCES FROM PUTATIVE H⁺-ATPases

Since the elucidation of the nucleotide sequence of the H^+-ATPase gene from *Saccharomyces cerevisiae* (Serrano *et al.*, 1986; Van Dyck *et al.*, 1990; Perlin *et al.*, 1989) and *Neurospora crassa* (Addison, 1986; Hager *et al.*, 1986), numerous additional H^+-ATPase gene sequences have been reported from fungi such as *Schizosacharomyces pombe* (Ghislain *et al.*, 1987; Ghislain and Goffeau, 1991), *Saccharomyces cerevisiae* (Schlesser *et al.*, 1988), *Candida albicans* (Monk *et al.*, 1991), *Zygosaccharomyces rouxii* (Watanabe *et al.*, 1991), as well as from the protozoal ciliate *Leishmania donovani* (Meade *et al.*, 1987) and plants such as *Nicotiana plumbaginifolia* (Boutry *et al.*, 1989; Perez *et al.*, 1992; Morieau and Boutry, personal communication),

Arabidopsis thaliana (Harper *et al.*, 1989, 1990; Pardo and Serrano, 1989; Houlné and Boutry, personal communication), or *Lycopersicom esculentum* (Ewing *et al.*, 1990). To our knowledge, a total of at least 17 entire sequences of presumed H^+-ATPase genes have been carried out from 1986 to 1991.

As seen in Fig. 2, the seven fungal ATPases align each other extremely well. The alignment of the nine plant ATPases is even more spectacular. If one excludes the first 31 to 172 (according to the species) N-terminal residues and the last 43 to 104 (according to the species) C-terminal residues, it is easy to align the plant and the fungal ATPase by introduction of 11 block deletions in the fungal sequences and 10 block deletions in the plant sequences. The deletions in fungal sequences ranging from 1 to 13 residues generally correspond to "insertions" in the plant sequences, and conversely the
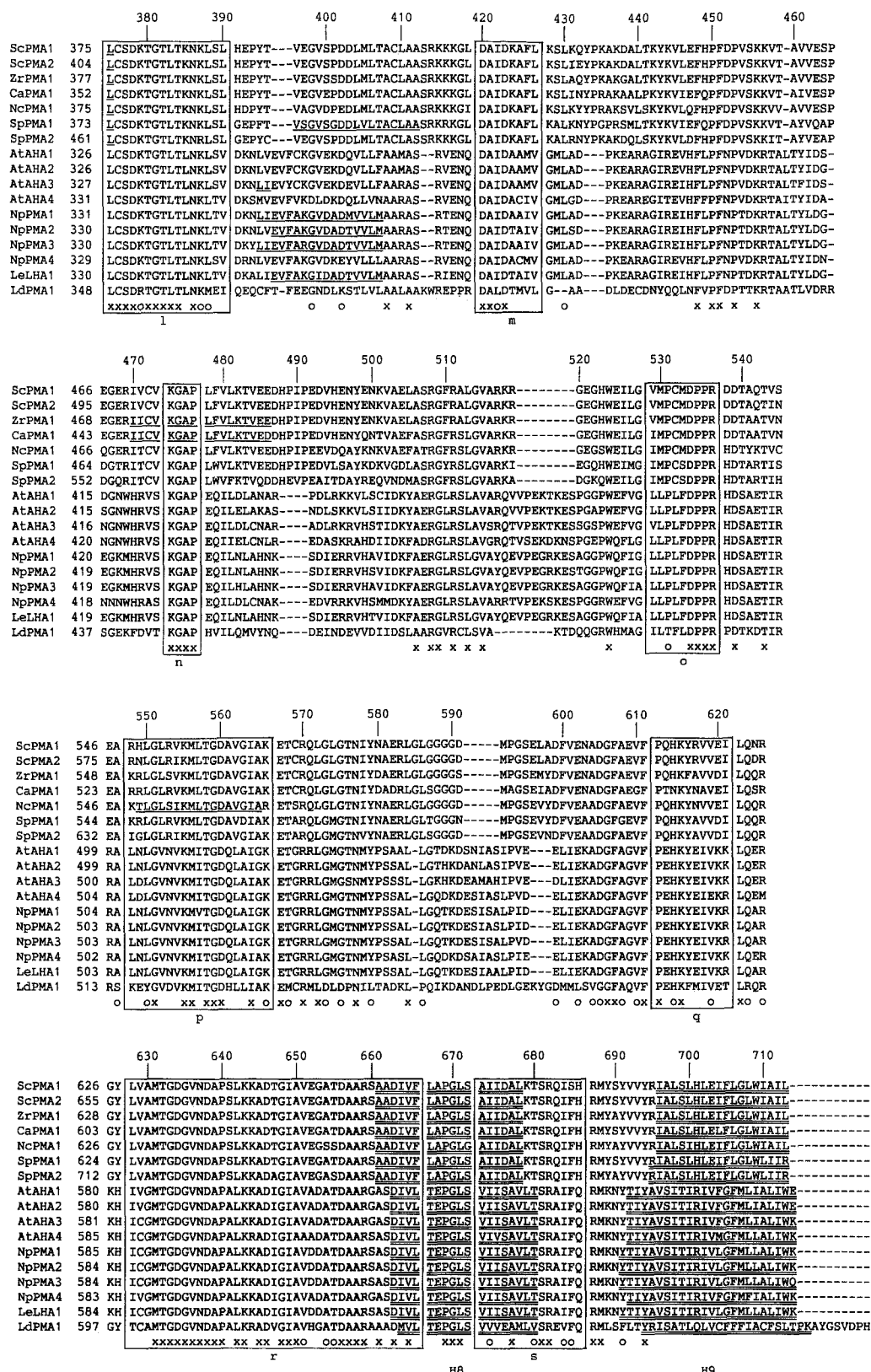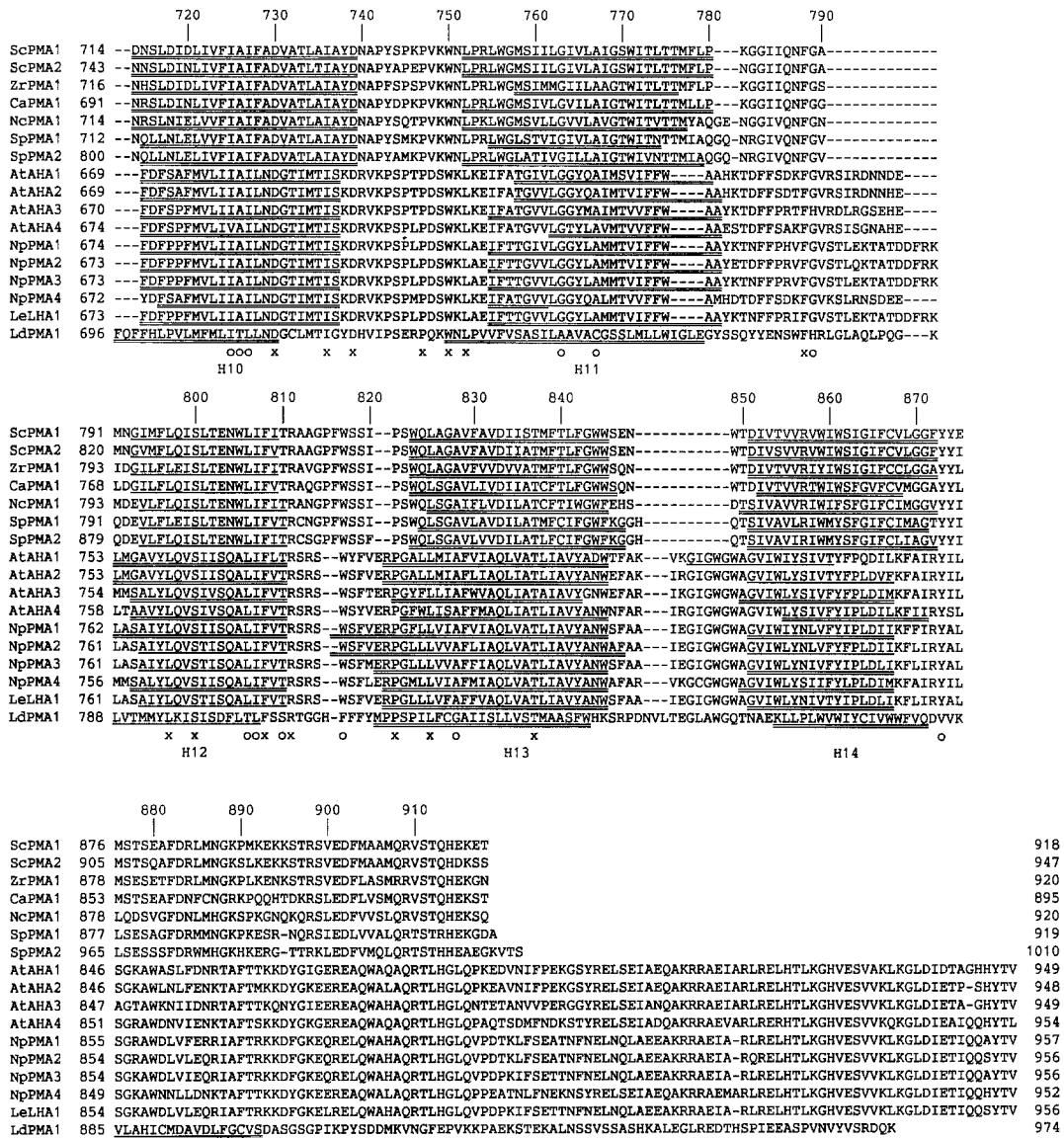
```
                          10        20        30        40        50        60        70        80
                          |         |         |         |         |         |         |         |
ScPMA1    1 ---------------------------MTDTSSSSSSSSASSVSAHQPTQEKPAKTYDDAASESSDDDDIDALIEELQSNHGVDDEDSDNDGPVAAGEARLVPEEYLQT
ScPMA2    1 MSSTEAKQYKEKPSKEYLHASDGDDPANNSAASSSSSSSTSTSASSSAAAVPRKAAAASAADDSDSDEDIDQLIDELQSNYGEGDESGEEEVRTDGVHAGQRVVPEKDLST
ZrPMA1    1 ------------------------MSDERITEKPPHQQPESEGEPVPEEEVEEETEEEVPDEQSSEDDDIDGLIDELQSQEAHEEAEEDDGPAAAGEARKIPEELLQT
CaPMA1    1 -----------------------------------MSATEPTNEKVDKIVSDDEDEDIDQLVADLQSNPGAGDEEEEEENDSSFKAVPEELLQT
NcPMA1    1 ----------------MADHSASAAPALSTNIESGKFDEKAAEAAAYQPKPKVEDDEDEDIDALIEDLESHDGHDAEEEEEEATPGGGRVVPEDMLQT
SpPMA1    1 ------------------MADNAGEYHDAEKHAPEQQAPPPQQPAHAAAPAQDDEPDDDIDALIEELFSEDVQEEQEDNDDAPAAGEAKAVPEELLQT
SpPMA2    1 -------------------------MQRNNGEGRPEGMHRISRFLHGNPFKNNASPQDDSTTRTEVYEEGGVEDSAVDYDNA
SpPMA2   58 SGNAAPRLTAAPNTHAQQANLQSGNTSITHETQSTSRGQEATTSPSLSASHEKPARPQTGEGSDNEDEDEDIDALIEDLYSQDQEEEQVEEEESPGPAGAAKVVPEELLET
AtAHA1    1 ---------------------------------------------------------------------------MSGLEDIKNETVDLEKIPIEEVFQQLK
AtAHA2    1 ---------------------------------------------------------------------------MSSLEDIKNETVDLEKIPIEEVFQQLK
AtAHA3    1 ---------------------------------------------------------------------------MASGLEDIVNENVDLEKIPIEEVFQQLK
AtAHA4    1 ---------------------------------------------------------------------------MAGNKDSSWDDIKNEGIDLEKIPIEEVLTQLR
NpPMA1    1 --------------------------------------------------------------------------MGEEKPEVLDAVLKEAVDLENIPIEEVFENLR
NpPMA2    1 --------------------------------------------------------------------------MGEKPEVLDAVLKETVDLENIPIEEVFENLR
NpPMA3    1 --------------------------------------------------------------------------MGEKPEVLDAVLKETVDLENIPIEEVFENLR
NpPMA4    1 --------------------------------------------------------------------------MAKAISLEEIKNETVDLEKIPIEEVFEQLK
LeLHA1    1 --------------------------------------------------------------------------MAEKPEVLDAVLKETVDLENIPIEEVFENLR
LdPMA1    1 -----------------------------------------------------------MSSKKYELDAAAFEDKPESHSDAEMTPQKPQRRQSVLSKAVSEHDERATGPATDPV


                   90        100       110       120       130       140       150       160       170
                   |         |         |         |         |         |         |         |
ScPMA1    83 DPSY GLT SDEVLKRRKKY GLN QMADEKESLVVKFVMFFVGPIQF VMEAAAILAAGL S-------DWVDFGVICGLLMLNAGVGFVQEFQAGSIVDELKKTLANT
ScPMA2   112 DPAY GLT SDEVARRRKKY GLN QMAEENESLIVKFLMFFVGPIQF VMEAAAILAAGL S-------DWVDVGVICALLLLNASVGFIQEFQAGSIVEELKKTLANT
ZrPMA1    85 DPSV GLS SDKVVNRRKKY GLN QMREESENLLVKFMFFIGPIQF VMEAAAVLAAGL E-------DWVDFGYICGLLLLNAFVGFIQEYQAGSIVDELKKTLANS
CaPMA1    60 DPRV GLT DDEVTKRRKRY GLN QMAEEQENLVLKFVMFFVGPIQF VMEAAAVLAAGL E-------DWVDFGVICALLLLNAFVGFIQEYQAGSIVDELKKTLANS
NcPMA1    83 DTRV GLT SEEVVERRRKY GLN QMKEEKENHFLKFLGFFVGPIQF VMEGAAVLAAGL E-------DWVDFGVICGLLLLNAVVGFVQEFQAGSIVDELKKTLALK
SpPMA1    81 DMNT GLT MSEVEERRKKY GLN QMKEELENPFLKFIMFFVGPIQF VMELAAALAAGL R-------DWVDFGVICALLLMLNAVVGFVQEYQAGSIVDELKKSLALK
SpPMA2   169 DPKY GLT ESEVEERRKKY GLN QMKEEKTNNIKKFLSFFVGPIQF VMELAAALAAGL R-------DWVDFGVICALLLLNATVGFVQEYQAGSIVDELKKTMALK
AtAHA1    28 CTRE GLT TQEGEDRIVIF GPN KLEEKKESKILKFLGFMWNPLSW VMEAAALMAIAL ANGDNRPPDWQDFVGIICLLVINSTISFIEENNAGNAAAALMAGLAPK
AtAHA2    28 CSRE GLT TQEGEDRIQIF GPN KLEEKKESKLLKFLGFMWNPLSW VMEMAAIMAIAL ANGDGRPPDWQDFVGIICLLVINSTISFIEENNAGNAAAALMAGLAPK
AtAHA3    29 CSRE GLS GAEGENRLQIF GPN KLEEKKESKLLKFLGFMWNPLSW VMEAAAIMAIAL ANGGGKPPDWQDFVGIVCLLVINSTISFIEENNAGNAAAALMAGLAPK
AtAHA4    33 CTRE GLT SDEGQTRLEIF GPN KLEEKKENKVLKFLGFMWNPLSW VMEAAIIGIAL ANGGGRPPDWQDFVGILCLLIINSTISFIEENNAGNAAAALMAGLAPK
NpPMA1    33 CTKE GLT ATAAQERLAIF GYN KLEEKKDSKLLKFLGFMWNPLSW VMEAAAIMAIAL ANGGGKPPDWQDFVGIITLLVINSTISFIEENNAGNAAAALMARLAPK
NpPMA2    32 CTKE GLS GPAAQERLAIF GYN KLEEKKSKFLKFLGFMWNPLSW VMEAAAIMAIAL ANGGGKPPDWQDFVGIITLLVINSTISFIEENNAGNAAAALMARLAPK
NpPMA3    32 CTKE GLT ATAAQERLSIF GYN KLEEKKSKFSKFLGFMWNPLSW VMEAAAIMAIAL ANGGGKPPDWQDFVGIILLLIINSTISFIEENNAGNAAAALMARLAPK
NpPMA4    31 CTRE GLS ADEGASRLQIF GPN KLEEKNESKILKFLGFMWNPLSW VMEAAAVMAIAL ANGDGCKPPDWQDFIGIICLLVINSTISFIEENNAGNAAAALMAGLAPK
LeLHA1    32 CTRE GLT ATAAQERLSIF GYN KLEEKKSKFLKFLGFMWNPLSW VMEAAAIMAIAL ANGGGKPPDWQDFVGIITLLIINSTISFIEENNAGNAAAALMARLAPK
LdPMA1    57 PPSK GLT TEEAAEELLKKY GRN ELPEKKTPSWLIYVRGLWGPMPA ALWIAIIIEFAL E-------NWPDGAILFAIQIANATIGWYETIKAGDAVAALKNSLKPT
             xx   o      x x    o      oo  o  x      ooo xo    x     ox x  o oo  x      o  o xx      x   oo
             a          b                         c                      ox x                  o
                                           H1                                   H2


                    190       200       210       220       230       240       250       260       270
                    |         |         |         |         |         |         |         |
ScPMA1   180 AV VIRDGQ LVEIPAN EVVPGDI LQLED GTIIPTD GRIVTEDCFLQ IDQSAITGESL AVDKHYGDQTFSSSTVKRGEGF MVVTATGDNTFVG RAAALVNK
ScPMA2   209 AT VIRDGQ LIEIPAN EVVPGEI LQLES GTIAPAD GRIVTEDCFLQ IDQSAITGESL AAEKHYGDEVFSSSTVKTGEAF MVVTATGDNTFVG RAAALVGQ
ZrPMA1   182 AT VIRDGS VQEAPAN EIVPGDI LKLED GTVIPAD GRLVTEECFLQ VDQSSITGESL AVDKHYGDEVFSSSTVKRGEGF MIVTATGDNTFVG RAASLVNA
CaPMA1   157 AL VVRNGQ VQEAPAN EVVPGDI LQLED GTVIPTD GRIVTDDAFLQ VDQSALTGESL AVDKHKGDQVFASSAVKRGEAF VVITATGDNTFVG RAAALVNK
NcPMA1   180 AV VLRDGT LKEIEAP EVVPGDI LQVEE GTIIPAD GRIVTDDAFLQ VDQSALTGESL AVDKHHKGDQVFASSAVKREGL VVITATGDNTFVG RAAALVNA
SpPMA1   178 AV VIREGQ VHELEAN EVVPGDI LKLDE GTIICAD GRVVTPDVHLQ VDQSAITGESL AVDKHYGDPTFASSGVKREGL MVVTATGDSTFVG RAASLVNA
SpPMA2   266 AS VLRDGR VKEIEAS EIVPGDI LHLDE GTICPAD GRLITKDCFLQ VDQSAITGESL AVDKHQNDTMYSSSTVKRGEAF MVVTATADSTFVG RAASLVGA
AtAHA1   132 TK VLRDGK WSEQEAA ILVPGDI VSIKL GDIIPAD ARLLEGDP-LK VDQSALTGESL PVTKHPGQEVFSGSTCKQGEIE AVVIATGVHTFFG KAAHLVDS
AtAHA2   132 TK VLRDGK WSEQEAA ILVPGDI VSIKL GDIIPAD ARLLEGDP-LK VDQSALTGESL PVTKHPGQEVFSGSTCKQGEIE AVVIATGVHTFFG KAAHLVDS
AtAHA3   133 TK VLRDGK WSEQEAA ILVPGDI VSIKL GDIIPAD ARLLEGDP-LK VDQSALTGESL PATKGPGEEVFSGSTCKQGEIE AVVIATGVHTFFG KAAHLVDS
AtAHA4   137 TK VLRDGK WSEQEAA ILVPGDI ISIKL GDIVPAD GRLLDGDP-LK VDQSALTGESL PVTKHPGQEVFSGSTCKQGELE AVVIATGVHTFFG KAAHLVDS
NpPMA1   137 AK VLRDGR WKEEDAA VLVPGDI ISIKL GDIIPAD ARLLEGDP-LK IDQSALTGESL PVTKGPGDGVYSGSTCKQGEIE AIVIATGVHTFFG KAAHLVDS
NpPMA2   136 AK VLRDGR WDEQDAA ILVPGDI ISIKL GDIIPAD ARLLEGDP-LK IDQSALTGESL PVTKGPGDGVYSGSTCKQGEIE AVVIATGVHTFFG KAAHLVDS
NpPMA3   136 AK VLRDGR WKEEDAA VLVPGDI ISIKL GDIIPAD ARLLEGDP-LK IDQSALTGESL PVTKGPGDGVYSGSTCKQGEIE AVVIATGVHTFFG KAAHLVDS
NpPMA4   135 TK VLRDGR WSEQEAA ILVPGDI ISVKL GDIIPAD ARLLEGDP-LK IDQSALTGESL PVTKNPGDGVYSGSTCKQGEIE AVVIATGVHTFFG KAAHLVDS
LeLHA1   136 AK VLRDGK WDEEDAS VLVPGDI ISIKL GDIIPAD ARLLEGDP-LK IDQSALTGESL PVTKGPGDGVYSGSTCKQGEIE AVVIATGVHTFFG KAAHLVDS
LdPMA1   154 AT VYRDSK WQQIDAA VLVPGDL VKLAS GSAVPAD CSINEG-V-ID VDEAALTGESL PVTMGPEHMPKMGSNVVRGEVE GTVQYTGSLTFFG KTAALLQS
             x x o  o x     xxxoo   x   x  o      o      xooo xxxxx      o        x o xx      o x  xx x    ox xo
             d      e                f                  g                         x                 h
                        H3                                                                   H4


             280       290       300       310       320       330       340       350       360       370
             |         |         |         |         |         |         |         |
ScPMA1   279 AA GGQG HFTEVLNGIGIIILLVLVIATLLLVWTACFYRTNG-IVRILRYTLGITIIG VPVGLPAVVTTT MAVG AAYLAKKQAIVQKLSAIESLA GVEI
ScPMA2   308 AS GVEG HFTEVLNGIGIIILLVLVIATLLLVWTACFYRTVG-IVSILRYTLGITIIG VPVGLPAVVTTT MAVG AAYLAKKQAIVQKLSAIESLA GVEI
ZrPMA1   281 AA GGQG HFTEVLNGIGILLVLVVITLLLIWTACFYRTVR-IVPILRYTLGITIVG VPVGLPAVVTTT MAGG AAYLAKKQAIVQKLSAIESLA GVEI
CaPMA1   256 AS AGTG HFTEVLNGIGTTLLVFVIVTLLVVWVACFYRTVR-IVPILRYTLAITIIG VPVGLPAVVTTT MAVG AAYLAKKQAIVQKLSAIESLA GVEI
NcPMA1   279 AS GGSG HFTEVLNGIGTILLILVIFTLLIVWVSSFYRSN-PIVQILEFTLAITIIG VPVGLPAVVTTT MAVG AAYLAEKQAIVQKLSAIESLA GVEV
SpPMA1   277 AA GGTG HFTEVLNGIGTILLVLVLLTLFCIYTAAFYRSVR-LARLLEYTLAITIIG VPVGLPAVVTTT MAVG AAYLAKKKAIVQKLSAIESLA GVEI
SpPMA2   365 AG QSQG HFTEVLNGIGTILLVLVILTLLCIYTAAFYRSVR-LAALLEYTLAITIIG VPVGLPAVVTTT MAVG AAYLAKKKAIVQKLSAIESLA GVEI
AtAHA1   230 TN -QVG HFQKVLTSIGNFCICSIAIGIAIEIVVMVPIQHRKYRDGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHRLSQQGAITKRMTAIEEMA GMDV
AtAHA2   230 TN -QVG HFQKVLTAIGNFCICSIAIGMVIEIIVMYPIQRRKYRDGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHKLSQQGAITKRMTAIEEMA GMDV
AtAHA3   231 TN -QVG HFQKVLTAIGNFCICSIAVGIAIEIVVMYPIQRRHYRDGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHKLSQQGAITKRMTAIEEMA GMDV
AtAHA4   235 TN -QEG HFQKVLTAIGNFCICSIAIGMLIEIVVMYPIQRRKYRDGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHRLSQQGAITKRMTAIEEMA GMDV
NpPMA1   235 TN -QVG HFQKVLTAIGNFCICSIAVGMIIEIIVMYPIQHRAYRPGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHRLAQQGAITKRMTAIEEMA GMDV
NpPMA2   234 TN -QVG HFQKVLTAIGNFCICSIAVGMIIEIIVMYPIQRRKYRPGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHRLAQQGAITKRMTAIEEMA GMDV
NpPMA3   234 TN -QVG HFQKVLTAIGNFCICSIAIGMLIEIIVMYPIQRRKYRDGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHRLSQQGAITKRMTAIEEMA GMDV
NpPMA4   233 TN -NVG HFQKVLTAIGNFCICSIAIGMLVEIIVMYPIQHRKYRPGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHRLAQQGAITKRMTAIEEMA GMDV
LeLHA1   234 TN -QVG HFQKVLTAIGNFCICSIAVGMIIEIIVMYPIQHRKYRPGIDNLLVLLIGG IPIAMPTVLSVT MAIG SHRLSQQGAITKRMTAIEEMA GMDV
LdPMA1   251 VE SDLG NIHVILRRVMFSLCAISFMLCMCCFIYLLARFYETFRHALQFAVVVLVVS IPIALEIVVTTT LAVG SKHLSKHKIIVTKLSAIEMMS GVNM
             x    oo  ox  oo                          o    o  o x  o  x      ox x  x   ox  xxx  o x
             i                                               j                k
                        H5                                          H6                    H7
```

Fig. 2.

```
              380       390        400       410       420      430       440       450       460
              |         |          |         |         |        |         |         |         |
ScPMA1  375 LCSDKTGTLTKNKLSL HEPYT---VEGVSPDDLMLTACLAASRKKKGL DAIDKAFL KSLKQYPKAKDALTKYKVLEFHPFDPVSKKVT-AVVESP
ScPMA2  404 LCSDKTGTLTKNKLSL HEPYT---VEGVSPDDLMLTACLAASRKKKGL DAIDKAFL KSLIEYPKAKDALTKYKVLEFHPFDPVSKKVT-AVVESP
ZrPMA1  377 LCSDKTGTLTKNKLSL HEPYT---VEGVSSDDLMLTACLAASRKKKGL DAIDKAFL KSLAQYPKAKGALTKYKVLEFHPFDPVSKKVT-AVVESP
CaPMA1  352 LCSDKTGTLTKNKLSL HEPYT---VEGVEPDDLMLTACLAASRKKKGL DAIDKAFL KSLINYPRAKAALPKYKVIEFQPFDPVSKKVT-AIVESP
NcPMA1  375 LCSDKTGTLTKNKLSL HDPYT---VAGVDPEDLMLTACLAASRKKKGI DAIDKAFL KSLKYYPRAKSVLSKYKVLQFHPFDPVSKKVV-AVVESP
SpPMA1  373 LCSDKTGTLTKRNLSL GEPFT---VSGVSGDDLVLTACLAASRKKKGL DAIDKAFL KALKNYPGPRSMLTKYKVIEFQPFDPVSKKVT-AYVQAP
SpPMA2  461 LCSDKTGTLTKRNLSL GEPYC---VEGVSPDDLMLTACLASSRKKKGL DAIDKAFL KALRNYPKAKDQLSKYKVLDFHPFDPVSKKIT-AYVEAP
AtAHA1  326 LCSDKTGTLTLNKLSV DKNLVEVFCKGVEKDQVLLFAAMAS--RVENQ DAIDAAMV GMLAD---PKEARAGIREVHFLPFNPVDKRTALTYIDS-
AtAHA2  326 LCSDKTGTLTLNKLSV DKNLVEVFCKGVEKDQVLLFAAMAS--RVENQ DAIDAAMV GMLAD---PKEARAGIREVHFLPFNPVDKRTALTYIDG-
AtAHA3  327 LCSDKTGTLTLNKLSV DKNLIEVYCKGVEKDEVLLFAARAS--RVENQ DAIDAAMV GMLAD---PKEARAGIREIHFLPFNPVDKRTALTFIDS-
AtAHA4  331 LCSDKTGTLTLNKLTV DKSMVEVFVKDLDKDQLLVNAARAS--RVENQ DAIDACIV GMLGD---PREAREGITEVHFFPFNPVDKRTAITYIDA-
NpPMA1  331 LCSDKTGTLTLNKLTV DKNLIEVFAKGVDADMVVLMAARAS--RTENQ DAIDAAIV GMLAD---PKEARAGIREIHFLPFNPTDKRTALTYLDG-
NpPMA2  330 LCSDKTGTLTLNKLTV DKNLVEVFAKGVDADTVVLMAARAS--RTENQ DAIDTAIV GMLSD---PKEARAGIREIHFLPFNPTDKRTALTYLDG-
NpPMA3  330 LCSDKTGTLTLNKLTV DKYLIEVFARGVDADTVVLMAARAS--RTENQ DAIDAAIV GMLAD---PKEARAGIREIHFLPFNPTDKRTALTYLDG-
NpPMA4  329 LCSDKTGTLTLNKLSV DRNLVEVFAKGVDKEYVLLLAARAS--RVENQ DAIDACMV GMLAD---PKEARAGIREVHFLPFNPVDKRTALTYIDN-
LeLHA1  330 LCSDKTGTLTLNKLTV DKALIEVFAKGIDADTVVLMAARAS--RIENQ DAIDTAIV GMLAD---PKEARAGIREIHFLPFNPTDKRTALTYLDG-
LdPMA1  348 LCSDRTGTLTLNKMEI QEQCFT-FEEGNDLKSTLVLAALAAKWREPPR DALDTMVL G--AA---DLDECDNYQQLNFVPFDPTTKRTAATLVDRR
            XXXXOXXXXX XOO    O  O   X  X     XXOX    O      X XX X  X
            1                                         m

              470        480       490       500       510       520       530       540
              |          |         |         |         |         |         |         |
ScPMA1  466 EGERIVCV KGAP LFVLKTVEEDHPIPEDVHENYENKVAELASRGFRALGVARKR-------GEGHWEILG VMPCMDPPR DDTAQTVS
ScPMA2  495 EGERIVCV KGAP LFVLKTVEEDHPIPEDVHENYENKVAELASRGFRALGVARKR-------GEGHWEILG VMPCMDPPR DDTAQTIN
ZrPMA1  468 EGERIICV KGAP LFVLKTVEEDHPIPEDVHENYENKVAELASRGFRALGVARKR-------GEGHWEILG VMPCMDPPR DDTAATVN
CaPMA1  443 EGERIICV KGAP LFVLKTVEDDHPIPEDVHENYQNTVAEFASRGFRSLGVARKR-------GEGHWEILG IMPCMDPPR DDTAATVN
NcPMA1  466 QGERITCV KGAP LFVLKTVEEDHPIPEEVDQAYKNKVAEFATRGFRSLGVARKR-------GEGSWEILG IMPCMDPPR HDTYKTVC
SpPMA1  464 DGTRITCV KGAP LWVLKTVEEDHPIPEDVLSAYKDKVGDLASRGYRSLGVARKI-------EGQHWEIMG IMPCSDPPR HDTARTIS
SpPMA2  552 DGQRITCV KGAP LWVFKTVQDDHEVPEAITDAYREQVNDMASRGFRSLGVARKA-------DGKQWEILG IMPCSDPPR HDTARTIH
AtAHA1  415 DGNWHRVS KGAP EQILDLANAR----PDLRKKVLSCIDKYAERGLRSLAVARQVVPEKTKESPGGPWEFVG LLPLFDPPR HDSAETIR
AtAHA2  415 SGNWHRVS KGAP EQILELAKAS----NDLSKKVLSIIDKYAERGLRSLAVARQVVPEKTKESPGAPWEFVG LLPLFDPPR HDSAETIR
AtAHA3  416 NGNWHRVS KGAP EQILDLCNAR----ADLRKRVHSTIDKYAERGLRSLAVSRQTVPEKTKESSGSPWEFVG VLPLFDPPR HDSAETIR
AtAHA4  420 NGNWHRVS KGAP EQIIELCNLR----EDASKRAHDIIDKFADRGLRSLAVGRQTVSEKDKNSPGEPWQFLG LLPLFDPPR HDSAETIR
NpPMA1  420 EGKMHRVS KGAP EQILNLAHNK----SDIERRVHAVIDKFAERGLRSLGVAYQEVPEGRKESAGGPWQFIG LLPLFDPPR HDSAETIR
NpPMA2  419 EGKMHRVS KGAP EQILNLAHNK----SDIERRVHSVIDKFAERGLRSLGVAYQEVPEGRKESTGGPWQFIG LLPLFDPPR HDSAETIR
NpPMA3  419 EGKMHRVS KGAP EQILHLAHNK----SDIERRVHAVIDKFAERGLRSLAVAYQEVPEGRKESAGGPWQFIA LLPLFDPPR HDSAETIR
NpPMA4  418 NNNWHRAS KGAP EQILDLCNAK----EDVRRKVHSMMDKYAERGLRSLAVARRTVPEKSKESPGGRWEFVG LLPLFDPPR HDSAETIR
LeLHA1  419 EGKMHRVS KGAP EQILNLAHNK----SDIERRVHTVIDKFAERGLRSLGVAYQEVPEGRKESAGGPWQFIA LLPLFDPPR HDSAETIR
LdPMA1  437 SGEKFDVT KGAP HVILQMVYNQ----DEINDEVVDIIDSLAARGVRCLSVA--------KTDQQGRWHMAG ILTFLDPPR PDTKDTIR
            XXXX        X XX X X X             X       O  XXXX  X   X
            n                                         o

             550       560       570       580       590       600       610       620
             |         |         |         |         |         |         |         |
ScPMA1  546 EA RHLGLRVKMLTGDAVGIAK ETCRQLGLGTNIYNAERLGLGLGGGD-----MPGSELADFVENADGFAEVF PQHKYRVVEI LQNR
ScPMA2  575 EA RNLGLRIKMLTGDAVGIAK ETCRQLGLGTNIYNAERLGLGGGD-----MPGSELADFVENADGFAEVF PQHKYRVVEI LQDR
ZrPMA1  548 EA KRLGLSVKMLTGDAVGIAK ETCRQLGLGTNIYDAERLGLGLGGGS-----MPGSEMYDFVENADGFAEVF PQHKFAVVDI LQQR
CaPMA1  523 EA RRLGLRVKMLTGDAVGIAK ETCRQLGLGTNIYDADRLGLSGGGD-----MAGSEIADFVENADGFAEGF PTNKYNAVEI LQSR
NcPMA1  546 EA KTLGLSIKMLTGDAVGIAR ETSRQLGLGTNIVNAERLGLGGGD-----MPGSEVYDFVEAADGFAEVF PQHKYNVVEI LQQR
SpPMA1  544 EA KRLGLRVKMLTGDAVDIAK ETARQLGMGTNIYNAERLGLTGGGN-----MPGSEVYDFVEAADGFGEVF PQHKYAVVDI LQQR
SpPMA2  632 EA IGLGLRIKMLTGDAVGIAK ETARQLGMGTNVYNAERLGLSGGGD-----MPGSEVNDFVEAADGFAEVF PQHKYAVVDI LQQR
AtAHA1  499 RA LNLGVNVKMITGDQLAIGK ETGRRLGMGTNMYPSAAL-LGTDKDSNIASIPVE---ELIEKADGFAGVF PEHKYEIVKK LQER
AtAHA2  499 RA LNLGVNVKMITGDQLAIGK ETGRRLGMGTNMYPSSAL-LGTHKDANLASIPVE---ELIEKADGFAGVF PEHKYEIVKK LQER
AtAHA3  500 RA LDLGVNVKMITGDQLAIAK ETGRRLGMGSNMYPSSSL-LGKHKDEAMAHIPVE---ELIEKADGFAGVF PEHKYEIVKK LQER
AtAHA4  504 RA LDLGVNVKMITGDQLAIGK ETGRRLGMGTNMYPSSAL-LGQDKDESIASLPVD---ELIEKADGFAGVF PEHKYEIEKR LQEM
NpPMA1  504 RA LNLGVNVKMVTGDQLAIGK ETGRRLGMGTNMYPSSAL-LGQTKDESISALPID---ELIEKADGFAGVF PEHKYEIVKR LQAR
NpPMA2  503 RA LNLGVNVKMITGDQLAIGK ETGRRLGMGTNMYPSSAL-LGQTKDESISALPID---ELIEKADGFAGVF PEHKYEIVKR LQAR
NpPMA3  503 RA LNLGVNVKMITGDQLAIGK ETGRRLGMGTNMYPSSAL-LGQTKDESISALPVD---ELIEKADGFAGVF PEHKYEIVKR LQAR
NpPMA4  502 RA LNLGVNVKMITGDQLAIAK ETGRRLGMGTNMYPSASL-LGQDKDSAIASLPIE---ELIEKADGFAGVF PEHKYEIVKK LQER
LeLHA1  503 RA LNLGVNVKMITGDQLAIGK ETGRRLGMGTNMYPSSAL-LGQTKDESIAALPID---ELIEKADGFAGVF PEHKYEIVKR LQAR
LdPMA1  513 RS KEYGVDVKMITGDHLLIAK EMCRMLDLDPNILTADKL-PQIKDANDLPEDLGEKYGDMMLSVGGFAQVF PEHKFMIVET LRQR
            O  OX  XX XXX   X O XO X XO O X O   X O          O  O OOXXO OX X OX  O  XO O
            p                           q

             630       640       650       660       670       680       690       700       710
             |         |         |         |         |         |         |         |         |
ScPMA1  626 GY LVAMTGDGVNDAPSLKKADTGIAVEGATDAARSAADIVF LAPGLS AIIDALKTSRQISH RMYSYVVYRIALSLHLEIFLGLWIAIL----------
ScPMA2  655 GY LVAMTGDGVNDAPSLKKADTGIAVEGATDAARSAADIVF LAPGLS AIIDALKTSRQIFH RMYSYVVYRIALSLHLEIFLGLWIAIL----------
ZrPMA1  628 GY LVAMTGDGVNDAPSLKKADTGIAVEGATDAARSAADIVF LAPGLS AIIDALKTSRQIFH RMYAYVVYRIALSLHLEIFLGLWIAIL----------
CaPMA1  603 GY LVAMTGDGVNDAPSLKKADTGIAVEGATDAARSAADIVF LAPGLS AIIDALKTSRQIFH RMYSYVVYRIALSLHLELFLGLWIAIL----------
NcPMA1  626 GY LVAMTGDGVNDAPSLKKADTGIAVEGSSDAARSAADIVF LAPGLG AIIDALKTSRQIFH RMYAYVVYRIALSIHLEIFLGLWIAIL----------
SpPMA1  624 GY LVAMTGDGVNDAPSLKKADTGIAVEGATDAARSAADIVF LAPGLS AIIDALKTSRQIFH RMYSYVVYRIALSLHLEIFLGLWLIIR----------
SpPMA2  712 GY LVAMTGDGVNDAPSLKKADAGIAVEGASDAARSAADIVF LAPGLS AIIDALKTSRQIFH RMYAYVVYRIALSLHLEIFLGLWLIIR----------
AtAHA1  580 KH IVGMTGDGVNDAPALKKADIGIAVADATDAARGASDIVL TEPGLS VIISAVLTSRAIFQ RMKNYTIYAVSITIRIVFGFMLIALIWE---------
AtAHA2  580 KH IVGMTGDGVNDAPALKKADIGIAVADATDAARGASDIVL TEPGLS VIISAVLTSRAIFQ RMKNYTIYAVSITIRIVFGFMLIALIWE---------
AtAHA3  581 KH ICGMTGDGVNDAPALKKADIGIAVADATDAARGASDIVL TEPGLS VIISAVLTSRAIFQ RMKNYTIYAVSITIRIVGFMLIALIWK---------
AtAHA4  585 KH ICGMTGDGVNDAPALKKADIGIAAADATDAARSASDIVL TEPGLS VIVSAVLTSRAIFQ RMKNYTIYAVSITIRIVMGFMLLALIWK---------
NpPMA1  585 KH ICGMTGDGVNDAPALKKADIGIAVDDATDAARSASDIVL TEPGLS VIISAVLTSRAIFQ RMKNYTIYAVSITIRIVLGFMLLALIWK---------
NpPMA2  584 KH ICGMTGDGVNDAPALKKADIGIAVDDATDAARSASDIVL TEPGLS VIISAVLTSRAIFQ RMKNYTIYAVSITIRIVLGFMLLALIWQ---------
NpPMA3  584 KH ICGMTGDGVNDAPALKKADIGIAVDDATDAARSASDIVL TEPGLS VIISAVLTSRAIFQ RMKNYTIYAVSITIRIVLGFMLLALIWK---------
NpPMA4  583 KH IVGMTGDGVNDAPALKKADIGIAVADATDAARGASDIVL TEPGLS VIISAVLTSRAIFQ RMKNYTIYAVSITIRIVFGFMFIALIWK---------
LeLHA1  584 KH ICGMTGDGVNDAPALKKADIGIAVDDATDAARSASDIVL TEPGLS VIISAVLTSRAIFQ RMKNYTIYAVSITIRIVLGFMLLALIWK---------
LdPMA1  597 GY TCAMTGDGVNDAPALKRADVGIAVHGATDAARAAAMVL TEPGLS VVVEAMLVSREVFQ RMLSFLTYRISATLQLVCFFFIACFSLTFKAYGSVDPH
            XXXXXXXXXX XX XX XXXO  OOXXXX X X X   XXX   O  X  OXX OO  XX  O  X
            r                              H8    s                        H9
```

Fig. 2. Continued.

```
              720       730       740       750       760       770       780       790
               |         |         |         |         |         |         |         |
ScPMA1  714 --DNSLDIDLIVFIAIFADVATLAIAYDNAPYSPKPVKWNLPRLWGMSIILGIVLAIGSWITLTTMFLP---KGGIIQNFGA------------
ScPMA2  743 --NNSLDINLIVFIAIFADVATLTIAYDNAPYAPEPVKWNLPRLWGMSIILGIVLAIGSWITLTTMFLP---NGGIIQNFGA------------
ZrPMA1  716 --NHSLDIDLIVFIAIFADVATLAIAYDNAPFSPSPVKWNLPRLWGMSIMMGIILAAGTWITLTTMFLP---KGGIIQNFGS------------
CaPMA1  691 --NRSLDINLIVFIAIFADVATLAIAYDNAPYDPKPVKWNLPRLWGMSIVLGVILAIGTWITLTTMLLP---KGGIIQNFGG------------
NcPMA1  714 --NRSLNIELVVFIAIFADVATLAIAYDNAPYSQTPVKWNLPKLWGMSVLLGVVLAVGTWITVTTMYAQGE-NGGIVQNFGN------------
SpPMA1  712 --NQLLNLELVVFIAIFADVATLAIAYDNAPYSMKPVKWNLPRLWGLSTVIGIVLAIGTWITNTTMIAQGQ-NRGIVQNFGV------------
SpPMA2  800 --NQLLNLELIVFIAIFADVATLAIAYDNAPYAMKPVKWNLPRLWGLATIVGILLAIGTWIVNTTMIAQGQ-NRGIVQNFGV------------
AtAHA1  669 ---FDFSAFMVLIIAILNDGTIMTISKDRVKPSPTPDSWKLKEIFATGIVLGGYQAIMSVIFFW----AAHKTDFFSDKFGVRSIRDNNDE----
AtAHA2  669 ---FDFSAFMVLIIAILNDGTIMTISKDRVKPSPTPDSWKLKEIFATGVVLGGYQAIMTVIFFW----AAHKTDFFSDTFGVRSIRDNNHE----
AtAHA3  670 ---FDFSPFMVLIIAILNDGTIMTISKDRVKPSPTPDSWKLKEIFATGVVLGGYMAIMTVVFFW----AAYKTDFFPRTFHVRDLRGSEHE----
AtAHA4  674 ---FDFSPFMVLIVAILNDGTIMTISKDRVKPSPLPDSWKLKEIFATGVVLGTYLAVMTVVFFW----AAESTDFFSAKFGVRSISGNAHE----
NpPMA1  674 ---FDFPPFMVLIIAILNDGTIMTISKDRVKPSPLPDSWKLAEIFTTGIVLGGYLAMMTVIFFW----AAYKTNFFPHVFGVSTLEKTATDDFRK
NpPMA2  673 ---FDFPPFMVLIIAILNDGTIMTISKDRVKPSPLPDSWKLAEIFTTGVVLGGYLAMMTVIFFW----AAYETDFFPRVFGVSTLQKTATDDFRK
NpPMA3  673 ---FDFPPFMVLIIAILNDGTIMTISKDRVKPSPLPDSWKLAEIFTTGVVLGGYLAMMTVIFFW----AAYKTNFFPRVFGVSTLEKTATDDFRK
NpPMA4  672 ---YDFSAFMVLIIAILNDGTIMTISKDRVKPSPMPDSWKLKEIFATGVVLGGYQALMTVVFFW----AMHDTDFFSDKFGVKSLRNSDEE----
LeLHA1  673 ---FDFPPFMVLIIAILNDGTIMTISKDRVKPSPLPDSWKLAEIFTTGVVLGGYLAMMTVIFFW----AAYKTNFFPRIFGVSTLEKTATDDFRK
LdPMA1  696 FQFFHLPVLMFMLITLLNDGCLMTIGYDHVIPSERPQKWNLPVVFVSASILAAVACGSSLMLLWIGLEGYSSQYYENSWFHRLGLAQLPQG---K
            OOO X     X X       X X X    O O                                           XO
            H10                              H11


              800       810       820       830       840             850       860       870
               |         |         |         |         |               |         |         |
ScPMA1  791 MNGIMFLQISLTENWLIFITRAAGPFWSSI--PSWQLAGAVFAVDIISTMFTLFGWWSEN----------WTDIVTVVRVWIWSIGIFCVLGGFYYE
ScPMA2  820 MNGVMFLQISLTENWLIFVTRAAGPFWSSI--PSWQLAGAVFAVDIIATMFTLFGWWSEN----------WTDIVSVVRVWIWSIGIFCVLGGFYYI
ZrPMA1  793 IDGILFLEISLTENWLIFITRAVGPFWSSI--PSWQLAGAVFVVDVVATMFTLFGWWSQN----------WTDIVTVVRIYIWSIGIFCCLGGAYYL
CaPMA1  768 LDGILFLQISLTENWLIFVTRAQGPFWSSI--PSWQLSGAVLIVDIIATCFTLFGWWSQN----------WTDIVTVVRTWIWSFGVFCVMGGAYYL
NcPMA1  793 MDEVLFLQISLTENWLIFITRANGPFWSSI--PSWQLSGAIFLVDILATCFTIWGWFEHS----------DTSIVAVVRIWIFSFGIFCIMGGVYYI
SpPMA1  791 QDEVLFLEISLTENWLIFVTRCNGPFWSSI--PSWQLSGAVLAVDIIATMFCIFGWFKGGH----------QTSIVAVLRIWMYSFGIFCIMAGTYYI
SpPMA2  879 QDEVLFLQISLTENWLIFITRCSGPFWSSF--PSWQLSGAVLVVDIILATLFCIFGWFKGGH----------QTSIVAVIRIWMYSFGIFCLIAGVYYI
AtAHA1  753 LMGAVYLQVSIISQALIFLTRSRS--WYFVERPGALLMIAFVIAQLVATLIAVYADWTFAK---VKGIGWGWAGVIWIYSIVTYFPQDILKFAIRYIL
AtAHA2  753 LMGAVYLQVSIISQALIFVTRSRS--WSFVERPGALLMIAFLIAQLIATLIAVYANWEFAK---IRGIGWGWAGVIWLYSIVTYFPLDVFKFAIRYIL
AtAHA3  754 MMSALYLQVSIVSQALIFVTRSRS--WSFTERPGYFLLIAFWVAQLIATAIAVYGNWEFAR---IKGIGWGWAGVIWLYSIVFYFPLDIMKFAIRYIL
AtAHA4  758 LTAAVYLQVSIVSQALIFVTRSRS--WSYVERPGFWLISAFFMAQLIATLIAVYANWNFAR---IKGIGWGWAGVIWLYSIVFYIPLDILKFIIRYSL
NpPMA1  762 LASAIYLQVSIISQALIFVTRSRS--WSFVERPGFLLVIAFVIAQLVATLIAVYANWSFAA---IEGIGWGWAGVIWIYNLVFYIPLDIIKFFIRYAL
NpPMA2  761 LASAIYLQVSTISQALIFVTRSRS--WSFVERPGLLLVVAFLIAQLVATLIAVYANWAFAA---IEGIGWGWAGVIWLYNLVFYFPLDIIKFLIRYAL
NpPMA3  761 LASAIYLQVSTISQALIFVTRSRS--WSFMERPGLLLVVAFFIAQLVATLIAVYANWSFAA---IEGIGWGWAGVIWLYNIVFYIPLDLIKFLIRYAL
NpPMA4  756 MMSALYLQVSIISQALIFVTRSRS--WSFLERPGMLLVIAFMIAQLVATLIAVYANWAFAR---VKGCGWGWAGVIWLYSIIFYLPLDIMKFAIRYIL
LeLHA1  761 LASAIYLQVSTISQALIFVTRSRS--WSFVERPGLLLVFAFFVAQLVATLIAVYANWSFAA---IEGIGWGWAGVIWLYNIVTYIPLDLIKFLIRYAL
LdPMA1  788 LVTMMYLKISISDFLTLFSSRTGGH-FFFYMPPSPILFCGAIISLLVSTMAASFWHKSRPDNVLTEGLAWGQTNAEKLLPLWVWIYCIVWWFVQDVVK
            X X    OOX OX    O    X  X  O         X                                          O
            H12                          H13                                          H14


              880       890       900       910
               |         |         |         |
ScPMA1  876 MSTSEAFDRLMNGKPMKEKKSTRSVEDFMAAMQRVSTQHEKET                                              918
ScPMA2  905 MSTSQAFDRLMNGKSLKEKKSTRSVEDFMAAMQRVSTQHDKSS                                              947
ZrPMA1  878 MSESETFDRLMNGKPLKENKSTRSVEDFLASMRRVSTQHEKGN                                              920
CaPMA1  853 MSTSEAFDNFCNGRKPQQHTDKRSLEDFLVSMQRVSTQHEKST                                              895
NcPMA1  878 LQDSVGFDNLMHGKSPKGNQKQRSLEDFVVSLQRVSTQHEKSQ                                              920
SpPMA1  877 LSESAGFDRMMNGKPKESR-NQRSIEDLVVALQRTSTRHEKGDA                                             919
SpPMA2  965 LSESSSFDRWMHGKHKERG-TTRKLEDFVMQLQRTSTHHEAEGKVTS                                         1010
AtAHA1  846 SGKAWASLFDNRTAFTTKKDYGIGEREAQWAQAQRTLHGLQPKEDVNIFPEKGSYRELSEIAEQAKRRAEIARLRELHTLKGHVESVAKLKGLDIDTAGHHYTV 949
AtAHA2  846 SGKAWLNLFENKTAFTMKKDYGKEEREAQWALAQRTLHGLQPKEAVNIFPEKGSYRELSEIAEQAKRRAEIARLRELHTLKGHVESVVKLKGLDIETP-SHYTV 948
AtAHA3  847 AGTAWKNIIDNRTAFTTKQNYGIEEREAQWAHAQRTLHGLQNTETANVVPERGGYRELSEIANQAKRRAEIARLRELHTLKGHVESVVKLKGLDIETA-GHYTV 949
AtAHA4  851 SGRAWDNVIENKTAFTSKKDYGKGEREAQWAQAQRTLHGLQPAQTSDMFNDKSTYRELSEIADQAKRRAEVARLRERHTLKGHVESVVKQKGLDIEAIQQHYTL 954
NpPMA1  855 SGRAWDLVFERRIAFTRKKDFGKEQRELQWAHAQRTLHGLQVPDTKLFSEATNFNELNQLAEEAKRRAEIA-RLRELHTLKGHVESVVKLKGLDIETIQQAYTV 957
NpPMA2  854 SGRAWDLVLEQRIAFTRKKDFGKEQRELQWAHAQRTLHGLQVPDTKLFSEATNFNELNQLAEEAKRRAEIA-RQRELHTLKGHVESVVKLKGLDIETIQQSYTV 956
NpPMA3  854 SGRAWDLVIEQRIAFTRKKDFGKEQRELQWAHAQRTLHGLQVPDPKIFSETTNFNELNQLAEEAKRRAEIA-RLRELHTLKGHVESVVKLKGLDIETIQQAYTV 956
NpPMA4  849 SGKAWNNLLDNKTAFTTKKDYGKEEREAQWALAQRTLHGLQPPEATNLFNEKNSYRELSEIAEQAKRRAEMARLRELHTLKGHVESVVKLKGLDIETIQQHYTV 952
LeLHA1  854 SGKAWDLVLEQRIAFTRKKDFGKELRELQWAHAQRTLHGLQVPDPKIFSETTNFNELNQLAEEAKRRAEIA-RLRELHTLKGHVESVVKLKGLDIETIQQSYTV 956
LdPMA1  885 VLAHICMDAVDLFGCVSDASGSGPIKPYSDDMKVNGFEPVKKPAEKSTEKALNSSVSSASHKALEGLREDTHSPIEEASPVNVYVSRDQK            974
```

Fig. 2. Continued. Alignment of putative P-type H^+-ATPase sequences. Amino acid sequences are from *Saccharomyces cerevisiae* (ScPMA1, ScPMA2), *Zygosaccharomyces rouxii* (ZrPMA1), *Candida albicans* (CaPMA1), *Neurospora crassa* (NcPMA1), *Schizosaccharomyces pombe* (SpPMA1, SpPMA2), *Arabidopsis thaliana* (AtAHA1-AtAHA4), *Nicotiana plumbaginifolia* (NpPMA1-NpPMA2), *Lycopersicom esculentum* (LeLHA1), *Leishmania donovani* (LdPMA1). See text for references. The alignment was initially performed by using the program CLUSTAL (Higgins and Sharp, (1988)) and then refined manually. The numbers on top of each block are positions in ScPMA1. Underlined sequences indicate predicted hydrophobic regions by the method of Klein et al. (1985): membrane integral sequences are underlined twice, and possibly membrane integral sequences are underlined once; these regions are labeled H1-H14. Conserved regions are derived from the sequence comparison in Fig. 1 are boxed and labeled from a to s. An x indicates total identity in all aligned sequences and a (O) indicates identity in all except one sequence. The overall identity is 13.3%, and the similarity is 24%.

plant "deletions" correspond to fungal insertions. These "insertions" and the N and C terminus represent a total of 22 segments which are not conserved among yeast and plants and which separate well-conserved domains. The evolutionary and functional

significances of these punctuations in the H^+-ATPase genes have still to be clarified.

The *Leishmania donovani* sequence (Ld PMA1) is obviously distinct from the plant and the yeast H^+-ATPase sequences. This raises the questions

whether or not it corresponds to a H^+-ATPase as initially proposed by Meade et al. (1987) on the basis of sequence homologies.

A total of 241 residues identical in at least 16 out of 17 sequences (to allow for sequencing or alignment errors) is marked by an asterisk or a circle in Fig. 2. As expected, many of these are located in the boxed regions a to s which are conserved in ATPases of different cation specificity. However, 118 residues were found identical in at least 16 presumed H^+-ATPases but not located in the conserved region a to s. This approach is prone to errors since the alignment depends partially on personal judgment. Nevertheless, it is remarkable that more than 100 identical residues can be identified which seem specifically conserved in the H^+-ATPase sequences determined from different species: five fungi, three plants, and one ciliate.

PREDICTIONS OF HYDROPHOBIC REGIONS

The number and location of the membrane-spanning segments of the H^+-ATPases are not known. Nevertheless, different topographies have been proposed for the yeast H^+-ATPase comprising either 5, 8, 9, or 10 membrane spans (Hager et al., 1986; Serrano et al., 1986; Serrano, 1988, 1989; Davis and Hammes, 1989; Goffeau and Green, 1990). It is clearly established that the NH_2 and COOH termini are both located in the cytoplasm (Mandala and Slayman, 1989; Hennessey and Scarborough, 1990). This excludes the model of Serrano (1989) and Davis and Hammes (1989) who concluded on the existence of an uneven number of membrane spans. A biochemical investigation carried out recently (Subrahmanyeswara Rao et al., 1991) concluded on the existence of up to 12 membrane spans. We are thus presently left with three possibilities: either 8, 10, or 12 transmembrane segments.

These discrepancies reflect the limitations of the different prediction methods for transmembrane segments. A recent study (Fasman and Gilbert, 1990) compared nine different predictive algorithms applied to the sequence of the reaction center from Rhodopseudomonas viridis, the molecular structure of which is known at 3 Å resolution (Deisenhofer et al., 1985). It was concluded that the method with least ambiguity and highest accuracy in determining the membrane sequence is that of Klein et al. (1985). However, the method is not perfect: it misses the membrane span of

the H-polypeptide, and three amphiphilic regions of the L and M subunits with periphery/integral odds between 1.0 to 13.4 are predicted to be "possibly integral" but do in fact span the membrane as determined by X-ray crystallography.

We have applied the Klein et al. (1985) method to the detection of hydrophobic segments of the 17 putative H^+-ATPase. This method determines 12–14 hydrophobic or amphiphilic segments in most sequences. For each gene product only 10 segments are generally predicted to be "integral" membrane segments, the others being predicted to be "possibly integral." These "integral" plus "possibly integral" segments are labelled H1 to H14 in Fig. 2. The segments H3 and H4 are considered "possibly integral" in plants and are predicted to be "peripheral" in yeasts. In contrast, H7 is predicted "possibly integral" in yeast and "peripheral" in plants. The segment H12 is predicted "possibly integral" in yeast and "integral" in plants. Only nine hydrophobic segments H1, H2, H5, H6, H8, H9, H10, H11, and H13 are predicted to be "integral" in all 17 sequences. The segment H14 might be put in the same category with 14 predictions as "integral" and three predictions as "possibly integral." These three later segments have peripheral/integral odds of 1.3, 4.0, and 21.6. Such odds are in the same order as those (from 2.5 to 13.4) given for the three amphiphilic membrane-spanning segments of the L and M subunits of the photosynthetic reaction center (Fasman and Gilbert, 1990). On this basis, we conclude on a consensus of 10 predicted membrane spans in all H^+-ATPase as we had already proposed before (Goffeau and Green, 1990; Van Dyck et al., 1990). It is, however, recognized that only straightforward biochemistry and structural data will be able to identify the real membrane topography.

SOME SUGGESTIONS FOR SITE-DIRECTED MUTAGENESIS

Site-directed mutagenesis of H^+-ATPase turns out to be a more difficult approach than first believed. Even though ingenious procedures have been developed to put the wild type and/or mutant ATPase genes under the control of inducible promoters (Portillo and Serrano, 1988; Nakamoto et al., 1991), it has not yet been possible to produce sufficient amounts of lethal mutant ATPase enzyme to carry out extensive biochemical studies. This is due to several reasons. The S. cerevisiae PMA1 gene is essential

Table I. Some Characteristic Residues Specifically Conserved in H$^+$-ATPases from Yeasts and Plants[a]

Domain	Boundaries (residue number in S. cerevisiae)	E or D	H	K or R	N or Q	P	W, F, or Y	S or T	C or M
Hydrophobic 1 (H$_1$ and H$_2$)	107 to 161	D107, D143, E162			N154	P123	F119		
Hydrophilic I	162 to 283	E190, E255		R215, K252				S249	
Hydrophobic 2 (H$_5$ and H$_6$)	284 to 345		H285				F286		
Hydrophilic II	346 to 660	D402, E567	H614		N577	P450, P453	Y579, F448 F607, F611	T543, T568	M688
Hydrophobic 3 (H$_8$ to H$_{14}$)	661 to 874	D730, D739		R687		P669, P747, P821	Y691, Y694, F788, W750	S800, T837	

[a]The residues underlined have been submitted to site-directed mutagensis (see text and review by R. Gaber, 1992).

(Serrano et al., 1986); therefore it is impossible to grow for more than a couple of divisions an ATPase mutant modified in an essential amino acid. Furthermore, a mixture of wild-type and lethal mutant ATPases is difficult to obtain in the same strain because of the presence of the PMA2 isogene which promotes intensive DNA repair of lethal mutations. Finally, no biochemical procedure exists today to separate purified wild-type from mutant ATPases. The very smart procedures developed by Nakamoto et al. (1991) for promoting specific accumulation of mutant enzyme into secretory vesicles might provide a solution to some of these problems but is probably not applicable to lethal mutation which might not be stably inserted in the endoplasmic reticulum. Anyway, an enormous number of possibilities of site-directed single amino acid substitutions exists: 918[20] for the H$^+$-ATPase of S. cerevisiae. Choices have thus to be made. The alignment of 17 H$^+$-ATPase of Fig. 2 might help in this respect. For instance, we suggest to give some priority to the residues listed in Table I. Our choice was based on the following rationale. One of the most interesting type of information we wish to obtain is the identification of residues specifically involved in H$^+$-transport. Therefore we eliminated all residues contained in the segment a to s conserved in all eukaryotic transport-ATPases. We also did not consider the Leishmania donovani ATPase because of the uncertainty of it being a H$^+$-ATPase and the numerous alignment ambiguities. Furthermore, we eliminated the N and C terminal as well as all the 20 "insertions" not conserved in both yeast and plant ATPases. We did not consider nonpolar residues such alanine, leucine, valine, isoleucine, and glycine because of their general unspecificity and unlikeliness to be

involved in H$^+$-binding either by chelation or by direct protonation. Finally, among the remaining residues which were conserved in at least 15 of the plant and fungal ATPases, we eliminated those which were homologous to residues conserved in the NA$^+$/K$^+$- and Ca$^+$-ATPases in approximately similar position (with some flexibility to allow for misalignment) even when not located in the conserved region a to s. Using this rather empirical and partially subjective screen, we identified 38 residues: six residues in the hydrophobic region covering the predicted hydrophobic segments H1 and H2, five residues in the small hydrophilic domain, two residues in the predicted hydrophobic segments H5 and H6, 13 residues in the large hydrophylic domain, and 12 residues located in the hydrophobic region including the predicted hydrophobic segments H8–H14. Among the latter, the apparently specific conservation of three proline and four aromatic residues is rather surprising and not noticed so far to our knowledge. Among the 38 residues screened out by our approach, only three have been mutated so far (see review by Gaber, 1992). D143 and D730 were found to be essential, whereas K252 gives vanadate-resistant ATPase. It must be noted that the residue D730, which was proposed by Clarke et al. (1989) to be a H$^+$-binding site similar to the Ca^{2+}-binding site of the endoplasmic reticulum, was pinpointed by our method. This is only an example of the usefulness of our alignment. These alignments will become more precise and informative as additional sequences become available. New information analysis will have to be developed and will provide new structural and functional predictions. Unfortunately many of these cannot be tested yet biochemically, but it is to be hoped that a general method for expression

of lethal mutations of yeast H^+-ATPase will become available in the near future.

REFERENCES

Addison, R. (1986). *J. Biol. Chem.* **261**, 14896–14901.
Addison, R., and Scarborough, G. A. (1981). *J. Biol. Chem.* **256**, 13165–13171.
Boutry, M., Michelet, B., and Goffeau, A. (1989). *Biochim. Biophys. Res. Commun.* **162**, 567–574.
Bowman, B. J., and Bowman, E. J. (1986). *J. Membr. Biol.* **94**, 83–93.
Bowman, B. J., Blasco, F., and Slayman, C. W. (1981). *J. Biol. Chem.* **256**, 12343–12349.
Clarke, D. M., Loo, T. W., Inesi, G., and MacLennan, D. H. (1989). *Nature (London)* **339**, 476–478.
Davis, B. D., and Hammes, G. G. (1989). *J. Biol. Chem.* **264**, 370–374.
Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1985). *Nature (London)* **318**, 618–624.
Dufour, J. P., and Goffeau, A. (1978). *J. Biol. Chem.* **253**, 7026–7032.
Ewing, N. N., Wimmers, L. E., Meyer, D. J., Chetelat, R. T., and Bennett, A. B. (1990). *Plant Physiol.* **94**, 1874–1881.
Fasman, G. D., and Gilbert, W. A. (1990). *Trends Biochem. Sci.* **15**, 89–92.
Feng, D. F., Johnson, M. S., and Doolittle, R. F. (1985). *J. Mol. Evol.* **21**, 112–125.
Foury, F., Amory, A., and Goffeau, A. (1981). *Eur. J. Biochem.* **119**, 395–400.
Gaber, R. (1992). *International Reviews of Cytology: Membrane Transporters and Receptors*, in press.
Ghislain, M., and Goffeau, A. (1991). *J. Biol. Chem.* **266**, 18276–18279.
Ghislain, M., Schlesser, A., and Goffeau, A. (1987). *J. Biol. Chem.* **262**, 17549–17555.
Goffeau, A. (1987). In *Molecular Basis of Membrane Transport*, UNESCO Workshop (Kotyk, A., ed.), printed for UNESCO by Czechoslovak Academy of Sciences, Prague pp. 32–37.
Goffeau, A. (1988). *Brazilian J. Med. Biol. Res.* **21**, 1233–1240.
Goffeau, A., and Boutry, M. (1986). *Microbiol. Sci.* **3**, 164–167.
Goffeau, A., and Green, N. M. (1990). In *Monovalent Cations in Biological Systems* (Pasternak, C. A. ed.), CRC Press, Boca Raton, Florida, pp. 155–169.
Goffeau, A., and Slayman, C. W. (1981). *Biochim. Biophys. Acta* **639**, 197–223.
Goffeau, A., Coddington, A., and Schlesser, A. (1989). In *Molecular Biology of the Fission Yeast* (Nasim, A., Young, P., and Johnson, B., eds.), Adademic Press, Orlando, Florida, pp. 397–429.
Green, N. M. (1989). *Biochem. Soc. Trans.* **17**, 819–822.
Hager, K. M., Mandala, S. M., Davenport, J. W., Speicher, D. W., Benz, E. J., and Slayman, C. W. (1986). *Proc. Natl. Acad. Sci. USA* **83**, 7693–9697.
Harper, J. F., Surowy, T. K., and Sussmann, M. R. (1989). *Proc. Natl. Acad. Sci. USA* **86**, 1234–1238.

Harper, J. F., Manney, L., DeWitt, N. D., Yoo, M. H., and Sussmann, M. R. (1990). *J. Biol. Chem.* **265**, 13601–13608.
Hennessey, J. P., Jr., and Scarborough, G. A. (1990). *J. Biol. Chem.* **265**, 532–537.
Higgins, D. G., and Sharp, P. M. (1988). *Gene* **73**, 237–244.
Houlné, G., and Boutry, M., personal cummunication.
Klein, P., Kanehisa, M., and DeLisi, C. (1985). *Biochim. Biophys. Acta* **815**, 468–476.
Koland, J. G., and Hammes, G. G. (1986). *J. Biol. Chem.* **261**, 5936–5942.
Malpartida, F., and Serrano, R. (1980), *FEBS Lett.* **111**, 69–72.
Mandala, S. M., and Slayman, C. W. (1989). *J. Biol. Chem.* **264**, 16276–16281.
Meade, J. C., Shaw, J., Lemaster, S., Gallagher, G., and Stringer, J. R. (1987). *Mol. Cell. Biol.* **7**, 3937–3946.
Monk, B. C., Kurtz, M. B., Marrinan, J. A., and Perlin, D. (1991). *J. Bacteriol.* **173**, 6826–6836.
Morieau, L., and Boutry, M., personal communication.
Nakamoto, R. K., and Slayman, C. W. (1989). *J. Bioenerg. Biomembr.* **21**, 621–632.
Nakamoto, R. K., Rao, R., and Slayman, C. W. (1991). *J. Biol. Chem.* **266**, 7940–7949.
Pardo, J. M., and Serrano, R. (1989). *J. Biol. Chem.* **264**, 8557–8562.
Perez, C., Michelet, B., Ferrant, V., Bogaerts, P., and Boutry, M. (1992). *J. Biol. Chem.*, **267**, 1204–1211.
Perlin, D. S., Harris, S. L., Seto-Young, D., and Haber, J. E. (1989). *J. Biol. Chem.* **264**, 21857–21864.
Portillo, F., and Serrano, R. (1988). *EMBO J.* **7**, 1793–1798.
Scarborough, G. A. (1985). *Microbiol. Rev.* **49**, 214–231.
Scarborough, G. A. (1992), In *Molecular Aspects of Membrane Proteins* (de Pont, J. J. H. H. M., ed.), Elsevier Science Publishers, Amsterdam, New York, in press.
Schlesser, A., Ulaszewski, S., Ghislain, M., and Goffeau, A. (1988). *J. Biol. Chem.* **263**, 19480–19487.
Serrano, R. (1984). *Curr. Top. Cell. Regul.* **23**, 87–126.
Serrano, R. (1988). *Biochim. Biophys. Acta* **947**, 1–28.
Serrano, R. (1989). *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 61–94.
Serrano, R. (1991). In *The Molecular and Cellular Biology of the Yeast Saccharomyces, Genome Dynamics, Protein Synthesis, and Energetics*, (Broach, J. R., Pringle, J. R., and Jones, E., eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, Chap. 9, pp. 523–585.
Serrano, R., Kielland-Brandt, M. C., and Fink, G. R. (1986). *Nature (London)* **319**, 689–693.
Shull, G. E., and Greeb, J. (1988). *J. Biol. Chem.* **263**, 8646–8657.
Shull, G. E., Schwartz, A., and Lingrel, J. B. (1985). *Nature (London)* **316**, 691–695.
Slayman, C. L. (1987). *J. Bioenerg. Biomembr.* **19**, 1–20.
Subrahmanyeswara Rao, U., Hennessey, J. P., and Scarborough, G. A. (1991). *J. Biol. Chem.* **266**, 14740–14746.
Van Dyck, L., Petretski, J. H., Wolasker, H., Rodrigues, G., Schlesser, A., Ghislain, M., and Goffeau, A. (1990). *Eur. J. Biochem.* **194**, 785–790.
Wach, A., Ahlers, J., and Gräber, P. (1990). *Eur. J. Biochem.* **189**, 675–682.
Watanabe, Y., Shiramizu, M., and Tamai, Y. (1991). *J. Biochem.* **110**, 237–240.