Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 1775–1779, March 1992
Plant Biology

Molecular cloning and sequence of cDNA encoding the pyrophosphate-energized vacuolar membrane proton pump of *Arabidopsis thaliana*

(cloning/inorganic pyrophosphatase/nucleotide sequence)

VAHE SARAFIAN*†, YONGCHEOL KIM*, RONALD J. POOLE†, AND PHILIP A. REA*‡

*Plant Science Institute, Department of Biology, University of Pennsylvania, Philadelphia, PA 19104; †Department of Biology, McGill University, 1205 Avenue Dr. Penfield, Montreal, PQ H3A 1B1, Canada

Communicated by André Jagendorf, November 1, 1991 (received for review September 16, 1991)

ABSTRACT **The energy-dependent transport of solutes across the vacuolar membrane (tonoplast) of plant cells is driven by two H^+ pumps: a vacuolar ("V-type") H^+-ATPase (EC 3.6.1.3) and a H^+-translocating (pyrophosphate-energized) inorganic pyrophosphatase (H^+-PPase; EC 3.6.1.1). The H^+-PPase, like the V-type H^+-ATPase, is abundant and ubiquitous in the vacuolar membranes of plant cells, and both enzymes make a substantial contribution to the transtonoplast H^+-electrochemical potential difference. Here, we report the cloning and sequence of cDNAs encoding the tonoplast H^+-PPase of *Arabidopsis thaliana*. The protein predicted from the nucleotide sequence of the cDNAs is constituted of 770 amino acids and has a molecular weight of 80,800. It is a highly hydrophobic integral membrane protein, and the structure derived from hydrophilicity plots contains at least 13 transmembrane spans. Since the tonoplast H^+-PPase appears to be constituted of one polypeptide species and genomic Southern analyses indicate that the gene encoding the M_r 80,800 polypeptide is present in only a single copy in the genome of *Arabidopsis*, it is suggested that the H^+-PPase has been cloned in its entirety. The lack of sequence identities between the tonoplast H^+-PPase and any other characterized H^+ pump or PP_i-dependent enzyme implies a different evolutionary origin for this translocase.**

The chemiosmotic hypothesis (1) contends that membrane-bound H^+ pumps constitute the primary transducers by means of which living cells interconvert light, chemical, and electrical energy. Through the establishment and maintenance of transmembrane electrochemical gradients, H^+ pumps energize the transport of other solutes or, in the special case of the energy-coupling membranes of mitochondria, chloroplasts, and bacteria, transduce the H^+ electrochemical gradient generated by membrane-linked anisotropic redox reactions to the synthesis of ATP (1). Given the multitude of biological reactions energized by ATP, primary H^+ translocation and the interconversions of ATP have come to be recognized as the principal generators of usable energy in the cell. Intriguing, therefore, is the fact that the vacuolar membrane (tonoplast) of plant cells contains not only a H^+-ATPase (EC 3.6.1.3) (2, 3) but also an inorganic pyrophosphate-energized H^+-pyrophophatase (H^+-PPase; EC 3.6.1.1) (2). Both enzymes catalyze inward electrogenic H^+ translocation (from cytosol to vacuole lumen), but the H^+-PPase is unusual in its exclusive use of PP_i as energy source (4).

The tonoplast H^+-PPase appears to be important for plant cell function: it is widespread, active, and abundant. The enzyme is ubiquitous in the vacuolar membranes of plant cells (2) and capable of establishing a H^+ gradient of similar, and often greater, magnitude than the H^+-ATPase on the same membrane (2, 5–7). The M_r 64,500–73,000 substrate ($MgPP_i$)-binding subunit of the H^+-PPase constitutes between 1% (8) and 10% (6, 7) of total vacuolar membrane protein and the purified enzyme has a turnover number of between 50 and 100 s^{-1}, depending on source and preparation (7, 8). When account is taken of the large size of the vacuole—it can account for 90–99% of total intracellular volume—the potential bioenergetic impact of the H^+-PPase is great.

Here we describe the molecular cloning of cDNAs encoding the tonoplast H^+-PPase of *Arabidopsis thaliana*.§ While the predicted structural characteristics of the protein encoded by the clones are consistent with the capacity of the enzyme for PP_i-dependent H^+ pumping, the deduced amino acid sequence of the polypeptide encoded does not show identities with any other sequenced ion pump. We therefore propose that this enzyme is a member of a class of ion translocase that has not been previously described.

MATERIALS AND METHODS

Materials. The plant materials were *Arabidopsis thaliana* cv. Columbia, *Vigna radiata* cv. Berken (mung bean), and *Beta vulgaris* L. cv. Detroit Dark (red beet). Rabbit polyclonal antibody raised against tonoplast H^+-PPase purified from etiolated hypocotyls of *Vigna radiata* was a kind gift from M. Maeshima (Hokkaido University, Japan) to V.S. and R.J.P. The *Arabidopsis* cDNA libraries constructed in λZAP and λgt10, respectively, were provided by Joe Ecker (University of Pennsylvania) and Nigel Crawford (University of California, San Diego).

Preparation of Membranes. Tonoplast vesicles were isolated from *Beta vulgaris* storage root and etiolated hypocotyls of *Vigna radiata* as described previously (6, 9). Microsomes were prepared from 10-day-old dark-grown seedlings of *Arabidopsis thaliana* by homogenization and differential centrifugation (9).

Preparation of Antibody for Immunoscreens. The rabbit polyclonal antibody used in these studies had been raised against total peak tonoplast H^+-PPase chromatographic fractions purified from etiolated hypocotyls of *Vigna* (6). To ensure monospecificity, the antibody was affinity-purified against the M_r 66,000 substrate ($MgPP_i$)-binding subunit of the H^+-PPase from *Vigna* (7) by the method of Sambrook *et al.* (ref. 10, pp. 18.17–18.18).

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

Abbreviation: H^+-PPase, H^+-translocating inorganic pyrophosphatase.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M81892).

Polypeptide Sequencing. *Beta* tonoplast vesicles were isolated (9), the H^+-PPase was purified by the methods of Britten *et al.* (8) and Sarafian and Poole (11), and aliquots (20–50 μg) of the purified enzyme were separated by one-dimensional SDS/PAGE (7) according to the recommendations of Matsudaira (12). Electrophoresed PPase was electrotransferred to 0.45-μm nitrocellulose filters and protein was reversibly stained with Ponceau-S (13). The protein band corresponding to the substrate-binding subunit of the H^+-PPase was excised immediately after destaining, stored wet at −85°C, and shipped on dry ice to the Harvard Microchemistry Facility (Cambridge, MA) for *in situ* tryptic digestion and gas-phase sequence analysis by the method of Aebersold *et al.* (13).

Screens of *Arabidopsis* λZAP and λgt 10 cDNA Libraries. An *Arabidopsis* cDNA expression library in λZAP was screened by standard procedures (ref. 10, pp. 12.16–12.20) to isolate the primary partial clone AVP-1 (AVP = *Arabidopsis* vacuolar pyrophosphatase), which was then employed for hybridization screens of a cDNA library constructed in λgt10 for the isolation of the near-full-length clone pAVP-3.

Genomic Blot Hybridizations. Genomic DNA was extracted from 14-day-old *Arabidopsis* seedlings (14). Samples (2 μg) of the DNA were digested with restriction enzyme, electrophoresed on 0.8% agarose gels, and transferred to nylon membrane filters (Nytran-45; Schleicher & Schuell) for the hybridizations (ref. 10, pp. 9.42–9.55). The filters were exposed to Kodak X-Omat AR x-ray film with intensifying screens for 12–16 hr at −85°C.

Northern Blot Hybridizations. Total RNA was extracted from 7-day-old dark-grown *Arabidopsis* seedlings by the phenol/SDS method (14), and poly(A)$^+$ RNA was isolated by oligo(dT)-cellulose chromatography (ref. 10, pp. 7.26–7.29). The poly(A)$^+$ RNA was size-fractionated by electrophoresis on 1% agarose gels containing 0.63 M formaldehyde and blotted onto nylon membrane filters (Nytran-45). The filters were prehybridized as described for the λgt10 screens and hybridized in the same solution containing random-primed AVP-1 cDNA as probe (specific activity 1×10^8 cpm/μg) for 15–20 hr at 42°C. The filters were washed and exposed to x-ray film as described above for the genomic blot hybridizations.

DNA Sequencing. DNA was sequenced by the dideoxynucleotide chain-termination method (15), using a Sequenase II kit (United States Biochemical). A series of nested deletions derived from the cDNA insert of pAVP-1 were obtained by using a double-stranded nested deletion kit (Pharmacia). All other sequencing reactions were performed with custom-synthesized oligonucleotide primers.

RESULTS

Selection of cDNA Clones Encoding H^+-PPase. Immunoblots demonstrated that the antibody raised and affinity-purified against the M_r 66,000 $MgPP_i$-binding subunit of the H^+-PPase from *Vigna* immunoreacts specifically with an M_r 66,800 polypeptide in microsomes prepared from *Arabidopsis*. Thus, to isolate cDNAs encoding the tonoplast H^+-PPase, the affinity-purified antibody was used to screen a size-fractionated [2–3 kilobases (kb)] *Arabidopsis* cDNA library constructed in the expression vector λZAP. Approximately 200,000 plaques were screened with antibody and 6 independent positive clones were obtained. After *in vivo* excision, the plasmids were found to contain inserts ranging in length from 1.6 to 2.1 kb. Restriction mapping confirmed that all 6 inserts correspond to the same gene. Sequence analysis of two of the clones, pAVP-1 and pAVP-2, showed both to contain cDNA inserts of about 2.1 kb with nonoverlapping 5′ and 3′ ends (Fig. 1).

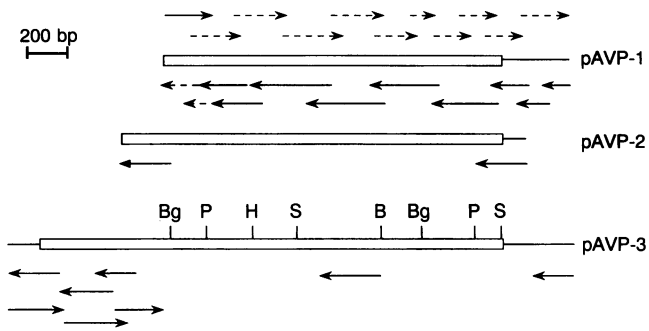

FIG. 1. Sequencing strategy and restriction map of clones pAVP-1, pAVP-2, and pAVP-3. Dashed arrows indicate inserts sequenced with T3 and T7 primers from nested deletions of pAVP-1. Solid arrows indicate sequences obtained from T3 and T7 primers, or specific custom-synthesized primers, using complete plasmid clones as templates. The restriction enzymes employed were *Bam*HI (B), *Bgl* II (Bg), *Hin*dIII (H), *Pvu* II (P), and *Sca* I (S). The insert boundaries demark the *Eco*RI sites. bp, Base pairs.

To obtain longer cDNA inserts encoding the tonoplast H^+-PPase, the cDNA insert of pAVP-1 was used as probe for filter hybridization screens of an *Arabidopsis* cDNA library constructed in λgt10. From 50,000 plaques, 2 positive clones containing cDNA inserts of 2.7 and 2.8 kb, respectively, were isolated and subcloned in pBluescript II (Stratagene). Restriction mapping demonstrated that both clones encode the same gene but contain inserts extending beyond those of pAVP-1 and pAVP-2 in both the 5′ and 3′ directions. The largest λgt10-derived clone, pAVP-3, was sequenced.

Nucleotide Sequence. The complete nucleotide sequence and predicted amino acid sequence of AVP-3 (Fig. 2) were determined by the sequencing strategy depicted in Fig. 1.

The methionine (ATG) codon at position 1 in the nucleotide sequence is identified as the translation initiation site. Only by invoking translation initiation at this point can the N-terminal sequence obtained directly by Maeshima and Yoshida (6) for the substrate-binding subunit of the H^+-PPase from *Vigna* (7) be accommodated by the deduced amino acid sequence of the protein encoded by the cDNA. Moreover, only the sequences contiguous with this, the first methionine codon (AAG ATG GT) in the near full-length clone, match the consensus sequence for plant initiator codons (R A/CA/C AUG GC) with R (purine) at −3 and G at +4 being the most critical (16). The open reading frame of the cDNA insert of pAVP-3 contains 2310 nucleotides followed by a translation termination codon (TAA) at position 2371 and 348 nucleotides of 3′ noncoding sequence (Fig. 2).

mRNA Analysis. The cDNA insert of pAVP-3 is near full length as indicated by the results of Northern blot analyses (Fig. 3). When poly(A)$^+$ RNA is isolated from 7-day-old dark-grown seedlings of *Arabidopsis*, size-fractionated, and hybridized with random-primed AVP-1 cDNA, a single RNA species with an approximate length of 2.8 kb is detected. The lengths of AVP-3 and mature *Arabidopsis* transcript therefore coincide.

Identity of cDNA Clones. The identity of the putative H^+-PPase cDNA clones from *Arabidopsis* is verified by the precise alignment of the deduced amino acid sequence of the protein encoded by AVP-3 with direct internal sequence data acquired from the M_r 64,500–67,000, $MgPP_i$-binding polypeptide of *Beta* (7, 8, 11) and N-terminal sequence data obtained from the corresponding polypeptide of *Vigna* (6).

In situ tryptic digestion (13) of the "M_r 67,000" (*Beta*-1) and "M_r 64,500" (*Beta*-2) substrate-binding subunits of the H^+-PPase purified by the methods of Sarafian and Poole (11) and Britten *et al.* (8), respectively, generates peptide fragments with identical sequences (Fig. 2). The direct sequence data from the M_r 64,500–67,000 subunit of *Beta* and the amino

```
-155     CTTAGATTTATCTTTGAGTCCCGAAACATCGAGGAACGCCTTCGAATCCCTCTCTCTCTGTGTGTGTTCTCTGTGTTCTCTCTCTCGCG   -67
 -66 CGAAGCGGTTCTCTTTCTTTTGTTTATTTGTTTTTATTTGTTTTTCTCTTATACGGAGGAGAGAAGATGGTGGCGCCTGCTTTGTTACCGGAG    27
                                                                  1-MetValAlaProAlaLeuLeuProGlu     9
  28 CTCTGGACGGAGATCCTTGTACCGATTTGTGCGGTGATTGGTATCGCCTTTTCGCTTTTCCAATGGTACGTTGTATCTCGCGTGAAACTCACC   120
  10 LeuTrpThrGluIleLeuValProIleCysAlaValIleGlyIleAlaPheSerLeuPheGlnTrpTyrValValSerArgValLysLeuThr    40
 121 TCTGACCTCGGCGCATCGTCTTCCGGTGGAGCTAACAATGGGAAGAATGGATACGGTGATTATCTAATCGAGGAAGAGGAAGGTGTTAATGAC   213
  41 SerAspLeuGlyAlaSerSerGlyGlyAlaAsnAsnGlyLysAsnGlyTyrGlyAspTyrLeuIleGluGluGluGluGlyValAsnAsp     71
 213 CAGAGTGTTGTCGCTAAGTGCGCTGAGATTCAGACTGCTATTTCCGAAGGTGCAACTTCATTCCTATTCACGGAGTACAAATATGTTGGTGTC   306
  72 GlnSerValValAlaLysCysAlaGluIleGlnThrAlaIleSerGluGlyAlaThrSerPheLeuPheThrGluTyrLysTyrValGlyVal   102
 307 TTCATGATTTTCTTTGCTGCTGTTATCTTTGTTTTCCTCGGCTCTGTTGAGGGATTCAGCACTGATAACAAGCCTTGTACTTACGACACCACC   399
 103 PheMetIlePhePheAlaAlaValIlePheValPheLeuGlySerValGluGlyPheSerThrAspAsnLysProCysThrTyrAspThrThr   133
 400 AGAACCTGCAAGCCTGCATTGGCTACTGCAGCTTTCAGTACCATTGCTTTCGTGCTTGGTGCTGTTACCTCTGTTCTATCTGGTTTCCTTGGG   492
 134 ArgThrCysLysProAlaLeuAlaThrAlaAlaPheSerThrIleAlaPheValLeuGlyAlaValThrSerValLeuSerGlyPheLeuGly   164
 493 ATGAAGATTGCTACATACGCTAATGCTAGGACCACTTTGGAGGCGAGGAAAGGTGTTGGAAAGGCGTTCATTGTTGCATTCAGGTCTGGTGCT   585
 165 MetLysIleAlaThrTyrAlaAsnAlaArgThrThrLeuGluAlaArgLysGlyValGlyLysAlaPheIleValAlaPheArgSerGlyAla   195
 586 GTGATGGGTTTCCTTCTTGCAGCGAGTGGTCTATTGGTGCTTTACATTACTATCAATGTGTTCAAGATCTATTACGGAGATGACTGGGAAGGT   678
 196 ValMetGlyPheLeuLeuAlaAlaSerGlyLeuLeuValLeuTyrIleThrIleAsnValPheLysIleTyrTyrGlyAspAspTrpGluGly   226
 679 CTTTTTGAGGCTATTACTGGTTATGGTCTTGGTGGGTCTTCCATGGCTCTCTTTGGCCGTGTTGGTGGTGGGATCTACACTAAGGCTGCTGAT   771
 227 LeuPheGluAlaIleThrGlyTyrGlyLeuGlySerSerMetAlaLeuPheGlyArgValGlyGlyGlyIleTyrThrLysAlaAlaAsp     257
 772 GTCGGCGCTGACCTTGTCGGTAAAATTGAGAGGAATATTCCAGAGGATGATCCAAGAAACCCAGCTGTCATTGCTGATAATGTCGGTGACAAT   864
 258 ValGlyAlaAspLeuValGlyLysIleGluArgAsnIleProGluAspAspProArgAsnProAlaValIleAlaAspAsnValGlyAspAsn   288
 865 GTTGGTGACATTGCTGGTATGGGATCTGATCTCTTTGGATCATATGCTGAAGCATCATGCGCTGCTCTTGTTGTTGCCTCGATCTCATCTTTC   957
 289 ValGlyAspIleAlaGlyMetGlySerAspLeuPheGlySerTyrAlaGluAlaSerCysAlaAlaLeuValValAlaSerIleSerSerPhe   319
 958 GGAATCAACCACGACTTCACTGCCATGTGCTACCCATTGCTCATCAGTTCAATGGGAATCTTGGTTTGTTTGATCACAACTCTCTTTGCCACT  1050
 320 GlyIleAsnHisAspPheThrAlaMetCysTyrProLeuLeuIleSerSerMetGlyIleLeuValCysLeuIleThrThrLeuPheAlaThr   350
1051 GACTTCTTTGAGATTAAGCTTGTCAAGGAGATTGAACCAGCATTGAAGAACCAGCTCATTATCTCAACTGTTATTATGACTGTTGGTATTGCT  1143
 351 AspPheGluIleLysLeuValLysGluIleGluProAlaLeuLysAsnGlnLeuIleIleSerThrValIleMetThrValGlyIleAla   381
1144 ATTGTGTCATGGGTTGGCTTACCGACCTCCTTTACCATCTTCAACTTTGGAACACAAAAAGTTGTCAAGAACTGGCAGCTATTCCTTTGTGTT  1236
 382 IleValSerTrpValGlyLeuProThrSerPheThrIlePheAsnPheGlyThrGlnLysValValLysAsnTrpGlnLeuPheLeuCysVal   412
1237 TGTGTTGGTCTTTGGGCTGGACTCATTATTGGTTTCGTCACTGAGTACTACACTAGTAACGCCTACAGCCCTGTGCAAGATGTTGCAGATTCA  1329
 413 CysValGlyLeuTrpAlaGlyLeuIleIleGlyPheValThrGluTyrTyrThrSerAsnAlaTyrSerProValGlnAspValAlaAspSer   443
1330 TGCAGAACTGGTGCAGCTACCAATGTTATCTTCGGCCTTGCTCTTGGTTACAAATCCGTCATTATTCCAATCTTTGCTATTGCTATCAGTATA  1422
 444 CysArgThrGlyAlaAlaThrAsnValIlePheGlyLeuAlaLeuGlyTyrLysSerValIleIleProIlePheAlaIleAlaIleSerIle   474
1423  TTCGTTAGCTTCAGCTTTGCTGCTATGTATGGTGTTGCTGTTGCTGCTCTTGGTATGCTCAGTACCATTGCCACTGGTTTGGCAATTGATGCT  1515
 475 PheValSerPheSerPheAlaAlaMetTyrGlyValAlaValAlaAlaLeuGlyMetLeuSerThrIleAlaThrGlyLeuAlaIleAspAla   505
1516 TATGGTCCCATCAGTGACAATGCTGGTGGTATTGCTGAAATGGCTGGAATGAGCCACCGCATCCGTGAAAGAACTGATGCTCTTGATGCCGCT  1608
 506 TyrGlyProIleSerAspAsnAlaGlyGlyIleAlaGluMetAlaGlyMetSerHisArgIleArgGluArgThrAspAlaLeuAspAlaAla   536
1609 GGAAACACCACTGCTGCTATTGGAAAGGGATTTGCCATTGGCTCTGCTGCCCTAGTCTCCTTGGCTCTCTTTGGTGCCTTTGTGAGCCGTGCA  1701
 537 GlyAsnThrThrAlaAlaIleGlyLysGlyPheAlaIleGlySerAlaAlaLeuValSerLeuAlaLeuPheGlyAlaPheValSerArgAla   567
1702 GGGATCCACACCGTAGATGTTTTGACCCCTAAAGTTATCATTGGGCTCCTTGTTGGTGCCATGCTTCCTTACTGGTTCTCTGCCATGACAATG  1794
 568 GlyIleHisThrValAspValLeuThrProLysValIleIleGlyLeuLeuValGlyAlaMetLeuProTyrTrpPheSerAlaMetThrMet   598
1795 AAGAGTGTGGGAAGTGCAGCTCTTAAGATGGTTGAAGAAGTTCGCAGGCAGTTCAACACCATCCCTGGACTTATGGAAGGAACCGCAAAACCA  1887
 599 LysSerValGlySerAlaAlaLeuLysMetValGluGluValArgArgGlnPheAsnThrIleProGlyLeuMetGluGlyThrAlaLysPro   629
1888 GACTACGCCACATGTGTCAAGATCTCCACCGATGCTTCCATCAAGGAAATGATACCTCCTGGTTGCCTTGTCATGCTCACACCTCTCATTGTT  1980
 630 AspTyrAlaThrCysValLysIleSerThrAspAlaSerIleLysGluMetIleProProGlyCysLeuValMetLeuThrProLeuIleVal   660
1981 GGTTTCTTCTTTGGAGTTGAGACCCTCTCTGGTGTCCTCGCCGGATCTCTTGTATCCGGTGTTCAGATCGCCATATCAGCATCTAACACTGGT  2073
 661 GlyPhePhePheGlyValGluThrLeuSerGlyValLeuAlaGlySerLeuValSerGlyValGlnIleAlaIleSerAlaSerAsnThrGly   691
2074 GGTGCCTGGGACAACGCCAAGAAATACATCGAGGCTGGTGTATCAGAGCACGCAAAGAGCCTTGGACCAAAGGGTTCAGAGCCACACAAGGCA  2166
 692 GlyAlaTrpAspAsnAlaLysLysTyrIleGluAlaGlyValSerGluHisAlaLysSerLeuGlyProLysGlySerGluProHisLysAla   722
2167 GCTGTGATTGGAGACACAATTGGAGACCCATTGAAGGATACTTCAGGACCTTCATTGAACATCCTCATCAAGCTCATGGCTGTTGAGTCTCTT  2259
 723 AlaValIleGlyAspThrIleGlyAspProLeuLysAspThrSerGlyProSerLeuAsnIleLeuIleLysLeuMetAlaValGluSerLeu   753
2260 GTCTTTGCTCCCTTCTTCGCCACTCACGGTGGTATCCTTTTCAAGTACTTCTAAACTCAATCCGAGGGAAGAAGATGACGATGATGAAGAAGA  2352
 754 ValPheAlaProPhePheAlaThrHisGlyGlyIleLeuPheLysTyrPhe-770
2353  AGAAGATGATGATGGCGATCGATTCTAAACTTTCTTTTTTACCATTCTTATTTTCGTTTACCGTAGGTGGTTAAAAAACCTTTTTGTTGATGA  2445
2446 GGCTCATTTAAAGAACCAACCAAATGATGTTTCTTTCTCTCACTCTCTGTCTTTCTGTTTTCTTTTTGTTCTGTTTAGAATTTAGAAATCCAC  2538
2539 CAAGTATTCGGTCGAGACTTGTTTTAGCCGTTACTTTCTGCTGCTTATATTTCCTAAATTGGTTGTCTTCTTCGAAACATAATTGGAATTTAT  2631
2632 TGTTACTGTTAGTCTAAAAAAAAAAAA  2658
```

```
Arabidopsis          1 ATGGTGGCGCCTGCTTTGTTACCGGAGCTCTGGACGGAGATCCTTGTACCGATTTGTGCGGTGATTGGTATCGCCTTTTCGCTTTTCCAATGGTAC
                     1 MetValAlaProAlaLeuLeuProGluLeuTrpThrGluIleLeuValProIleCysAlaValIleGlyIleAlaPheSerLeuPheGlnTrpTyr
                              :    | :  |  |  : |     |  |  |  |  : |  :     |  |  |     |  |  |     |  |  |  |
Vigna                          GlyAlaAlaIleLeuProAspLeuGlyThrGluIleLeuIleProVal---AlaValIleThrIleAlaPheAlaLeuPheGlnTrpLeu

Arabidopsis        763 GCTGCTGATGTCGGCGCTGACCTTGTCGGTAAAATTGAG
                   255 AlaAlaAspValGlyAlaAspLeuValGlyLysIleGlu
                        |  |  |  |  |  |  |  |  |  |  |  |  :  |
Beta-2 (NT25-2)         AlaAlaAspValGlyAlaAspLeuValGlyLysValGlu

Arabidopsis       1588 ACTGATGCTCTTGATGCCGCTGGAAACACCACTGCTGCTATTGGAAAG
                   530 ThrAspAlaLeuAspAlaAlaGlyAsnThrThrAlaAlaIleGlyLys
                        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Beta-1 (NT25)           ThrAspAlaLeuAspAlaAlaGlyAsnThrThrAlaAlaIleGlyLys
Beta-2 (NT22)              AspAlaCysAspAlaAlaGlyAsnThrThrAlaAlaIleGly

Arabidopsis       1699 GCAGGGATCCACACCGTAGATGTTTTGACCCCTAAA
                   567 AlaGlyIleHisThrValAspValLeuThrProLys
                        |  |  :  |  |  |  |  |  |  |  |
Beta-1 (NT31)           AlaSerIleGlnThrValAspValLeuThrProLys

Arabidopsis       2166 GCAGCTGTGATTGGAGACACAATTGGAGACCCATTGAAGGATACTTCAGGACCTTCATTGAACATCCTCATCAAG
                   722 AlaAlaValIleGlyAspThrIleGlyAspProLeuLysAspThrSerGlyProSerLeuAsnIleLeuIleLys
                        |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
Beta-1 (NT53)           AlaAlaValIleGlyAspThrIleGlyAspProLeuLysAspThrSerGlyProSerLeuAsnIleLeu---Lys
Beta-2 (NT41)                ValIleGlyAspThrIleGlyAsp---Leu---Asp
```

FIG. 2. Nucleotide sequence of *Arabidopsis* cDNA encoding H^+-PPase and predicted amino acid sequence of polypeptide. (*Upper*) Total sequence of cDNA insert of pAVP-3. The underlined amino acids align with the direct sequence data from the $MgPP_i$-binding subunit of the H^+-PPase isolated from *Beta* and *Vigna*. (*Lower*) Alignments of direct sequence data acquired from $MgPP_i$-binding subunit of the tonoplast H^+-PPase isolated from *Beta* and *Vigna* with deduced amino acid sequence AVP-3. The sequences of the tryptic fragments derived from the "M_r 67,000" (*Beta*-1) and "M_r 64,500" (*Beta*-2) polypeptides of the H^+-PPase purified from *Beta* by the methods of Sarafian and Poole (11) and Britten *et al.* (8), respectively, were obtained after SDS/PAGE and electrotransfer to nitrocellulose (13). The N-terminal sequence data for the enzyme from *Vigna* were taken from Maeshima and Yoshida (6).

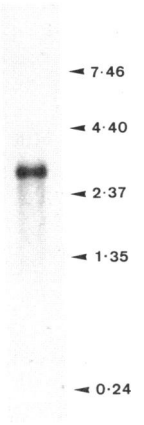

FIG. 3. Northern blot analysis of *Arabidopsis* tonoplast H^+-PPase. One microgram of poly-$(A)^+$ RNA isolated from 7-day-old dark-grown *Arabidopsis* seedlings was electrophoresed, blotted, and probed with radiolabeled cDNA insert AVP-1. The positions of the RNA molecular size markers (kb) are indicated.

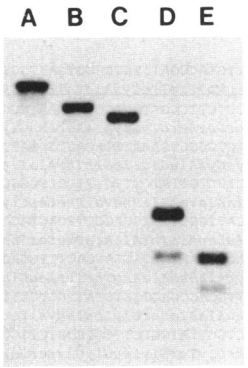

FIG. 4. Genomic Southern analysis of tonoplast H^+-PPase. *Arabidopsis* genomic DNA digested with *Xba* I (lane A), *Eco*RV (lane B), *Eco*RI (lane C), *Hin*dIII (lane D), or *Bgl* II (lane E) was electrophoresed, blotted, and hybridized with ^{32}P-labeled AVP-1. The positions of the DNA molecular weight markers (kb) are indicated.

acid sequence deduced from the nucleotide sequence of pAVP-3 show complete identity over a total span of 66 amino acid residues except for two conservative (Val → Ile, Gln → His) substitutions at positions 266 and 570 and one nonconservative (Ser → Gly) substitution at 568. Comparison of the deduced N-terminal sequence of the open reading frame of the cDNA insert of pAVP-3 and the N terminus of the substrate-binding subunit of the H^+-PPase from *Vigna* (6), on the other hand, reveals 19 identities and 5 conservative substitutions within a span of 30 amino acid residues starting at position 3 (Fig. 2).

Number of Tonoplast H^+-PPase Genes in *Arabidopsis*. A Southern blot analysis of *Arabidopsis* genomic DNA, using the 2.1-kb cDNA insert of pAVP-1 as radioactive probe, was performed to enumerate the genes encoding the substrate-binding subunit of the H^+-PPase in this organism (Fig. 4). Of the five enzymes chosen for the genomic digestions, two (*Hin*dIII and *Bgl* II) each have one restriction site within the 2.1-kb fragment, while the other three (*Xba* I, *Eco*RV, and *Eco*RI) do not (Fig. 1). Hence, providing that the region, or regions, of the genome corresponding to the 2.1-kb fragment is devoid of introns containing identical restriction sites, a single band from the *Xba* I, *Eco*RV, and *Eco*RI digests and two bands from each of the *Hin*dIII and *Bgl* II digests would be expected if the H^+-PPase is encoded by a single gene. This is exactly what is found. Genomic DNA digested with *Xba* I, *Eco*RV, or *Eco*RI and probed with the 2.1-kb fragment yields intensely hybridizing bands at 7, 8, and 12 kb, respectively. The *Hin*dIII and *Bgl* II digests, on the other hand, each contain two hybridizing bands at 2.6 and 1.83 kb and 1.78 and 1.4 kb, respectively.

Deduced Amino Acid Sequence. The open reading frame of AVP-3 encodes a 770 amino acid polypeptide with a pI of approximately 4.95 and a predicted M_r of 80,800. While the estimated M_r of the protein encoded by the insert is approximately 8000 larger than the highest apparent M_r (73,000) determined by SDS/PAGE of the enzyme from *Vigna* (6), 13,800–16,300 greater that the estimated M_r (64,500–67,000) of the corresponding subunit from *Beta* (8, 11), and about 14,000 greater than the M_r 66,800 immunoreactive polypeptide reported here for *Arabidopsis*, the anomalous migration of hydrophobic membrane proteins on SDS gels is not unusual. The binding of nonsaturating amounts of SDS by membrane proteins is common and frequently accompanied by large shifts in apparent M_r as a result of the exposure of charged amino acid residues and/or irregularities in the shape of the SDS–protein complex (17).

Computer-assisted hydrophilicity plots of the *Arabidopsis* H^+-PPase amino acid sequence establish that the cDNA insert of pAVP-3 encodes an extremely hydrophobic integral membrane protein (Fig. 5). Membrane-associated α-helices were determined by hydrophobic moment analysis (19) and some of the secondary structural characteristics of the hydrophilic sequences were assigned by the Garnier method (20). The model contains 13 transmembrane spans, all of which are amphipathic ("multimeric") and may interact to

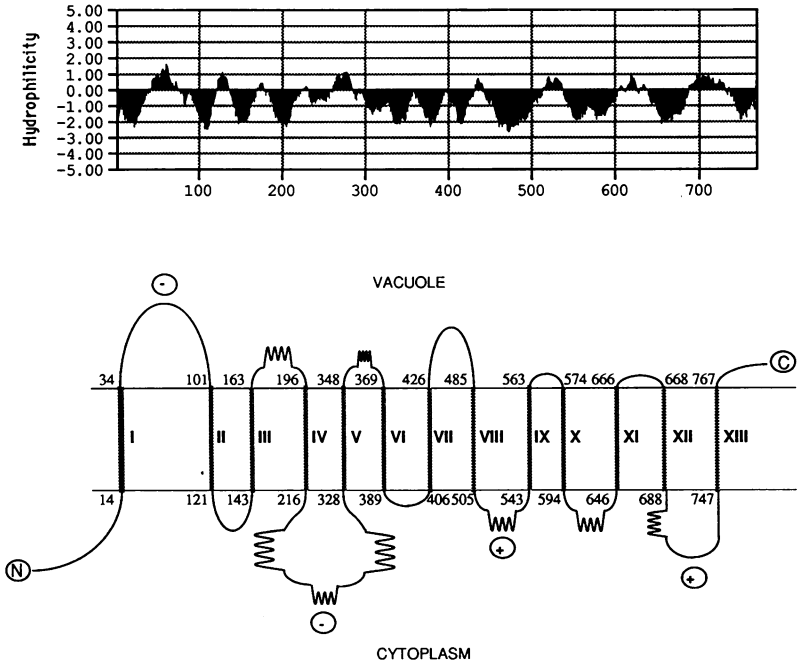

FIG. 5. Computer-assisted hydrophilicity plot (*Upper*) and topographic model (*Lower*) of *Arabidopsis* tonoplast H^+-PPase amino acid sequence. The plot was calculated according to Kyte and Doolittle (18) with the MacVector program over a running window of 20 amino acid residues. Values above the median represent hydrophilic segments; values below the median represent hydrophobic segments. Numbers I through XIII represent the putative membrane-spanning segments depicted in the tentative topographic model of the tonoplast H^+-PPase. Transmembrane segments, all of which appear to be multimeric on the basis of their hydrophobic moment, were predicted by the HELIXMEM program of PC/GENE. The structure of the nontransmembrane regions was examined with both the GARNIER and GGBSM programs of PC/GENE. The structures indicated are ww, α-helix; —, random coil; ⊖, clusters of negative charge; ⊕, clusters of positive charge; N, amino terminus; and C, carboxyl terminus.

stabilize the structure through the formation of interhelical H bonds and salt bridges. In addition, several of the hydrophilic domains are characterized by the presence of clusters of charged residues that might participate in anion (PP_i^{4-}, $MgPP_i^{2-}$) or cation (K^+, Mg^{2+}, Ca^{2+}) binding. Thus, the hydrophilic segment linking transmembrane spans I and II contains four contiguous Glu residues (positions 64–67), the segment linking spans IV and V contains four acidic residues in a stretch of eight amino acids (positions 222–229), the segment between spans VIII and IX contains the sequence Arg-Xaa-Arg-Xaa-Arg (positions 525–529), and the penultimate hydrophilic domain, linking spans XII and XIII, contains a preponderance of Lys residues. The overall orientation of the M_r 81,000 polypeptide is depicted as shown in Fig. 5 in accord with the "positive-inside rule" (21) wherein, for the majority of polytopic membrane proteins, most of the positively charged amino acids are disposed towards the cytoplasmic face of the membrane.

Computer searches of both the nucleotide sequence of the cDNA insert of pAVP-3 and the deduced amino acid sequence of the polypeptide encoded by the clone against the GenBank, Genpept, and Swiss-Prot data bases (as of September, 1991) yielded no detectable homology between this pump and any other sequenced ion translocase or PP_i-dependent enzyme.

CONCLUSIONS

The $MgPP_i$-binding subunit of the tonoplast H^+-PPase appears to be the sole polypeptide constituting the functional enzyme complex. This component, alone, copurifies with PPase activity during detergent solubilization and chromatography (6–8, 11) and is the only polypeptide of tonoplast vesicles to undergo $MgPP_i$-protectable, free PP_i-potentiated labeling by *N*-[^{14}C]ethylmaleimide (7, 8). In addition, selective purification of the $MgPP_i$-binding subunit of the tonoplast H^+-PPase from *Vigna* and its incorporation into artificial phospholipid bilayers generate proteoliposomes capable of both $MgPP_i$ hydrolysis and $MgPP_i$-dependent H^+ translocation (C. J. Britten and P.A.R., unpublished results). Together with the amino acid sequence data derived from pAVP-3, which show that the $MgPP_i$-binding subunit of the tonoplast H^+-PPase is highly hydrophobic and possesses multiple transmembrane spans, these findings strongly suggest that subunits in addition to the M_r 81,000 polypeptide encoded by AVP-3 need not be implicated to account for the capacity of the tonoplast H^+-PPase for PP_i-energized transtonoplast H^+ translocation.

On the basis of the results reported here and its biochemical characteristics, the tonoplast H^+-PPase must be ascribed to its own category of H^+ translocase. Sequence identities with any other characterized ion translocase are lacking; the type-specific inhibitors azide, orthovanadate, and bafilomycin, which selectively and strongly inhibit F-, P-, and V-type ATPases, respectively, are without effect on the tonoplast H^+-PPase (2); and none of the subunits of F-, P-, and V-type ATPases are immunologically crossreactive with the $MgPP_i$-binding subunit of the enzyme (P.A.R., unpublished results; V.S. and R.J.P., unpublished results). Similarly, close phylogenic links between the tonoplast H^+-PPase and soluble PPases are unlikely. All known soluble PPases have subunit molecular weights different from the tonoplast H^+-PPase—20,000 for the enzymes from prokaryotes (22) and 32,000–42,000 for the enzymes from eukaryotes (e.g., refs. 23, 24)—and none of the known sequences for soluble PPases [*Arabidopsis* (25), *Escherichia coli*, (22), *Kluveromyces lactis*, (26), and *Saccharomyces cerevisiae* (27)] align with the deduced sequence of the tonoplast H^+-PPase. The recently cloned catalytic M_r 28,000–30,000 subunit of the membrane-bound mitochondrial H^+-PPase (28) may also be eliminated as a homolog: the subunit from *Saccharomyces* is 49% identical to the soluble PPase from the same source (28) and shows no sequence identities with the tonoplast H^+-PPase from *Arabidopsis*, and the corresponding enzyme from rat liver mitochondria does not crossreact with antibody raised against the tonoplast H^+-PPase from *Vigna* (29).

Evaluation of a potentially more promising evolutionary relationship between the tonoplast H^+-PPase and the reversible H^+-translocating PPase (H^+-PP_i synthase) of phototrophic bacteria awaits the acquisition of sequence data from the latter enzyme. The existence of a H^+-PP_i synthase on the energy-coupling membranes of phototrophic bacteria, notably *Rhodospirillum rubrum*, has been known for some time, but only recently has it been shown that this translocase is an integral membrane protein with an apparent M_r of 56,000 (30). Two features of the M_r 56,000 polypeptide are significant: it is immunologically crossreactive with the $MgPP_i$-binding subunit of the tonoplast H^+-PPase (E. J. Kim, Y.K., and P.A.R., unpublished results) and, unlike the M_r 28,000–30,000 peripheral catalytic subunit of the mitochondrial H^+-PPase (28), it, alone, is capable of mediating both $MgPP_i$ hydrolysis and H^+ translocation (30). Genomic screens of *Rhodospirillum* using the cDNA insert of pAVP-1 and derivatives thereof as probes should allow us to examine the exciting possibility that the H^+-PPases of phototrophic bacteria and plant vacuoles have common evolutionary origins.

We thank Drs. Joe Ecker, Nigel Crawford, M. Maeshima, W. Lane, and Margaret Ahmad for the provision of some of the materials employed in this work. This work was supported by the National Science Foundation (Grant DCB-9005330), the University Research Foundation, University of Pennsylvania, and a grant from the Biological Sciences Research Group, University of Pennsylvania, awarded to P.A.R.; V.S. was partly supported by the Natural Sciences and Engineering Research Council, Canada, by a grant awarded to R.J.P.

1. Mitchell, P. (1961) *Chemiosmotic Coupling in Oxidative and Photosynthetic Phosphorylation* (Glyn Research, Bodmin, Cornwall, England).
2. Rea, P. A. & Sanders, D. (1987) *Physiol. Plant* **71**, 131–141.
3. Nelson, N. & Taiz, L. (1989) *Trends Biochem. Sci.* **14**, 113–116.
4. Rea, P. A. & Poole, R. J. (1986) *Plant Physiol.* **81**, 126–129.
5. Rea, P. A. & Poole, R. J. (1985) *Plant Physiol.* **77**, 46–52.
6. Maeshima, M. & Yoshida, Y. (1989) *J. Biol. Chem.* **264**, 20068–20073.
7. Rea, P. A., Britten, C. J. & Sarafian, V. (1992) *J. Biol. Chem.*, in press.
8. Britten, C. J., Turner, J. C. & Rea, P. A. (1989) *FEBS Lett.* **256**, 200–206.
9. Rea, P. A. & Turner, J. C. (1990) *Methods Plant Biochem.* **3**, 385–405.
10. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
11. Sarafian, V. & Poole, R. J. (1989) *Plant Physiol.* **91**, 34–38.
12. Matsudaira, N. (1988) *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Academic, New York).
13. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. & Kent, S. B. H. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6970–6974.
14. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Lutcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F. & Sceele, G. A. (1987) *EMBO J.* **6**, 43–48.
17. Maddy, A. H. (1976) *J. Theor. Biol.* **62**, 315–326.
18. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
19. Eisenberg, D., Scharz, E., Komaromy, M. & Wall, J. (1984) *J. Mol. Biol.* **179**, 125–142.
20. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
21. von Heijne, G. & Gavel, Y. (1988) *Eur. J. Biochem.* **174**, 671–678.
22. Lahti, R., Pitkaranta, T., Valve, E., Ilta, I., Kukko-Kalske, E. & Heinonen, J. (1988) *J. Bacteriol.* **170**, 5901–5907.
23. Cohen, S. A., Sterner, R., Keim, P. S. & Henrikson, R. L. (1978) *J. Biol. Chem.* **253**, 889–897.
24. Krishnan, V. A. & Gnanam, A. (1988) *Arch. Biochem. Biophys.* **260**, 277–284.
25. Kieber, J. J. & Signer, E. R. (1991) *Plant Mol. Biol.* **16**, 345–348.
26. Stark, M. J. R. & Milner, J. S. (1989) *Yeast* **5**, 35–50.
27. Kolakowski, L. F., Schloesser, M. & Cooperman, B. S. (1988) *Nucleic Acids Res.* **22**, 10441–10452.
28. Lundin, M., Baltscheffsky, H. & Ronne, H. (1991) *J. Biol. Chem.* **266**, 12168–12172.
29. Maeshima, M. (1991) *Eur. J. Biochem.* **196**, 11–17.
30. Nyren, P., Nore, B. F. & Strid, A. (1991) *Biochemistry* **30**, 2883–2887.