BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **251,** 252–255 (1998)
ARTICLE NO. RC989452

Identification of a Histidine Acid Phosphatase (*phy*A)-like Gene in *Arabidopsis thaliana*

Edward J. Mullaney and Abul H. J. Ullah
Southern Regional Research Center, ARS, USDA, New Orleans, Louisiana 70124

Received August 31, 1998

A close examination of the protein sequence encoded by the *Arabidopsis thaliana* gene F21M12.26 reveals the gene product to be a phosphomonoesterase, acid optimum (EC 3.1.3.2). A subclass of this broad acid phosphatase is also known as 'histidine acid phosphatase.' This is the first sequence-based evidence for a 'histidine acid phosphatase' in a dicotyledon. One important member of this class of enzymes is *Aspergillus niger (ficuum)* phytase, which came into prominence for its commercial application as a feed additive. The putative protein from *A. thaliana* gene F21M12.26 shares many important features of *Aspergillus* phytase, namely, size, active-site sequence, catalytic dipeptide and ten cysteine residues located in the key areas of the molecule, but lacks all nine N-glycosylation sites.

Myo-inositol hexakisphosphate phosphohydrolase, phytase, is a member of broad acid optimum phosphomonoesterase. This group is also referred to as 'acid phosphatase.' While all phytases show acid phosphatase-like activity by hydrolyzing an array of phosphomonoesters, i.e. sugar phosphates, nucleotide phosphates, etc., the reverse is not true. We have purified acid phosphatase in *A. niger,* which cannot accept phytic acid as a substrate (1). The acid phosphatases, orthophosphoric-monoester phosphohydrolase, that are able to hydrolyze phosphomonoesters from simple sugars and nucleotides but not from a myo-inositol backbone were given the Enzyme Commission nomenclature EC 3.1.3.2. Acid phosphatases that are able to liberate inorganic acid phosphates from a myo-inositol backbone are referred to as phytase and were assigned the Enzyme Commission designation EC 3.1.3.8, myo-inositol-hexakisphosphate 3-phosphohydrolase, or 3.1.3.26, myo-inositol-hexakisphosphate 6-phosphohydrolase (2). Microbial phytases, including that from *A. niger,* hydrolyze preferentially the third phosphate group of the myo-inositol hexakisphosphate. The plant phytase shows a preference for hydrolyzing the sixth phosphate group of the same molecule. The microbial and plant phytases are grouped under 3-phytase (EC 3.1.3.8) and 6-phytase (3.1.3.26), respectively (3, 4).

An active phytase had been reported in *A. niger (ficuum)* NRRL 3135 as early as 1968 by Shieh and Ware (5). This fungal enzyme has been the focus of study for over a decade by numerous groups (6). Unlike plant phytases, which are difficult to purify, *A. niger* phytase is well characterized and its sequence determined by Edman degradation and also deduced from DNA sequencing (7, 8). The gene coding for this enzyme was named *phy*A (8). This phytase had two pH optima, pH 5.0 and pH 2.5, respectively. The temperature optima was found to be 58°C. Later, we identified another acid phosphatase activity from the same fungal culture filtrate. This enzyme was markedly different from the *phy*A gene product, having only one pH optimum at pH 2.5 (9). Since then the enzyme was shown to be another efficient phytase (1). Because of both kinetic and sequence differences, the enzyme was referred to as *phy*B gene product (10). Both phyA and phyB proteins share a common octa-peptide active site region (11, 12) and have 10 cysteines (7, 10). In phyA, all ten cysteines are known to participate in forming five crucial disulfide bridges (13).

In this paper we have examined the deduced amino acid sequence of the *A. thaliana* gene F21M12.26 and compared the sequence with that of both *phy*A and *phy*B gene product. Based on this comparison we infer that the gene F21M12.26 from *A. thaliana* may code for a 'histidine acid phosphatase' with phytase activity.

RESULTS AND DISCUSSION

Active Site Similarities

Searching the Genbank database for sequence similarities with the septa-peptide active site sequence of *A. niger,* RHGXRXP, led to the identification of *A. thaliana* gene F21M12.26 as a putative histidine acid phosphatase. The septa peptide active site region was

0006-291X/98 $25.00

```
phyA  ASRNQSSCD-T-VDQGY-QCFSETSHLWG--QYAPFFSLANESVISPEVPAGCRVTFAQVLSRHGARYP
phyB  AAIPQSTQE-KQFSQEFRDGYSILKH-YG--GNGPYSERVSYG-IARDPPTGCEVDQVIMVKRHGERYP
A.Th  TVWIILLCLFV-VSQAD-GGFDVRHHLSTVTRYSTSKDVTQNLIEGSNVPSECTPIHLNLVARHGTRSP

phyA  TDSKGK-------KYSAIEEI-QQNATTFDGKYAFLKTYNYS----LGAD-------DLTPF-GEQEL
phyB  SPSAGK-------SIEEALAKVYSINTTEYKGDLAFLNDWTYY----VPNECYYNAETTSGPYAGLLDA
A.Th  TKKRLRELESLAGRFKEIVRDA-EARKLPSDKIPGWLGQWKSPWEGKVKGG-------ELIRQ-GEDEL

phyA  VNSGIKFYQRYESL--TRNIVPFIRSSGSSRVIASGKKFIEGFQSTKLKDPRAQPGQSSPKIDVYISEA
phyB  YNHGNDYKARYGHLWNGETVVPFF-SSGYGRVIETARKFGEGFFGYNYS---------TNAALNIISES
A.Th  YQLGIRVRERFPSL--FEEDYHPDVYTIRATQIPRASASAVAFGMGLFSE-KGNLGPGRNRAFAVTSEN

phyA  SSSNNTLDP-GTCTVFED-SELADTVEANFTATFVPSIRQRLENDLSGVTLTDTEVTYLMDMCSFDTIS
phyB  EVMGADSLT-PTCDTDND-QTTCDNLTYQLPQ-FKVAA-ARLNSQNPGMNLTASDVYNLMVMASFELNA
A.Th  RASDTKLRFFECCQNYKSYRKAKEPAVDKLKEPVLNKITASVAKRY-DLKFTKQDISSLWFLCKQVALL

phyA  TSTVDTKLSPFCDLFTHDEWINYDYLQSLKKYYGHGAGNPLGPTQG-----VGYANELIARLTHSPVHD
phyB  RP-----FSNWINAFTQDEWVSFGYVEDLNYYYCAGPGDKNMAAVG-----AVYANASLTLLNQGPK--
A.Th  EWTDD--LEVF-LLKGYGNSLNYKMGVPLLEDVLHSMEEAIKAREEKLPPGSYEKARL-RFAHAETIVP

phyA  DTSS--NHTLDSSPATFPLNSTLYADFSHDNGIISIL-FALGLYNGTKPLSTTTVENITQTDGFSSAW-
phyB  ----------EAGP--------LFFNLAHDTNITPILAALGVLIPNEDLP---LDRVAFGNPYSIGN-I
A.Th  PFSCLLGLFLDGSEFEKIQKEKPLELPPQPPKTRDFRGSTMAPFGCNNILVLYSCPAESSPKYFVQVLH

phyA  -TVPFASRLYVEMMQCQA----EQ-EPLVRVLVNDRVVPLHGCPVDALGRCTRDSFVR-GL--------
phyB  -IVPMGGHLTIERLSCQATALSDK-GTYVRLVLNEAVLPFNDCTSGPGYSCPLANYTSILNKNLPDYTT
A.Th  NEHPIAVPGCDGKDFCPL----EDFKAKVVTPHLKHAFDNL-CNADLNDLKQKPASSK--LSILS-SWL

phyA  ------SFARSGGDW---------------AEC-----FA*
phyB  TCNVSASYPQYLSFWWNYNTTTELNYRSSPIACQEGDAMD*
A.Th  ----FGSSHDTEL*
```

FIG. 1. Amino acid alignment of *A. niger* phyA, phyB, and derived amino acid sequence from *A. thaliana* gene F21M12.26 (A. Th). The active site motif (RHGXRXP) and the HD motif are underlined. Regions of conserved sequence are highlighted.

identified earlier as the hallmark of histidine acid phosphatase (14).

The sequence of *A. thaliana* gene F21M12.26 was obtained from GenBank, accession number 2160177, from chromosome 1. The sequence was deposited with Genbank as part of the Arabidopsis genome sequencing program. Up until this reporting no function was attributed to the gene product. However, the gene hybridizes to one of the expressed sequence tags or EST prepared for chromosome 1 of *A. thaliana* (15, 16).

When the amino acid sequence of *A. niger* phytase (phyA) was compared with *A. thaliana* F21M12.26 by using ALIGN program Release 1.6c (17), the fungal phyA protein only showed 15.1% sequence homology. Despite the low total level of homology both the proteins shared the common active site sequences and several other conserved regions (Figure 1). We also compared the amino acid sequences of *A. thaliana* gene F21M12.26 with that of *A. niger* pH 2.5 optimum acid phosphatase, which is also known as phyB, with ALIGN program. The results are also shown in Figure 1. Although, both the proteins contain septa-peptide active site belonging to histidine acid phosphatase, the overall sequence homology was only 16.4%.

Histidine-Aspartic (H-D) Acid Motif

In other histidine phosphatases a dipeptidic region containing His-Asp residues was shown to be important for catalysis (18). These two-residues are thought to be involved in donation of a proton to the substrate leaving group during the formation of phosphohistidine (18). Fungal phyA has three “HD” regions, while both fungal phyB, and *A. thaliana* putative acid phosphomonoesterase have one “HD” region (Figure 1).

Conservation of Cysteines

Although cysteines (Cys) are not implicated in having any role in catalysis, they were shown to be very important for folding in *Aspergillus* phytase (13). All ten Cys residues form five disulfide bridges (13, 19). The pairing of Cys 8-17, 48-391, 192-442, 241-259, and 413-421, is probably critical for maintaining the optimal three-dimensional structure of the fungal phytase (13). Of all the five disulfide bridges present in *A. niger* phytase the one which seems to be most crucial is the pairing between Cys 48 and Cys 391. This disulfide bridge brings two β-sheets, A and E, closer together (19). In addition, the septa-peptide active site sequence

is located at the end of β-sheet A. These two Cys were found in the identical position not only in *A. niger* phyA and phyB proteins, but also in the deduced amino acid sequence of *A. thaliana* gene F21M12.26.

The other important finding is that both *A. niger* phyA and phyB proteins contain 10 Cys. The deduced sequence of *A. thaliana* gene F21M12.26 contains 10 Cys. The relative position of these Cys are also maintained throughout the molecule.

Conservation of Histidines

Histidines (His) of acid phosphatases are implicated in catalysis of phosphomonoesters (20). When His residues of fungal phytase were modified by diethyl pyrocarbonate (5 mM), the enzyme lost about 85% activity in 30 minutes at room temperature (21). The fungal phyA protein contains 10 His residues and the putative histidine acid phosphatase from *A. thaliana* has eleven. Moreover, four His residues of both the proteins occupy the same location. The fungal phyB protein, however, only has six His residues and only two of them show sequence homology with fungal phyA and putative *A. thaliana* histidine acid phosphatase.

Total Length of Histidine Acid Phosphatases

The total residues of *A. niger* phyA and phyB proteins, both belonging to histidine acid phosphatase, were reported to be 448 and 450, respectively (8, 10). The deduced sequence of *A. thaliana* gene F21M12.26 codes for a 468 amino acid protein. This size is in agreement with the size expected for other members of histidine acid phosphatase.

N-Glycosylation Sites

A search of the putative phytase encoded by *A. thaliana* revealed no N-glycosylation signal motif (NXS/T). Nine N-glycosylation signal motifs are present in *A. niger* phyA enzyme (8). Chemical sequencing of this enzyme also confirmed the presences of the nine N-glycosylation sites (7).

Table 1 summarizes the sequence and residue similarities among *A. niger* phyA, phyB, and the deduced amino acid sequence of the *A. thaliana* gene. From the active site motif, and Cys residue distributions in the deduced sequence of *A. thaliana* gene F21M12.26, it is indicated that the gene codes for a phyA type of histidine acid phosphate. This is the first documentation of any plant protein sequence that show a significant homology to fungal phyA or phyB protein.

Recently the *phy*A gene from a monocotyledon, maize, was reported (22). The cDNA encoding maize phytase contains an open reading frame of 387 amino acids with five Cys residues. Analysis of the these two plant amino sequences reveal the active site motif RHGXRXP is more conserved in the dicotyledon *Arabidopsis* gene than in the monocotyledon maize seedling phytase phyA. The homology between maize phytase and this putative *Arabidopsis* phytase is limited to the same 33 amino acid region that maize phytase shares with fungal phytase.

TABLE 1

Properties of a Putative *A. thaliana* and Two Fungal Histidine Acid Phosphatases

Parameters	Proteins and deduced protein		
	Arabidopsis thaliana	Fungal phyA	Fungal phyB
Total residues*	464	448	460
Active site motif 'RHGXRXP'	Yes	Yes	Yes
'HD' motif	Yes	Yes	Yes
Cys residue	10	10	10

* Minus signal peptide.

As more genomes are being sequenced and deduced ORF proteins generated, it is becoming increasingly important to identify these new proteins. In this communication we have provided a minimum criteria for a subclass of acid phosphomonoesterase referred to as 'histidine acid phosphatase'.

REFERENCES

1. Ullah, A. H. J., and Phillippy, B. Q. (1994) *J. Agric. Food Chem.* **42,** 423–425.
2. International Union of Biochemistry (1984) Enzyme Nomenclature: Recommendations (1978) of the Nomenclature Committee of the International Union of Biochemistry, pp.282–284. Academic Press, San Diego.
3. Cosgrove, D. J. (1969) *Ann. NY Acad. Sci.* **165,** 677–686.
4. Johnson, L. F., and Tate, M. E. (1969) *Ann. NY Acad. Sci.* **165,** 526–532.
5. Shieh, T. R., and Ware, J. H. (1968) *Appl. Microbiol.* **16,** 1348–1351.
6. Wodzinski, R. J., and Ullah, A. H. J. (1996) *Adv. Appl. Microbiol.* **42,** 263–302.
7. Ullah, A. H. J., and Dischinger, H. C. Jr. (1993) *Biochem Biophys. Res. Commun.* **192,** 747–753.
8. Van Hartingsveldt, W., Van Zeijl, C. M. J., Harteveld, G. M., Gouka, R. J., Suykerbuyk, M. E. G., Luiten, R. G. M., Van Paridon, P. A. , Selton, G. C. M., Veenstra, A. E., Van Gorcom, R. F. M., and Van den Hondel, C. A. M. J. J. (1993) *Gene* **127,** 87–94.
9. Ullah, A. H. J., and Cummins, B. J. (1987) *Prep. Biochem.* **17,** 397–422.
10. Ehrlich, K. C., Montalbano, B. G., Mullaney, E. J., Dischinger, H. C., Jr., and Ullah, A. H. J. (1993) *Biochem. Biophys. Res. Commun.* **195,** 53–57.
11. Ullah, A. H. J., Cummins, B. J., and Dischinger, H. C., Jr. (1991) *Biochem. Biophys. Res. Commun.* **178,** 45–53.
12. Ullah, A. H. J., and Dischinger, H. C., Jr. (1993) *Biochem Biophys. Res. Commun.* **192,** 754–759

13. Ullah, A. H. J., and Mullaney, E. J. (1996) *Biochem Biophys. Res. Commun* **227,** 311–317.
14. Ullah, A. H. J., and Dischinger, H. C., Jr. (1995) *Ann. NY Acad. Sci* **750,** 51–57.
15. Hayashida, N., Sumi, Y., Wada, T., Handa, H., and Shinozaki, K. (1995) *Gene* **165,** 155–161.
16. Rounsley, S. D., Glodek, A., Sutton, G., Adams, M. D., Somerville, C. R., Venter, J. C., and Kerlavage, A. R. (1996) *Plant Physiol* **112,** 1177–1183.
17. Pearson, W. R. (1990) *Methods Enzymol.* **183,** 63–98.
18. Ostanin, K., Harms, E., Stevis, P. E., Kuciel, R., Zhou, M., and Van Etten, R. L. (1992) *J. Biol. Chem.* **267,** 22830–22836.
19. Kostrewa, D., Gruninger-Leitch, F., D'Arcy, A., Broger, C., Mitchell, D., and van Loon, A. P. (1997) *Nat. Struct Biol.* **4,** 185–190.
20. Van Etten., R. L. (1982) *Ann. N. Y. Acad. Sci.* **390,** 27–51.
21. Ullah, A. H. J., and Dischinger, H. C. Jr. (1992) *Ann. NY Acad. Sci.* **672,** 45–51.
22. Maugenest, S., Martinez, I., and Lescure, A. M. (1997) *Biochem J.* **332,** 511–517.