Vol. 195, No. 1, 1993 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 31, 1993 Pages 53-57

Identification and cloning of a second phytase gene (*phyB*) from *Aspergillus niger (ficuum)*

Kenneth C. Ehrlich, Beverly G. Montalbano, Edward J. Mullaney,
H. Charles Dischinger, Jr., and Abul H.J. Ullah

Southern Regional Research Center, United States Department of Agriculture, New Orleans,LA 70179

Received July 6, 1993

SUMMARY: An *Aspergillus niger (ficuum)* genomic DNA λEMBL3 library was probed with a 354-bp DNA fragment obtained by polymerase chain reaction of *A. niger* DNA with oligonucleotides based on partial amino acid sequence of a pH 2.5 optimum acid phosphatase. A clone containing a 1605 bp segment (*phyB*) encoding the 479 amino acid enzyme was isolated and found to contain four exons. Global alignment revealed 23.5% homology to *Aspergillus niger* phytase (PhyA); four regions of extensive homology were identified. Some of these regions may contain catalytic sites for phosphatase function.

Phytase, *myo*-inositol hexakisphosphate phosphohydrolase (E.C. 3.1.3.8) is an acid phosphatase capable of hydrolyzing phytic acid (*myo*-inositol hexakisphosphate) as well as a number of other organophosphate substrates (1). There has been much interest in the use of fungal phytase to increase the bioavailability of phosphate in plant seed meals intended as feed for monogastric animals (2,3). A phytase (PhyA) with a pH optimum of 5.0 has been isolated from the culture filtrate of *Aspergillus niger* NRRL 3135, biochemically characterized (1), partially sequenced by chemical sequencing (4), and the gene for the enzyme cloned (5,6). A second phosphatase with pH optimum 2.5 was also isolated from the same culture filtrate, and was found to have different properties from those of PhyA (7). This second enzyme was previously only characterized as an acid phosphatase, but recently has been found to hydrolyze phytic acid at pH 2.5 (but not at pH 5), and therefore is a second phytase (PhyB) from this strain of *Aspergillus* (Ullah and Phillippy, unpublished results). In this communication we report the cloning and sequencing of the gene for this enzyme.

METHODS

Protein purification and sequence determination

PhyB was purified by soft-gel chromatography as previously described (7) followed by reverse-phase C-8 HPLC. Primary structures of peptides obtained by cyanogen bromide, endoproteinase Glu-C, and clostripain digestion were determined using a Porton 2090 gas phase protein sequencer as previously described (8).

Abbreviations:
Phy-phytase, PCR-polymerase chain reaction, HPLC-high performance liquid chromatography.

0006-291X/93 $4.00

Gene cloning and DNA sequencing

Non-degenerate oligonucleotides (809 and 810) were synthesized, taking into account the expected codon bias for highly expressed genes in *Aspergillus spp.* (9). These oligonucleotides corresponded to the N-terminal amino acid sequence, FSYGAAI, of the purified protein and an internal peptide, MGGHLTIE (Figure 1). They were used in a PCR reaction with Taq polymerase (Amplitaq from Perkin-Elmer, Norwalk, CT) to amplify, from *A. niger* DNA, a 1258-bp portion of the *phyB* gene. A second set of oligonucleotides (nested primers 822 and 828) was used to amplify a portion of the gene obtained in the first PCR reaction. The main PCR product (354 bp) in the second amplification reaction was radiolabeled and used as a hybridization probe to screen a λEMBL3 library of *A. niger* DNA. One clone (λAN1) was selected, which, after partial sequencing, contained an open-reading frame whose deduced protein sequence was the same as that found for PhyB by chemical sequencing (8). In order to determine intron-exon boundaries, and the 3'- and 5'- termini of the gene, messenger RNA was isolated from *A. niger* mycelia grown on starch medium capable of inducing production of high levels of phytase (1). The mRNA was used to prepare first-strand cDNA following the method described in a kit from Pharmacia (Piscataway, NJ) in which the mRNA was primed with oligo dT as well as oligonucleotide primers adjacent to intron-exon boundaries.

Sequencing was performed by subcloning the 1258-bp PCR product into pCRII (Invitrogen, San Diego, CA). Following this, complete sequencing of the double-strand genomic fragment in both directions was accomplished using Sequenase Version 2 (U.S. Biochemical Corp., Cleveland, OH) starting with M13 universal primer and -40 reverse primer (10) followed by gene-walking using specific oligonucleotides. A clone containing the N-terminal portion of the gene was prepared by subcloning into pBluescript (Stratagene, La Jolla, CA) a *Sph*I/*Eco*RI fragment of λAN1 DNA. The C-terminal portion of the gene was determined from a pBluescript clone containing a *Bam*HI fragment of λAN1 DNA. Both of these clones were sequenced using specific oligonucleotide primers. Sequencing of first-strand cDNA used the fmol sequencing kit of Promega (Madison, WI) and ^{32}P-labeled oligonucleotides chosen to read sequence across the intron-exon boundaries.

RESULTS AND DISCUSSION

The open reading frame of the *phyB* gene encodes a 479 amino acid polypeptide (Figure 1); correspondence between DNA-deduced primary structure and chemically derived protein sequence is shown in bold type. The calculated molecular weight of the unglycosylated molecule is 52601 daltons. At least eight possible glycosylation sites are present in the enzyme. The mature protein was unblocked and, by sequencing the N-terminus, began at FSYGAA. A leader peptide sequence of approx. 20 amino acids is consistent with the signal peptide found in PhyA (6).

The ALIGN program of Pearson and Lipman (11) was used to compare the sequence of this phosphatase to that of PhyA (6); alignment revealed 23.5 % identical amino acids and 46% conservative substitutions (Figure 2). Four regions share a particularly high degree of homology (shown in boxes in Figure 2). Two regions correspond to previously characterized domains believed to be necessary for phosphatase catalytic activity (12,13). The first, beginning at amino acid 81, contains the motif RHG*X*R*X*P and is believed to be the phosphate acceptor region while the second, at amino acid 336, contains an "HD" dipeptide thought to be involved in donation of a proton to the substrate leaving group during the formation of the phosphohistidine (12). PhyA has three "HD" regions (at amino acids 287, 333, and 361) that may be involved in catalysis, while PhyB has only one site (at amino acid 337). How these and other differences in the structure of the two similar phytases from the same organism relate to catalytic activity will be the subject of further studies by site-directed mutagenesis.

```
-134 GCATGCTGGACCGCAATCTCCGATCGCCGGGTATAAAAGGTCCTCCAAACCCCTCTCGGTCGATATGTACCCCGCTCGTC
 -54 ATCTCCAATCCTCTCGAGAGCACCTTCTCCAGCTTTTGTCAATTGTACCTTGCAATGCCTCGCACCTCTCTCCTCACCCT
   1                                                          M  P  R  T  S  L  L  T  L
                                                ----------809------►-822--►
  27 GGCCTGTGCTCTGGCCACGGGCGCATCCGCTTTCTCCTACGGCGCTGCCATTCCTCAGTCGACCCAGGAGAAGCAGTTCT
  10  A  C  A  L  A  T  G  A  S  A  F  S  Y  G  A  A  I  P  Q  S  T  Q  E  K  Q  F

 107 CTCAGGAGTTCCGAGATGGCTACAGCATCTTGAAGCACTACGGTGGTAACGGACCCTACTCCGAGCGTGTGTCCTACGGT
  36 S  Q  E  F  R  D  G  Y  S  I  L  K  H  Y  G  G  N  G  P  Y  S  E  R  V  S  Y  G

 187 ATCGCTCGCGATCCCCCGACCGGCTGCGAGGTCGATCAGGTCATCATGGTCAAGCGTCACGGAGAGCGCTACCCGTCCCC
  63  I  A  R  D  P  P  T  G  C  E  V  D  Q  V  I  M  V  K  R  H  G  E  R  Y  P  S  P

 267 TTCAGCCGGCAAGAGCATCGAAGAGGCCCTGGCCAAGGTCTACAGCATCAACACTACTGAATACAAGGGCGACCTGGCCT
  90  S  A  G  K  S  I  E  E  A  L  A  K  V  Y  S  I  N  T  T  E  Y  K  G  D  L  A
                                                                 ◄--------828-----
 347 TCCTGAACGACTGGACCTACTACGTCCCTAATGAGTGCTACTACAACGCCGAGACCACCAGCGGTCCCTACGCCGGTTTG
 116 F  L  N  D  W  T  Y  Y  V  P  N  E  C  Y  Y  N  A  E  T  T  S  G  P  Y  A  G  L

 427 CTGGACGCGTACAACCATGGCAACGATTACAAGGCTCGCTACGGCCACCTCTGGAACGGCGAGACGGTCGTGCCCTTCTT
 143  L  D  A  Y  N  H  G  N  D  Y  K  A  R  Y  G  H  L  W  N  G  E  T  V  V  P  F  F

 507 TTCTAGTGGCTACGGACGTGTCATCGAGACGGCCCGCAAGTTCGGTGAGGGTTTCTTTGGCTACAACTACTCCACCAACG
 170  S  S  G  Y  G  R  V  I  E  T  A  R  K  F  G  E  G  F  F  G  Y  N  Y  S  T  N

 587 CTGCCCTCAACATCATCTCCGAGTCCGAGGTCATGGGCGCGGACAGCCTCACGCCCACCTGTGACACCGACAACGACCAG
 196 A  A  L  N  I  I  S  E  S  E  V  M  G  A  D  S  L  T  P  T  C  D  T  D  N  D  Q

 667 ACCACCTGCGATAACCTGACCTACCAGCTGCCCCAGTTCAAGGTCGCTGCTGCCCGCCTAAACTCCCAGAACCCTGGCAT
 223  T  T  C  D  N  L  T  Y  Q  L  P  Q  F  K  V  A  A  A  R  L  N  S  Q  N  P  G  M

 747 GAACCTCACCGCATCTGATGTCTACAACCTGATGGgtatgtgattacggaacaatcattggctcaaacatccagctgaca
 250  N  L  T  A  S  D  V  Y  N  L  M►

 827 gcatcccagTTATGGCTTCCTTTGAGCTCAATGCTCGTCCCTTCTCTAACTGGATCAACGCCTTTACCCAGGACGAATGG
 267         V  M  A  S  F  E  L  N  A  R  P  F  S  N  W  I  N  A  F  T  Q  D  E  W

 907 GTCAGCTTCGGCTACGTTGAGGATTTGAACTACTACTACTGCGCTGGgtgagtttaccatttgatcgattattgtctcgg
 285  V  S  F  G  Y  Y  V  E  D  L  N  Y  Y  Y  C  A  G►

 987 atcagcaactaacgatcgacagTCCCGGTGACAAGAACATGGCTGCTGTGGGTGCCGTCTACGCCAACGCCAGTCTTACC
 301                          P  G  D  K  N  M  A  A  V  G  A  V  Y  A  N  A  S  L  T

1067 CTCCTGAACCAAGGACCCAAGGAAGCCGGCCCCTTGTTCTTCAACTTgtacgttctcggcaggatcaaagtctcacaaaa
 320  L  L  N  Q  G  P  K  E  A  G  P  L  F  F  N  L►

1147 agaaactcttcactaacatagagtagTGCCCACGACACCAACATCACCCCCATCCTCGCCGCCCTAGGCGTCCTCATCCC
 336                              A  H  D  T  N  I  T  P  I  L  A  A  L  G  V  L  I  P
                                                                    ◄-----810---
1227 CAACGAGGACCTGCCTCTTGACCGCGTCGCCTTCGGCAACCCCTACTCGATCGGCAACATCGTGCCCATGGGTGGCCATC
 354  N  E  D  L  P  L  D  R  V  A  F  G  N  P  Y  S  I  G  N  I  V  P  M  G  G  H
      -------
1307 TGACCATCGAGCGTCTCAGCTGCCAGGCCACCGCTCTCTCGGACAAGGGTACCTACGTGCGTCTGGTGCTGAACGAGGCC
 380 L  T  I  E  R  L  S  C  Q  A  T  A  L  S  D  K  G  T  Y  V  R  L  V  L  N  E  A

1387 GTGCTCCCCTTCAACGACTGTACGTCCGGACCGGGCTACTCCTGCCCTCTGGCCAACTACACCTCCATCCTGAACAAGAA
 407  V  L  P  F  N  D  C  T  S  G  P  G  Y  S  C  P  L  A  N  Y  T  S  I  L  N  K  N

1467 TCTGCCCGACTACACGACCACCTGCAACGTGTCTGCGTCCTACCCGCAGTATCTGAGCTTCTGGTGGAACTACAACACCA
 434  L  P  D  Y  T  T  T  C  N  V  S  A  S  Y  P  Q  Y  L  S  F  W  W  N  Y  N  T

1547 CGACGGAGCTGAACTATCGCTCTAGCCCTATTGCGTGCCAGGAGGGTGATGCTATGGACTAGATACAGAGGGGTAGAACC
 460 T  T  E  L  N  Y  R  S  S  P  I  A  C  Q  E  G  D  A  M  D  *

1627 GGGATACTTTAGTGATGATTGATATTCAAGTTTGGGTGGTGGCGGTCACCTTGTTAATAGTCTTGTACAGTCATGCGGTG
1707 AATATAAATAATGGATAAGAC
```

Figure 1. The nucleotide and deduced amino acid sequence of the *A. niger phyB* gene (GenBank Accession No. L20567). Amino acid sequences that were confirmed by direct amino acid sequencing are shown in bold type. Putative N-glycosylation sites are underlined. Locations for the oligonucleotide primers used to obtain a PCR fragment from genomic DNA are shown by dashed lines. Locations of intron-exon boundaries (the intron is shown is lower-case type) were confirmed by sequencing first strand cDNA prepared from oligo-dT primed *A. niger* mRNA. The location of the probable polyadenylation signal is doubly underlined.

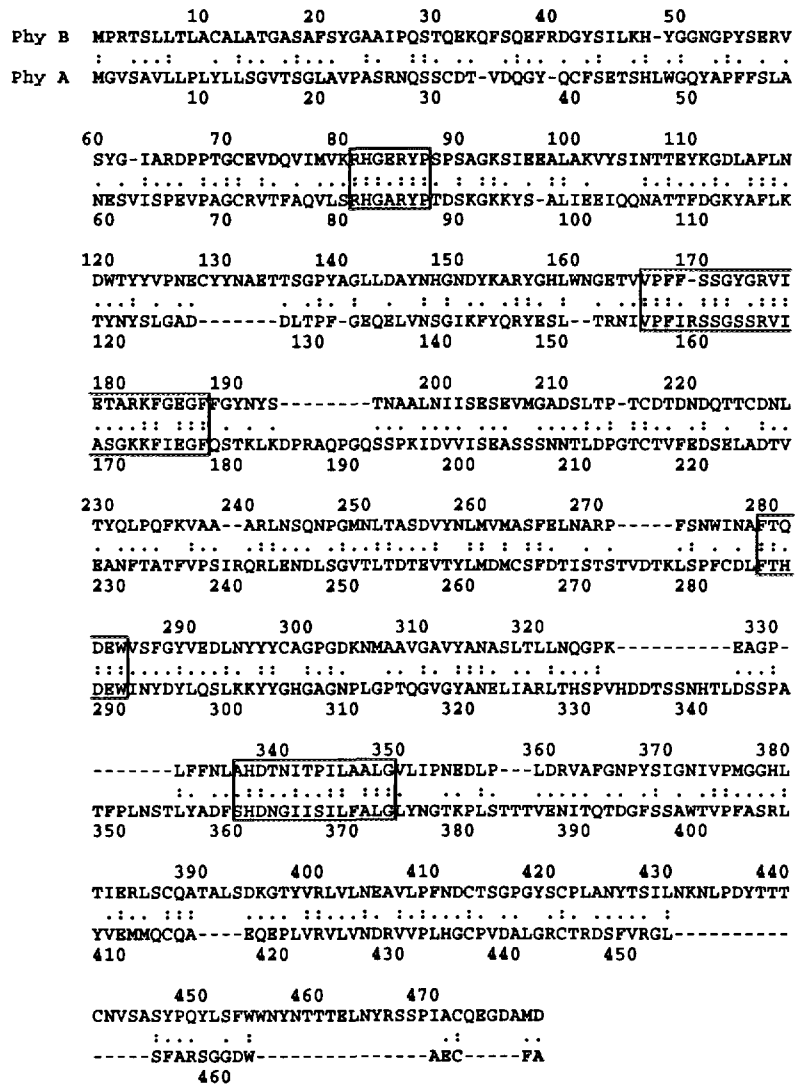

Figure 2. Alignment of the PhyB with the PhyA protein sequence. Alignment was accomplished by the method of Pearson and Lipman (11) using the program ALIGN. Amino acid homology is indicated by a colon while conservative substitutions are identified by a period. Regions with a high degree of homology, some of which correspond to putative catalytic domains previously characterized (12) in the *E. coli* AppA protein, are boxed. Note that PhyA has three domains with the "HD" amino acid motif while PhyB has only one.

REFERENCES

1. Ullah, A.H.J. and Gibson, D.M. (1987) Prep. Biochem. **17**:63-71.
2. Zyta, K. (1992) World J. Microbiol. Biotechnol. **8**:467-472.
3. Cromwell, G.L. (1991) Feedstuffs **62**:14-16.
4. Ullah, A.H.J., Cummins, B.J., Dischinger, H.C. (1991) Biochem. Biophys. Res. Commun. **178**:45-53.
5. Mullaney, E.J., Gibson, D.M., Ullah, A.H.J. (1991) Appl. Microbiol. Biotechnol. **35**:611-614.
6. 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., van den Hondel, C.A.M.J.J. (1993) Gene **127**:87-94.

7. Ullah, A.H.J. and Cummins, B.J. (1987) Prep. Biochem. **17**:397-422.
8. Ullah, A.H.J. and Dischinger, Jr., H.C. (1993) Biochem. Biophys. Res. Commun. **192**:754-759.
9. Lloyd, A.T. and Sharp, P.M. (1991) Mol. Gen. Genet. **230**:288-294.
10. Maniatis, T., Fritsch, E.F., Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, New York.
11. Pearson, W.R., and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. **85**:2444-2448.
12. Ostanin, K., Harms, E., Stevis, P.E., Kuciel, R., Zhou, M-M., and van Etten R.L. (1992) J. Biol. Chem. **267**:22830-22836.
13. Bazan, J.F., Fletterick, R.J., Pilkis, S.J. (1989) Proc. Natl. Acad. Sci. USA, **86**:9642-9646.