Cloning and characterization of a cDNA encoding a maize seedling phytase

Sébastien MAUGENEST†, Isabelle MARTINEZ‡ and Anne-Marie LESCURE†§

†Laboratoire de Biologie des Semences, INRA INA-PG, F-78026 Versailles Cédex, France, and ‡BIOCEM-Groupe Limagrain, 24 Avenue des Landais, 63170 Aubière Cédex, France

During germination, maize seedlings express a phytase able to hydrolyse the large amount of phytin stored in the dry seed. Previous studies allowed purification and characterization of this enzyme as a homodimer of 38 kDa subunits [Labouré, Gagnon and Lescure, Biochem. J. (1993) **295**, 413–419]. In the present work, an antibody against the purified maize phytase has been used to screen a maize seedling cDNA expression library. Several positive clones containing an insert of about 1400 bp were isolated. The nucleotide sequence of the insert of one of these clones has been established. This cDNA, called phy S11, was 1335 bp long and contained an open reading frame of 387 amino acids. The sequence of N-terminal residues (23 amino acids) of the purified phytase has been established. These residues are found at positions 19–41 of the amino acid sequence encoded by phy S11. This confirms that this cDNA codes for the maize phytase. The deduced amino acid sequence appears to be very different from those of published *Aspergillus niger* phytases; however, an homologous region of 33 amino acids was detected. This region of the fungal sequence contains the RHGxRxP consensus motif found in various high molecular mass acid phosphatases and believed to be the acceptor site for phosphate. Expression of the phy S11 cDNA in *Escherichia coli* allowed the production of the phytase subunit and its assembly to give a protein of the same size as the native phytase. The time course of phy S11 mRNA accumulation during germination showed that no transcript was present in dry seeds. The mRNA accumulated during the first day of germination, to reach a maximum after 2 days (radicle protrusion), and then decreased in young seedlings. Genomic Southern blot analyses suggest the existence of at least two genes and genetic mapping reveals two loci separated by 1 cM on chromosome 3 of maize. The cloning of this first cDNA coding for a plant phytase, will allow the isolation of the corresponding genes and the study of their regulation during germination.

INTRODUCTION

Most of the phosphorus present in seeds is in the form of phytin, a complex salt of *myo*-inositol hexaphosphoric acid (phytic acid). Phytin is therefore a store for phosphorus, carbohydrate and a variety of cations. In maize grains, phytate phosphorus represents up to 88 % of total phosphorus [1]. Mobilization of phytin is due to phytases, a special type of phosphatase, that are able to hydrolyse phosphate from phytic acid as well as from other phosphorylated substrates [2]. Phytase activities have been reported in a wide range of seeds. Generally, very low amounts of endogenous activity can be detected in ungerminated seeds, except in wheat, rye and their hybrid, triticale [3]. A large increase in activity concomitant with a decrease in phytin occurs at the beginning of germination. Little information is available about the mechanisms that control the level of phytase activity in seeds.

The primary interest in phytase comes from the food industry. Phytin from dry seed meals is not digested by monogastric animals in the absence of exogenous phytase and is considered to be an anti-nutritional factor [4]. Extensive research has been undertaken in order to allow phytate digestion by adding exogenous phytases to such meals. Plant phytases are normally produced in amounts insufficient for their use in industrial processes. Extracellular phytases produced by the fungus *Aspergillus niger* may be obtained in large amounts and have been found to be well suited for use as an animal feed additive [5,6]. Recently, cDNAs coding for two different phytases, Phy A and Phy B, from *A. niger* have been sequenced [7,8] and transgenic tobacco plants expressing ectopically the fungal phytase Phy A have been obtained [9,10]. These authors envision using the same genetic engineering approach to produce mature seeds for animal feed containing large amounts of active *Aspergillus* phytase.

So far, no cDNA coding for a plant phytase has been reported and the molecular mechanisms that control the regulation of phytase expression during seed formation or germination are not known. Because of the importance of phytic acid as a storage form of phosphate in maize seeds, a better knowledge of these mechanisms would be of fundamental and applied interest.

In a previous paper [11], we have reported the purification and characterization of a phytase which accumulated in maize seedlings during the first days of germination. The native protein appeared to be a homodimer with 38 kDa subunits. This enzyme had most of the enzymic characteristics of other plant acid phytases. A rabbit antiserum was raised against a homogeneous maize phytase preparation. Western blot analyses have shown that the increase of phytase activity observed during the first days of maize seed germination corresponded to synthesis of the enzyme *de novo*.

We report here the isolation and characterization of a cDNA coding for the subunit of this maize seed phytase. This constitutes the first nucleotide sequence for a plant phytase.

MATERIALS AND METHODS

Maize seed germination conditions

Two homozygous lines of *Zea mays* were successively utilized for this work. The cDNA library was constructed with RNAs extracted from *Z. mays* c.v. MO17 (Maïsadour, Mont-de-Marsan, France). The RNA utilized for the Northern blots was

Abbreviations used: SSD, single-seed-descent; RFLP, restriction fragment length polymorphism; IPTG, isopropyl β-D-thiogalactoside.

The nucleotide sequence reported will appear in the GenBank/EMBL Data Bank under accession number U75531.

§ To whom correspondence should be addressed.

extracted from *Z. mays* c.v. AMO406 (Maïsadour). Maize seeds were soaked for 12 h in aerated distilled water at room temperature and then transferred into wet sand at 26 °C, with a 16 h photoperiod (50 μE/s per m^2).

cDNA library construction and screening

Total RNA was purified from frozen embryos (scutellum and embryonic axis) of 3- to 4-day-old maize seedlings, using guanidine hydrochloride [12]. The poly(A)$^+$ fraction was enriched using oligo(dT) spin columns (Pharmacia). Double-stranded cDNA was synthesized using a cDNA kit (Promega) and an oligo(dT)-*Not*I primer adaptor. After the ligation of *Eco*RI adaptators (Pharmacia), the cDNA was digested with *Not*I and directionally cloned into λgt11 (Sfi-Not) (Promega). About 1×10^6 independent recombinants were obtained after packaging (Stratagene) and infection of *Escherichia coli* strain LE 392 (Promega). The λgt11 cDNA library was amplified in *E. coli* RY1090 strain. Immunological screening of phage plaques was performed using the ProtoBlot Immunoscreening System from Promega according to the supplier's protocol, with a crude rabbit antiserum directed against the maize purified phytase [11]. After dilution (1:1000 for the first screening and 1:5000 for the following screening steps), the antiserum was treated with *E. coli* extracts as indicated in the Promega guideline. Several positive recombinant phages were isolated after screening 1.5×10^5 plaques. These positive phages were then submitted to a second screen with a more specific antibody purified by affinity on the phytase subunit according the procedure described by Lin et al. [13]. The size of the inserts was estimated by PCR reaction with primers U5 (5′-AACAGCTATGACCATG-3′) and U3 (5′-GTAAAACGAACGGCCAGT-3′), using amplification conditions described by Marin et al. [14]. The inserts of the positive phages were then subcloned in the *Eco*RI/*Not*I sites of pBS-SK+ (Stratagene). The nucleotide sequences were determined on double-stranded DNA using the PRISM Ready Reaction Kit (Applied Biosystems), following the instructions of the manufacturer and using a Perkin Elmer Cetus DNA thermal cycler PEC480 and an ABI 373A sequencer (Applied Biosystems).

Phage plaque hybridizations with the partial ^{32}P-labelled phy AM10 cDNA as probe, in order to isolate full-length cDNAs, were performed on 3.6×10^5 plaques according to the technique of Sambrook et al. [15] with the following modifications: the filter was immersed, DNA-side up, in the denaturing solution for 10 min and then for 6 min in the neutralizing solution.

The total sequence analysis of the 1.4 kb insert phy S11 was carried out on both strands by the dideoxy chain termination method [16], using M13 mp18 and mp19 single-stranded templates (the high content of GC in this insert did not allow its sequencing on double-stranded DNA).

Sequence analyses were performed using the UWGCG package (Program Manual for the Wisconsin Package, University of Wisconsin Genetics Computer Group, Madison, WI, U.S.A.).

RNA gel blot analyses

Total RNA was extracted using guanidine hydrochloride [12] from frozen embryos at various stages of germination. Equal amounts of RNA were size fractionated on a 1.2% (w/v) agarose gel containing formaldehyde [15] and transferred to a Hybond N membrane (Amersham). The *Eco*RI/*Not*I phy S11 insert labelled with [α-^{32}P]dCTP by the random priming technique, was used as a probe. Prehybridizations and hybridizations were performed at 42 °C in: 50% (v/v) formamide/5 × Denhardt/5 × SSPE (0.15 M NaCl/10 mM sodium phosphate, pH 7.4/1 mM EDTA)/0.5% (w/v) Sarkosyl/salmon sperm DNA (100 μg/ml). The membranes were washed twice at 42 °C with 2 × SSC (0.15 M NaCl/0.015 M sodium citrate)/0.1% (w/v) SDS and once at 60 °C for 10 to 30 min with 0.2 × SSC/0.1% SDS. The blots were exposed at −80 °C to Amersham MP films between intensifying screens.

Southern blot analysis and genetic mapping

DNA used for Southern blot analysis was isolated from seedlings of *Z. mays* c.v. B73 using the procedure described by Rogers and Bendich [17]. After digestion by appropriate enzymes, DNA fragments were separated on agarose gel, denatured and blotted onto Hybond N+ membranes. The labelled *Eco*RI/*Not*I phy S11 insert was used as a probe. Prehybridization and hybridization were carried out in a buffer containing 6 × SSC/0.5% SDS/5 × Denhardt/denatured herring sperm DNA (25 μg/ml). The blots were prehybridized for 6 h and hybridized with denaturated, labelled probe overnight at 68 °C. They were subsequently washed once in 2 × SSC, once in 2 × SSC/0.1% SDS and finally in 0.1 × SSC/0.1% SDS. Autoradiography was as described for Northern blots.

For genetic mapping, 58 single-seed-descent (SSD) maize lines derived from the cross A188 × D7 were used as the mapping population. The SSD population was produced from the A188 × HD7 hybrid by successive selfing, up to the F6 generation [18]. A188 is a North American public line related to the Minnesota group and HD7 is a doubled haploid line derived from synthetic population with Chinese geographical origin [19]. Hybridization was performed with ^{32}P-labelled phy S11 cDNA as a probe on blots containing DNA from SSD lines, digested with *Hin*dIII or *Eco*RI. DNA extraction, Southern blotting and probe hybridization were as described above. Linkage analysis was done by using Mapmaker, version 2.0, computer program for Macintosh [20] and map distances were calculated with the Kosambi function.

Expression of the phy S11 encoded protein in *E. coli*

A 1150 bp *Nco*I–*Ase*I fragment of the cDNA phy S11 (see Figure 1) was cloned between the *Nco*I and *Nde*I sites of the expression vector pET14(b) (Novagen). Expression of the protein was obtained by cloning the recombinant vector in *E. coli* strain BL21 (Novagen).

Extraction of proteins under denaturing or non-denaturing conditions was performed according to the procedures described by the supplier. Electrophoresis under denaturating conditions was performed on a 12.5% (w/v) polyacrylamide/SDS gel. Estimation of the molecular mass of native protein in *E. coli* was done using a 4–20% (w/v) polyacrylamide gradient gel run at 100 V for 24 h as previously reported [11]. Western blot analyses were conducted with the affinity-purified antibody, using anti-rabbit IgG immunoglobulins coupled with peroxidase (Biosys). The detection was performed with the chemiluminescent ECL detection system (Amersham). Phosphatase activity on non-denaturing polyacrylamide gels was detected as previously described [11].

N-terminal sequence determination

Purified maize phytase was electrophoresed by SDS/PAGE (12.5% gel) and electro-transferred onto a PVDF membrane 'Problott' (Applied Biosystems), followed by Amido Black staining to reveal the 38 kDa subunit. Automated Edman degradation of the protein was performed with an Applied Biosystems

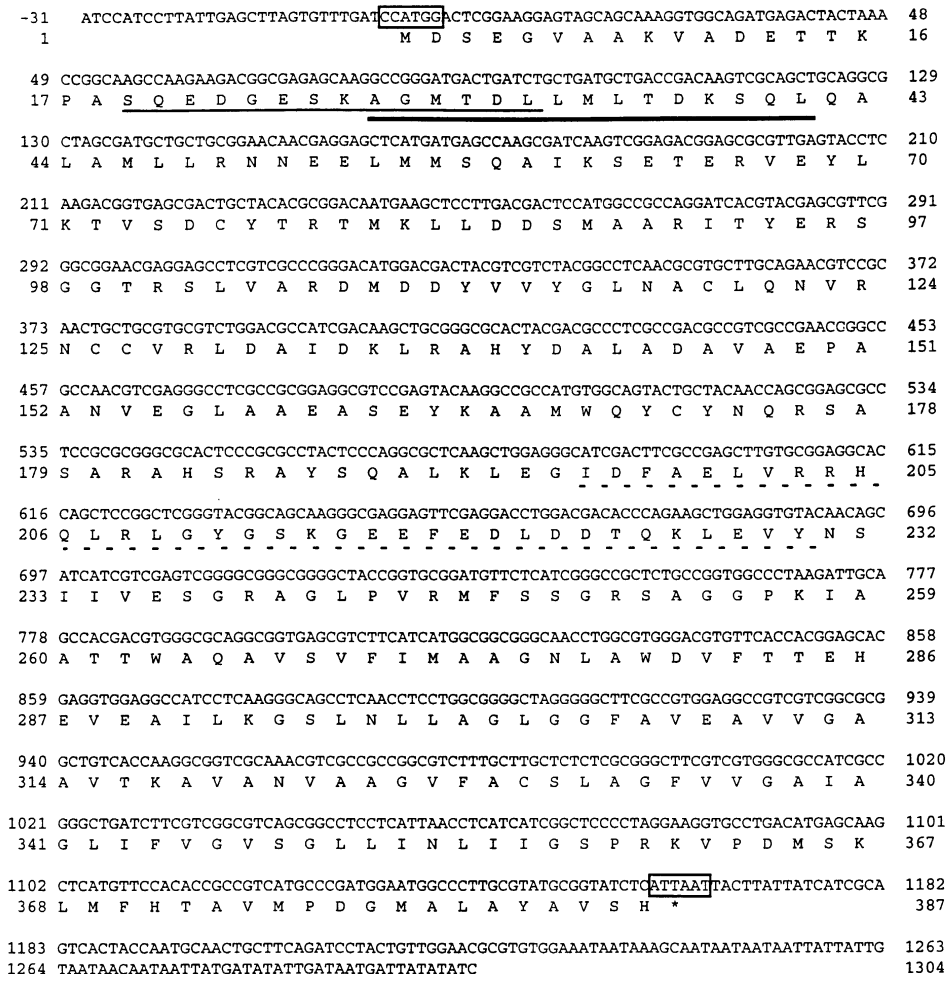

Figure 1 Nucleotide and deduced amino acid sequences of the phy S11 cDNA

Underlined amino acids correspond to the two N-terminal sequences obtained from the purified maize phytase. The amino acid sequence homologous to the 33 amino acid sequence of *A. niger* Phy A phytase [7] is underlined with a broken line. The *Nco*I (CCATGG) and *Ase*I (ATTAAT) sites utilized for cloning in the expression vector pET 14(b) are boxed.

473 A sequencer using the reagents and the methods of the manufacturer.

RESULTS

Isolation and characterization of a cDNA encoding a maize seedling phytase

A cDNA library was constructed in the expression vector λgt11 (Sfi-Not) using cDNA synthesized from polyadenylated mRNA isolated from 3–4-day-old maize seedlings. The initial screening with the crude antiserum raised against maize phytase resulted in the isolation of 12 positive clones. Among these clones, nine were found to contain the same insert of 500 bp and three other clones named pAM-10, -15 and -16, contained distinct inserts of respectively 1000, 600 and 1200 bp. When tested by Northern blots on total mRNA from 3–4-day-old maize seedlings, no hybridization was observed with the insert of 500 bp, which appeared to be an artifact of cloning. The inserts of pAM10, pAM15 and pAM16 were found to hybridize respectively with mRNAs of about 1400, 1300 and 1800 bp. We have previously found [11] that the crude antiserum was not very specific and cross-reacted with at least four peptides besides the phytase subunit. When pAM-10, -15 and -16 were tested with the specific antibody, obtained by affinity purification of the phytase subunit, only clone pAM10 gave a positive response. The corresponding insert was subcloned in pBS-SK+ and sequenced. This cDNA, designated phy AM10, contained 930 bp with an incomplete open reading frame coding for 139 amino acids. Labelled phy AM10 cDNA was then used as a probe to screen the same library in order to seek full-length clones. Several positive clones containing an insert of about 1400 bp were isolated. The insert of one of these clones, designated phy S11 cDNA, was entirely sequenced. The nucleotide sequence of phy S11 cDNA contains 1335 bp with a high GC content (65%) and exhibits an open reading frame corresponding to a 387 amino acid polypeptide (Figure 1). The presence of a stop codon (TGA) two codons upstream of the first ATG suggests that this ATG is the initiation codon and thus that the coding sequence is complete. This cDNA presents 31 bp of 5′- and 147 bp of 3′-untranslated sequences. A putative polyadenylation signal identical with the animal consensus sequence AATAAA is observed 63 nucleotides upstream from the poly(A) tail (position 1236–1241) (Figure 1). The consensus distance between polyadenylation signals and polyadenylation sites is 27 ± 9 nucleotides [21]. Therefore this consensus hexamer is probably not the polyadenylation signal used for the phy S11

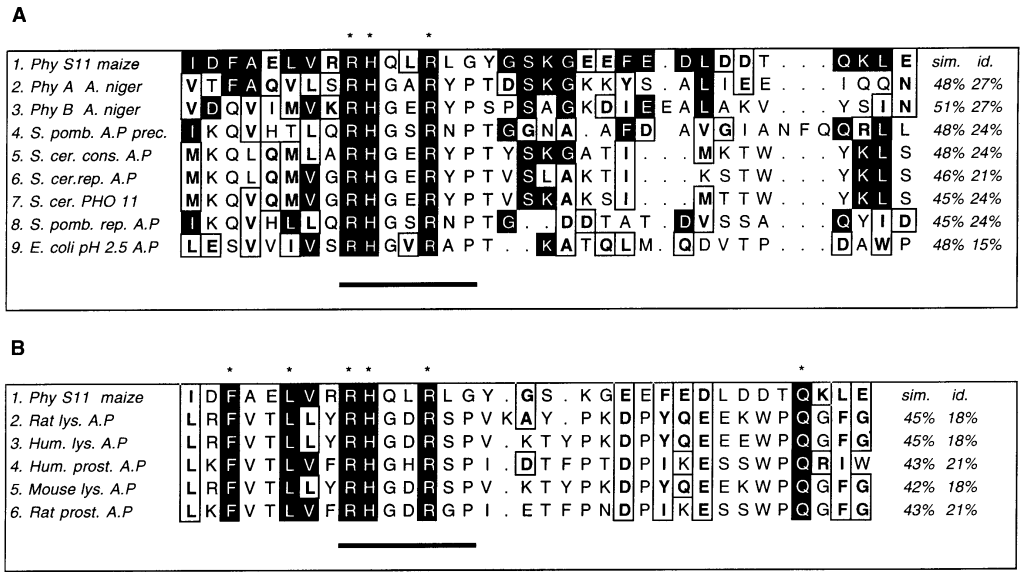

Figure 2 Comparison of the 33 amino acid homologous region of the protein sequence deduced from phy S11 cDNA and those of published high molecular mass acid phosphatase (A.P) sequences

(**A**) 1, *Z. mays* phytase, the present study; 2, *A. niger* phytase A [7]; 3, *A. niger* phytase B [8]; 4, *Schizosaccharomyces pombe* A.P precursor [33]; 5, *Saccharomyces cerevisiae* constitutive A.P [34]; 6, *S. cerevisiae* repressible A.P [35]; 7, *S. cerevisiae Pho*11 A.P [36]; 8, *S. pombe* repressible A.P [37]; 9, *E. coli* pH 2.5 A.P [38]. (**B**) 1, Maize Phy S11; 2, rat lysosomal A.P [39]; 3, human lysosomal A.P [31]; 4, human prostatic A.P [40]; 5, mouse lysosomal A.P [41]; 6, rat prostatic A.P [42]. Acid phosphatases were searched in the NCBI protein database using Nentrez software. Sequences were compared using the Bestfit program of GCG. Identical residues are shown in black and white and conserved residues are in boxes. Asterisks correspond to residues that are conserved in all the sequences presented in alignment. Percentages of similarity or of identity are given by the Bestfit comparison. The RHGxRxP motif is underlined.

cDNA. It is known that in plants polyadenylation can occur at multiple sites [22]. Other putative signals AATAAT are detected in the 3′-untranslated sequence of phy S11. The sequence located 29 nucleotides upstream from the poly(A) tail (position 1271–1276) may constitute the signal used for this cDNA polyadenylation.

Comparison of the nucleotide sequences of phy AM10 and phy S11 cDNAs showed that the two sequences were homologous (95% identity) but not identical. Some punctual differences appeared essentially localized in the 3′-untranslated sequences suggesting that the two cDNAs could derive from two different genes.

The size of the encoded protein deduced from the amino acid sequence is 41 kDa, which corresponds to the size of the phytase subunit detected by Western blot analysis of the total protein extracted from young seedlings [11]. The subunit of the phytase purified according the procedure previously described [11] is always slightly smaller (about 38 kDa), suggesting that specific cleavage occurs during purification (see also Figure 5, lanes 6 and 7). The N-terminal sequence was determined on the purified subunit. The 38 kDa peptide appears to be a mixture of two peptides: a minor one with an N-terminus XXEDGESKAGMTDL, and a major one beginning with the amino acid sequence AGMTDLLMLTDKSQL. These two sequences, which partially overlap, are localized respectively at positions 19–32 and 27–41 on the amino acid sequence deduced from the phy S11 cDNA (Figure 1), confirming that this cDNA does encode the phytase subunit and that the first ATG is the initiation codon.

A computer search for similarities between the deduced amino acid sequence and other proteins in several databases, using the FastA program of the UWGCG package, has shown 96% identity with a 162 bp undetermined EST of *Z. mays* etiolated seedlings (accession number: T20338). A 55% similarity (30% identity) has also been found with an amino acid sequence encoded by a cDNA from *Hordeum vulgare* (accession number: X82937), which has been reported to be induced by jasmonate, and whose function is unknown [23].

When the amino acid sequence encoded by the phy S11 cDNA was compared with that of the *A. niger* phytase Phy A using the Bestfit program of the UWGCG package, the only significant area of similarity was found to correspond to a 33 amino acid segment (underlined with a broken line on Figure 1). The corresponding sequence of Phy A contains the RHGxRxP consensus motif, characteristic of high molecular mass acid phosphatases from yeast, *E. coli* and some mammals, and believed to be the phosphate acceptor region [24,25]. Figure 2 shows the alignment of the 33 amino acid sequence of the maize Phy S11 (lane 1) with the corresponding sequences of acid phosphatases from micro-organisms (Figure 2A) and from mammals (Figure 2B). The alignment shows that residues R, H and R of the RHGxRxP consensus motif are present in that region of maize phytase sequence. Furthermore, other conserved residues are found in the sequences located upstream and downstream of this motif. The conserved residues observed in the two alignments (with phosphatases from micro-organisms and from mammals) are not the same. The highest conservation is found with sequences of Phy A and Phy B *Aspergillus* phytases: the maize sequence presents 48% similarity (27% identity) with the corresponding sequence of the Phy A phytase (Figure 2A, lane 2) and 51% similarity with that of Phy B phytase (lane 3).

Southern blot analysis and genetic mapping

The phy S11 cDNA was hybridized to *Eco*RI, *Eco*RV, *Dra*I and *Hin*dIII digested genomic DNA. No corresponding sites for

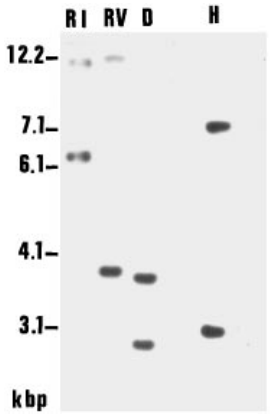

Figure 3 Southern blot hybridization of maize genomic DNA

Genomic DNA (10 μg) was digested with the restriction enzymes *Eco*RI (RI), *Eco*RV (RV), *Dra*I (D) and *Hin*dIII (H). Phy S11 cDNA has been used as the radioactive probe. The positions of molecular mass markers are indicated on the left.

these enzymes are present in the phy S11 cDNA. Two hybridizing bands were observed in each digest (Figure 3), suggesting that at most two genes were present in the maize genome.

Phy S11 cDNA was then characterized in a restriction fragment length polymorphism (RFLP) mapping population to establish its chromosomal location. Hybridization on *Eco*RI and *Hin*dIII digests revealed in each case two polymorphic fragments in the A188 and HD7 parental lines. These hybridization data allowed the genetic mapping which revealed two loci homologous to phy S11 cDNA sequence, located on the long arm of chromosome 3, just near the centromere. These loci, named *Phy I* and *Phy II*, are tightly linked; the genetic distance evaluated between them is 1 cM. They lie between RFLP markers *umc*010 and *umc*026 [26]. *Phy I* mapped at 4.6 cM from *umc*010. *Phy II*, located at the same position as the *bnl*06.06 marker, mapped at 4.2 cM from *umc*026.

Analysis of the expression of the phytase gene at various stages of maize seed germination

To investigate whether the accumulation of phytase reported in maize seedlings during the first days of germination [11] was related to an increase of the corresponding mRNA content, levels of transcripts were determined by Northern blot analysis (Figure 4). No phy S11 transcript was detected in total RNA from dry seeds or from seeds soaked for 12 h in aerated water (lanes 1 and 2). The phy S11 cDNA was found to hybridize to a unique mRNA of about 1400 bp, which appeared during the first day of germination at 26 °C (lane 3), reached its highest level at day 2 (lane 4), and then decreased progressively from day 3 to day 5 (lanes 5 to 7). Only a very slight hybridization was observed with mRNA from mature leaves (lane 8). These results show that the accumulation of phytase is correlated to an increase of the corresponding mRNA at the beginning of germination.

Expression of the phy S11 cDNA in *E. coli*

The complete coding sequence was inserted in frame at the *Nco*I site of the pET-14b vector, under the control of the bacteriophage T7 transcription and translation signals. The resulting plasmid, pET-S11, was transferred into the expression host BL21, in

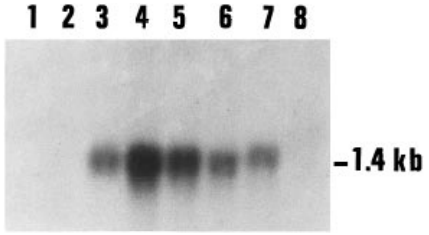

Figure 4 Time course of the accumulation of phy S11 mRNA at different stages of maize seed germination

Total RNA was extracted from dry seeds (lane 1), seeds soaked for 12 h in aerated water (lane 2) and seeds transferred into wet sand and sampled after 1 (lane 3), 2 (lane 4), 3 (lane 5), 4 (lane 6) and 5 (lane 7) days. Lane 8 corresponds to RNA extracted from mature leaves. In each case 10 μg of RNA was subjected to Northern blot analysis, using labelled phy S11 cDNA as the probe.

which the expression of T7 RNA-polymerase is inducible by isopropyl β-D-thiogalactoside (IPTG). Total proteins from non-transformed or transformed *E. coli*, before and after induction with IPTG, were extracted under denaturing conditions and analysed on SDS/PAGE. Staining of the gel with Coomassie Blue revealed a weak band at 41 kDa, which was present only after induction by IPTG (not shown). When the same extracts were analysed by Western blot (Figure 5), the purified antibody cross-reacted with a polypeptide from the induced *E. coli* extracts (lanes 3, 4 and 5), which migrated exactly at the same level as the polypeptide detected in the total proteins extracted from 5 day maize seedlings (lane 6). Interestingly, the protein synthesized in *E. coli* undergoes cleavages, resulting in a peptide of the same size as the subunit of the purified phytase. When soluble proteins were separated on PAGE under non-denaturing conditions and analysed by Western blot, a protein migrating at the same level as native maize phytase was observed, but the signal was very low (not shown). In order to concentrate this protein, the soluble extract from *E. coli* was loaded onto a small column of SP-Trisacryl and eluted with 300 mM NaCl according the procedure described to concentrate maize phytase [11]. To determine precisely the size of the protein, the eluted fraction was analysed on a 4–20 % (w/v) polyacrylamide gel gradient, in parallel with different concentrations of purified maize phytase. Western blot analyses of the gel show that a protein from *E. coli* extract eluted from SP-Trisacryl under the same conditions as maize phytase, cross-reacted with the specific antibody and co-migrated on the polyacrylamide gradient at the same level as native maize phytase

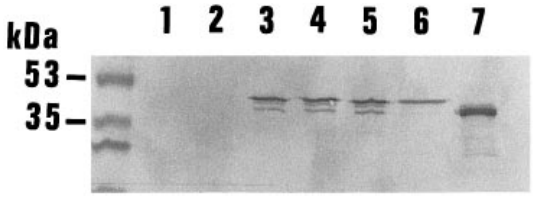

Figure 5 Analysis of the expression of phy S11 cDNA in *E. coli*

Total protein (10 μg per lane) extracted from *E. coli* BL21 was separated on SDS/12.5% (w/v) PAGE and electroblotted onto a nitrocellulose membrane. Blots were probed with the purified antibody. Lane 1: non-transformed BL21. Lane 2: BL21 transformed with the recombinant pET14, without induction by IPTG. Lanes 3–5: BL21 transformed with the recombinant pET14 after 1, 2 or 3 h of induction. Lane 6 corresponds to 10 μg of total protein from 5-day-old maize seedlings and lane 7 to 1 μg of purified maize seedling phytase.

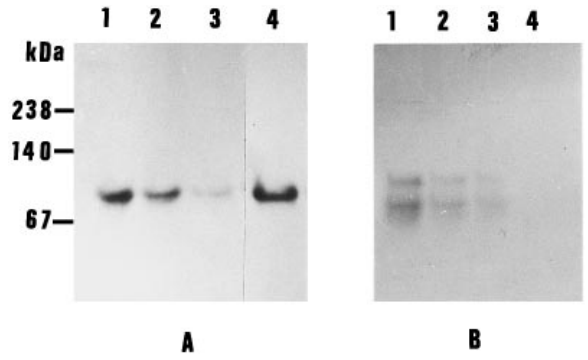

Figure 6 Analysis by electrophoresis on a 4–20% polyacrylamide gradient gel of the native protein encoded by the phy S11 cDNA in *E. coli*

(**A**) Western blot analysis: lanes 1–3 correspond to 400, 160 and 80 ng of a partly purified maize phytase preparation. Lane 4: 1 μg of *E. coli* soluble protein extract after chromatography on SP-Trisacryl and elution by 300 mM NaCl. (**B**) Phosphatase activity revealed on the gel with 1-naphthyl phosphate coupled with Fast Blue RR. Lanes 1– 4 correspond to the same lanes in (**A**). The lower band (80 kDa) corresponds to phytase. The upper band corresponds to a small degree of contamination of this preparation by an acid phosphatase of 100 kDa [11].

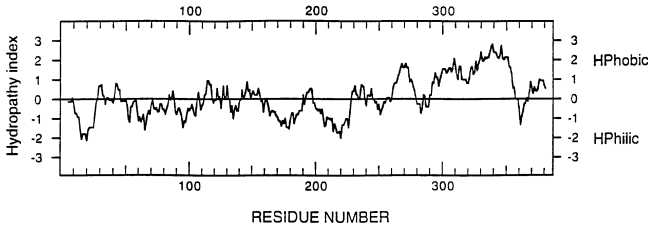

Figure 7 Hydropathy plot of the phy S11 protein sequence

Hydrophobicity values were obtained according to Kyte and Doolittle [43].

(Figure 6A). When phosphatase activity was tested on a parallel gel, activity was observed in the lanes corresponding to the different concentrations of maize-purified phytase (Figure 6B, lanes 1–3). No activity was observed at the level of the *E. coli* protein (Figure 6B, lane 4), although the amount of the *E. coli* protein was about 10 times higher than the lower concentration of maize phytase (Figure 6, lanes 3 and 4). These results show that the subunits synthesized in *E. coli* are assembled to give a dimer of the same size as native maize phytase, but this recombinant protein remains inactive.

DISCUSSION

Immunoscreening of a maize seedling cDNA library allowed the isolation of a cDNA called phy S11, encoding a 41 kDa polypeptide that corresponds to the size of the phytase subunit detected by Western blots in total protein extracts from seedlings. The N-terminal sequence of the purified phytase subunit appears as a mixture of two sequences corresponding to amino acids 19–32 and 27–41 of the sequence deduced from the phy S11 cDNA. This result is in agreement with the differences observed between the size of the subunit shown by Western blot on total protein extracts from seedlings and that of the purified protein, indicating that the observed cleavages occur during the purification process. Similar cleavages are also observed when the protein is produced in *E. coli*. The correspondence between the amino acid sequences determined on the purified protein and that deduced from the nucleotide sequence of the phy S11 cDNA confirms that this cDNA does encode the phytase subunit.

This work constitutes the first report of a sequence coding for a plant phytase. Ehrlich and Montalbano [27] and Gellatly and Lefebvre [28] have mentioned the cloning and sequencing of phytase cDNAs respectively from soybean seeds and from potato tubers. However, the corresponding sequences have never been published and thus comparison of the maize cDNA with other plant phytase cDNAs was not possible.

The only available nucleotide sequences correspond to phytases from *A. niger* [7,18]. Microbial phytases belong to class-3 phytases, which initially remove orthophosphate from position 3 of phytic acid, whereas plant phytases belong to class-6 phytases, which catalyse the removal of orthophosphate from position 6. Furthermore, microbial phytases are extracellular enzymes that are secreted under phosphate starvation, whereas plant phytases hydrolyse endogenous phytin contained inside the plant cells. Previous comparisons between purified *A. ficuum* and soybean phytases [2] showed that few physical characteristics at the protein level appeared to be in common between both enzymes. Mature fungal proteins are glycoproteins and the polypeptide sequence of Phy A phytase contains ten potential N-glycosylation sites [7]. No carbohydrate residues were detected in soybean- [29] or in maize- [11] purified phytases, by staining with periodic acid Schiff reagent, and no consensus motif for attachment of sugar moieties was found in the maize amino acid sequence. The amino acid sequence of Phy A and Phy B *Aspergillus* phytases revealed the presence of signal peptides in both fungal enzymes. No evidence for a typical signal peptide has been found in the N-terminal sequence of Phy S11. The observation that the protein synthesized in *E. coli* from the complete coding sequence of phy S11 migrated at the same level as the phytase subunit detected in the total protein extract from maize seedlings (Figure 5) confirms that no cleavage of signal peptide occurs *in vivo*.

When the amino acid sequence encoded by phy S11 cDNA was compared with that of *A. niger* Phy A, the only significant similarity concerned a 33 amino acid sequence which contains the RHGxRxP motif considered as the phosphate acceptor region [25]. The two Arg and the His residues of the consensus are conserved in the maize sequence. These residues have been shown to be essential for phosphatase activity [30]. Therefore the presence of these residues in the sequence of the maize phytase may be related to its phosphatase activity. It is noteworthy that the consensus motif is generally localized near the N-terminus of microbial acid phosphatases (position 81–87 in *A. niger* Phy A), whereas it is located between positions 204–210 in the maize sequence.

The hydropathy plot of the Phy S11 protein sequence (Figure 7) reveals the presence of about 50 hydrophobic amino acid residues (residues 290–355) close to the COOH-terminus. Such a hydrophobic sequence may have a membrane anchoring function [31]. Phytin is generally sequestered as inclusions inside single membrane enclosed organelles called protein bodies in seeds and storage bodies in pollen. In lily pollen, a cytochemical localization of phytases by the lead capture technique [32] has shown that pH 5 and pH 8 phytases of pollen were localized at the membrane of the storage bodies. Although pH 8 phytase was found to be an integral component of the membrane, the pH 5 phytase appeared to be more loosely attached, as a peripheral component, to the storage body membrane. No similar experiments have been performed to localize phytases inside the seeds. The presence of a hydrophobic region at the COOH-terminus of Phy S11 suggests that a similar anchorage might exist. The strong divergences observed between the amino acid sequences of the microbial and

the maize phytases, suggests that the homologous enzyme could be more adapted to hydrolyse endogenous phytin present inside maize seeds.

Expression in *E. coli* using the pET14 vector allowed the production of the phytase subunit as shown by Western blot after electrophoresis under denaturing conditions (Figure 5). The electrophoresis of the proteins on a non-denaturing polyacrylamide gradient gel showed that a protein from *E. coli* extract was revealed by Western blot at exactly the same level as native, purified phytase. Therefore the dimerization appears to occur in *E. coli*. The failure to express activity may be due to an incorrect folding of the subunits in the bacteria resulting in a non-functional protein. Alternatively, a post-translational modification of the protein necessary for the activity would not occur in *E. coli*. Experiments are currently being performed to try to express phy S11 cDNA in a eukaryotic host.

The estimation by Northern blots of the levels of phytase mRNA during the germination process has shown that no mRNA was present in dry seeds or in vegetative organs (leaves). This mRNA appeared very early during germination since the maximum level was detected at day 2, which corresponds to the time of radicle extrusion, suggesting that phytin hydrolysis is an important step of the early germination process. Genetic mapping data have revealed two loci homologous to the phy S11 cDNA sequence, located on chromosome 3 and separated by 1 cM. Therefore two phytase genes could be present in the maize genome. Gene isolation and transcription studies are necessary to point out whether these two genes are functional. However, the punctual nucleotide sequence divergences observed between phy AM10 and phy S11 cDNAs suggest that these two cDNAs could be encoded by distinct genes.

The cloning of the phy S11 cDNA provides a useful molecular tool to isolate the corresponding genes and to investigate their regulation and their function during seed germination.

We thank Michel Caboche for his continuous support and helpful discussions. We are very grateful to Alberto Quesada for precious help during the cDNA library construction, and to Pascual Perez and Gaëlle Baudot (BIOCEM–Limagrain) for their contribution to the genetic mapping. We thank Philippe Grappin for help with sequence alignment and Mark Tepfer and Loïc Lepiniec for critical reading of the manuscript. S.M. was the recipient of a fellowship from INRA. This work was partly supported by a grant from the MESR (ACC-SV).

REFERENCES

1 O'Dell, B. L., de Boland, A. R. and Koirtyohann, S. M. (1972) J. Agric. Food Chem. **20**, 718–721
2 Gibson, D. M. and Ullah, A. B. J. (1990) in Inositol Metabolism in Plants, pp. 77–92, Wiley-Liss, New York
3 Pointillard, A. (1994) INRA-Productions animales **7**, 29–39
4 Graf, E. (ed.) (1986) Phytic Acid, Chemistry and Applications, Pillsburry Co., Pilatus Press, Minneapolis, MN, USA
5 Nair, V. C. and Duvnjak, Z. (1990) Appl. Microbiol. Biotechnol. **34**, 183–188
6 Simons, P. C. M., Wersteegh, H. A. J., Jongbloed, A. W., Kemme, P. A. and Verschoor, G. J. (1990) Br. J. Nutr. **64**, 525–540
7 van Hartingsveldt, W., van Zeijl, C. M. J., Harteveld, G. M., Gouka, R. J., Suyberbuyk, M. E. G., Luiten, R. G. M., van Paridon, P. A., Selton, G. C. M., Veenstra, E. E., van Gorcom, R. F. M. and van den Hondel, C. A. M. J. J. (1993) Gene **127**, 87–94
8 Ehrlich, K. C., Montalbano, B. G., Mullaney, E. J., Dischinger, Jr., H. C. and Ullah, A. H. J. (1993) Biochem. Biophys. Res. Commun. **195**, 53–57
9 Pen, J., Verwoerd, T. C., van Paridon, P. A., Beudeker, R. F., van den Elzen, P. J. M., Geerse, K., van der Klis, J. D., Versteegh, H. A. J., van Ooyen, A. J. J. and Hoekema, A. (1993) Bio/Technology **11**, 811–814
10 Verwoerd, T. C., van Paridon, P. A., van Ooyen, A. J. J., van Lent, J. W. M., Hoekema, A. and Pen, J. (1995) Plant Physiol. **109**, 1199–1205
11 Labouré, A. M., Gagnon, J. and Lescure, A. M. (1993) Biochem. J. **295**, 413–419
12 Logemann, J., Schell, J. and Willmitzer, L. (1987) Anal. Biochem. **163**, 16–20
13 Lin, M., Turpin, D. H. and Plaxton, N. C. (1989) Arch. Biochem. Biophys. **269**, 219–227
14 Marin, E., Nussaume, L., Quesada, A., Sotta, B., Gonneau, M. and Marion-Poll, A. (1996) EMBO J. **15**, 2331–2342
15 Sambrook, J., Maniatis, T. and Fritsch, E. F. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
16 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463
17 Rogers, S. O. and Bendich, A. J. (1988) Plant Mol. Biol., Manual **A6**, 69–76
18 Murigneux, A., Barloy, D., Leroy, P. and Beckert, M. (1993) Theor. Appl. Genet. **86**, 837–842
19 Barloy, D., Denis, L. and Beckert, M. (1989) Maydica **34**, 303–308
20 Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. and Newburg, L. S. O. (1987) Genomics **1**, 174–181
21 Joshi, C. P. (1987) Nucleic Acids Res. **15**, 9627–9641
22 Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1986) Nucleic Acid Res. **14**, 2229–2240
23 Leopold, J., Hause, B., Lehmann, J., Graner, A., Parthier, B. and Wasternack, C. (1996) Plant, Cell Environ. **19**, 675–684
23a Reference deleted
24 van Etten, R. L., Davidson, R., Stevis, P., MacArthur, H. and Moore, D. (1991) J. Biol. Chem. **266**, 2313–2319
25 Piddington, C. S., Houston, C. S., Paloheimo, M., Cantrell, M., Miettinen-Oinonen, A., Nevalainen, H. and Rambosek, J. (1993) Gene **133**, 55–62
26 Maize Genetics Cooperation Newsletters (1996) Department of Agronomy and U.S. Department of Agriculture, University of Missouri, Columbia, MO, U.S.A. **70**, 125
27 Ehrlich, K. C. and Montalbano, B. G. (1992) Plant Physiol., supplement **99**, abstract 515
28 Gellatly, K. S. and Lefebvre, D. D. (1990) Plant Physiol., Suppl. **93**, abstract 562
29 Gibson, D. M. and Ullah, A. H. J. (1988) Arch. Biochem. Biophys. **260**, 503–513
30 Ostanin, K., Harms, E. H., Stevis, P. E., Kuciel, R., Zhou, M. M. and van Etten, R. L. (1992) J. Biol. Chem. **267**, 22830–22836
31 Pohlmann, R., Krentler, C., Schmidt, B., Schroeder, W., Lorkowski, G., Cully, J., Mersmann, G., Geier, C., Waheed, A., Gottschalk, S., Crzeschik, K. H., Hasilik, A. and von Figura, K. (1988) EMBO J. **7**, 2343–2350
32 Baldi, B. G., Scott, J. J., Everad, J. D. and Loewus, F. A. (1988) Plant Sci. **56**, 137–147
33 Elliott, S., Chang, C., Schweingruber, M. E., Schaller, J., Rickli, E. E. and Carbon, J. (1986) J. Biol. Chem. **261**, 2936–2941
34 Bajwa, W., Meyhack, B., Rudolph, H., Schweingruber, A. M. and Hinnen, A. (1984) Nucleic Acids Res. **12**, 7721–7739
35 Arima, K., Oshima, T., Kubota, I., Nakamura, N., Mizunaga, T. and Tohe, A. (1983) Nucleic Acids Res. **11**, 1657–1672
36 Chen, J. Y., Gong, Y and Ao, S. Z. (1989) Acta Biochim. Biophys. Sin. **21**, 437–444
37 Yang, J. and Schweingruber, M. E. (1990) Curr. Genet. **18**, 269–272
38 Dassa, J., Marck, C. and Boquet, P. L. (1990) J. Bacteriol. **172**, 5497–5500
39 Himeno, M., Fujita, H., Noguchi, Y., Kono, A. and Kato, K. (1989) Biochem. Biophys. Res. Commun. **162**, 1044–1053
40 Tailor, P. G., Govidan, M. V. and Patel, P. C. (1990) Nucleic Acids Res. **18**, 4928
41 Geier, C., von Figura, K. and Pohlmann, R. (1991) Biol. Chem. Hoppe-Seyler **372**, 301–304
42 Roiko, K., Jaenne, O. A. and Vihko, P. (1990) Gene **89**, 223–229
43 Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. **157**, 105–132

Received 29 July 1996/24 October 1996; accepted 1 November 1996