BIOCELL
2006, 30(1): 15-25

ISSN 0327 - 9545
PRINTED IN ARGENTINA

Genomic cloning and characterization of a PPA gene encoding a mannose-binding lectin from *Pinellia pedatisecta*

JUAN LIN[1], XUANWEI ZHOU[2], JIONG FEI[2], ZHIHUA LIAO[1], WANG JIN[1], XIAOFEN SUN[1,*] AND KEXUAN TANG [1,2,*]

1. State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan-SJTU-Nottingham Plant Biotechnology R&D Center, Morgan-Tan International Center for Life Sciences, Fudan University, Shanghai 200433, People's Republic of China;
2. Plant Biotechnology Research Center, Fudan-SJTU-Nottingham Plant Biotechnology R&D Center, School of Agriculture and Biology, Shanghai Jiaotong University, Shanghai 200030, People's Republic of China;

Keywords: Genomic cloning, Genomic walker technology, Mannose-binding lectin, *Pinellia pedatisecta, Pinellia pedatisecta* agglutinin (PPA)

ABSTRACT: A gene encoding a mannose-binding lectin, *Pinellia pedatisecta* agglutinin (PPA), was isolated from leaves of *Pinellia pedatisecta* using genomic walker technology. The *ppa* contained an 1140-bp 5'-upstream region, a 771-bp open reading frame (ORF) and an 829-bp 3'-downstream region. The ORF encoded a precursor polypeptide of 256 amino acid residues with a 24-amino acid signal peptide. There were one putative TATA box and six possible CAAT boxes lying in the 5'-upstream region of *ppa*. The *ppa* showed significant similarity at the nucleic acid level with genes encoding mannose-binding lectins from other *Araceae* species such as *Pinellia ternata*, *Arisaema heterophyllum*, *Colocasia esculenta* and *Arum maculatum*. At the amino acid level, PPA also shared varying homology (ranging from 40% to 85%) with mannose-binding lectins from other plant species, such as those from *Araceae*, *Alliaceae*, *Iridaceae*, *Lillaceae, Amaryllidaceae* and *Bromeliaceae*. The cloning of the *ppa* gene not only provides a basis for further investigation of PPA's structure, expression and regulation mechanism, but also enables us to test its potential role in controlling pests and fungal diseases by transferring the gene into tobacco and rice in the future.

Introduction

The lectins are plant proteins possessing at least one-catalytic domain that bind reversibly to a specific mono- or oligosaccharides. One of the main physiological roles of plant lectin is to mediate defense response in plant (Van Damme *et al*., 1998). Monocot mannose-binding lectin refers to a superfamily of strictly mannose-specific lectin. Numerous members of this superfamily have been characterized and cloned from species of the families *Alliaceae*, *Amaryllidaceae*, *Orchidaceae*, *Bromeliaceae*, *Liliaceae*, *Araceae* (Van Damme *et al*., 1998) and *Iridaceae* (Van Damme *et al*., 2000). The majority of the cloned lectins are homo-oligomeric lectins synthesized as preproproteins, which are converted into the mature lectin polypeptides by the co-translational cleavage of a signal peptide and the posttranslational remova1 of a C-terminal peptide (Van Damme *et al*., 1991). Aside from the homo-oligomeric lectins, there are three types of hetero-oligomeric forms

Address correspondence to: Xiaofen Sun and Kexuan Tang. State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200433, CHINA.
Fax: (+86-21) 65643552. E-mail: xfsun1@sohu.com or kxtang1@yahoo.com or kxtang1@sohu.com
Received on January 29, 2005. Accepted on October 27, 2005.

of lectins. The first type is the hetero-dimer that results from a noncovalent association between two different (but highly homologous) subunits of about 12 kDa, both of which are derived from separate preproproteins that undergo a processing similar to that of precursors of the homo-oligomeric lectins (e.g. bulb agglutinin from *Allium ursinum*, AUA) (Van Damme *et al*., 1993). The second type is the hetero-dimer or hetero-tetramer, which is composed of two different types of subunits that are derived from a single precursor with two distinct lectins domains. In some cases the two subunit types are highly homologous (e.g. bulb agglutinin from *Allium sativum*, ASA) (Van Damme *et al*., 1992) and in others the homology between the two domains is much lower (e.g. agglutinin from *Arum maculatum*, AMA) (Van Damme *et al*., 1995). The third type is the hetero-octamer, which is a tetramer of four identical subunits of 28 kDa containing two separate domains (e.g. tulip lectin from *Tulipa hybrid,* TxLCI) (Van Damme *et al*., 1996).

Except for AKA (Fei *et al.*, 2003), the mannose-binding lectin from *Amorphophallus konjac*, and TDA (AY347940), the mannose-binding lectin from *Typhonium divaricatum*, being the homo-oligomeric lectins, all other mannose-binding lectins isolated from *Araceae* family so far belong to hetero-tetramer lectins, such as the agglutinins from *Colocasia esculenta* (CEA) (Hirai *et al*., 1993; Bezerra *et al*., 1995), *A. maculatum* (AMA), *Xanthosoma sagittifolium* (XSA), *Dieffenbachia sequina* (DSA) (Van Damme *et al.,* 1995), *Arisaema heterophyllum* (AHA) (Zhao *et al.*, 2003) and *Pinellia ternata* (PTA) (Yao *et al*., 2003).

The snowdrop lectin (*Galanthus nivalis* agglutinin, GNA) is the first reported and extensively studied mannose-binding lectin (Van Damme *et al.,* 1987). During past 15 years GNA has been studied as the model of mannose-binding lectins for the biochemical properties, molecular structures, carbohydrate-binding specificities and biological activities (Van Damme *et al.,* 1998). At present, over 50 mannose-binding lectins from different plant species have been purified and characterized in some details. However, until now there are few reports on studies regarding the structures of the genomic forms and the promoter analyses of mannose-binding lectin genes.

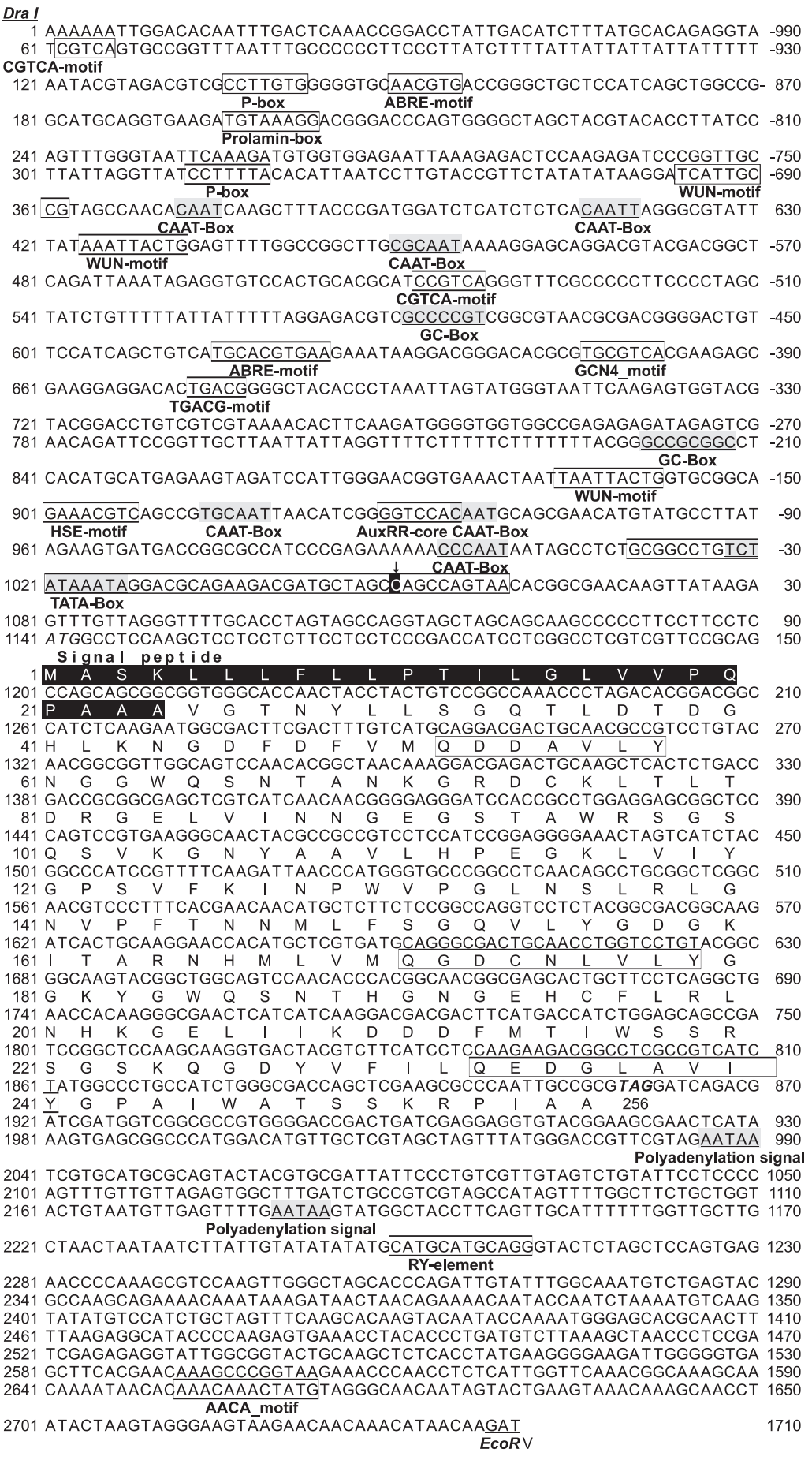
Pinellia pedatisecta agglutinin (PPA) is a very basic protein accumulated in the tuber of *Pinellia pedatisecta*, an *Araceae* species, and is widely used as Chinese traditional medicine. PPA was a tetrameric protein of 40 kDa composed of two polypeptide chains that are slightly different in size and great different in pI (Sun *et al.*, 1995). Previous studies showed PPA exhibited various pharmacological and biological activities such as termination of pregnancy (Tao *et al.*, 1981) and anti-tumor activity (Sun *et al.*, 1992; Zhu *et al.*, 1999). Recent insect bioassay studies showed that PPA had significant insecticidal activities towards cotton aphids (*Aphis gossypii* Glover) and peach potato aphids (*Myzus persicae* Sulzer) when incorporated into artificial diets at 1.2 g/l and 1.5 g/l, respectively (Huang *et al.*, 1997; Pan *et al.*, 1998). The insecticidal activities of PPA were very similar to that of GNA, making PPA a potential candidate in controlling aphids by genetic engineering. Until now, there is no report on the cloning of *ppa* gene, either in genomic form or in cDNA form. In this paper, we reported for the first time on the cloning of the genomic gene sequence of the *ppa* using genomic walker technology and on the studies for the structure of the genomic form and the promoter analysis of *ppa*, as the example of mannose-binding lectin genes. The sequence analyses on the molecular provided useful information not only on the structure of mannose-binding lectin gene, but also on the regulation of gene expression.

Materials and Methods

Plant materials

The corms of *P. pedatisecta* were collected from Jinhua, Zhejiang province, China. The corms were

FIGURE 1. The genomic DNA sequence and the deduced amino acid sequence of *P. pedatisecta* agglutinin gene (*ppa*). The start codon (ATG) was indicated italically and the stop codon (TAG) was bolded italically. Mannose-binding motifs (QXDXN/LXVXY) were boxed. The putative one TATA-box, six CAAT-boxes, two GC-boxes and two polyadenylation signals (AATAA) were underlined with gray background. The inducible elements and the tissue or developmental stage specific factors were boxed. The predicted single peptide sequence was shown in black background. The upright arrowheads indicated the start site of transcription (first letter A)


Dra I
CGTCA-motif
P-box
ABRE-motif
Prolamin-box
P-box
WUN-motif
CAAT-Box
CAAT-Box
WUN-motif
CAAT-Box
CGTCA-motif
GC-Box
ABRE-motif
GCN4_motif
TGACG-motif
GC-Box
WUN-motif
HSE-motif
CAAT-Box
AuxRR-core
CAAT-Box
CAAT-Box
TATA-Box
Signal peptide
Polyadenylation signal
Polyadenylation signal
RY-element
AACA_motif
EcoRV

grown in pots in the greenhouse under standard conditions. Leaves were collected from two-month-old seedlings. The materials were stored at -70°C until use.

DNA isolation

Genomic DNA was extracted by the method of Dellaporta *et al.* (1983). Leaf materials (1 g) were homogenized in liquid nitrogen in a pre-cooled mortar, transferred to a 50 ml tube containing 5.0 ml of extraction buffer [1 volume DNA extraction buffer (140 ml 0.25M sorbitol, 100 ml 1M Tris pH 8.2, 200 ml 0.25M EDTA, 560 ml MQ water), 1 volume nucleic lysis buffer (200 ml 1 M Tris pH 7.5, 200 ml 0.25 M EDTA, 400 ml 5 M NaCl, 20 g CTAB, 200 ml MQ water), 0.1 volume sarkosyl (10%) and 0.02 M Na-bisulfate] and mixed gently. The mixture was then placed in 65°C water both for 1 h and the homogenate was extracted with 7.5 ml of chloroform: isoamyl alcohol (24:1). The aqueous phase was removed to a new tube containing 1 volume cold isopropanol and shaken gently until DNA precipitation. The precipitation was pelleted by centrifugation at 8,000 rpm for 10 min and the pellet (DNA) was washed with 70% ethanol, dried and resuspended in 500 µl TE buffer.

Construction of GenomeWalker DNA libraries

GenomeWalker DNA libraries were constructed using the Universal GenomeWalker™ Kit (CLONTECH Laboratories, Inc., USA). The genomic DNA was completely digested with different blunt-end restriction enzymes (*Dra*I, *EcoR*V, *Pvu*II, *Stu*I) separately and the DNA fragments were then ligated separately to the GenomeWalker adaptor. The adaptor-ligated genomic DNA fragments were referred to for convenience as GenomeWalker 'libraries'.

PCR amplification

DNA fragment was amplified by PCR using primers PPAF (5'-AAGCTCCTCCTCTT CCTCCTCC-3') and PPAR (5'-GGAGGATGAAGACGTAGTCACC-3'). The primers were designed according to conservative nucleic acid sequences possessed by other mannose-binding lectins from *Araceae* using the software of primer premier 5.0. The PCR amplification was performed in a PTC-100™ programmable (MJ Research, INC, USA) for 30 cycles (94°C for 30 sec, 55°C for 1 min, 72°C for 1 min) followed by extension for 5 min at 72°C.

The amplification of upstream sequence of *ppa* genomic DNA consisted of two PCR amplification steps per library. The primary PCR used the outer adaptor primer AP (5'-GTAATACGACTCACTATAGGGC-3') provided by the kit and an outer, gene-specific primer 5GSP (5'-GCATGACAAAGTCGAAGTCGCCATTC-3'). The amplification was performed for 7 cycles (25 sec at 94°C, 3 min at 72°C) and then 32 cycles (25 sec at 94°C, 3 min at 67°C) followed by extension for 7 min at 67°C. The primary PCR mixture was diluted and used as the template for nested PCR with the nested adaptor primer NAP (5'-ACTATAGGGCACGCGTGGT-3') provided by the kit and a nested gene-specific primer 5NGSP (5'-GCTGCTGGCTGCGGAACGACGAGG-3'). The amplification was performed for 5 cycles (25 sec at 94°C, 3 min at 72°C, and then 20 cycles (25 sec at 94°C, 3 min at 67°C) followed by extension for 7 min at 67°C.

The amplification of downstream sequence of *ppa* genomic DNA consisted of two PCR amplification steps per library. The primary PCR used the outer adaptor primer AP and an outer, gene-specific primer 3GSP (5'-GCAACGTCCCTTTCACGAACAA CATG-3'). The nested PCR used the nested adaptor primer NAP and a nested gene-specific primer 3NGSP (5'-CCACAAGGGCGAACTCATCATCAAGG-3'). The conditions of PCR reactions were the same as mentioned above.

All the PCR products were purified using Gel Extraction Mini Kit (Watson, China), ligated to pMD18-T vectors (TaKaRa, China), transformed into *E. coli* strain DH5α and then sequenced with DYEnamic Direct dGTP Sequencing Kit (Amersham) by a 373A DNA sequencer.

Sequence analysis

The encoding amino acid sequence of *ppa* genomic DNA was deduced with DNA tools 5.0. The analysis and comparison of the deduced amino acid sequence with published sequences of mannose-binding lectins were performed with blastp (Standard Protein-Protein BLAST) on NCBI (www. ncbi.nlm.nih.gov) and Vector NTI Suite 8.0. The conserved domains were searched with RPS-BLAST (Search the Conserved Domain Database) on NCBI. Promoter motifs and transcription start site of 5' upstream were analyzed using the PlantCARE database (a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences). Two and three-dimensional structure predictions of PPA were performed with ExPASy Proteomics tools (//cn.expasy.org./tools/#secondary. and //cn.expasy.org. /swissmod /SWISS- MODEL.html).

Molecular evolution analysis

Phylogenetic analysis of PPA and mannose-binding lectins from other plant species belonging to families *Araceae*, *Iridaceae*, *Amaryllidaceae*, *Bromeliaceae* and *Orchidaceae* family were aligned with CLUSTAL W (1.82) using default parameters and subsequently a phylogenetic tree was constructed by the neighbor-joining method (Thompson *et al.*, 1994).

Results and Discussion

Cloning of the ppa gene

A DNA fragment, designated PPAD1, was amplified by PCR with primers PPAF and PPAR and showed a 700 bp band in a 1.0 % agarose gel. Sequence analysis revealed that the fragment was homology with other mannose-binding lectins from *Araceae*. According to the sequence of this fragment, four specific primers (5GSP and 5NGSP, 3GSP and 3NGSP) were designed and synthesized. The DNA fragments of upstream and downstream sequences were amplified using GenomeWalker DNA libraries as templates. A specific band of about 1200 bp was amplified using *Dra*I-digested genomic DNA as template and primers NAP and 5NGSP. Sequence analysis showed that 59 bp of this fragment was overlapped to PPAD1 5'-end sequence. A specific band of about 1000 bp was amplified using *Eco*RI-digested genomic DNA as template and primers NAP and 3NGSP. Sequence analysis showed that 95 bp of this fragment was overlapped to PPAD1 3'-end sequence. Finally, sequence analysis revealed that the obtained entire *ppa* genomic DNA (GenBank Accession No. AY451853) was 2740-bp containing a 771-bp gene-coding region, an 1140-bp of 5'-upstream and 829-bp of 3'-downstram regions (Fig. 1).

Characterization of the ppa gene-coding region

The *ppa* gene-coding region contained a precursor of 256 amino acids with a deduced molecular weight of 27.9 kDa and pI of 7.47. The gene-coding region of *ppa* had three mannose-binding motifs (QDNY). The amino acid sequences of the motifs I and II of PPA were the same as those of GNA, while the amino acid sequences of motif III was different from those of GNA,

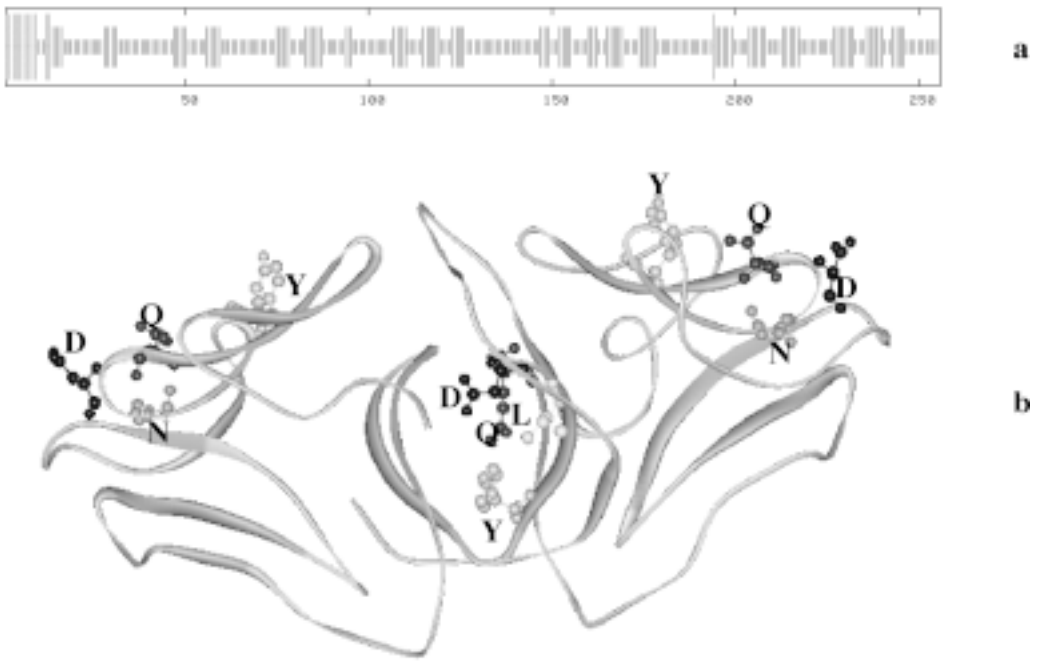


FIGURE 2. The two- and three-dimensional structures of the predicted PPA polypeptide. a) The two-dimensional structure. á-helix and extended strand were denoted as vertical long bars and vertical short bars respectively, with the horizontal line presenting the random coil running through the whole molecule. b) The three-dimensional structure. â-sheets and random coils were indicated in dark and light patches respectively. The amino acid residues QDN/LY constituting the three mannose-binding sites were signified with line-bead spatial configurations.

in which Asn (N) was substituted by Leu (L). A 24-amino acid signal peptide was predicted in PPA based on the rules of predicting signal peptide (Heijne, 1986), which was in agreement with those in most reported mannose-binding lectins from family *Amaryllidaceae* (Van Damme *et al.*, 1998) (Fig. 1). The PPA signal peptide contained highly hydrophobic amino acid residues (79.2%), implying it was a secretary putative signal peptide. Similar secretary signal peptide sequences were also reported in other mannose-binding lectins from *Araceae* such as AMA (Van Damme *et al.,* 1995), PTA (Yao *et al.,* 2003), AHA (Zhao *et al.,* 2003) and CEA (Bezerra *et al.,* 1995). After deleting the signal peptide, the deduced amino acid sequence indicated a proprotein with a molecular weight of 25.4 kDa and a pI of 7.08 and post-translationally processed into two mature pep-

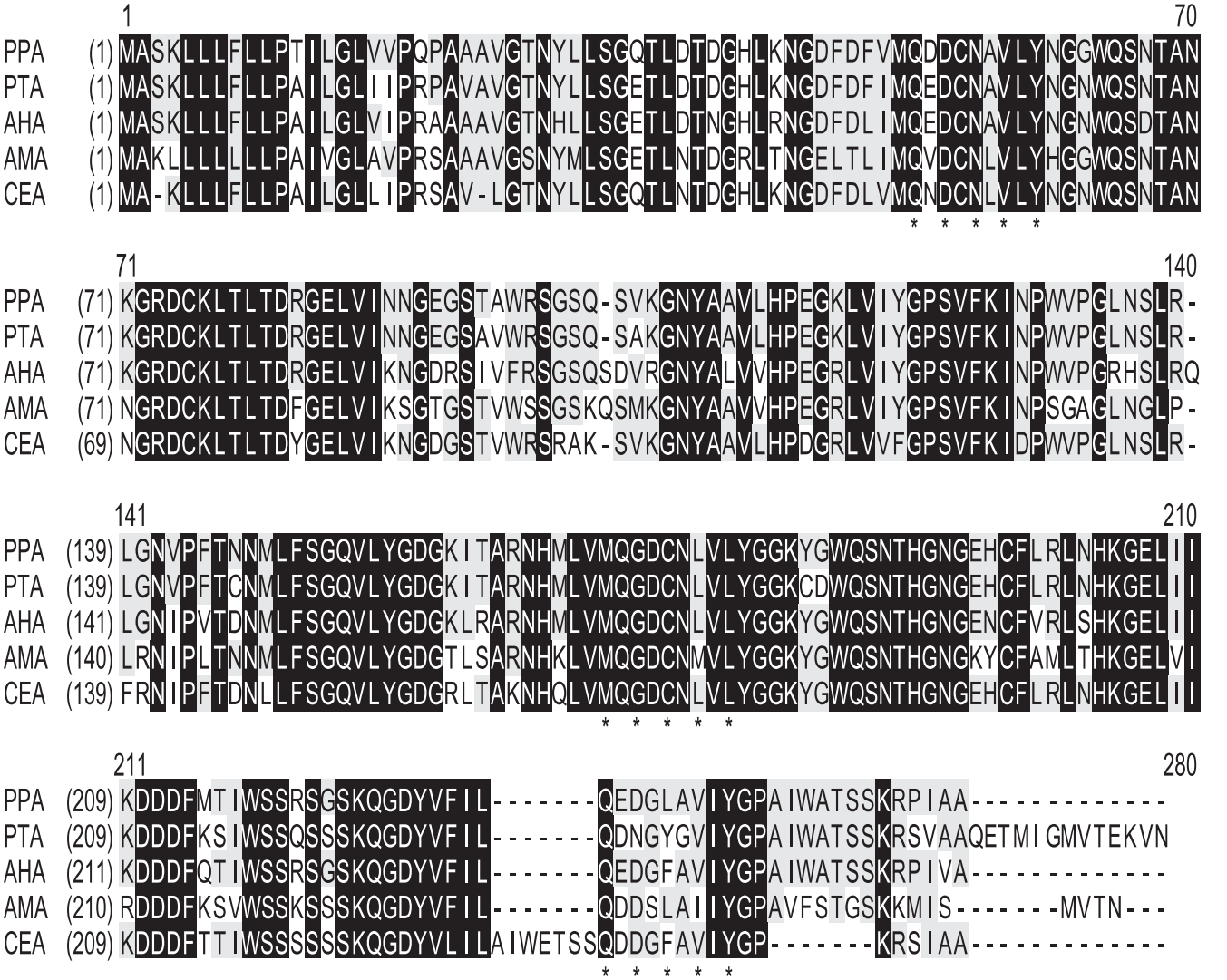


FIGURE 3. Multiple alignment of amino acid sequence of PPA with those of other mannose-binding lectins from *Araceae* species. The alignment was performed with the Vector NTI Suite 8.0 by using mannose-binding lectin sequences of PTA (*Pinellia ternata*) (AY191305), AHA (*Arisaema heterophyllum*) (AAP50524), AMA (*Arum maculatum*) (AAC48997), CEA (*Colocasia esculenta*) (BAA03722) and PPA (*Pinellia pedatisecta*). Gaps were introduced for optimal alignment and maximum similarity between all compared sequences. The identical amino acids among all the aligned sequences were shown in black background and the identical amino acids with those in PPA were shown in gray background. Mannose-binding sites were shown with '*'.

tides with a molecular weight of 12.5 kDa and 12.9 kDa, respectively, based on the database search with Blast RPS on NCBI.

Two and three-dimensional structure predictions of PPA were conducted (Fig. 2). Based on the Hierarchical Neural Network method result, PPA proprotein was composed of 4.69% alpha helix, 34.77% extended strand and 60.55% random coil. The alpha helix lied mostly in signal peptide. Penetrating through most parts of the PPA, random coil was the most abundant structural element of PPA, while extended strands were intermittently distributed in the proprotein. The amino acids (QDN) of mannose-binding motifs consisted of random coil, the amino acid (Y) consisted of extended strand (Fig. 2a). Swiss-Model structure prediction resulted in a similar folding mode and spatial configuration of PPA to GNA (Barre *et al*., 2001) (Fig. 2b). PPA proprotein was also composed of three sub-domains, each with a conserved mannose-binding motif. The first two sub-domains lied in both sides of the three-dimensional structure and the third sub-domain lied in the middle of the structure.

The genomic cloning of *ppa* using genomic walker technology revealed that this lectin belonged to the superfamily of monocot mannose-binding lectins. The analysis results showed that PPA consisted of a signal peptide and two similar tandem arrayed domains with a reasonable sequence identity/similarity to GNA.

Homology analysis

The database search and analysis with blastp showed that the deduced amino acid sequence of PPA had higher homologies with hetero-oligomeric mannose-binding lectins than with homo-oligomeric mannose-binding lectins. The comparison of deduced amino acid sequence of PPA with those hetero-tetramer mannose-binding lectins through search with Blastp showed PPA had highest identities with the mannose-binding lectins from *Araceae* species. The percentages of identity were from 86% to 76%, such as 86% (222/256), 80% (208/258), 83% (191/230) and 76% (183/239) identities to PTA (AY191305), AHA (AAP50524), CEA (BAA03722) and AMA (AAC48997), respectively (Fig. 3). The deduced PPA had also identities with mannose-binding lectins from other plant families such as *Alliaceae*, *Hyacinthaceae*, *Iridaceae* and *Lillaceae*. The percentages of identity were from 43% to 31%, such as 43% (103/237), 42% (104/244), 39% (99/248), 34% (90/263) and 34% (81/237) identities to CSA (*Crocus sativus* agglutinin) (AF347116), CVA (*Crocus vernus* agglutinin) (AF233285), THA (*T. hybrid* agglutinin) (S62648), ASA (S23496) and HHA (*Hyacinthoides hispanica* agglutinin) (AAD16404), respectively. It appeared therefore that the obvious sequence similarity of PPA with previously isolated monocot mannose-binding lectins from *Araceae* supported that mannose-binding lectins of *Araceae* belonged to hetero-tetramers (Van Damme *et al.*, 1995).

According to the NCBI conserved domain search, the predicted PPA possessed two-conserved domains, called PPA-DOM1 and PPA-DOM2. PPA-DOM1 was between T_{27} and W_{130} and PPA-DOM2 was between D_{146} and S_{250}. They were 43% (47/107) identity. One belonged to agglutinin family (pfam01453) (probable mannose binding) (CD-Length = 110 residues) and the members of this family were plant lectins that also contained a number of S-locus glycoproteins. The other belonged to B-lectin (cd00028) (Bulb-type mannose-specific lectin) (CD-Length = 116 residues) and the members of this family were involved in a-D-mannose recognition and contained a consensus sequence motif (QXDXNXVXY). The comparison of the sequence of PPA-DOM1 and PPA-DOM2 with those conserved domains through search from NCBI and GNA showed the higher identity in a consensus sequence motif (QXDXNXVXY) (Fig. 4). One consensus sequence motif was quite matching to the QXDXNXVXY sequence in PPA-DOM1 and PPA-DOM2. However, the hydrophobic residue, Val, was identical with the exception of one residue that Val-63 in GNA was replaced by Leu-66 in PPA DOM2. The function of Val was to interact with C3 and C4 of mannose through hydrophobic interactions.

Taking into consideration that the native PPA lectin had a molecular mass of <50 kDa, one could reasonably assume that PPA consisted of two identical (two-domain) protomers. Consequently, the lectin behaved as a hetero-tetramer consisting of two polypeptides corresponding to domain 1 and two polypeptides corresponding to domain 2 of LECPPA.

In the past, multiple mannose-binding lectins have been isolated and characterized. According to previous studies, the lectins belonged to homo-oligomeric lectins containing three, two or one conservative mannose-binding motif (QXDXNXVXY) (Ramachandraiah *et al.,* 2000). Until now, only nine monocot mannose-binding lectins composed of either intact or cleaved two-domain protomers have been identified, that were ASA from *Alliaceae*, HHA from *Lillaceae*, CVA from *Iridaceae*, CSA from *Iridaceae*, THA from *Lillaceae*, AMA from *Araceae*, CEA from *Araceae*, PTA from *Araceae* and AHA from *Araceae.* These lectins possessed two do-

mains, called DOM1 and DOM2. DOM1 contained one consensus mannose-binding motif (QXDXNXVXY) and DOM2 contained two, one or no consensus mannose-binding motif (QXDXNXVXY). For example, HHA and ASA contained two consensus mannose-binding motifs. AHA, PTA, CEA, AMA, CSA and CVA contained one consensus mannose-binding motif. THA contained no consensus mannose-binding motif (data no shown). These motifs were essential for mannose binding property. The number of mannose-binding motif was correlated with the capability of binding mannose. Comparison of this mannose-binding motif of PPA with

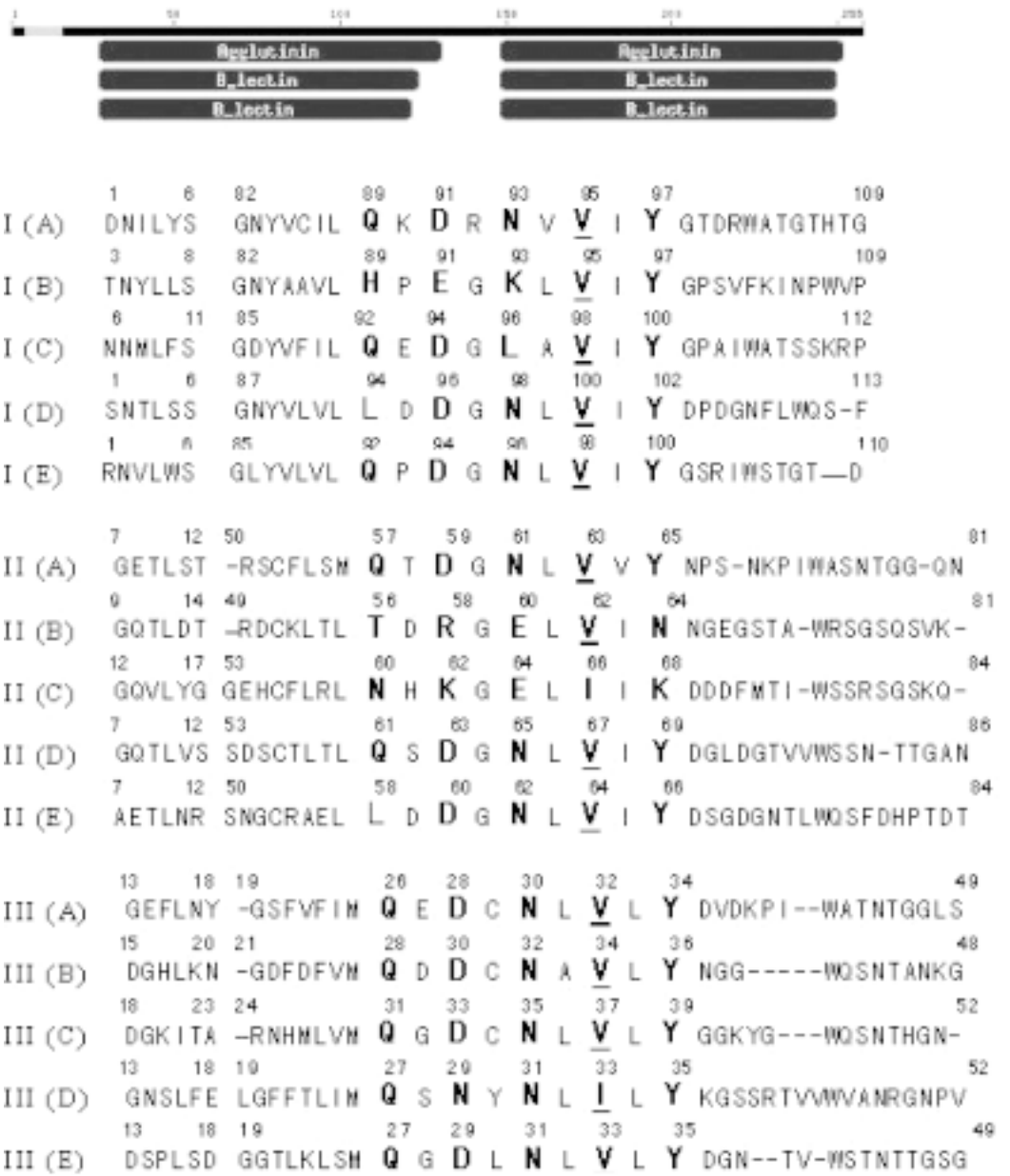


FIGURE 4. Alignment of the amino acid sequence stretches forming the three subdomains I (A), II (A) and III (A) of GNA (*Galanthus nivalis*) (P30617) with those PPA-DOM1 [I (B), II (B) and III (B)], PPA-DOM2 [I (C), II (C) and III (C)], AGGLUTININ search from GenBank [I (D), II (D) and III (D)] and B-LECTIN search from GenBank [I (E), II (E) and III (E)]. The conserved amino acid residues forming the mannose-binding site of GNA and the corresponding residues of PPA, AGGLUTININ and B-LECTIN are shown in bold. The conserved Val residue participating in the mannose-binding sites of GNA is in bold and underlined.

GNA and other lectins from *Araceae* showed the first two motifs were matching while the third motif was varied in which Asn (N) was substituted by Leu (L) (Fig. 3). The result showed that PPA possessed two domains, DOM1 contained one consensus mannose-binding motif and DOM2 also contained one consensus mannose-binding motif.

Characterization of the ppa 5'-upstream region

The 5'-upstream region of *ppa* had high content of A+T (52.54%), which was also found in other mannose-binding lectin genes, such as 5' regulatory region of CEA gene (58% of A+T) (AF178113). Analysis of the promoter sequence of PPA using the PlantCARE database identified a conserved transcription start site at -91 bp position upstream from the start codon ATG. Usually, the structures of 5' gene flanking region of eukaryotes were comprised of four parts: the site of start transcription, TATA box, CAAT box and GC box. TATA box was generally located at -32±7 bp positions upstream from the start of transcription. The consensus sequence for the TATA box was [T (CG) TATA (TA) A_{1-3}(CT) A] that was important for eukaryotic transcription (Joshi*,* 1987). A TATAA sequence was found to be located at -26 bp position upstream from the start of transcription in PPA, which might be important for the transcription control as well. The consensus sequence for TATA box was TCTATAAATA.

Six CAAT boxes were identified at the -51, -112, -132, -596, -641 and -676 bp positions upstream from the start of transcription (Fig. 1). The CAAT box was sometimes important for the efficiency of eukaryotic transcription (Benoist *et al.*, 1980). Usually, CAAT boxes were found at the -77±10 bp positions upstream from the start of transcription, although longer intervals have also been found earlier. It was reported previously that another possible plant consensus control element was present in zein genes and other plant gene promoters: (CT)A_{2-5}(GT)NGA_{2-4}(CT)(CT) (Halling *et al.,* 1985). In the present study two sequences resembling the above consensus sequence were identified in the PPA at the -56 and -169 bp positions upstream from the start of transcription, which were C*CC*GAGAAAAAACC and *G*AA*CG*GTGAAACT (the italic represented the base of dissimilarity).

There were two GC boxes (GCCGCGGC and GCCCCGT) located at -213 and -474 bp positions respectively upstream from the start of transcription (Fig. 1).

In addition to Cis-acting element, some other elements were also identified in the 5' flanking region of the PPA, which included inducible elements by physiological and environmental factors, e.g. abscisic acid responsiveness (ABRE: aACGTg or tgcACGTgaa), auxin responsiveness (AuxRR-core: GGTCcac), MeJA-responsiveness (CGTCA-motif: CGTCa and TGACG-motif: TGACg), heat stress responsiveness (HSE: aGAAAcgtc), gibberellin-responsive element (P-box: CCTTgtg), Prolamin regulatory element (Prolamin_box: tgtAAAGg), wound-responsive element (WUN-motif: tCATTgccg), and the tissue or developmental stage specific factors, e.g. endosperm expression (GCN4_motif: tgcGTCA and Skn-1_motif: GTCAt) (Fig. 1).

An earlier research about characterization and function of *Araceae* tuber lectin (*Arum maculatum* agglutinin, AMA) revealed that the lectin was a major storage protein and the storage role could be recruited for a defense-related function when necessary (Van Damme *et al.,* 1995). Our research demonstrated that the 5'-upstream region of *ppa* genomic gene possessed some elements which were inducible by physiological and environmental factors including plant stress regulators such as salicylic acid and MeJA that could induce defense-related gene expression. These elements might be the structural bases of the lectin to possess defense-related functions. In the 5'-upstream region, *ppa* genomic gene possessed organ-specific elements, e.g. endosperm expression motif. These elements may be related with the property of the lectin to have specific expression in storage tissues (e.g. tuber or bulb).

Characterization of the ppa 3'-downstream region

The 3'-downstream region of the gene contained 56.09% of A+T. This value was slightly lower than those of other lectin genes, such as ricin gene (70% of A+T). There were two sequences in the 3'-downstream region of the *ppa* gene that resembled the dual plant gene polyadenylation signals reported before (Halling *et al.,* 1985). To polyadenylation signals of plants, the first AATAAA was always found from the 25th to 44th bp downstream from the stop codon and the second $AATAA_{1-3}$ was found to be present from the 16th to 35th bp upstream from the polyadenylation site. The two sequences were found to be located at the 124th bp and 268th bp downstream from the stop codon in the *ppa*. Besides Cis-acting element, some other elements belonging to tissue or developmental stage specific factors, e.g. seed-specific regulation (RY-element: CATGcagg), endosperm-specific negative expression (AACA_motif: aAACAaactatg) were also identified in

the 3'-flanking region of the *ppa* genomic sequence. In the 3'-downstream region, *ppa* genomic gene possessed organ-specific elements, e.g. the seed-specific regulation motif. These elements may be same as organ-specific of 5'-upstream related with the property of the lectin to have specific expression in storage tissues (e.g. tuber or bulb).

Molecular evolution analysis

The phylogenetic tree analysis demonstrated that mannose-binding lectins were derived from a common ancestor in the evolution and evoluted into two groups (Fig. 5). One group contained mannose-binding lectins belonging to homo-oligomeric lectins and the other group contained mannose-binding lectins belonging to hetero-oligomeric lectins (Fig. 5). Among hetero-oligomeric lectin group, *P. pedatisecta* and other *Araceae* species, such as *C. esculenta*, *A. maculatum* and *A. heterophyllum* clustered in one sub-group. Interestingly, according to the phylogenetic tree, PPA together with other reported hetero-oligomeric lectins were placed in a group. The phylogenetic tree analysis demonstrated that PPA belonged to hetero-oligomeric lectin.

In this article, a genomic sequence of *ppa* was presented and characterized for the first time. Multiple alignments showed that the deduced PPA was homologous with other known mannose-binding lectins, and it contained conserved motifs possessed by plant mannose-binding lectin family. The cloning of the *ppa* gene

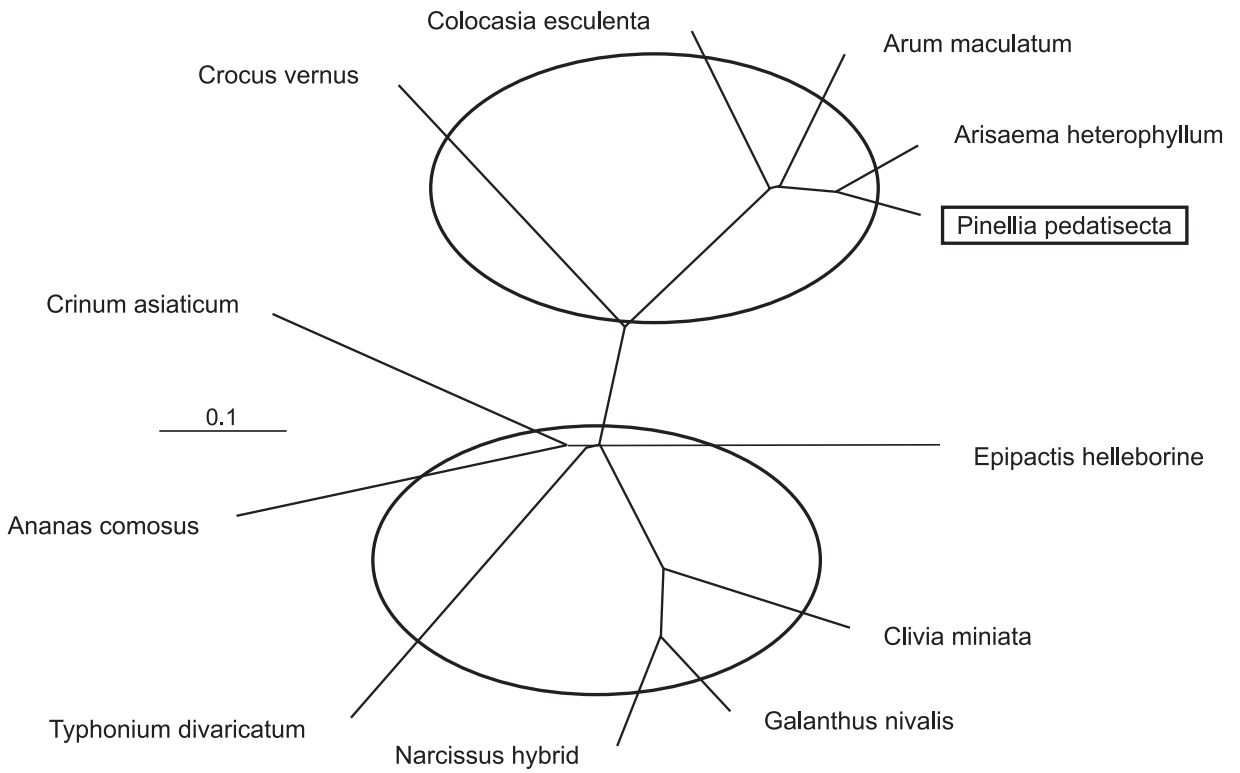


FIGURE 5. The phylogeny of mannose-binding lectins from *Araceae*, *Iridaceae*, *Amaryllidaceae*, *Bromeliaceae* and *Orchidaceae* family. A phylogenetic tree was drawn using the CLUSTAL W program. The mannose-binding lectin sequences were downloaded from NCBI: *Pinellia pedatisecta* from *Araceae* (AY451853); *Colocasia esculenta* from *Araceae* (D16173); *Arum maculatum* from *Araceae* (U12197); *Arisaema heterophyllum* from *Araceae* (AY289926); *Typhonium divaricatum* from *Araceae* (AY347940); *Crinum asiaticum* from *Amaryllidaceae* (AY212158); *Narcissus hybrid* from *Amaryllidaceae* (M88117); *Clivia miniata* from *Amaryllidaceae* (L16511); *Galanthus nivalis* from *Amaryllidaceae* (AAL07474); *Ananas comosus* from *Bromeliaceae* (AY098512); *Crocus vernus* from *Iridaceae* (AF233283); *Epipactis helleborine* from *Orchidaceae* (U07787).

provides us a basis to further investigate PPA's structure, expression and regulation mechanism, and enables us to test its potential role in controlling pests and fungal diseases by transferring the gene into tobacco and rice in the future.

Acknowledgements

This research was funded by China National Transgenic Plant Research and Commercialization Project, China National "863" High-Tech Program, China Ministry of Education and Shanghai Science and Technology Committee.

References

Barre A, Bourne Y, Van Damme EJM, Peumans WJ, Rougé P (2001). *Mannose-binding plant lectins: Different structural scaffolds for a common sugar-recognition process*. Biochimie 83: 645-651.

Benoist C, O'Hare K, Breathnach R, Chambon P (1980). *The ovalbumin gene-sequence of putative control regions.* Nucl Acids Res. 8: 127-142.

Bezerra IC, Castro LA, Neshich G, de Almeida ER, de Sa MF, Mello LV, Monte-Neshich DC (1995). *A corm-specific gene encodes tarin, a major globulin of taro (Colocasia esculenta L. Schott).* Plant Mol Biol. 28: 137-144.

Dellaporta SL, Wood JM, Hicks JB (1983). *A plant DNA minipreparation.* Plant Mol Biol Rep. 1: 19-21.

Fei J, Liao ZH, Chai YR, Pang YZ, Yao JH, Sun XF, Tang KX (2003). *Molecular cloning and characterization of a novel mannose-binding lectin gene from Amorphophallus konjac.* Mol Biol Rep. 30: 177-183.

Halling KC, Halling AC, Murray EE, Ladin BF, Houston LL, Weaver RF (1985). *Genomic cloning and characterization of a ricin gene from Ricinus communis.* Nucl Acids Res. 13: 8019-8033.

Heijne GV (1986). *A new method for predicting signal sequence cleavage sites.* Nucl Acids Res. 14: 4683-4690.

Hirai M, Nakamura K, Imai T, Sato T (1993). *cDNAs encoding for storage proteins in the tubers of taro (Colocasia esculenta Schott).* Jpn J Genet. 68: 229-236.

Huang DF, Pan YH, Zhang SX, Cao JP, Yang XM, Zhang J, Yi WZ (1997). *The discovery of insecticidal protein against aphids from Pinellia pedatisecta and P. ternata.* Sci. Agri. Sin. 30: 94.

Joshi CP (1987). *An inspection of the domain between putative TATA box and translation start site in 79 plant genes.* Nucl Acids Res. 15: 6643-6653.

Pan YH, Zhang SX, Cao JP, Yang XM, Zhang J, Ying WZ, Huang DF (1998). *The isolation, purification of Pinellia pedatisecta lectin and its activity on aphid-resistance*. Prog Nat Sci. 8: 502-505.

Ramachandraiah G, Chandra NR (2000). *Sequence and structural determinants of mannose recognition.* Protein Struct Funct Genet. 39: 358-364.

Sun GX, Ding SG, Qian YJ (1992). *The extraction and chemical analysis of proteins from Pinellia pedatisecta and their inhibitory effects on the mouse sarcoma-180.* J Shanghai Med Univ. 19: 17-20.

Sun GX, Ding SG, Qian YJ (1995). *Purification and characterization of Pinellia pedatisecta lectin.* J Shanghai Med Univ. 22: 299-302.

Tao ZJ, Xu QY, Wu KZ, Lian SH, Sun D (1981). *Isolation, crystallization, biological activities and some chemical characteristics of pinellin.* Acta Biochim Biophys Sin. 13: 77-82.

Thompson JD, Higgins DG, Gibson TJ (1994). *Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucl Acids Res. 22: 4673-4680.

Van Damme EJM, Allen AK, Peumans WJ (1987). *Isolation and characterization of a lectin with exclusive specificity towards mannose from snowdrop (Galanthus nivalis) bulbs.* FEBS Lett. 215: 140-144.

Van Damme EJM, Kaku H, Perini F, Goldstein IJ, Peeters B, Yagi F, Decock B, Peumans WJ (1991). *Biosynthesis, primary structure and molecular cloning of snowdrop (Galanthus nivalis L.) lectin.* Eur J Biochem. 202: 23-30.

Van Damme EJM, Smeets K, Torrekens S, Van Leuven F, Goldstein IJ, Peumans WJ (1992). *The closely related homomeric and heterodimeric mannose-binding lectins from garlic are encoded by one-domain and two-domain lectin genes, respectively.* Eur J Biochem. 206: 413-420.

Van Damme EJM, Smeets K, Torrekens S, Van Leuven F, Peumans WJ (1993). *The mannose-specific lectins from ramsons (Allium ursinum L.) are encoded by three sets of genes.* Eur J Biochem. 217: 123-129.

Van Damme EJM, Goossens K, Smeets K, Van Leuven F, Verhaert P, Peumans WJ (1995). *The major tuber storage protein of Araceae species is a lectin: Characterization and molecular cloning of the lectin from Arum maculatum.* Plant Physiol. 107: 1147-1158.

Van Damme EJM, Briké F, Winter HC, Van Leuven F, Goldstein IJ, Peumans WJ (1996). *Molecular cloning of two different mannose-binding lectins from tulip bulbs.* Eur J Biochem. 236: 419-427.

Van Damme EJM, Peumans WJ, Barre A, Rougé P (1998). *Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles.* Crit Rev Plant Sci. 17: 575-692.

Van Damme EJM, Astoul CH, Barre A, Rougé P, Peumans WJ (2000). *Cloning and characterization of a monocot mannose-binding lectin from Crocus vernus (family Iridaceae)*. Eur J Biochem. 267: 5067-5077.

Yao JH, Zhao XY, Liao ZH, Lin J, Chen ZH, Chen F, Song J, Sun XF, Tang KX (2003). *Cloning and molecular characterization of a novel lectin gene from Pinellia ternata.* Cell Res. 13: 301-308.

Zhao XY, Yao JH, Liao ZH, Zhang HY, Chen F, Lu YQ, Song J, Sun XF, Yu SQ, Tang KX (2003). *Molecular cloning of a novel mannose-binding lectin gene from Arisaema heterophyllum.* Plant Sci. 165: 55-60.

Zhu MW, Zhou KM, Ding SS (1999). *Total proteins of Pinellia pedatisecta effects in ovarian cancer cell lines and human umbilical cord blood hematopoietic progenitors.* J Shanghai Med Univ. 26: 455-458.