Molecular cloning and characterization of soybean peroxidase gene families

Huabang Chen, Richard A. Vierling *  
Indiana Crop Improvement Association and Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA

Received 2 July 1999; received in revised form 25 August 1999; accepted 25 August 1999

Abstract

Plant peroxidases play major roles in many physiological processes. A soybean seedbud (21 days after flowering) Uni-ZAP XR cDNA library was screened with a peroxidase-specific probe. The probe was generated by 3′ rapid amplification of cDNA ends with soybean seedbud total RNA and a degenerate primer derived from a plant peroxidase conserved amino acid region (distal heme ligand). Positive clones were recovered by PCR using the degenerate peroxidase-specific primer and the vector primer T7 flanking the cloning site. Four cDNAs, designated GmEpa1, GmEpa2, GmEpb1, and GmEpb2, contained 1298, 1326, 1171, and 1145 nucleotides, excluding poly(A) tail, and encoded mature proteins of 303, 303, 292, and 292 amino acids, respectively. The four predicted amino acid sequences showed homology to other peroxidases. GmEpa1 and GmEpa2 exhibited 97% amino acid identity, GmEpb1 and GmEpb2 exhibited 93% amino acid identity, and GmEpa1 and GmEpb1 exhibited 47% amino acid identity. GmEpa1 and GmEpb1 were expressed as fusion proteins in Escherichia coli. The recombinant fusion proteins were sequestered in inclusion bodies and active forms of the two denatured proteins were recovered after in vitro folding in a medium containing heme, urea and Ca++. GmEpa1 and GmEpa2 messages were detected in developing seed and root, while GmEpb1 and GmEpb2 messages were present in root, leaf, stem and seed pod. These cDNAs and cDNA-specific primers will allow investigations into peroxidase’s role in development, stress response and in other physiological processes. © 2000 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Peroxidase; cDNA; Differential expression; Soybean

1. Introduction

Peroxidases, a ubiquitous class of protein, are enzymes whose primary function is to oxidize a variety of hydrogen donors. Peroxidase has been studied extensively in higher plants and has been implicated in a variety of physiological processes including lignin biosynthesis [1], extensin polymerization [2], disease and pathogen response [3], wound healing [4], and response to air pollutant stress [5].

Most higher plants possess a number of different peroxidase isozymes. In the tobacco plant there are at least 12 distinguishable isozymes [1]. The soybean peroxidase family consists of ~ 20 isozymes [6]. The large number of peroxidase isozymes could represent gene copy numbers through polymorphism or post-translational modification. These many possibilities of so-called isozyme expression make it difficult to study the actual function of peroxidase. Peroxidase expression has been reported to be tissue-specific, developmentally regulated, and influenced by environmental factors [6].

Considering the vast number of studies in this field, the physiological function of individual members within this class of enzymes is only par-
tially understood. The lack of knowledge is due, in part, to the complexity of peroxidase isozymes expressed within a given plant tissue. Cloning of plant peroxidase genes could help in elucidating the physiological role of individual peroxidase isoforms during plant normal development and in response to various environmental factors. Understanding the expression pattern of individual peroxidase isozymes may lead to understanding their function. Without gene-specific probes, separation of the expression patterns of closely related peroxidases may be difficult due to similarity in sequence and transcript size. Cloning of plant peroxidase genes also could open the possibility of transgenic studies and in vitro site-mutagenesis. Without gene-specific probes, separation of the expression patterns of closely related peroxidase isozymes may lead to understanding their function. Without gene-specific probes, separation of the expression patterns of closely related peroxidase isozymes may lead to understanding their function.

2. Materials and methods

2.1. cDNA library construction and screening

Total RNA was extracted from soybean (Glycine max cul. Resnik) seedbuds 21 days after flowering as previously described [7]. Poly(A)-enriched RNA was prepared from total RNA using PolyATract (Promega) and the cDNA library was constructed in the unidirectional vector Uni-ZAP XR (Stratagene). A plant peroxidase specific primer (PSP) was designed from a plant peroxidase conserved amino acid region (distal heme ligand, HFHDCFV) (5’ CA(C/T)TT(T/C)CA(C/T)GA(C/T)TG(C/T)TT(C/T)GT 3’) [8]. The plant peroxidase-specific probe was generated using the 3’ RACE system with soybean seed bud total RNA and PSP as described by the manufacturer (GIBCO/BRL) except that hot-start PCR was performed. The PCR-RACE products were cloned into pCR™ II plasmid (Invitrogen). DNA from 20 clones was purified and digested with EcoRI, fractionated by electrophoresis on a 1% agarose gel, and blotted on a nylon membrane and probed with [γ-32P]dATP-end-labeled PSP. A single positive clone was random-prime-labeled with [α-32P]dCTP and used for primary screening of the cDNA library. Prehybridization was conducted in 6 × SSPE, 5 × Denhardt’s, 0.5% (w/v) SDS, 100 µg/ml denatured salmon sperm DNA, and 50% (v/v) formamide at 42°C for 2 h. Hybridizations were done overnight and the conditions were the same as those in prehybridization, except that 1 × Denhardt’s solution was used. Membranes were washed three times for 30 min each with 0.1 × SSPE and 0.1% (w/v) SDS at 65°C.

PCR using PSP and the vector primer T7 was used to purify single phage clones. Phage particles were eluted by incubating primary picks and/or single plagues in 500 µl of SM buffer (100 mM NaCl, 10 mM MgSO4, 0.01% (w/v) gelatin in 50 mM Tris pH 7.5) at room temperature for 2 h. The PCR cycling parameters were 94°C for 4 min followed by 30 cycles of 1 min at 94°C, 1 min at 57°C, and 1 min at 72°C, and then followed by a final extension at 72°C for 5 min. PCR reaction conditions were 1 × reaction buffer (500 mM KCl, 100 mM Tris–HCl, pH 9.0 and 1.0% (v/v) Triton X-100), 1.5 mM MgCl2, 200 µM each dNTP, two units of Taq DNA polymerase, 1 µM each primer and 2 µl of phage particle elution in 50 µl total volume.

2.2. DNA sequencing and analysis

DNA sequencing of both strands was performed using Sequenase Kit 2.0 (USB) and SK and KS primers. Synthetic primers corresponding to internal sequences of cDNA were made to complete sequencing. Sequences were optimally aligned using the PILEUP program of the Wisconsin GCG software package, and the DISTANCES program was used to determine the similarity percentage. The dendrogram was generated using the clustering algorithm in PILEUP, which aligns the most similar sequences prior to alignment with more distantly related sequences.

2.3. Reverse transcription PCR (RT-PCR)

cDNA specific primers designed from 3’ untranslated regions of each peroxidase cDNA and PSP were used in reverse transcript PCR (RT-PCR) to study expression patterns. For GmEpa1, GmEpa2, GmEph1, and GmEph2 the primers were 5’ AAATTAACCTACGCTGTTGGG 3’, 5’ GGAACCCACTTATTCATCG 3’, 5’ CC-
CAAGACATGCTTGAGAT 3', and 5' AAGTTCACTACTTCTAAC 3', respectively.

A total of 5 µg of total RNA from different soybean tissues was used for synthesizing the first strand of cDNA using SUPERSCRIPT™ II RNase H− REVERSE TRANSFERASE as suggested by the manufacturer (GIBCO/BRL). RT-PCR conditions were the same as those in 3' RACE except that the annealing temperature for GmEPb2 was 45°C. A total of 20 µl of PCR products was electrophoresed on a 1% (w/v) agarose gel and visualized with ethidium bromide.

2.4. In vitro folding of bacterially expressed GmEPa1 and GmEPb1 proteins

The open reading frames including the 5' leader sequences of GmEPa1 and GmEPb1 were PCR-amplified and cloned into the pET-34b(+) expression vector (Novagen). The primers complementary to the lower strands were designed with a BamH I site at the 5' ends and the primers complementary to the upper strands with a Xho I site (GmEPa1: 5' GACCGATCCATGGGAAGCAACTTGAGGTTTTTG 3', 5' GACCTCAGTTAGCTATTTATAAATGCACAATG 3'; GmEPb1: 5' GACCGATCCATGGCTGTACCTCGAGTTAGCTATTTATAAATGCACAATG 3'; GmEPb2: 5' GACCGATCCATGGCTGTACCTCGAGTTAGCTATTTATAAATGCACAATG 3'). PCR products were digested with BamH I and Xho I and ligated to pET-34b(+) digested with the same two enzymes. The constructs were then transformed into BL21 (DE3) competent E. coli cells.

An overnight culture of 5 ml of BL21 was inoculated into a 100-ml culture containing 50 µg/ml kanamycin. The inoculated culture was grown at 37°C with vigorous shaking until an OD<sub>600</sub> of ~1.0. The 100-ml culture was then split into 2 x 50-ml cultures. IPTG was added to one of the 50-ml cultures to a final concentration of 1 mM. The other culture was used as an uninduced control. The 2 x 50-ml cultures were further grown for 5 h after IPTG induction. Different temperatures and media additives were used during the bacterial growth. Isolation of total cell proteins, inclusion bodies and cellulose binding domain (CBD) fused peroxidase were performed according to the manufacturer (Novagen). Inclusion bodies were washed twice with buffer (200 mM Tris–Cl pH 8.0, 100 mM EDTA, 100 mM DTT, 10% (v/v) Triton X-100) before being fully solubilized in 10 vol. of 6 M urea, 1 mM DTT in 50 mM Tris–Cl (pH 8.0).

A single-step dilution was used for the denatured protein refolding. A total of 10 µg (1µg/µl) of the inclusion body prep was slowly diluted in 190 µl of PBS, which once diluted, contained 2 M urea, 5 mM CaCl₂, 10 µM hemin and 0.1 mM DTT (PBS: 137 mM NaCl, 1.47 mM KH₂PO₄, 8.10 mM Na₂HPO₄, and 2.68 mM KCl, pH 8.0). After overnight incubation at room temperature, 20-µl aliquots of the folding mixtures were transferred to the wells of a microtiter plate, and peroxidase activity was monitored using substrate tetramethylbenzadine as described by Vierling and Wilcox [9].

3. Results

3.1. Isolation of soybean peroxidase cDNA

The vast number of plant peroxidase sequences documented and the rapid amplification of cDNA ends (RACE) technique made possible the generation of a plant peroxidase-specific probe. A degenerate plant peroxidase-specific primer (PSP) corresponding to a highly conserved region, distal heme ligand (HFHDFCV) was synthesized. Using PSP and anchor primer complementary to the poly(dT) end of the cDNA, the 3' RACE experiment resulted in amplification of a major DNA band of ~900 bp (data not shown). The fragment was cloned and one of the clones was used as probe to screen the cDNA library. Approximately 2 x 10<sup>5</sup> recombinant phages from the soybean seedbud cDNA primary library were screened using the plant peroxidase-specific probe. A total of 25 clones were obtained by primary screening, and 11 positive clones were recovered after two rounds of PCR using PSP and T<sub>7</sub> vector primers. The four longest clones, designated GmEPa1, GmEPa2, GmEPb1 and GmEPb2, were further analyzed.

3.2. Nucleotide and deduced protein sequences of the soybean peroxidase cDNA

GmEPa1, GmEPa2, GmEPb1, and GmEPb2 have been registered in the EMBL Nucleotide Database under the accession numbers U51191, 150 (2000) 129–137 131.
U51192, U51193 and U51194. They contained 1298, 1326, 1171 and 1145 nucleotides, excluding poly(A) tail, with 86, 82, 59 and 38-bp 5′ untranslated leaders, and 240, 272, 173 and 168-bp 3′ untranslated regions, respectively. Two copies of the putative polyadenylation signals AATAAG are present at nucleotides 25 and 81 upstream of the poly(A) tail in GmEpa1, and 45 and 112 bases upstream of the poly(A) tail in GmEpa2. There was only one copy of the putative polyadenylation signal AATAAA 42 bases upstream of the poly(A) tail in GmEpb1 and 20 bases upstream in GmEpb2.

The open reading frames (ORFs) of GmEPa1, GmEPa2, GmEPb1 and GmEPb2 were 972, 972, 942 and 942 bp long. The deduced amino acid sequences encoded by the four ORFs are shown in Fig. 1. It was predicted from these sequences that the proteins were synthesized as preproteins of 324, 324, 314 and 314 amino acids with hydrophobic putative signal sequences of 21, 21, 22 and 22 residues, respectively. The mature proteins from GmEpa1, GmEpa2, GmEpb1 and GmEpb2 were designated as a1, a2, b1 and b2. Cleavage of putative signal sequences releases mature proteins of 303, 303, 292, and 292 residues with theoretical M, of 33333, 33333, 32412 and 32412 Da. The theoretical pIs of mature a1, a2, b1 and b2 were 6.96, 7.41, 10.04 and 9.05, respectively. There were six putative glycosylation sites specified by N-X-T/S at amino acid residues 36, 69, 128, 142, 183 and 214 in a1 and a2, and four putative glycosylation sites at residues 70, 142, 185 and 195 in b1 and b2. Peroxidases a1 and a2 had the [Q L X X F Y] motif at the NH2 terminus that was a feature...
Fig. 2. (A) Northern blot analysis of GmEPa1 expression in root (R), seedpod (Sp), stem (S), leaf (L), and developing seed coat (Ds). (B) Ethidium bromide stained RNA gel indicating roughly equal loading of total RNA. (C) Ethidium bromide stained RT-PCR products amplified from cDNA synthesized from total RNA of R, Sp, S, L, and Ds. The two primers were PSP and GmEPa1-specific primer. (D) Southern blot analysis of RT-PCR products amplified using PSP and GmEPa1-specific primers. The probe was 32P-labelled GmEPa1.

3.3. Differential expressions of peroxidase mRNA

Total RNA from leaf, stem, root, seedpod, and developing seed were probed with a 300-bp genespecific Kpn-Tif I fragment from the 3' end of GmEpa1. Transcripts of ~1400 nucleotides from GmEpa1 were present in developing seed and root (Fig. 2). Since both the coding regions and the non-coding regions of the four cDNAs are highly homologous, RT-PCR experiments were conducted to study the differential expressions of peroxidase mRNA. The divergence at the 3' end of GmEPa1 and GmEPa2, GmEPb1 and GmEPb2, enabled us to design gene-specific PCR primers (Fig. 3). cDNA specific primers designed from this region of each peroxidase cDNA and PSP were used in RT-PCR to study expression patterns. The RT-PCR results (Fig. 2) for GmEPa1 were consistent with the above Northern blotting analysis for GmEPa1, and the RT-PCR products were also confirmed by probing the products with GmEPa1 (Fig. 2). Based on the results of cDNA-specific primers, transcripts from GmEpa2 were also detected in root and developing seed, and transcripts from GmEpb1 and GmEpb2 were detected in root, stem, leaf, and seedpod (Fig. 4).

3.4. In vitro folding of bacterially expressed GmEPa1 proteins

In order to obtain expression of soybean peroxidases, the ORFs of GmEPa1 and GmEPb1 including the leader sequences were cloned into the expression vector pET-34b(+). Since the two sequences showed the same features in both expression and in vitro folding, only GmEPa1 is reported here. Peroxidase a1 was found toxic to E. coli growth upon early IPTG induction, even though the peroxidase was produced as a CBD-peroxidase fusion protein (Figs. 5 and 6). The E. coli culture containing the vector construct was then induced by IPTG at OD600 = 1.0 to maximize the fusion protein production. As shown in Fig. 6, a fusion protein of 60 kDa was produced and the fusion protein was accumulated in inclusion bodies under all conditions tested. After four rounds of sonication and washing, the urea-solubilized inclusion prep gave >90% pure fusion proteins as judged by SDS–PAGE (Fig. 6). As can be seen in Figs. 7 and 8, the recovery of peroxidase activity was critically dependent on the addition of hemin and on the concentration of urea, with ~2 M being optimal at pH 8.0. The addition of oxidized glutathione (GSSG) inhibited correct folding (Fig. 8). No difference in folding efficiency existed among solubilized inclusion bodies obtained under different bacteria growing conditions (data not shown).

4. Discussion

We have cloned four soybean peroxidase cDNAs. The similarity of the predicted amino acid from GmEPa1 and GmEPa2 to other plant peroxidases ranges from 54% (cotton, L08199) to 78%

Mohan et al. [10] reported the induction of L13653 by wounding and pathogen attack, which suggested these peroxidases may play a role in wound healing and/or plant defense response. Although GmEPa1 and GmEPa2 also are expressed in seed coat, they share only about 46% amino acid identity with two other soybean seed coat peroxidases (U41657 and AF014502) [11]. The similarity of the predicted amino acid from GmEPb1 and GmEPb2 to other plant peroxidases ranges from 47% (soybean, U41657) to 64% (turnip, B23116). Further studies are needed to address the physiological roles of these genes.

It is predicted from the cDNA sequences that all four proteins are initially synthesized as preproteins with predominantly hydrophobic amino acid signal sequences, suggesting that the mature proteins could be secreted through cell membranes. The four soybean peroxidases differ in the number of glycosylation sites, their pIs, and their expression patterns. Other plant peroxidases vary greatly in the number of glycosylation sites, from one in peanut and turnip to eight in horseradish [12]. The heterogeneity of glycosylation indicates that peroxidases exist in different glycosylated forms or glycoforms. Lerouge et al. [13] reported that N-glycosylation is a major modification of plant proteins, and glycosylation is necessary for an efficient secretion of plant glycoproteins. Variability in N-linked oligosaccharide chain number and location also may be adaptively important for fine tuning catalytic properties of the functional

---

Fig. 3. (A) Comparison of the nucleotide sequences at the 3’ end of GmEPa1 and GmEPa2. Gene-specific primers were designed from the highlighted areas. (B) Comparison of the nucleotide sequences at the 3’ end of GmEPb1 and GmEPb2. Gene-specific primers were designed from the highlighted areas.
enzyme molecule. However, a glycosylation site (Asp-185) is common to most peroxidases. Wan and Huystee [14] reported that oligosaccharides were the major antigenic sites of glycoproteins during the production of antibodies. We initially screened the cDNA expression library with an anti-soybean seed coat monoclonal antibody and failed to obtain any positive clones. One possible reason could be that the *E. coli* expression system lacks glycosylation machinery, and non-glycosylation of peroxidase may result in loss of the epitope or the antibody recognized a conformational epitope.

Peroxidase genes comprise multigene families in the plant genomes. In order to analyze each member of a gene family, the availability of gene-specific probes is essential. The divergence at the 3' ends of the highly homologous soybean peroxidase coding regions enabled us to design gene-specific primers and study the expression of highly ho-

---

**Fig. 4.** Ethidium bromide stained RT-PCR products amplified from cDNA synthesized from total RNA of root (e, j, p), seedpod (d, i, o), stem (c, h, n), leaf (b, g, l), and developing seed coat (a, f, k). The primer pairs used were PSP and cDNA-specific primers derived from the 3' untranslated regions of *GmEpa2* (I), *GmEpb1* (II) and *GmEpb2* (III), respectively. Lane M was the 1-kb DNA marker.

**Fig. 5.** Growth of *E. coli* with different treatments. ○, BL21; □, BL21 + plasmid; △, BL21 + plasmid + insert; ◊, BL21 + plasmid + insert + IPTG.

**Fig. 6.** Expression of CBD-soybean peroxidase in *E. coli* and SDS–PAGE of different fractions. (1) Protein molecular weight marker. (2) Total protein of BL21. (3) Total protein of BL21 with plasmid containing insert but no IPTG induction. (4) Total protein of BL21 with plasmid containing insert and IPTG induction. (5) Bacteria growth temperature of 25°C. (6) Bacteria growth temperature of 30°C. (7) Bacteria growth temperature of 37°C. (8) Bacteria growth temperature of 37°C and addition of δ-ALA. (9) Bacteria growth temperature of 37°C and addition of hemin.
mologous genes. The expression patterns of these genes based on both Northern analysis and RT-PCR indicated differential expression of soybean peroxidase genes, which was expected based on expression patterns of other plant peroxidases. It is not apparent why some organisms have several expressed peroxidase genes, but one possibility is that different isoforms have different functions. However, different isoforms can be produced by post-translational modification, suggesting that different genes might not be necessary to provide different functions. A second possibility is that multiple genes can allow for greater regulatory flexibility. Some genes may be expressed in tissue-specific organs or at specific stages, and the expression of the genes may be determined by different signals. This appears to be the case in soybean as we have demonstrated that the four soybean peroxidases are expressed in different tissues. The use of gene-specific probes will allow us to analyze individual gene expression independent of the other highly related gene sequences. And the availability of these genes also will allow us to further study their catalytic mechanisms by site-mutagenesis, and their physiological roles by transgenic studies.

The cDNA open reading frame encoding soybean peroxidase was found to be growth inhibiting upon expression in E. coli, even when expressed in an inactive form. Thus, it is unlikely that the growth inhibiting activity is mediated by peroxidase activity generating oxygen radicals. Baronek-Roxa and Eriksson [15] found that the first 64 N-terminal amino acids of a mature, neutral horseradish peroxidase was responsible for its toxicity to E. coli growth. This region was highly conserved among peroxidases, but no homology was found when this region was compared to other protein sequences. Kempf et al. [16] reported that formation of heme-containing holoenzyme cytochrome P450 2D6 was strictly dependent on addition of the heme precursor δ-aminolevulinic acid to the E. coli culture. The addition of δ-aminolevulinic acid and hemin to the E. coli culture, however, did not result in functional expression of soybean peroxidase. Unlike previous studies [17,18], GSSG was inhibitory in this study. One possible explanation is that soybean peroxidases have seven or eight cysteine residues that form disulfide bridges. The addition of GSSG may facilitate the formation of incorrect, mispaired intra and/or intermolecular disulfide bridges among the cysteines of the reduced, inactive polypeptide. Whether refolding of the completely reduced plant enzyme would occur more readily is uncertain. The native enzyme differs from the recombinant enzyme in being glycosylated and in having a blocked N-terminus, whereas the N-terminal amino acid of the recombinant enzyme is presumably methionine. Glycosylation may serve to prevent irreversible aggregation of partially un-
folded intermediates, and the lack of glycosylation in *E. coli* system may explain, in part, the low efficiency of refolding. Meanwhile, the addition of an anti-seed-coat peroxidase monoclonal antibody did not affect the in vitro protein refolding (data not shown). Further refinement of the expression/folding/activation of soybean seed coat peroxidase is needed to provide sufficient active protein for characterization of both the wild type and site-directed mutants.

All of the purified sequences contained the expected peroxidase conserved regions and the cDNA expressed in *E. coli* was confirmed to have peroxidase activity. The use of gene-specific probes will allow the analysis of individual gene expression independent of highly related gene sequences. This will allow for greater discrimination in studying tissue specificity, inducibility and regulation of related soybean peroxidase genes.

References