THE BULLETIN OF OKAYAMA UNIVERSITY OF SCIENCE No. 42 A pp. 47-51 (2006)

Analysis of the nucleotide sequence for a thaumatin-like protein of *Polygonum tinctorium* and deduced amino acid sequence

Toshitsugu YUBISUI, Hiromi IKEDA, Yoshitsugu SHIMAZOE,

Terumi IKEGAMI and Yoshiko MINAMI

Department of Biochemistry,
Faculty of Science,
Okayama University of Science,
1-1 Ridai-cho, Okayama 700-0005, Japan

(Received September 20, 2006; accepted November 6, 2006)

Expression of sequence tag (EST) of a cDNA library of *Polygonum tinctorium* was analyzed to reveal the gene expression in the plant. Redundancy of ESTs related to chloroplast was found, and then those related to mitochondria and nuclear DNAs were also found. Among of them, an EST for a thaumatin-like protein was found, which is homologous to a secretary sweet tasting protein and also known to show antifungal activity. The nucleotide sequence for the thaumatin-like protein was analyzed to be 1320 base pairs including 5'- and 3'-noncoding sequence, and an open reading frame of 765 base pairs for 255 amino acid residues. The analyzed nucleotide sequence and the deduced amino acid sequence were homologous to those of other plant thaumatin-like proteins.

Keywords: Thaumatin-like protein; *Polygonum tinctorium*; EST; nucleotide sequence; amino acid sequence.

INTRODUCTION

EST analysis of various organisms was performed to analyze the gene expression in the developing tissues and on the time-dependent stages (Adams *et al.*, 1991, 1992). Creating the gene expression profiles is a useful method to characterize the particular cells and tissues (Matsubara *et al.*, 1993; Nishiguchi *et al.*, 1996). Thus, the systematic EST analysis is rather a better approach than the cloning of a single gene to understand the experimental organism as a whole. EST analysis on various organisms was already reported, and the databases were accumulated for general use.

Polygonum tinctorium (*P. tinctorium*) is an indigo plant, in which indoxyl is accumulated in vacuoles in the leave cells as glucoside compound. The glucoside compound of indoxyl, indican (indoxyl β-glucoside) is cleaved by b-glucosidase in chloroplast when the cells were destroyed, and the free indoxyl is air-oxidized to form the insoluble indigo pigment (Minami *et al.*, 1997, 1999). In our laboratory, various genes for proteins have been cloned and characterized (Minami *et al.*, 1997, 1999). Those include β-glucosidase and transketolase (Minami *et al.*, 1999, 2004).

In this study, the EST analysis was performed to examine the gene expression in the grown leaves of *P. tinctorium*. EST analysis carried out in this study revealed that a redundancy of chloroplast gene expression was noted. Among of them, a cDNA for thaumatin-like protein was found.

The sweet tasting protein, thaumatin was first found in arils of *Thaumatococcus daniellii* Beneth, a plant native to tropical West Africa and was studied on its exceptionally sweet taste (Van der Wel and Loeve, 1972). Thaumatin is composed of 207 amino acid residues, and its isoelectric point was estimated to be about 12. The sweetness of the protein is reported to be about one hundred thousand of sucrose on a molecular base (Van der Wel and Loeve, 1972). Based on these interesting properties of thaumatin, many studies were reported on its protein structure to reveal the critical region for the sweetness. To reveal the structural base of the sweetness of these proteins, the crystal structure of thaumatin was determined by Ko *et al.* (1994). Other

than thaumatin, other sweet proteins were also reported, and these are brazzein and monellin (Tancredi *et al.*, 2004).

Tancredi *et al.* (2004) studied on the interaction of sweet proteins with their receptor to find out the assumed "sweet finger". They prepared the assumed hairpin loop as the sweet finger, but failed to obtain the sweet peptide. From these results, they suggested that the sweet protein seems to bind the receptor different from the low molecular sweetners. Recently, Masuda *et al.* (2004) cloned a cDNA for thaumatin from *Thaumatococcus daniellii*, and expressed the recombinant protein in methylotrophic yeast, *Pichia pastoris* to characterize the protein. The recombinant thaumatin showed sweetness as well as the naturally occurring sweet protein. They analyzed function of the N-terminal and C-terminal peptides relating to the sweetness, but those were not.

Thaumatin-like protein has very similar structure to that of thaumatin, but the function is different, most of thaumatin-like proteins show the anti-fungal activity and endo-β-1,3-glucanase activity (Greiner *et al.*, 1999; Menu-Bouaouiche *et al.*, 2003). *Various plants* express many isoforms of thaumatin, and also thaumatin-like protein in various tissues.

This paper describes the EST analysis of *P. tinctorium*, and the nucleotide sequence for the thaumatin-like protein found in the cDNA library of the plant. The nucleotide sequence and the deduced amino acid sequence are very similar to those of other plant thaumatin or thaumatin-like proteins.

MATERIAL AND METHODS

Materials

All endonucleases used in this study were obtained from New England Biolabs, Inc. (Beverly, MA). Ex Taq DNA polymerase and kit reagent for polymerase chain reaction (PCR) was purchased from TaKaRa (Kyoto, Japan). Other reagents used in this study were all reagent grade.

cDNA library

A cDNA library of *P. tinctorium* was constructed from the mRNA of young leaves of the plant with a double-stranded Uni-ZAP XR vector (Stratagene, La Jolla, CA) following the manufacture's protocol as described previously (Minami *et al.*, 1999). Phage forming unit was determined to be 3×10^6/ml when the experiments started.

Analysis of the nucleotide sequence and amino acid sequence

Homologies of the nucleotide sequences with the known proteins were analyzed by the Blastx program with the protein database through Internet. The amino acid sequences of ESTs were deduced from the nucleotide sequence using the GENETYX software.

EST analysis

Phage was added to *Escherichia coli* (*E. coli*, XL1-Blue) culture for infection, and spread the mixture on the agar plate to form plaques. Phage DNAs were isolated from the plaques on the agar plates, and were used as the DNA template for PCR to amplify the insert DNA. The amplified DNAs were examined by agarose gel electrophoresis, and those with sizes of 0.3-1.5 k base pairs were used for the nucleotide sequence analysis.

Polymerase chain reaction (PCR)

The insert DNA in the phage vector was amplified by PCR with Ex Taq DNA polymerase or Taq Gold DNA polymerase. The reaction was performed with an ASTEC thermal cycler, model PC707 (Fukuoka, Japan).

Nucleotide sequence analysis

Nucleotide sequence of DNA was analyzed with Big Dye Cycle Sequencing Ready reaction for DNA (PE Applied Biosystem, CA) and ABI Prism DNA sequencer (PE Applied Biosystem, CA).

RESULTS AND DISCUSSIONS

EST analysis

Total of 185 clones of phage DNA were isolated form a cDNA library to analyze the insert DNA. The insert DNA in the phage DNA vector was amplified by PCR, and then the amplified DNA was examined by agarose gel electrophoresis. DNA with 0.3-1.5 kbp in sizes was pooled for the nucleotide sequence analysis. Of 185 clones isolated, 150 clones had the insert DNA with 0.3-1.5 kbp in size. Of 150 clones with the insert DNA, 71 clones showed a single band in the agarose gel electrophoresis, and those were then used as the DNA template for the nucleotide sequence analysis.

The insert DNA sequence was analyzed by using the T3 vector primer to amplify the DNA, and applied to a DNA sequencer. The determined nucleotide sequences were examined the homology with the known protein sequences in the database. Thus, the homology obtained by the search through Internet was shown in Table 1. Among the homologies obtained (43 clones), the valuable data with significant similarity were 18, and the redundancy of gene expression of chloroplast, cytoplasm, and nuclear genes were observed as shown in Table 1.

Table 1. EST of *P. tinctortum*, putatively identified with known proteins

No.	Homologous proteins identified
1.	*Arabidopsis thaliana* putative protein
2.	Phosphoribosylamidoimidazole-succinocarboxiamide synthetase
3.	Histone H1
4.	Photosystem I reaction center subunit V, chloroplast precursor
5.	Ribulose bisphosphate carboxylase small chain 1
6.	60S ribosomal protein L18Plexin B1 precursor
7.	Fructose-bisphosphate aldolase, chloroplast precursor
8.	S-adenosylmethionine decarboxylase proenzyme
9.	Cytochrome B6-F complex iron-sulfur subunit, chloroplast precursor
10.	Elongation factor 1-b
11.	Probable glutathione S-transferase
12.	Ribulose bisphosphate carboxylase small chain SSU26, chloroplast precursor
13.	Transketolase
14.	Ribulose bisphosphate carboxylase/oxygenase activase, chloroplast precursor
15.	Chalcone synthetase 8
16.	Thaumatin-like protein 1a precursor
17.	60S ribosomal protein L11
18.	Fructose-bisphosphate aldolase, chloroplast precursor

Nucleotide sequence and deduced amino acid sequence of thaumatin-like protein

Among the 18 clones with a significant homology was found, the DNA for the thaumatin-like protein was found and characterized the DNA in this study. As shown in Fig. 1, the DNA sequence of the thaumatin-like protein of *P. tinctorium* was composed of 1,320 bp, including the coding sequence of 765 bp for 255 amino acid residues. The DNA contains the initial Met codon, 5'-noncoding sequence of 146 bp and 3'-noncoding sequence of 306 bp with poly (A) sequence at 3'-terminal side. The sequence around the initial Met codon (CTGATGG) fits well the Kozak's motif. Fig. 1 shows the deduced amino acid sequence from the DNA sequence. When the deduced amino acid sequence of *P. tinctorium* was aligned and compared with other plant thaumatin-like proteins, significant homologies with other proteins were observed as shown in Fig. 2. Amino acid sequence of thaumatin-like protein of *P. tinctorium* was most similar to that of *Malus domestica*, 55% identity and 67% homology. Very similar results were also obtained with those of *Pyrus pyrifolia*, *Puruns persica*, and also with *Arabidopsis thaliana* as shown in Fig.2.

As shown in Fig. 1 and 2, this protein from various plants contain 15 cysteine residues, probably forming many disulfide bridges to stabilize the protein conformation. The other sweetness protein, brazzein also contain disulfide bridges, and chemical modification of cysteines lead to loss of sweetness (Ming and Hellekant, 1994). Different from thaumatin, the sweetness protein, most thaumatin-like proteins of plants are not endowed the sweet taste, and show the antifungal activity. As shown in Fig. 3, the amino acid sequence of thaumatin-like protein of *P. tinctorium* showed a high similarity to that of thaumatin from *Thaumatococcus daniellii*, but shows some different properties. Thaumatin-like protein of *P. tinctorium* has 255 amino acids

(207 residues for thaumatin), and an estimated isoelectric point of 7.7 (12 for thaumatin). Probably the thaumatin-like protein of *P. tinctorium* is also functioning as anti-fungal protein in response to fungal infection. Recently, thaumatin-like protein was found first from animals, *Schistocerca gregaria* and *Locusta migratoria* (Brandazza *et al.*, 2004). And through the search of the genome database, a similar sequence was identified from the nematode *Caenorhabditis*. The thaumatin-like proteins in animals are also considered to play a defense role against pathogens (Brandazza *et al.*, 2004).

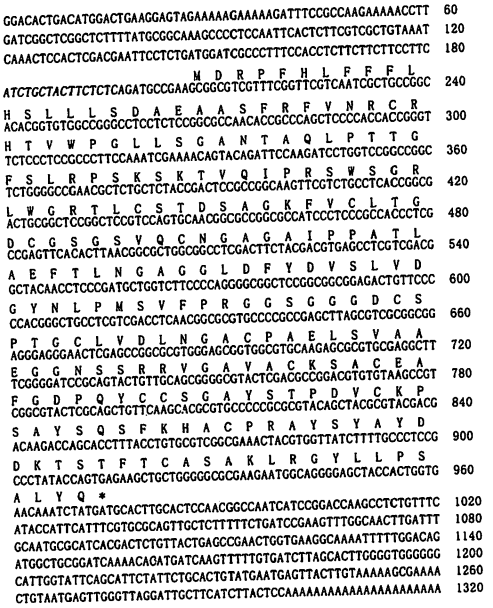

Figure 1. Nucelotide sequence of thaumatin-like protein of *P. tinctorium*, and deduced amino acid sequence

* shows the stop codon.

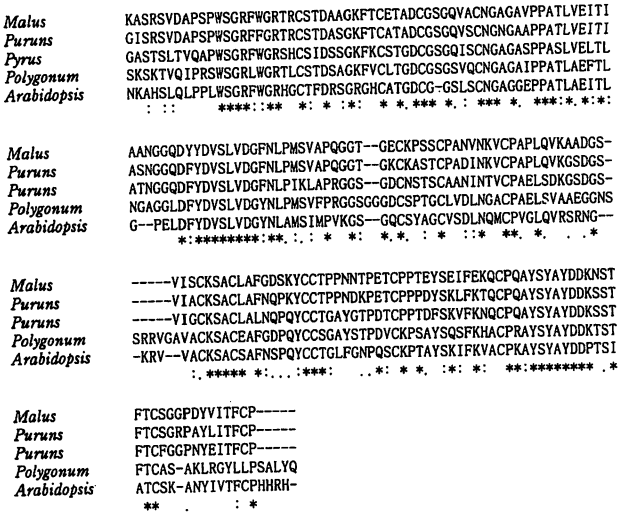

Figure 2. Comparison of thaumatin-like protein of *P. tinctorium* with those of various plants.

* shows the identical residues ; .show the homologous residues.

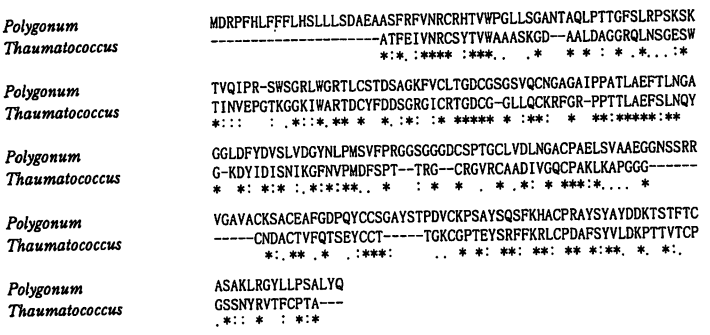

Figure 3. Comparison of amino acid sequence of thaumatin-like protein of *P. tinctorium* and that of thaumatin from *Thamatococcus daniellii*.

* shows the identical residues ; .show the homologous residues.

As a conclusion of this study, among the DNA sequences determined from 70 clones, only 18 clones showed the significant homology with known proteins, 27 clones did not hit any known proteins, and the others showed very low similarity

to known proteins. These DNA without any significant similarity to known proteins may be specific for the *P. tinctorium* proteins. Analysis of such DNA may lead to interesting functions of the plant.

REFERENCES

Adams MD, Dubnick M, Kerlavage AR, Moreno R, Kelly JM, Utterback TR, Nagle JW, Fields S, Venter JC (1992) Sequence identification of 2,375 human brain genes. *Nature* **355:** 632-634

Adams MD, Kelly JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC (1991) Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252:** 1651-1656

Assadi-Porter FM, Aceti DJ, Markley JL (2000) Sweetness determinant sites of brazzein, a small heat-stable sweet-tasting protein. *Arch. Biochem. Biophys.* **376:** 259-265

Brandazza A, Angeli S, Tegoni M, Cambillau C, Pelosi P (2004) Plant stress proteins of the thaumatin-like family discovered in animals. *FEBS Lett.* **572:** 3-7

Grenier J, Potvin C, Trudel J, Asselin A (1999) Some thaumatin-like proteins hydrolyze polymeric beta-1,3-glucans. *Plant J.* **19:** 473-480

Jin Z, Danilova V, Assadi-Porter FM, Aceti DJ, Markley JL, Hellekant G (2003) Critical region for the sweetness of brazzein. *FEBS Lett.* **544:** 33-37

Ko TP, Day J, Greenwood A, McPherson A (1994) Structure of three crystal forms of the sweet protein thaumatin. *Acta Cryst.* **D50:** 813-825

Masuda T, Tamaki S, Kaneko R, Wada R, Fujita Y, Mehta A, Kitabatake N (2004) Cloning, expression, and characterization of recombinant sweet-protein thaumatin II using the methylotrophic yeast *Pichia pastoris*. *Biotech. Bioeng.* **85:** 761-769

Matsubara K, Okubo K (1993) cDNA analysis in the human genome project. *Gene* **135:** 265-274

Meru-Bouaouiche L, Vreit C, Peumans WJ, Barre A, Van Damme EJ, Rouge P (2003) A molecular basis for the endo-beta 1,3-glucanase activity of the thaumatin-like proteins from edible fruits. *Biochimie* **85:** 123-131

Minami Y, Fujita K, Matsubara H (2004) cDNA cloning and mRNA expression of transketolase. *The Bulletin of Okayama Univ. Sci.* **40A:** 59-66

Minami Y, Shigeta Y, Tokumoto U, Tanaka Y, Yonekura-Sakakibara K, Oh-oka H, Matsubara H (1999) Cloning, sequencing, characterization, and expression of a β-glucosidase from the indigo plant. *Plant Sci.* **142:** 219-226

Minami Y, Takao H, Kanafuji T, Miura K, Kondo M, Hara-Nishimura I, Nishimura M, Matsubara H (1997) β-Glucosidase in the indigo plant: Intracellular localization and tissue specific expression in leaves. *Plant Cell Physiol.* **38:** 1069-1074

Ming D, Hellekant G (1994) Brazzein, a new high-potency thermostable sweet protein from *Pentadiplandra brazzeana* B. *FEBS Lett.* **355:** 106-108

Nishiguchi S, Sakuma R, Nomura M, Zou Z, Jearanaisilavong J, Joh T, Yasunaga T, Shimada K (1996) A catalogue of genes in mouse embryonal carcinoma F9 cells identified with expression sequence tags. *J. Biochem.* **119:** 749-767

Tancredi T, Pastore A, Salvadori S, Esposito V, Temussi PA (2004) Interaction of sweet proteins with their receptor. A conformational study of peptides corresponding to loops of brazzein, monellin and thaumatin. *Eur. J. Biochem.* **271:** 2231-2240

Temussi PA, (2002) Why are sweet protein sweet? Interaction of brazzein, monellin and thaumatin with the T1R2-T1R3 receptor. *FEBS Lett.* **526:** 1-4