

S23-009

Cloning of a cDNA encoding the 22-kDa protein in chloroplasts of spores of *Osmunda japonica*

H Inoue¹, H Tamura¹, H Kamachi¹, K Nakayama²

¹*Department of Environmental Biology and Chemistry, Faculty of Science, Toyama University, Toyama 930-8555, Japan. inoue@sci.toyama-u.ac.jp*

²*Department of Anatomy, School of Medicine, Shinshu University, Matsumoto 390-0802, Japan. kohzona@gipac.shinshu-u.ac.jp*

Keywords: desiccation-tolerant, fern spores, germination, LEA protein, *Osmunda*

Introduction

Spores of several ferns are used as model systems for investigating cellular differentiation in plants. Some species of fern have chloroplasts in their spores. Very few studies have focused on the spore specific chloroplast proteins. We reported that five proteins in the soluble fraction of chloroplasts and three proteins in the thylakoid membranes in chloroplasts from green spores of the fern *Osmunda japonica* decreased during spore germination (Inoue et al. 1995). The 22-kDa protein, whose abundance decreases in thylakoid membranes during germination over 48 h, has been purified in the previous study and partial amino acid sequences were determined (Inoue et al. 2000). Amino-terminal short sequences of proteolytic fragments suggest that the protein may have similar features to a cDNA corresponding to cold hardening-induced *Chlorella* gene (Joh et al. 1995). The specific role of the 22-kDa protein in chloroplasts remains presently unknown. In this study, we tried to elucidate all the amino acid sequence of the 22-kDa protein by cloning of a cDNA encoding the protein, as by way of the study on the role of abundant proteins in chloroplasts of fern green spores.

Materials and methods

Sporophylls of *Osmunda japonica* were collected in southwestern Toyama in April 1999 and the spores were harvested and stored at -20 °C.

Total RNA was extracted from quiescent spores using a QuickPrep Total RNA Extraction Kit (Pharmacia Biotech). Polyadenylated RNA was prepared using Oligotex-dT30 <Super> mRNA Purification Kit (TaKaRa), following the manufacture protocol. RT-PCR was performed using Gene Amp RNA PCR Kit (Perkin Elmer). 3'RACE and 5'RACE reaction was performed using Marathon cDNA Amplification Kit (Clontech). Four degenerate primers were synthesized according to partial amino acid sequences of polypeptide fragments of the 22-kDa protein (Inoue et al. 2000). DNA sequence was obtained using Thermo Sequenase Cy5.5 dye terminator cycle sequencing kit and Gene Rapid (Amersham Pharmacia Biotech).

Results

Figure 1 shows the DNA sequence and deduced amino acid sequence of cDNA for the 22-kDa protein. The coding region starts with ATG at nucleotide 286 and ends at nucleotide 873. Downstream of the open reading frame there is a polyadenylation signal AATAAA, 18 bp upstream of the poly A sequence. The precursor protein of the protein is composed of 196

amino acids including the chloroplast signal peptide with 16 amino acids and the mature protein with 180 amino acids. The mature protein of the 22-kDa protein is a basic protein that is rich in Ala (19%), Gly (16%) and Lys (10%). The deduced amino acid composition is near with the experimental value reported previously (Inoue et al. 1995). The calculated molecular mass of the mature protein is 17,791 Da in discrepancy with the molecular mass of 22 kDa estimated by SDS-PAGE. This difference could be caused by over estimation in SDS-PAGE of the protein with a high isoelectric point as reported for ABA-induced proteins (Hong et al. 1988). The isoelectric point of the mature protein deduced from cDNA was estimated at 9.9, and this agrees with the experimental isoelectric point value of about 10.

A search of the protein database using BLAST program revealed significant similarities to a number of homologs relating to group 3 LEA proteins (Table 1). The higher significant proteins were hypothetical proteins that were found in open reading frames in chromosomes of *Deinococcus radiodurans*, *Drosophila melanogaster* and *Caenorhabditis elegans*, but not genes of plants. Next, some plant genes including cDNA corresponding to cold hardening-induced *Chlorella* genes, a LEA protein of *Glycine max* and a putative protein of *Arabidopsis thaliana* followed these.

Discussion

The LEA proteins are found in seeds during maturation and desiccation and known to express in plant tissues exposed to several stresses such as desiccation and cold. Most LEA proteins show biased amino acid compositions, rich in Ala and Gly and lacking Cys and Try. The 22-kDa protein possesses these amino acid characteristics. The LEA proteins include five distinct families. The group 3 proteins contain repeated motifs 11 amino acids long with the consensus sequence TAQAAKEKAXE. The group 3 LEA protein shows that Ala and Thr predominate in positions, 2, 4, 5 and 9. Positions 3, 7, and 11 are occupied principally by either negatively charged residue or Gln. Positions 6 and 8 prefer positively charged residues (Dure III et al. 1989). For instance, the deduced amino acid sequence of pHV A1 encoding a mRNA rapidly induced by ABA

```

CGT TGA TTT GGT GGT TAG GCG CTG CAA TCG AGT TGC CTC ATC ANG
GCA GCC TCC GCA ACT GCA GCT CAG TCC CTC ACT GCT AGT GCT CCT GCT ACT GCA ATC CGC
GAT ACC TCG TCC TTA TCT TGC TCC TCC TCC CCC TTT GTT GCA TAT ACT GTT CAC GCT GCA
CCC TTG AAA TTC CAC CAG GCA ACA ACT AAG GCC AGG CTT GTG GCA GCA AAA TGC AGT CTT
TCT GAG GAC CAG CTCTCT GGC AAG GCT GCA GCT GAT CGA CGG CAA GTG TTA GCA GGG CTA
ATG GCA GCC ACT GCA GCA TTG GCT GTG TCT TCC GAT TTA GCT TTG GCC GCA TCT GGA CCT
1 M A A T A A L A V S S D L A L A A S G P 20
GGC ATT AAT ATT GGG GCA CAA GGA ACA GCT GAG AAG GCC AAC GAT CTC CTC AAA GGC GCT
21 G I N I G A Q G T A E K A N D L L K G A 40
GCA GAC AAG CTC AAC ATT CAA GAT GCA CCT AAA CGG TTT GGC CCA GGC ATT CCAGGT GCC
41 A D K L N I Q D A P K R F G P G I P G A 60
GCA AAA GAT GCA ACG AAG GTT GCT CAG AAA GCT GGC AGT GGA GCC ATC GGG GAC TTG CAA
61 A K D A T K V A Q K A G S G A I G D L Q 80
GCA GGG GCC ACG GAT GTC ACC AGG CAG GCC CGT CAG AAT GTC GAA GAC ACA GCT CGAAGG

```



```

81 A G A T D V T R Q A R Q N V E D T A R R 100
   ACT GGT GGG CTA TTT GGA AAT GCC AAG GAT AAT GCT GGC GGT GTT GCT GGC AAC GTG CAA
101 T G G L F G N A K D N A G G V A G N V Q 120
   GAC GGT GCC AAG AAC ATC CTA GGT CAG GTG CAG GGC AAA AAG GAT GAC GCC AGA AGTGCA
121 D G A K N I L G Q V Q G K K D D A R S A 140
   GCA AAA GAC GTC GTA GGC CAA GCG CAG AGC AAA GCG AAC GAC GGG AAA GGAGTGTGGGC
141 A K D V V G Q A Q S K A N D G K G V L G 160
   TCT ATC CAG GAA GGC ATT CAG GGC TTC ACG GAG AGC GTC GCT CAG AAG GGT TCC GATGCA
161 S I Q E G I Q G F T E S V A Q K G S D A 180
   AAG ACT GCA GCC GAA GGA GCT GCA AGC AAT GCT CAA AAG GCA CTG CAG TGA TAT GAT ACT
181 K T A A E G A A S N A Q K A L Q STOP    196
   GAA CCA ATA GGT AGG GCA ATC TTA ATT TGT AAC GCG GCT CGA TGG AAT TCG CCA AGC GAG
   CTC CGA GAG ATT TGT GTA TAA CTA CTC TCA CAA GGA AAA ACA AAA AGG CGA GTG TGA CCG
   TCA ATT CCC CAC CCG TCC GTG GAT TTG TGT ATA ACT ACT CTC ACA AGG AAA AGC TTT GTG
   GCT TTA CCG AGT ACT GTA TAC ATA TTC ACC GAG ACG AAT AAA ACT GCT TTT GAG AAA AAA
   AAA AAA AAA AAA AAA AAA AAA

```

Fig. 1. DNA sequence and deduced amino acid sequence of cDNA of the 22-kDa protein. The signal peptide with 16 amino acid is underlined and amino acids are numbered. The polyadenylation signal, AATAAA, is also underlined.

Table 1. Database similarities to the 22-kDa mature protein having 180 amino acids.

Protein	Length	Expect	Identities	(%)
Hypothetical protein of <i>Deinococcus radiodurans</i>	298	9e-10	50/183	(27%)
CG9682 gene product of <i>Drosophila melanogaster</i>	456	2e-08	54/162	(33%)
Ce- LEA of <i>Caenorhabditis elegans</i>	733	4e-08	48/186	(25%)
LEA homolog of <i>Chlorella vulgaris</i>	178	2e-05	36/156	(23%)
LEA protein of <i>Glycine max</i>	458	2e-04	40/168	(23%)
Putative protein of <i>Arabidopsis thaliana</i>	347	4e-04	34/106	(32%)

in barley aleurone layers contains nine repeats 11 amino acids long and an arrange such as AKEK with two positive amino acids (Hong et al. 1988). The 22-kDa protein has slightly these essential characteristics, because the pair of lysine is not observed in the amino acid sequence. Although it is an imperfect motif, the 11-mer motifs containing AKD and occupying by nonpolar amino acid in positions 1, 2, 5 and 9 are found out in four places. Though most LEA proteins express in cytoplasmic location, some LEA homologs induced by

cold acclimation have transit peptides that target proteins to the stromal compartment of chloroplasts (Lin and Thomashow 1992, Joh et al. 1995).

Raghavan and Kamalay (1993) analyzed two spore specific clones from cDNA library of the fern *Onoclea sensibilis*. The predicted amino-acid sequence of pOSS194 has a striking similarity to the early light-inducible proteins expressed during the greening of etiolated pea and barley seedlings. The sequence of pOSS68 (pir: S29941) shows some homology to proteins encoded by LEA mRNA of angiosperm embryos. The predicted amino-acid sequence of pOSS68 is a fragment with 90 amino acids long. Because significant similarity was not observed between the 22-kDa protein from spores of *Osmunda japonica* and pOSS68 fragment from spores of *Onoclea sensibilis*, the pOSS68 protein is different from this 22-kDa protein. *Deinococcus radioduran* is a Gram-positive and radiation-resistant bacterium. It is known that functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation (Mattimore and Battista 1996). The 22-kDa protein may be a novel type LEA homolog relating to a desiccation-tolerant protein that distributes to bacteria, plants and invertebrate animal widely.

References

- Dure III L, Crouch M, Harada J, Ho TD, Mundy J, Quatrano R, Thomas T, Sung ZR (1989) *Plant Molecular Biology* **12**, 475-486.
- Hong B, Uknes SJ, Ho TD (1988) *Plant Molecular Biology* **11**, 495-506.
- Inoue H, Takano A, Asano Y, Katoh Y (1995) *Physiologia Plantarum* **95**, 465-471.
- Inoue H, Kamachi H, Yamaya D, Oguma M, Noguchi M (2000) *Physiologia Plantarum* **109**, 129-136.
- Joh T, Honjoh K, Yoshimoto M, Funabashi J, Miyamoto T, Hatano S (1995) *Plant Cell Physiology* **36**, 85-93.
- Lin C, Thomashow M F (1992) *Plant Physiology* **99**, 519-525.
- Mattimore V, Battista J R (1996) *Journal of Bacteriology* **178**, 633-637.
- Raghavan V, Kamalay JC (1993) *Planta* **189**, 1-9.