

## ▶ PLANT GENE REGISTER PGR99-134

**Clay Carter and Robert W. Thornburg** (1999) Characterization of a Gymnosperm-like Germin-like Protein (GLP7) Gene from *Arabidopsis thaliana* (Accession No. AF170550). (PGR99-134) *Plant Physiol.* **121**: 312

### **Characterization of a Gymnosperm-like Germin-like Protein (GLP7) Gene from *Arabidopsis thaliana* (Accession No. [AF170550](#))**

Clay Carter and Robert W. Thornburg\*

Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011

\*Corresponding Author; FAX : (515) 294-0453;  
Email [thorn@iastate.edu](mailto:thorn@iastate.edu)

## ▶ REPORT

Germin-like proteins (GLPs) are a large family of glycoproteins that share identity with wheat germin, a homo-oligomeric protein, consisting of 26 kDa subunits. Germin is primarily found in germinating embryos of cereal plants (Lane, 1993) and has oxalate oxidase activity (Lane, et al., 1993; Dumas, et al., 1993). Oxalate oxidase catalyzes the oxidative breakdown of oxalate into CO<sub>2</sub> and H<sub>2</sub>O<sub>2</sub>. A number of *in planta* roles for germin have been proposed. These include participation in cell wall restructuring during embryo growth, certain biotic and abiotic stress responses (e.g. fungal pathogens, salt, heavy metals) and as a possible receptor for microbial adhesin proteins (Lane, 1993; Hurkman and Tanaka, 1996; Smit, et al., 1992).

GLPs have been identified in all species of plants examined to date going as far back as the gymnosperms. The reported gymnosperm GLPs are expressed in embryogenic cell cultures of *Pinus caribaea* and *P. radiata* (Domon, et al, 1995; Bishop-Hurley, et al., 1998), and preliminary southern blot evidence indicates that the gymnosperm GLP PcGER1 is probably unique in the pine genome (Neutelings, et al., 1998). This is in contrast with the broad divergence of GLPs among the angiosperms.

Recently a large family of GLPs containing at least 27 genes were identified in *Arabidopsis thaliana* (Membré, et al., 1997, Carter, et al., 1998, Sato, et al., 1998, Carter and Thornburg, 1999, in press). Cladistic analysis of the deduced protein sequences reveal that 25 of the 27 *Arabidopsis* GLPs fall into one of three major subfamilies (Carter and Thornburg, 1999, in press). The two outliers are At-MAC9.4 (Nakamura, 1999) and At-GLP7 (Carter, et al., 1998).

Little information is available about the At-MAC9.4 GLP, however, there are other GLPs that share identity with At-GLP7. These sequences that share high identity with At-GLP7 are the gymnosperm GLPs, Pc-GER1 (Domon, et al., 1995) and Pr-GLP1 (Bishop-Hurley, et al., 1998). When all complete GLP sequences are compared, only GLP7 falls into the same clade as these gymnosperm GLPs.

Because angiosperms are generally thought to have originated from some primitive gymnosperm, it is especially

interesting to determine whether the *Arabidopsis* GLP7 is an evolutionary remnant that is important in the evolution of germin like proteins. To further investigate this question, we have isolated and characterized the GLP7 gene from *Arabidopsis* and report it here.

A partial *Arabidopsis thaliana* GLP7 cDNA (GenBank Accession No. [U75202](#)) had been previously isolated and characterized (Carter, et al., 1998). An *Arabidopsis thaliana* genomic library was kindly provided by Dr. Daniel Voytas (Voytas, et al., 1990). This library was screened with the partial GLP7 cDNA and five positively hybridizing plaques were purified to homogeneity. One lambda phage that cross-reacted strongly with the GLP7 cDNA was mapped with a battery of restriction enzymes. From this analysis a 5.4 kb EcoRI fragment was chosen for isolation and sequencing. This fragment was subcloned into the plasmid pUC19 to form the vector pRT458.

The insert of the pRT458 vector was sequenced and found to contain the GLP7 gene. Sequencing reactions were performed using the Applied Biosystems Prism Dye-deoxy Cycle Sequencing Kit. The reactions were run on an Applied Biosystems Prism 377 DNA sequencer (Perkin-Elmer Corp.). Sequencing of the insert was initiated from known vector sequences. Based upon these sequencing runs, additional sequencing primers were designed to extend new sequence. Sequence rounds were continued until the gene was completely sequenced. The entire sequence of the GLP7 gene was independently confirmed on the opposite strand. DNA sequences were performed in duplicate or triplicate for each run.

The single reported GLP7 cDNA was truncated in the 5' end. Comparison of the cDNA with the gene sequence analysis reveals that the cDNA lacks 55 N-terminal amino acids. The GLP7 gene encodes the entire 217 amino acid protein which has a pI of 6.95 and a Mr of 23,323 Da. Amino acids 1 to 17 are predicted to encode a classical signal sequence that should result in extracellular targeting of the GLP7 protein. Amino acids 19 to 217 encode the mature protein with a pI of 6.17 and a Mr of 21,399 Da. A single site of N-glycosylation occurs at Asn(75). Like the GLP1 (Carter and Thornburg, 1998) and GLP3b (Carter, et al., 1998) genes, the GLP7 gene does not contain an intron. The wheat germin genes also lack introns (Lane, et al., 1991). This lack of introns is unusual for most *Arabidopsis* GLP genes. Most *Arabidopsis* GLP genes contain an intron located near amino acid 40 (Carter and Thornburg, unpublished data). The expression patterns of the GLP7 gene are currently being investigated in transgenic plants.

Journal Paper No. J-18540 of the Iowa Agriculture and Home Economics Experiment Station Station, Ames, Iowa.

## ► **TABLE I**

### **Characteristics of the GLP7 gene from *Arabidopsis thaliana***

#### **Organism:**

*Arabidopsis thaliana* var Landsberg erecta

#### **Gene Product:**

Germin-like Protein 7 (GLP7) related to oxalate oxidase from cereals

#### **Characteristics of the GLP7 gene:**

The sequenced DNA fragment was a 5.4 kb EcoRI fragment containing the entire GLP7 gene. This GLP7 gene was also recently identified on a Chromosomal BAC T10024 (GenBank # [AC007067](#)) sequenced by the *Arabidopsis thaliana* Genome Center (Shinn, et al., 1999). That work indicates that the gene is located on chromosome 1.

**Promoter**

3192 nucleotides of the promoter are located in this EcoRI fragment [1..3192]. A putative TATA box having (10/13) to canonical plant TATA box (Joshi, 1987) was identified [3067..3079]. No significant repeats, palindromes, or hairpin loops were detected.

**Coding region**

650 nucleotides [3193..3843] terminates with an ochre codon (TAA) [3844..3846]. The GLP7 gene does not contain an intron.

**3' untranslated region**

260 nucleotides [3844..4102] were examined. There is a single consensus poly A signal [3962..3967]; however, this is quite distant (135 nt) from the polyA addition site at [4102]. There is no obvious polyadenylation signal in the vicinity of the polyA site that is utilized in the single cDNA that has been isolated ([U75202](#)). There is a single hairpin loop [ACTCAATTC-C-GAATTGAGT] within the 3' untranslated region [3990..4007], but no significant palindromes or repeats were detected.

**3' flanking region**

1335 nucleotides [4103..5437] showed no significant repeats, palindromes, or hairpin loops within the 3' flanking region.

**Differences between Landsberg Erecta and Columbia:**

Only a single nucleotide difference was identified as position 3966 of the genomic sequence where an A was identified in the genomic Landsberg clone and a T was identified at this position in the Columbia cDNA. This single nucleotide difference occurs in the 3' untranslated region of the cDNA and alters a potential poly A adenylation site (AATAAA) that occurs in Landsberg Erecta and is altered (AATATA) in Columbia.

## LITERATURE CITED

**Bishop-Hurley SL, Walter C, Gardner RC** (1998) Isolation and expression of abundant mRNAs during somatic embryogenesis of *Pinus radiata*. (Genbank Accession [AF049065](#)).

**Carter C, Graham R, Thornburg RW** (1998) *Arabidopsis thaliana* contains a large family of germin-like proteins: Characterization of cDNA and genomic sequences encoding 12 unique family members. *Plant Mol. Biol.* **38**:929-943

**Carter C, Thornburg RW** (1998) Cloning and characterization of the *Arabidopsis thaliana* Germin-like Protein 1 gene ([AF090733](#)). *Plant Physiol.*

**Carter C, Thornburg RW** (1999) Germin-like proteins: Structure, Phylogeny and Function. *J. Plant Biol.* (In Press)

**Domon J-M, Dumas B, Lainé E, Meyer Y, Alain D, David H** (1995) Three glycosylated polypeptides secreted by several embryogenic cell cultures of pine show highly specific serological affinity to antibodies directed against the wheat germin apoprotein monomer. *Plant Physiol.* **108**:141-148

**Dumas B, Sailland A, Cheviet JP, Freyssinet G, Pallett K** (1993) Identification of barley oxalate oxidase as a germin-like protein. *C.R. Acad Sci III* **316**:793-798

**Hurkman WJ, Tanaka CK** (1996) Germin gene expression is induced in wheat leaves by powdery mildew infection. *Plant Physiol* **111**:735-739

**Lane BG** (1993) Oxalate, germin and the extracellular matrix of higher plants. *FASEB Journal* **7**:294-301

**Lane BG, Bernier F, Dratewka-Kos E, Shafai R, Kennedy TD, Caron P, Munro JR, Vaughan T, Walters D, Altomare F** (1991) Homologies between members of the germin gene family in hexaploid wheat and similarities between these wheat germinals and certain *Physarum* spherulins. *J Biol. Chem.* **266**:10461-10469

**Lane BG, Dunwell JM, Ray JA, Schmitt MR, Cuming AC** (1993) Germin, a protein marker of early plant development is an oxalate oxidase. *J. Biol.Chem.* **268**:12239-12242

**Membré N, Berna A, Neutelings G, David A, David H, Staiger D, Vásquez JS, Raynal M, Delseny M, Bernier F** (1997) cDNA sequence, genomic organization and differential expression of three *Arabidopsis* genes for germin/oxalate oxidase-like proteins. *Plant Mol. Biol.* **35**:459-469

**Nakamura Y** (1999) Structural analysis of *Arabidopsis thaliana* chromosome 5 IV. GenBank Accession [AB010069](#)

**Neutelings G, Domon J, Membre N, Bernier F, Meyer Y, David A, David H** (1998) Characterization of a germin-like protein gene expressed in somatic and zygotic embryos of pine (*Pinus caribaea* Morelet). *Plant Mol Biol* **38**:1179-1190

**Sato S, Kaneko T, Kotani H, Nakamura Y, Asamizu E, Miyajima N, Tabata S** (1998) Structural analysis of *Arabidopsis thaliana* chromosome 5. IV. Sequence features of the regions of 1,456,315 bp covered by nineteen physically assigned P1 and TAC clones. *DNA Research* **5**:41-54

**Shinn P, Brooks S, Buehler E, Chao Q, Dunn P, Khan S, Kim C, Walker M, Altafi H, Araujo R, Conn L, Conway AB, Gonzalez A, Hansen NF, Huizar L, Kremenetskaia I, Lenz C, Li J, Liu S, Luross S, Rowley D, Schwartz J, Toriumi M, Vysotskaia V, Yu G, Davis RW, Federspiel NA, Theologis A, Ecker JR** (1999) Genomic sequence for *Arabidopsis thaliana* BAC T10O24 from Chromosome 1. (GenBank Accession [AC007067](#)).

**Smit G, Swart S, Lugtenberg BJJ, Kijne JW** (1992) Molecular mechanisms of attachment of *Rhizobium* bacteria to plant roots. *Mol. Microbiol* **6**:2897-2903

**Voytas DF, Konieczny A, Cummings MP, Ausubel FM** (1990) The structure, distribution and evolution of the Ta1 retrotransposable element family of *Arabidopsis thaliana*. *Genetics* **126**:713-721

[Plant Gene Register Index](#)

[Plant Physiology ONLINE](#)

© 1999 American Society of Plant Physiologists