

Plant Gene Register

Nucleotide Sequence and Putative Regulatory Elements of a Nodule-Development-Specific Member of the Soybean (*Glycine max*) Chalcone Synthase Multigene Family, *Gmchs 7*¹

Shinji Akada, Shain-dow Kung, and Shyam K. Dube*

Center for Agricultural Biotechnology and Department of Botany, University of Maryland, College Park, Maryland 20742

Certain plant flavonoids identified in exudates of the roots of leguminous plants are highly effective inducers of *nod* genes. In soybean (*Glycine max* L. Merr) the isoflavone daiz-ein has been identified as a naturally occurring inducer of *Bradyrhizobium japonicum nod* genes (Long, 1989). Estabrook and Sengupta-Gopalan (1991) have demonstrated that a gene encoding CHS, a key enzyme of flavonoid biosynthesis, is specifically induced during nodulation of soybean. They isolated RNA from nodules and sequenced a partial cDNA clone (cEUC2) encoding exon 2 of CHS (Estabrook and Sengupta-Gopalan, 1991). On the other hand, in the course of characterizing the soybean *chs* multigene family, we found a family member, gene 7, having a sequence almost identical with that published for exon 2 and the 3' flanking region of cEUC2 (Fig. 1). Thus, we call gene 7 a nodule-development-specific member of the soybean *chs* family.

A comparison of the nucleotide sequence of the protein-coding region of gene 7 with those of genes 1, 3, and 4 (Akada et al., 1991, and refs. therein) shows that the former has a substantially lower similarity to the latter three genes, which among themselves have >98% similarity (Table I). It is also interesting to note that the similarity of gene 7 to bean and kudzu vine *chs* genes (Ryder et al., 1987; Nakajima et al., 1991) is considerably higher than to genes 1, 3, and 4 at both nucleotide and amino acid sequence levels (Table I).

Gene 7 has a protein-coding exon 1 of 178 bp, an intron of 445 bp, and an exon 2 of 992 bp, giving rise to an open reading frame of 1170 bp encoding a 389-amino acid polypeptide. The size of the intron is larger than that of genes 1, 3, or 4 by 323 to 324 bp, and the size of exon 2 is larger by 3 bp. Various putative regulatory elements are found in the 5' flanking sequence of gene 7 (Table I). This may imply that the regulation of gene 7 may be quite sophisticated and responsive to a variety of abiotic and biotic stresses.

ACKNOWLEDGMENT

We thank Professor Hahlbrock for a gift of parsley cDNA clone.

Received September 30, 1992; accepted November 24, 1992.

Copyright Clearance Center: 0032-0889/93/102/0321/03.

The GenBank accession number for the sequence reported in this article is M98871.

LITERATURE CITED

- Akada S, Kung SD, Dube SK (1991) The nucleotide sequence of gene 1 of the soybean chalcone synthase multi-gene family. *Plant Mol Biol* 16: 751–752
- Estabrook EM, Sengupta-Gopalan C (1991) Differential expression of phenylalanine ammonia-lyase and chalcone synthase during soybean nodule development. *Plant Cell* 3: 299–308
- Gilmartin PM, Sarokin L, Memelink J, Chua N-H (1990) Molecular light switches for plant genes. *Cell* 2: 369–378
- Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res* 15: 6643–6653
- Loake GJ, Faktor O, Lamb CJ, Dixon RA (1992) Combination of H-box [CCTACC(N)₂CT] and G-box (CACGTG) *cis* elements is necessary for feed-forward stimulation of a chalcone synthase promoter by the phenylpropanoid-pathway intermediate *p*-coumaric acid. *Proc Natl Acad Sci USA* 89: 9230–9234
- Long SR (1989) Rhizobium-legume nodulation: life together in the underground. *Cell* 56: 203–214
- Mcknight S, Tjian R (1986) Transcriptional selectivity of viral genes in mammalian cells. *Cell* 46: 795–805
- Nakajima O, Akiyama T, Hakamatsuka T, Shibuya M, Noguchi H, Ebizuka Y, Sankawa U (1991) Isolation, sequence and bacterial expression of a cDNA for chalcone synthase from the cultured cells of *Pueraria lobata*. *Chem Pharm Bull* 39: 1911–1913
- Ryder TB, Hedrick SA, Bell JN, Liang X, Clouse SD, Lamb CJ (1987) Organization and differential activation of a gene family encoding the plant defence enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol Gen Genet* 210: 219–233
- Tsukaya H, Ohshima T, Naito S, Chino M, Komeda Y (1991) Sugar-dependent expression of the *CHS-A* gene for chalcone synthase from petunia in transgenic *Arabidopsis*. *Plant Physiol* 97: 1414–1421

¹ Partially supported by the Maryland Agricultural Experiment Station.

* Corresponding author; fax 1-301-314-9075.

Abbreviations: CHS, chalcone synthase; *chs*, gene encoding CHS.

Table I. Characteristics of *Gmchs7* from soybean

Species and Cultivar:

Glycine max (L.) Merr, cv Williams.

Gene:

Gmchs7, encoding gene 7 of the soybean CHS multigene family.

Cloning:

A genomic library of *G. max* cv Williams cloned in a bacteriophage λ EMBL-3 was initially screened at a low stringency of hybridization with a parsley cDNA probe for CHS. One positive clone, λ Gmw15, was further characterized by sequence analysis. A 5'-gene-specific probe derived from this clone was found to hybridize with a 15-kb *Hind*III fragment on a Southern blot of genomic DNA. The nucleotide sequence of a 2688-bp *Mbo*I-*Taq*I DNA fragment (Fig. 1) derived from λ Gmw15 was found to contain the entire coding sequence as well as 796-bp 5' and 276-bp 3' flanking sequences.

Techniques of Sequencing:

The 2688-bp *Mbo*I-*Taq*I DNA region of λ Gmw15 was digested with several restriction endonucleases to generate overlapping subclones in bacteriophage M13. The nucleotide sequence of both strands was determined by the dideoxy chain termination method.

GC Content:

36.7% overall; 46.8% in the protein-coding region.

Codon Usage:

Third base frequency: T>A>C>G. Codon preferences: AAG for Lys, AAT/C for Ile, CCA/T for Pro, GCA/T/C for Ala.

Nucleotide and amino acid sequence similarity (%):

	Nucleotide Sequence					Amino Acid Sequence				
	<i>Gmchs1</i>	<i>Gmchs3</i>	<i>Gmchs4</i>	Bean	Kudzu	<i>Gmchs1</i>	<i>Gmchs3</i>	<i>Gmchs4</i>	Bean	Kudzu
<i>Gmchs7</i>	81.6	81.6	80.6	90.7	96.2	90.2	89.9	90.0	95.4	97.9

Putative regulatory elements in the 5'-flanking region of *Gmchs7*:

Element	Sequence	Distance Upstream of ATG	Reference
		<i>bp</i>	
TATA box	TATATAT	-114	Joshi, 1987
H box	ACCTACCC	-143	Loake et al., 1992
G box	CACGTG	-157	Gilmartin et al., 1990
CCAT box	CCAAT	-304	Mcknight and Tjian, 1986
Sugar box 2	ATGATGtAg(t) ₁₁ aAAaaaGAt	-496	Tsukaya et al., 1991
Sugar box 3	gAATCAataaatTTA	-577	Tsukaya et al., 1991
GT-1 box	TTAACC (reversed)	-624	
	GGTAAA	-646	Gilmartin et al., 1990
AT-1 box	cATTATTaTTAag	-650	
	TAT(A) ₅ TAATT (reversed)	-727	Gilmartin et al., 1990