cDNA and Genomic DNA Clonings of Chalcone Synthase from Pueraria lobata

Osamu Nakajima, Masaaki Shibuya, Takashi Hakamatsu, Hiroshi Noguchi, Yutaka Ebizuka,* and Ushio Sankawa

Faculty of Pharmaceutical Sciences, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113, Japan.
Received June 19, 1995; accepted September 6, 1995

Four cDNAs encoding chalcone synthase (CHS), the key enzyme in flavonoid biosynthesis, were isolated from Pueraria lobata cells challenged with yeast extract elicitor using bean CHS cDNA as a probe. The longest clone contained a complete open reading frame of 1170 bp which would predict a protein of about 43 kDa. The others were not full-length clones. Using isolated cDNA as a probe, Southern blot hybridization of genomic DNA fragments revealed the presence of multiple CHS genes in the P. lobata genome. We cloned and sequenced one CHS genomic clone, gCHS14, whose 5' untranslated region showed homology with the bean CHS gene CHS15 and included the several reported sequences characteristic of stress response.

Key words chalcone synthase; cDNA cloning; genomic cloning; DNA sequence analysis; cis-acting elements; elicitor

Leguminosae plants produce characteristic isoflavono- noid phytoalexins against stress or microbial infection.1) Chalcone synthase (CHS; EC 2.3.1.74), which catalyzes the formation of chalcone from p-coumaroyl and malonyl CoAs, is the first pathway-specific enzyme in flavonoid biosynthesis.2) Therefore, CHS is regarded as the meta- bolic regulatory point in isoflavonoid phytoalexin bio- synthesis. We have been studying isoflavonoid production using Pueraria lobata cells, and found that Pueraria lobata cells differentially responded to yeast extract, CuCl2 and endogenous elicitor prepared by the hydrolysis of P. lobata cell wall fraction with a fungal endopolygalacturonase, and produced different isoflavonoids3,4); we also found that the transcript of CHS increases upon stimulation by yeast extract.5) Recent studies have demonstrated that transcriptional activation is a major factor regulating flavonoid biosynthesis in response to external stimuli, including UV irradiation and fungal elicitors and during developmental stages.6-8) Isolating the genomic CHS gene is indispensable for investigating the regulatory mechanism of flavonoid biosynthesis. As part of our continuing studies on isoflavonoid biosynthesis of P. lobata, we cloned cDNA and the genomic DNA of the CHS gene from the cultured cells of P. lobata.

MATERIALS AND METHODS

Plant Material and Elicitor Treatment Cell cultures of Pueraria lobata were maintained as previously described.3) Autoclaved yeast extract was added to a final concentration of 0.33 mg/ml. Cells were harvested 5 h after elicitation by filtration, frozen in liquid nitrogen and stored at -80 ºC until use.

Standard Recombinant DNA Techniques These were essentially as described by Maniatis et al. (1989).9) pBlue- script II (Stratagene), pUC18 or pUC118 (Takara) were mainly used for subcloning.

Isolation of Poly(A)+ RNA Total RNA for cDNA cloning was isolated from 10 g of frozen cells following the reported method.10) Poly(A)+ RNA was purified by using an oligo(dT)-cellulose column (Pharmacia).

* To whom correspondence should be addressed.

cDNA Cloning Procedure cDNA was synthesized with 6 µg of poly(A)+ RNA using a cDNA Synthesis Kit (Boehringer Mannheim). Double-stranded cDNA was fractionated through Sephacryl S-1000 (Pharmacia). cDNA longer than 0.5 kb was recovered, an EcoRI adapter (Amersham) was attached, then the whole was phosphorylated and ligated to αgt10 arms (Stratagene). The resulting DNA was packaged (Gigapack II Gold, Stratagene). E. coli strain C600Hfr was infected with the packaging mixture. A library containing 2×105 clones was screened using bean CHS cDNA.7)

Isolation of Genomic DNA Genomic DNA was isolated from P. lobata cultured cells following the reported method10) starting with 40 g of frozen cells.

Isolation of CHS Genomic Clones Genomic DNA isolated from P. lobata cells was partially digested with Sau3A1 and ligated into the BamHI site of EMBL 3. 2×105 clones were screened with 1.5-kb EcoRI fragment derived from pICHIS I as a probe. 19 Positive clones were obtained and analyzed by Southern blot hybridization.

Southern Blot Hybridization Genomic DNA (10 µg), digested with restriction enzymes, was separated by electrophoresis in 0.8% agarose gel. The DNA were transferred to Gene Screen Plus (Du Pont) and hybridized with 32P-labeled cDNA probes generated by the random primer method9) using full-length cDNA or upstream from BgII (position at 321) of cDNA (cDNA corresponding to the region from -96 to 321) as a template. Hybridization and washing were carried out according to the recommendation for Gene Screen Plus. Hybridization was performed in 50% formamide, 1 M NaCl, 1% sodium dodecyl sulfate (SDS), 10% Dextran sulfate and 15 µg/ml of denatured salmon sperm DNA. Hybridization was continued for 24 h at 42 ºC after 2 h preincubation in the same buffer without the hybridization probe. Prior to autoradiography, blots were washed twice in 2×SSC, 0.1% SDS for 15 min and twice in 0.2×SSC, 0.1% SDS for 15 min at 42 ºC.

DNA Sequence Analysis DNA sequence analysis was carried out by the dideoxy chain-termination method9) on double- or single-stranded DNA using a Hitachi DNA
seqencer, SQ-3000, and the sequencing kit for Sequenase Ver. 2.0 (Toyobo). Helper phage M13KO7 and E. coli strain MV1184 were used routinely.9)

RESULTS

Elicitor-Induced CHS cDNA Clones A λgt10 cDNA library was constructed with poly(A)+ RNA from P. lobata cells challenged with yeast extract elicitor to stimulate isoflavonoid biosynthesis. From approximately 2×10² clones, 4 clones were isolated using bean CHS cDNA as a probe. The positive phages were subcloned into pUC18 and designated as pPICHIS I, II, III and IV. The longest cDNA clone, pPICHIS I had an insert 1455 bp

\[
\begin{align*}
\text{TCTTTCTACTAGTAAAGTTACGTGCTAG} \\
\text{CTGTGACTTGTGTGATCGGATCAAAATCGACGATACTTTGATATCATATCATCTACATACTCAAATTGAAAGAGAAA} \\
\text{ATGACTGATGGTGTGATCGGAAAGCTGATCTCTAATCGGAAAGAGAGATACTGACATGACCACT} \\
\text{GCAAACCCCAACAACTTGTTGTAGCAGGACACCATCTACACGAGGCAAGGCGACCCCAAGCATAAATCTGACTGAC} \\
\text{ATGACACGAGCTCAAGAGAAATCTCCAGGCGCTGGTGAACAGGCTTACATGATCGACTGACAGGAC} \\
\text{MTELKEKFPQRCMDCKSMIMKRYML} \\
\text{ACCGGAGAAGCTTGATCAGGACCAACACATGTTGATCAATGAGCCACTCCTTCTTGGGAGTTCAGGAGACAC} \\
\text{TSEEILKEMCNMCAYMPSLPDRQD} \\
\text{ATGACTGATGGTGTGATCGGAAAGCTGATCTCTAATCGGAAAGAGAGATACTGACATGACCACT} \\
\text{GCAAACCCCAACAACTTGTTGTAGCAGGACACCATCTACACGAGGCAAGGCGACCCCAAGCATAAATCTGACTGAC} \\
\text{MVVVVEVPKLGKEAAATKIAKEWQQ} \\
\text{AAGCTCAAGATTACCCCACTCTTGATTTGCACCAAGAGTTGTGATGACATGACAGGAC} \\
\text{KKQLQTRVYKRMQYQCGAPG} \\
\text{TGACATGACAGGACACTTCTGATGACAGGAC} \\
\text{TVLRILAENKGNFARVLLVCSE} \\
\text{ATCAGCTCAGTCAGTACCTGTCGAGCAACTCTCATTTGATAGTTCGAGCTCAGTAGATTTGGCGAAGCATTGTTTGG} \\
\text{1TAVTFRGPSDTHLSDLVQALFG} \\
\text{GATGACAGTCAGTACCTGTCGAGCAACTCTCATTTGATAGTTCGAGCTCAGTAGATTTGGCGAAGCATTGTTTGG} \\
\text{DCAGAAAVIVGSDPQVEKPLYELV} \\
\text{TGACTGCACAAACAACTGTCGAGAAATCCAGGCACTACAGATGTGCTATGAGTCATGCGTAAA} \\
\text{TAAQTIAPSDGSAIDGHRLREVGLT} \\
\text{TTCATCTCCTTTGATCAGGACACATGTTGCTTCAAGAAACCTTTTGGAGTTCATGACCTAC} \\
\text{FHLKLDVPGISKNIKDALKFRAFN} \\
\text{CCACGTAAACCTTGTGATTTGACACCTCTTACCCAGCTGCGCGTGCAATTTTGGACCAC} \\
\text{PLNISYDSNFIAHPGPGPAIQLQ} \\
\text{GTGACAGTCAGTACCTGTCGAGCAACTCTCATTTGATAGTTCGAGCTCAGTAGATTTGGCGAAGCATTGTTTGG} \\
\text{VQKLGKPMKATRDVLSDYG} \\
\text{TGACTGCACAAACAACTGTCGAGAAATCCAGGCACTACAGATGTGCTATGAGTCATGCGTAAA} \\
\text{MSSACVLPILDERMRRKASENGLKT} \\
\text{ACAGGATCAAGAGCTTACATGGTGGTGGGTAGTTGCTTGTGTCGAGCTACCTGCTGACATGTTTGGCG} \\
\text{TGEGLEWGVLPGFPGLTIETVVL} \\
\text{CTGACTGTGGCGCTTGAGGCACTATTATATTATATATTATATTATACCTACATTTCACGTTTGGCG} \\
\text{ACAGGGATGACATGGTGGTGGGTAGTTGCTTGTGTCGAGCTACCTGCTGACATGTTTGGCG} \\
\text{RSVAY} \\
\text{TGAGATTGTAACCCACACCAACCAACCAACAAAAAACACTCGTTTTAGAGTTGTTGTGCTAGTTTATATTAA} \\
\text{ATAATCATAGTTGGGCCCTTTTATAGCTTAATATAACGATCATTTTTGAAAGAAAA} \\
\end{align*}

Fig. 1. Nucleotide and Deduced Amino Acid Sequences of pPICHIS I

Nucleotides are numbered with the proposed translational start point designated as 1. The asterisk indicates the stop codon.
in length and contained a complete open reading frame of 1170 bp encoding 389 amino acids which would predict a protein of about 43 kDa (Fig. 1). Within the translated regions, the *P. lobata* and *P. vulgaris* CHS cDNA sequences shared 90.6% homology at the nucleotide sequence level and 93.4% homology at the amino acid sequence level without deletions or insertions.

The other clones, pPICHS II, III and IV were 5’ truncated clones (625 bp, 838 bp and 474 bp long, respectively). Their sequences were completely identical to that of pPICHS I, except that they lacked parts of the sequences in the 3’ untranslated regions just before poly(A) tails, by 18 bp, 15 bp and 8 bp, respectively. It is not clear whether the 4 cDNAs were derived from different mRNA species.

### Genomic Organization of CHS

The result of Southern blots analysis is shown in Fig. 2. Since full-length CHS cDNA of *P. lobata* hybridized to several fragments of genomic DNA digested with *Eco*RI, *Hind*III or *Bam*HI, respectively, CHS seems to be encoded by a gene family in the *P. lobata* genome. In the analysis of *Hind*III digestion, the intensity of a 20-kb fragment exhibited double or triple the intensity of 4.5, 2.5 and 2.0-kb fragments. If it is assumed that each of these weak bands (4.5, 2.5, 2.0-kb fragments) corresponds to 1 copy of the CHS gene, the number of copies of CHS genes in the *P. lobata* genome could be estimated to be about 6—7. This seems to be a reasonable estimation, since isolated genomic clones were made into more than four groups (described below) and bean CHS genes are composed of 6—8 copies.

### Cloning and Sequence of Genomic Clones

*A P. lobata* genomic library was screened using the *P. lobata* CHS cDNA as a probe. We cloned 19 positive clones from about 2 × 10⁶ recombinant EMBL3 phages and designated them gCHS1-19. The polymorphism of these clones was analyzed by Southern blot hybridization. All clones revealed different patterns due to partial digestion by *Sau*3A at the construction of a library, but were made into more than 4 groups (data not shown). Among the positive clones, gCHS1 and gCHS14 hybridized particularly strongly to both a full-length and partial length (upstream from 321) of cDNA. Expecting that gCHS1 and gCHS14 should be inducible CHS genes responding to the elicitor stimuli, we further analyzed gCHS1 and gCHS14. The fragment of 3.6-kb *Eco*RI—*Sal*I digestion from gCHS1, and 3.8-kb and 4.3-kb *Eco*RI—*Bgl*II fragments from gCHS14, which hybridized to CHS cDNA, were subcloned into pUC18 and these fragments were sequenced. The sequences of the first exons and some of the introns of gCHS1 and gCHS14 were completely identical, and the sequences of the coding region of both clones perfectly matched that of cDNA (data not shown). Thus, we concluded that gCHS1 and gCHS14 originated

---

**Fig. 2.** Southern Blot Hybridization Analysis of CHS Sequences in the *Pueraria lobata* Genome

Genomic DNA (10 μg/lane) prepared from *P. lobata* cells was digested with *Eco*RI (lane 1), *Hind*III (lane 2) or *Bam*HI (lane 3). The fragments were separated by 0.8% agarose gel electrophoresis, and blotted prior to hybridization with a 32P-labeled full-length CHS cDNA.

---

**Fig. 3.** Restriction Maps and Sequencing Strategies for gCHS1 and 14

The solid box, the open box and the line in subclone indicate the translated region, intron and the untranslated region, respectively. Arrows indicate the directions and lengths of the determined sequences. Restriction enzymes are *Apa*I; *Bgl*I; *Bgl*II; *Eco*RI; *Sau*3A; *Sal*I; *Sst*I. Sp1, Spe1.
from the same CHS gene and were thus inducible genes. Since gCHS1 lacked the 5' untranslated region, the 5' untranslated region of gCHS14 was sequenced (Fig. 3).

**Promoter Sequence** The nucleotide sequences of the 5' untranslated region of *P. lobata* gCHS14 and bean CHS15 are shown in Fig. 4. Nucleotides were numbered with the translational start point as 1. Since the sequence which extended from -171 to -98 in gCHS14 is highly homologous to that of the bean CHS promoter from -52 to 18, which includes a TATA box, we deduced that the TATA box is located from -145 to -142, though it is still possible that the TATA box is located from -149 to 147 or from -147 to -144. The sequence which extended from -399 to -355 in gCHS14 is also highly homologous to the sequence found in the bean CHS15 promoter.

Distinct sequences were found in the 5' untranslated region of gCHS14, which include ACTCACCTACC located at -195 to -184, and CCACAAACTTC located at -295 to -285. These are also observed in the bean CHS15 promoter and correspond to the conserved motifs in the promoters of several genes involved in phenylpropanoid metabolism referred to as (T/A)CT(C/A)ACTT- (C/A)(C/A) and CCA/A(C/A)CTTC/C/TCC) by R. Lois et al.

The sequence, CCAATTATTTGCTATGTAATGAA

Fig. 4. Alignment of 5' Untranslated Sequence of gCHS14 with Bean CHS15 Promoter

Nucleotides of gCHS14 are numbered with the translational start point as 1. The sequence of bean CHS promoter are from ref. 11. Hyphens are inserted to maximize homology. Identical nucleotides are denoted by horizontal bars. Putative cis-acting elements referred to in the content are underlined.
located at −379 to −355 in the 5′ untranslating region of gCHS14 is very similar to the sequence, GATCAC-
CAATTATTGTTACCTAAATCTAACCGT, which was thought to form a conserved or overlapping silencer and
activator in the bean CHS15 promoter, and is defined as box III.  

DISCUSSION

pPICHSI had 90.6% homology at the nucleotide
sequence level to bean CHS cDNA, which was proven to be inducible CHS cDNA in response to the stimuli.  
And as previously shown by northern blot analysis,  
CHS mRNA was markedly induced by yeast extract elicitation in P. lobata cells. Therefore, it could be
concluded that the cDNA, pPICHS I isolated from P. lobata cells challenged with yeast extract, was derived from the
inducible transcript in response to the elicitation.

Southern blot analysis showed that the genome of
P. lobata cells contains multiple CHS genes which might be
responsible for the stress response or the biosynthesis of
color pigment, or might be silent genes. The presence of
several CHS genes is the same as the case in bean, where
the haploid genome of bean contains about 6–8 copies
of CHS genes.  
We screened the genomic library using
pPICHS I as a probe, and isolated 19 clones. Two of them
were strongly hybridized to the probe. Since one of them,
gCHS1 lacks the upstream sequence of the initiation
codon, the other clone, gCHS14, was analyzed. The
sequences of the first exon and the 5′ untranslated region
of gCHS14 matched perfectly to those of CHS cDNA,
pPICHS I. This is one of the bases suggesting that gCHS14
is responsible for the stress response.

In Leguminosae plants, the promoter sequences of the
CHS genes were cloned from bean and soybean,  
but they have no overall homology. The 5′ untranslated
region of P. lobata gCHS14 is more homologous to the
bean CHS15 promoter than the soybean CHS promoter,
but does not have high overall homology. However, they
have several highly homologous sequences of 10–30 bp
in length, and a few of them have been characterized
(Table 1).

The presumptive TATA box is located from −145 to
−142 in the 5′ untranslated region of P. lobata gCHS14.
This is surrounded by a sequence of about 30 bp which is
highly homologous to the sequence around the TATA
box in the bean CHS15 promoter.  
Conserved motifs found in the promoters of several
genes involved in phenylpropanoid metabolism were
reported by R. Lois et al.  
These conserved motifs were assumed to be the binding sites of regulatory factors
which promote gene activation as responses to stress
stimuli such as UV light or an elicitor.  
For example, the conserved motifs were revealed by in vivo
footprints in the parsley PAL-1 promoter.  
And in the bean CHS15 promoter, the motifs were located in the region which
included an elicitor-regulated activator.  
Similarly, in the case of snapdragon CHS promoter, a minimal light
responsive element contained the conserved motifs and
deletion analysis showed the importance of the motif
(TACCAT).  
Also in the 5′ untranslated region of P. lobata gCHS14, there are the related sequences,
ACTCACCTACC located at −195 to −184, and
CCACGAACTCG located at −295 to −284 corresponding to the conserved motifs. These sequences are
considered to function as elicitor-regulated activators in
the promoter of P. lobata gCHS14.

The role of the sequence ACCAATTATTGTTAC-
TAACTTAA was demonstrated by deletion analysis in
the bean CHS15 promoter and defined as box III.  

<table>
<thead>
<tr>
<th>Table 1. List and Properties of Presumptive cis-Elements Found in the 5′ Untranslated Region of P. lobata gCHS14</th>
</tr>
</thead>
</table>
| **(I)** Conserved motif
| Found in the promoters of several genes involved in phenylpropanoid metabolism; reference 14.
| Presumptive binding sites of regulatory factors promoting gene activation as response to stress stimuli.
| **(i)** 3′ Portion
| Consensus sequence
| T C C C
| CT ACCTA C
| A A A A
| **P. lobata gCHS14**
| ACTCACCTACC (located at −195 to −184)
| **(ii)** 5′ Portion
| Consensus sequence
| A A C C
| CCA C C AAC CC
| C T T T
| **P. lobata gCHS14**
| CCACGAACTCG (located at −295 to −284)
| **(II)** Box III
| Identified in the promoter of bean CHS15; reference 13.
| Conserved, or overlapping silencer and activator.
| **Bean CHS15**
| ACCAATTATTGTTACTAATTTAA
| **P. lobata gCHS14**
| CCCAATTATTGCTTAGTAAATGAAA (located at −379 to −355)
| **(III)** G-box
| Found in many plant promoters; reference 19.
| Binding site of a family of basic domain/leucine zipper transcription factors.
| **Consensus sequence**
| CACGTG
| **P. lobata gCHS14**
| ATGCACGTGATA (located at −206 to −195)
| **(IV)** H-box
| Found in the promoters of many stress-inducible phenylpropanoid biosynthetic genes; reference 19.
| Necessary for stimulation of the bean CHS15 promoter by p-coumaric acid combined with G-box.
| **Consensus sequence**
| CCTACC(N)₂CT
| **P. lobata gCHS14**
| CCTACC(N)₂CG (located at −190 to −176)
III was supposed to form a consecutive or overlapping silencer and activator. Also the box III is the binding site of SBF-1, which is identical to or closely related to GT-1.\textsuperscript{18} In the 5' untranslated region of \textit{P. lobata} gCHS14, a similar sequence, \texttt{CCTACCTTTGATGTAATGGA}, located at −379 to −355, is observed. This sequence probably acts as a silencer and activator in a similar manner in the promoter of \textit{P. lobata} gCHS14.

Combination of the H-box (CCTACC(N)\textsubscript{2}CT) and the G-box (CAGTGG) cis-elements has been reported to be necessary for feed-forward stimulation of the bean CHS15 promoter by \textit{p}-coumaric acid.\textsuperscript{19} The G-box is found in a wide variety of plant promoters such as snapdragon and parsley CHS genes, and in genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase,\textsuperscript{20} and they bind to a family of basic domain/leucine zipper transcription factors.\textsuperscript{21} The H-box is also found in the promoters of a number of stress-inducible phenylpropanoid biosynthetic genes.\textsuperscript{14} Mutation of box II (CCACGTGGCC), located in the parsley CHS promoter containing the G-box, results in a loss of UV light-responsive expression.\textsuperscript{22} The sequence related to the G-box, ATGCACCCTGATA, and the sequence related to the H-box, CCTACCTCTCCATCG, can be found in the 5' untranslated region of \textit{P. lobata} gCHS14 at positions −206 to −195, and −190 to −176, respectively.

As described above, the 5' untranslated region of gCHS14 contains several sequences characteristic to the stress response found in the bean CHS promoter. Thus, we consider that gCHS14 is the elicitor-responsive CHS gene in \textit{P. lobata} cells. This is supported by the fact that the 5' untranslated region of gCHS14 functions as a stress responsive promoter in the hairy root of tobacco (unpublished data).

As bean and \textit{P. lobata} are very close among Leguminosae plants and the homology between the CHS cDNAs of bean and \textit{P. lobata} is very high (90.6% at the nucleotide level), we had expected that the promoter regions of their genomic CHS genes should show high similarity. But the promoter regions of their genomic CHS genes did not show much similarity. They have only several highly homologous sequences of 10—30 bp in length. It is not clear whether only these short sequences play the key role in the regulation of response to a stimulii. It is possible that other sequences also take part in the regulation. To make this clear, it is indispensable to do an element-analysis.

\textit{P. lobata} cells mainly produce 5-deoxy-type isoflavonoids.\textsuperscript{23} Recently it was demonstrated that isoliquiritigenin, a deoxy-type chalcone, is synthesized by the co-action of CHS and reductase (CHR).\textsuperscript{24-26} It is very interesting how the expression of the two enzymes are regulated, being induced concomitantly by elicitors. The cloning of genomic CHR genes is in progress in our laboratory.

**Acknowledgment** The authors thank Dr. C. J. Lamb of the Salk Institute for providing us with cDNA for the CHS of \textit{P. vulgaris}; Dr. Y. Ozeki of the College of Arts and Sciences, The University of Tokyo, for helpful suggestions; and Dr. T. Nishimura of the Faculty of Medicine, The University of Tokyo, for many valuable suggestions.

**REFERENCES AND NOTES**

10. The nucleotide sequences were submitted to DDBJ, EMBL, and GenBank nucleotide sequence databases. The accession number of pPiCHS and gCHS14 are D01023 and D63855, respectively.