

The nucleotide sequence of gene 3 of the soybean chalcone synthase multigene family

S.Akada, S.D.Kung and S.K.Dube*

Center for Agricultural Biotechnology and Department of Botany, 1206 H.J.Patterson Hall, University of Maryland, College Park, MD 20742, USA

Submitted August 7, 1990

EMBL accession no. X53958

Chalcone synthase (CHS) plays a central role in the biosynthesis (1) of a class of secondary plant metabolites, called flavonoids (2). In soybean, *Glycine max* (L) Merr., the CHS genes comprise a multigene family (3) and the nucleotide sequence of one member of the family has been reported (4).

We have now determined the nucleotide sequence of another member of the CHS multigene family, presently designated as gene 3 (3). The sequence of 4047 nucleotides presented in Fig. 1 contains the entire coding sequence of gene 3 and a large part (2523 nt) of the intergenic region connecting this gene with another member of the family reported earlier (4). Several features in the sequence are noteworthy: 1) the gene contains 2 exons, the first having 178 nt and the second 989 nt, 2) the splice sites of the intron determined by GT/AG rule (5) are between nt 413 and 414 and between nt 535 and 536 giving an intron of 122 nt, 3) the putative TATA box is 98 nt upstream

from the first exon, 4) the region between the termination of the coding sequence of gene 3 and that of the other member of the family deduced from our previous work (4) spans 2683 nt.

ACKNOWLEDGEMENT

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

REFERENCES

1. Hahlbrock, K. and Grisebach, H. (1979) *Ann. Rev. Plant Physiol.* **30**, 105–130.
2. Cody, V., Middleton, E., Harborne, J.B. and Beretz, A. (eds) (1988) *Flavonoids in Biology and Medicine*. A.R. Liss, New York.
3. Wingender, R., Röhrig, H., Hörigke, C., Wing, D. and Schell, J. (1989) *Mol. Gen. Genet.* **218**, 315–322.
4. Akada, S., Kung, S.D. and Dube, S.K. (1990) *Nucl. Acids Res.* **18**, 3398.
5. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* **50**, 349–383.

```

1  aacatagtcac tgcagtgatc aatatagttg agagaaaaca cagaacacag ccaattcggt agaggaaaca tgctcatcat ctactcagta ctcacctacc cacttcaagt tcaactgtct
121 atctattcat atatatatcc ccacccttcc aaaccacttt gcaacatcca tccaagcctt tcttttccca gctactacac ttctattctt tgccttcagaa aattaactag ctaggATGGT
241 CAGTGTGTGA GAGATCCGTA ATGCACAACG TGCAGAGGGC CCTGCCACTG TCATGGCTAT TGGCACCACA ACTCCTCCAA ACTGTGTGCA TCAGAGTACC TATCCTGACT ATTATTCCG
S V E E I R N A Q R A E G P A T V M A I G T A T P P N C V D Q S T Y P D Y Y F R
361 SATCACCAAC AGCGAGCACA TGACCGAGCT CAAAGAAAAC TTCAAGCGCA TGTgtaagat atatatctct ctcctttctt catttcttta tacaatatgt atattgctta ttttcaaat
I T N S E H M T E L K E K F K R M
481 attcctttga tttgattagt gatattaatg aaatttaatt tattattttc atcagGTGAT AAGTCAATGA TTAAGAAGCG ATACATGTAC TTAATGAAG AAATCCTGAA AGAGAATCCG
601 AGTGTGTTG CTTATCGGC ACCTTGCTTG GATGCAAGGC AAGACATGGT GGTGTGGAG GTACCAAAAG TGGGAAAAGA GGCTGCAACT AAGGCAATCA AGGAATGGGG TCAACCCAA
S V C A Y M A P S L D A R Q D M V V V E V P K L G K E A A T K E I K E N P
721 TCCAGATTA CCATCTCAT CTTTGCACG ACTAGTGGTG TCGACATGCC TGGTGTGAT TATCAGCTCA CTAAACTATT AGGCCTTCGC CCCTCCGCTCA AGCGTTACAT GATGTACCA
S K I T H L I F C T T S G V D M P G A D Y Q L T K L L G L R P S V K R Y M M Y Q
841 CAAGGCTGCT TTGCCGGTGG CACGGTGCTT CGTTTGGCCA AAGACCTCGC TGAAAACAAC AAGGGTGCTC GCGTGTCTGT CGTTTGTCTT GAGATCACCG CAGTCACATT CCGCGGCCA
Q G C F A G G T V L R L A K D L A E N N K G A R V L V V C S E I T A V T F R G P
961 ACTGACACCC ATCTTGATAG CTTTGTGGTG CAAGCCTTGT TTGGAGATGG TGCAGCCGCT GTCATTTGTT GATCAGACCC CTTACCACTT GAAAGCCTT TGTTCAGCT TGTCTGGACT
T D T H L D S L V G Q A L F G D G A A A V I V G S D P L P V E K P L F V L W T
1081 GCCCAACAA TCCTTCCAGA CAGTGAAGGG GCTATTGATG GACACCTTGG CGAAGTTGGT CTCATTTCCT ATCTCTCTCA GGATGTTCCT GGACTTCAT CCAAGATTAT TGAGAAGGCC
A Q T I L P D S E G A I D G H L G E V G L T F H L L K D V P G L I S K N I E K A
1201 TTGGTTGAAG CCTTCCAACC CTGGGAATC TCCGATTACA ATTCTATTCT CTGGATTGCA CACCCTGGTG GACCCGCAAT TTGGACCAA GTTGAGGCTA AGTTAGGCTAT TGAAGCTGAA
L V E A F Q P L G I S D Y N S I F W I A H P G P A I L D Q V E A K L G L K P E
1321 AAAATGGAAG CTACTAGRA TGTGCTCAGC GAGTATGGTA ACATGTCAAG TGCATGTGTG CTATTCTACT TGGATCAAA GAGGAAGAAA TCAATAGAAA ATGGACTTGG CACAACCGT
K H E A T Y R H V L S E Y G N M S S A C V L F I L D Q M R K K S I E N G L G T T G
1441 GAAGGCTTGG ACTGGGTGTG GCTATTGTGT TTCGGCCCTG GACTCACCGT TGAGACTGTT GTGCTCCGCA GTGTCTCTGT Ctaatccata tatcttctgtc aagaccaagt ccttttcttt
E G L D W G V L F G F G P G L T V E T V V L R S V T V
1561 cttattcatt ctttcattgt gagtttgaag aatgtattct ttctcttctt ttttttccct gcttccctca catgcttctt tttttatata agtaataaag aacatgttat attactatta
1681 atgcaattaa gttttcacaa aattcatcta tatattcagt tgcgaagtga atatctcttg tccaatttca tctttattga aagattcttt gtgacttgct ataaaagtga ctgtagaact
1801 ttgcaggttt atggcttaac tgtgtaaaga gaatagtgcc aattaacgta gtccacttag gcttaatttt tttttttttt ttttgtgttg attgagttat ctaactgtga ctgtgtgaga
1921 aaaaattatt aaaaaagga ttaataagta tcaagtaaac ctatagtgta agaaattgca ctctctacct tgattactta atgaacaaaa tctaagtcga caaagtgggt ggaactgctt
2041 tgcattatt tgccttaca gagttaacta ctaaccaca atagataaca tttatccatg actctgtatt aacactttaa aataaatggt tatgagttgt agtatgggct
2161 ctaaggttag attttattgt tcagtttgta tcatgctctc agcttatggt gtgcatcaat atgtgaaatc tattactcaa tttgtttcag tttgtattac attttcattg tgtgcttttg
2281 atttaataca tgaataatca taacgcagtt tatcacttta ggtagttttt atcatattcc agtttatctt gtgcttctgg ttgagtatca gattgcagtt tttatcagat tgtagtttat
2401 ggtgtgcttc tgattcaata tgaataattt gtgacacaat ttgtttcagt ttttatctta ttgtaattta tgatttggtt ctatctgtgt atcatattag ttttgttgtt cgggtgatgg
2521 tgtgtctgga aatttataac aatctcatgg ggtgccaat ttgctctgct gctctggcca tatgattcta atcatgggat atcatgttat tctcgtttat ggtgtgtcaa
2641 ttttttttta cgtctgtacc aatttgattc ttttttaggg gccttggttt atgtgaaat ctgacacaa ttaggagacat ttgttagttt gtttcagcca tggagttttg tgaagtatta
2761 ataccaacca atttacttat gtaatgagca ttttaaaacc gaatgacctt ttaaaacctt aatattttga caacataaaa gcataatttt gatactctat gatactctat ttgtcattag tttcattagt caataacctg
2881 cacaaccaag ttgtaattct tatcacacc tttaaactctg acaactata aatatctagt aaacttagta gctaaagaaa tgcacttcta tcttgattac ttaagaattt aaatctagt
3001 tgggacaaaa agtatgttca ccatcttgga ttattgattt tacaattttg taactcagtt aaaaatgttc attgttactt ttgcaataca atcatgactc atctgttctt tattaggagc gggggattat ccatcaacct
3121 accaggtata gtataggctc caattgttag attagttagt attttctgat aaaaatgttc attgttactt ttgcaataca atcatgactc atctgttctt tattaggagc gggggattat ccatcaacct
3241 aatatgataa aaacataact atcatattat tgaagataat caatagttag tggcccaata ggtcaaaaaa ttaacccata tctgtgaagt ttactgaagt ataatatag atagcaagac tttcacttag tattaggagc gggggattat ccatcaacct
3361 caagaataac aggaataatg ttggttggtc aataacaatt tctttaattc gatgtgggac caacataaaa ttaacccata tctgtgaagt ttactgaagt ataatatag atagcaagac tttcacttag tattaggagc gggggattat ccatcaacct
3481 gacgttaaat gaagatctgc caagttttat ttaaacacca tctttaattc ccttactgaa tctacataat ttttcttact gcggtttact tgatacttat tgaatctgtt catataatgt tgtatatcac agtcaaatca
3601 gataggactg acggcaatta aggtggtttt ccaacatatt ttgatacttt ttttttactt caataatccc aacactttcc tagtatagat caattgtgta taagctt
3721 gatgacgtag gaagtgcatt ttcttagcta ctaggtttac ttgatacttt ttttttactt gcggtttact tgatacttat tgaatctgtt catataatgt tgtatatcac agtcaaatca
3841 ctacacaaag acagaaattt ctggactaa ctgaattgaa ttaattgtcca atactctctt caataatccc tagtatagat caattgtgta taagctt
3961 tcttttgttg gagatgaatt tggtaaaatt aaaaataga atattcactt tgcactgtga tagtatagat caattgtgta taagctt

```

* To whom correspondence should be addressed