

Nucleotide sequence of one member of soybean chalcone synthase multi-gene family

S.Akada, S.D.Kung, S.K.Dube*
Center for Agricultural Biotechnology, 1206 H.J.Patterson Hall, University of Maryland, College Park, MD 20742, USA

Submitted April 27, 1990 EMBL accession no. X52097

Flavonoids constitute a class of secondary plant metabolites (1) and chalcone synthase (CHS) by catalysing the formation of a key intermediate, naringenin-chalcone, plays a central role in their biosynthesis (2). In soybean, Glycine max (L) Merr., CHS genes comprise a multigene family (3). The family is composed of a minimum of six members as estimated by Southern hybridization of soybean genomic DNA digested with HindIII, which cuts outside the coding region (Fig. 1). The probe for hybridization was parsley cDNA (4).

We have determined the nucleotide sequence (Fig. 2) of one family member contained within a 2,317 bp HindIII fragment marked by an arrow in Fig. 1. 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 1047 and 1048 and between nt 1168 and 1169, giving an intron of 121 nt. 3) The putative TATA box is 115 nt upstream from the first exon. 4) The putative light responsive element (6) is 351 nt upstream from the first exon. 5) The nucleotide sequence of the protein coding region has 72% homology with the nt sequence of parsley CHS while the derived amino acid sequence has 85% homology with that of parsley CHS (4).

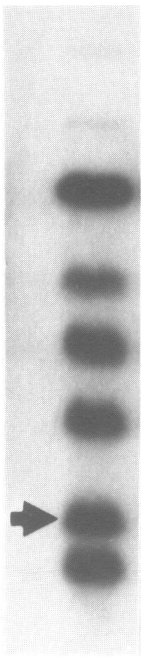

Figure 1.

```
   1  aagcttcatc acccacttat tccaaagata aagttcagtt taatcccctc ccaaaccaaa taaattatga agtagttcac agccacacat gtctataatc
 101  tcaaactaat atttatataa cacatattaa aaattattaa tttatgatta cttgattata tattacataa aaattaatat agtgtaagaa ccaagataaa
 201  tcataatcat ttaataattt ctcttcagac caacataacc acgaccagtt tctttcatga gagagaagat aagagaaaaa atgtttttca atttttttta
 301  aaaaagaatt taatattagt ctttgaaatt tttaagcacc atggaggtga aaaaaataga tatccatata atggacagga tatctgaatt gcaaaaaaat
 401  catgaatctc ttgtttaaaa acagttttat ttaaaacatt tatttttat tggaatgttt tcaagatgat aaatgagaca aatcaatcaa tcagacttgg
 501  tattaaaaac aaataatttc ctcgtgacat ttttttttc ataaacataa ctcaactaaa gaaaaaaaaa cagaaaatta aaacccggtt atttgctgat
 601  cattaggaaa agaaaaaaaa atgggttggt aagtataact ataatgggga gaatcagcgg tctacttaga catgcggtgg gtgcacacca caagcgcagt
 701  cagagaaagg aagcatgcac tgcatctacc ttaatctacc tacccacact tttctatata tatatatcca cccttccaag ccactttgca acatccatcc
 801  aagccttttc tttcgtagat agctactact tcactttcat cctttgctcc agaaaattaa ctagctaggA TGGTGAGTGT TGAAGAGATT CGTAAGGCGC
                                                                  M  V  S  V  E  E  I  R  K  A
 901  AACGTGCAGA AGGCCCTGCC ACTGTCATGG CTATTGGCAC CGCCACTCCT CCCAACTGCG TGGATCAGAG TACCTATCCT GACTATTATT TCCGCATCAC
      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  I  T
1001  CAACAGCGAG CACATGACCG AGCTCAAAGA AAAATTCAAA CGCATGTgta agatatctct ctcttttatc ctatcttcat ttcattatat aatatgcatg
        N  S  E  H  M  T  E  L  K  E  K  F  K  R  M
1101  ttgcttattt ccaacatata cctttgattt cattaatgat atcaatgaaa tttaatttat tatttcagGT GATAAGTCGA TGATTAAGAA GCGATACATG
                                                                  C  D  K  S  M  I  K  K  R  Y  M
1201  TACTTAAACG AAGAGATCCT GAAGGAGAAT CCCAGTGTTT GTGCATATAT GGCACCTTCG TTGGATGCAA GGCAAGACAT GGTGGTTATG GAGGTACCAA
      Y  L  N  E  E  I  L  K  E  N  P  S  V  C  A  Y  M  A  P  S  L  D  A  R  Q  D  M  V  V  M  E  V  P
1301  AGTTGGGAAA AGAGGCTGCA ACTAAGGCAA TCAAGGAATG GGGTCAACCC AAGTCCAAGA TTACCCATCT CATCTTTTGC ACCACTAGTG GTGTCGACAT
      K  L  G  K  E  A  A  T  K  A  I  K  E  W  G  Q  P  K  S  K  I  T  H  L  I  F  C  T  T  S  G  V  D  M
1401  GCCTGGTGCT GATTATCAGC TCACTAAACT ATTAGGCCTT CGTCCCTCCG TCAAGCGTTA CATGATGTAC CAACAAGGCT GCTTTGCCGG TGGCACGGTG
        P  G  A  D  Y  Q  L  T  K  L  L  G  L  R  P  S  V  K  R  Y  M  M  Y  Q  Q  G  C  F  A  G  G  T  V
1501  CTTCGTTTGG CCAAAGACCT CGCTGAAAAC AACAAGGGTG CTCGCGTGCT TGTCGTTTGT TCTGAGATCA CCGCAGTCAC ATTCCGCGGC CCAACTGACA
        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  T  D
1601  CCCATCTTGA TAGCCTTGTG GGTCAAGCCT TGTTTGGAGA TGGTGCAGCC GCTGTCATTG TTGGATCAGA CCCCTTACCA GTTGAAAAGC CTTTGTTTCA
      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  Q
1701  GCTTGTCTGG ACTGCCCAGA CAATCCTTCC AGACAGTGAA GGGGCTATTG ATGGACACCT TCGCGAAGTT GGTCTCACTT TCCATCTCCT CAAGGATGTT
        L  V  W  T  A  Q  T  I  L  P  D  S  E  G  A  I  D  G  H  L  R  E  V  G  L  T  F  H  L  L  K  D  V
1801  CCTGGACTCA TCTCCAAGAA TATTGAGAAG GCCTTGGTTG AAGCCTTCCA ACCCTTGGGA ATCTCCGATT ACAATTCTAT CTTCAGGATT GCACACCCTG
        P  G  L  I  S  K  N  I  E  K  A  L  V  E  A  F  Q  P  L  G  I  S  D  Y  N  S  I  F  R  I  A  H  P
1901  GTGGACCCGC AATTTTGGAC CAAGTGGAGG CTAAGTTAGG CTTGAAGCCT GAAAAAATGG AAGCTACTAG GCATGTGCTC AGCGAGTATG GTAACATGTC
      G  G  P  A  I  L  D  Q  V  E  A  K  L  G  L  K  P  E  K  M  E  A  T  R  H  V  L  S  E  Y  G  N  M  S
2001  AAGTGCATGT GTGCTATTCA TCTTGGATCA AATGCGGAAG AAATCAATAG AAAATGGACT TGGCACAACC GGCGAAGGCC TTGACTGGGG TGTGCTATTT
        S  A  C  V  L  F  I  L  D  Q  M  R  K  K  S  I  E  N  G  L  G  T  T  G  E  G  L  D  W  G  V  L  F
2101  GGTTTCGGTC CTGGACTCAC TGTTGAGACT GTTGTACTCC GCAGTGTCAC TGTCtaatca tatatattga gcaagaacac agatccttct tttcttctta
        G  F  G  P  G  L  T  V  E  T  V  V  L  R  S  V  T  V  .
2201  tgtattattg ctttttttagt ttgaaaaatg tattctttct cttttgcttt ctcacattct tctttttttgt ataccagtaa acactaaacg agaacacatc
2301  ttattattaa tgcaattaag ctt
```

Figure 2.

ACKNOWLEDGEMENT

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

REFERENCES

1. Cody,V., Middleton,E., Harborne,J.B. and Beretz,A. (eds.) (1988) *Flavonoids in Biology and Medicine*. Publ. A.R.Liss, NY.
2. Heller,W. and Hahlbrock,K. (1980) *Arch. Biochem. Biophys.* **200**, 617–619.
3. Wingender,R., Röhrig,H., Höricke,C., Wing,D. and Schell,J. (1989) *Mol. Gen. Genet.* **218**, 315–322.
4. Reimold,U., Kröger,M., Kreuzaler,F. and Hahlbrock,K. (1983) *EMBO J.* **2**, 1801–1805.
5. Breathnach,R. and Chambon,P. (1981) *Ann. Rev. Biochem.* **50**, 349–383.
6. Schultze-Lefert,P., Dangl,J.L., Becker-Andre,M., Hahlbrock,K. and Schulz,W. (1989) *EMBO J.* **8**, 651–656.

* To whom correspondence should be addressed