Nucleotide sequence of a ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit gene (rbcS) in rice

Yong Xie and Ray Wu

Section of Biochemistry, Cornell University, Ithaca, NY 14853, USA
Submitted April 8, 1988

Accession no. X07515

Previously, we described the cloning and preliminary analysis of a rice *rbcS* gene from a lambda EMBL4 genomic library of rice (*Oryza sativa*, IR26) (1). We now present a 1100 bp DNA sequence including the protein coding regions, an intron and the 5' and 3' non-coding regions of a rice *rbcS* gene using the dideoxynucleotide chain-termination procedure (2). The structural organization of this rice *rbcS* gene shares the common features of two other graminaceous plants, wheat (3) and corn (4).

```

GGTCTGGAGA TGGCAACAGS GCCACAACCC ACGAGCCCGG CGGACACCA CCGCGCGCGT TGAGCCAGCC 70
ACAAACGCCC GCGGATAGGC GCGCGCACG CGGCCAATCC TACCACATCC CGCGCTCCG CGGCTCGAGC 140
GCGGTGCCAT CCGATCCGCT GAGTTTGGC TATTATACG TACCGCGGA GCGTGTGTC AGAGAGTGA 210
TCTCAGAG TACTCGAGCA AAGAAGGAGA GAGCTTGGTG AGCTGCAGAG ATG GCC CCC TCC GTG 275
                                     M A P S V
ATG GCG TCG TCG GCC ACC ACC GTC GCT CCC TTC CAG GGC TCA AGT CCA CCG CCG GCA 332
M A S S A T T V A P F Q G S S P P P A
TGC CGT CGC CCG CCG TCC GAA CTC CAG CTT CGG CAA CGT CAG CAT GGC GGC AGG ATC 389
C R R P P P S E L Q L R Q R Q B G G R I
AGG TGC ATG CAG GTAATACCT ACTGATCCAA CACACATTCT TCTTCTTCTT CTCTCTCTTA 451
R C M Q
ACCAACATTA ACCAACAACCT CAATTATCGT TTATTCAITGAG GTG TGG CCG ATT GAG GGC ATC 514
                                     V W P I E G I
AAG AAG TTC GAG ACC CTC TCC TAC CTG CCA CCG CTC ACC GTG GAG GAC CTC CTG AAG 571
K K F E T L S Y L P P L T V E D L L K
CAG ATC GAG TAC CTA GCT CCG TTC CAA GTG GTG CCG TGC CTC GAG TTC AGC AAG GTC 628
Q I E Y L A P F Q V V P C L E F S K V
GGA TTT GTC TAC CGT GAG AAC CAC AAG TCC CCT GGA TAC TAC GAC GGC AGG TAC TGG 685
G F V Y R E N H K S F G Y D G R Y W
ACC ATG TGG AAG CTG CCC ATG TTC GGG TGC ACC GAC GCC AOC CAG GTC GTC AAG GAG 742
T M W K L P M F G C T D A T Q V V K E
CTC GAG GAG GCC AAG AAG GCG TAC CCT GAT GCA TTC GTC CGT ATC ATC GGC TTC GAC 799
L E E A K K A Y P D A F V R I I G F D
AAC GTT AGG CAG GTG CAG CTC ATC AGC TTC ATC GCC TAC AAC CCG GGC TGC GAG GAG 856
N V R Q V Q L I S F I A Y N P G C E E
TCT GGT GGC AAC TAA GCGTCATCG TCATATATAG CCTGTTTAA TTGTTTCATCT CTGATTCGAT 921
S G G N U
GATGTCACCC ACCTGTTTTC GTGAGTTCOC AGTTTGTTTC ATCGTCITTT GATTTTACCG GCGGTGCTCT 991
GCTTTTGTTC TTCTTTTCA OCTGATTCCT TCTCTGACTT GATGTAAGAG TGGTATCTGC TACGACTATA 1061
TGTGTGTTGG GTGAGGCATA TGTGAATGAA ATCTATGAAA GCTCGGCT 1110

```

FIGURE LEGEND: Sequence of a rice *rbcS* gene. The putative "CAAT" box and the putative "TATA" box are underlined. The coding sequence is 519 bp long (from 261 to 401 and 494 to 871), interrupted by a 93 bp intron. The deduced precursor protein is 173 amino acids long, including a putative transit peptide of 45 amino acids. The arrow indicates the probable site of cleavage of the transit peptide.

ACKNOWLEDGEMENTS: This work was supported by research grants RF84066, Allocation No. 3, from the Rockefeller Foundation and GM29179 from NIH.

REFERENCES:

1. Xie, Y., Peng, Z.G., Cai, Y.X. & Wu, R. (1987) *Scientia Sinica (Series B)* **30**, 706-718.
2. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
3. Broglie, R., Coruzzi, J., Lampa, G., Keith, B. and Chua, N.H. (1983) *Biotechnology* **1**, 55-61.
4. Lebrun, M., Waksman, G. and Freyssinet, G. (1987) *Nucleic Acids Res.* **15**, 4360.