

Sequence and transcript map of barley chloroplast *psbA* gene

Scott K. Boyer and John E. Mullet

Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843-2128, USA  
 Submitted July 7, 1988 Accession no. X07942

The barley chloroplast *psbA* gene and 3'-exon of *trnK* were located on a 2.19 kbp *EcoRI* DNA fragment located 7 kbp from previously mapped *psbD-psbC* genes (Nuc Acids Res. 15:5217). Both genes are transcribed toward the chloroplast DNA inverted repeat. Transcript mapping localized the *psbA* RNA 5'-end 80 nucleotides upstream from the *psbA* open reading frame (nucleotide 923) and downstream from consensus -10 and -35 transcription promoter elements (boxed). *PsbA* transcript 3'-ends were located 87, 104 and 122 nucleotides downstream from the UAA termination codon. Each transcript 3'-end is located at the base of a putative RNA stem loop structure which may function in transcription termination or to stabilize the RNA. The barley *psbA* open reading frame is highly homologous to *psbA* genes from other plants (i.e., 98.3% homologous to tobacco *psbA*). While most *psbA* gene products lack lysine, the barley *psbA* sequence predicts a single lysine at amino acid 238.

```

1  GAATTCATTTCTCATAGACTCGAATGAAAAATTCGATACCATAGTCCCCGCTACTCTCCTCATAGGATACTTACAAAAGCTCAATTTGTACTGGA
   N S F L I D T R M K K F D T I V P A T L L I G Y L S K A Q F C T G
101 TCGGGGCCTCCTATTAGTAAACCCATTGGACGGATTTATCAGATTGGGATATTCTTGATCGAATTTGGCGGATATGTAGAAATCTTTTTCATATCATGA
   S G H P I S K P I W T D L S D W D I L D R F G R I C R N L F H Y H
201 GTGGATCTCGAAAAACGGACTTTGTATCGACTAAAGTATATACTTCGACTTTTCATGGCGTAGAAGCTTTAGCTCGTAAACATAAAAAGCCGGTACGAAC
   S G S S K R T L Y R L K Y I L R L S C A R T L A R K H K S T V R T
301 TTTTATGCAACGATGGGTTCGGCATTTTGAAGAATTTTACGGAAGAAGAGCAAGTTTTTCTTGATGTCCACAAAACAACCTCTTTTCTTTC
   F M Q R L G S A F L E E F F T E E E Q V F S L M F T K T T L F S F
401 AGTGGATCACACACTGAGCCCGATTTTGGTATTTGGATATTAGGTATCAATGACCTGGTCAACCCCTTAATTAATCATAGACAAAACATAAAGCAG
   S G S H T E P Y L V F G Y R Y Q
501 GAAAGGGTGTATAAATGATCAAGAAAAAATTTTCATATTTTTCATTCTGAAATGTCCTTTTATAATAAAGAGTAGGTGAATCAACTACTAATTA
   AAAATTAGTAGAAGCTTCCCTTTTGAATAGAATTTGGCTATTTCTACATAGGGAAGTCGGTCAAGTCAAAAATGCAAGCAGCATTGGGGAGGGGATT
701 TTTCTCTATTGTAACAAGGAAGAATATCTACTCCATCCGACTAGTTTCCGGGTCGAGTCCCGGCAACCCATATAGAAAAGCCCATCAAAGTTTTTA
   l-----trnK Exon-----J
801 ACTTTTACTCCTTCATTACAATAACAAAATTTGGTTGGTAAATTTATATATGGATAGCCAATCTTTGGGCTGACTTGGTGTGACATTTGGTATAT
   901 AGTCTATGTATACCTGTTAAATAACAATCCTCTATTATCTATATTCTAGTTAATACGTGTGCTGGGAGTCTCGCAATTTGAATAAACCAAGATCTTA
   1001 CCATGACTGCAATTTTAGAGAGACGCGAAAGTACAAGCCTGTGGGGTCGCTTCTGCAACTGGATTAAGTACGACTGAAATCGCTTTACATCGGATGGTT
   M T A I L E R R E S T S L W G R F C N W I T S T E N R L Y I G W F
1101 CGGTGTTTGGATGATCCCTACCTTATTGACCGCAACTCTGTATTTATATCGCCTTCATCGCTGCCCTCCAGTAGATATTTGATGGTATTCGGAGCCT
   G V L M I P T L L T A T S V F I I A F I A A P P V D I D G I R E P
1201 GTTTCGTGCTTTACTTTATGAAAAACAATATATCTCTGGTGCATATATCTACTTCTGCGGCGATCGGATGCACTTTTACCAATTTGGGAAGCTG
   V S G S L L Y G N N I I S G A I I P T S A A I G H P I W E A
1301 CATCTGTGATGAGTGGTATACAATGGTGGTCTTATGAGCTAATGTTCTACACTTCTTACTTGGTGTAGCTGTTATATGGGTCGTGAGTGGGAAGT
   A S V D E W L Y N G G P Y E L I V L H F L L G V A C Y M G R E W E L
1401 TAGTTCCGCTCTGGGTATGCGCTTCTGGATTCGATATTGAGCTGCTGTCATATTGAGCTGCTGTCGAGCTGCTACTGCTGTTTCTTCTGATTTACCCATTTGGTCAAGGA
   S F R L G M R P W I A V A Y S A P V A A A T A V F L I Y P I G Q G
1501 AGCTTTTCTGATGGTATGCTTTAGGAATCTCTGGTACTTCAACTTTATGATTGATATCCAGGCAGACACAATCTTATGATCCATCCATCCCATGT
   S F S D G M P L I S G T F N F M I V F Q A E H N H P F H M
1601 TAGGTGATGCTGGTGTATTCGGGCTTCCCTATTCAGTCTATGATGATGCTTCCCTGGTAACCTCTAGTTGATCAGGGAAGTACTGAAATGAATCTGC
   L G V A G V F G G S L F S A M H G S L V T S S L I R E T T E N E S A
1701 TAATGAGGGTTACAAAATTTGGTCAAGAGGAGAGCACTTATAAATTTGGCTGCTCATGGTATTTTGGCCGATTAATCTTCCAATATGCTAGTATTCAC
   N E G Y K F G Q E E E T Y N I V A A N G Y F G R L I F Q Y A S F N
1801 AACTCTGCTTCTTACACTTCTTCTGGCTGCTGGCTGTAGTAGGAATCTGGTCTACTGCTTTAGGTATTAGTACTATGGCTTTCAACTAAATGGTT
   N S R S L H F L A A W P V V G I W F T A L G I N M A F N L N G
1901 TCAATTTCAACCAATCTGTAGTGTAGTAGCAAGTCCGCTTATTAATCTTGGGCTGATATCATCAACCGTGCTAACCTGGTATGGAAGTAAAGCAGCA
   F N F N Q S V V D S Q G R V I N T W A D I N R A N L G M E V M H E
2001 ACGTAATGCTCACACTTCCCTCTAGACTAGCTGCTGTTGAAGTCCAGCTTAAATGGTAAGGTTTTTCTGCTCAACATATAAGAAATTTTTGAAGAA
   R N A H N F P L D L A A V E V P A I N G
2101 GAAAAGACAAAATACCAATATCTGTCTAGCAAGATTTGGGATTTTTGAACTTTTTTCTTCAATCTTCTTACTTCAAGATTC
    
```