Nucleotide sequence of the *rpo*A gene in wheat chloroplast DNA

Sean M.Hird[1,2], Tristan A.Dyer[2] and John C.Gray[1]

[1]Botany School, University of Cambridge, Downing Street, Cambridge CB2 3EA and [2]Institute of Plant Science Research (Cambridge Laboratory), Maris Lane, Trumpington, Cambridge CB2 2JB, UK

Submitted June 29, 1989 EMBL accession no. X15595

The nucleotide sequence of a wheat (*Triticum aestivum* cv. Mardler) chloroplast gene (*rpo*A) encoding a protein homologous to the α subunit of RNA polymerase from *Escherichia coli* (1) has been determined. The gene is located at the end of the 27.2kbp *Sal*I fragment (S1) nearer the inverted repeat sequence (2), and is largely contained within the 1.7kbp *Bam*HI fragment (B20). The 3' end of the coding region is located 186bp from the 3' end of the *pet*D gene. The gene encodes a protein of 339 amino acid residues. The deduced amino acid sequence contains 29% residues identical to the *E. coli* protein (1), 47% identical to the *Marchantia polymorpha* chloroplast protein (3), 69-71% identical to the chloroplast proteins from spinach, tobacco and pea (4-6), and 86% identical to the maize chloroplast protein (7).

```
                                                                                  rpoA >>
                                                                                 M  V  R  E  V  V  A  G  S  T  Q  T  L  Q  W  K  C  V
   1 CCCGCTTCAACTTTCATTTCAAGAGAAATAAACGATTCAATGATAAAATAATAATACTAGTCTATTATGGTTCGAGAGGTGGTAGCAGGATCCACTCAAACACTACAGTGGAAGTGTGTT  120

      E  S  R  V  D  S  K  R  L  Y  Y  G  R  F  I  L  S  P  L  R  K  G  Q  A  D  T  V  G  I  A  L  R  R  A  L  L  G  E  I  E
 121 GAATCAAGAGTAGATAGTAAGCGCCTTTATTATGGTCGTTTCATTCTGTCCCCGCTTAGAAAAGGTCAAGCGGACACCGTCGGTATTGCCTTGCGAAGAGCTTTACTTGGAGAAATAGAA  240

      G  T  C  I  T  R  A  K  F  G  S  V  P  H  E  Y  S  T  I  A  G  I  E  E  S  V  Q  E  I  L  L  N  L  K  E  I  V  L  R  S
 241 GGAACATGTATCACACGTGCAAAATTTGGGAGCGTGCCGCACGAATATTCTACAATAGCAGGTATTGAAGAATCCGTACAAGAAATTTTACTAAATTTGAAAGAAATTGTATTGAGAAGT  360

      N  L  Y  G  V  R  D  A  S  I  C  V  K  G  P  R  Y  I  T  A  Q  D  I  I  L  P  P  S  V  E  I  V  D  T  A  Q  P  I  A  N
 361 AATCTCTATGGAGTTAGAGACGCATCCATTTGCGTCAAAGGTCCTAGATACATAACTGCTCAAGATATCATCTTACCGCCTTCCGTAGAAATCGTTGATACGGCACAACCTATAGCTAAC  480

      L  T  E  P  I  D  F  C  I  D  L  Q  I  K  R  D  R  G  Y  Q  T  E  L  R  K  N  Y  Q  D  G  S  Y  P  I  D  A  V  S  M  P
 481 TTGACAGAGCCCATTGATTTCTGTATTGATTTACAGATCAAGAGAGATCGTGGATATCAGACAGAACTCAGAAAGAACTATCAAGATGGAAGTTATCCTATAGATGCTGTATCCATGCCT  600

      V  R  N  V  N  Y  S  I  F  S  C  G  N  G  N  E  K  L  E  I  L  F  L  E  Y  G  S  N  G  S  L  T  P  K  E  A  L  Y  E  A
 601 GTTCGAAATGTGAATTATAGTATTTTTTCTTGTGGGAATGGAAATGAAAAACTCGAGATACTTTTTCTAGAATATGGATCTAATGGAAGTCTAACCCCTAAGGAAGCACTTTATGAGGCT  720

      S  R  N  L  I  D  L  F  L  P  F  L  H  A  E  E  E  G  A  S  F  E  E  N  K  N  R  F  T  P  P  L  F  T  F  Q  K  R  L  T
 721 TCTCGTAATTTGATTGATTTATTTCTTCCTTTTCTACACGCGGAGGAAGAGGGCGCTAGTTTCGAAGAAAATAAAAACAGGTTTACTCCACCCCTTTTTACTTTTCAAAAAAGATTAACT  840

      N  L  K  K  N  K  K  G  I  P  L  N  C  I  F  I  D  Q  L  E  L  T  S  R  T  Y  N  C  L  K  R  A  N  I  H  T  L  L  D  L
 841 AATCTAAAGAAAAACAAAAAAGGAATTCCATTGAATTGTATTTTTATTGATCAATTAGAATTGACTTCTAGAACGTATAATTGTCTAAAAAGGGCCAATATACATACACTATTGGACCTT  960

      L  S  K  T  E  E  D  L  L  R  I  D  S  F  R  M  E  D  R  K  H  I  W  D  T  L  E  K  H  L  P  I  D  L  L  K  N  K  L  C
 961 TTGAGTAAGACTGAAGAAGATCTTCTGAGAATTGACAGTTTTCGTATGGAAGATAGAAAACATATATGGGACACTCTAGAGAAGCATCTCCCAATTGATTTACTTAAGAATAAACTCTGC 1080

      I
1081 ATTTAAATCCATTGCAATTTTTTTTCTTCTTCTTTTTCGAGATAGAATAAAAAGAAAACAATTTAGCATCTTTGGTACAATCTATATTTTCGCGAAATGGATCATAATAAAATGGATTTT 1200

                                                                                                                << petD
                                                                                               F  L  G  L  T  L  S  K  D  I  P  L
1201 AGGTATCTAGGGAAGATTCACTTGGAAGTAACTATTTCCTAGATACCTATGGTACTTCACGGTTGAATGAATCAAAAATATCTAAAAAAGACCTAAAGTTAATGATTTATCAATGGGTAA 1320
```

References
1. Ovchinnikov, Y.A. *et al.* (1977) FEBS Lett. 76, 108-111
2. Bowman, C.M. *et al.* (1981) Mol. Gen. Genet. 183, 93-101
3. Ohyama, K. *et al.* (1986) Nature 322, 572-574
4. Sijben-Muller, G. *et al.* (1986) Nucleic Acid Res. 2, 1029-1044
5. Shinozaki, K. *et al.* (1986) EMBO J. 5, 2043-2049
6. Purton, S. and Gray, J.C. (1989) Mol. Gen. Genet. 217, 77-84
7. Ruf, M. and Kossel, H. (1988) Nucleic Acids Res. 16, 5741-5754