

The second largest subunit of RNA polymerase II from *Arabidopsis thaliana*

Rob Larkin and Tom Guilfoyle
Department of Biochemistry, University of Missouri-Columbia, Columbia, MO 65211, USA

Received December 31, 1992; Accepted January 5, 1993 EMBL accession nos Z19120, Z19121

Three forms of DNA dependent RNA polymerase (RNAP) exist in the nucleus and are frequently referred to as I, II, and III. The enzymes are most thoroughly studied in yeast, particularly RNAP II, which is composed of 12 subunits some of which are homologous to or shared with RNAPs I and III. Many of the characteristics of the yeast enzyme appear to be conserved among other species (1).

In plants, one and two dimensional gels (2), immunological studies (3), and isolation of genes and cDNAs encoding RNAP subunits (4, 5) largely confirm the evolutionary conservation of RNAP subunit composition. Here we report the sequence of the second largest subunit of RNAP II from *Arabidopsis thaliana* ecotype Columbia (*AtRP140*). A segment of *AtRP140* was isolated by PCR-mediated amplification of *Arabidopsis* genomic DNA using degenerate oligonucleotides derived from conserved amino acid sequences among homologous yeast, *Drosophila*, *E.coli*, and chloroplast subunits (6, 7). Subsequently, *AtRP140* was isolated on a single λEMBL3 *Arabidopsis* genomic clone. Overlapping cDNA clones constituting the entire *AtRP140* ORF and 3′ UTR were isolated from a λYES *Arabidopsis* cDNA library and by PCR-mediated amplification of first strand cDNA. Overlapping subclones constituting 9064 bp of genomic DNA were sequenced on both strands. cDNA clones were sequenced on one strand to confirm all exon/intron boundaries and the polyadenylation site. Genomic Southern blotting experiments suggest that the subunit is encoded by a single copy gene. Northern blotting experiments indicate that the subunit is expressed as a 3.8 kb mRNA.

The derived amino acid sequence predicts a protein of 1188 residues with a MW of 135,051 and a pI of 8.25. The amino acid sequence is 66, 62, 39, and 25 percent identical to the second largest subunits of *Drosophila* RNAP II (7, 8), yeast RNAP II (6), yeast RNAP III (9), and yeast RNAP I (10), respectively. Apparent functional domains that have been reported for homologous subunits are conserved in the *Arabidopsis* subunit. Examples include ribonuclease (11), nucleotide binding (6), and the zinc finger motifs (12). Other invariant residues of functional significance are also conserved. Examples include residues homologous to the *E.coli* β subunit residues K1065 and H1237 which are important for viability and promoter clearance (13).

AtRP140 contains 24 introns. The homologous genes from yeast and *Drosophila* contain no introns (6) and three introns (7, 8), respectively. The first intron of *AtRP140* is 776 bp long, while others range in size from 75 to 165 bp. The first intron contains an inverted repeat and four 92 bp direct repeats, the first of which is degenerate in the last 36 bp. A TATAAA motif was found 116 bp upstream of the translational start site. The *AtRP140* putative promoter region contains three sets of direct repeats that range from 13 bp to 31 bp. A 52 bp sequence resembling a homopyrimidine/homopurine element (14) is also present.

ACKNOWLEDGEMENTS

We thank Harry Klee for the *Arabidopsis* λEMBL3 genomic library, John Mulligan and Ronald Davis for the *Arabidopsis*, λYES cDNA library, and Craig Chandler for technical assistance. This work was supported by National Institutes of Health grant GM37950 to T.J.G.

REFERENCES

1. Young,R.A. (1991) *Annu. Rev. Biochem.* **60**, 689-715.
2. Guilfoyle,T.J. (1983) in Jacob,S.T. (ed.), *Enzymes of Nucleic Acid Synthesis and Modification*. CRC Press, Inc., Boca Raton, FL, Vol. **2**, pp. 1-42.
3. Guilfoyle,T.J. *et al.* (1984) *J. Biol. Chem.* **254**, 640-648.
4. Dietrich,M.A. *et al.* (1990) *Plant Mol. Biol.* **15**, 207-223.
5. Ulmasov,T. and Guilfoyle,T. (1992) *J. Biol. Chem.* **267**, 23165-23169.
6. Sweetser,D. *et al.* (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1192-1196.
7. Falkenburg,D. *et al.* (1987) *J. Mol. Biol.* **195**, 929-937.
8. Sitzler,S. *et al.* (1991) *Gene* **100**, 155-162.
9. James,P. *et al.* (1991) *J. Biol. Chem.* **266**, 5616-5624.
10. Yano,R. *et al.* (1991) *Mol. Cell. Biol.* **11**, 754-764.
11. Shirai,T. and Go,M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9056-9060.
12. Treich,I. *et al.* (1991) *J. Biol. Chem.* **266**, 21971-21976.
13. Mustaev,A. *et al.* (1991) *J. Biol. Chem.* **266**, 23972-23931.
14. Hoffman,E.K. *et al.* (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2705-2709.

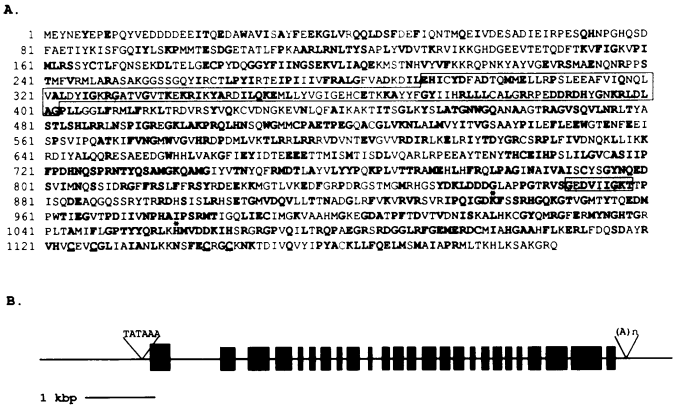

Figure 1. A. Amino acid sequence derived from *AtRP140*. Identities among *Arabidopsis*, *Drosophila*, and yeast RNAP II subunits are in *bold*. Ribonuclease similarity is *boxed*. The nucleotide binding motif is *shaded*. Residues homologous to *E.coli* β subunit K1065 and H1237 are denoted with an *asterisk* above each residue. Cysteine residues of the zinc finger motif are *underlined*. **B.** Diagram of *AtRP140*. Translated exons are indicated with black boxes. Positions of the polyadenylation site and the putative TATAAA box are indicated.