Genomic Nucleotide Sequence of a Gene from Arabidopsis thaliana Encoding a Protein Homolog of Escherichia coli RecA

Marie-Noelle Binet, Mahasin Osman, and André T. Jagendorf*

Plant Biology Section, Cornell University, Ithaca, New York 14853

In Escherichia coli and many other prokaryotes, the RecA protein is essential for homologous recombination and for a variety of S.O.S. responses to DNA damage (Little and Mount, 1982; Kowalczykowski, 1991). Previously, Cerutti et al. (1992) found a chloroplast-localized strand recombination activity resembling that of bacterial RecA and a protein band cross-reacting with antibody to E. coli RecA. In addition, they described what is known to be the first demonstration of a homolog of the bacterial recA gene in a cDNA library from Arabidopsis thaliana, which included a putative transit peptide and probably codes for the protein with those activities.

We present here the structure of the recA-At gene isolated from an A. thaliana genomic library (Table I) and corresponding to the previously described recA cDNA (Cerutti et al., 1992). The 5′ flanking region has promoter elements corresponding to presumed TATA box and CAT box sequences. The gene has 12 AT-rich introns interrupting the coding sequence. The percentage of AT nucleotides varies from 62.5% (sixth intron) to 72.8% (eighth intron). The GT and AG dinucleotides invariant at the 5′ and 3′ splices are present at the ends of each intron. The predicted size of the precursor RNA molecule transcribed from the recA-At gene is approximately 3216 nucleotides long and the splicing by cleavage of introns can produce a mature RNA of 1400 nucleotides. The first intron did not separate exactly the putative transit peptide-processing cleavage site; it came 93 bp downstream of the probable start of the mature protein. The coding sequence of the combined exons corresponds to a protein of 439 amino acids (precursor form). The first exon (323 nucleotides long) contains the 5′ noncoding region (74 nucleotides) and the beginning of the coding sequence (249 nucleotides long), including a predicted chloroplast transit peptide of 52 amino acids followed by the 31 NH₂-terminal amino acids of the mature protein. The predicted amino terminus of the mature protein (Cerutti et al., 1992) is based on sequence similarity with a conserved motif at the cleavage site for the stromal processing protease.

Table I. Characteristics of recA-At gene

| Organism: Arabidopsis thaliana, Columbia strain. | Source: Obtained from a EMBL-3 genomic library (from Clontech, Palo Alto, CA) using as a probe the 149-bp HindIII-EcoRI restriction fragment (bases 36–185, coding for the chloroplast transit peptide region) of the Arabidopsis cDNA clone recA (Cerutti et al., 1992). Four overlapping clones were isolated containing a common 6-kb SauIII fragment hybridizing with the probe used for the screening. Deletion subcloning and complete sequencing on both strands by the dideoxy-chain termination method using T7 DNA polymerase and universal or specific oligodeoxyribonucleotides as primers. | 
| Genomic Organization and Gene Copy Number: | Confirmation: Identity with cDNA sequence (GenBank accession number M98039) | 
| recA-At gene isolated from an A. thaliana genomic library (Table I) and corresponding to the previously described recA cDNA (Cerutti et al., 1992). | Features of Gene Structure: An open reading frame of 1317 bp interrupted by 12 AT-rich introns. | 
| Clone Type, Designation: | Features of Protein Structure: Open reading frame of 1317 bp encoding a protein of 439 amino acid residues (precursor form), which encodes a 52-amino acid chloroplast transit peptide and a 387-amino acid mature protein. | 
| Genomic sequence; full length (3849 bp) with 423 bp of the 5′ untranslated sequence and 297 bp of 3′ untranslated sequence; clone recA-At. | | 

1 Supported in part by grant 91-37301-6421 from the U.S. Department of Agriculture/National Research Initiative Competitive Grants Program.
2 Present address: Laboratoire de Phyto-Biologie Cellulaire, B.P. 138, 21004 Dijon Cedex, France.
3 Present address: Department of Biochemistry, Cell, and Molecular Biology, Cornell University, Ithaca, NY 14853.
4 Corresponding author; fax 1-607-255-5407.
quence (data not shown). It may correspond to a protein similar to RecA but functional in some other cellular compartment, such as the nucleus or mitochondria.

**ACKNOWLEDGMENTS**

We thank Dr. R. Last for the genomic library and numerous members of Dr. J. Nasrallah's laboratory for use of equipment and technical advice.

Received May 6, 1993; accepted May 18, 1993.

The EMBL, GenBank, and DDBJ accession number for the sequence reported in this article is L15229.

**LITERATURE CITED**

