Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 4604–4608, June 1989
Genetics

A gene encoding the tryptophan synthase β subunit of *Arabidopsis thaliana*

(higher plants/amino acid biosynthesis)

MARY B. BERLYN*†, ROBERT L. LAST, AND GERALD R. FINK

Whitehead Institute for Biomedical Research and Biology Department, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, MA 02142

Contributed by Gerald R. Fink, March 21, 1989

ABSTRACT **The DNA sequence of a tryptophan synthase gene and the flanking 5′ and 3′ regions has been determined for *Arabidopsis thaliana*. The sequence encodes only the β subunit domain, indicating that α and β subunits are specified by separate genes. The gene contains four introns and encodes 470 amino acid residues. The plant amino acid sequence is highly conserved with respect to corresponding microbial sequences. The NH_2-terminal amino acid sequence is characteristic of chloroplast transit peptides. Identity of the sequences of the genomic exons and a cDNA clone and the presence of cellular RNA corresponding in size and 5′ sequence to the gene indicate that the gene is expressed. Analysis of *Arabidopsis* genomic DNA suggests the presence of a second gene for the β subunit.**

Our understanding of the regulation of the tryptophan biosynthetic pathway comes primarily from the analysis of the expression of tryptophan synthase genes in bacteria and fungi. Tryptophan synthase [L-serine hydro-lyase (adding indoleglycerol-phosphate); EC 4.2.1.20] catalyzes the last step of the pathway, conversion of indoleglycerol phosphate to tryptophan. The enzyme contains two functional domains: tryptophan synthase α (TRPA) catalyzes the conversion of indole glycerol phosphate to indole and tryptophan synthase β (TRPB) catalyzes the conversion of indole plus serine to tryptophan. In bacteria each of these domains is encoded by a separate gene, whereas in yeast and *Neurospora* both domains are encoded by a single gene (1–5). The domains of the two bacterial genes can be aligned with the single yeast gene: the NH_2-terminal domain of *TRP5* is homologous to the *trpA* gene and the CO_2 terminal is homologous to the *trpB* gene. These homologies have prompted the speculation that the fungal gene evolved by fusion of *trpA* to *trpB*. Gene fusions have also been proposed to account for other bifunctional enzymes of the tryptophan pathway (6, 7).

Although tryptophan has a special role in plants as a precursor of the auxin indoleacetic acid, the biochemistry of tryptophan synthesis and its regulation are not well understood because the biosynthetic enzymes and the genes encoding them have not been isolated from plants. Biochemical studies on plant tryptophan synthase were not sufficient to establish whether TRPA and TRPB activities are encoded by one or two genes. The individual activities were separated during purification steps (8–10), but, as in *Neurospora* and yeast fractionations, the low activities may represent results of proteolysis, and attenuated protein fragments may provide partial activity (11–13).

We report the cloning and sequencing of an *Arabidopsis* gene for TRPB.‡ Absence of any adjacent open reading frames homologous to TRPA indicates that in plants, as in *Escherichia coli*, TRPA and TRPB subunits are encoded by separate genes. The sequence also suggests that the NH_2-terminal portion of the gene encodes a chloroplast transit peptide. Genomic hybridization provides evidence for a second gene for TRPB in *Arabidopsis*.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Cultivars, Libraries, Plasmids, and Strains. The Columbia wild-type cultivar of *Arabidopsis thaliana* was the source of plant tissue. The plasmid pYE(trp5) containing the yeast *TRP5* gene (4) was provided by C. Yanofsky (Stanford University). The *Arabidopsis* (Landsberg *erecta* ecotype) genomic library in λEMBL4 was a gift from E. Meyerowitz (California Institute of Technology) and the Columbia cDNA library in λgt10 was a gift from N. Crawford (University of California, San Diego). M13 derivatives, pUC vectors (14, 15), and pGEM-Z3 (Promega) vectors were grown in *E. coli* K-12 strains JM109 or JM103 cured of phage P1 (14) or DH5 (Bethesda Research Laboratories and ref. 16), and λ derivatives were propagated in DB1317 (17). The plasmid pMBT2 was constructed by inserting a 4.2-kb *Eco*RI fragment from the strongly hybridizing genomic phage clone λM1g into the *Eco*RI site of pUC18 (14). Plasmid pMBT3 was constructed by ligation of the 2.1-kilobase (kb) cDNA insert from λGT10 phage λCD2 into the *Eco*RI site of pUC118 (15).

Hybridization Probes and Conditions. The yeast fragment used to probe the genomic library was the *Eco*RI fragment extending from nucleotide 865–1384 of the *TRP5* gene (figure 2 of ref. 4). The probe used to screen the cDNA library was a 1.1-kb *Sac* I–*Xho* I fragment (Fig. 1, nucleotides 740–1827) of the pMBT2 plasmid. Hybridization probes were synthesized by using random hexamer primers (18). For each library, ≈80,000 phage were screened (19) on nylon filters (Biotrans, ICN). For low-stringency hybridization screening, standard formamide prehybridization solutions (20) were modified by including 25% (wt/vol) formamide and 100 mg of heat-denatured calf thymus DNA per ml. Hybridizations also included 5% (wt/vol) dextran sulfate (omitted for genomic filters) and labeled probe (5×10^6–10^7 cpm). After hybridization, filters were washed at 42°C with several changes of 25% formamide/5× SSC (1× SSC is 0.15 M sodium chloride/0.015 M sodium citrate, pH 7)/0.1% SDS.

Sequencing. A series of nested deletions derived from the 2.2-kb *Eco*RI/*Xho* I fragment (Fig. 1) cloned into M13mp19 (14), from the 4.2-kb *Eco*RI fragment cloned into pUC118 (15), and from the cDNA insert of pMBT3 were obtained by the exonuclease III method of Henikoff (21), modified by use of exonuclease VII in place of S1 nuclease. Dideoxy chain-

Abbreviations: TRPA, tryptophan synthase α subunit; TRPB, tryptophan synthase β subunit.

*On leave from Lehman College and Graduate Center, City University of New York.
†Present address: Biology Department and School of Forestry, Yale University, New Haven, CT 06511.

‡The sequence reported in this paper has been deposited in the EMBL/GenBank data base (accession no. M23872).

termination sequencing reactions were run, with modified T7 DNA polymerase (22) by use of the Sequenase protocol (United States Biochemical).

Genomic DNA and RNA Preparations. Genomic DNA and total RNA were prepared from leaves and stems of 18-day-old soil-grown plants by using the protocol of Ausubel *et al.* (20). Those protocols (20) were also used for RNA blotting, primer extension, and ribonuclease protection experiments. Ribonuclease protection probes corresponding to antisense RNA extending from (*i*) nucleotide 265 or (*ii*) nucleotide 30 to the *Ssp* I site at nucleotide −240 in Fig. 1 were synthesized by SP6 polymerase with the pGEM-Z3 plasmids (Promega).

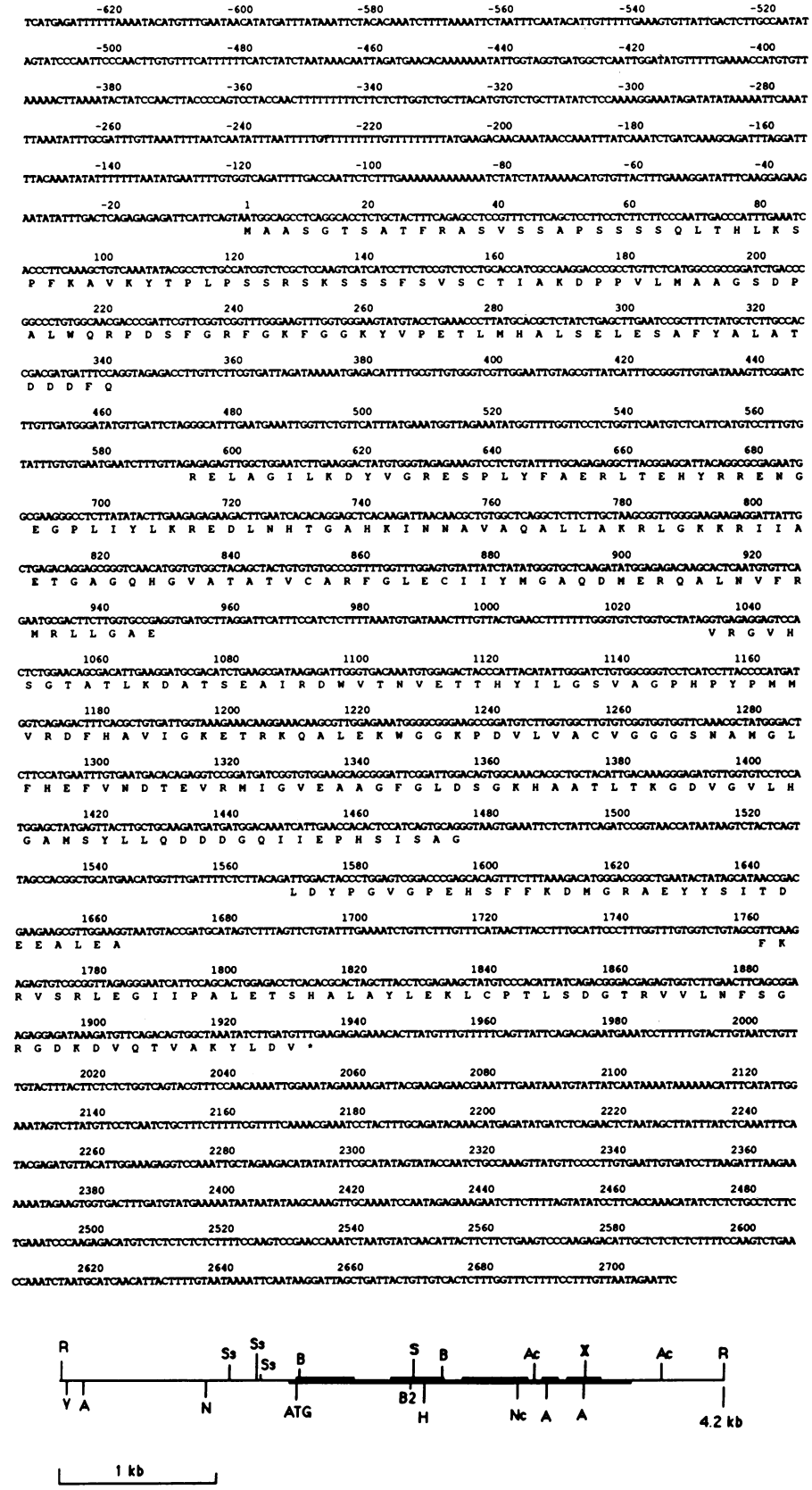

FIG. 1. (*Upper*) Nucleotide sequence of the gene and surrounding 5′ and 3′ regions. The first ATG within the reading frame is designated nucleotide 1, and the deduced amino acid sequence is shown. Nucleotide sequence was obtained for both strands of the 4.2-kb region from pMBT2. The entire sequence is available through GenBank. An 800-base pair (bp) sequence with 80 adenosines is present at the 5′ end of the cDNA clone pMBT3. This sequence appears to be a cDNA cloning artifact because corresponding sequence is not found in either phage DNA, genomic sequences 5′ to our clone, or the RNA analyses. (*Lower*) The restriction map of the 4.2-kb region. Closed bars represent exons, and open bars represent noncoding regions. ATG indicates the site of the first methionine codon. The symbols used for restriction enzymes shown are as follows: A, *Ava* I; Ac, *Acc* I; B, *Ban* I; B2, *Ban* II; H, *Hin*cII; N, *Nde* I; Nc, *Nco* I; R, *Eco*RI; S, *Sac* I; Ss, *Sst* I; V, *Eco*RV; and X, *Xho* I.

Materials. Restriction enzymes and exonuclease III were from New England Biolabs, modified T7 DNA polymerase (Sequenase) was from United States Biochemical, and exonuclease VII and Moloney murine leukemia virus reverse transcriptase were from Bethesda Research Laboratories. Random hexamers were from P-L Biochemicals, and RNases A and T1 and Klenow fragment were from Boehringer Mannheim.

RESULTS

Nucleotide Sequence and Deduced Amino Acid Sequence. The nucleotide sequence of the cloned genomic DNA encodes a 470-amino acid sequence with strong homology to the TRPB domain of yeast tryptophan synthase (Fig. 1). The reading frame is interrupted by four intervening sequences with 5′ GT-AG 3′ borders. Positions of these introns were confirmed by sequencing of a cDNA clone (see subsequent section) identical to the exons in Fig. 1.

Comparison of amino acid sequences of TRPB of *Arabidopsis* and six microbial species (Fig. 2) shows 50–65% overall identity with greater conservation in many regions. An extensive conserved segment (peptide 81–100 in Fig. 2) spans the pyridoxal phosphate-binding region delimited by Pratt and DeMoss (27). Three residues essential for binding pyridoxal phosphate in *Neurospora* and *E. coli*—the histidine residues at positions 88 and 92, lysine at position 93, as well as the essential cysteine residue at position 237 in Fig. 2 (corresponding to residues 82, 86, 87, and 230 in the original *E. coli* sequence; refs. 27 and 28)—are invariant in all the known TRPB sequences. The region surrounding the pyridoxal phosphate-binding site is highly conserved, although the alignment is interrupted by the insertion of amino acids 71–75 (Fig. 2) in the *Arabidopsis* sequence. Other conserved sites include arginine residue 154 that binds the α carboxyl group of serine and basic residues at positions 147 and 156 (29). The NH_2-terminus is exceptional in showing very little homology to microbial sequences and could be a chloroplast transit peptide (see *Discussion*).

Two Related Sequences in the *Arabidopsis* Genome. The strongly hybridizing bands on a filter of *Arabidopsis* genomic DNA probed with the cloned *Arabidopsis* gene (Fig. 3) correspond in size to bands in the restriction digests of the phage λM1g identified by hybridization with the yeast *TRP5* probe. The structure of the genomic tryptophan synthase region, inferred from this pattern, is identical to that of the recombinant phage. Less intense bands can also be seen (Fig. 3) by using either the genomic or cDNA clone as a hybridization probe under conditions of high stringency. Preliminary DNA sequence analysis of one of these fragments (R.L.L., unpublished data) shows that it has significant nucleotide and inferred amino acid homology to the TRPB gene reported here.

Characterization of the Transcript. Several lines of evidence demonstrate that the cloned gene is transcribed and delineate the transcription unit. The cDNA sequence (begin-

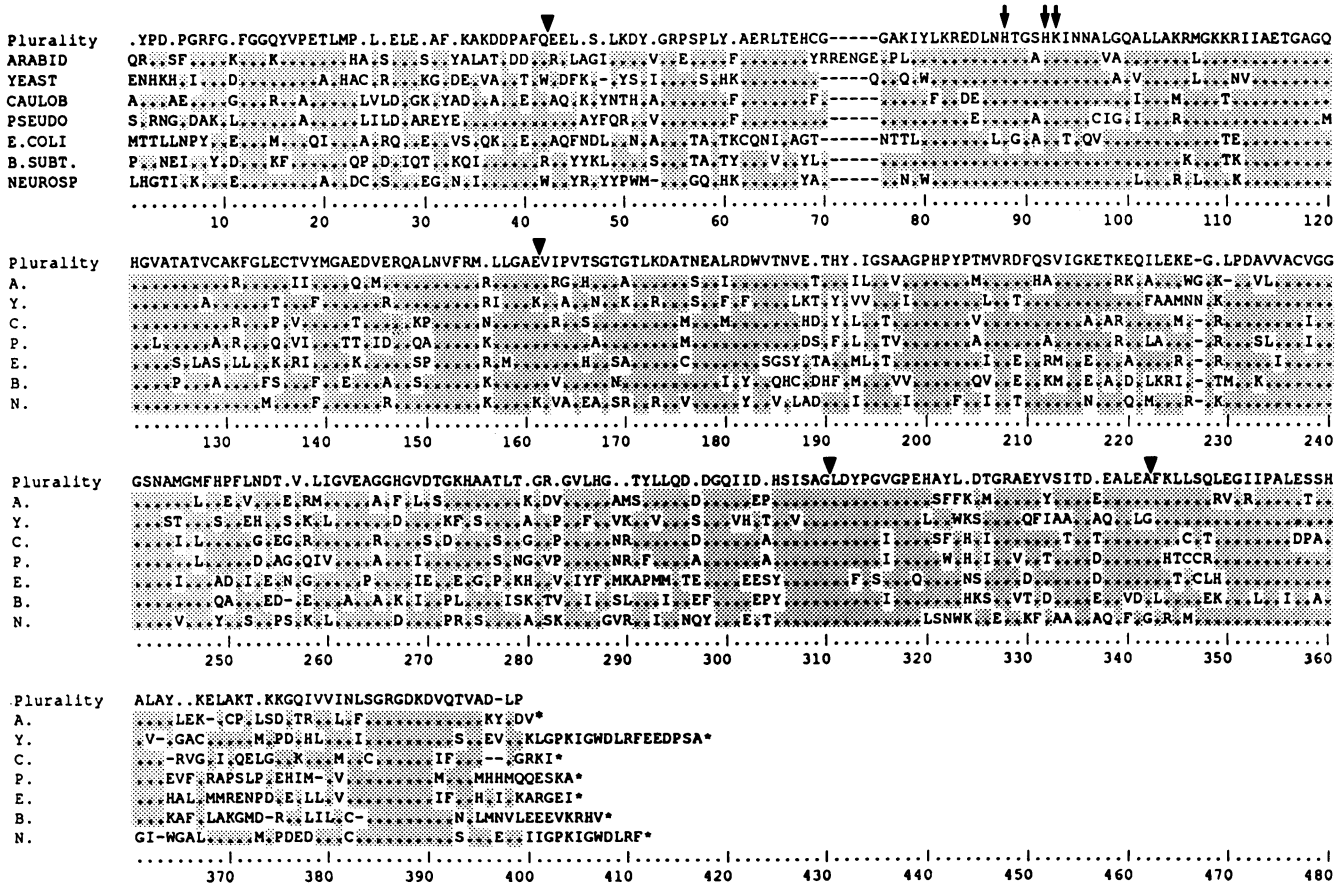

FIG. 2. Comparison of amino acid sequences of TRPB or B-domain regions of seven organisms; one-letter amino acid code is used. Amino acids diverging from the plurality sequence are shown. Dots and shaded areas indicate identity with the plurality sequence. The organisms are: *A. thaliana* (ARABID, A.), *Saccharomyces cerevisiae* (YEAST, Y.) (4), the Gram-negative bacteria *Caulobacter crescentus* (CAULOB, C.) (23), *Pseudomonas aeruginosa* (PSEUDO, P.) (24), and *E. coli* (3), the Gram-positive *Bacillus subtilis* (B. SUBT., B) (25), and *Neurospora crassa* (NEUROSP, N.) (5). The *Salmonella typhimurium* sequence (26) is not included because of its near-identity with the *E. coli* sequence. Positions of introns in the *Arabidopsis* gene are shown as triangles above the plurality sequence, and arrows indicate binding sites in the pyridoxal-binding region. The other TRPB sequences contain no introns.

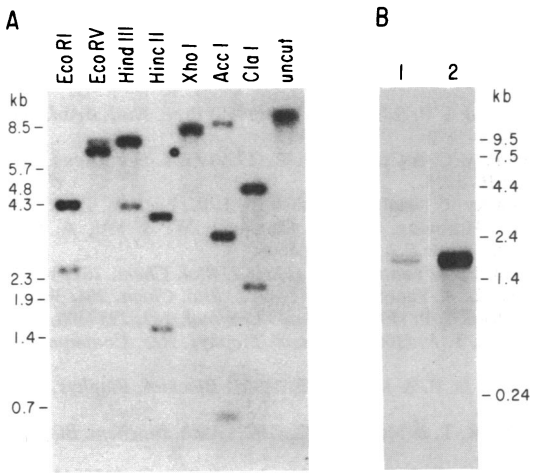

FIG. 3. (*A*) Genomic DNA of *Arabidopsis* hybridized with the 1.1-kb *Sac* I–*Xho* I fragment from genomic clone pMBT2. The restriction endonuclease used to digest the DNA is shown above each lane, and the position of DNA size markers is indicated on the left in kb. The same bands were seen by using the *Sac* I–*Xho* I fragment of the cDNA clone as probe. For the DNA digested with the enzymes *Eco*RI, *Eco*RV, *Hin*dIII, and *Cla* I (which do not cleave at any site within the cloned fragment), the strong bands correspond to single bands in the corresponding digests of λM1g hybridized with the yeast probe (data not shown). The genomic fragments seen for *Acc* I, *Hin*cII, and *Xho* I are also consistent with the restriction map of the cloned gene (Fig. 1). (*B*) RNA gel blot hybridization of *Arabidopsis* RNA [10 μg of total RNA, lane 1, and 1 μg of poly(A)-enriched RNA (20), lane 2] with the *Xho* I–*Sac* I fragment of pMBT3. Markers shown are the 0.24- to 9.5-kb RNA ladder from Bethesda Research Laboratories.

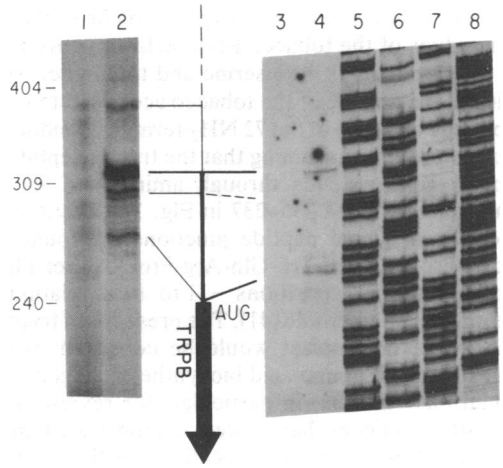

FIG. 4. Primer-extension and RNase protection experiments. Reaction products were subjected to electrophoresis on 6% acrylamide-urea gels (20). The position of denatured DNA size standards is shown in kb on the left. Primer-extension reactions were a modification of the procedure of Ausubel *et al.* (20). The oligonucleotide used in the primer-extension experiment shown corresponded to nucleotides 189–217 in Fig. 1 and was hybridized with 25 μg of total *Arabidopsis* RNA in 20 μl of Tris/EDTA buffer, pH 7.5 (20)/250 mM KCl at 75°C for 90 min and then brought to room temperature. After adding 38 μl of reverse transcriptase buffer (25 mM KCl/50 mM Tris, pH 7.5/10 mM dithiothreitol/3.5 mM $MgCl_2$/100 μg of bovine serum albumin per ml/0.5 mM each of dNTP) and 400 units of Moloney murine leukemia virus reverse transcriptase, the reaction was incubated at 37°C for 1 hr. Antisense RNA corresponding to the genomic region extending from nucleotide 265 in Fig. 1 to the *Ssp* I site at −240 was hybridized with cellular RNA and treated with ribonuclease A (20). Lanes: 1, ribonuclease protection without *Arabidopsis* RNA; 2, ribonuclease protection of *Arabidopsis* RNA; 3, primer extension without *Arabidopsis* RNA; 4, primer extension of *Arabidopsis* RNA; 5–8, dideoxy chain-termination sequencing ladder by using the 189–217 primer with single-stranded template from the 4.2-kb *Eco*RI genomic fragment cloned into pUC118.

ning with nucleotide −31 in Fig. 1) is identical to the exons of the genomic clone, showing that the gene is transcribed. The observation of a 1.6-kb poly(A)-enriched mRNA that hybridizes with the cDNA probe (Fig. 3*B*) is consistent with the size of the cDNA clone and the characterization of the *in vivo* 5′ ends of the mRNA. The 5′ ends of the mRNA were defined by primer-extension and nuclease protection analysis. Three RNase protection products correspond to 5′ ends identified by primer-extension analysis (Fig. 4, lane 2). The two prominent primer-extension products have 5′ ends that map to positions −44 and −2 with respect to the first AUG of the mRNA (Fig. 4, lane 4). The least abundant primer-extension product suggests a 5′ end at position −29. The largest of the four nuclease protection fragments has no counterpart in the primer-extension analysis. A possible TATA box (TATAAAAA, −77 in Fig. 1) is found upstream from these 5′ ends. The 3′ end of the cDNA extends 196 bp beyond the putative termination codon, ending with three adenosine residues that cannot be accounted for in the genomic DNA sequence (positions 2127–2129). The observation that this short poly(A) tail lies 25–44 nucleotides downstream from three copies of the canonical poly(A) addition signal AATAAA (30, 31) suggests a 3′ end of the transcription unit between 2127–2129. These data are all consistent with the proposed 5′ and 3′ ends of the mRNA.

DISCUSSION

The *Arabidopsis* Gene Encodes a Highly Conserved TRPB. The *Arabidopsis* clone obtained by hybridization to the *TRP5* gene of yeast has extensive amino acid homology to the highly conserved TRPB domain from microorganisms. The degree of conservation of the gene is striking (Fig. 2). For example, in a 45-amino acid sequence surrounding the pyridoxal phosphate-binding region, the plurality sequence is seen for all but three positions in *B. subtilis* and four positions in *Neurospora* and *Arabidopsis* (82–126 in Fig. 2).

The small size of the introns (83, 92, 96, and 250 bp) is typical of those found in other *Arabidopsis* genes. For example, four of six introns of the alcohol dehydrogenase gene and five of seven *enol*pyruvylshikimate synthase introns have lengths of 75–87 bp, and the remaining two introns in each case range from 103–235 nucleotides (32, 33). By contrast, the *enol*pyruvylshikimate synthase introns of *Petunia hybrida* range from 82–2300 bp (34) and the alcohol dehydrogenase introns of maize from 86–535 bp (35). In each of the four introns the GT/AG rule (36) is followed, and polypyrimidine tracts are found 5–30 bases from the 3′ ends. Three of the introns (1, 3, and 4) are of the pyrimidine-rich type (37). The sequences of the 5′ junctions conform, in general, to the broadly drawn consensus sequence for plant introns (38).

The First Exon Encodes a Chloroplast Transit-Like Peptide. The NH_2 terminus shown in Fig. 1 is predicted from the cDNA and transcriptional analysis if we assume that translation begins at the first AUG on the message. Both the primer-extension and nuclease protection experiments suggest the presence of major transcription starts at −44 and −2. Translation initiating from the first AUG produces an amino acid sequence divergent from the yeast or bacterial sequences, but with features characteristic of chloroplast transit peptides (39). In the 72 residues of exon 1 preceding the region of homology with yeast there are 20 serine, 9 alanine, and 4 threonine residues, with the threonine and serine residues accounting for 22 of the first 54 NH_2-terminal amino

acids (41%). This composition is quite similar to that of the first 60 residues of the tobacco acetolactate synthase transit peptide (40), which has 40% serine and threonine. Arginine and lysine make up 12% of the tobacco acetolactate synthase transit peptide and 11% of the 72 NH_2-terminal residues of the *Arabidopsis* protein. Assuming that the transit peptide of the *Arabidopsis* gene extends through amino acid 79 (corresponding to nucleotides 235–237 in Fig. 1), several features characteristic of signal peptide junctions are found in the sequence -Pro-Ala-Leu-Trp-Gln-Arg-Pro-Asp-Ser-Phe-Gly-Arg-, corresponding to positions −11 to +1 in relation to the putative mature polypeptide (41). The presence of tryptophan synthase in the chloroplast would be consistent with biochemical studies on amino acid biosynthesis, in general, and tryptophan biosynthesis, in particular (for review, see ref. 42). Transit sequences have been demonstrated in other amino acid pathway enzymes localized to the chloroplast such as acetolactate synthase (40), glutamine synthetase (43), and *enol*pyruvylshikimate synthase (33, 44).

A Second TRPB Gene in *Arabidopsis*. The presence in genomic DNA of hybridizing bands in addition to those predicted for this clone suggests the possibility of a second gene. Sequence analysis of the 2.5-kb *Eco*RI fragment seen in Fig. 3 shows it to encode a sequence very similar to the TRPB gene described in this report (unpublished results). Tryptophan synthase isozymes have not been demonstrated in any of the microorganisms where it has been studied genetically. However, there are genetic, enzymological, and DNA sequence data demonstrating the presence of isozymes for several other biosynthetic enzymes in plants (33, 40, 43, 45). The presence of two TRPB genes in *Arabidopsis* is unexpected because tryptophan auxotrophs have been isolated (46).

Genes for TRPA and TRPB Are Separate in *Arabidopsis*. Our analysis of the *Arabidopsis* gene provides the first evidence for a eukaryotic TRPB gene without an adjacent TRPA domain. Only two eukaryotic tryptophan synthase genes, those of *Neurospora* (5) and yeast (4), have been sequenced. The *Arabidopsis* organization contrasts with that in yeast and *Neurospora*, in which a single gene encodes a polypeptide with both A and B domains. Our evidence for separate TRPA and TRPB genes in *Arabidopsis* includes the sequence of the isolated gene, the cDNA, and the regions flanking the gene. The cDNA clone, which represents virtually the entire transcription product of the gene, does not encode sequences with TRPA homology. The presence of consensus TATA and poly(A) signals at appropriate sites within noncoding 5′ and 3′ regions of the gene and the cDNA clone provides support for the conclusion that the TRPB transcript is monocistronic. The genomic *Eco*RI subclone in pMBT2 includes 4.2 kb of *Arabidopsis* DNA that encompasses flanking regions 777 bp downstream of the end of the cDNA and 1500 bp upstream of the beginning of the transcript. These regions were sequenced and found to contain open reading frames of only 100–200 bases that could encode sequences unrelated to the products of the *trpA* gene of *E. coli* or the A region of the yeast *TRP5* gene. Although we do not know the location of the gene for TRPA, it is clear from the sequence analysis that the gene for TRPB is not fused to the α-subunit gene. The independent expression of the *Arabidopsis* tryptophan synthase subunits raises the possibility of differential regulation of each subunit.

The authors acknowledge the skillful technical assistance of Erick Kauffman and Cherie Day. We are grateful to Michael Christman and Brian Keith for their critical reading of the manuscript. This work was supported by National Science Foundation Grant DCB 8416894 (to G.R.F., American Cancer Society Professor), National Science Foundation Visiting Professorship for Women Program Grant RII-8600419 (to M.B.B.), and National Science Foundation Plant Molecular Biology Postdoctoral Fellowship Grant DCB-8508804 (to R.L.L.).

1. Crawford, I. P. & Yanofsky, C. (1958) *Proc. Natl. Acad. Sci. USA* **44**, 1161–1170.
2. Yanofsky, C. & Crawford, I. P. (1959) *Proc. Natl. Acad. Sci. USA* **45**, 1016–1026.
3. Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., van Cleemput, W. & Wu, A. M. (1981) *Nucleic Acids Res.* **9**, 6647–6668.
4. Zalkin, H. & Yanofsky, C. (1982) *J. Biol. Chem.* **257**, 1491–1500.
5. Burns, D. & Yanofsky, C. (1989) *J. Biol. Chem.* **264**, 3840–3848.
6. Crawford, I. P. (1987) *Methods Enzymol.* **142**, 293–300.
7. DeMoss, J. A. (1965) *Biochem. Biophys. Res. Commun.* **18**, 850–857.
8. Delmer, D. P. & Mills, S. E. (1968) *Biochim. Biophys. Acta* **167**, 432–443.
9. Nagao, R. T. & Moore, T. C. (1972) *Arch. Biochem. Biophys.* **149**, 402–413.
10. Chen, J. & Boll, W. G. (1971) *Can. J. Bot.* **49**, 1155–1163.
11. Manney, T. R. (1968) *Genetics* **60**, 719–733.
12. Tsai, H. & Suskind, S. R. (1972) *Biochim. Biophys. Acta* **284**, 224–340.
13. Schott, E. H. & Holzer, H. (1974) *Eur. J. Biochem.* **42**, 61–64.
14. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33**, 103–119.
15. Vieira, J. & Messing, J. (1987) *Methods Enzymol.* **153**, 3–11.
16. Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580.
17. Wertman, K. F., Wyman, A. R. & Botstein, D. (1986) *Gene* **49**, 253–262.
18. Feinberg, A. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
19. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
20. Ausubel, F., Brent, R. E., Kingston, R., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1983) *Current Protocols in Molecular Biology* (Wiley, New York).
21. Henikoff, S. (1984) *Gene* **28**, 351–359.
22. Tabor, S. & Richardson, C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
23. Ross, C. M. & Winkler, M. E. (1988) *J. Bacteriol.* **170**, 757–768.
24. Hadero, A. & Crawford, I. P. (1986) *Mol. Biol. Evol.* **3**, 191–204.
25. Henner, D. J., Bank, L. & Shimotsu, H. (1984) *Gene* **34**, 169–177.
26. Crawford, I. P., Nichols, B. P. & Yanofsky, C. (1980) *J. Mol. Biol.* **142**, 489–502.
27. Pratt, M. L. & DeMoss, J. A. (1988) *J. Biol. Chem.* **263**, 6872–6876.
28. Higgins, W., Miles, E. W. & Fairwell, T. (1980) *J. Biol. Chem.* **155**, 512–517.
29. Tanizawa, K. & Miles, E. W. (1983) *Biochemistry* **22**, 3594–3603.
30. Fitzgerald, M. & Shenk, T. (1981) *Cell* **24**, 251–260.
31. Heidecker, G. & Messing, J. (1986) *Annu. Rev. Plant Physiol.* **37**, 439–466.
32. Chang, C. & Meyerowitz, E. M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1408–1412.
33. Klee, H. J., Muskopf, Y. M. & Gasser, C. S. (1987) *Mol. Gen. Genet.* **210**, 437–442.
34. Gasser, C. S., Winter, J. A., Hironaka, C. M. & Shah, D. M. (1988) *J. Biol. Chem.* **263**, 4280–4289.
35. Dennis, E. S., Sachs, M. M., Gerlach, W. L., Finnegan, E. & Peacock, W. J. (1948) *Nucleic Acids Res.* **13**, 727–743.
36. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
37. Hanley, B. A. & Schuler, M. A. (1988) *Nucleic Acids Res.* **16**, 7159–7176.
38. Brown, J. W., Feix, G. & Frendewey, D. (1986) *EMBO J.* **5**, 2749–2758.
39. Schmidt, G. W. & Mishkind, M. L. (1986) *Annu. Rev. Biochem.* **55**, 879–912.
40. Lee, K. Y., Townsend, J., Tepperman, J., Black, M., Chui, C., Mazur, B., Dunsmuir, P. & Bedbrook, J. (1988) *EMBO J.* **7**, 1241–1248.
41. von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243–251.
42. Miflin, B. J. & Lea, P. J. (1982) *Encycl. Plant Physiol. New Ser.* **14A**, 5–64.
43. Tingey, S. V., Tsai, F.-Y., Edwards, J. W., Walker, E. L. & Coruzzi, G. M. (1988) *J. Biol. Chem.* **263**, 9651–9657.
44. della-Cioppa, G., Bauer, S. C., Klein, B. K., Shah, D. M., Fraley, R. T. & Kishore, G. M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 6873–6877.
45. Chaleff, R. S. & Bascomb, N. F. (1987) *Mol. Gen. Genet.* **210**, 33–38.
46. Last, R. L. & Fink, G. R. (1988) *Science* **240**, 305–310.