

▶ PLANT GENE REGISTER PGR99-044

Brock D. Weers and Robert W. Thornburg (1999) Characterization of the cDNA and Gene for an *Arabidopsis thaliana* Uracil Phosphoribosyltransferase (Accession No. AF116860). (PGR99-044) Plant Physiol. **119**: 1567

Characterization of the cDNA and Gene for an *Arabidopsis thaliana* Uracil Phosphoribosyltransferase (Accession No. [AF116860](#))

Brock D. Weers and Robert W. Thornburg*

Department of Biochemistry and Biophysics, Iowa State University, Ames, Iowa 50011

*Corresponding author
FAX 1-515-294-0453
Email: thorn@iastate.edu

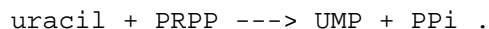
▶ REPORT

Journal Paper No. J-18270 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa Project No. 3202.

Keywords: UPRTase, uracil phosphoribosyltransferase, nucleotide salvage, pyrimidine metabolism, Arabidopsis

To supply pyrimidine nucleotides for the synthesis of RNA and DNA, cells can synthesize these nucleotides via the *de novo* pyrimidine biosynthetic pathway. This pathway, however, is energetically expensive, directly requiring the hydrolysis of two ATPs and additionally requiring several others due to the involvement of PRPP. Therefore, cells have developed the strategy to repeatedly reuse the preformed nucleotides. In general, cells at rest can meet their metabolic requirements by salvage of these preformed nucleotides, perhaps supplemented to a small degree with newly synthesized nucleotides to balance nucleotide degradation. It is only when cells are undergoing rapid growth or development that large amounts of new nucleotides are required thereby necessitating a large flux through the *de novo* biosynthetic pathway.

The salvage of pyrimidines therefore permits the reutilization of the preformed nucleotides to meet normal metabolism. Both pyrimidine ribosides and pyrimidine free bases are salvaged in these processes. Cytidine deaminase can convert cytidine into uridine (Kafer and Thornburg, in preparation) which is phosphorylated to form UMP (Deng and Ives, 1975). Similarly thymidine is phosphorylated to form TMP by thymidine kinase (Rudd and Fites, 1972). The free base, cytosine, cannot be salvaged in plants because plants lack the enzyme cytosine deaminase (Stougaard, 1993), however, uracil is converted directly into UMP by the action of the enzyme uracil phosphoribosyltransferase which transfers the phosphoribosyl moiety from PRPP to uracil to form UMP (Bressan, et al., 1978):



In animal cells, this enzyme, uracil phosphoribosyltransferase, is also required to activate some chemotherapeutic agents such as 5-fluorouracil or 5-fluorocytosine (Hoeprich, et al., 1974). Because of this redundancy in the *de novo* and the salvage pathways, many slow growing organisms, particularly parasites, rely exclusively on pyrimidine salvage. Such organisms prey metabolically on their host, utilizing preformed nucleotides available from their host. Indeed, many obligate parasites such as *Mycoplasma sp.*, *Borrelia burgdorferi*, and *Chlamydia psittaci* have lost the genes required for the *de novo* synthesis of nucleotides

(Boursaux-Eude, et al., 1997; McClarty and Qin, 1993; Neale, et al., 1983). They do not need them because they can obtain these important components from their host.

We are interested in the interaction of pyrimidine biosynthesis with pyrimidine salvage in Arabidopsis. Therefore, we searched the Arabidopsis EST libraries for other clones with similarity to the yeast uracil phosphoribosyltransferase. In this screen we identified the clone 146J9T7. This clone was obtained from the ABRC (ABRC, 1995), sequenced, characterized and deposited in GenBank as Accession Number [AF116860](#).

The UPT1 cDNA was characterized in this study. It was identified as an expressed sequence tag (146J9T7) in the ABRC collection due to its similarity with the yeast uracil phosphoribosyltransferase (GenBank [U10398](#)). After the clone had been sequenced, blast searches (Altschul, et al., 1990) revealed that the highest identity was found with a series of proteins identified as phosphoribosyltransferases. A comparison was made with this sequence compared with all known phosphoribosyltransferases from *Escherichia coli* and separately with all known phosphoribosyltransferases from *Saccharomyces cerevisiae*. These analyses demonstrated that the UPT1 cDNA was more similar to the *E. coli* and *S. cerevisiae* uracil phosphoribosyltransferases than to phosphoribosyltransferases specific for orotate phosphoribosyltransferase, adenine phosphoribosyltransferase, xanthine/guanine phosphoribosyltransferases, nicotinic acid phosphoribosyltransferase, quinolinic acid phosphoribosyltransferase, amido phosphoribosyltransferase, ATP phosphoribosyltransferase, or to phosphoribosyl pyrophosphate synthase. On the basis of these analyses, the 146J9T7 clone was considered to be an *Arabidopsis thaliana* uracil phosphoribosyltransferase.

Blast searches also identified an Arabidopsis thaliana chromosomal BAC clone (M3E9) sequenced by the EU Arabidopsis sequencing project (Bevan, et al., 1998) that had 100% identity with this cDNA. Therefore, we also included this sequence in our analysis of the Arabidopsis UPT1 gene.

DNA sequence reactions were performed using the Applied Biosystems Prism Dye-deoxy Cycle Sequencing Kit. The reactions were run on an Applied Biosystems Prism 377 DNA sequencer (Perkin-Elmer Corp.). The sequencing was initiated from known vector sequences. On the basis of these runs, primers specific to the UPT1 cDNA were constructed to extend the DNA sequence. DNA sequences were performed in duplicate or triplicate for each run. Each strand was completely sequenced, including sequencing through all restriction sites, and the entire sequence of the cDNA was confirmed on the opposite strand.

The 146J9T7 cDNA was found to be 914 nucleotides in length and lacked approximately 100 nucleotides encoding 25 amino acids from the 5' end of the cDNA. The poly A site of this cloned mRNA was found at nucleotide 892. Upstream of this site, we could not identify a classical poly adenylation signal; however, one potential sequence, AAGGAA, was identified at 865 to 870 of this cDNA, 22 nucleotides upstream from the poly A site. No significant palindromes or hairpins were found within the cDNA; however, a 42 bp duplication containing only 5 mismatches (84% duplicated identity) was found within the 3' end of the UPT1 cDNA.

The coding region of the UPT1 clone contains an open reading frame of 173 codons. Based upon sequence identity with the M3E9 BAC clone, the missing 25 N-terminal amino acids were identified and are included in our analysis. The total UPRTase 1 protein is 198 amino acids long. The protein sequence deduced from this open reading frame is a 21,961 Da polypeptide with a pI of 6.16. When this protein was analyzed for intracellular location using the PSORT tool (Nakai and Kanehisa, 1992), it was predicted to be cytoplasmic as is the next step in pyrimidine metabolism, UMP kinase (Zhou, et al., 1998). Further, uridine nucleotides are also cytoplasmic (Dancer, et al., 1990).

UPRTases have been identified from a wide variety of sources including four species of bacteria and three species of eukaryotes, including tobacco. Comparison of these sequences with the GCG tool "PileUp" (GCG, v. 9.0) revealed that these uracil phosphoribosyltransferase sequences define two distinct gene subfamilies. The *Arabidopsis thaliana* UPRTase1 is most closely related to the *Toxoplasma gondii* and *S. cerevisiae* enzymes. Interestingly, the tobacco enzyme is more closely related to the bacterial enzymes than it is to the eukaryotic enzymes.

Although the Arabidopsis UPRTase 1 protein shared only a limited degree of identity with the *Salmonella typhemurium* orotate phosphoribosyltransferase gene (20.9% identity), structural predictions on each of these proteins indicates that the C-terminal half of the molecule folds into a triple beta-sheet/alpha-helix structure. Similar folding patterns are found in other phosphoribosyltransferases (Eads, et al., 1997). The conserved lysine at position 73 of the *S. typhemurium* orotate phosphoribosyltransferase, that extends into the active site and interacts with either the 5' phosphate of OMP or the 2-hydroxyl and alpha-phosphoryl oxygen of PRPP in their respective substrate complexes (Ozturk, et al., 1995) is structurally conserved as Lysine 79 of the Arabidopsis uracil phosphoribosyltransferase. Thus it appears that both structurally and functionally, the Arabidopsis uracil phosphoribosyltransferase is similar to other phosphoribosyltransferases.

The chromosomal BAC clone (M3E9 = GenBank [AL022223](#)) was identified in blast searches using the UPRTase cDNA sequence. The M3E9 clone shows 100% identity with the 146J9T7 clone. This clone maps to Arabidopsis chromosome 4 at approximately 71 cM. The cDNA aligns with genomic sequences located from 22380 to 20939 on the negative strand of the M3E9 clone. The sequence of this gene permitted us to identify the first 25 amino acids of the UPRTase 1 protein, which are missing from the 146J9T7 clone. Alignment of the cDNA with the genomic sequence reveals the presence of seven introns within the cDNA. All of these introns are short, being between 71 and 137 bp long. They are polydisperse throughout the cDNA.

Flanking the UPT1 gene at the 5' end is an open reading frame whose sequence is similar to a hypothetical protein from *S. pombe*. The hypothetical protein is transcribed on the opposite strand from the UPT1 gene defining a bifunctional, divergent promoter of 3,114 nucleotides in length. Downstream of the UPT1 gene is a fructose-bisphosphate aldolase gene. We searched the EST databases using each of these flanking sequences for ESTs to establish whether these flanking genes were transcribed. No ESTs for either clone were identified. The ATG of the fructose-bisphosphate aldolase gene is located only 386 nt downstream from the polyA site of the UPT1 gene. This gene is transcribed from the same strand as the UPT1 gene leaving only a very short promoter for the fructose-bisphosphate aldolase gene.

Within the 5' flanking region, we identified a putative TATA box at [-140..-126] relative to the UPT1 methionine start codon. This suggests a long 5' UTR of 50 to 80 nt. This TATA box shows about 70% identity (9/13) with the canonical plant TATA box (Joshi, 1987). In addition, 12 root motif boxes "ATATT" were identified between 23023 and 21481 of the M3E9 BAC (= -1712 to -170 from the UPT1 methionine start codon) along with 10 light specific I boxes "GATAA" located between 23634 and 21515 of the M3E9 BAC (= -2323 to -204 from the UPT1 methionine start codon). Although it is not yet clear whether these signals regulate this gene, their presence may indicate that the UPT1 gene is expressed in both aerial and underground portions of the plant as would be expected based upon the function of the gene.

TABLE I

Characteristics of the UPT1 cDNA and gene

Organism:

Arabidopsis thaliana

Nomenclature:

cDNA ID = 146J9T7 (GenBank Accession # [AF116860](#))

Gene Name UPT1

Genomic Clone ID = BAC clone M3E9 (GenBank # [AL022223](#))

Chromosomal Location = Chromosome 4 near 71 cM

Gene Product:

Uracil phosphoribosyltransferase, UPRTase, EC [2.4.2.9](#) identification by sequence comparison to known genes

Characteristics of the UPT1 cDNA:

(Positions are numbered relative to the cDNA sequence) The sequenced DNA fragment of the UPT1

cDNA was a 969 bp fragment. 5' untranslated region could not be determined due to an incomplete cDNA at the 5' end.

Coding region

519 nucleotides [<1..519]

Terminates with an amber codon (TAG) [520..522]

3' untranslated region

388 nucleotides [523..914]

Poly A signal [865..870]

Poly A addition site at 892 of this sequence

A 42 bp duplication [747..788] and [800..841] was found within the 3' end of the UPT1 cDNA.

Characteristics of the UPT1 gene

(Positions are numbered relative to the M3E9 BAC clone and where appropriate to the location of the cDNA start site. Note: the UPT1 gene is on the negative strand of the M3E9 P1 clone, therefore the numbering of the genomic sequences proceeds from larger to smaller numbers)

5' flanking sequences

Promoter

3114 nt of the sequences adjacent to the UPT1 start codon were analyzed

A putative TATA box was identified at [21437..21451]

12 root motif boxes were identified between [21481..23023]

10 light specific I boxes were identified between [21515..23634]

Coding region structure

(identified by comparison with the UPT1 cDNA)

location length splice site

Exon I [<22546..22472] >75 nt CTG[^] Leu25

Intron I [22471..22381] 91 nt

Exon 2 [22380..22320] 61 nt G[^]GA Gly46

Intron 2 [22319..22200] 120 nt

Exon 3 [22199..22139] 49 nt AG[^]T Ser66

Intron 3 [22138..22037] 102 nt

Exon 4 [22036..21943] 94 nt GAG[^] Gln97

Intron 4 [21942..21872] 71 nt

Exon 5 [21871..21799] 73 nt G[^]GA Gly122

Intron 5 [21798..21662] 137 nt

Exon 6 [21661..21579] 80 nt TCA[^] Ser149

Intron 6 [21578..21462] 117 nt

Exon 7 [21461..20942] 542 nt

Stop codon [21314..21312]

3' UTR [21311..20942] 392 nt

3' flanking region

No palindromes or hairpin loops were detected within the 3' flanking region; however, a 42 bp duplicated region [21087..21052] and [21034..21009] is described above for the 3' end of the cDNA.

LITERATURE CITED

ABRC (1995) Arabidopsis Biological Resource Center: Seed and DNA Stock List. Library information for ESTs. pp. 239

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410

Bevan M, Vandenbol M, Jallet C, Portetelle D, Hoheisel J, Mewes HW, Mayer K, Schueller C (1998) *Arabidopsis thaliana* DNA chromosome 4, BAC clone M3E9 (ESSAII project). GenBank [AL022223](#) - M3E9.60

Boursaux-Eude C, Margarita D, Gilles A, Barzu O, Saint Girons I (1997) Borrelia burgdorferi uridine kinase: an enzyme of the pyrimidine salvage pathway for endogenous use of nucleotides. *FEMS Microbiol Lett* **151**:257-261

Bressan RA, Murray MG, Gale JM, Ross CW (1978) Properties of pea seedling uracil phosphoribosyltransferase and its distribution in other plants [pinto beans, soybeans, cereals]. *Plant Physiol.* **61**:442-446

Dancer J, Neuhaus HE, Stitt M (1990) Subcellular compartmentation of uridine nucleotides and nucleoside-5'-diphosphate kinase in leaves. *Plant Physiol.* **92**:637-641

Deng Q-I, Ives DH (1975) Non-allosteric regulation of the uridine kinase from seeds of *Zea mays*. *Biochem. Biophys. Acta* **377**:84-94

Eads J, Ozturk D, Wexler T, Grubmeyer C, Sacchettini J (1997) A new function for a common fold: the crystal structure of quinolinic acid phosphoribosyltransferase. *Structure* **5**:47-58

Hoeprich PD, Ingraham JL, Kelcker E, Winship MJ (1974) Development of resistance to 5-fluorocytosine in *Candida parapsilosis* during therapy. *J. Inf. Dis.* **130**:112-118

Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucl. Acids Res.* **15**:6643-6653

McClarty G, Qin B (1993) Pyrimidine metabolism by intracellular *Chlamydia psittaci*. *J Bacteriol* **175**:4652-4661

Nakai K, Kanehisa M (1992) A knowledge base for predicting protein localization sites in eukaryotic cells. *Genomics* **14**:897-911

Neale G, Mitchell A, Finch L (1983) Enzymes of pyrimidine deoxyribonucleotide metabolism in *Mycoplasma mycoides* subsp. *mycoides*. *J Bacteriol* **156**:1001-1005

Ozturk D, Dorfman R, Scapin G, Sacchettini J, Grubmeyer C (1995) Locations and functional roles of conserved lysine residues in *Salmonella typhimurium* orotate phosphoribosyltransferase. *Biochemistry* **34**:10755-10763

Rudd TP, Fites RC (1972) Association of thymidine and uridine kinase activities with changes in nucleic acid levels during peanut fruit ontogeny. *Phytochem.* **11**:1631-1636

Stougaard J (1993) Substrate-dependent negative selection in plants using a bacterial cytosine deaminase gene. *Plant J* **3**:755-761

Zhou L, Lacroute F, Thornburg R (1998) Cloning, expression in *Escherichia coli* and characterization of *Arabidopsis thaliana* uridine 5'-monophosphate/cytidine 5'-monophosphate kinase. *Plant Physiol* **117**:245-253

[Plant Gene Register Index](#)

[Plant Physiology ONLINE](#)