°°PLANT GENE REGISTER PGR00-043

Sanggyu Park and Robert W. Thornburg (2000) Characterization of a cDNA Encoding UMP Synthase from *Oryza sativa* (AF210322, AF210323, AF210324, and AF210325) (PGR00-043) Plant Physiol. **122**: 1457

Characterization of a cDNA Encoding UMP Synthase from *Oryza sativa* (AF210322, AF210323, AF210324, and AF210325)

Sanggyu Park and Robert W. Thornburg*

Department of Agricultural Chemistry, Taegu University, 15 Naeriri, Jinryang-myon, Kyongsan-si, Kyungbuk 712-714 S. Korea (S.P.); Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, Iowa 50011 (S.P., R.W.T.)

*Corresponding Author; Email: thorn@iastate.edu
FAX: (515) 294-0453

°°REPORT

UMP synthase is a bifunctional enzyme that catalyzes the final two steps of the de novo pyrimidine biosynthetic pathway. In bacteria and lower eukaryotes, these two steps, orotate phosphoribosyl transferase (OPRTase) and orotidylate decarboxylase (ODCase), are separate enzymes that are encoded by separate cistrons or by individual genes. In higher eukaryotes, including plants, however, these genes have become fused during evolution into a single transcriptional unit that encodes a single bifunctional enzyme, UMP synthase (Nasr, et al., 1994, Maier, et al., 1995).

In all cases, with the exception of some parasitic protozoans, such as *Trypanosoma cruzi*, the N-terminal half of the enzyme shares high identity with OPRTase, while the C-terminal half shares identity with ODCase (Suttle, et al., 1988, Schoeber, et al., 1993, Nasr, et al., 1994, Maier, et al., 1995). In *T. cruzi* the structure of the UMP synthase gene is exactly opposite with the N-terminal portion sharing identity with the ODCase and the C-terminal portion, OPRTase (Gao, et al., 1999). This suggests that the bifunctional UMP synthase gene has arisen several times in the course of evolution.

The evolutionary advantage of this bifunctionality is generally considered to be twofold. First, coordinated regulation of multiple transcriptional units is more taxing than for a single transcriptional unit, and second, the product from the first reaction, orotidine monophosphate (OMP), is channeled directly from the active site of the first enzyme to the active site of the second enzyme without disassociation from the enzyme, thereby improving the efficiency of the two-step catalytic process.

We have had a long interest in pyrimidine metabolism in plants, focusing primarily on dicotyledonous species (Shi and Thornburg, 1993, Maier, et al., 1995, Santoso and Thornburg, 1995, Zhou, et al., 1997, Santoso and Thornburg, 1998, Weers and Thornburg, 1998, Zhou, et al., 1998, Zhou and Thornburg, 1998, Park and Thornburg, 1999, Weers and Thornburg, 1999). Because of this interest, we have begun a comparison of the pyrimidine biosynthetic enzymes from dicots and monocots, and in this report we characterized the cDNAs for

UMP synthase from rice. We have previously characterized the rice UMP kinase cDNAs (Park and Thornburg, 1999).

To identify the rice UMP synthase cDNAs, the MAFF (Ministry of Agriculture, Forestry and Fisheries) DNA database (Japanese Rice Genome Project) was searched for clones that had identity to the Nicotiana tabacum UMP synthase gene (Maier, et al., 1995). Four clones were identified with high identity to the tobacco sequence. Each of these clones was kindly supplied to us by the MAFF DNA stock center. Sequencing 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.). Sequencing of the inserts was initiated from known vector sequences. Based upon these sequencing runs, additional sequencing primers were designed to extend new sequence. This process was continued until the gene was completely sequenced. The entire sequence was independently confirmed on the opposite strand.

The inserts of each of the four clones contained a UMP synthase cDNA. Three of the cDNAs shared an identical sequence, and the composite cDNA was termed Os-umps1. Individual clones were labeled Os-umps1a (AF210322), Os-umps1b (AF210323), and Os-umps1c (AF210324). The other cDNA (AF210325) had almost 90% identity to Os-umps1 at the nucleotide level and was termed Os-umps2.

Two of the cDNAs, Os-umps1b and Os-umps1c, were full-length (1,744 and 1,750 nucleotides, respectively) and possessed a 5' UTR. In the case of Os-umps1b, the 5' untranslated region (UTR) was 84 nucleotides long, and in Os-umps1c, the 5' UTR was 94 nucleotides long. These regions were identical in sequence. Neither of these had any significant direct or indirect repeats or palindromes. The other two cDNAs were incomplete-Os-umps1a was 1,173 nucleotides long and the Os-umps2 cDNA was 1,450 nucleotides long.

The three Os-umps1 clones had 3' UTRs with identical sequences. The Os-umps1b 3' UTRs was 5 nucleotides longer (210 nucleotides) than the 1a and 1c 3' UTRs (205 nucleotides). The 3' end of Os-umps2 cDNA was 191 nucleotides. In neither 3' UTR were any direct or indirect repeats or palindromes observed. No consensus polyA signal was identified within the umps1 cDNAs; however, a sequence (AATACC) that resembled the polyA sequence was identified near the polyA sites. The umps1 cDNAs use two different polyA sites downstream from the AATACC sequence. Two of the cDNAs, Os-umps1a and Os-umps1c, used a site 20 nucleotides from this polyA signal, while the polyA tail of Os-umps1b was 25 nucleotides from the AATACC signal. This indicates heterogeneity in the 3' end processing as has been observed for a number of other plant genes (Carter, et al., 1998).

The composite Os-UMPS1 cDNA, prepared from the three individual clones, encoded a 477 amino acid protein. The encoded protein had high identity with other eukaryotic UMP synthases. We prepared an identity profile of the following eukaryotic UMP synthases to compare the rice UMP synthase with other eukaryotic UMP synthases.

A multiple sequence format was obtained using the Wisconsin GCG (GCG, Version 9.0) DNA analysis program, "pileup". The amino acid sequences of the following species were compared: *Arabidopsis thaliana* (GenBank Accession Number X71842), *Bos taurus* (X65125), *Caenorhabditis elegans* (Z29443), *Dictyostelium discoideum* (X07560), *Drosophila melanogaster* (L00968), *Homo sapiens* (NM_000373), *Mus musculus* (P13439), *Naegleria gruberi* (L08073), *Nicotiana tabacum* (U22260), *Oryza sativa* umps1 (AF210323), and *Oryza sativa* umps2 (AF210325). This analysis revealed three clades with two outliers. The mammalian sequences (human, mouse, and bovine) were the most identical and formed a single clade. Next, the plant sequences formed an individual clade with two branches that correspond to the dicot UMP synthases (Arabidopsis and tobacco) and the monocot UMP synthases (Os-umps1 and Os-umps2). The third clade was formed from the *Dictyostelium* and *Drosophila* sequences. The *Caenorhabditis* and *Naegleria* sequences

were outliers of this analysis. Like other plant UMP synthase enzymes, The N-terminal half (amino acids 1-220) of the Os-UMPS1 cDNA shares high identity with OPRTase (23% identity with yeast ura5) and the C-terminal half (amino acids 220-477) shares high identity with ODCase (49% identity with yeast ura3).

To date, there have been no solved crystal structures of a eukaryotic UMP synthase; however, crystal structures of the *Salmonella typhemurium* OPRTase complexed with both reactants (orotate and PRPP) and products (OMP) have been solved (Scapin, et al., 1994, Scapin, et al., 1995). As a result, six catalytically important amino acid residues have been identified. These residues are: Lys(26), which interacts with the ribose-5-phosphate moiety of PRPP and OMP (Grubmeyer, et al., 1993); Lys(100), which interacts with b-pyrophosphate in the symmetry related adjacent subunit (Grubmeyer, et al., 1993, Scapin, et al., 1994, Ozturk, et al., 1995, Scapin, et al., 1995) Lys(103), which interacts with the pyrophosphate of PRPP (Grubmeyer, et al., 1993); Arg(155), which forms two hydrogen bonds between the guanidine sidechain and the orotate C4 oxygen (Scapin, et al., 1994, Scapin, et al., 1995); Glu(123), which may be involved in Mg^{2+} binding (Black and Hruby, 1992); and Asp(124), which interacts with the ribose moiety of PRPP (Scapin, et al., 1994, Scapin, et al., 1995). In addition, the carboxylates of Glu(123) and Asp(124) are also thought to stabilize an oxycarbonium-like intermediate as the anomeric carbon of phosphoribose swings through a large arc across the binding cavity from its position in the E-PRPP-orotate complex to the E-OMP complex (Scapin, et al., 1994, Scapin, et al., 1995). Similar stabilization of oxycarbonium ions are found in glycosidase catalysis in which carboxylate residues within the active site stabilize the oxocarbonium transition state intermediate (Kajimoto, et al., 1991, Tanaka, et al., 1994). This mechanism for OPRTase is consistent with the large primary and secondary tritium kinetic isotope effects that have been observed (Goitein, et al., 1978). Each of these residues important in the S. typhemurium enzymatic mechanism is absolutely conserved in all of the eukaryotic UMP synthases. In the rice Os-umps1 these residues are conserved as Lys(27), Lys(94), Lys(97), Glu(119), Asp(120), and Arg(151).

Similar studies have been performed on the yeast OMP decarboxylase that identify a catalytic role for Lys(93) in this enzyme (Smiley and Jones, 1992). This residue is conserved in the rice UMP synthase as Lys(307). Other mutations in the human UMP synthase identify Arg(96), Val(109), and Gly(429) as amino acid residues which that are required for enzymatic activity (Suchi, et al., 1997). These residues are also conserved throughout all other eukaryotic UMP synthases, although the hydrophobic Val(109) is replaced by isoleucine in some enzymes. These residues are conserved in the rice enzyme as Arg(93), Ile(105), and Gly(416). The conservation of these residues implies that the enzyme encoded by the Os-umps1 cDNA has both OPRTase and ODCase enzyme activities that function by a mechanism similar to that of the salmonella and yeast enzymes.

The Os-UMPS2 cDNA was incomplete at its 5' end, apparently lacking about 20 amino acids. It shares high identity with the Os-UMPS1 cDNA throughout the coding region (overall 87.4% nucleotide identity), but the 3' untranslated regions are completely different, indicating that these are independent genes. The most remarkable feature of this cDNA, however, is a 155 nucleotide deletion in the middle of the coding region. This deletion removes amino acids 63 through 114 of the UMP synthase protein. Thus this deletion removes approximately 30% of the OPRTase coding region. Further, this deletion causes a frameshift that results in another interesting caveat in the analysis of this cDNA. Assuming that the OPRTase domain is the correct reading frame, then the encoded protein would be approximately 150 amino acids. The 5' end of the cDNA is missing in this clone, so the exact length cannot be determined. This 150 amino acid protein corresponds almost exactly to the full OPRTase domain of the Os-UMPS2 gene. However, this protein is unrecognizable as an OPRTase. It lacks the middle third, and the C-terminal portion is out of frame. Blast searches (Altschul, et al., 1990) of this frameshifted protein do not identify any related proteins in GenBank. Further, if this was the correct reading frame, one would expect that the ODCase domain would not be highly conserved and certainly not maintained as an open reading frame. Yet not only is the ODCase domain conserved as an open reading frame, it is more highly conserved than the OPRTase domain at the amino acid level. Upstream of the deletion the two cDNAs

share 75.0% identity at the amino acid level, while downstream, the proteins share 87.4% amino acid identity. Clearly the ODCase domain is more highly conserved than the remnants of the OPRTase domain. The nucleotide identity of these two regions is relatively constant. Upstream of the deletion, the two cDNAs share 91.2% nucleotide identity, while downstream, the identity is 87.0% at the nucleotide level.

Because the ODCase domain appeared to be more highly conserved, we held the ODCase domain as the correct reading frame and searched in this same reading frame throughout the OPRTase domain for stop codons that would interrupt this reading frame. No stop codons were identified. This ODCase reading frame is intact throughout this cDNA. Because this cDNA is lacking approximately 20 codons at its 5' end, we cannot verify that this reading frame is complete and initiated by a methionine start codon. Comparison of this 20 codon stretch of the Os-UMPS1 cDNA does not identify any methionine codons in this reading frame; however, within this reading frame of the Os-UMPS1 cDNA there are four codons, any of which could give rise to an in-frame methionine start codon by a single nucleotide change. Based upon this analysis, we predict that the Os-UMPS2 cDNA encodes a UMP synthase deletion that results in an inactive OPRTase subunit but with an active ODCase subunit.

°°ACKNOWLEDGMENTS

Journal Paper No. J-18772 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 3202. This work was supported by a grant from the USDA, Hatch Act, and State of Iowa funds.

°°TABLE I

Characteristics of the UMP synthase mRNA from Oryza sativa

Organism:

Oryza sativa var. Nipponbare

Gene Product:

UMP synthase converts orotic acid into UMP in a two-step process. The first step couples a phosphoribosyl moiety with orotic acid to form OMP. The second step involves a decarboxylation of OMP to form UMP.

Clone Identity:

	Os-umps1a	**Os-umps1b**	**Os-umps1c**	**Os-umps2**
MAFF	C60441	C50057	C12584	C2074
DDBJ	RICC60441A	RICC50057A	RICC12584A	RICC2074A
EMBL	OSC60441_1A	OSC50057_1A	OSC12584_2A	OSC2074_1A
GenBank	C28243	C26785	C26554	D23033
This work	AF210322	AF310323	AF210324	AF210325

Characteristics of the UMP kinase cDNAs:

	Os-umps1a	Os-umps1b	Os-umps1c	Os-umps2
Length	**1173**	**1744**	**1750**	**1450**
5' UTR	--	[1..84]	[1..94]	--
Coding	[<1..948]	[85..1515]	[95..1525]	[<3..1208]
Stop Codon	[949..951]	[1516..1518]	[1526..1528]	[1209..1210]
3'UTR	[952..1156]	[1519..1727]	[1529..1733]	[1211..1402]
polyA signal	[1131..1135]	[1698..1703]	[1703..1713]	[1383..1388]
polyA site	1156	1727	1733	1402

°°LITERATURE CITED

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

Black M, Hruby D (1992) Site-directed mutagenesis of a conserved domain in vaccinia virus thymidine kinase. Evidence for a potential role in magnesium binding. J. Biol. Chem. **267**:6801-6806

Carter C, Graham R, Thornburg RW (1998) Arabidopsis thaliana contains a large family of germin-like proteins: Characterization of cDNA and genomic sequences encoding 12 unique family members. Plant Mol. Biol. **38**:929-943

Gao G, Nara T, Nakajima-Shimada J, Aoki T (1999) Novel organization and sequences of five genes encoding all six enzymes for de novo pyrimidine biosynthesis in Trypanosoma cruzi. J Mol Biol **285**:149-161

GCG (Version 9.0) Wisconsin Package.

Goitein R, Chelsky D, Parsons S (1978) Primary and secondary 3H substrate kinetic isotope effects for some phosphoribosyltransferases. J. Biol. Chem. **253**:2963-2971

Grubmeyer C, Segura E, Dorfman R (1993) Active site lysines in orotate phosphoribosyltransferase. J Biol Chem **268**:20299-20304

Kajimoto T, Liu KK-C, Pederson RL, Zhong Z, Ichikawa Y, Porco JA, Wong C-H (1991) Enzyme-catalyzed aldol condensation for asymmetric synthesis of azasugars: Synthesis, evaluation, and modeling of glycosidase inhibitors. J. Am. Chem. Soc. **113**:6187-6196

Maier T, Zhou L, Thornburg R (1995) Nucleotide sequence of a cDNA encoding UMP synthase from Nicotiana tabacum (GenBank U22260). Plant Physiol **108**:1747

Nasr F, Berthauche N, Dufour M-E, Minet M, Lacroute F (1994) Heterospecific cloning of Arabidopsis thaliana cDNAs by direct complementation of pyrimidine auxotrophic mutants of Saccharomyces cerevisiae. I. Cloning and sequence analysis of two cDNAs catalysing the second, fifth and sixth steps of the de novo pyrimidine biosynthesis pathway. Mol. Gen. Genet. **244**:23-32

Ozturk D, Dorfman R, Scapin G, Sacchettini J, Grubmeyer C (1995) Structure and function of

Salmonella typhimurium orotate phosphoribosyltransferase: protein complementation reveals shared active sites. Biochemistry **34**:10764-10770

Park S, Thornburg RW (1999) Characterization of UMP Kinase cDNAs from Oryza sativa (AF187062, AF187063). (PGR99-174). Plant **Physiol 121**:1384

Santoso D, Thornburg R (1998) UMP synthase is transcriptionally regulated by pyrimidine levels in Nicotiana plumbaginifolia. Plant Physiol **116**:815-821

Santoso D, Thornburg RW (1995) The use of a heterologous antibody to characterize UMP synthase mutant of tobacco cells. Menara **Perkebunan 63**:71-79

Scapin G, Grubmeyer C, Sacchettini J (1994) Crystal structure of orotate phosphoribosyltransferase. Biochemistry **33**:1287-1294

Scapin G, Ozturk D, Grubmeyer C, Sacchettini J (1995) The crystal structure of the orotate phosphoribosyltransferase complexed with orotate and alpha-D-5-phosphoribosyl-1-pyrophosphate. **Biochemistry 34**:10744-10754

Schoeber S, Simon D, Schwenger B (1993) Sequence of the cDNA encoding bovine uridine monophosphate synthase. Gene **124**:307-308

Shi N-Q, Thornburg RW (1993) Construction of a UMP synthase expression cassette from Dictyostelium discoideum. Gene **127**:199-202

Smiley J, Jones M (1992) A unique catalytic and inhibitor-binding role for Lys93 of yeast orotidylate decarboxylase. **Biochemistry 31**:12162-12168

Suchi M, Mizuno H, Kawai Y, Tsuboi T, Sumi S, Okajima K, Hodgson M, Ogawa H, Wada Y (1997) Molecular cloning of the human UMP synthase gene and characterization of point mutations in two hereditary orotic aciduria families. Am J Hum Genet **60**:525-539

Suttle DP, Bugg BY, Winkler JK, Kanalas JJ (1988) Molecular cloning and nucleotide sequence for the complete coding region of human UMP synthase. Proc. Natl. Acad. Sci. USA **85**:1754-1758

Tanaka Y, Tao W, Blanchard JS, Hehre EJ (1994) Transition state structures for the hydrolysis of a-D-glucopyranosyl fluroide by retaining and inverting reactions of glycosidases. J. Biol **Chem. 269**:32306-32312

Weers B, Thornburg RW (1998) Characterization of the cDNA and gene encoding the Arabidopsis thaliana AMP kinase (AF082882). Plant Physiol. **118**:711

Weers B, Thornburg RW (1999) Characterization of the cDNA and gene encoding the Arabidopsis thaliana Uracil Phosphoribosyltransferase (AF116860). Plant Physiol. **119**:1567

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

Zhou L, Lacroute F, Thornburg RW (1997) DNA sequence of dihydroorotase from Arabidopsis thaliana (AF000146). Plant Physiol. **114**:1569

Zhou L, Thornburg RW (1998) Site specific mutagenesis of conserved residues in the phosphate binding loop of the Arabidopsis UMP/CMP kinase alter ATP and UMP binding. Arch. Biochem. **Biophys. 358**:297-302

Plant Gene Register Index Plant Physiology ONLINE

' 1999 American Society of Plant Physiologists