

# Plant Gene Register PGR98-175

**Brock Weers and Robert Thornburg** (1998) Characterization of the cDNA and Gene for the *Arabidopsis thaliana* GDP-Mannose Pyrophosphorylase (Accession No. AF076484).. (PGR98-175) Plant Physiol. **118**: 1101

---

## **Characterization of the cDNA and Gene for the *Arabidopsis thaliana* GDP-Mannose Pyrophosphorylase (Accession No. [AF076484](#)).**

Brock Weers and Robert Thornburg (\*)

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

(\*) Corresponding author: Robert Thornburg

FAX: 1-515-294-0453

Email: thorn@iastate.edu

Journal Paper No. J-17993 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa  
Project No. 3202.

Keywords: GDP-mannose, glycoprotein biosynthesis, GTP Mannose-1 phosphate guanylyltransferase

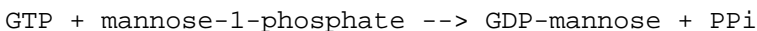
Mannose is involved in a variety of intracellular processes including the biosynthesis of polysaccharides, the modification of glycoproteins and the biosynthesis of vitamin C. Of these processes, the glycosylation of proteins may be the most physiologically important, because protein glycosylation is thought to be necessary for proper processing and secretion of numerous proteins.

Glycosylation of proteins is a post-translational modification that frequently occurs with extracellular proteins. Details of the mechanisms of protein glycosylation have been well worked out in both plants and animals. Briefly, a lipid-linked oligosaccharide composed of N-acetylglucosamine, mannose and glucose is synthesized, then transferred en block to the receptor protein. The addition of the monosaccharides to the lipid-linked oligosaccharide is performed by individual glycosyltransferases. The formation of a glycosidic bond by these transferases is a highly exergonic process and requires high free-energy input. Under physiological conditions the  $\Delta G^{\circ}$  for these reactions is on the order of 16 kJ/mol. Because of this highly exergonic process in the biosynthesis of these bonds, high-energy substrate intermediates are utilized for the biosynthesis of the oligosaccharide linkages. These high-energy intermediates are in the form of nucleotide diphosphosaccharides.

The nucleotide diphosphate at the sugar's anomeric carbon is a good leaving group and thereby facilitates the formation of a glycosidic bond in the reactions catalyzed by glycosyltransferases. Each of the common five nucleotide bases is utilized in the formation of nucleotide diphosphosaccharides (Sharon, 1975). Other than glucose, which can be associated with UDP, TDP, ADP and GDP, a particular sugar is normally associated with only one nucleotide. The majority of all sugars utilize the UDP to form the nucleotide diphosphosaccharide. The other nucleotides are associated with only a limited number of sugars. For example, sialic acid and KDO are associated with CMP, rhamnose is associated with TDP, and fucose and mannose are associated with GDP.

The enzymes that catalyze the formation of nucleotide diphosphosaccharides have the general nomenclature of nucleotide diphosphosaccharide pyrophosphorylases. They form the nucleotide diphosphosaccharide from the nucleotide triphosphate and monosaccharide-1-phosphate. In the case of GDP-mannose pyrophosphorylase,

the reaction is given as:



This reaction is reversible with the equilibrium towards the substrates; however, because a greater free energy is associated with the hydrolysis of pyrophosphate, the reaction proceeds towards the formation of GDP-Mannose. Because of the principal role of these pyrophosphorylases, they are thought to act as key sites for the regulation of oligosaccharide synthesis (Duffus and Duffus, 1984).

The importance of GDP-mannose pyrophosphorylase in intracellular plant biochemistry is demonstrated by the recent proposal that mannose appears to be involved in the biosynthesis of ascorbic acid (Wheeler *et al.*, 1998). In this pathway, GDP-mannose pyrophosphorylase produces GDP-mannose from mannose-1-phosphate. The GDP-mannose is then converted into GDP-L-galactose by an unusual 3, 5-diepimerase. The DGP-L-galactose is subsequently converted into L-galactose-1-phosphate, free L-galactose, L-galactono-1, 4-lactone, and finally L-ascorbic acid. The conversion of radiolabeled mannose into L-galactose is more rate limiting than the conversion of L-galactose into ascorbate. Thus, in addition to its role in glycoprotein biosynthesis, GDP-mannose pyrophosphorylase also appears to play a crucial role in the biosynthesis of vitamin C.

#### Clone Identification

The *Arabidopsis thaliana* GDP-mannose pyrophosphorylase cDNA was characterized in this study. It was identified as an expressed sequence tag (132P5T7) in the ABRC collection (ABRC, 1995) due to its similarity with the UDP-glucose pyrophosphorylase. 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 GDP-mannose pyrophosphorylases. On the basis of this high degree of identity, this clone was considered to be the *Arabidopsis thaliana* GDP-mannose pyrophosphorylase.

BLAST searches also identified an *Arabidopsis thaliana* chromosomal BAC (T5I7) that had near 100% identity with this cDNA. Therefore, we also included this sequence in our analysis of the Arabidopsis GMP1 gene.

#### cDNA Sequencing

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.). Sequence was initiated from known vector sequences. On the basis of these runs, primers specific to the cDNA sequence were synthesized 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.

#### Characteristics of the GDP-mannose pyrophosphorylase cDNA

The cDNA was found to be 1490 nucleotides in length and encoded the full length GDP-mannose pyrophosphorylase mRNA, including a long, 152 bp, 5' untranslated region. The initiation ATG start codon is located at position [153..155]. There are three lines of evidence indicating that this is the correct start codon. First, this is the first ATG in the 5' UTR of the mRNA. Second, the deduced N-terminal amino acid sequence of the GDP-mannose pyrophosphorylase is identical to the N-terminus of GDP-mannose pyrophosphorylases from *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Synechocystis sp.*, and *Aquifex aeolicus*. Finally, there are several in-frame stop codons within the 5' UTR indicating that the coding region is unlikely to extend further 5' beyond the initiation methionine at 153 to 155.

#### Characteristics of the Deduced amino Acid Sequence

The coding region of the Arabidopsis GDP-mannose pyrophosphorylase contains an open reading frame of 361 codons. The protein sequence deduced from this open reading frame is a 39,449 Da polypeptide with a pI of 6.28. When this protein was analyzed for intracellular location using the PSORT tool (Nakai and Kanehisa, 1992), this protein was predicted to be a peripheral membrane protein that has a non-cleavable signal sequence at its N-terminus. Such proteins are often targeted to the ER and subsequently sorted into vesicles. Because glycosylation occurs in the Golgi apparatus, the intracellular localization predicted by PSORT seems to fit very well with the biochemical function of the GDP-mannose pyrophosphorylase.

GDP-mannose pyrophosphorylases have been identified from seven species including archaeobacteria, eubacteria, blue-green algae, and eukaryotic sources. Comparison of these sequences with the GCG tool "PileUp" (GCG, v. 9.0) revealed that the eukaryotic GDP-mannose pyrophosphorylases form a distinctly different subset from the prokaryotic enzymes.

The deduced 361 amino acid sequence of the GDP-mannose pyrophosphorylase shares 40.9 %, 39.6% , and 44.3 % identity with GDP-mannose pyrophosphorylases from *S. pombe* , *S. cerevisiae* , and *C. elegans* respectively.

#### Characteristics of the GDP-mannose pyrophosphorylase gene

The chromosomal BAC T5I7 was identified in Blast searches using the GDP-mannose pyrophosphorylase cDNA sequence. This BAC maps to Arabidopsis chromosome II near 72 cM. The cDNA aligns with genomic sequences located from nucleotides 31244 to 33512 of the T5I7 BAC. This alignment reveals the presence of four introns including a large intron (538 bp) located within the 5' untranslated leader sequence at nucleotide 127 of the cDNA. The remaining 3 introns are short (81, 88, and 86 bp, respectively) and are located in the 5' half of the cDNA.

We searched for potential regulatory elements within the flanking regions of the GDP-mannose pyrophosphorylase gene. Within the 5' flanking region, we identified the putative TATA box at [-42..-30] relative to the start of the GDP-mannose pyrophosphorylase cDNA. We also identified a potential auxin A box regulatory element (McClure *et al.* , 1989) TGATAAAAA located at [30456..30465] and potential light regulatory elements similar to the rbcS light-regulated box 1 (Green *et al.* , 1987) TTTCAA located at [30368..30374] and [30650..30656].

---

#### **Table 1.** Characteristics of the GMP1 cDNA and gene

##### Organism:

Arabidopsis thaliana

##### Nomenclature:

cDNA ID = 132P5T7 (GenBank Accession # [AF076484](#))

Gene Name GMP1

Genomic Clone ID = BAC T5I7 (GenBank Accession # AC003000)

Chromosomal Location = Chromosome II near 72 cM

##### Gene Product:

GDP-mannose pyrophosphorylase; GTP Mannose-1-phosphate guanylyltransferase identification by sequence comparison to known genes

##### Characteristics of the GMP1 cDNA:

(positions are numbered relative to the cDNA sequence) The sequenced DNA fragment was 1490 bp long and encoded the entire GMP1 cDNA.

##### 5' untranslated region

152 nucleotides of the 5' UTR were identified [1..152]

##### coding region

1083 nucleotides [153..1235]

terminates with an opal codon (TGA) [1236..1238]

##### 3' untranslated region

252 nucleotides [1239..1490]

poly A signal [1457..1463]

poly A addition site at 1475 of this sequence

### Characteristics of the GMP1 gene

(positions are numbered relative to the BAC T5I7 and where appropriate to the location of the cDNA start site)

### Promoter

2265 nt of the promoter adjacent to the cDNA were analyzed  
potential Auxin A box - TGATAAAAA [30456..30465] = [-788..-779]  
potential light regulated element - TTTCAAA [30368..30374]=[-876..-870]  
potential light regulated element - TTTCAAA [30650..30656]=[-594..-588]  
TATA box - TATATAAACAAAA [31202..31214] = [-42..-30]  
cDNA start [31244]  
no significant repeats, palindromes, or hairpin loops were detected within the 3' flanking region

### coding region structure

(identified by comparison with the GMPI cDNA)  
5' UTR [31244..31933]  
Exon I [31244..31370]  
Intron I [31371..31908] 538 nt long, after position 127 of 5' UTR  
Exon 2 [31909..32044]  
Intron 2 [32045..32125] 81 nt long  
Exon 3 [32126..32188]  
Intron 3 [32189..32276] 88 nt long  
Exon 4 [32277..32516]  
Intron 4 [32517..32602] 86 nt long  
Exon 5 [32603..33512]  
3' UTR [33275..33512]

### 3' flanking region

poly A signal - TATAAA [33492..33497]  
poly A site - [33512]  
no significant repeats, palindromes, or hairpin loops were detected within the 3' flanking region

### Differences between cDNA and genomic clones

cDNA position 2 (cDNA = G, genomic = A) located in 5' UTR  
cDNA position 1277 (cDNA = "-", genomic = T) located in 3' UTR

---

## Literature Cited

**ABRC:** (1995) Library information for ESTs. in "Arabidopsis Biological Resource **Center:** Seed and DNA Stock List". The Ohio State University, 1735 Neil Avenue, 309 Botany & Zoology Bldg. Columbus, OH 43210 USA pp 239

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

**Duffus C, Duffus J:** (1984) Carbohydrate metabolism in plants. Longman, London, pp. 183

GCG: (Version 9.0) Wisconsin Package, Genetics Computer Group, Madison, WI.

**Green PJ, Kay SA, Chua N-H:** (1987) Sequence-specific interactions of a pea nuclear factor with light responsive elements upstream of the *rbcS-3A* gene. *EMBO J* **6**:2543-2549

**McClure BA, Hage G, Brown CS, Gee MA, Guilfoile TJ:** (1989) Transcription, organization and sequence of an auxin-regulated gene cluster in soybean. *Plant Cell* **1**:229-239

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

**Sharon N:** (1975) *Complex carbohydrates: Their chemistry, biosynthesis and functions. A set of Lecture Notes.* Addison-Wesley Publishing Co., London, pp. 466

**Wheeler GL, Jones MA, Smirnoff N** (1998) The biosynthetic pathway of vitamin C in higher plants. *Nature* **393**:365-369



[Return to Plant Gene Register Index](#)



[Return to Plant Physiology ONLINE](#)