JOURNAL OF BACTERIOLOGY, Nov. 1991, p. 7042–7045
0021-9193/91/217042-04$02.00/0

Vol. 173, No. 21

Nucleotide Sequence and Expression Analysis of the *Acetobacter xylinum* Uridine Diphosphoglucose Pyrophosphorylase Gene

GAUTE BREDE, ESPEN FJÆRVIK, AND SVEIN VALLA*

Center for Molecular Biology, The University of Trondheim, Medisinsk Teknisk Senter, N-7005 Trondheim, Norway

Received 29 May 1991/Accepted 2 September 1991

The nucleotide sequence of the *Acetobacter xylinum* uridine diphosphoglucose pyrophosphorylase gene was determined; this is the first procaryotic uridine diphosphoglucose pyrophosphorylase gene sequence reported. The sequence data indicated that the gene product consists of 284 amino acids. This finding was consistent with the results obtained by expression analysis in vivo and in vitro in *Escherichia coli*.

The enzyme uridine diphosphoglucose pyrophosphorylase (UDPGP) catalyzes the formation of uridine diphosphoglucose (UDPG), which is the glucosyl donor for the synthesis of cellulose (1, 2) and many other polysaccharides (20). UDPGP-deficient *Escherichia coli* mutants are altered in the carbohydrate part of the lipopolysaccharide layer (7, 21); this alteration affects the synthesis of the flagella (11). UDPGP is also involved in galactose utilization (7), modification of bacteriophage DNA (19), bacterial osmoregulation (8), and developmental regulation in the slime mold *Dictyostelium discoideum* (5). Despite the important role of UDPGP in metabolism, the primary structure of the gene encoding this enzyme has not previously been reported for any procaryotic species.

Bacterial strains and plasmids used are described in Table 1. The *Acetobacter xylinum* UDPGP gene (*celA*) was originally cloned as a 2.8-kb *Hin*dIII fragment by complementation of cellulose-negative mutants (24). Templates for sequencing of the cloned fragment were obtained by constructing exonuclease III deletion derivatives of plasmid pKF10 and a derivative of this plasmid in which the insert was cloned in the opposite orientation. The plasmids were digested with *Kpn*I and *Sal*I (unique sites in the polylinker) prior to exonuclease III degradation from the *Sal*I site. Initial experiments using transposon mutagenesis showed that the complementing function was localized within a 1.5-kb *Hin*dIII-*Xho*I region of the 2.8-kb fragment (data not shown), and the 1.5-kb complementing region was subjected to DNA sequence analysis by the method of Sanger et al. (18). The relevant part of this sequence is shown in Fig. 1. One long open reading frame (ORF) was identified, starting at either GTG (nucleotides [nt] 110 to 112 or nt 185 to 187) or ATG (nt 197 to 199) and terminating at TGA (nt 962 to 964). Potential Shine-Dalgarno (SD) sequences (9) associated with two of the putative translation initiation codons were identified at positions 97 to 100 (GAGG) and 192 to 195 (AGGA). RNA secondary structure predictions (25) indicated two stem-loop structures involving nt 961 to 1129. This predicted structure 3′ of the translational stop codon might be involved in termination of transcription.

In order to identify the biologically active initiation codon, we used an expression system based on the plasmid vector pT7-5. This plasmid contains a promoter (ϕ10) which is specifically recognized by the bacteriophage T7 RNA polymerase, and the system was used to obtain regulated expression of the UDPGP gene from this promoter. Two pKF10 deletion (from the *celA* upstream side) derivatives constructed for the DNA sequence analysis were used as a source of *celA* DNA for the expression analysis. The *celA*-containing *Sac*I (site in the pKF10 polylinker)-*Xho*I (site in *celA*) fragments from these two derivatives were subcloned into the pT7-5 polylinker (after digestion with *Sac*I plus *Sal*I), generating pGB31 and pGB119. In addition, we also constructed plasmid pGB89 by cloning the 1.1-kb *celA*-containing *Nsi*I-*Hin*dIII fragment from pGB31 into the pT7-5 polylinker (digested with *Pst*I plus *Hin*dIII). The natures of all three constructs are shown in Fig. 2.

Transcription from the ϕ10 promoter in the pGB constructs is dependent on the presence of the T7 RNA polymerase, which is expressed from a temperature-inducible promoter (λ P_L) in a separate replicon (pGP1-2). Exclusive expression from the ϕ10 promoter is achieved by first inducing transcription of the T7 RNA polymerase and then inhibiting the activity of the host RNA polymerase with rifampin (22).

For expression analysis, the pGB plasmids (and pGP1-2) were transformed into the UDPGP-deficient *E. coli* strain FF4001. Cells were grown in L-broth (24) containing 0.2%

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Description[a]	Reference
E. coli FF4001	*galU95*; derivative of MC4100	8
pKF10	pUC19 with a 2.8-kb *Hin*dIII insert; *celA*$^+$ Ap^r	24
pT7-5	ColE1 replicon, contains T7 ϕ10 promoter upstream of a polylinker; Ap^r	4
pGP1-2	P15A replicon, encoding T7 RNA polymerase transcribed from λ P_L under control of the thermolabile repressor *c*I857; Km^r	22
pGB31	pT7-5 carrying nt 31 through nt 1165 from *celA*	This study
pGB89	pT7-5 carrying not 89 through nt 1165 from *celA*	This study
pGB119	pT7-5 carrying nt 119 through nt 1165 from *celA*	This study

[a] Ap^r, ampicillin resistance; Km^r, kanamycin resistance.

* Corresponding author.

```
                                                    CACGCTACA
  10 TCATTGCCATCTGGCAACGCCCGCTTGCCTGTTTCCGTCATGTAAACAGT

  60 GGGCCGGTCGGTTTTGCCTGTTCCTGTCATGCATCTTGAGGTAAATATTA

 110 GTGATTAAGCCCCTTAAAAAAGCCGTATTGCCGGTTGCCGGCCTTGGAAC
       V  I  K  P  L  K  K  A  V  L  P  V  A  G  L  G  T

 160 ACGCTTTCTGCCCGCCACCAAGTGCGTGCCCAAGGAAATGCTGACCGTTG
       R  F  L  P  A  T  K  C  V  P  K  E  M  L  T  V  V

 210 TTGACCGTCCGCTGATCCAGTATGCGATTGACGAGGCACGCGAAGCCGGG
       D  R  P  L  I  Q  Y  A  I  D  E  A  R  E  A  G

 260 ATCGAGGAATTCTGCCTCGTTTCCAGCCGGGGCAAGGATTCCCTGATCGA
       I  E  E  F  C  L  V  S  S  R  G  K  D  S  L  I  D

 310 TTATTTCGACATTTCCTACGAACTCGAAGACACGCTGAAGGCCCGCAAGA
       Y  F  D  I  S  Y  E  L  E  D  T  L  K  A  R  K  K

 360 AGACATCGGCACTGAAGGCCCTGGAAGCAACCCGCGTCATCCCGGGCACC
       T  S  A  L  K  A  L  E  A  T  R  V  I  P  G  T

 410 ATGCTGTCCGTTCCGCCAGCAGGAACCGCTGGGCCTTGGCACGCCATCTG
       M  L  S  V  P  P  A  G  T  A  G  P  W  H  A  I  W

 460 GTGTGCGCGTGAGTTCATTGGCAACGACCCGTTCGCCATCCTGCTGCCCG
       C  A  R  E  F  I  G  N  D  P  F  A  I  L  L  P  D

 510 ATGACGTGGTGCAGAGCAAGAAGTCATGCATCGGCCAGCTGGTTGAAGTC
       D  V  V  Q  S  K  K  S  C  I  G  Q  L  V  E  V

 560 TACAACAAGACCGGCGGCAACGTACTGGCCGTGACCGAAGTCCCGCGTGA
       Y  N  K  T  G  G  N  V  L  A  V  T  E  V  P  R  E

 610 GCAGACCGGCAGCTATGGCATCCTTGATGTCGGCAAGGACGACGGCAAGA
       Q  T  G  S  Y  G  I  L  D  V  G  K  D  D  G  K  T

 660 CCGTCGAGGTCAAGGGCCTGGTTGAAAAGCCCGACCCGAAGGACGCACCG
       V  E  V  K  G  L  V  E  K  P  D  P  K  D  A  P

 710 TCCACCCTGTCCGTGATCGGTCGCTACGTGCTGACCGCCGACGTGCTGAA
       S  T  L  S  V  I  G  R  Y  V  L  T  A  D  V  L  K

 760 GCACCTGGCCAAGCTGGAAAAGGGCGCAGGCGGCGAAGTGCAGCTGACCG
       H  L  A  K  L  E  K  G  A  G  G  E  V  Q  L  T  D

 810 ACGCCATGGCCAAGACCATCGGCCACGTACCGTTCCACGGCTATCGCTAC
       A  M  A  K  T  I  G  H  V  P  F  H  G  Y  R  Y

 860 GAAGGCAAGCGCTTCGACTGCGGCAGCAAGATCGCTTCCTGGAAGCCCAG
       E  G  K  R  F  D  C  G  S  K  I  A  S  W  K  P  R

 910 ATCGCCTTTGCGCTGGAGCGTGAGGAACTGGCTCCCGGCGTGCGTGAATT
       S  P  L  R  W  S  V  R  N  W  L  P  A  C  V  N  S

 960 CCTGACGAAGTACAAGTGATCTTCCGGCACCGTTTTGATCCGACCAGCCT

1010 GCGGGAATATGATATCCGTGGGACGGTCGGCAAGACACTGGGGCCTGAAG

1060 ACGCCTATGCCATCGGCCGCACGGTTCGCAAGCGTCGTGGCCGGTGACGC

1110 GGCAGGACTGTCGTCGTCGGCTCTAGACGGACGCCTTTACGTCTCCCGCC

1160 CTCGAG
```

FIG. 1. Nucleotide sequence of the *A. xylinum* UDPGP gene (*celA*). Nucleotide 1 is located 387 bp from the *Hin*dIII site proximal to the starting point used for generation of the deletion plasmids.

glucose at 30°C to a cell density of approximately 5×10^8 cells per ml prior to induction. Heat induction of transcription from the ϕ10 promoter was performed by the method of Tabor and Richardson (22), and UDPGP activities were measured in crude enzyme extracts as described previously (6).

The results of the expression analysis (Table 2) demonstrated that pGB31, but not pGB89 or pGB119, expressed UDPGP in the absence of the T7 RNA polymerase. These results indicated that the native *celA* promoter was eliminated in the latter two constructs.

Analysis of enzyme extracts prepared from heat-induced cells carrying pGP1-2 and the different pGB derivatives showed that only the cells containing pGB89 expressed UDPGP at a high level. Since there is no SD sequence in the ϕ10-polylinker region of pT7-5, the presence of a biologically

TABLE 2. Expression analysis of the *celA* gene

Plasmid	UDPGP activity[a]		
	pGP1-2 absent	pGP1-2 present	
		Noninduced	Induced
pGB31	296	365	590
pGB89	9	401	3,000
pGB119	11	5	7

[a] Expressed as micromoles of substrate transformed per minute per gram of total protein. Enzyme extracts were prepared from FF4001 cells containing the specified plasmids.

active SD sequence in pGB89 was indicated. Extracts prepared from uninduced cells containing pGP1-2 and pGB89 also exhibited significant UDPGP activity. This uninduced expression is probably caused by the leakage synthesis known to occur in this system (22). There was no expression of UDPGP activity in induced cells containing pGP1-2 and pGB119, and thus it could be concluded that the biologically active SD sequence is eliminated in pGB119.

The UDPGP activity in cells containing pGB31 and pGP1-2 was not significantly stimulated after induction of T7 RNA polymerase synthesis. The reason for this is not known, but it seems possible that the host RNA polymerase binds to the native *celA* promoter and thereby represses the activity of the T7 RNA polymerase. This interpretation is supported by the fact that binding of the host RNA polymerase is known to take place also in the presence of rifampin (13) and that the binding affinity is much higher than for the T7 RNA polymerase (3).

All data from the expression analysis are thus consistent with the hypothesis that the putative SD sequence at nt 97 to 100 is biologically active and that translation starts at the GTG at position 110 to 112. The corresponding ORF encodes a protein with a predicted molecular weight of 30,900 (284 amino acids).

The results of the expression analysis showed that the native *celA* promoter can initiate transcription in *E. coli*, and we have previously reported that the *celB* gene is expressed in *E. coli* from an *A. xylinum* promoter (6). It has been shown that foreign promoters active in *E. coli* usually show significant homology to the σ^{70} consensus sequence (23), and a sequence homologous to the −10 region of *E. coli* σ^{70} promoters (12) could be identified at position 50 to 55 (TGTAAA). Since we have not performed a transcriptional analysis of the expression from the native *celA* promoter, the biological function of the putative −10 region is at present unknown.

In order to visualize the protein encoded by the identified ORF, we performed an in vitro coupled transcription-translation analysis of a purified 1.15-kb *Sac*I-*Xho*I fragment (containing the sequence from nt 1 down to the *Xho*I site) obtained from a deletion derivative of pKF10. The result of this analysis is shown in Fig. 3 and demonstrates that the DNA fragment directed the synthesis of a protein with a molecular weight of about 30,000, which correlates well with the calculated molecular weight deduced from the identified ORF.

It has been shown previously that the UDPGP from *Salmonella typhimurium* exhibits substantial deoxythymidine diphosphoglucose pyrophosphorylase (dTDPGP) activity (15), and conversely, dTDPGP from *Pseudomonas aeruginosa* also shows some UDPGP activity (14). We have tested the activity of the enzyme expressed from pGB89 in

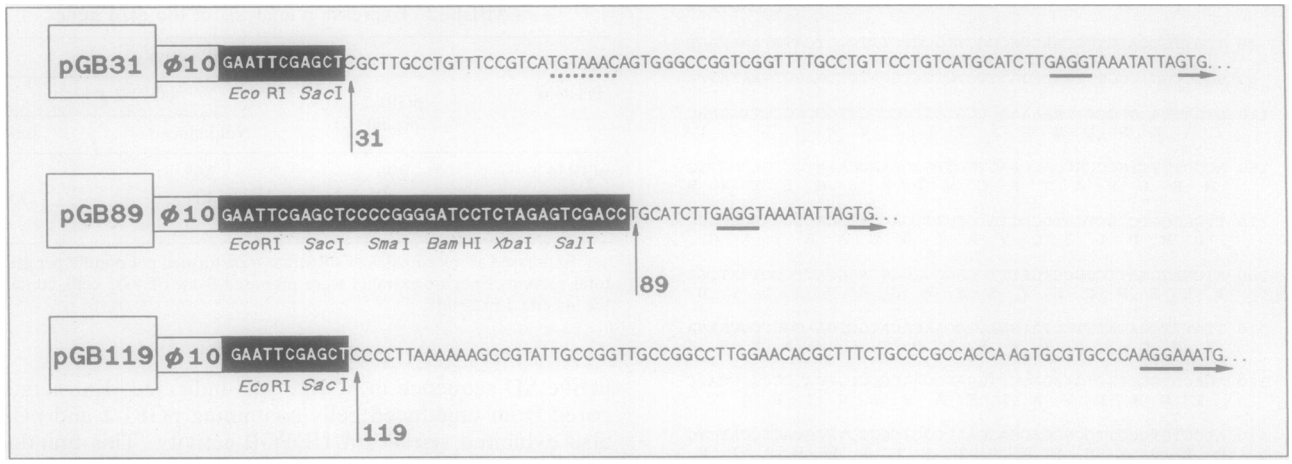

FIG. 2. Nucleotide sequences of expression plasmids in the region involved in initiation of transcription and translation. Putative SD sequences are underscored with a solid line, whereas a potential −10 region for a σ^{70}-like promoter sequence is underlined with a dotted line. Horizontal arrows are used to indicate the two potential translation initiation codons; vertical arrows indicate the position of the first nucleotide belonging to the *celA* sequence. The nucleotides in the black background belong to the pT7-5 polylinker sequence.

induced cells of FF4001 with respect to the substrates adenosine diphosphoglucose, cytidine diphosphoglucose, guanosine diphosphoglucose, and deoxythymidine diphosphoglucose. Due to the high levels of UDPGP expressed in these cells, the analysis could be performed directly in crude enzyme extracts. All substrates were supplied at a concentration of 1.0 mM, and assay conditions were otherwise as described previously (6). The analysis showed that the rates of pyrophosphorolysis of the alternative substrates were less than 0.6% of the UDPGP activity. The UDPGP expressed from *celA* in *E. coli* thus appears to be highly specific for UDPG, at least with respect to the substrates tested.

The GENE/EMBL sequence data base was searched for protein sequences having homology to the putative polypeptide encoded by *celA* (16), but no sequences with significant similarities were identified; the same was true for a search of the nucleotide sequence data base. The only UDPGP genes (cDNA) sequenced previously originate from potato tuber (10) and the slime mold *D. discoideum* (17) and encode polypeptides with molecular weights of 51,800 and 57,900, respectively. These two proteins show significant (42%) sequence homology with each other (10), while a direct comparison between the *celA* sequence and the two eucaryotic genes did not lead to identification of sequences with significant similarities. The possibility that there also exists a more procaryotelike UDPGP gene in *D. discoideum* cannot be excluded, however, since two distinct UDPGP mRNAs have been identified in this organism (5).

In *S. typhimurium* several forms of UDPGP also appear to exist, but it is assumed that there is only one structural gene (*galU*) encoding the enzyme, while another gene (*galF*) encodes a protein that modifies the *galU* protein into several forms (15). It is not known whether a similar system exists in *A. xylinum*, but the molecular weight of the polypeptide deduced from the *celA* sequence does not correlate with that reported for any of the *S. typhimurium* UDPGP isozymes.

Nucleotide sequence accession number. These data have been submitted to the GenBank data base and have been assigned accession number M76548.

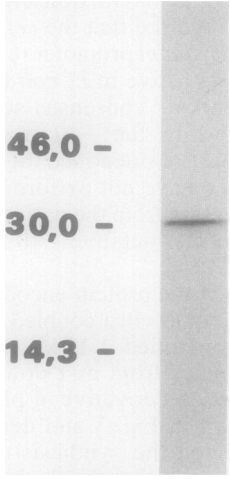

FIG. 3. Autoradiograph of a sodium dodecyl sulfate-polyacrylamide gel displaying the in vitro-translated *celA* gene product. A DNA fragment containing nt 1 through nt 1165 was purified and translated (in the presence of [^{35}S]methionine) by using a procaryotic transcription-translation kit (Amersham). Two micrograms of *celA* DNA was added to one reaction mixture (total volume, 35 μl), and 1 μl of this solution was applied on the gel. The numbers represent the molecular weights (in thousands) of three proteins in Amersham's molecular weight standard CFA 626.

This work was supported by a grant from The University of Trondheim.

REFERENCES

1. **Aloni, Y., D. P. Delmer, and M. Benziman.** 1982. Achievement of high rates of *in vitro* synthesis of 1,4-β-D-glucan: activation by cooperative interaction of the *Acetobacter xylinum* enzyme system with GTP, polyethylene glycol, and a protein factor. Proc. Natl. Acad. Sci. USA **79**:6448–6452.
2. **Callaghan, T., and M. Benziman.** 1984. High rates of *in vitro* synthesis of 1,4-β-D-glucan in cell free preparations from *Phaseolus aureus*. Nature (London) **311**:165–167.
3. **Chamberlin, M., and J. Ring.** 1973. Characterization of T7-specific ribonucleic acid polymerase. J. Biol. Chem. **248**:2245–2250.
4. **Durland, R. H., and D. R. Helinski.** 1990. Replication of the broad-host-range plasmid RK2: direct measurement of intracel-

lular concentrations of the essential TrfA replication proteins and their effect on plasmid copy number. J. Bacteriol. **172**:3849–3858.

5. **Fishel, B. R., J. A. Ragheb, A. Rajkovic, B. Haribabu, C. W. Schweinfest, and R. P. Dottin.** 1985. Molecular cloning of a cDNA complementary to a UDP-glucose pyrophosphorylase mRNA of *Dictyostelium discoideum*. Dev. Biol. **110**:369–381.
6. **Fjærvik, E., K. Frydenlund, S. Valla, Y. Huggirat, and M. Benziman.** 1991. Complementation of cellulose-negative mutants of *Acetobacter xylinum* by the cloned structural gene for phosphoglucomutase. FEMS Microbiol. Lett. **77**:325–330.
7. **Fukasawa, T., K. Jokura, and K. Kurahashi.** 1963. Mutations in *Escherichia coli* that affect uridine diphosphate glucose pyrophosphorylase activity and galactose fermentation. Biochim. Biophys. Acta **74**:608–620.
8. **Giæver, H. M., O. B. Styrvold, I. Kaasen, and A. R. Strøm.** 1988. Biochemical and genetic characterization of osmoregulatory trehalose synthesis in *Escherichia coli*. J. Bacteriol. **170**: 2841–2849.
9. **Gold, L., and G. Stormo.** 1987. Translational initiation, p. 1302–1307. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
10. **Katsube, T., Y. Kazuta, H. Mori, K. Nakano, K. Tanizawa, and T. Fukui.** 1990. UDP-glucose pyrophosphorylase from potato tuber: cDNA cloning and sequencing. J. Biochem. **108**:321–326.
11. **Komeda, Y., T. Icho, and T. Iino.** 1977. Effects of *galU* mutation on flagellar formation in *Escherichia coli*. J. Bacteriol. **129**:908–915.
12. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
13. **McClure, W. R., and C. L. Cech.** 1978. On the mechanism of rifampicin inhibition of RNA synthesis. J. Biol. Chem. **253**: 8949–8956.
14. **Melo, A., and L. Glaser.** 1965. The nucleotide specificity and feedback control of thymidine diphosphate D-glucose pyrophosphorylase. J. Biol. Chem. **240**:398–405.
15. **Nakae, T., and H. Nikaido.** 1971. Multiple molecular forms of uridine diphosphate glucose pyrophosphorylase from *Salmonella typhimurium*. J. Biol. Chem. **246**:4386–4411.
16. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
17. **Ragheb, J. A., and R. P. Dottin.** 1987. Structure and sequence of a UDP glucose pyrophosphorylase gene of *Dictyostelium discoideum*. Nucleic Acids Res. **15**:3891–3906.
18. **Sanger, F., S. Nicklen, A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
19. **Shedlovsky, A., and S. Brenner.** 1963. A chemical basis for the host-induced modification of T-even bacteriophages. Proc. Natl. Acad. Sci. USA **50**:300–305.
20. **Shibaev, V. N.** 1986. Biosynthesis of bacterial polysaccharide chains composed of repeating units. Adv. Carbohydr. Chem. Biochem. **44**:277–339.
21. **Sundararajan, T. A., A. M. C. Rapin, and H. M. Kalckar.** 1962. Biochemical observations on *E. coli* mutants defective in uridine diphosphoglucose. Proc. Natl. Acad. Sci. USA **48**:2187–2193.
22. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.
23. **Thomas, C. M., and F. C. H. Franklin.** 1989. Gene expression signals in gram-negative bacteria, p. 229–245. *In* C. M. Thomas (ed.), Promiscuous plasmids of gram-negative bacteria. Academic Press Limited, London.
24. **Valla, S., D. H. Coucheron, E. Fjærvik, J. Kjosbakken, H. Weinhouse, P. Ross, D. Amikam, and M. Benziman.** 1989. Cloning of a gene involved in cellulose biosynthesis in *Acetobacter xylinum*: complementation of cellulose-negative mutants by the UDPG pyrophosphorylase structural gene. Mol. Gen. Genet. **217**:26–30.
25. **Zuker, M.** 1981. Optimal computer folding of large RNA molecules using thermodynamics and oscilliary information. Nucleic Acids Res. **9**:133–148.