JOURNAL OF BACTERIOLOGY, Sept. 1987, p. 4355–4360
0021-9193/87/094355-06$02.00/0

Vol. 169, No. 9

Biosynthesis of Bacterial Glycogen: Primary Structure of *Salmonella typhimurium* ADPglucose Synthetase as Deduced from the Nucleotide Sequence of the *glgC* Gene

PATRICK S. C. LEUNG† AND JACK PREISS‡*

Department of Biochemistry and Biophysics, University of California, Davis, California 95616, and Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824

Received 9 March 1987/Accepted 10 June 1987

The nucleotide sequence of a 1.4-kilobase-pair fragment containing the *Salmonella typhimurium* LT2 *glgC* gene coding for ADPglucose synthetase was determined. The *glgC* structural gene contains 1,293 base pairs, having a coding capacity of 431 amino acids. The amino acid sequence deduced from the nucleotide sequence shows that the molecular weight of ADPglucose synthetase is 45,580. Previous results of the total amino acid composition analysis and amino acid sequencing (M. Lehmann and J. Preiss, J. Bacteriol. 143:120–127, 1980) of the first 27 amino acids from the N terminus agree with that deduced from nucleotide sequencing data. Comparison of the *Escherichia coli* K-12 and *S. typhimurium* LT2 ADPglucose synthetase shows that there is 80% homology in their nucleotide sequence and 90% homology in their deduced amino acid sequence. Moreover, the amino acid residues of the putative allosteric sites for the physiological activator fructose bisphosphate (amino acid residue 39) and inhibitor AMP (amino acid residue 114) are identical between the two enzymes. There is also extensive homology in the putative ADPglucose binding site. In both *E. coli* K-12 and *S. typhimurium* LT2, the first base of the translational start ATG of *glgA* overlaps with the third base TAA stop codon of the *glgC* gene.

ADPglucose synthetase (EC 2.7.7.27) is an allosteric enzyme in the glycogen biosynthetic pathway of eubacteria (23, 24). Among the enteric bacteria, ADPglucose synthetase is activated by glycolytic intermediates with fructose 1,6-bisphosphate as the activator and AMP, ADP, and P_i as inhibitors (23). The enzyme catalyzes the synthesis of ADP glucose from glucose 1-phosphate and ATP in the reaction glucose 1-phosphate + ATP $\rightleftharpoons$ ADPglucose + PP_i. This reaction is the first unique step in bacterial glycogen biosynthesis.

In *Escherichia coli*, the structural genes for ADPglucose synthetase (*glgC*), glycogen synthase (*glgA*), and branching enzyme (*glgB*) are mapped at 75 units on the genetic map, flanked by the *asd* (aspartic semialdehyde dehydrogenase) and *glpD* (glycerol phosphate dehydrogenase) genes (1). Okita et al. (20) have cloned the structural genes of glycogen biosynthetic enzymes of *E. coli* K-12 into the *Pst*I site of pBR322, and the nucleotide sequences of the *glgC* (2), *glgB* (3), and *glgA* genes (10) have been determined. *Salmonella typhimurium*, being closely related to *E. coli*, shows a lot of similarities in glycogen biosynthesis (12). The ADPglucose synthetases of *E. coli* and *S. typhimurium* are similar in that (i) they have similar subunit and native molecular weights; (ii) they have the same spectrum of activators and inhibitors; (iii) they have immunological cross-reactivity; (iv) of the first 27 amino acids in the N terminus 25 are identical; (e) genetically, the *glg* genes of both are clustered around 75 units on their genetic maps and are cotransducible with *asd* and *glpD* genes (28). Recently, we have cloned the *glgC* and *glgA* genes from *S. typhimurium* (14). This paper is a report of the nucleotide sequence, the deduced amino acid sequence, and codon usage pattern of ADPglucose synthetase from *S. typhimurium*. Its deduced amino acid sequence and amino acid composition are compared with those of *E. coli* ADPglucose synthetase.

MATERIALS AND METHODS

Bacteria, phage strains, and plasmids. The bacteria, phage strains, and plasmids used in this study are as follows: *E. coli* K-12 JM101 [*supE thi* Δ(*lac-proAB*) F′ *traD36 proAB lacI*q Δ*M15*], *E. coli* K-12 JM103 [*supE thi* Δ(*lac-proAB*) *strA endA sbcA hsdR* F′ *traD36 proAB lacI*q Δ*M15*]; bacteriophages M13 mp8, 9, 10, and 11 (17); and plasmid pPL301, which contains the *S. typhimurium glgC* and *glgA* genes on a 5.8-kilobase-pair insert on the *Sal*II site of pBR322 (14).

Media, commercial enzymes, and chemicals. Luria broth, YT medium, and YT soft agar were prepared as described previously (14, 18). Ampicillin was added at 25 μg/ml (final concentration) for the maintenance of plasmid pPL301.

Restriction endonucleases *Cla*I, *Dde*I, *Hin*dIII, *Hin*fI, *Pst*I, and *Pvu*II are from Bethesda Research Laboratories, Inc., as were T4 DNA ligase and the large fragment of DNA polymerase I. T4 polynucleotide kinase was from Pharmacia Fine Chemicals and P-L Biochemicals, Inc., calf intestine alkaline phosphatase was from Boehringer Mannheim Corp., and restriction enzyme *Bss*HII was from New England Biolabs. All enzymes were used as recommended by the manufacturers.

Isopropyl-β-D-thiogalactopyranoside was from Sigma Chemical Co., and 5-bromo-4-chloro-3-indoyl-galactoside was from Bethesda Research Laboratories, Inc. Radioactive nucleotides [γ-^{32}P]ATP (>5,000 Ci/mmol) and [α-^{32}P]dATP (>800 Ci/mmol) were from Amersham Corp.

DNA sequencing. DNA fragments containing the *S. typhimurium glgC* gene were purified from plasmid pPL301 (14) as described previously (2). DNA sequencing was done by

* Corresponding author.
† Present address: Division of Rheumatology and Allergy, School of Medicine, University of California, Davis, CA 95616.
‡ Present address: Biochemistry Department, Michigan State University, East Lansing, MI 48824.

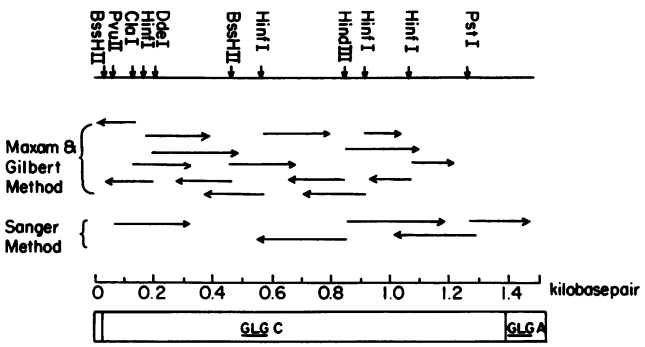

FIG. 1. DNA sequencing strategy of the *S. typhimurium* LT2 *glgC* gene. Restriction endonucleases cleavage sites are represented by short vertical arrows on a thin line. The arrows beneath the line indicate the direction and extent of the sequence. The physical locations of *glgC* and *glgA* genes are represented in the box.

```
          10          20          30          40          50          60
          *           *           *           *           *           *
CCT TGA GAG GTA AAA AAG GAG TTA ACT ATG GTG AGT TTA GAG AAG AAC GAT CGT GTA ATG
                                    Met Val Ser Leu Glu Lys Asn Asp Arg Val Met

          70          80          90          100         110         120
          *           *           *           *           *           *
TTG GCG CGC CAG CTG CCA TTG AAA TCT GTT GCC CTA ATC CTG GCC GGC GGC CGC GGC ACG
Leu Ala Arg Gln Leu Pro Leu Lys Ser Val Ala Leu Ile Leu Ala Gly Gly Arg Gly Thr

          130         140         150         160         170         180
          *           *           *           *           *           *
CGT CTA AAA GAT TTA GCA AAC AAA CGC GCC AAA CCA GCC GTC CAC TTT GGT GGG AAG TTT
Arg Leu Lys Asp Leu Ala Asn Lys Arg Ala Lys Pro Ala Val His Phe Gly Gly Lys Phe

          190         200         210         220         230         240
          *           *           *           *           *           *
CGC GTC ATC GAT TTC GCC TTA TCT AAT TGT CTG AAC TCC GGG ATT CGC CGT ATC GGC GTG
Arg Val Ile Asp Phe Ala Leu Ser Asn Cys Leu Asn Ser Gly Ile Arg Arg Ile Gly Val

          250         260         270         280         290         300
          *           *           *           *           *           *
ATC ACT CAG TAT CAG TCC CAT ACG CTG GTG CAG CAT ATT CAG CGC GGC TGG TCA CTG TTT
Ile Thr Gln Tyr Gln Ser His Thr Leu Val Gln His Ile Gln Arg Gly Trp Ser Leu Phe

          310         320         330         340         350         360
          *           *           *           *           *           *
AGC GAA GAG ATG AAC GAA TTT GTC GAT CTG CTA CCA GCC CAA CAG CGT ATG AAG GGC GAA
Ser Glu Glu Met Asn Glu Phe Val Asp Leu Leu Pro Ala Gln Gln Arg Met Lys Gly Glu

          370         380         390         400         410         420
          *           *           *           *           *           *
AAC TGG TAT CGC GGC ACG GCA GAC GCG GTG ACC CAG AAC CTG GAT ATT ATT CGT CGC TAT
Asn Trp Tyr Arg Gly Thr Ala Asp Ala Val Thr Gln Asn Leu Asp Ile Ile Arg Arg Tyr

          430         440         450         460         470         480
          *           *           *           *           *           *
AAA GCG GAA TAT GTC GTC ATC CTG GCA GGC GAT CAT ATC TAC AAG CAG GAC TAC TCG CGT
Lys Ala Glu Tyr Val Val Ile Leu Ala Gly Asp His Ile Tyr Lys Gln Asp Tyr Ser Arg

          490         500         510         520         530         540
          *           *           *           *           *           *
ATG TTG ATC GAT CAC GTC GAA AAA GGC GCG CGT TGC ACG GTG GCC TGT ATG CCG GTG CCG
Met Leu Ile Asp His Val Glu Lys Gly Ala Arg Cys Thr Val Ala Cys Met Pro Val Pro

          550         560         570         580         590         600
          *           *           *           *           *           *
ATC AAA GAA GCG ACG GCG TTC GGC GTG ATG GCG GTC GAT GAA AGC GAC AAG ATT ATT GAT
Ile Lys Glu Ala Thr Ala Phe Gly Val Met Ala Val Asp Glu Ser Asp Lys Ile Ile Asp
```

FIG. 2. Complete nucleotide sequence of the antisense strand and deduced amino acid sequence of the *S. typhimurium glgC* gene. The DNA sequence was determined from a 1.4-kilobase fragment of pPL301 as described in Materials and Methods. The box indicates the potential ribosome binding site. The deduced amino acid sequence of the *glgC* structural protein is shown below the DNA sequence. ** indicates the end of the *glgC* structural protein. The ATG start codon of the *glgA* structural protein is underlined.

the method of Maxam and Gilbert (17) and the method of Sanger et al. (26), and the sequencing strategy is shown in Fig. 1. DNA fragments were cloned into M13 mp phages as described by Messing (18).

RESULTS AND DISCUSSION

Nucleotide sequence of *S. typhimurium glgC* gene. The nucleotide sequence of a 1.4-kilobase-pair fragment containing the structural gene of *S. typhimurium* ADPglucose synthetase was determined. The sequencing strategy is shown in Fig. 1. More than 90% of the sequence was determined by the method of Maxam and Gilbert (17), and about 40% was determined by the method of Sanger et al. (26), allowing sequencing of both the sense and antisense strands with considerable overlapping. The fragments that were cloned into M13 phages for sequencing were *Pvu*II-*Pst*I (cloned into M13 mp11), *Pvu-Hin*dIII (cloned into M13 mp10), *Hin*dIII-*Pst*I (cloned into M13 mp8 and mp9), and *Pst*I-*Sal*I (cloned into M13 mp8). The complete nucleotide sequence of the *S. typhimurium glgC* structural gene is shown in Fig. 2. It contains 1,293 base pairs and has the coding capacity for a protein of 431 amino acids. The calculated molecular weight

```
          610         620         630         640         650         660
          *           *           *           *           *           *
TTT GTC GAA AAA CCG GCG AAT CCC GCA ATG CTT GGT GAC GCC AGC AAA TCG CTG GCC AGT
Phe Val Glu Lys Pro Ala Asn Pro Ala Met Leu Gly Asp Ala Ser Lys Ser Leu Ala Ser

          670         680         690         700         710         720
          *           *           *           *           *           *
ATG GGC ATT TAC GTT TTT GAC GCC GAT TAC CTG TAT GAA TTG CTG GCG GCA GAC GAT AAA
Met Gly Ile Tyr Val Phe Asp Ala Asp Tyr Leu Tyr Glu Leu Leu Ala Ala Asp Asp Lys

          730         740         750         760         770         780
          *           *           *           *           *           *
GAT GAC GCT TCC AGC CAC GAT TTC GGT AAA GAC ATT ATC CCC AAA ATC ACC CGC GAA GGT
Asp Asp Ala Ser Ser His Asp Phe Gly Lys Asp Ile Ile Pro Lys Ile Thr Arg Glu Gly

          790         800         810         820         830         840
          *           *           *           *           *           *
ATG GCT TAC GCG CAT CCT TTC CCG CTC TCC TGC GTG CAG TCC GAT CCA CAA GCC GAA CCG
Met Ala Tyr Ala His Pro Phe Pro Leu Ser Cys Val Gln Ser Asp Pro Gln Ala Glu Pro

          850         860         870         880         890         900
          *           *           *           *           *           *
TAC TGG CGC GAT GTA GGT ACG CTG GAA GCT TAC TGG AAG GCG AAC CGA GAT TTA GCC TCG
Tyr Trp Arg Asp Val Gly Thr Leu Glu Ala Tyr Trp Lys Ala Asn Arg Asp Leu Ala Ser

          910         920         930         940         950         960
          *           *           *           *           *           *
GTG ACG CCG CAG CTG GAT ATG TAT GAC CAG AAC TGG CCT ATC CGT ACG CAT ATG GAA TCG
Val Thr Pro Gln Leu Asp Met Tyr Asp Gln Asn Trp Pro Ile Arg Thr His Met Glu Ser

          970         980         990         1000        1010        1020
          *           *           *           *           *           *
CTA CCG CGA GCG AAA TTC GTG CAG GAA CGC TCC GGT AGC CAC GGT ATG ACG CTG AAG TCG
Leu Pro Arg Ala Lys Phe Val Gln Glu Arg Ser Gly Ser His Gly Met Thr Leu Lys Ser

          1030        1040        1050        1060        1070        1080
          *           *           *           *           *           *
TTG GTT TTC GGC GGC TGC ATT ATC TCC GGT TCG GTG GTG GTG CAA TCT GTG CTC TTC CCA
Leu Val Phe Gly Gly Cys Ile Ile Ser Gly Ser Val Val Val Gln Ser Val Leu Phe Pro

          1090        1100        1110        1120        1130        1140
          *           *           *           *           *           *
CGG GTG AGA ATA AAT TCT TTT TGT AAT ATT GAT TCG GCA GTG TTG TTA CCT GAG GTT TGG
Arg Val Arg Ile Asn Ser Phe Cys Asn Ile Asp Ser Ala Val Leu Leu Pro Glu Val Trp

          1150        1160        1170        1180        1190        1200
          *           *           *           *           *           *
GAT GGG CGC TCC TGC CGT TTA CGC TGC TGT GTT ATT GAC CGT GCC TGT ATT ATC CCG GAA
Asp Gly Arg Ser Cys Arg Leu Arg Cys Cys Val Ile Asp Arg Ala Cys Ile Ile Pro Glu

          1210        1220        1230        1240        1250        1260
          *           *           *           *           *           *
GGC ATG GTG ATT GGT GAA AAT GCG GAA GAG GAC GCT CGT CGT TTC TAC CGT TCA GAA GAA
Gly Met Val Ile Gly Glu Asn Ala Glu Glu Asp Ala Arg Arg Phe Tyr Arg Ser Glu Glu

          1270        1280        1290        1300        1310        1320
          *           *           *           *           *           *
GGT ATT GTA CTG GTC ACG CGT GAA ATG CTG CGC AAA CTG CAG GGC CAT AAA CAG GAG CGA
Gly Ile Val Leu Val Thr Arg Glu Met Leu Arg Lys Leu Gln Gly His Lys Gln Glu Arg

          1330        1340        1350        1360        1370        1380
          *           *           *           *           *           *
TA ATG CAG GTT TTA CAT GTA TGT TCA GAG ATG TTC CCC CTG CTG AAG ACT GGG GGG CTG GCG
** Met Gln Val Leu His Val Cys Ser Glu Met Phe Pro Leu Leu Lys Thr Gly Gly Leu Ala

     1390        1400        1410
     *           *           *
GAT GTG ATA GGC GCG TTG CCT GCG CGA C
Asp Val Ile Gly Ala Leu Pro Ala Arg
```

TABLE 1. Amino acid composition of ADPglucose synthetases

Amino acid	Amino acid residues in ADPglucose synthetase from:		
	E. coli (deduced from DNA sequence)	*S. typhimurium*	
		Amino acid analysis[a]	Deduced from DNA sequence
Arg	32	31.6 ± 1.7	33
Lys	20	21.3 ± 1.7	22
His	10	8.2 ± 0.1	10
Cys	9	10.4	10
Asp	29	45.2 ± 3.3[b]	30
Asn	19		13
Glu	29	43.6 ± 2.6[c]	25
Gln	13		17
Thr	13	15.5 ± 0.4	13
Ser	30	39.4 ± 1.1	28
Pro	21	19.6 ± 0.6	18
Gly	27	31.0 ± 1.6	28
Ala	29	37.1 ± 2.7	36
Val	39	34.0 ± 1.5	34
Met	14	12.3 ± 0.7	15
Ile	24	26.3 ± 0.4	28
Leu	37	35.9 ± 1.6	36
Tyr	15	15.1 ± 0.9	14
Phe	15	13.2 ± 0.3	15
Trp	6	5.6 ± 0.3	6

[a] Determined in reference 12.
[b] Number of Asp and Asn residues.
[c] Number of Glu and Gln residues.

of the *glgC* gene product is 45,580 and is in accordance to the value of 48,000 reported by Lehmann and Preiss (12).

The amino acid sequence of ADPglucose synthetase was deduced from the nucleotide sequence data (Fig. 2). Previous results of amino acid sequencing of the first 27 amino acids from the N terminus agree with predictions from nucleotide sequencing data. The total amino acid composition determined from the deduced amino acid sequence (Table 1) is in accordance with the total amino acid composition data determined by acid hydrolysis and reported by Lehmann and Preiss (12). Most of the deduced values are within 1 to 2 standard deviations or identical to the observed value. The only striking difference is with serine, where there can be an overestimation in amino acid analysis data. Of the 431 amino acids of ADPglucose synthetase, 15.1% are basic amino acids, 12.8% are acidic amino acids, 28.5% are polar uncharged amino acids, and 43.6% are nonpolar amino acids.

Nucleotide sequence analysis. There is a potential Shine-Dalgarno sequence (27) of AGGAG at 11 bases upstream from the translational start site of *glgC*. Deduced amino acid sequence data, which agree with the C-terminal analysis of the *S. typhimurium* ADPglucose synthetase (11) and the N-terminal amino acid sequence data of the *E. coli* glycogen synthase (6), indicate that the translational start of *glgA* follows immediately at the translational stop site of *glgC*. There is an overlapping of a single base; with the third base of the UAA stop codon being the first base of ATG start codon of the *glgA* structural gene. Therefore, there may be translational regulation on the expression of the *glgC* and *glgA* genes, and thus it is not surprising to see that *glgC* and *glgA* genes are coordinately expressed at the stationary phase of growth as reported by Steiner and Preiss (28).

glgC and *glgA* genes are mapped between 74 and 75 units on the *S. typhimurium* genetic linkage map (25, 28). The cloning of *glgC* and *glgA* genes on the same DNA fragment and the finding of the translational start of *glgA* follows immediately after the *glgC* translational stop support the genetic mapping results reported by Steiner and Preiss (28).

The G+C content of the *S. typhimurium glgC* gene is 52% and is in accordance with the average value of G+C content of 50 to 53% of the enteric bacteria and the average value of G+C value of 52% of the total *S. typhimurium* genome reported by Normore and Brown (19).

Codon usage. The codon usage of the *S. typhimurium glgC* gene is shown in Table 2. The codon usage is not random. There are preferences of CTG over CTT, CTC, and CTA for leucine; GTG over GTT, GTC, and GTA for valine; ACG over ACT, ACC, and ACA for threonine; CCG over CCT, CCC, and CCA for proline, AAA over AAG for lysine; CAG over CAA for glutamine; GAA over GAG for glutamic acid; GAT over GAC for aspartic acid; and GCC and GCG over GCT and GCA for alanine. Pyrimidine nucleotides are preferred at the wobble position of isoleucine, arginine, and glycine; G is preferred at the wobble position of leucine, valine, threonine, and glutamine; and G and C are preferred at the wobble position of valine and alanine. Besides, for isoleucine ATA is used only once, whereas ATT and ATC are used at approximately the same frequency. Similar codon usage pattern have been reported in the *araB*, *araA*, and *araD* genes of *S. typhimurium* LT2 (15, 16). Grosjeans and Fiers (4) suggested a general rule of codon usage in frequently expressed genes and rarely expressed genes in *E. coli*. It is correlated with the abundance of transfer RNA species and the occurrence of the respective codons in the structural genes (7, 8).

Interestingly, codons that were used only once, CGG for arginine, AGA for arginine, and ATA for isoleucine, are all clustered in amino acids residues 353 to 356. Moreover, CTC for isoleucine, which is used only twice, is at amino acid residue 350. The clustering of these rarely used codons around a certain area of the protein may serve to regulate the expression of the protein at the translational level. The relative abundance of these tRNA species may reflect the metabolic condition of the cell and hence may be a signal governing the expression of the protein.

Comparison of nucleotide sequences and deduced amino acid sequences of *S. typhimurium* and *E. coli glgC* genes. There is 80% homology in nucleotide sequence between *E. coli* and *S. typhimurium glgC* genes, showing that they have very similar codon usage. Moreover, it is interesting to note that in both bacteria the first base of the translational start of *glgA* (ATG) overlaps with the third base of the translational stop (TAA) of *glgC* (Fig. 2).

The deduced amino acid sequences of the *S. typhimurium* and *E. coli glgC* genes are compared in Fig. 3. There is 90% homology in their deduced amino acid sequence, and most of the changes are conservative. Of 45 differences in their amino acid sequences, 16 involve only one base change, 25 involve two base changes, and only 4 of them involve three base changes. Lehmann and Preiss (12) reported that ADPglucose synthetases from *E. coli* and *S. typhimurium* differ in their first 27 amino acid residues in the N terminus at residues 9 and 10. There was some uncertainty about residue 17 because gas chromatography and thin-layer chromatography indicated that it is glutamic acid, whereas back-hydrolysis of the phenylthiohydantion derivative showed that it may be glutamic acid or proline (12). Our results from nucleotide sequencing showed that amino acid residue 17 is proline in the *S. typhimurium* enzyme. The two differences in amino acid residues 9 and 10 among ADPglucose synthetase are histidine in *E. coli* to arginine in *S. typhimurium* at residue 9 and leucine in *E. coli* to valine in *S. typhimurium* at

TABLE 2. Codon usage of the *S. typhimurium glgC* gene

Codon	Amino acid	No. of residues	% of total	Codon	Amino acid	No. of residues	% of total
TTT	Phe	7	1.6	TCT	Ser	4	0.9
TTC	Phe	8	1.9	TCC	Ser	8	1.9
TTA	Leu	6	1.4	TCA	Ser	2	0.5
TTG	Leu	6	1.4	TCG	Ser	7	1.6
TAT	Tyr	6	1.4	TGT	Cys	5	1.2
TAC	Tyr	8	1.9	TGC	Cys	5	1.2
TAA		1		TGA		0	
TAG		0		TGG	Trp	6	1.4
CTT	Leu	1	0.2	CCT	Pro	3	0.7
CTC	Leu	2	0.5	CCC	Pro	2	0.5
CTA	Leu	4	0.9	CCA	Pro	5	1.2
CTG	Leu	17	3.9	CCG	Pro	8	1.9
CAT	His	6	1.4	CGT	Arg	14	3.2
CAC	His	4	0.9	CGC	Arg	14	3.2
CAA	Gln	3	0.7	CGA	Arg	3	0.7
CAG	Gln	14	3.2	CGG	Arg	1	0.2
ATT	Ile	14	3.2	ACT	Thr	1	0.2
ATC	Ile	13	3.0	ACC	Thr	2	0.5
ATA	Ile	1	0.2	ACA	Thr	0	0.0
ATG	Met	15	3.5	ACG	Thr	10	2.3
AAT	Asn	5	1.2	AGT	Ser	2	0.5
AAC	Asn	8	1.9	AGC	Ser	5	1.2
AAA	Lys	15	3.5	AGA	Arg	1	0.2
AAG	Lys	7	1.6	AGG	Arg	0	0.0
GGT	Val	5	1.2	GCT	Ala	4	0.9
GTC	Val	9	2.1	GCC	Ala	13	3.0
GTA	Val	3	0.7	GCA	Ala	6	1.4
GTG	Val	17	3.9	GCG	Ala	13	3.0
GAT	Asp	19	4.4	GGT	Gly	10	2.3
GAC	Asp	11	2.6	GGC	Gly	15	3.5
GAA	Glu	20	4.6	GGA	Gly	0	0.0
GAG	Glu	5	1.2	GGG	Gly	3	0.7

residue 10. The change in amino acid character is conservative because both arginine and histidine are basic amino acids and both leucine and valine are nonpolar in nature. It therefore retains the ionic character and hydrophobicity in this portion of the enzyme. Moreover, the change from leucine to valine involves only one base change, from TTA to GTA.

The deduced amino acid sequences of ADPglucose synthetase from *E. coli* and *S. typhimurium* are also compared with respect to their allosteric sites. Previous studies in the *E. coli* enzyme (9, 21, 22) showed that pyridoxal phosphate can be specifically reduced onto lysine residue 39 and that this phosphopyridoxylation increases the enzymatic activity in the absence of the allosteric activator fructose bisphosphate in the reaction mixture, therefore suggesting that the lysine residue constitutes at least a portion of the activator binding site. Since fructose bisphosphate is the physiological activator of both the *E. coli* and *S. typhimurium* ADPglucose synthetase, it is not surprising that the lysine residue is retained in the same position of the primary structure of the enzyme. Moreover, the conservation of basic amino acid residues in the vicinity of lysine residue 39, especially arginine residues in position 29, 32, and 40, may be involved in the ionic interaction with the phosphorylated activators of ADPglucose synthetase as suggested by Haugen et al. (5), Parsons and Preiss (21, 22), and Kappel and Preiss (9). The similar features in the primary structure of ADPglucose synthetase in *E. coli* and *S. typhimurium* suggest that their mechanism in fructose bisphosphate activation is very similar if not identical.

Inhibitor and substrate binding sites in the *glgC* enzyme were also studied with azido-AMP and azido-ATP analogs (11) and was shown to involve specifically tyrosine residue 114. This tyrosine residue is also conserved in the *E. coli* and *S. typhimurium* enzymes. Moreover, the amino acid sequence upstream and downstream from the tyrosine residue is highly conserved without a single change in amino acid sequence.

Lysine residue 195 of the *E. coli* ADPglucose synthetase was shown to be an ADPglucose-protected site of the enzyme from inactivation by reductive phosphopyridoxylation by Parson and Preiss (21). Modification by pyridoxal phosphate on the lysine residue was blocked by incubation with the substrate ADPglucose. Therefore, this region was suggested to be involved in the binding of ADPglucose or ATP. This lysine residue is also conserved in the *S. typhimurium* enzyme.

Comparison of the deduced amino acid sequence of ADPglucose synthetase of *E. coli* and *S. typhimurium* shows that the allosteric sites, fructose bisphosphate-, AMP-, ATP-, and ADPglucose-protected sites, are all conserved and thus suggest that the reaction mechanisms in the

```
                    20                  40                  60
MVSLEKNDHL MLARQLPLKS VALILAGGRG TRLKDLTNKR AKPAVHFGGK FRIIDFALSN
MVSLIKNDRV MLARQLPLKS VALILAGGRG TRLKDLANKR AKPAVHFGGK FRVIDFALSN

                    80                  100                 120
CINSGIRRMG VITQYQSHTL VQHIQRGWSF FNEEMNEFVD LLPAQQRMKG ENWYRGTADA
CLNSGIRRIG VITQYQSHTL VQHIQRGWSL FSEEMNEFVD LLPAQQRMKG ENWYRGTADA

                    140                 160                 180
VTQNLKIIRR YKAEYVVILA GDHIYKQDYS RMLIDHVEKG VRCTVVCMPV PIEEASAFGV
VTQNLKIIRR YKAEYVVILA GDHIYKQDYS RMLIDHVEKG ARCTVACMPV PIKEATAFGV

                    200                 220                 240
MAVDENDKTI EFVEKPANPP SMPNDPSKSL ASMGIYVFDA DYLYELLEED DRDENSSHDF
MAVDESDKII DFVEKPANP- AMLGDASKSL ASMGIYVFDA DYLYELLAAD DKDDASSHDF

                    260                 280                 300
GKDLIPKITE AGLAYAHPFP LSCVQSDPDA EPYWRDVGTL EAYWKANLDL ASVVPKLDMY
GKDIIPKITE EGMAYAHPFP LSCVQSDPQA EPYWRDVGTL EAYWKANRDL ASVTPQLDMY

                    320                 340                 360
DRNWPIRTYN ESLPPAKFVQ DRSGSHGMTL NSLVSGGCVI SGSVVVQSVL FSRVRVNSFC
DQNWPIRTHM ESLPRAKFVQ ERSGSHGMTL KSLVFGGCII SGSVVVQSVL FPRVRINSFC

                    380                 400                 420
NIDSAVLLPE VWVGRSCRLR RCVIDRACVI PEGMVIGENA EEDARRFYRS EEGIVLVTRE
NIDSAVLLPE VWDGRSCRLR CCVIDRACII PEGMVIGENA EEDARRFYRS EEGIVLVTRE

MLRKL-GHKQER*
MLRKLQGHKQWR*
```

FIG. 3. Comparison of amino acid sequence of ADPglucose synthetase of *E. coli* K-12 and *S. typhimurium* LT2. Amino acids are represented by the one-letter amino acid code. The amino acid sequence of *E. coli* ADPglucose synthetase is shown in the top line, and that of the *S. typhimurium* enzyme is shown on the bottom line. * indicates the end of sequence. The amino acids underlined are those that are different in the two enzymes.

ADPglucose synthetases are very similar if not identical between the two bacteria.

Comparison of nucleotide sequence and the deduced amino acid sequence has allowed us to investigate the molecular mechanisms of reaction of ADPglucose synthetase. Further experiments are in progress on the nucleotide sequencing of a cloned ADPglucose synthetase from an *E. coli* allosteric mutant (13) and on site-directed mutagenesis of the allosteric enzyme to gain further insights into the reaction mechanism at the molecular level.

ACKNOWLEDGMENTS

This work was supported in part by Public Health Service grants AI 05520 and AI 22835 from the National Institutes of Health.

LITERATURE CITED

1. **Bachman, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
2. **Baecker, P. A., C. E. Furlong, and J. Preiss.** 1983. Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* ADPglucose synthetase as deduced from the nucleotide sequence of the *glg*C gene. J. Biol. Chem. **258**:5084–5088.
3. **Baecker, P. A., E. Greenberg, and J. Preiss.** 1986. Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* 1,4-α-D-glucan:1,4-α-D-glucan 6-α-D-(1,4-α-D-glucano)-transferase as deduced from the nucleotide sequence of *glg*B gene. J. Biol. Chem. **261**:8738–8743.
4. **Grosjeans, H., and W. Fiers.** 1982. Potential codon usage in procaryotic gene: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.
5. **Haugen, T., A. Ishaque, and J. Preiss.** 1978. Biosynthesis of bacterial glycogen. Characterization of the subunit structure of *Escherichia coli* B glucose-1-phosphate adenyl transferase. J. Biol. Chem. **251**:7880–7885.
6. **Holmes, E., C. Boyer, and J. Preiss.** 1982. Immunological characterization of *Escherichia coli* B glycogen synthase and branching enzyme and comparison with enzymes from other bacteria. J. Bacteriol. **151**:1444–1453.
7. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. **146**:1–21.
8. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for the synonymous codon choice that is optimal for the *Escherichia coli* translational system. J. Mol. Biol. **151**:389–409.
9. **Kappel, W. K., and J. Preiss.** 1981. Biosynthesis of bacterial glycogen. Purification and characterization of ADPglucose pyrophosphorylase with modified regulatory properties from *Escherichia coli* B mutant CL1136-504. Arch. Biochem. Biophys. **209**:15–28.
10. **Kumar, A., C. E. Larsen, and J. Preiss.** 1986. Biosynthesis of bacterial glycogen. Primary structure of *Escherichia coli* ADPglucose:α-1,4-glucan-4-glucosyltransferase as deduced from the nucleotide sequence of the *glg*A gene. J. Biol. Chem. **261**:16256–16259.
11. **Larsen, C. E., Y. M. Lee, and J. Preiss.** 1986. Covalent modification of the inhibitor-binding site(s) of *Escherichia coli* ADPglucose synthetase. J. Biol. Chem. **261**:15402–15409.
12. **Lehmann, M. and J. Preiss.** 1980. Biosynthesis of bacterial glycogen: purification and properties of *Salmonella typhimurium* LT-2 adenosine diphosphate glucose pyrophosphorylase. J. Bacteriol. **143**:120–127.
13. **Leung, P., Y. M. Lee, E. Greenberg, K. Esch, S. Boylan, and J. Preiss.** 1986. Cloning and expression of the *Escherichia coli glgC* gene from a mutant containing an ADPglucose pyrophosphorylase with altered allosteric properties. J. Bacteriol. **167**:82–88.
14. **Leung, P. S. C., and J. Preiss.** 1987. Cloning of the ADPglucose pyrophosphorylase (*glgC*) and glycogen synthase (*glgA*) structural genes from *Salmonella typhimurium* LT2. J. Bacteriol. **169**:4349–4354.
15. **Lin, H. C., S. P. Lei, and G. Wilcox.** 1985. The *ara*BAD operon of *Salmonella typhimurium* LT2. I. Nucleotide sequence of *ara*B and its primary structure of its product, ribulokinase. Gene **34**:111–112.
16. **Lin, H. C., S. P. Lei, and G. Wilcox.** 1985. The *ara*BAD operon of *Salmonella typhimurium* LT2. II. Nucleotide sequence of *ara*A and primary structure of its product, L-arabinose isomerase. Gene **34**:123–128.
17. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing of end-labelled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
18. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
19. **Normore, W. M., and J. R. Brown.** 1970. *In* H. A. Saker (ed.), Handbook of biochemistry: selected data for molecular biology. Chemical Rubber Company, Cleveland, Ohio.
20. **Okita, T. W., R. L. Rodriguez, and J. Preiss.** 1981. Biosynthesis of bacterial glycogen: cloning of the glycogen biosynthetic enzyme structural genes of *Escherichia coli*. J. Biol. Chem. **256**:6944–6952.
21. **Parsons, T. F., and J. Preiss.** 1978. Biosynthesis of bacterial glycogen. Incorporation of pyridoxal phosphate into the allosteric activator site and an ADPglucose protected pyridoxal phosphate binding site of *Escherichia coli* B ADPglucose synthase. J. Biol. Chem. **253**:6197–6202.
22. **Parsons, T. F., and J. Preiss.** 1978. Biosynthesis of bacterial glycogen: Isolation and characterization of the pyridoxal phosphate allosteric activator site and the ADPglucose protected pyridoxal phosphate binding site of *Escherichia coli* B ADPglucose synthase. J. Biol. Chem. **253**:7638–7645.
23. **Preiss, J.** 1984. Bacterial glycogen synthesis and its regulation. Annu. Rev. Microbiol. **38**:419–458.
24. **Preiss, J., and D. A. Walsh.** 1981. The comparative biochemistry of glycogen and starch, p. 199–314. *In* V. Ginsburg and P. Robbins (ed.), Biology of carbohydrates, vol. 1. John Wiley and Sons, Inc., New York.
25. **Sanderson, K. E., and J. R. Roth.** 1983. Linkage map of *Salmonella typhimurium* edition VI. Microbiol. Rev. **47**:410–

453.
26. **Sanger, F., S. Nicklen, and A. R. Carlson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
27. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementary to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71**:1342–1346.
28. **Steiner, K. E., and J. Preiss.** 1977. Biosynthesis of bacterial glycogen: genetic and allosteric regulation of glycogen biosynthesis in *Salmonella typhimurium* LT2. J. Bacteriol. **129**: 246–253.