Volume 243, number 2, 193–198 FEB 06713 January 1989

Molecular cloning and sequencing of the glycogen phosphorylase gene from *Escherichia coli*

Yong-Lark Choi, Makoto Kawamukai*, Ryutaro Utsumi, Hiroshi Sakai and Tohru Komano

Laboratory of Biochemistry, Department of Agricultural Chemistry, Kyoto University, Kyoto 606 and **Laboratory of Applied Microbiology, Department of Agricultural Chemistry, Shimane University, Matsue 690, Japan*

Received 30 September 1988

The *glgP* gene, which codes for glycogen phosphorylase, was cloned from a genomic library of *Escherichia coli*. The nucleotide sequence of the *glgP* gene contained a single open reading frame encoding a protein consisting of 790 amino acid residues. The glgP gene product, a polypeptide of M_r 87 000, was confirmed by SDS-polyacrylamide gel electrophoresis. The deduced amino acid sequence showed that homology between *glgP* of *E. coli* and rabbit *glgP*, human *glgP*, potato *glgP*, and *E. coli malP* was 48.6, 48.6, 42.3, and 46.1%, respectively. Within this homologous region, the active site, glycogen storage site, and pyridoxal-5'-phosphate binding site are well conserved. The enzyme activity of glycogen phosphorylase increased after introduction on a multicopy of the *glgP* gene.

Glycogen phosphorylase; Gene cloning; Nucleotide sequence; Gene *glgP*; (*E. coli*)

1. INTRODUCTION

More than 50 years of research on glycogen phosphorylase in eukaryotes has yielded a rich harvest of ideas and principles, and the *glgP* genes have been isolated from several eukaryotic organisms [1–7], but this enzyme is as yet unknown in prokaryotes. There are at least three glycogen-synthesizing enzymes of bacteria, glycogen synthase, branching enzyme, and ADP-glucose pyrophosphorylase, which are the gene products of *glgA*, *glgB*, and *glgC*, respectively [8–11]. The glycogen catabolizing enzyme, glycogen phosphorylase, of bacteria has not yet been found.

We cloned *glgP*, the structural gene of glycogen phosphorylase of *E. coli*, and also described the nucleotide sequence, the deduced amino acid sequence, and the expression of the *glgP* gene. The primary structure of the glgP protein is compared with those of other organisms.

2. MATERIALS AND METHODS

2.1. *Bacterial strain and plasmids*

We used *E. coli* JM109 (*recA1*, *lac*, *pro*, *endA1*, *gyrA96*, *thi*, *hsdR17*, *relA1*, ⟨F′, *traD36*, *proA*$^+$*B*$^+$⟩) [12]. The source of the *glgP* gene sequenced in this study was the recombinant plasmid pGC1 and pGTC12 (fig.1), which was cloned from the genomic library of the *E. coli* into pUC18 [12]. Derivatives of pGTC12

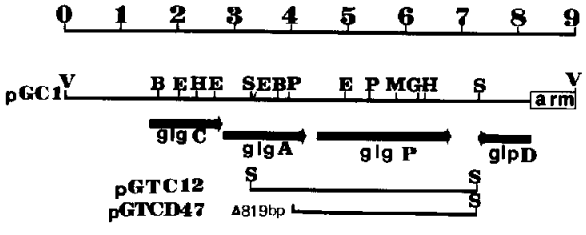

Fig.1. Physical map of the *glg* operon-*glpD* region of the *E. coli* chromosome. The initiation site and direction of transcription of the *glgC*, *A*, *P* and *glpD* genes are shown. B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; H, *Hin*dIII; M, *Sma*I; P, *Pst*I; S, *Stu*I; V, *Eco*RV.

Correspondence address: T. Komano, Laboratory of Biochemistry, Department of Agricultural Chemistry, Kyoto University, Kyoto 606, Japan

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50

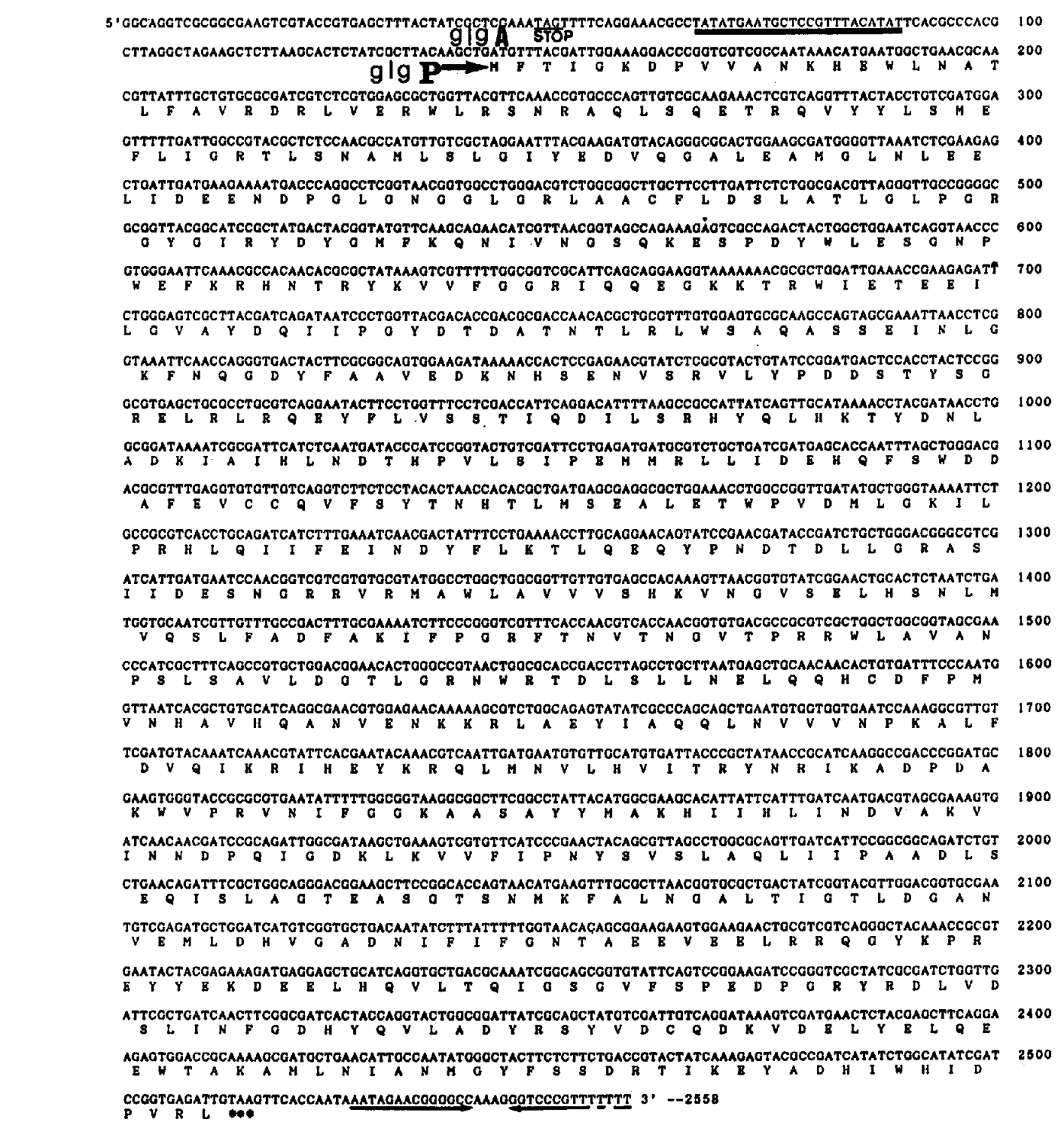

Fig.2. Nucleotide sequence of the *glgP* gene and its flanking region, and the amino acid sequence deduced for glycogen phosphorylase. The initiation sites of the translation are indicated by the thick arrows. The last 525 bp have been presented previously [20]. A candidate for the cAMP-CRP complex binding site is thickly underlined. Dyad symmetry thought to be involved in the transcriptional termination is indicated by the thin arrow.

with deletions by exonuclease III were constructed by standard procedures.

λ phage 16A5 of the *E. coli* genomic library was supplied by Kohara [13].

2.2. *Enzymes and chemicals*

All enzymes used for DNA treatment and analysis were purchased from Takara Shuzo, Boehringer and Sigma. Radioactively labeled materials were from Amersham International.

2.3. *DNA sequencing*

Determination of DNA sequence was performed by the dideoxy chain termination method using the bacteriophages M13mp18 and mp19 [14]. The 4.0 kb *Stu*I fragment of pGC1 was subcloned into M13 vectors. Deletion mutants were constructed by exonuclease III and mung bean nuclease. Plasmid DNA and phage DNA were prepared as in [15].

2.4. *Enzyme activity and analysis*

The crude enzyme extracts were prepared from cells grown overnight in 5 ml of M9 media containing 0.4% glycogen. The harvested cells were suspended in 1 ml of 25 mM sodium phosphate buffer (pH 7.5) containing 0.5% Triton X-100. After sonication for 90 s with intermediate cooling, the debris was removed by centrifugation. The cell extracts were assayed for glycogen and dextrin phosphorylase activity as in [16].

2.5. *Protein analysis*

Protein was measured by the method of Bradford [17]. SDS-polyacrylamide gel electrophoresis (PAGE) was done as in [18,19].

2.6. *Gel retardation assay*

An 869 base pair *Eco*RI fragment of GTCD47 was labeled at the 5′ ends using T4 polynucleotide kinase and [γ-^{32}P]dATP. The labeled DNA was digested with *Spl*I and the two fragments were purified by 5% PAGE. The isolated DNA fragments (<0.2 pmol) were incubated with the purified CRP protein (0.4 pmol) and cAMP (0.2 pmol) for 10 min at 25°C in a buffer containing 40 mM Tris-Cl, 10 mM $MgCl_2$, 100 mM KCl, 1 mM EDTA, 100 μg/ml BSA, pH 8.0, then analysed by 5% PAGE.

3. RESULTS

3.1. *Cloning of the* glgP *gene*

We have previously shown that a part of the *glgP* gene is located downstream from the *glpD* gene [20]. DNA from a recombinant phage showing a positive hybridization signal with the radioactive probe containing the C-terminus of *glgP* was digested with several restriction endonucleases. A positive 9.0-kb *Eco*RV fragment of the phage DNA was ligated into the *Sma*I site of pUC18 to construct pGC1. A detailed restriction map of this plasmid is presented in fig.1, and part of the map was consistent with those of the *glgAC* and *glpD* genes [8,10,20]. This clone was positive upon staining colonies with iodine for the presence of glycogen (not shown). These results show that the *glgP* gene is located between the *glgA* and *glpD* genes. The 4.0-kb *Stu*I fragment of the pGC1 was subcloned into pUC18, and this recombinant plasmid was designated pGTC12 (fig.1). The specific activity of glycogen phosphorylase of an *E. coli* strain harboring pGTC12 increased 5- to 7-fold (table 1).

Table 1

Specific activity of glycogen and maltodextrin phosphorylase

	Plasmid				
	pGC1	pGTC12	pGTCD9	pGTCD47	pUC18
Glycogen	0.007	0.033	0.058	0.071	<0.005
Maltodextrin	0.008	0.032	0.060	0.065	<0.006

Specific activity is defined as μmol of product, D-glucose 1-phosphate, from substrate and orthophosphate/mg protein per min

3.2. *Nucleotide sequence*

Fig.2 shows the nucleotide sequence of the *glgP* gene along with the 3′ and 5′ flanking regions. This sequence contained an open reading frame of 2370 base pairs coding for a polypeptide of 790 amino acid residues. The open reading frame terminated with the sequence TAA. Two other ORFs of the *glgA* and *glpD* were found upstream and downstream from the *glgP* gene, respectively, and this was consistent with the previous results [8,20].

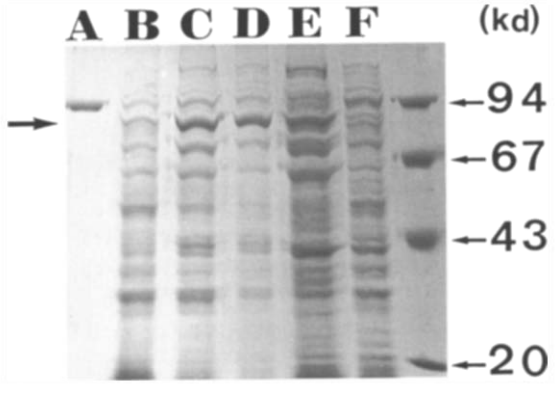

Fig.3. SDS-polyacrylamide gel electrophoretic patterns showing expression of the *E. coli glgP* gene. The thick arrow indicates the position of the band of glycogen phosphorylase of *E. coli*. Lanes: A, rabbit glycogen phosphorylase; B, pUC18; C, pGTCD9; D, pGTCD47; E, pGTC12; F, pGC1.

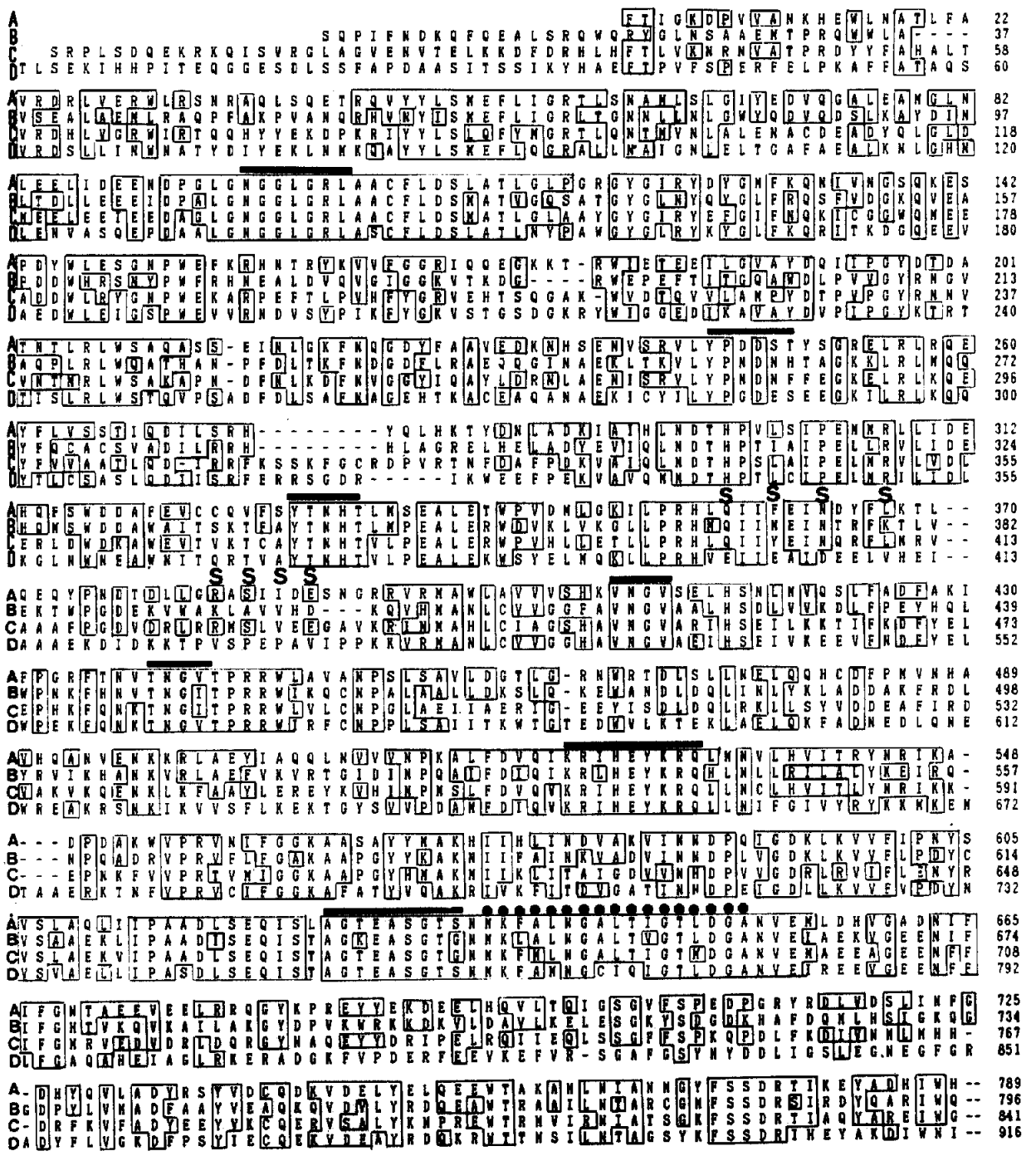

Fig.4. Comparison of the primary structure among glycogen phosphorylases from different organisms. Amino acids identical to those of the glgP protein of *E. coli* are shown in the boxes. Amino acid residues involved in catalytic or regulatory interaction of phosphorylase are marked as follows: solid bar, active site (glucose binding); S, glycogen storage site; ---, pyridoxal-phosphate cofactor binding site. Lanes: A, *glgP* of *E. coli*; B, *malP* of *E. coli*; C, *glgP* of rabbit muscle; D, glucan phosphorylase of potato.

A possible stem and loop structure (ΔG = −16.0 kcal) characteristic of the prokaryotic transcriptional termination site was found in the 3′ noncoding region of the *glgP* gene. A potential cAMP-CRP binding site is located at 55 base pairs upstream from the translational starting point of the *glgP* gene.

3.3. *Expression of the* glgP *gene*

The specific activities of glycogen and dextrin phosphorylase in the strains of *E. coli* with and without the multicopy plasmids containing the *glgP* gene are summarized in table 1. *E. coli* JM109 harboring pGTC12 showed 5- to 7-fold higher levels of glycogen phosphorylase activity than that of JM109 harboring the vector plasmid pUC18. To increase the enzyme activity, we constructed the derivative plasmids pGTCD9 and pGTCD47, which had 650 bp and 819 bp deleted upstream of the pGTC12. The enzyme activities of JM109 harboring pGTCD9 and pGTCD47 increased 12- to 14-fold, respectively. All these plasmids containing the *glgP* gene had also specified the maltodextrin phosphorylase (*malP*) activity. These results confirm the idea that the location of the *glgP* gene should be more finely specified based on the results of the deletion mutants. Fig.3 shows the electrophoretic patterns of SDS-PAGE of whole cell proteins from *E. coli*. The gene product (thick arrow) of cloned *glgP* was obtained as shown by comparative 10% SDS-PAGE. The molecular mass of 87 kDa is consistent with that deduced from the DNA sequence.

Expression of *glgP* in eukaryotes is regulated by the cAMP system [21,22]. To find whether a binding site for CRP existed in the *glgP* regulatory region, a gel retardation assay for CRP was done. Purified CRP was incubated in the presence of cAMP with the *Eco*RI-*Spl*I fragment. There was a specific and CRP-dependent retardation in electrophoresis of the fragment bearing the *glgP* flanking region, but the retardation was not observed in the absence of cAMP (fig.5), neither was retardation in electrophoresis detected in the control fragment. This result implied that expression of the *glgP* gene was dependent on the cAMP-CRP complex.

3.4. *Properties of the deduced glgP protein*

The primary structure of glycogen phosphorylase, deduced from the *glgP* nucleotide sequence is shown in fig.2. Assuming the initiating formyl methionine residue was not removed, the glgP protein comprised 790 amino acids and had a predicted molecular mass of 90147 Da. The deduced amino acid sequences of *glgP* genes of *E. coli* and several other organisms are compared in fig.4. Homologies in the amino acid sequences between *E. coli glgP* and rabbit *glgP* [3], human *glgP* [5], potato *glgP* [6], and *E. coli malP* [23] were 48.6, 48.6, 42.3, and 46.1%, respectively.

Extensive biochemical and crystallographic study of rabbit muscle phosphorylase found several regulating regions [2,4,24,25]. In the catalytic region, the pyridoxal-5-phosphate binding site and the active site are extensively conserved in the *E. coli* enzyme (fig.4). At the glycogen storage site, only two of the eight residues involved in substrate binding are different from those of eukaryotes. Asp 247, which forms an important part of the loop interaction with Arg 531, is also conserved. Purines bind in a hydrophobic pocket formed by the Phe 249 and Tyr 569; one of these residues (Tyr) is conserved in the *E. coli* enzyme. The glucose binding loop (Asn 248) is not conserved. The poor alignment of glucose and purine binding sites suggested that the mode of regulation of the *E. coli* enzyme is different from that of the eukaryotic enzyme.

4. DISCUSSION

In this report, we have shown the existence of a glycogen phosphorylase in *E. coli*. While Okita et al. [26] reported cloning of the *glgA, B,* and *C* genes from *E. coli*, they thought that downstream of the *glgA* gene would be a *glgR* region coding for a repressor-like protein of glycogen biosynthesis. It is fortuitous that the truncated form of the *glgP* gene is located next to the 3′ end of *glpD* gene on the *E. coli* chromosome and that the operon of glycogen synthesis was closely linked to the *glpD* gene [20]. Though the *glgP* gene cannot be selected directly, we have cloned it from the *E. coli* genomic library [13]. The glycogen phosphorylase activity increased 5–14-fold in *E. coli* strains with a hybrid plasmid containing the *glgP* gene. Nevertheless, overproduction of the glycogen phosphorylation of the *E. coli* remains under the control of the cAMP-CRP complex (fig.5). To understand

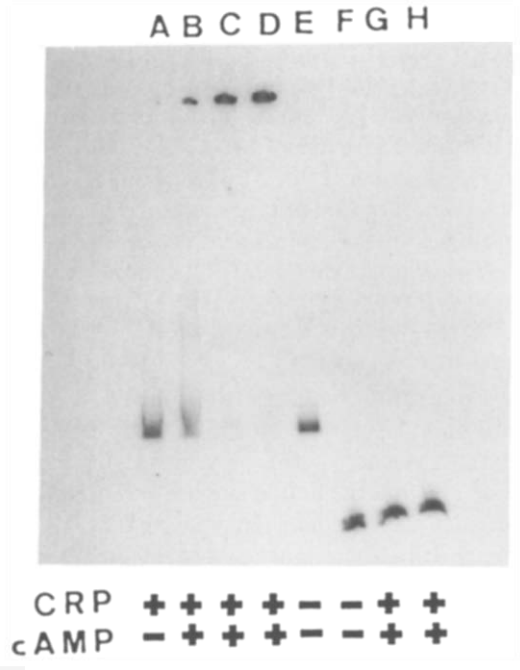

Fig.5. Gel retardation assay for CRP binding to the *glgP* regulation region. DNA segments used for the complex formation: lanes A–E, the DNA segment from the regulatory region of *glgP*; lanes F–H, the DNA segment from the *glgP* coding sequence (note nos 313–605 in fig.2). CRP concentration: B, 0.2 pmol; A,C and G, 0.4 pmol; D and H, 0.6 pmol.

the regulation of *glgP* gene, it will be necessary to study the promoter and regulator region at the transcriptional level.

At the N- and C-termini, the length of *E. coli* glycogen phosphorylase is 36 and 10 amino acid residues shorter than the rabbit muscle enzyme, respectively, and the two regions are not homologous with the *glgP* gene family. A striking structural difference in several species of eukaryotes occurs at the C-terminus of the enzyme [6,26]. The difference in N- and C-termini strongly suggests that this region does not play a major role in the enzyme function although the N-terminus of the eukaryote enzyme contains a covalent phosphorylation site.

Despite the difference in mode of regulation of the glycogen phosphorylase activity between eukaryotes and prokaryotes, the active site, the glycogen storage site, and the pyridoxal-5-phosphate binding site are extensively conserved. This makes glycogen phosphorylase highly attractive for investigators of the evolution of regulatory mechanisms.

Acknowledgements: We would like to thank Dr H. Aiba for supplying purified CRP and Dr Y. Kohara for the genomic library of *E. coli*.

REFERENCES

[1] Fletterick, R.J. (1980) Annu. Rev. Biochem. 49, 31–61.
[2] Fletterick, R.J. and Sprang, S.R. (1982) Acc. Chem. Res. 15, 361–369.
[3] Nakano, K., Hwang, P.K. and Fletterick, R.J. (1986) FEBS Lett. 204, 283–287.
[4] Hwang, P.K. and Fletterick, R.J. (1986) Nature 324, 80–84.
[5] Newgard, C.B., Littman, D.R., Genderen, C.V., Smith, M. and Fletterick, R.J. (1988) J. Biol. Chem. 263, 3850–3857.
[6] Nakano, K. and Fukui, T. (1986) J. Biol. Chem. 261, 8230–8236.
[7] Newgard, C.B., Nakano, K., Hwang, P.K. and Fletterick, R.J. (1986) Proc. Natl. Acad. Sci. USA 83, 8132–8136.
[8] Baecker, P.A., Furlong, C.E. and Preiss, J. (1983) J. Biol. Chem. 258, 5084–5088.
[9] Baecker, P.A., Greenberg, E. and Preiss, J. (1986) J. Biol. Chem. 261, 8738–8743.
[10] Kumar, A., Larsen, C.E. and Preiss, J. (1986) J. Biol. Chem. 261, 16256–16259.
[11] Leung, P.S.C. and Preiss, J. (1987) J. Bacteriol. 169, 4349–4360.
[12] Yanisch-Perron, Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
[13] Kohara, Y., Akiyama, K. and Isono, K. (1987) Cell 50, 495–508.
[14] Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, pp.79–96, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[16] Helmreich, E. and Cori, C.F. (1964) Proc. Natl. Acad. Sci. USA 51, 131–138.
[17] Bradford, M. (1976) Anal. Biochem. 72, 248–254.
[18] Larson, T.J., Ludtke, D.N. and Bell, R.M. (1984) J. Bacteriol. 160, 711–717.
[19] Laemmli, U.K. (1970) Nature 227, 680–685.
[20] Choi, Y.L., Kawase, S., Kawamukai, M., Utsumi, R., Sakai, H. and Komano, T. (1988) submitted.
[21] Cohen, P. (1982) Nature 296, 613–620.
[22] Smith, S.B., White, H.D., Siegel, J.B. and Krebs, E.G. (1981) Proc. Natl. Acad. Sci. USA 78, 1591–1595.
[23] Palm, D., Goerl, R., Weidinger, G., Zeier, R., Fisher, B. and Schinzel, R. (1987) Z. Naturforsch. Teil C 42, 394–400.
[24] Fletterick, R.J. and Madsen, M.B.A. (1980) Rev. BIochem. 831–861.
[25] Dombradi, V. (1981) Int. J. Biochem. 13, 125–139.
[26] Okita, T.W., Rodriguez, R.L. and Preiss, J. (1981) J. Biol. Chem. 256, 6944–6952.