Glycerol catabolism in *Bacillus subtilis*: nucleotide sequence of the genes encoding glycerol kinase (*glpK*) and glycerol-3-phosphate dehydrogenase (*glpD*)

CHRISTINA HOLMBERG,* LENA BEIJER, BLANKA RUTBERG and LARS RUTBERG

Department of Microbiology, University of Lund, Sölvegatan 21, S-223 62 Lund, Sweden

(Received 22 May 1990; revised 10 July 1990; accepted 9 August 1990)

The *glpPKD* region of the *Bacillus subtilis* chromosome was cloned in its natural host in plasmid pHP13. The *glpPKD* region contains genes required for glycerol catabolism: *glpK* coding for glycerol kinase, *glpD* coding for glycerol-3-phosphate (G3P) dehydrogenase and *glpP*, proposed to code for a positively acting regulatory protein. The cloned 7 kb fragment carries wild-type alleles of *glpK*, *glpD* and *glpP*. It can also complement a strain deleted for the entire *glpPKD* region. The wild-type alleles were mapped to different subfragments, establishing the gene order *glpP*-*glpK*-*glpD*. The nucleotide sequence of *glpK* and *glpD* was determined. Immediately upstream of *glpK*, an additional open reading frame was found, possibly being part of the same operon. Putative transcription terminators were found in the region between *glpK* and *glpD* and downstream of *glpD*. In a coupled *in vitro* transcription/translation system, two proteins were found, corresponding in size to those predicted from the deduced amino acid sequences of glycerol kinase and G3P dehydrogenase (54 kDa and 63 kDa, respectively).

Introduction

Many bacteria can use glycerol or glycerol 3-phosphate (G3P) as the sole source of carbon and energy. The main pathway of glycerol dissimilation involves a glycerol kinase (EC 2.7.1.30) which phosphorylates glycerol to G3P, and a G3P dehydrogenase (EC 1.1.99.5) which oxidizes G3P to dihydroxyacetone phosphate, an intermediate in glycolysis. This pathway is the only one known in *Bacillus subtilis*, the bacterium employed in this study (Mindich, 1968; Lindgren & Rutberg, 1974; Lin, 1976). Glycerol catabolism has been studied extensively in *Escherichia coli*, where four probable operons, constituting a glycerol (*glp*) regulon, have been identified. The *glpFK* operon codes for a membrane protein which facilitates diffusion of glycerol across the cytoplasmic membrane, and for glycerol kinase (Pettigrew *et al.*, 1988; Sweet *et al.*, 1990). *glpTQ* codes for G3P permease and for a periplasmic glycerophosphodiester phosphodiesterase (Larson *et al.*, 1982, 1983). This operon is

closely linked to but inversely oriented to *glpACB*, coding for the anaerobic G3P dehydrogenase (Ehrmann *et al.*, 1987; Cole *et al.*, 1988). A fourth linkage group contains four genes: *glpR* encoding the *glp* repressor, *glpD* encoding the aerobic G3P dehydrogenase, and *glpE* and *glpG*, two genes of unknown functions (Schweizer *et al.*, 1985, 1986; Schweizer & Larson, 1987). Regulation of the *E. coli* *glp* regulon is complex. The *glp* genes are repressed to various degrees by the *glp* repressor. Binding of G3P to the repressor decreases its affinity for the operator regions, thus allowing for induction of the *glp* regulon. The *glp* regulon is sensitive to glucose repression and is dependent on the cAMP–CRP complex for full expression (Lin, 1976). At least *glpD* and *glpACB* are also influenced by regulatory circuits involved in the general control of aerobic and anaerobic metabolism (Iuchi *et al.*, 1990). In bacteria other than *E. coli* relatively little is known about the organization and regulation of genes involved in glycerol catabolism.

In *B. subtilis*, four *glp* genes have been mapped. Three genes coding for glycerol kinase (*glpK*), an NAD-independent G3P dehydrogenase (*glpD*) and a regulatory element (*glpP*) are closely linked and map at 75° on the *B. subtilis* chromosomal map (Lindgren & Rutberg, 1974). The fourth gene, *glpT*, codes for G3P permease and maps at 15° (Lindgren, 1978). In *B. subtilis*, as in *E. coli*, *glpK*, *glpD* and *glpT* are repressed by glucose and

Abbreviations: Cm, chloramphenicol; Em, erythromycin; G3P, glycerol 3-phosphate; ORF, open reading frame; RBS, ribosome-binding site; TBAB, tryptose blood agar base.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number M34393.

induced by G3P. Mutations in *glpP* render the former three genes non-inducible, which has led to the suggestion that *glpP* codes for a positively acting regulatory protein (Lindgren & Rutberg, 1976). In order better to understand the genetic organization and regulation of glycerol catabolism in *B. subtilis*, we have started to clone the *glp* genes. In a recent report we described the cloning in *E. coli* of part of the *glpPKD* region (Holmberg & Rutberg, 1989). The results indicated that a plasmid carrying the intact *glpPKD* region could not be recovered in *E. coli*. In the present communication we describe the cloning of the entire *glpPKD* region in its natural host and present the nucleotide sequence of *glpK* and *glpD*.

Methods

Strains and plasmids. These are listed in Table 1. For M13, *E. coli* JM101 or JM103 were used as hosts. pUC18 or 19 subclones were maintained in *E. coli* JM83.

Media. All bacterial strains were kept on tryptose blood agar base (TBAB) plates. The minimal salts solution was that of Anagnostopoulos & Spizizen (1961). Glucose or glycerol was added at a concentration of 5 g l⁻¹. When glycerol was the sole source of carbon and energy,

citrate was omitted from the minimal salt solution. Minimal plates contained 15 g agar l⁻¹. Antibiotics were added to the following concentrations: ampicillin 50 mg l⁻¹, chloramphenicol (Cm) 12.5 mg l⁻¹ (*E. coli*) or 5 mg l⁻¹ (*B. subtilis*), erythromycin (Em) 5 mg l⁻¹ and phleomycin 0.15 mg l⁻¹.

Transformation. Competent *B. subtilis* cells were prepared as described by Arwert & Venema (1973). *E. coli* cells were made competent by the CaCl₂ method described by Mandel & Higa (1970).

DNA techniques. Chromosomal *B. subtilis* DNA was prepared essentially as described for plasmid DNA by Canosi *et al.* (1978). Plasmid DNA was prepared by the alkaline lysis method of Ish-Horowicz & Burke (1981). When preparing plasmid DNA from *B. subtilis*, two steps were added to the procedure. The cells were treated with lysozyme (5 mg ml⁻¹) prior to lysis with NaOH/SDS and then the DNA preparation was treated with pronase (0.1 mg ml⁻¹) prior to precipitating the plasmid DNA. Southern blots and agarose gel electrophoresis were performed as described by Maniatis *et al.* (1982). DNA fragments separated on agarose gels were isolated using GeneClean (Bio 101 Inc.). Restriction endonucleases and T4 DNA ligase were purchased from Boehringer Mannheim and used as recommended by the manufacturer.

Enzyme assays. *B. subtilis* cell-free extracts were prepared as described by Lindgren & Rutberg (1974). Cells were grown at 37 °C in NSMP medium (Fortnagel & Freese, 1968). At early exponential phase (OD₆₀₀ = 0.3), the culture was divided into two parts and glycerol (5 g l⁻¹) was added to one part. After 2 h incubation, samples were

Table 1. Bacterial strains and plasmids

Strain/plasmid	Genotype/phenotype	Source/reference
<i>B. subtilis</i>		
W168	Wild-type	J. Spizizen (University of Arizona)
BR95	<i>trpC2 pheA1 ilvC2</i>	Our collection
LUG3621	<i>trpC2 pheA1 glpK21</i>	
LUG4806	<i>trpC2 glpD6</i>	
LUG2518	<i>trpC2 ilvC2 glpP18</i>	
LUG2512	<i>trpC2 ilvC2 glpP12</i>	
LUG2509	<i>trpC2 ilvC2 glpP9</i>	
LUG2506	<i>trpC2 ilvC2 glpP6</i>	
LUG3111	<i>trpC2 glpK11 recE4</i>	
LUG4108	<i>trpC2 glpD8 recE4</i>	This work
LUG0401	<i>trpC2 pheA1 ilvC2 ΔglpPKD ble</i>	
<i>E. coli</i>		
MM294	<i>pro thi endA hsr_x hsr_m⁺</i>	Amann <i>et al.</i> (1983)
JM83	<i>ara Δ(lac-proAB) rpsL φ80 lacZΔM15</i>	Yanisch-Perron <i>et al.</i> (1985)
JM101	<i>thi Δ(lac-proAB) (F' traD36 proAB lacP⁺ΔM15)</i>	Yanisch-Perron <i>et al.</i> (1985)
JM103	<i>endA1 hsdR supE thi-1 sbcB15 strA Δ(lac-proAB) (F' traD36 proAB lacP⁺ΔM15)</i>	Messing <i>et al.</i> (1981)
Plasmid/phage		
pHP13	Cm ^R Em ^R	Haima <i>et al.</i> (1987)
pLUM20	Ap ^R Tc ^R Cm ^R	Holmberg & Rutberg (1989)
M13	mp18/mp19	Norlander <i>et al.</i> (1983)
pUC18/19	Ap ^R	Norlander <i>et al.</i> (1983)
pHSG575/576	Cm ^R	Takeshita <i>et al.</i> (1987)
pUB110	Km ^R Pm ^R	Gryczan <i>et al.</i> (1978)

taken from both cultures. Glycerol kinase was measured as described by Freedberg & Lin (1973) but with the buffer at pH 9.0 instead of 9.5. G3P dehydrogenase was measured as described by Lin *et al.* (1962). The amount of protein was determined by the Lowry method, with bovine serum albumin as standard.

DNA sequencing. The nucleotide sequence was determined by the dideoxy-chain-termination method of Sanger *et al.* (1977). DNA fragments were cloned into M13mp18/19 or pUC18/19. Sequencing was performed using Sequenase (United States Biochemical Corporation) and M13/pUC or *glp* specific primers. The nucleotide sequence obtained was analysed using the GCG Sequence Analysis Software Package, version 5.3 (Devereux *et al.*, 1984).

Other techniques. Plasmid-encoded proteins were analysed in a coupled *E. coli* *in vitro* transcription/translation assay using the prokaryotic DNA-directed translation kit from Amersham and used as recommended by the manufacturer. The protein products, labelled with L-[³⁵S]methionine (29.6 TBq mmol⁻¹), were separated by SDS-PAGE according to Neville (1971). Bovine serum albumin (68 kDa), catalase (58 kDa), fumarase (48 kDa), carbonic anhydrase (29 kDa), myoglobin (17 kDa) and lysozyme (14 kDa) were used as molecular mass markers.

Results and Discussion

Cloning of *glpP*, *glpK* and *glpD*

We have previously described the cloning in *E. coli* of a 3.2 kb *B. subtilis* chromosomal DNA fragment which contains wild-type alleles of *glpK* and *glpD* mutations (Holmberg & Rutberg, 1989). The cloning was done using pHV32 and the resulting plasmid was called pLUM20. The insert of pLUM20 weakly complements a *B. subtilis glpK* mutation but not a *glpD* mutation in a *recE4* background. Attempts to clone the entire *glpPKD* region from *B. subtilis* in *E. coli* using different strategies, including use of λ vectors and low-copy-number plasmids, have all failed, indicating that the *glpPKD* region cannot be stably maintained in *E. coli*. We have therefore used the *E. coli*-*B. subtilis* shuttle plasmid pHP13 to clone the *glpPKD* region in its natural host. pHP13 confers resistance to Cm and Em in both bacteria.

Fig. 1. Physical map of the *glp* region with the chromosomal fragments represented in pLUM20 and pLUM30. Restriction endonuclease sites: A, *AccI*; C, *Clal*; D, *DraI*; E, *EcoRI*; H, *HindIII*; N, *NcoI*; P, *PstI*; Pv, *PvuII*; S, *SphI*; Sa, *SacI*. The *Clal* site is subject to *dam* methylation in *E. coli*.

Southern blots of restricted *B. subtilis* chromosomal DNA using pLUM20 as a probe had indicated that *glpK* and *glpD* are contained within a 7 kb *EcoRI*-*PstI* fragment. Chromosomal *B. subtilis* W168 DNA was digested with *EcoRI* and *PstI*. Fragments ranging from 6 to 10 kb were enriched by sucrose-gradient centrifugation and ligated to pHP13, which had been digested with *EcoRI* and *PstI*. The ligation mixture was used to transform *B. subtilis* LUG2518 (*glpP18*), and GlpP⁺ transformants able to grow on glycerol as the sole carbon source were selected. The transformants were replica-plated to TBAB-plates containing Cm. Fifty-one Cm^R clones were found among 5000 GlpP⁺ transformants. Plasmid DNA was prepared from Glp⁺ Cm^R transformants and used to transform LUG2518 to Cm^R or GlpP⁺. One plasmid, called pLUM30, which showed 100% linkage between *cat* and *glpP18* was isolated and used in further experiments. pLUM30 also transforms *glpK* and *glpD* mutants to Glp⁺. The plasmid carries a 6.8 kb *EcoRI*-*PstI* insert and is stably maintained in *B. subtilis* in the presence of Cm or Em.

Characterization of pLUM30

A restriction map of the cloned *glp* region is shown in Fig. 1. A comparison between Southern blots of

Table 2. Activity of glycerol kinase and G3P dehydrogenase in LUG0401 carrying pHP13 or pLUM30

Strain	Inducer	Specific activity*	
		Glycerol kinase	G3P dehydrogenase
LUG0401(pHP13)	—	<0.5	<0.5
	20 mM-glycerol	<0.5	<0.5
LUG0401(pLUM30)	—	4	4
	20 mM-glycerol	12	9

* Specific activities are expressed as nmol substrate converted min⁻¹ (mg protein)⁻¹.

Fig. 2. Autoradiogram of [^{35}S]methionine-labelled proteins from *in vitro* transcription/translation assays, separated on a 10–16% (w/v) gradient acrylamide gel. Approximately 10^5 c.p.m. were loaded on each lane. Lane 1 contains proteins derived from pLUM30 DNA, lane 2 proteins from pHP13 DNA. Std, molecular mass standards; these were stained with Coomassie brilliant blue.

restricted *B. subtilis* chromosomal DNA and the insert in pLUM30 indicates that the insert represents one continuous segment of the chromosome. The 3 kb *Cla*I–*Pst*I fragment in pLUM30 is identical to that of the insert of pLUM20, as indicated by Southern blots (data not shown). pLUM30 also complements a *glpK* and a *glpD* mutation in a *recE4* background. To fully test the functionality of the *glp* genes in pLUM30, a deletion mutant was constructed. The 5 kb *Sac*I–*Pst*I fragment in pLUM30 was deleted from the chromosome of *B. subtilis*

BR95 and replaced with a DNA fragment from pUB110 containing the *ble* gene, according to principles previously described (Fridén *et al.*, 1987), to give strain LUG0401. Southern blots of restricted chromosomal DNA from LUG0401 and BR95, using pLUM30 as a probe, were used to verify the nature of the deletion in LUG0401. This strain is unable to grow on glycerol and has no detectable activities of glycerol kinase or G3P dehydrogenase. Transformation of LUG0401 by pLUM30 restores the ability to grow on glycerol. Glycerol kinase and G3P dehydrogenase activities were found in extracts of glycerol-induced LUG0401 (pLUM30) (Table 2). These results demonstrate that pLUM30 contains functional *glpK* and *glpD* genes and the control regions required for their induction.

In a coupled *in vitro* transcription/translation assay, pLUM30 DNA was found to encode two proteins, of 63 kDa and 54 kDa. These proteins were not found when pHP13 DNA was used (Fig. 2). We suggest that the 63 kDa protein represents G3P dehydrogenase and the 54 kDa protein represents glycerol kinase (see below).

In order to map the relative positions of *glpP*, *glpK* and *glpD* on the 6.8 kb *Eco*RI–*Pst*I fragment in pLUM30, subfragments were cloned and tested for their ability to rescue various *glp* mutations by means of recombinational repair. The results are summarized in Fig. 3. Most fragments were subcloned in *E. coli*. However, the *Hind*III–*Cla*I fragment and the *Hind*III–*Nco*I fragment, both containing at least part of *glpP*, could not be cloned intact in *E. coli*. They were therefore subcloned in *B. subtilis* using pHP13. The sites of mutation in *glpP6*, *glpP9* and *glpP12* are located on a 800 bp *Hind*III–*Nco*I fragment. The *glpP18* mutation, however, cannot be placed within a defined subfragment. Only pLUM30 can rescue this mutation, indicating that *glpP18* may not represent a point mutation. The *glpK21* mutation appears to be located on a 800 bp *Cla*I–*Acc*I fragment and *glpD6* on a 600 bp *Sph*I–*Pvu*II fragment. These results establish the gene order *glpPKD*, which is in agreement with the results of earlier transformation crosses (Lindgren & Rutberg, 1974). The relative order of *glpK* and *glpD* does not agree, however, with results previously obtained using nuclease *Bal*31-generated deletions of the insert in pLUM20 (Holmberg & Rutberg, 1989). We do not understand the reason for this inconsistency. However, the gene order presented here is clearly correct since it is verified by the nucleotide sequence (see below). The above results position *glpPKD* on a 4.7 kb *Hind*III fragment. Attempts to clone this fragment in the *E. coli* low-copy-number vectors pHSG575/576 have only yielded plasmids where parts of the *glp* region have been deleted (data not shown). This confirms that an intact *glp* region cannot be maintained in *E. coli*.

Fig. 3. Relative order of *glpP*, *glpK* and *glpD* markers on the cloned fragment. Fragments were subcloned into pHP13 or pUC18/19. The generated plasmids were tested for their ability to transform various Glp mutants (carrying the mutations *glpP6*, *glpK21* and *glpD6*, respectively) to Glp⁺. Mutants carrying the *glpP9* and *glpP12* mutations gave the same result as the mutant carrying the *glpP6* mutation. Restriction endonuclease sites: A, *AccI*; C, *ClaI*; E, *EcoRI*; H, *HindIII*; N, *NcoI*; P, *PstI*; Pv, *PvuII*; S, *SphI*.

Nucleotide sequence

The above experiments indicate that *glpK* and *glpD* are contained on a 4.2 kb *NcoI*–*EcoRI* fragment (Fig. 1). The nucleotide sequence of this fragment was determined using subcloned fragments from pLUM20 and/or pLUM30. The nucleotide sequence is shown in Fig. 4. Four major open reading frames (ORFs) were found.

The first ORF starts upstream of the *NcoI* site and ends at position 679 with a TAA stop codon. The predicted amino acid sequence of ORF1 indicates that it codes for a hydrophobic protein. In *E. coli*, the *glpF* gene which codes for the glycerol facilitator protein is situated upstream of *glpK* and is part of the same operon (Sweet *et al.*, 1990). Based on the hydrophobic character of the predicted ORF1 protein and the positioning of ORF1 immediately upstream of *glpK* (see below), we speculate that the ORF1 protein may be involved in entry of glycerol. The gene for G3P permease has been mapped to another part of the *B. subtilis* chromosome (Lindgren, 1978).

The second ORF starts with an ATG at position 698. Preceding this ATG by 12 nucleotides there is a potential ribosome-binding site (RBS), -AAAGGGG GA-, with a calculated $\Delta G = -91$ kJ (Tinoco *et al.*, 1973), a value in the range of strong *B. subtilis* RBSs (Hager & Rabinowitz, 1985). This ORF ends at position 2188 with a TAA stop codon. It is predicted to code for a protein with a molecular mass of 55 kDa, which may correspond to the 54 kDa protein found in the *in vitro* transcription/translation assay using pLUM30 DNA. The predicted amino acid sequence shows 62% identity with that predicted for the *glpK* protein of *E. coli* (Pettigrew *et al.*, 1988). This fact, together with the finding that the

glpK21 allele maps within this ORF, makes us confident that it represents the *B. subtilis glpK* gene. Glycerol kinase is an ATP-requiring enzyme. Comparisons of the amino acid sequences of a number of ATP-binding proteins have revealed common motifs, suggested to be important for ATP binding (Fry *et al.*, 1986; Higgins *et al.*, 1986). Two such regions have been identified in the *E. coli* glycerol kinase. One of these is conserved in both the *E. coli* and the *B. subtilis* proteins (amino acids 153–165 of *glpK* in Fig. 4). The other region, which contains a glycine-rich motif, is not conserved in *B. subtilis*. However, a possible glycine-rich consensus region is found at a different position (amino acids 258–266) in the *B. subtilis* enzyme. The gene presumably encoding glycerol kinase in *Streptomyces coelicolor*, *gylA*, has recently been cloned and the 5' part of the coding sequence has been published (Smith & Chater, 1988). Surprisingly, the predicted N-terminal part of the GylA protein shows no significant similarity to those predicted for either the *E. coli* or the *B. subtilis* glycerol kinase.

The third ORF starts with an ATG at position 2329 and ends with a TAA at position 3996. Potential start codons are also found at positions 2332, 2368 and 2380. Preceding the first ATG codon by 10 nucleotides there is a potential RBS, -AAGGAGG- ($\Delta G = -74$ kJ). No obvious RBS precedes the ATG codons at positions 2368 and 2380. Assuming that the ATG at position 2329 is the start codon, this ORF can code for a protein with a molecular mass of 63 kDa, which may correspond to the 63 kDa protein found in the *in vitro* transcription/translation assay using pLUM30 DNA. A comparison of the predicted amino acid sequence of this ORF with those of the aerobic (GlpD) and anaerobic (GlpA) G3P dehydrogenases of *E. coli* (T. J. Larson, personal communication;

1	<i>NcoI</i> ccatggcggatacgcggttgccgcatcagcgggtgccatttgaatccgg	1101	TCTCCGGCACGAAGGTGAAGTGGATTTTAGACAATGTGAAGGCGCAAGA S G T K V K W I L D N V E G A R	151
51	ORF1 cgctaacgatagcgccttgctattgtaggagattttccttggaagaggtt	1151	GAAAAAGCGGAAAAAGGCGAGCTGCTGTTTGGAACGATTGATACGTGGCT E K A E K G E L L F G T I D T W L	168
101	cgggtttatattgcggcgcaaatgatcggagcGATCATCGGTGCGGTGAT I I G A V I	1201	CATTTGGAAAAATGTCAGGCGGAAAAGCGCATGTGACCGATTACTCCAATG I W K M S G G K A H V T D Y S N A	185
151	TATTTATCTGCATTACCTCCCGCACTGGAAGTCAACGGATGATCCCGCTG I Y L H Y L P H W K S T D D P A A	1251	CCTCAAGAACTGATGTTTAAATTTACGATTTAAAATGGGACGATCAA S R T L M F N I Y D L K W D D Q	201
201	CCAAGCTGGGTGTTTCTCAACAGGCGCCAGCATTCCGCATACATTGCA K L G V F S T G P S I P H T F A	1301	CTGCTCGACATTCTAGGCGTACCGAAATCCATGCTCCCTGAAGTGAAGCC L L D I L G V P K S M L P E V K P	218
251	AACGTTTTAAGCGAAGTGATTGGGACATTGTCTTGTACTTGAATCTT N V L S E V I G T F V L V L G I L	1351	GTCCTCTCATGTGTATGCGGAGACTGTTGATTACTTCTTCGAAAAA S S H V Y A E T V D Y H F F G K N	235
301	GGCCATAGGTGCAATCAATTTACAGAAGGACTTAATCCTTAATCGTCG A I G A N Q F T E G L N P L I V G	1401	ATATCCCGATTGCTGGAGCGGAGGCGACGACGAGTCCGATTGTTCCGGC I P I A G A A G D Q Q S A L F G	251
351	GTTTCCTCATTGTAGCAATCGGTATTCTCTGGGAGGCCACCGGCTAT F L I V A I G I S L G G T T G Y	1451	<i>SphI</i> CAGGCATGCTTTGAAGAAGCATGGGAAAAACACTTACGGCACAGGATG Q A C F E E G M G K N T Y G T G C	268
401	GCTATCAATCCTGCACGTGACTTAGGTCGCGGATCGCCACGCTTTTCT A I N P A R D L G P R I A H A F L	1501	TTTCATGCTGATGAATACGGGGAAAAAGCAATTAAGTCCGAACATGGGC F M L M N T G E K A I K S E H G L	285
451	TCCGATTCGGGGGAAGGGTCATCAAACGGAATACGCATGGGTCCGG P I P G K G S S N W K Y A W V P V	1551	TTTTGACAACAATCGCTTGGGGCATTGACGGAAAAAGTGAACATGCGTTA L T T I A W G I D G K V N Y A L	301
501	TAGTCGGCCGATTTTAGCGGATCATTCGGCGGTGATTTTACAACGCT V G P I L G G S F G G V F Y N A	1601	GAAGGGAGCATTTTGTGCGAGGCTCTGCCATCCAGTGGCTTAGAGACGG E G S I F V A G S A I Q W L R D G	318
551	<i>DraI</i> GCATTTAAAGGGACATCACAAGCAGCTTCTGGATTGTAAGCGTTATATT A F K G H I T S S F W I V S V I L	1651	TTTGAGAATGTTCCAGGATTCATCGCTAAGCGAATCTTATGCAGAAAAAG L R M F Q D S S L S E S Y A E K V	335
601	GGTTGTGGTATTGTTAGGACTCTATGTTTATACGAAATCACATTCTGCTA V V V L L G L Y V Y T K S H S A K	1701	TGGATTCAACTGACGGCGTGTATGTTGTTCCAGCATTTGTCGGACTGGGA D S T D G V Y V V P A F V G L G	351
651	AAACATTATCAAAATCTAATATATTTAATCAAGGGGAGACATCTTATG T L S N S K Y I * M	1751	ACGCCTTACTGGGACAGCGATGTGCGCGGTTTCGGTTTTTCGGCTGACAAG T P Y W D S D V R G S V F G L T R	368
701	GAAACGTACATTTATCCTTAGATCAGGGGACGACAAGTTCAAGAGCGAT E T Y I L S L D Q G T T S S R A I	1801	AGGGACAACAAAAGAGCACTTTATCCGTCGCACTGGAGTCATTGGCTT G T T K E H F I R A T L E S L A Y	385
751	TCTGTTTAAATAAGGCAAAATGTCCACTCTGCTCAAAAGGAATTTA L F N K E G K I V H S A Q K E F T	1851	ATCAGACCAAGATGTGCTTGACGCAATGGAAGCAGATTCAAACATTTCA Q T K D V L D A M E A D S N I S	401
801	CACAATACTTCCCGCATCTGGCTGGGTTGAGCATAATGCCAATGAAATT Q Y F P H P G W V E H N A N E I	1901	<i>AccI</i> TTAAAGACGCTCCGTGTAGACGGAGGAGCTGTAAAAACAATTTCTAAT L K T L R V D G G A V K N N F L M	418
851	TGGGGCTCTGTCCTCGCGGTTATCGCCTCAGTCATCTCTGAATCAGGAAT W G S V L A V I A S V I S E S G I	1951	GCAGTTCAAAGGAGACCTGTTGAATGTTCTGTGGAGCGCCGGAATTA Q F Q G D L L N V P V E R P E I N	435
901	CAGCGCTTCTCAAATTGCCGGCATCGGCATCACGAACGCGCGAGACGA S A S Q I A G I G I T N Q R E T T	2001	ATGAAACACTGCACCTTGGCGCGGCTTATTTGGCGGGTATCGCTGTGGGA E T T A L G A A Y L A G I A V G	451
951	CGGTTGTGTGGGATAAGATACAGGAAGTCTGTCTATAATGCAATCGTT V V W D K D T G S P V Y N A I V	2051	TTCTGGAAGGACCGTTCTGAAATCGCGAACCAGTGAATCTGGATAACG F W K D R S E I A N Q W N L D K R	468
1001	TGGCAGTCCAGACAGAGCTCGCGCATTTGTGAGGAACCTCGTGAAAAAGG W Q S R Q T S G I C E E L R E K G	2101	GTTTGAGCCTGAATTGGAAGAAGAAAAACGAAATGAGCTGTATAAAGGCT F E P E L E E E K R N E L Y K G W	485
1051	ATATAATGATAAATTCAGAGAAAAACAGGCGCTTTTAAATCGATCCTTACT Y N D K F R E K T G L L I D P Y F	2151	<i>DraI</i> GGCAAAAAGCGTGAAAGCAGCTATGGCTTTTAAATAAAGTAATACTATG Q K A V K A A M A F K *	496

2201	GTATAATGGTTACAAGTTAATAAGAACGGTCTGAGATGAGGAGAGACCA				
	<i>SphI</i>				
2251	CAGCACCAAGTGTAAAGCATGCACTTTGGCTGTTGTGGTCTCTTTTCTA				
	-----> <-----				
2301	TTTACCGTGACAACAAGGAGGAAACGTAATGATGAATCATCAATTTTCAA				
	M M N H Q F S S	8			
	<i>SphI</i> <i>glpD</i>				
2351	GTCCTGAAAGAGATCGCATGCTGACAGACATGACAAAAAATCATATGAC				
	L E R D R M L T D M T K K T Y D	24			
2401	CTATTTATTATCGGAGGAGGAATTACAGGAGCCGGAACAGCTCTTGACGC				
	L F I I G G G I T G A G T A L D A	41			
2451	GGCATCAAGGGGAATGAAGGTCGCACTAAGCGAAATGCAGGACTTTGCGG				
	A S R G M K V A L S E M Q D F A A	58			
2501	CGGGAACATCAAGCCGATCAACGAAGCTGGTACATGGCGGCTTGCGCTAT				
	G T S S R S T K L V H G G L R Y	74			
	<i>DraI</i>				
2551	TTAAACAATTTGAAGTGAATAAGTGGTGGTGGGAGGAGCGGGC				
	L K Q F E V K M V A E V G K E R A	91			
2601	GATTGTGTATGAAACGGCCCGCATGTCACTACCCGAATGGATGCTGC				
	I V Y E N G P H V T T P E W M L L	108			
2651	TTCCGTTTCATAAAGGCGGCACATTCGGTTTCATTACAACATCTATTGGG				
	P F H K G G T F G S F T T S I G	124			
2701	TTAAGGGTTTATGACTTCCTTGAGGTGTGAAAAGTCAGAACGAAGAAG				
	L R V Y D F L A G V K K S E R R S	141			
	<i>SphI</i>				
2751	CATGCTTTCAGCAAAAGAACGCTGCAAAAAGAGCCTTTGGTGA AAAAG				
	M L S A K E T L Q K E P L V K K D	158			
2801	ACGGCTTAAAGGCGGCGGCTACTATGTGGAATACCGCACTGACGATGCG				
	G L K G G G Y Y V E Y R T D D A	174			
2851	AGACTGACCATCGAAGTCATGAAGGAAGCGGTTAAATTCGGGGCAGAGCC				
	R L T I E V M K E A V K F G A E P	191			
2901	TGTGAATTACTCCAAAGTGAAGGAGCTCTTTACGAAAAGGCAAGCGG				
	V N Y S K V K E L L Y E K G K A V	208			
2951	TCGGCGTATTAATTGAAGATGTGCTGACAAAGAAAGATATAAAGTGAT				
	G V L I E D V L T K K E Y K V Y	224			
3001	GCGAAAAAATTGTCAATGCTACAGGCCCTTGGGTCGATCAGCTCAGAGA				
	A K K I V N A T G P W V D Q L R E	241			
3051	AAAAGACCATTCGAAAACGGAAGCATTGTCAGCATACAAAAGGCATTTC				
	K D H S K N G K H L Q H T K G I H	258			
3101	ACCTTGATTTTGACCACTGCTCTTTCCGCTGAAACAGGCTGTATATTTTC				
	L V F D Q S V F P L K Q A V Y F	274			
3151	GATACACCTGATGGCCGGATGGTATTTGCGATTCTCGTGAAGGCAAAAC				
	D T P D G R M V F A I P R E G K T	291			
3201	ATACGTGGGAACAACAGACACTGTTTACAAAGAGGCGCTGGAGCATCCGC				
	Y V G T T D T V Y K E A L E H P R	308			
3251	GGATGACAACGGAGGATCGTGATTATGTATCAATCAATCAATTACATG				
	M T T E D R D Y V I K S I N Y M	324			
	<i>PvuII</i>				
3301	TTCCCGGAAGTGAATATCACTGCGAATGACATCGAATCCAGCTGGGCGGG				
	F P E L N I T A N D I E S S W A G	341			
3351	GCTGCGTCCGCTGATTCATGAAGAAGGCAAGATCCTTCTGAAATTTTAC				
	L R P L I H E E G K D P S E I S R	358			
3401	GGAAAGACGAGATTTGGACATCTGACTCAGGCTGATCACCATTGCCGGC				
	K D E I W T S D S G L I T I A G	374			
3451	GGAAACTGACTGGATACAGAAAAATGGCGGACGACATCGTTGATCTTGT				
	G K L T G Y R K M A D D I V D L V	391			
3501	CCGTGATCGCTTAAAGAAGAGGGCGAAAGGATTTTGACCATGTAAAG				
	R D R L K E E G E K D F G P C K T	408			
3551	CGAAAAACATGCCAATCTCTGGCGGGCACGtcggcgGTTGAAAAATCTT				
	K N M P I S G G H V G G S K N L	424			
3601	ATGTCCTTCGTTACCGCGAAAAACAAAAGGAATTCGACCGGGTTTATC				
	M S F V T A K T K E G I A A G L S	441			
3651	AGAAAAAGACGCAAAACAGCTTCGCGATCAGATACGGCTCTAACGTAGATC				
	E K D A K Q L A I R Y G S N V D R	458			
3701	GCGTCTTTGACCGGGTAGAAGCGCTGAAAGATGAGGCCGCAAAACGCAAC				
	V F D R V E A L K D E A A K R N	474			
3751	ATTCGGTTTCATATTCCTGCTGAGGCAGAATACAGTATAGAAGAAGAGAT				
	I P V H I L A E A E Y S I E E E M	491			
3801	GACTGCAACCCCTGCTGACTTCTTTGTCGCGAGAACGGGACGTTTATTT				
	T A T P A D F F V R R T G R L F F	508			
3851	TTGATATCAATTGGGTGAAGACATATAAAGATGCCGTTATTGATTTATG				
	D I N W V R T Y K D A V I D F M	524			
3901	AGCGAGCGATTCCAATGGGATGAGCAGGCGAAAAACAACATACAGAAAA				
	S E R F Q W D E Q A K N K H T E N	541			
	<i>HindIII</i>				
3951	CCTCAACAAGCTTTTACACGATGCGGTCGTACCGCTTGAGCAATAAATCA				
	L N K L L H D A V V P L E Q *	555			
	<i>PstI</i>				
4001	TAACGGGCTGTCTGCAGCCCGTTATTTCTTTTACGTGCCGAAAGGGGGA				
	-----> <-----				
4051	GATCTCAGGTTATCAAGTGACAGAGGCCACCGCTTCATATAAAATATGG				
	-35 -10				
4101	CATAAGCTGATACATAGGAGGACGAATATGACTTGGAGAAGAGCTATGA				
	M T W R K S Y E				
	ORF4				
4151	ACGCTGGAACAGACAGAACATTTAGATCTGGAATTAAGAGCGCCTTA				
	R W K Q T E H L D L E L K E R L I				
4201	TTGAATTAGAGGAGATGAACAGGCCCTTGAGGACTGTTTCTATAAAGAC				
	E L E G D E Q A L E D C F Y K D				
	<i>EcoRI</i>				
4251	CTTGAATTC				
	L E F				

Fig. 4. Nucleotide sequence of part of ORF1, *glpK* and *glpD* of *B. subtilis*. All restriction sites fully overlapped. Both strands were sequenced except for 132 bp in the 5' part. These 132 bp are written in lower-case letters. Potential RBSs and a putative promoter sequence are underlined and the putative transcription terminators are shown as dyad symmetries. Left-margin numbers refer to the nucleotide sequence, right-margin numbers to amino acid sequences.

Fig. 5. Comparison between predicted flavin-binding sequences of the *E. coli* G3P dehydrogenases and the amino-terminal region of GlpD of *B. subtilis*.

Cole *et al.*, 1988) revealed 32% and 25% identity, respectively. Both *E. coli* enzymes have FAD as a co-factor and a typical ADP-binding $\beta\alpha\beta$ -fold (Wierenga *et al.*, 1986) is located at the N-terminal parts of both proteins. As shown in Fig. 5, this sequence is highly conserved also in the protein of the ORF, indicating the presence of a similar cofactor. Transformation data place the *glpD6* allele within this ORF. These results together identify the third ORF as *glpD*. The entire *glpD* sequence is present in both pLUM20 and pLUM30. In pLUM20, however, we found a deletion of an A:T pair around position 3198. This deletion generates an in-frame TGA stop codon at position 3254. We believe that the mutation in pLUM20 was selected for in the cloning procedures and may contribute to the stability of pLUM20 in *E. coli*. The finding of the frame-shift mutation in pLUM20 *glpD* explains why the plasmid cannot complement a *glpD* mutation.

The fourth ORF starts at position 4128 with an ATG start codon, which is preceded by 9 bp by a potential RBS, -AGGAGG- ($\Delta G = -69$ kJ). Upstream of this ATG, a putative promoter sequence is found, GTGACA (-35) and TAAAT (-10). It corresponds well to the consensus sequence of the major σ factor of *B. subtilis*, σ^A , with differences of one nucleotide in the -35 region and one in the -10 region. Insertion of a 0.3 kb *HindIII*-*EcoRI* fragment (see Fig. 1) containing this promoter-like sequence in the proper orientation before a promoter-less *xylE* gene results in expression of this gene in a glycerol-independent manner (data not shown). Only the N-terminal part of this ORF is shown, but an additional 1000 bp have been cloned and partially sequenced. The identity or function of the protein encoded by this ORF is not known.

No obvious σ^A -type promoter has been found upstream of *glpK* or *glpD*. The TAA stop codon of ORF1 is located 18 nucleotides upstream of the proposed *glpK* ATG start codon, suggesting that ORF1 may be part of the same transcriptional unit as *glpK*. The proposed *glpK* TAA stop codon is separated from the proposed ATG start codon of *glpD* by 140 nucleotides. Within this region there is an inverted repeat which can form a stem-loop structure with a 22 bp stem ($\Delta G = -151$ kJ) and a loop of six nucleotides and which is followed by a run of Ts. This region has the characteristics of a rho-

independent transcription terminator (Rosenberg & Court, 1979). The role of this inverted repeat (if any) in the regulation of *glpD* expression is as yet unknown. An interesting possibility is that the inverted repeat serves to block transcription of *glpD* from a promoter located upstream of *glpK* in uninduced cells and that induction of *glpD* may involve anti-termination of transcription in the *glpK*-*glpD* interval. Anti-termination of transcription has recently been found important in the control of, for example, the *E. coli* *bgl* operon (Mahadevan & Wright, 1987) and the *B. subtilis* *sac* regulon (Crutz *et al.*, 1990). Following the *glpD* TAA stop codon by three nucleotides there is a second inverted repeat ($\Delta G = -95$ kJ) followed by a run of Ts, which may constitute a *glpD* transcription terminator. This suggestion is somewhat supported by the presence of a glycerol-independent promoter preceding ORF4.

Further work is now oriented towards characterizing the transcriptional organization of *glpK* and *glpD* and the structure and role of the GlpP protein in regulating expression of genes in glycerol catabolism in *B. subtilis*.

We thank Rune-Pär Nilsson for constructing the deletion mutant and Ingrid Stål for expert technical assistance. We thank Dr T. J. Larson for communicating unpublished sequence data. This project was supported by grants from the Swedish Medical Research Council, the Emil and Wera Cornell Foundation, the Crafoord Foundation and from Kungliga Fysiografiska Sällskapet.

References

- AMANN, E., BROSIUS, J. & PTASHNE, M. (1983). Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene* **25**, 167-178.
- ANAGNOSTOPOULOS, C. & SPIZIZEN, J. (1961). Requirements for transformation in *Bacillus subtilis*. *Journal of Bacteriology* **81**, 741-746.
- ARWERT, F. & VENEMA, G. (1973). Transformation in *Bacillus subtilis*. *Molecular and General Genetics* **123**, 185-198.
- CANOSI, U., MORELLI, G. & TRAUTNER, T. A. (1978). The relationship between molecular structure and transformation efficiency of some *S. aureus* plasmids isolated from *B. subtilis*. *Molecular and General Genetics* **166**, 259-267.
- COLE, S. T., EIGLMEIER, K., AHMED, S., HONORE, N., ELMES, L., ANDERSON, W. F. & WEINER, J. H. (1988). Nucleotide sequence and gene-polypeptide relationships of the *glpABC* operon encoding the anaerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. *Journal of Bacteriology* **170**, 2448-2456.

- CRUTZ, A.-M., STEINMETZ, M., AYMERICH, S., RICHTER, R. & LE COQ, D. (1990). Induction of levansucrase in *Bacillus subtilis*: an antitermination mechanism negatively controlled by the phosphotransferase system. *Journal of Bacteriology* **172**, 1043–1050.
- DEVEREUX, J., HAEBERLI, P. & SMITHES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.
- EHRMANN, M., BOOS, W., ORMSETH, E., SCHWEIZER, H. & LARSON, T. J. (1987). Divergent transcription of the *sn*-glycerol-3-phosphate active transport (*glpT*) and anaerobic *sn*-glycerol-3-phosphate dehydrogenase (*glpA glpC glpB*) genes of *Escherichia coli* K-12. *Journal of Bacteriology* **169**, 526–532.
- FORTNAGEL, P. & FREESE, E. (1968). Analysis of sporulation mutants. II. Mutants blocked in the citric acid cycle. *Journal of Bacteriology* **95**, 1431–1438.
- FREEDBERG, W. B. & LIN, E. C. C. (1973). Three kinds of controls affecting the expression of the *glp* regulon in *Escherichia coli*. *Journal of Bacteriology* **115**, 816–823.
- FRIDÉN, H., HEDERSTEDT, L. & RUTBERG, L. (1987). Deletion of the *Bacillus subtilis* *sdh* operon. *FEMS Microbiology Letters* **41**, 203–206.
- FRY, D. C., KUBY, S. A. & MILDVAN, A. S. (1986). ATP-binding site of adenylate kinase: mechanistic implications of its homology with *ras*-encoded p21, F_1F_0 -ATPase, and other nucleotide-binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 907–911.
- GRYCZAN, T. J., CONTENTE, S. & DUBNAU, D. (1978). Characterization of *Staphylococcus aureus* plasmids introduced by transformation into *Bacillus subtilis*. *Journal of Bacteriology* **134**, 318–329.
- HAGER, P. W. & RABINOWITZ, J. C. (1985). Translational specificity in *Bacillus subtilis*. In *The Molecular Biology of the Bacilli*, vol. II, pp. 1–32. Edited by D. A. Dubnau. New York: Academic Press.
- HAIMA, P., BRON S. & VENEMA, G. (1987). The effect of restriction on shotgun cloning in *Bacillus subtilis* Marburg. *Molecular and General Genetics* **209**, 335–342.
- HIGGINS, C. F., HILES, I. D., SALMOND, G. P. C., GILL, D. R., DOWNIE, J. A., EVANS, I. J., HOLLAND, I. B., GRAY, L., BUCKEL, S. D., BELL, A. W. & HERMODSON, M. A. (1986). A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature, London* **323**, 448–450.
- HOLMBERG, C. & RUTBERG, B. (1989). Cloning of the glycerol kinase gene of *Bacillus subtilis*. *FEMS Microbiology Letters* **58**, 151–156.
- ISH-HOROWICZ, D. & BURKE, J. F. (1981). Rapid and efficient cosmid cloning. *Nucleic Acids Research* **9**, 2989–2998.
- IUCHI, S., COLE, S. T. & LIN, E. C. C. (1990). Multiple regulatory elements of the *glpA* operon encoding anaerobic glycerol-3-phosphate dehydrogenase and the *glpD* operon encoding aerobic glycerol-3-phosphate dehydrogenase in *Escherichia coli*: further characterization of respiratory control. *Journal of Bacteriology* **172**, 179–184.
- LARSON, T. J., SCHUMACHER, G. & BOOS, W. (1982). Identification of the *glpT*-encoded *sn*-glycerol-3-phosphate permease of *Escherichia coli*, an oligomeric integral membrane protein. *Journal of Bacteriology* **152**, 1008–1021.
- LARSON, T. J., EHRMANN, M. & BOOS, W. (1983). Periplasmic glycerophosphodiester phosphodiesterase of *Escherichia coli*, a new enzyme of the *glp* regulon. *Journal of Biological Chemistry* **258**, 5428–5432.
- LIN, E. C. C. (1976). Glycerol dissimilation and its regulation in bacteria. *Annual Review of Microbiology* **36**, 535–578.
- LIN, E. C. C., KOCH, J. P., CHUSED, T. M. & JORGENSEN, S. E. (1962). Utilization of L- α -glycerophosphate by *Escherichia coli* without hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America* **48**, 2145–2150.
- LINDGREN, V. (1978). Mapping of a genetic locus that affects glycerol-3-phosphate transport in *Bacillus subtilis*. *Journal of Bacteriology* **133**, 667–670.
- LINDGREN, V. & RUTBERG, L. (1974). Glycerol metabolism in *Bacillus subtilis*: gene–enzyme relationships. *Journal of Bacteriology* **119**, 431–442.
- LINDREN, V. & RUTBERG, L. (1976). Genetic control of the *glp* system in *Bacillus subtilis*. *Journal of Bacteriology* **127**, 1047–1057.
- MAHADEVAN, S. & WRIGHT, A. (1987). A bacterial gene involved in transcription antitermination: regulation at a rho-independent terminator in the *bgl* operon of *E. coli*. *Cell* **50**, 485–494.
- MANDEL, M. & HIGA, A. (1970). Calcium-dependent bacteriophage DNA infection. *Journal of Molecular Biology* **53**, 159–162.
- MANIATIS, T., FRITSCH, E. F. & SAMBROOK, J. (1982). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MESSING, J., CREA, R. & SEEBURG, P. H. (1981). A system for shotgun DNA sequencing. *Nucleic Acids Research* **9**, 309–321.
- MINDICH, L. (1968). Pathway for oxidative dissimilation of glycerol in *Bacillus subtilis*. *Journal of Bacteriology* **96**, 565–566.
- NEVILLE, D. M., JR (1971). Molecular weight determination of protein–dodecyl sulfate complexes by gel electrophoresis in a discontinuous buffer system. *Journal of Biological Chemistry* **246**, 6328–6334.
- NORRANDER, J., KEMPE, T. & MESSING, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**, 101–106.
- PETTIGREW, D. W., MA, D.-P., CONRAD, C. A. & JOHNSON, J. R. (1988). *Escherichia coli* glycerol kinase. *Journal of Biological Chemistry* **263**, 135–139.
- ROSENBERG, M. & COURT, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Annual Review of Genetics* **13**, 319–353.
- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SCHWEIZER, H. & LARSON, T. J. (1987). Cloning and characterization of the aerobic *sn*-glycerol-3-phosphate dehydrogenase structural gene *glpD* of *Escherichia coli* K-12. *Journal of Bacteriology* **169**, 507–513.
- SCHWEIZER, H., BOOS, W. & LARSON, T. J. (1985). Repressor for the *sn*-glycerol-3-phosphate regulon of *Escherichia coli* K-12: cloning of the *glpR* gene and identification of its product. *Journal of Bacteriology* **161**, 563–566.
- SCHWEIZER, H., SWEET, G. & LARSON, T. J. (1986). Physical and genetic structure of the *glpD*–*malT* interval of the *Escherichia coli* K-12 chromosome. *Molecular and General Genetics* **202**, 488–492.
- SMITH, C. P. & CHATER, K. F. (1988). Structure and regulation of controlling sequences for the *Streptomyces coelicolor* glycerol operon. *Journal of Molecular Biology* **204**, 569–580.
- SWEET, G., GANDOR, C., VOEGELE, R., WITTEKINDT, N., BEUERLE, J., TRUNIGER, V., LIN, E. C. C. & BOOS, W. (1990). Glycerol facilitator of *Escherichia coli*: cloning of *glpF* and identification of the *glpF* product. *Journal of Bacteriology* **172**, 424–430.
- TAKESHITA, S., SATO, M., TOBA, M., MASAHASHI, W. & HASHIMOTO-GOTOH, T. (1987). High-copy number and low-copy number plasmid vectors for *lacZ*-complementation and chloramphenicol- or kanamycin-resistance selection. *Gene* **61**, 63–74.
- TINOCO, I., JR, BORER, P. N., DENGLER, B., LEVINE, M. D., UHLENBECK, O. C., CROTHERS, D. M. & GRALLA, J. (1973). Improved estimation of secondary structure in ribonucleic acids. *Nature New Biology* **246**, 40–41.
- WIERENGA, R. K., TERPSTRA, P. & HOL, W. G. J. (1986). Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$ -fold in proteins, using an amino acid sequence fingerprint. *Journal of Molecular Biology* **187**, 101–107.
- YANISCH-PERRON, C., VIEIRA, J. & MESSING, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119.