JOURNAL OF BACTERIOLOGY, Dec. 1990, p. 7227–7240
0021-9193/90/127227-14$02.00/0

Vol. 172, No. 12

Sequence and Genetic Organization of a *Zymomonas mobilis* Gene Cluster That Encodes Several Enzymes of Glucose Metabolism

WENDY O. BARNELL, KYUNG CHEOL YI, AND TYRRELL CONWAY*

School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588-0118

Received 10 April 1990/Accepted 7 September 1990

The *Zymomonas mobilis* genes that encode glucose-6-phosphate dehydrogenase (*zwf*), 6-phosphogluconate dehydratase (*edd*), and glucokinase (*glk*) were cloned independently by genetic complementation of specific defects in *Escherichia coli* metabolism. The identity of these cloned genes was confirmed by various biochemical means. Nucleotide sequence analysis established that these three genes are clustered on the genome and revealed an additional open reading frame in this region that has significant amino acid identity to the *E. coli* xylose-proton symporter and the human glucose transporter. On the basis of this evidence and structural analysis of the deduced primary amino acid sequence, this gene is believed to encode the *Z. mobilis* glucose-facilitated diffusion protein, *glf*. The four genes in the 6-kb cluster are organized in the order *glf*, *zwf*, *edd*, *glk*. The *glf* and *zwf* genes are separated by 146 bp. The *zwf* and *edd* genes overlap by 8 bp, and their expression may be translationally coupled. The *edd* and *glk* genes are separated by 203 bp. The *glk* gene is followed by tandem transcriptional terminators. The four genes appear to be organized in an operon. Such an arrangement of the genes that govern glucose uptake and the first three steps of the Entner-Doudoroff glycolytic pathway provides the organism with a mechanism for carefully regulating the levels of the enzymes that control carbon flux into the pathway.

Zymomonas mobilis is a gram-negative bacterium which, during evolution, has become highly specialized for growth in plant saps with a high sugar content (37, 51, 53). This obligately fermentative organism possesses remarkably simple carbon and energy metabolism. *Z. mobilis* is only able to utilize glucose, fructose, and sucrose, which are converted to the sole fermentation products ethanol and carbon dioxide. Yet, in terms of biosynthetic capabilities, this organism has only two growth factor requirements. *Z. mobilis* uses the Entner-Doudoroff pathway exclusively for conversion of carbohydrates to pyruvate and the decarboxyclastic mechanism for ethanol production, with the key enzyme pyruvate decarboxylase. The enzymes responsible for this fermentation compose as much as 50% of the total soluble protein. *Z. mobilis* is totally dependent on substrate-level phosphorylation for energy production and, due to its use of the Entner-Doudoroff pathway, obtains only a single mole of ATP per mole of glucose fermented. For this organism, rapid carbon flux is necessitated by inefficient energy production and is facilitated by high levels of the pathway enzymes.

Despite enormous carbon flux, *Z. mobilis* must keep the levels of toxic metabolic intermediates low while providing sufficient pools of precursor metabolites for biosynthetic pathways. The physiology and biochemistry of this organism are dictated by these constraints on metabolism. The glycolytic enzymes appear to operate near their maximal capacity, and there is no substantial allosteric control of physiologically irreversible enzymes (53). It can be concluded that carbon flux is limited solely by the maximal activity of the pathway enzymes. Biochemical evidence indicates that glucokinase activity is the most important rate-limiting step. Intracellular accumulation of glucose-6-phosphate also implicates glucose-6-phosphate dehydrogenase as catalyzing a flux-generating step (8).

Although much is known regarding the biochemistry of glycolytic enzymes, both in general and for *Z. mobilis* specifically, there are few systematic studies of the genetic mechanisms of control of glycolytic enzyme expression. Also, there is little knowledge of the mechanisms governing the variations in the levels of glycolytic enzymes. For the most part, central pathways for carbohydrate catabolism appear to be constitutive. However, Wolf and co-workers (56) have stressed that even an invariable level of enzyme must be the result of an active regulatory process. Furthermore, several of the glycolytic enzymes are present in various organisms and tissues in relatively constant proportions (41). Again, some form of genetic regulation must exist that stabilizes the levels of glycolytic enzymes with respect to one another. It is our goal to examine the genetic component of glycolytic flux control in *Z. mobilis*. To this end, we have been cloning the genes that encode the pathway enzymes for use as tools to study the contribution of gene expression to flux control at each step of the pathway.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The strains and plasmids used in this study are described in Table 1. *Z. mobilis* CP4 was grown at 30°C in complex medium containing 100 g of glucose per liter as described previously (39). *Escherichia coli* strains were routinely grown in Luria broth without added carbohydrate (36) or in M63 minimal medium with appropriate supplements (52) at 37°C. *E. coli* RW231R is an *eda*$^+$ revertant of strain RW231 that was selected by restored ability to grow on minimal glucuronic acid medium. Carbohydrates were filter sterilized and added to basal medium (0.2% final concentration). Media were solidified by the addition of agar (1.5%). Antibiotic-resistant transformants were selected by the addition of ampicillin (50 mg/liter). Gluconate bromthymol blue indicator plates were prepared as described previously (7). 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) (20 mg/liter) was used

* Corresponding author.

TABLE 1. Plasmids and strains used

Strain or plasmid	Relevant genotype	Source or reference
Zymomonas mobilis CP4	Prototroph	39
Escherichia coli		
DH5α	*lacZM15 recA*	Bethesda Research Laboratories
CH5α F′	*lacZM15 rec*A F′	Bethesda Research Laboratories
DF214	Δ(*zwf edd*) *pgi*::Mu	54
RW231	Δ(*zwf edd*) *eda gnd*	55
RW231R	Δ(*zwf edd*) *gnd*	This study
ZSC113	*ptsM12 ptsG22 glk*	18
W3110	Wild type	6
Plasmids		
pUC18	*bla lacI′Z′*[a]	58
pBluescript II	*bla lacI′Z′* f1 origin	Stratagene
pTC111	*zwf glf*	This study
pTC112	*zwf glf*	This study
pTC120	*edd glk*	This study
pTC150	*edd glk*	This study

[a] Incomplete *lacI* and *lacZ* genes.

to identify recombinant plasmids with DNA insertions that inactivated β-galactosidase activity in *E. coli* DH5α.

DNA methods. Transformation, restriction mapping, and subcloning were done by standard methods (45, 49). Small-scale plasmid isolations were prepared by a modification of the method of Birnboim and Doly (11) as described previously (49). Restriction enzymes and DNA-modifying enzymes were used according to the recommendations of the manufacturers. Genomic DNA from *Z. mobilis* CP4 was isolated as described previously (13). A library of 4- to 6-kb fragments of *Z. mobilis* genomic DNA, generated by partial digestion with *Sau*3A, was made by ligation into the *Bam*HI site of pUC18 as described previously (16). This ligation was amplified by transformation of *E. coli* DH5α and produced a library consisting of 10,000 clones (75% insertion frequency).

Cloning of *Z. mobilis* genes for glucose-6-phosphate dehydrogenase (*zwf*), 6-phosphogluconate dehydratase (*edd*), and glucokinase (*glk*). Transformation of the *Z. mobilis*(pUC18) library into *E. coli* DF214 [*pgi*::Mu Δ(*zwf-eda*) (54)] with selection for restored ability to grow on glucose minimal medium allowed the rescue of the *Z. mobilis* gene for glucose-6-phosphate dehydrogenase (pTC111, *zwf*$^+$). The nature of this clone was confirmed by the ability to transform *E. coli* DF214 to the glucose-positive phenotype at high frequency and by assay of enzyme activity.

Transformation of the *Z. mobilis*(pUC18) library into *E. coli* RW231 (*gnd zwf edd* [55]) with selection for restored ability to grow on gluconate allowed the rescue of the *Z. mobilis edd* gene that encodes 6-phosphogluconate dehydratase (pTC120). This clone was able to transform *E. coli* RW231 to a gluconate-positive phenotype at high frequency and was yellow on gluconate bromthymol blue indicator plates (*edd*$^+$ clones turn yellow [7]). Confirmation of the *edd* clone was obtained by assay of enzyme activity.

The *Z. mobilis* gene that encodes glucokinase (*glk*, pTC150) was cloned by complementation of specific defects in glucose phosphorylation in an *E. coli* mutant as described by Fukuda et al. (28). *E. coli* ZSC113 is defective in glucokinase, glucose phosphotransferase, and mannose phosphotransferase and forms white colonies on glucose-MacConkey agar plates (18). The *Z. mobilis* library in pUC18 was screened for *glk* by conversion of *E. coli* ZSC113 to formation of pink colonies on glucose-MacConkey agar. The nature of the cloned glucokinase from *Z. mobilis* was confirmed by enzyme assay.

Enzyme assays. *E. coli* cells were prepared for enzyme assays by washing in 50 mM potassium phosphate buffer (pH 6.8) and resuspension in the same buffer to an A_{550} of 1.0. Cell suspension (2 ml) was pelleted, suspended in phosphate buffer, and sonicated for 15 s. Glucose-6-phosphate dehydrogenase was assayed as described previously (31). Phosphoglucose isomerase was assayed as described previously (30). 6-Phosphogluconate dehydrogenase was assayed by the method of Wolf et al. (56). 6-Phosphogluconate dehydratase was assayed as described by Lessie and Neidhardt (35), except that sonicated extracts were spun in a microcentrifuge at 14,000 × *g* for 5 min to remove nonspecific NADH oxidase activity. Glucokinase activity was measured as described previously (22). Enzyme activities are expressed in international units (micromoles per minute per milligram of total cell protein). Protein concentrations were assayed by the method of Layne (34).

SDS-polyacrylamide gel electrophoresis. Sodium dodecyl sulfate (SDS)-gel electrophoresis and Coomassie blue staining of gels were described previously (15).

Native polyacrylamide gel electrophoresis and activity staining of glucose-6-phosphate dehydrogenase and glucokinase. Sonicated cell extracts were electrophoresed on 6% native gels according to the instructions of the manufacturer of the gel electrophoresis apparatus (Bio-Rad Laboratories, Richmond, Calif.). Native gels were stained for glucokinase and glucose-6-phosphate dehydrogenase activity as described previously (29).

Southern probing. Agarose gels (0.8%) to be used for Southern blotting were electrophoresed and stained with ethidium bromide by standard methods (49). Capillary transfer of the DNA to Genescreen Plus hybridization transfer membranes and hybridization at 42°C in the presence of 50% formamide were done according to the protocols of the manufacturer (catalog no. NEF-976; Dupont, NEN Research Products, Boston, Mass.). DNA hybridization probes were labeled by random primed DNA labeling, using a kit.

DNA sequence analysis. The *Z. mobilis* gene cluster was sequenced by the dideoxy method (46), by using a Sequenase kit or by using the automated sequencing method developed by Brumbaugh et al. (12). The *Z. mobilis* DNA insert in pTC111 was subcloned into the phagemid vectors pBluescript II SK+ and SK− to facilitate sequence analysis of both strands. These constructions were made by digesting pTC111 with *Xba*I (in the polylinker region of pUC18) and *Ssp*I (upstream of *glf*), isolating the fragment from an agarose gel, and ligating into pBluescript II SK+ and SK− digested with *Xba*I and *Sma*I. The construction in pBluescript II SK+ was designated pTC112. In addition, a 3-kb *Nru*I-to-*Xba*I fragment from pTC150 was subcloned into pBluescript II SK+ and SK− (digested with *Sma*I and *Xba*I). A series of deletion subclones were generated from each of these four subclones with exonuclease III by using an Erase-a-Base kit. Deletion subclones that averaged 250 bp were subjected to sequence analysis as described above. A total of 52 different deletion subclones were compiled. Sequence data were analyzed by using the University of Wisconsin Genetic Computer Group Sequence Analysis Software Package version 6.1 (20). The subclones from pTC111 and pTC150 overlapped by 556 bp within the *edd* gene and clearly indicated that pTC111 and pTC150 contain contiguous sequences. The sequence data presented in this

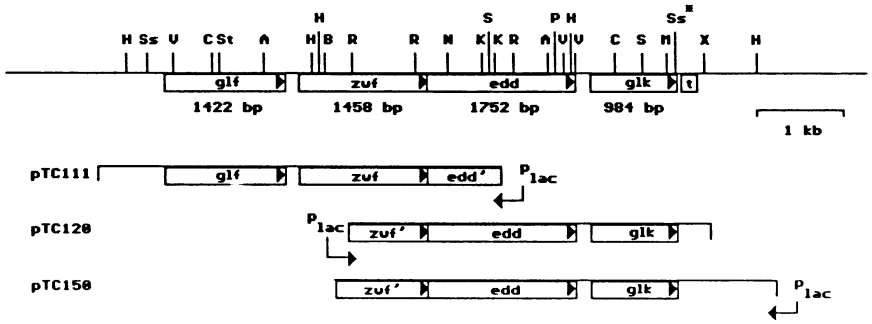

FIG. 1. Restriction map of the *Z. mobilis* glucose gene cluster. Restriction sites and location of open reading frames are drawn to scale on the top line. Abbreviations of restriction sites are as follows: H, *Hin*dIII; Ss, *Ssp*I; V, *Eco*RV; C, *Cla*I; St, *Stu*I; A, *Apa*I; B, *Bam*HI; R, *Eco*RI; N, *Nru*I; K, *Kpn*I; S, *Sma*I; P, *Pst*I; M, *Mlu*I; X, *Xba*I; Ss* denotes six clustered *Ssp*I sites. The genomic DNA inserts in original library clones of this region are shown below the restriction map. Direction of transcription from the *lac* promoter is indicated by arrows. A 1-kb size marker is shown.

report span 6,440 bp of DNA starting at the *Ssp*I site upstream of *glf* and ending with the *Xba*I site downstream of *glk*.

Enzymes and chemicals. Restriction enzymes and DNA-modifying enzymes were obtained from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.). The Sequenase and random primed DNA labeling kits were ordered from U.S. Biochemical Corp. (Cleveland, Ohio). The Erase-a-Base kit was purchased from Promega Biotec (Madison, Wis.). pBluescript II vectors were obtained from Stratagene (La Jolla, Calif.). Radioactive compounds were purchased from Dupont, NEN Research Products. Biochemicals were obtained from Sigma Chemical Co. (St. Louis, Mo.).

Nucleotide sequence accession number. The entire nucleotide sequence shown in Fig. 5 is available on the GenBank data base, accession no. M37982.

RESULTS

Cloning and characterization of gene encoding glucose-6-phosphate dehydrogenase. The general cloning strategy that was used to obtain the *Z. mobilis* genes described in this report was based on genetic complementation of *E. coli* mutants with specific defects in central metabolism. *E. coli* DF214 is unable to grow on glucose minimal medium due to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase mutations. The *Z. mobilis* library was transformed into *E. coli* DF214 and plated on minimal glucose plates containing ampicillin. Clones that were able to grow on glucose were obtained with a frequency of approximately 1 in 500. Over 150 glucose-positive colonies appeared, 35 of which were tested further. Plasmid DNAs prepared from each of these 35 clones were capable of transforming *E. coli* DF214 to a glucose-positive phenotype at high frequency. Enzyme assays of the clones were used to distinguish those that expressed glucose-6-phosphate dehydrogenase (*zwf*) from those that expressed phosphoglucose isomerase (*pgi*). Of the 35 original clones, 1 was found to be zwf^+ (pTC111) and the remainder were pgi^+. A study of the *Z. mobilis pgi* gene will be published elsewhere.

A restriction map of pTC111 was constructed (Fig. 1). The glucose-6-phosphate dehydrogenase activity expressed by *E. coli* DF214(pTC111) was active with either NAD (specific activity of 0.60) or NADP (specific activity of 0.34), as has been reported for *Z. mobilis* (48). The direction of transcription of the *zwf* gene in pTC111 was in the opposite orientation with respect to the pUC18-derived *lac* promoter. The subclone pTC112 contained the *zwf* gene in the same orientation as the *lac* promoter, and *E. coli* DF214(pTC112) expressed glucose-6-phosphate dehydrogenase at 20-fold-higher levels than *E. coli* DF214(pTC111). An extract of *E. coli* DF214(pTC111) was subjected to SDS-polyacrylamide gel electrophoresis (Fig. 2). A prominent band with an approximate size of 48,000 Da was observed. This is somewhat smaller than the size of glucose-6-phosphate dehydrogenase from *Z. mobilis* as determined by gel filtration (48) and is also somewhat smaller than the size deduced from the DNA sequence (reported below). An additional reading frame that lies immediately upstream of *zwf* on pTC111 was identified by sequence analysis. This gene is thought to encode the glucose-facilitated diffusion protein (*glf*, see below), which has a deduced size of approximately 50,000 Da. It is possible that both these proteins occupy a single band on the denaturing gel. A native polyacrylamide gel stained for glucose-6-phosphate dehydrogenase activity (zymogram; Fig. 3) clearly shows a prominent band with the same mobility as that from an extract of *Z. mobilis*. The origin of the *zwf* clone from *Z. mobilis* was confirmed by Southern gel analysis of genomic DNA digested with several restriction enzymes (Fig. 4). The specific gene probe used for hybridization was a 321-bp *Bam*HI-to-*Eco*RI fragment. This *Z. mobilis* probe did not hybridize to genomic DNA prepared from *E. coli* W3110 (digested with *Eco*RI) but did hybridize to restriction fragments of *Z. mobilis* DNA that matched the restriction map shown in Fig. 1. Hybridization to unique restriction fragments in the total genomic digests indicated that the *zwf* gene is present in a single copy. This in fact has been shown to be true for all the glycolytic genes from *Z. mobilis* that have been studied so far (23; unpublished data, this laboratory).

Cloning and characterization of gene encoding 6-phosphogluconate dehydratase. The *Z. mobilis* gene that encodes 6-phosphogluconate dehydratase (Entner-Doudoroff dehydratase, *edd*) was cloned by genetic complementation of a mutant that is defective in gluconate metabolism, *E. coli* RW231. Twelve original library clones with restored ability to grow on gluconate minimal medium were obtained (with a frequency of 1 in 2,000). Theoretically, clones of either phosphogluconate dehydrogenase (*gnd*) or *edd* could complement the defect in *E. coli* RW231. None of the 12 gluconate-positive clones expressed 6-phosphogluconate dehydrogenase activity. This is in keeping with the finding that *Z. mobilis* CP4 does not possess phosphogluconate dehydro-

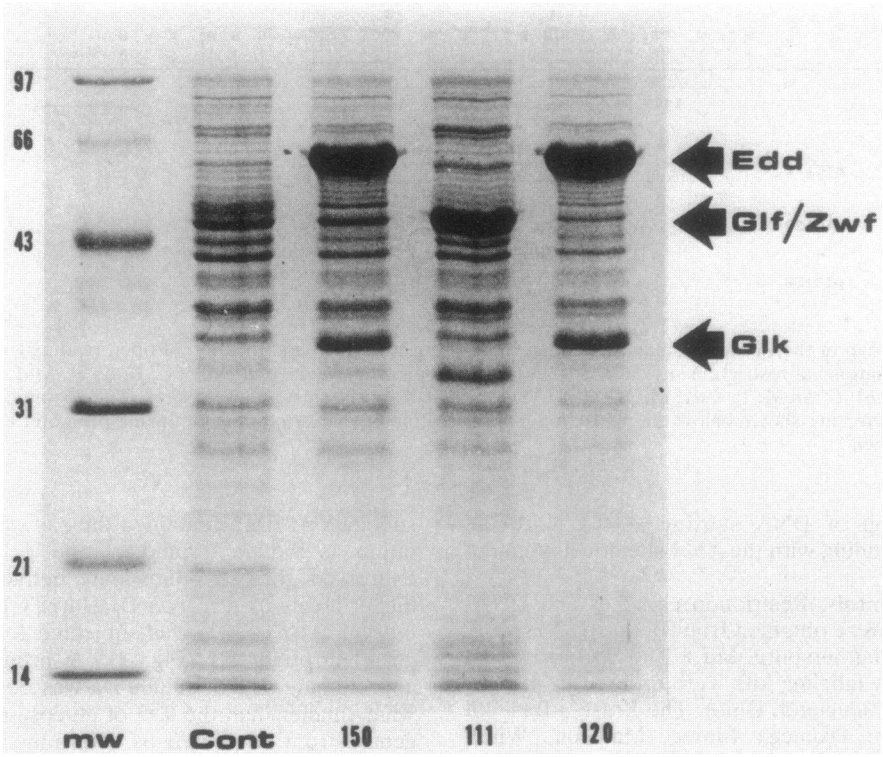

FIG. 2. SDS-polyacrylamide gel of cell extracts from strains expressing the cloned *Z. mobilis* genes. Molecular weights (in thousands are provided) on the left. Lanes are as follows: mw, molecular weight markers stained with Coomassie blue; Cont, *E. coli* DH5α(pUC18); 150, *E. coli* DH5α(pTC150); 111, *E. coli* DH5α(pTC111); 120, *E. coli* DH5α(pTC120). Overexpressed proteins presumed to be Edd, Glf, Zwf, and Glk are denoted with arrows on the right. Each lane was loaded with 10 μg of total cell protein.

genase activity when assayed with NADP. We believe that the low level of phosphogluconate dehydrogenase-like activity that was detected in *Z. mobilis* when assayed with NAD was most likely the result of an artifact of the enzyme assay (unpublished data). Phibbs (42) has shown that apparent 6-phosphogluconate oxidation by *Pseudomonas aeruginosa* is not catalyzed by 6-phosphogluconate dehydrogenase but rather is the result of conversion of 6-phosphogluconate to glyceraldehyde-3-phosphate and pyruvate by the Entner-Doudoroff enzymes with subsequent oxidation of glyceraldehyde-3-phosphate by NAD-specific glyceraldehyde-3-phosphate dehydrogenase.

Four of the gluconate-positive clones produced yellow colonies on gluconate bromthymol blue indicator plates, indicating high-level expression of *edd*. Transformation of *E. coli* RW231 with plasmid DNA prepared from these clones resulted in a pleiotropic gluconate-positive phenotype on gluconate bromthymol blue indicator plates. *E. coli* RW231 is defective in both steps of the Entner-Doudoroff pathway (*edd* and *eda* genes). During this investigation, it was found that all four of the gluconate-positive clones contained the *Z. mobilis edd* gene only and had acquired a mutation that restored the native *E. coli eda* gene to allow growth on gluconate. We were able to easily select for a revertant of *E. coli* RW231 that was *edd eda*$^+$ and did not show the pleiotropic gluconate-positive phenotype when transformed with the *Z. mobilis edd* gene (strain RW231R; see Materials and Methods). Plasmid DNA from one of the gluconate-positive clones, pTC120, was selected for further study. *E. coli* RW231R(pTC120) expressed fairly high levels of 6-phosphogluconate dehydratase (specific activity of 0.12) and overproduced a protein of 63,000 Da that was easily visual-

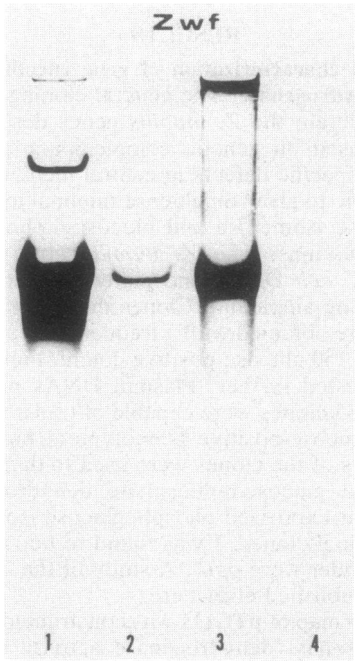

FIG. 3. Zymogram of cell extracts electrophoresed on a native polyacrylamide gel stained for glucose-6-phosphate dehydrogenase activity. Lanes: 1, *E. coli* DF214(pTC111); 2, *E. coli* DH5α; 3, *Z. mobilis* CP4; 4, *E. coli* DF214. Each lane was loaded with 5 μg of total cell protein.

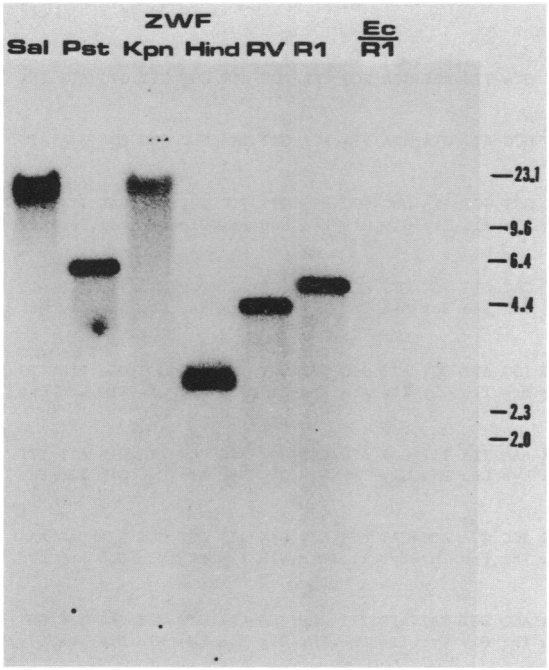

FIG. 4. Southern gel analysis of total genomic DNA with a *zwf*-specific hybridization probe. Size markers are in kilobase pairs on right. Lanes from left are *Z. mobilis* DNA digested with *Sal*I, *Pst*I, *Kpn*I, *Hin*dIII, *Eco*RV, and *Eco*RI. Last lane on right is *E. coli* DNA digested with *Eco*RI.

ized on an SDS-polyacrylamide gel (Fig. 2). In addition, this clone produced a protein of approximately 35,000 Da and assayed positive for glucokinase activity. Comparison of the restriction map of pTC120 with that of pTC150 showed that the two clones overlapped the same region of the *Z. mobilis* genome and confirmed that *glk* and *edd* are linked (Fig. 1). Confirmation of the origin of the *edd* gene from *Z. mobilis* was made by Southern gel analysis as described above for the *zwf* gene. The gene-specific hybridization probe was a 449-bp *Eco*RI-to-*Pst*I fragment. This probe hybridized only to unique restriction fragments of the *Z. mobilis* DNA and not *E. coli* (data not shown).

Cloning and characterization of gene encoding glucokinase. The *Z. mobilis* glucokinase gene (*glk*) was isolated after transformation of the gene library into *E. coli* ZSC113 with selection for pink colony formation on glucose-MacConkey agar. Strain ZSC113 is a triple mutant deficient in group translocation of glucose and mannose, as well as glucokinase. Over 120 faintly pink colonies obtained by this approach were assayed for glucokinase activity. Only one of these was glk^+ and was designated pTC150. This clone was found on a plate that had been incubated at 37°C overnight and left on the discard tray for 3 days. *E. coli* ZSC113(pTC150) expressed fairly high levels of glucokinase (specific activity of 0.77), despite being cloned in the opposite orientation as the *lac* promoter (Fig. 1). The cloned *Z. mobilis* glucokinase was able to use glucose, but not fructose or gluconate, and ATP, GTP, or CTP as substrates. This is in keeping with the reported specificities of the enzyme (22). Extracts from *E. coli* ZSC113(pTC150) were electrophoresed on an SDS-polyacrylamide gel to reveal two prominent bands with molecular weights of 35,000 and 63,000. The former is similar to the size of glucokinase as determined by gel filtration (48). The latter corresponds to the *edd* gene product, as described above. The origin of the *glk* gene from *Z. mobilis* was confirmed in two ways. A gene-specific 287-bp *Sma*I-to-*Mlu*I fragment was used for Southern gel analysis of total genomic DNA digests from *Z. mobilis* and *E. coli* and hybridized only to *Z. mobilis* DNA (data not shown). A native polyacrylamide gel stained for glucokinase activity demonstrated that *E. coli* DH5α(pTC150) produced two different glucokinase enzymes, one with the mobility of the native *E. coli* enzyme and the other with the mobility of the *Z. mobilis* enzyme (data not shown). In addition, a band with intermediate mobility was observed, possibly the result of heterodimer formation.

Nucleotide sequence of the *glf zwf edd glk* gene cluster. (i) *zwf* gene. The *Z. mobilis zwf* gene that encodes glucose-6-phosphate dehydrogenase consists of an open reading frame of 1,458 bp which corresponds to a protein of 485 amino acids with a molecular weight of 53,858 (Fig. 5). Comparison of the primary amino acid sequence of *Z. mobilis* glucose-6-phosphate dehydrogenase, deduced from the *zwf* coding region, with the human X-linked glucose-6-phosphate dehydrogenase (40) indicated 36% identity. In addition, a portion of the *Z. mobilis* enzyme is 43% identical to a 42-residue active site peptide prepared from purified glucose-6-phosphate dehydrogenase that was isolated from *Leuconostoc mesenteroides* (10). Further proof that the *Z. mobilis zwf* gene encodes glucose-6-phosphate dehydrogenase came from enzyme assays on *zwf* deletion subclones. Deletions that extended into the gene from either end eliminated activity (data not shown). The *zwf* reading frame is preceded by a strong Shine-Dalgarno sequence, separated by 8 bp from the AUG initiation codon (50). The *zwf* reading frame is preceded by a long open reading frame that stops 146 bp upstream of *zwf* (described below). Immediately upstream of the *zwf* Shine-Dalgarno sequence there is an inverted repeat followed by a string of four Ts that resembles a simple rho-independent transcriptional terminator (44, 57). The fact that glucose-6-phosphate dehydrogenase activity is expressed in *E. coli* containing pTC111 or pTC112 suggests that this is not an efficient transcriptional terminator. The *zwf* gene stops with the codon UGA. This stop codon is located within the downstream *edd* reading frame that begins 8 bp upstream of the *zwf* stop codon and is not within the *zwf* register.

(ii) *edd* gene. The *Z. mobilis edd* gene is 1,752 bp long and encodes 6-phosphogluconate dehydratase (Entner-Doudoroff dehydratase). The *edd* coding region corresponds to a protein of 583 amino acids with a molecular weight of 62,924 (Fig. 5). A search of the GenBank database using the deduced amino acid sequence for *edd* as the query sequence revealed a match to the *E. coli* dihydroxyacid dehydratase that is encoded by the *ilvD* gene (17). A comparison of the putative *Z. mobilis* 6-phosphogluconate dehydratase and *E. coli* dihydroxyacid dehydratase as aligned by the UWGCG Gap program is shown in Fig. 6. These proteins are 31% identical and 56% similar. Overall similarity between the first 100 amino acids of the *Z. mobilis* protein and the first 60 amino acids of the *E. coli* protein is poor, and several gaps were introduced by the computer program to obtain maximal alignment. The remaining sequences of the two proteins were found to be more similar, but the introduction of a substantial 19-amino-acid gap in the *Z. mobilis* protein at amino acid residue 400 was required for optimal alignment. The *E. coli* protein extends 34 amino acids beyond the carboxy terminus of the *Z. mobilis* protein, and in total, the *E. coli* protein is 33 amino acids longer than the *Z. mobilis* dehydratase.

```
1 T TGT TCC AAT AAT GAT AAG GAG GTT CCA AAA TCG GTT CAA TAG GAG TAA TCC GTA TTC AGT TTA TCG CCA AGG ATT CGG TTT GTG ATG TTA  91

92  TTT TTG GTC AGA AAC TAA AAT AAG ACC AAT GTT TAA CAT TGC CGA TAC TCG GCG ATT GTA AGA TTT ACA GAT TAA GGC GGG AGA GGA ATC  181

182  GCC ATG AGT TCT GAA AGT AGT CAG GGT CTA GTC ACG CGA CTA GCC CTA ATC GCT GCT ATA GGC GGC TTG CTT TTC GGT TAC GAT TCA GCG  271
    GLF  Met Ser Ser Glu Ser Ser Gln Gly Leu Val Thr Arg Leu Ala Leu Ile Ala Ala Ile Gly Gly Leu Leu Phe Gly Tyr Asp Ser Ala

272  GTT ATC GCT GCA ATC GGT ACA CCG GTT GAT ATC CAT TTT ATT GCC CCT CGT CAC CTG TCT GCT ACG GCT GCG GCT TCC CTT TCT GGG ATG  361
     Val Ile Ala Ala Ile Gly Thr Pro Val Asp Ile His Phe Ile Ala Pro Arg His Leu Ser Ala Thr Ala Ala Ala Ser Leu Ser Gly Met

362  GTC GTT GTT GCT GTT TTG GTC GGT TGT GTT ACC GGT TCT TTG CTG TCT GGC TGG ATT GGT ATT CGC TTC GGT CGT CGC GGC GGA TTG TTG  451
     Val Val Val Ala Val Leu Val Gly Cys Val Thr Gly Ser Leu Leu Ser Gly Trp Ile Gly Ile Arg Phe Gly Arg Arg Gly Gly Leu Leu

452  ATG AGT TCC ATT TGT TTC GTC GCC GCC GGT TTT GGT GCT GCG TTA ACC GAA AAA TTA TTT GGA ACC GGT GGT TCG GCT TTA CAA ATT TTT  541
     Met Ser Ser Ile Cys Phe Val Ala Ala Gly Phe Gly Ala Ala Leu Thr Glu Lys Leu Phe Gly Thr Gly Gly Ser Ala Leu Gln Ile Phe

542  TGC TTT TTC CGG TTT CTT GCC GGT TTA GGT ATC GGT GTC GTT TCA ACC TTG ACC CCA ACC TAT ATT GCT GAA ATT CGT CCG CCA GAC AAA  631
     Cys Phe Phe Arg Phe Leu Ala Gly Leu Gly Ile Gly Val Val Ser Thr Leu Thr Pro Thr Tyr Ile Ala Glu Ile Arg Pro Pro Asp Lys

632  CGT GGT CAG ATG GTT TCT GGT CAG CAG ATG GCC ATT GTG ACG GGT GCT TTA ACC GGT TAT ATC TTT ACC TGG TTA CTG GCT CAT TTC GGT  721
     Arg Gly Gln Met Val Ser Gly Gln Gln Met Ala Ile Val Thr Gly Ala Leu Thr Gly Tyr Ile Phe Thr Trp Leu Leu Ala His Phe Gly

722  TCT ATC GAT TGG GTT AAT GCC AGT GGT TGG TGC TGG TCT CCG GCT TCA GAA GGC CTG ATC GGT ATT GCC TTC TTA TTG CTG CTG TTA ACC  811
     Ser Ile Asp Trp Val Asn Ala Ser Gly Trp Cys Trp Ser Pro Ala Ser Glu Gly Leu Ile Gly Ile Ala Phe Leu Leu Leu Leu Leu Thr

812  GCA CCG GAT ACG CCG CAT TGG TTG GTG ATG AAG GGA CGT CAT TCC GAG GCT AGC AAA ATC CTT GCT CGT CTG GAA CCG CAA GCC GAT CCT  901
     Ala Pro Asp Thr Pro His Trp Leu Val Met Lys Gly Arg His Ser Glu Ala Ser Lys Ile Leu Ala Arg Leu Glu Pro Gln Ala Asp Pro

902  AAT CTG ACG ATT CAA AAG ATT AAA GCT GGC TTT GAT AAA GCC ATG GAC AAA AGC AGC GCA GGT TTG TTT GCT TTT GGT ATC ACC GTT GTT  991
     Asn Leu Thr Ile Gln Lys Ile Lys Ala Gly Phe Asp Lys Ala Met Asp Lys Ser Ser Ala Gly Leu Phe Ala Phe Gly Ile Thr Val Val

992  TTT GCC GGT GTA TCC GTT GCT GCC TTC CAG CAG TTA GTC GGT ATT AAC GCC GTG CTG TAT TAT GCA CCG CAG ATG TTC CAG AAT TTA GGT  1081
     Phe Ala Gly Val Ser Val Ala Ala Phe Gln Gln Leu Val Gly Ile Asn Ala Val Leu Tyr Tyr Ala Pro Gln Met Phe Gln Asn Leu Gly

1082  TTT GGA GCT GAT ACG GCA TTA TTG CAG ACC ATC TCT ATC GGT GTT GTG AAC TTC ATC TTC ACC ATG ATT GCT TCC CGT GTT GTT GAC CGC  1171
      Phe Gly Ala Asp Thr Ala Leu Leu Gln Thr Ile Ser Ile Gly Val Val Asn Phe Ile Phe Thr Met Ile Ala Ser Arg Val Val Asp Arg

1172  TTC GGC CGT AAA CCT CTG CTT ATT TGG GGT GCT CTC GGT ATG GCT GCA ATG ATG GCT GTT TTA GGC TGC TGT TTC TGG TTC AAA GTC GGT  1261
      Phe Gly Arg Lys Pro Leu Leu Ile Trp Gly Ala Leu Gly Met Ala Ala Met Met Ala Val Leu Gly Cys Cys Phe Trp Phe Lys Val Gly

1262  GGT GTT TTG CCT TTG GCT TCT GTG CTT CTT TAT ATT GCA GTC TTT GGT ATG TCA TGG GGC CCT GTC TGC TGG GTT GTT CTG TCA GAA ATG  1351
      Gly Val Leu Pro Leu Ala Ser Val Leu Leu Tyr Ile Ala Val Phe Gly Met Ser Trp Gly Pro Val Cys Trp Val Val Leu Ser Glu Met

1352  TTC CCG AGT TCC ATC AAG GGC GCA GCT ATG CCT ATC GCT GTT ACC GGA CAA TGG TTA GCT AAT ATC TTG GTT AAC TTC CTG TTT AAG GTT  1441
      Phe Pro Ser Ser Ile Lys Gly Ala Ala Met Pro Ile Ala Val Thr Gly Gln Trp Leu Ala Asn Ile Leu Val Asn Phe Leu Phe Lys Val

1442  GCC GAT GGT TCT CCA GCA TTG AAT CAG ACT TTC AAC CAC GGT TTC TCC TAT CTC GTT TTC GCA GCA TTA AGT ATC TTA GGT GGC TTG ATT  1531
      Ala Asp Gly Ser Pro Ala Leu Asn Gln Thr Phe Asn His Gly Phe Ser Tyr Leu Val Phe Ala Ala Leu Ser Ile Leu Gly Gly Leu Ile

1532  GTT GCT CGC TTC GTG CCG GAA ACC AAA GGT CGG AGC CTG GAT GAA ATC GAG GAG ATG TGG CGC TCC CAG AAG TAG TTA AAC TTG CTT TGG  1621
      Val Ala Arg Phe Val Pro Glu Thr Lys Gly Arg Ser Leu Asp Glu Ile Glu Glu Met Trp Arg Ser Gln Lys End

1622  CTG AAT CCT TTT GTC TTT TTT AGA TAA GTC TTA ACC AAT TAT ACT TTT TGT TTA CAA CGA TGG TAT AAA GCG GGC GGA CAG GCT AAA AAC  1711
```

FIG. 5. Nucleotide sequence of the *Z. mobilis* glucose gene cluster, including *glf*, *zwf*, *edd*, and *glk*. Sequence begins at the *Ssp*I site, and numbering begins with first base sequenced. The beginning of each reading frame is labeled with the gene designation in boldface capital letters. The Shine-Dalgarno sequences are underlined. The point at which the *zwf* and *edd* genes overlap is indicated and designated with a shill (/), and the *edd* start codon within the *zwf* gene is underlined and overlined. The stop codons are labeled in boldface (END). The reported sequence ends at the *Xba*I site. Potential transcriptional terminators and stem-loop structures are overlined.

```
1712  AGG CTA AAA GGA TTC GGC CTC TGT TTT AAG GAC GAG AAT A   ATG ACA AAT ACC GTT TCG ACG ATG ATA TTG TTT GGC TCG ACT GGC GAC  1799
                                                            ZWF Met Thr Asn Thr Val Ser Thr Met Ile Leu Phe Gly Ser Thr Gly Asp

1800  CTT TCA CAG CGT ATG CTG TTG CCG TCG CTT TAT GGT CTT GAT GCC GAT GGT TTG CTT GCA GAT GAT CTG CGT ATC GTC TGC ACC TCT CGT  1889
      Leu Ser Gln Arg Met Leu Leu Pro Ser Leu Tyr Gly Leu Asp Ala Asp Gly Leu Leu Ala Asp Asp Leu Arg Ile Val Cys Thr Ser Arg

1890  AGC GAA TAC GAC ACA GAT GGT TTC CGT GAT TTT GCA GAA AAA GCT TTA GAT CGC TTT GTC GCT TCT GAC CGG TTA AAT GAT GAC GCT AAA  1979
      Ser Glu Tyr Asp Thr Asp Gly Phe Arg Asp Phe Ala Glu Lys Ala Leu Asp Arg Phe Val Ala Ser Asp Arg Leu Asn Asp Asp Ala Lys

1980  GCT AAA TTC CTT AAC AAG CTT TTC TAC GCG ACG GTC GAT ATT ACG GAT CCG ACC CAA TTC GGA AAA TTA GCT GAC CTT TGT GGC CCG GTC  2069
      Ala Lys Phe Leu Asn Lys Leu Phe Tyr Ala Thr Val Asp Ile Thr Asp Pro Thr Gln Phe Gly Lys Leu Ala Asp Leu Cys Gly Pro Val

2070  GAA AAA GGT ATC GCC ATT TAT CTT TCG ACT GCG CCT TCT TTG TTT GAA GGG GCA ATC GCT GGC CTG AAA CAG GCT GGT CTG GCT GGT CCA  2159
      Glu Lys Gly Ile Ala Ile Tyr Leu Ser Thr Ala Pro Ser Leu Phe Glu Gly Ala Ile Ala Gly Leu Lys Gln Ala Gly Leu Ala Gly Pro

2160  ACT TCT CGC CTG GCG CTT GAA AAA CCT TTA GGT CAA GAT CTT GCT TCT TCC GAT CAT ATT AAT GAT GCG GTT TTG AAA GTT TTC TCT GAA  2249
      Thr Ser Arg Leu Ala Leu Glu Lys Pro Leu Gly Gln Asp Leu Ala Ser Ser Asp His Ile Asn Asp Ala Val Leu Lys Val Phe Ser Glu

2250  AAG CAA GTT TAT CGT ATT GAC CAT TAT CTG GGT AAA GAA ACG GTT CAG AAT CTT CTG ACC CTG CGT TTT GGT AAT GCT TTG TTT GAA CCG  2339
      Lys Gln Val Tyr Arg Ile Asp His Tyr Leu Gly Lys Glu Thr Val Gln Asn Leu Leu Thr Leu Arg Phe Gly Asn Ala Leu Phe Glu Pro

2340  CTT TGG AAT TCA AAA GGC ATT GAC CAC GTT CAG ATC AGC GTT GCT GAA ACG GTT GGT CTT GAA GGT CGT ATC GGT TAT TTC GAC GGT TCT  2429
      Leu Trp Asn Ser Lys Gly Ile Asp His Val Gln Ile Ser Val Ala Glu Thr Val Gly Leu Glu Gly Arg Ile Gly Tyr Phe Asp Gly Ser

2430  GGC AGC TTG CGC GAT ATG GTT CAA AGC CAT ATC CTT CAG TTG GTC GCT TTG GTT GCA ATG GAA CCA CCG GCT CAT ATG GAA GCC AAC GCT  2519
      Gly Ser Leu Arg Asp Met Val Gln Ser His Ile Leu Gln Leu Val Ala Leu Val Ala Met Glu Pro Pro Ala His Met Glu Ala Asn Ala

2520  GTT CGT GAC GAA AAG GTA AAA GTT TTC CGC GCT CTG CGT CCG ATC AAT AAC GAC ACC GTC TTT ACG CAT ACC GTT ACC GGT CAA TAT GGT  2609
      Val Arg Asp Glu Lys Val Lys Val Phe Arg Ala Leu Arg Pro Ile Asn Asn Asp Thr Val Phe Thr His Thr Val Thr Gly Gln Tyr Gly

2610  GCC GGT GTT TCT GGT GGT AAA GAA GTT GCC GGT TAC ATT GAC GAA CTG GGT CAG CCT TCC GAT ACC GAA ACC TTT GTT GCT ATC AAA GCG  2699
      Ala Gly Val Ser Gly Gly Lys Glu Val Ala Gly Tyr Ile Asp Glu Leu Gly Gln Pro Ser Asp Thr Glu Thr Phe Val Ala Ile Lys Ala

2700  CAT GTT GAT AAC TGG CGT TGG CAG GGT GTT CCG TTC TAT ATC CGC ACT GGT AAG CGT TTA CCT GCA CGT CGT TCT GAA ATC GTG GTT CAG  2789
      His Val Asp Asn Trp Arg Trp Gln Gly Val Pro Phe Tyr Ile Arg Thr Gly Lys Arg Leu Pro Ala Arg Arg Ser Glu Ile Val Val Gln

2790  TTT AAA CCT GTT CCG CAT TCG ATT TTC TCT TCT TCA GGT GGT ATC TTG CAG CCG AAC AAG CTG CGT ATT GTC TTA CAG CCT GAT GAA ACC  2879
      Phe Lys Pro Val Pro His Ser Ile Phe Ser Ser Ser Gly Gly Ile Leu Gln Pro Asn Lys Leu Arg Ile Val Leu Gln Pro Asp Glu Thr

2880  ATC CAG ATT TCT ATG ATG GTG AAA GAA CCG GGT CTT GAC CGT AAC GGT GCG CAT ATG CGT GAA GTT TGG CTG GAT CTT TCC CTC ACG GAT  2969
      Ile Gln Ile Ser Met Met Val Lys Glu Pro Gly Leu Asp Arg Asn Gly Ala His Met Arg Glu Val Trp Leu Asp Leu Ser Leu Thr Asp

2970  GTG TTT AAA GAC CGT AAA CGT CGT ATC GCT TAT GAA CGC CTG ATG CTT GAT CTT ATC GAA GGC GAT GCT ACT TTA TTT GTG CGT CGT GAC  3059
      Val Phe Lys Asp Arg Lys Arg Arg Ile Ala Tyr Glu Arg Leu Met Leu Asp Leu Ile Glu Gly Asp Ala Thr Leu Phe Val Arg Arg Asp

3060  GAA GTT GAG GCG CAG TGG GTT TGG ATT GAC GGA ATT CGT GAA GGC TGG AAA GCC AAC AGT ATG AAG CCA AAA ACC TAT GTC TCT GGT ACA  3149
      Glu Val Glu Ala Gln Trp Val Trp Ile Asp Gly Ile Arg Glu Gly Trp Lys Ala Asn Ser Met Lys Pro Lys Thr Tyr Val Ser Gly Thr

3150  TGG GGG CCT TCA ACT GCT ATA GCT CTG GCC GAA CGT GAT GGA GTA ACT TGG TAT GAC TGA   T CTG CAT TCA ACG GTA GAA AAG GTT ACC  3237
      Trp Gly Pro Ser Thr Ala Ile Ala Leu Ala Glu Arg Asp Gly Val Thr Trp Tyr Asp End     Leu His Ser Thr Val Glu Lys Val Thr
                                                                 EDD Met Thr Asp

3238  GCG CGC GTT ATT GAA CGC TCG CGG GAA ACC CGT AAG GCT TAT CTG GAT TTG ATC CAG TAT GAG CGG GAA AAA GGC GTA GAC CGT CCA AAC  3327
      Ala Arg Val Ile Glu Arg Ser Arg Glu Thr Arg Lys Ala Tyr Leu Asp Leu Ile Gln Tyr Glu Arg Glu Lys Gly Val Asp Arg Pro Asn
```

FIG. 5—*Continued*

The *edd* AUG start codon lies within the 3′ end of the upstream *zwf* gene, although the two reading frames are in different registers. Likewise, the Shine-Dalgarno sequence for the *edd* gene is buried within the *zwf* coding region, 8 bp upstream of the *edd* start codon. The sequence between the GGAG ribosome-binding site and the start codon is A+T-rich, which is conducive to efficient translation initiation (50). The high level of *edd* expression that was obtained with *E. coli* RW231R(pTC150), which is devoid of the entire 5′ end of the *zwf* gene, suggests that translation of the *zwf* gene is not required for *edd* translation. However, these results were obtained with *Z. mobilis* genes being expressed in *E.*

```
3328  CTG TCC TGT AGT AAC CTT GCT CAT GGC TTT GCG GCT ATG AAT GGT GAC AAG CCA GCT TTG CGC GAC TTC AAC CGC ATG AAT ATC GGC GTC  3417
      Leu Ser Cys Ser Asn Leu Ala His Gly Phe Ala Ala Met Asn Gly Asp Lys Pro Ala Leu Arg Asp Phe Asn Arg Met Asn Ile Gly Val

3418  GTG ACT TCC TAC AAC GAT ATG TTG TCG GCT CAT GAA CCG TAT TAT CGC TAT CCG GAG CAG ATG AAA GTA TTT GCT CGC GAA GTT GGC GCA  3507
      Val Thr Ser Tyr Asn Asp Met Leu Ser Ala His Glu Pro Tyr Tyr Arg Tyr Pro Glu Gln Met Lys Val Phe Ala Arg Glu Val Gly Ala

3508  ACG GTT CAG GTC GCC GGT GGC GTG CCT GCT ATG TGC GAT GGT GTG ACC CAA GGT CAG CCG GGC ATG GAA GAA TCC CTG TTT AGC CGC GAT  3597
      Thr Val Gln Val Ala Gly Gly Val Pro Ala Met Cys Asp Gly Val Thr Gln Gly Gln Pro Gly Met Glu Glu Ser Leu Phe Ser Arg Asp

3598  GTC ATC GCT TTG GCT ACC AGC GTT TCT TTG TCT CAT GGT ATG TTT GAA GGG GCT GCT CTT CTC GGT ATC TGT GAC AAG ATT GTC CCT GGT  3687
      Val Ile Ala Leu Ala Thr Ser Val Ser Leu Ser His Gly Met Phe Glu Gly Ala Ala Leu Leu Gly Ile Cys Asp Lys Ile Val Pro Gly

3688  CTG TTG ATG GGC GCT CTG CGT TTC GGT CAC CTG CCG ACC ATT CTG GTC CCA TCA GGC CCG ATG ACG ACT GGT ATC CCG AAC AAA GAA AAA  3777
      Leu Leu Met Gly Ala Leu Arg Phe Gly His Leu Pro Thr Ile Leu Val Pro Ser Gly Pro Met Thr Thr Gly Ile Pro Asn Lys Glu Lys

3778  ATC CGT ATC CGT CAG CTC TAT GCT CAG GGT AAA ATC GGC CAG AAA GAA CTT CTG GAT ATG GAA GCG GCT TGC TAC CAT GCT GAA GGT ACC  3867
      Ile Arg Ile Arg Gln Leu Tyr Ala Gln Gly Lys Ile Gly Gln Lys Glu Leu Leu Asp Met Glu Ala Ala Cys Tyr His Ala Glu Gly Thr

3868  TGC ACC TTC TAT GGT ACG GCA AAC ACC AAC CAG ATG GTT ATG GAA GTC CTC GGT CTT CAT ATG CCA GGT TCG GCA TTT GTT ACC CCG GGT  3957
      Cys Thr Phe Tyr Gly Thr Ala Asn Thr Asn Gln Met Val Met Glu Val Leu Gly Leu His Met Pro Gly Ser Ala Phe Val Thr Pro Gly

3958  ACC CCG CTC CGT CAG GCT CTG ACC CGT GCT GCT GTG CAT CGC GTT GCT GAA TTG GGT TGG AAG GGC GAT GAT TAT CGT CCG CTT GGT AAG  4047
      Thr Pro Leu Arg Gln Ala Leu Thr Arg Ala Ala Val His Arg Val Ala Glu Leu Gly Trp Lys Gly Asp Asp Tyr Arg Pro Leu Gly Lys

4048  ATC ATT GAC GAA AAA TCA ATC GTC AAT GCT ATT GTT GGT CTG TTG GCA ACC GGT GGT TCC ACC AAC CAT ACC ATG CAT ATT CCG GCT ATT  4137
      Ile Ile Asp Glu Lys Ser Ile Val Asn Ala Ile Val Gly Leu Leu Ala Thr Gly Gly Ser Thr Asn His Thr Met His Ile Pro Ala Ile

4138  GCT CGT GCT GCT GGT GTT ATC GTT AAC TGG AAT GAC TTC CAT GAT CTT TCT GAA GTT GTT CCG TTG ATT GCC CGC ATT TAC CCG AAT GGC  4227
      Ala Arg Ala Ala Gly Val Ile Val Asn Trp Asn Asp Phe His Asp Leu Ser Glu Val Val Pro Leu Ile Ala Arg Ile Tyr Pro Asn Gly

4228  CCG CGC GAC ATC AAT GAA TTC CAG AAT GCA GGC GGC ATG GCT TAT GTC ATC AAA GAA CTG CTT TCT GCT AAT CTG TTG AAC CGT GAC GTC  4317
      Pro Arg Asp Ile Asn Glu Phe Gln Asn Ala Gly Gly Met Ala Tyr Val Ile Lys Glu Leu Leu Ser Ala Asn Leu Leu Asn Arg Asp Val

4318  ACG ACC ATT GCC AAG GGC GGT ATC GAA GAA TAC GCC AAG GCT CCG GCA TTA AAT GAT GCT GGC GAA TTG GTA TGG AAG CCA GCT GGC GAA  4407
      Thr Thr Ile Ala Lys Gly Gly Ile Glu Glu Tyr Ala Lys Ala Pro Ala Leu Asn Asp Ala Gly Glu Leu Val Trp Lys Pro Ala Gly Glu

4408  CCT GGT GAT GAC ACC ATT CTG CGT CCG GTT TCT AAT CCT TTC GCA AAA GAT GGC GGT CTG CGT CTC TTG GAA GGT AAC CTT GGC CGT GCA  4497
      Pro Gly Asp Asp Thr Ile Leu Arg Pro Val Ser Asn Pro Phe Ala Lys Asp Gly Gly Leu Arg Leu Leu Glu Gly Asn Leu Gly Arg Ala

4498  ATG TAC AAG GCC AGT GCA GTT GAT CCT AAA TTC TGG ACT ATC GAA GCA CCG GTT CGC GTC TTC TCT GAC CAA GAC GAT GTT CAG AAA GCC  4587
      Met Tyr Lys Ala Ser Ala Val Asp Pro Lys Phe Trp Thr Ile Glu Ala Pro Val Arg Val Phe Ser Asp Gln Asp Asp Val Gln Lys Ala

4588  TTC AAG GCT GGC GAA TTG AAC AAA GAC GTT ATC GTT GTT GTT CGT TTC CAG GGC CCG CGC GCA AAC GGT ATG CCT GAA TTG CAT AAG CTA  4677
      Phe Lys Ala Gly Glu Leu Asn Lys Asp Val Ile Val Val Val Arg Phe Gln Gly Pro Arg Ala Asn Gly Met Pro Glu Leu His Lys Leu

4678  ACC CCG GCT TTG GGT GTT CTG CAG GAT AAT GGC TAC AAA GTT GCT TTG GTA ACT GAT GGT CGT ATG TCC GGT GCT ACC GGT AAA GTT CCG  4767
      Thr Pro Ala Leu Gly Val Leu Gln Asp Asn Gly Tyr Lys Val Ala Leu Val Thr Asp Gly Arg Met Ser Gly Ala Thr Gly Lys Val Pro

4768  GTT GCT TTG CAT GTC AGC CCA GAA GCT CTT GGC GGT GGT GCC ATC GGT AAA TTA CGT GAT GGC GAT ATC GTC CGT ATC TCG GTT GAA GAA  4857
      Val Ala Leu His Val Ser Pro Glu Ala Leu Gly Gly Gly Ala Ile Gly Lys Leu Arg Asp Gly Asp Ile Val Arg Ile Ser Val Glu Glu

4858  GGC AAA CTT GAA GCT TTG GTT CCA GCT GAT GAG TGG AAT GCT CGT CCA CAT GCT GAA AAA CCG GCT TTC CGT CCG GGA ACC GCG CGA ATT  4947
      Gly Lys Leu Glu Ala Leu Val Pro Ala Asp Glu Trp Asn Ala Arg Pro His Ala Glu Lys Pro Ala Phe Arg Pro Gly Thr Ala Arg Ile
```

FIG. 5—*Continued*

coli and may have no bearing on the in vivo events of *zwf-edd* expression in *Z. mobilis*. Analysis of the *edd* sequence revealed a substantial stem-loop structure 32 bp downstream of the start codon. This structure consists of a 14-bp inverted repeat in which 12 of the 14 bases can pair, with a 4-bp loop formed (Fig. 5). The *edd* reading frame ends with the stop codon UGA. An open reading frame begins 203 bp downstream of the *edd* stop codon (*glk*, described below). Another substantial stem-loop structure is present 115 bp downstream of the *edd* gene (Fig. 5). This structure consists of an exact 13-base inverted repeat separated by 4 bases that would form a loop. The ability to express both *edd* and *glk* from pTC150 in *E. coli* suggests that neither of these stem-loop structures serves as a transcriptional terminator.

(iii) *glk* gene. The *Z. mobilis glk* gene consists of an open reading frame of 984 bp and encodes a protein of 327 amino

```
4948  GTT TGA TAT CTT CCG TCA GAA CGC TGC TAA AGC TGA AGA CGG TGC AGT CGC AAT ATA TGC AGG TGC CGG TAT CTA ATT TTT CCA GCG AAA  5037
      Val End

5038  AAT TGT AGA CTT GGA CTT TGT AAT CTT ATT TTC TGG TAT AGG GCT ATC CCT ATA CCA GAA CTG AAA TCA GAC TTA TTT TAC CTG TTG GGT  5127

5128  AGC CTT CTG ATT TTA GAA AGG AAT TAT T   ATG GAA ATT GTT GCG ATT GAC ATC GGT GGA ACG CAT GCG CGT TTC TCT ATT GCG GAA GTA  5215
                                      GLK Met Glu Ile Val Ala Ile Asp Ile Gly Gly Thr His Ala Arg Phe Ser Ile Ala Glu Val

5216  AGC AAT GGT CGG GTT CTT TCT CTT GGA GAA GAA ACA ACT TTT AAA ACG GCA GAA CAT GCT AGC TTG CAG TTA GCT TGG GAA CGT TTC GGT  5305
      Ser Asn Gly Arg Val Leu Ser Leu Gly Glu Glu Thr Thr Phe Lys Thr Ala Glu His Ala Ser Leu Gln Leu Ala Trp Glu Arg Phe Gly

5306  GAA AAA CTG GGT CGT CCT CTG CCA CGT GCC GCA GCT ATT GCA TGG GCT GGC CCG GTT CAT GGT GAA GTT TTA AAA CTT ACC AAT AAC CCT  5395
      Glu Lys Leu Gly Arg Pro Leu Pro Arg Ala Ala Ala Ile Ala Trp Ala Gly Pro Val His Gly Glu Val Leu Lys Leu Thr Asn Asn Pro

5396  TGG GTA TTA AGA CCA GCT ACT CTG AAT GAA AAG CTG GAC ATC GAT ACG CAT GTT CTG ATC AAT GAC TTC GGC GCG GTT GCC CAC GCG GTT  5485
      Trp Val Leu Arg Pro Ala Thr Leu Asn Glu Lys Leu Asp Ile Asp Thr His Val Leu Ile Asn Asp Phe Gly Ala Val Ala His Ala Val

5486  GCG CAT ATG GAT TCT TCT TAT CTG GAT CAT ATT TGT GGT CCT GAT GAA GCG CTT CCT AGC GAT GGT GTT ATC ACT ATT CTT GGT CCG GGA  5575
      Ala His Met Asp Ser Ser Tyr Leu Asp His Ile Cys Gly Pro Asp Glu Ala Leu Pro Ser Asp Gly Val Ile Thr Ile Leu Gly Pro Gly

5576  ACG GGC TTG GGT GTT GCC CAT CTG TTG CGG ACT GAA GGC CGT TAT TTC GTC ATC GAA ACT GAA GGC GGT CAT ATC GAC TTT GCT CCG CTT  5665
      Thr Gly Leu Gly Val Ala His Leu Leu Arg Thr Glu Gly Arg Tyr Phe Val Ile Glu Thr Glu Gly Gly His Ile Asp Phe Ala Pro Leu

5666  GAC AGA CTT GAA GAC AAA ATT CTG GCA CGT TTA CGT GAA CGT TTC CGC CGC GTT TCT ATC GAA CGC ATT ATT TCT GGC CCG GGT CTT GGT  5755
      Asp Arg Leu Glu Asp Lys Ile Leu Ala Arg Leu Arg Glu Arg Phe Arg Arg Val Ser Ile Glu Arg Ile Ile Ser Gly Pro Gly Leu Gly

5756  AAT ATC TAC GAA GCA CTG GCT GCC ATT GAA GGC GTT CCG TTC AGC TTG CTG GAT GAT ATT AAA TTA TGG CAG ATG GCT TTG GAA GGT AAA  5845
      Asn Ile Tyr Glu Ala Leu Ala Ala Ile Glu Gly Val Pro Phe Ser Leu Leu Asp Asp Ile Lys Leu Trp Gln Met Ala Leu Glu Gly Lys

5846  GAC AAC CTT GCT GAA GCC GCT TTG GAT CGC TTC TGC TTG AGC CTT GGC GCT ATC GCT GGT GAT CTT GCT TTG GCA CAG GGT CGA ACC AGT  5935
      Asp Asn Leu Ala Glu Ala Ala Leu Asp Arg Phe Cys Leu Ser Leu Gly Ala Ile Ala Gly Asp Leu Ala Leu Ala Gln Gly Arg Thr Ser

5936  GTT GTT ATT GGC GGT GGT GTC GGT CTT CGT ATC GCT TCC CAT TTG CCA GAA TCT GGT TTC CGT CAG CGC TTT GTT TCA AAA GGA CGC TTT  6025
      Val Val Ile Gly Gly Gly Val Gly Leu Arg Ile Ala Ser His Leu Pro Glu Ser Gly Phe Arg Gln Arg Phe Val Ser Lys Gly Arg Phe

6026  GAA CGC GTC ATG TCC AAG ATT CCG GTT AAG TTG ATT ACT TAT CCG CAG CCT GGA CTG TTG GGT GCG CAG CTG CCT ATG CCA ACA AAT ATT  6115
      Glu Arg Val Met Ser Lys Ile Pro Val Lys Leu Ile Thr Tyr Pro Gln Pro Gly Leu Leu Gly Ala Gln Leu Pro Met Pro Thr Asn Ile

6116  CTG AAG TTG AAT AAT ATT TTT TAA TAT TAT GAA CTG AAT TTA AGA GGC TGC CTT CCG ATA AAA TCG GGA GGT GGC CTT TTT TAT ATT TTT  6205
      Leu Lys Leu Asn Asn Ile Phe End

6206  TAC TAA AAA ATG AAG ACA AAA AAG TCT TAA GTA AGA ATA ATA TTA TTA TTA ACT TTT GAT ATA TTT TGT ATT AGT TCT TCT TGG TGA AGA  6295

6296  ATT ATT TTT GAT AAA TTT TGT CTA ATA TCC TAT ATT TTA AAT ATT TTT TAT AAT GTT TTT TTA ATA AAA TTG ACG TGA TAT TTA GGG AGT  6385

6386  TGT GTA GAA AAA TGA GAT AAT ATT TAG AAT TAT TGA TTT AAT TCT ATC TAG AGC G  6440
```

FIG. 5—*Continued*

acids with an aggregate molecular weight of 35,422 (Fig. 5). The reading frame begins with an AUG start codon that is preceded 6 bp upstream by a Shine-Dalgarno sequence. The sequence between the ribosome-binding site and the start codon consists entirely of AT residues. The *glk* reading frame ends with the stop codon UAA. The sequence down stream of *glk* is extremely A+T-rich, contains numerous strings of Ts, and contains at least three structures that resemble simple rho-independent transcriptional terminators (44, 57). The most obvious of these begins 27 bp downstream of the *glk* stop codon (Fig. 5). Comparison of the deduced amino acid sequence for *Z. mobilis* glucokinase with glucokinase from *Saccharomyces cerevisiae* (2) and rats (4) revealed sufficient identity to suggest that they are similar proteins, but also revealed some striking differences. The *Z. mobilis* glucokinase is considerably smaller (327 amino acids) than yeast and rat glucokinases, which are 501 and 465 amino acids long, respectively. The *Z. mobilis* glucokinase is 21% identical and 49% similar to yeast glucokinase and is 22% identical and 50% similar to rat glucokinase. The *Z. mobilis* glucokinase contains a region that is similar to the ATP-binding domains of yeast and rat glucokinases, but in *Z. mobilis* this region begins at amino acid 7, whereas in yeast glucokinase the ATP-binding domain begins at amino acid 86 and in rat glucokinase the region begins at amino acid 78. Both the rat and yeast glucokinases contain putative glucose-binding domains, but no such region was found in the *Z. mobilis* glucokinase in either orientation.

```
Zm    1  MTDLHSTVEKVTARVIERSRETRKAYLDLIQYEREKGVDRPNLSCSNLAHGFAAMNGDKPALRDFNRMNIGVVTSYNDMLSAHEPYYRYPEQMKVFAREV  100
Ec    1                                MPKYRSATTTHGRNMAGARALWRATGMTDA.....DFGKPIIAVVNSFTQFVPGH.......VHLRDLGKLV   60

Zm  101  GATVQVAGGVP......AMCDGVTQGQPGMEESLFSRDVIALATSVSLSHGMFEGAALLGI..CDKIVPGLLMGALRFGHLPTILVPSGPMTTG......  186
Ec   61  AEQIEAAGGVAKEFNTIAVDDGIAMGHGGMLYSLPSRELI..ADSVEYMVNAHCADAMVCISNCDKITPGMLMASLRL.NIPVIFVSGGPMEAGKTKLSD  157

Zm  187  .IPNKEKIRIRQLYAQGKIGQKELLDMEAACYHAEGTCTFYGTANTNQMVMEVLGLHMPGSAFVTPGTPLRQALTRAAVHRVAELG...WKGDDYRPLGK  282
Ec  158  QIIKLDLVDAMIQGADPKVSDSQSDQVERSACPTCGSCSGMFTANSMNCLTEALGLSQPGNGSLLATHADRKQLFLNAGKRIVELTKRYYEQNDESALPR  257

Zm  283  IIDEKSIVNAIVGL.LATGGSTNHTMHIPAIARAAGVIVNWNDFHDLSEVVPLIARIYPNGPR.DINEFQNAGGMAYVIKELLSANLLNRDVTTI....A  376
Ec  258  NIASKAAFENAMTLDIAMGGSTNTVLHLLAAAQEAEIDFTMSDIDKLSRKVPQLCKVAPSTQKYHMEDVHRAGGVIGILGELDRAGLLNRDVKNVLGLTL  357

Zm  377  KGGIEEYAKAPALNDAGELVWKPA.....GEPGDDTILRPVSNPFAKDGGLRLLEGNLGRAMY..KASAVDPKFWTIEAPVRVFSD  455
Ec  358  PQTLEQYDVMLTQDDAVKNMFRAGPAGIRTTQAFSQDCRWDTLDDDRANGCIRSLEHAYSKDGGLAVLYGNFAENGCIVKTAGVDDSILKFTGPAKVYES  457

Zm  456  QDDVQKAFKAGELNKDVIVVVRFQGPRAN.GMPELHKLTPALGVLQDNGYKVALVTDGRMSGATGKVPVALHVSPEALGGGAIGKLRDGDIVRISVEEGK  554
Ec  458  QDDAVEAILGGKVVAGDVVVIRYEGPKGGPGMQEMLYPTSFLKSM.GLGKACALITDGRFSGGTSGLSIG.HVSPEAASGGSIGLIEDGDLIAIDIPNRG  555

Zm  555  LEALVPADEWNARPHAEKPAFRPGTARIV*  583
Ec  556  IQLQVSDAELAARREAQDA..RGDKAWTPKNRERQVSFALRAYASLATSADKGAVRDKSKLGG*  616
```

FIG. 6. Amino acid comparison of the putative *Z. mobilis edd* gene product (Zm) with the *E. coli ilvD* gene product (Ec). Identity is indicated by a vertical line (|); gaps are indicated by a dot (·).

(iv) *glf* gene. The reading frame that lies upstream of the *zwf* gene is 1,422 bp long and encodes a protein of 473 amino acids with a calculated molecular weight of 50,199. When this coding region was used as a query sequence to the GenBank database, two significant matches were made (Fig. 7). The *Z. mobilis* deduced amino acid sequence is 42% identical and 66% similar to the *E. coli xylE* gene that encodes a xylose-proton symport protein which mediates uptake of xylose (19). In addition, the *Z. mobilis* protein is 29% identical and 55% similar to the human glucose transporter (38). A hydropathy plot of the *Z. mobilis* protein indicated that it contains only 17% hydrophilic residues (33). Examination of this plot showed that the *Z. mobilis* protein contains 12 hydrophobic domains that are predicted, on the basis of a hydrophobic moment plot (data not shown), to be membrane-spanning domains (24). Thus, the *Z. mobilis* protein fits precisely the model for both the *E. coli* xylose-proton symporter and the human glucose transporter as membrane-spanning, carbohydrate transport proteins. On the basis of these data, we believe that the *Z. mobilis* gene in question codes for a glucose-facilitated diffusion protein and we have named the gene *glf*. The presence of a facilitated diffusion system for glucose entry into *Z. mobilis* was established by DiMarco and Romano (21). There are no obvious open reading frames in the 185 bp upstream of the *glf* gene, and the gene cluster appears to begin with this gene. The *glf* reading frame begins with an AUG codon which is preceded by a good Shine-Dalgarno sequence, but this is within a region that is G+C-rich. For this reason, the *glf* gene may not be translated as efficiently as some of the other glycolytic genes.

DISCUSSION

The data in this report provide genetic and biochemical evidence that we cloned four genes that are involved in the uptake and early steps in glucose catabolism by *Z. mobilis*. Although three of the genes were cloned independently, numerous experiments, including nucleotide sequence analysis, established that *glf*, *zwf*, *edd*, and *glk* are clustered in a 6-kb region of the *Z. mobilis* genome. Knowledge of the structure of these genes and the deduced primary amino acid sequences allows comparisons at the protein level of how closely related *Z. mobilis* glycolytic enzymes are to those from other organisms. To our knowledge, these are the first sequences reported for any of these genes in any procaryote. This information will be used as the basis for a detailed examination of the mechanisms of expression of these genes.

The *Z. mobilis zwf* gene that encodes glucose-6-phosphate dehydrogenase was cloned by genetic complementation of a specific defect in glucose metabolism in *E. coli* DF214 (52). Proof of the identity of the cloned gene comes from biochemical assay of the enzyme produced by *E. coli* DF214(pTC111), SDS-polyacrylamide gel electrophoresis, and native polyacrylamide gel electrophoresis accompanied by staining the gel for glucose-6-phosphate dehydrogenase activity. Proof that the cloned *zwf* gene did indeed come from *Z. mobilis* was provided by Southern gel analysis. Comparison of the deduced *Z. mobilis* glucose-6-phosphate dehydrogenase primary amino acid sequence with that of the human enzyme (40) revealed 36% identity, slightly lower than the 43% identity to a 42-residue active site peptide of *L. mesenteroides* (10). To our knowledge, this is the first

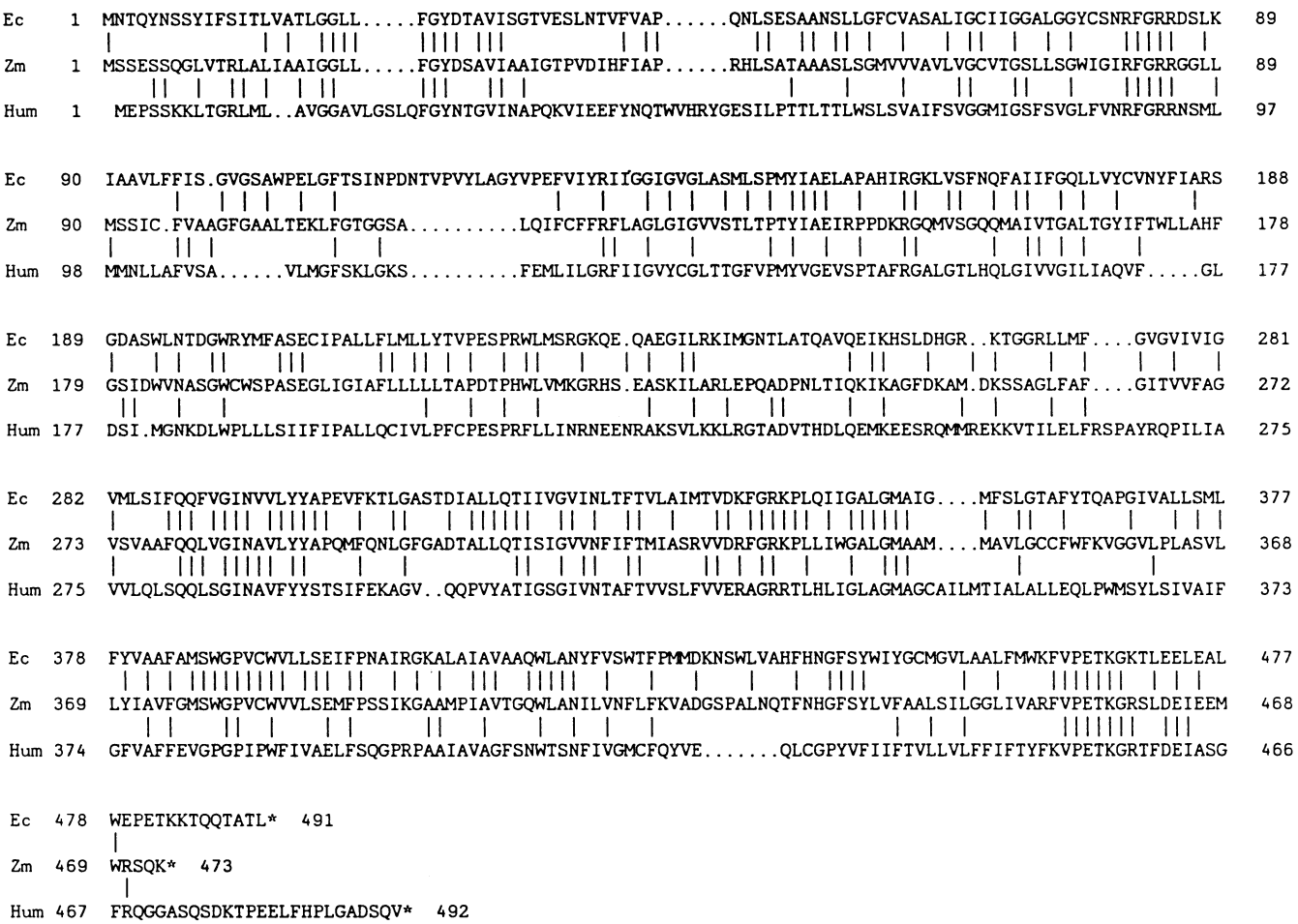

FIG. 7. Amino acid identity of *E. coli xylE* gene product (Ec), putative *Z. mobilis glf* gene product (Zm), and the human glucose transporter (Hum). Identity is indicated by a vertical line (|); gaps are indicated by a dot (·).

nucleotide sequence of a procaryotic glucose-6-phosphate dehydrogenase gene to be reported. The *Z. mobilis zwf* gene begins with an AUG initiation codon and a strong ribosome-binding site, in keeping with its concentration of over 1% of the total soluble protein (48). The glycolytic enzymes together total between 35 and 50% of the soluble protein in *Z. mobilis* (53). All the *Z. mobilis* glycolytic genes that have been studied thus far possess similar strong ribosome-binding sites (32, 43). This is in contrast to the few genes of biosynthetic and salvage pathways in this organism that have been studied so far, which are expressed at far lower levels than the glycolytic pathway and possess weak ribosome-binding sites (32). It can be concluded that one reason for high-level expression of the glycolytic enzymes in *Z. mobilis* is because of efficient translation initiation.

The *Z. mobilis edd* gene encodes 6-phosphogluconate dehydratase, one of two key enzymes of the Entner-Doudoroff pathway (25). The cloned gene was obtained by genetic complementation of a specific defect in gluconate metabolism in *E. coli* RW231 (55). Biochemical assay of the *Z. mobilis* 6-phosphogluconate dehydratase in *E. coli* RW231 and overexpression of a 63,000-Da protein served to confirm the cloning of the *edd* gene. Southern gel analysis showed that the gene was indeed cloned from *Z. mobilis*. Linkage of *edd* to *zwf* was established by comparison of the restriction maps of pTC111 and pTC120, as well as by nucleotide sequence analysis. The *edd* gene begins with a strong ribosome-binding site and AUG start codon, both of which lie within the 3′ end of the *zwf* gene. Expression of the *Z. mobilis edd* gene in *E. coli* suggests that translation of *zwf* is not required for translation of *edd*. We intend to examine the possibility that the *Z. mobilis zwf* and *edd* genes are cotranslationally coupled.

The fact that the *Z. mobilis* protein is related to the *E. coli* dihydroxyacid dehydratase (*ilvD* gene product) is extremely interesting (17). *E. coli* also possesses the Entner-Doudoroff pathway and has an *edd* gene that is certainly distinct from *ilvD* (5). It is highly likely that the *edd* gene of *E. coli* is closely related to the *Z. mobilis edd* gene. We are currently sequencing the *E. coli edd* gene for comparison with *E. coli ilvD* and *Z. mobilis edd*.

The *Z. mobilis glk* gene that encodes glucokinase was cloned by genetic complementation of an *E. coli* mutant that is deficient in glucose phosphorylation (28). Biochemical assay and a zymogram were used to confirm that the cloned gene did indeed code for glucokinase, and Southern gel analysis was used to confirm that the cloned gene hybridized to *Z. mobilis* DNA only. Linkage of the *glk* and *edd* genes was established by assay of glucokinase and 6-phosphogluconate dehydratase activity in appropriate *E. coli* strains containing pTC120 or pTC150, by comparison of the restriction maps of the two independently obtained clones, and by

sequence analysis. The *glk* reading frame begins with an AUG initiation codon and a strong ribosome-binding site. The *glk* gene is followed by a highly A+T-rich region that contains three or more structures that resemble rho-independent simple terminators. The *Z. mobilis* glucokinase is much smaller than those from *S. cerevisiae* (2) and rats (4) and also appears in other ways to be very different from these. The *Z. mobilis* enzyme possesses an ATP-binding domain, but this begins at amino acid residue 7, approximately 75 amino acid residues closer to the N terminus than in yeast and rat glucokinases. The *Z. mobilis* glucokinase does not contain a consensus glucose-binding domain as do the yeast and rat enzymes (4). It is not possible to predict whether these glucokinases arose by convergent or divergent evolution (26). The existence of vastly different proteins which catalyze the same reaction is interesting and not unprecedented in *Z. mobilis*. *Z. mobilis* possesses two alcohol dehydrogenase enzymes with vastly different primary amino acid sequences that have apparently arisen by convergent evolution to similar function (14, 16).

An open reading frame that lies 146 bp upstream of the *zwf* gene was identified by nucleotide sequence analysis and found to be significantly similar to the *E. coli xylE* gene that encodes a xylose-proton symporter (19) and to the gene that codes for the human glucose transporter (38). A hydrophobic moment plot of the *Z. mobilis* protein predicts that 12 separate hydrophobic regions in this protein are membrane-spanning domains. On this basis, the *Z. mobilis* protein appears to have a structure similar to that of the xylose-proton symporter and the human glucose transporter, both of which are membrane-spanning carbohydrate transport proteins. *Z. mobilis* obtains only a single mole of ATP per mole of glucose fermented and therefore cannot expend energy for glucose transport (53). A facilitated diffusion system for glucose transport in *Z. mobilis* was predicted (9) and later confirmed by DiMarco and Romano (21). We believe that the reading frame that lies upstream of *zwf* in *Z. mobilis* encodes the glucose-facilitated diffusion protein (*glf* gene). Experiments to confirm the biochemical function of the *glf* gene product are under way.

Examination of the genetic organization of these four clustered genes that are involved in consecutive steps of glucose metabolism raises several provocative questions regarding the mechanism of expression of these genes and the role of gene expression in control of carbon flux into and through the Entner-Doudoroff glycolytic pathway. The genes in this cluster are organized in the order *glf*, *zwf*, *edd*, *glk*. The proximity of these genes suggests that they form an operon and are transcribed on a polycistronic mRNA. The finding that the *zwf* and *edd* genes overlap makes this very likely, although the presence of a promoter internal to the *zwf* gene, or elsewhere within the gene cluster, has not yet been ruled out. The gap between the *glf* and *zwf* genes is 146 bp, and the gap between the *edd* and *glk* genes is 203 bp. In both cases, this is shorter than the 222-bp intergenic spacing of the *gap* and *pgk* genes which form an operon in *Z. mobilis* (23). The presence of a strong transcriptional terminator downstream of the *glk* gene indicates that the operon ends at that point.

The clustering of these genes in *Z. mobilis* is not unprecedented. In *P. aeruginosa*, the *glk*, *zwf*, *edd*, and *eda* genes are strictly coinducible and are clustered, along with *gltB*, which is thought to encode the glucose-binding protein that is involved in uptake (42). In *E. coli*, the *zwf*, *edd*, and *eda* genes are cotransducible but are each regulated by independent means (27). In *E. coli*, the *glk* gene is unlinked to this region, as are the carbohydrate transport genes (5). The *Z. mobilis* gene cluster is similar to that of both these organisms, but with significant differences. In *Z. mobilis*, the glucose-facilitated diffusion protein is probably coregulated with the other genes in the cluster. Also, the *Z. mobilis eda* gene is not a part of the gene cluster and, on the basis of Southern gel analysis, appears to be further than 5 kb from the cluster (unpublished data, this laboratory). Questions of the overall organization of the genes that encode the enzymes of central metabolism await physical mapping of the *Z. mobilis* genome. It will be interesting to see whether the *gap* operon and other genes of interest are located near to the *glf* gene cluster.

There are several sequences within the *Z. mobilis* gene cluster that might serve to modulate transcriptional readthrough or would impart very interesting secondary structure to the transcript for these genes. In the intergenic region between the *glf* and *zwf* genes, there is a single weak rho-independent terminator. The ability to express *zwf* from pTC111 suggests that this terminator does not completely terminate transcription, but this observation is made without knowledge of whether internal promoters exist and is made in *E. coli*. Possible regulation of transcriptional termination at this point in vivo in *Z. mobilis* remains to be tested. It is interesting to speculate that this terminator might serve as a control point for modulating the level of the facilitated diffusion protein with respect to the levels of the pathway enzymes. This may be necessary for growth under conditions of glucose limitation, since the K_m for glucose uptake is 5 to 15 mM (21), while the K_m of glucokinase for glucose is 0.2 to 0.5 mM (22, 48). There are two sequences within the gene cluster that would allow substantial stem-loop structures to form in the transcript. The first of these is immediately downstream of the *edd* start codon, and the second is in the intergenic region between the *edd* and *glk* genes. Experiments designed to localize the promoters for the *glf* gene cluster as well as to elucidate the involvement of these structures in gene regulation are in progress. A similar stem-loop structure within the *Z. mobilis pdc* transcript seems to serve as a site for transcript processing, based on primer extension mapping of the 5′ end of the mRNA to this region (15). A similar strategy for differential expression of genes within an operon appears to operate in *Rhodobacter capsulatus* (1).

One last feature of the *glf* gene cluster deserves attention, the overlap of the *zwf* and *edd* genes. *Z. mobilis* differs from *E. coli* in the arrangement of these genes. Despite the tight linkage of *zwf* and *edd* in *E. coli*, recent evidence indicates that these genes are separated by more than 100 bp (R. Wolf, personal communication). These genes code for the two enzymes that convert glucose-6-phosphate to 2-keto-3-deoxy-phosphogluconate (KDPG), which is a toxic metabolite (27). While *Z. mobilis* requires unusually high carbon flux to produce sufficient energy for growth, the toxicity of KDPG dictates that the levels of this metabolic intermediate must be kept low. The physiological levels of the enzymes that lead to KDPG production, glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydratase, are fairly well balanced at 1.4 and 1.8 μmol/min/mg of total cell protein, respectively (3). The physiological level of KDPG aldolase is 3.0 μmol/min/mg of total cell protein, an activity that is high enough to keep KDPG levels relatively low (3). These activities are consistent with the possibility that *zwf* and *edd* are coregulated and that *eda* is regulated separately from the *glf* gene cluster. The arrangement of the *zwf* and *edd* genes in an overlapping motif suggests that their expres-

sion is translationally coupled (59). The specific activity of purified glucose-6-phosphate dehydrogenase is 500 μmol/min/mg of protein (56), whereas the specific activity of purified 6-phosphogluconate dehydratase is 245 μmol/min/mg of protein (47). The measured in vivo activities can only be obtained if the latter enzyme is present in roughly twofold-higher amounts than the former enzyme. Therefore, tight translational coupling of *zwf* and *edd* probably does not occur in *Z. mobilis*, but translational coupling in addition to independent initiation of *edd* might ensure that appropriate expression of the two genes is achieved. This would be analogous to the situation for translation of the *E. coli trpB* and *trpA* genes (59). This possibility is being tested.

ACKNOWLEDGMENTS

We thank R. Wolf and J. Osterman for helpful discussions during this work and B. Bachmann for kindly providing strains. We greatly appreciate the technical assistance of Denise Jones-Kilpatrick. Some sequence was obtained by the UNL Center for Biotechnology DNA Sequencing Facility. Computer time was provided courtesy of the UNL Computer Resources Center.

This work was supported in part by the UN Foundation Layman Fund, Biomedical Research support grant RR-07055 from the National Institutes of Health, and the U.S. Department of Energy (DE-FG02-90ER20002).

LITERATURE CITED

1. **Adams, C. W., M. Forrest, S. Cohen, and J. T. Beatty.** 1989. Structural and functional analysis of transcriptional control of the *Rhodobacter capsulatus puf* operon. J. Bacteriol. **171**:473–482.
2. **Albig, W., and K.-D. Entian.** 1988. Structure of yeast glucokinase, a strongly diverged specific aldo-hexose-phosphorylating isoenzyme. Gene **73**:141–152.
3. **Algar, E. M., and R. Scopes.** 1985. Studies on cell-free metabolism: ethanol production by extracts of *Zymomonas mobilis*. J. Biotechnol. **2**:275–287.
4. **Andreone, T. L., R. Printz, S. Pilkis, M. Magnuson, and D. Granner.** 1989. The amino acid sequence of rat liver glucokinase deduced from cloned cDNA. J. Biol. Chem. **264**:363–369.
5. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
6. **Bachmann, B.** 1987. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1190–1219. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Esherichia coli* and *Salmonella typhimurium*: Cellular and molecular biology. American Society for Microbiology, Washington, D.C.
7. **Baker, H. V., and R. E. Wolf.** 1983. Growth rate-dependent regulation of 6-phosphogluconate dehydrogenase level in *Escherichia coli* K-12: β-galactosidase expression in *gnd-lac* operon fusion strains. J. Bacteriol. **153**:771–781.
8. **Barrow, K. D., J. G. Collins, R. S. Norton, P. L. Rogers, and G. M. Smith.** 1984. ^{31}P nuclear magnetic resonance studies of the fermentation of glucose to ethanol by *Zymomonas mobilis*. J. Biol. Chem. **259**:5711–5716.
9. **Belaich, J. P., J. Senez, and M. Murgier.** 1968. Microcalorimetric study of glucose permeation in microbial cells. J. Bacteriol. **95**:1750–1757.
10. **Bhadbhade, M. M., M. J. Adams, T. G. Flynn, and H. R. Levy.** 1987. Sequence identity between a lysine-containing peptide from *Leuconostoc mesenteroides* glucose-6-phosphte dehydrogenase and an active site peptide from human erythrocyte glucose-6-phosphate dehydrogenase. FEBS Lett. **211**:243–246.
11. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1520.
12. **Brumbaugh, J. A., L. R. Middendorf, D. L. Grone, and J. L. Ruth.** 1988. Continuous, on-line DNA sequencing using oligodeoxynucleotide primers with multiple fluorophores. Proc. Natl. Acad. Sci. USA **85**:5610–5614.
13. **Byun, M. O.-K., J. B. Kaper, and L. O. Ingram.** 1986. Construction of a new vector for the expression of foreign genes in *Zymomonas mobilis*. J. Ind. Microbiol. **1**:9–15.
14. **Conway, T., and L. O. Ingram.** 1989. Similarity of *Escherichia coli* propanediol oxidoreductase (*fucO* product) and an unusual alcohol dehydrogenase from *Zymomonas mobilis* and *Saccharomyces cerevisiae*. J. Bacteriol. **171**:3754–3759.
15. **Conway, T., Y. A. Osman, J. I. Konnan, E. M. Hoffman, and L. O. Ingram.** 1987. Promoter and nucleotide sequences of the *Zymomonas mobilis* pyruvate decarboxylase. J. Bacteriol. **169**:949–954.
16. **Conway, T., G. W. Sewell, Y. A. Osman, and L. O. Ingram.** 1987. Cloning and sequencing of the alcohol dehydrogenase II gene from *Zymomonas mobilis*. J. Bacteriol. **169**:2591–2597.
17. **Cox, J. L., B. Cox, V. Fidanza, and D. Calhoun.** 1987. The complete nucleotide sequence of the *ilv*GMEDA cluster of *Escherichia coli* K-12. Gene **56**:185–198.
18. **Curtis, S. J., and W. Epstein.** 1975. Phosphorylation of D-glucose in *Escherichia coli* mutants defective in glucosephosphotransferase, mannosephosphotransferase, and glucokinase. J. Bacteriol. **122**:1189–1199.
19. **Davis, E. O., and P. Henderson.** 1987. The cloning and DNA sequence of the gene *xyl*E for xylose-proton symport in *Escherichia coli* K-12. J. Biol. Chem. **262**:13928–13932.
20. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
21. **DiMarco, A. A., and A. Romano.** 1985. D-Glucose transport system of *Zymomonas mobilis*. Appl. Environ. Microbiol. **49**:151–157.
22. **Doelle, H. W.** 1982. Kinetic characteristics and regulatory mechanisms of glucokinase and fructokinase from *Zymomonas mobilis*. Eur. J. Appl. Microbiol. Biotechnol. **14**:241–246.
23. **Eddy, C. K., J. Mejia, T. Conway, and L. O. Ingram.** 1989. Differential expression of *gap* and *pgk* genes within the *gap* operon of *Zymomonas mobilis*. J. Bacteriol. **171**:6549–6554.
24. **Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall.** 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. **179**:125–142.
25. **Entner, N., and M. Doudoroff.** 1952. Glucose and gluconic acid oxidation of *Pseudomonas saccharophila*. J. Biol. Chem. **196**:853–862.
26. **Fothergill-Gilmore, L. A.** 1986. The evolution of the glycolytic pathway. Trends Biochem. Sci. **11**:47–51.
27. **Fraenkel, D. G.** 1987. Glycolysis, pentose phosphate pathway, and Entner-Doudoroff pathway, p. 142–150. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
28. **Fukuda, Y., S. Yamaguchi, M. Shimosaka, K. Mutata, and A. Kimura.** 1983. Cloning of the glucokinase gene in *Escherichia coli* B. J. Bacteriol. **156**:922–925.
29. **Harris, H., and D. A. Hopkinson.** 1976. Handbook of enzyme electrophoresis in human genetics. American Elsevier Publishing Co., Inc., New York.
30. **Hylemon, P. B., N. R. Krieg, and P. V. Phibbs, Jr.** 1974. Transport and catabolism of D-fructose by *Spirillum itersonii*. J. Bacteriol. **117**:144–150.
31. **Hylemon, P. B., and P. V. Phibbs, Jr.** 1972. Independent regulation of hexose catabolizing enzymes and glucose transport activity in *Pseudomonas aeruginosa*. Biochem. Biophys. Res. Commun. **48**:1041–1048.
32. **Ingram, L. O., C. K. Eddy, K. Mackenzie, T. Conway, and F. Alterthum.** 1989. Genetics of *Zymomonas mobilis* and ethanol production. Dev. Ind. Microbiol. **30**:53–69.
33. **Kyte, J., and R. F. Doolittle.** 1982. A simple method of displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
34. **Layne, E.** 1957. Spectrophotometric and turbidometric methods for measuring proteins. Methods Enzymol. **3**:447–454.
35. **Lessie, T., and F. C. Neidhardt.** 1967. Adenosine triphosphate-linked control of *Pseudomonas aeruginosa* glucose-6-phosphate

dehydrogenase. J. Bacteriol. **93**:1337–1345.
36. **Luria, S. E., and M. Delbruck.** 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28**:491–511.
37. **Montenecourt, B. S.** 1985. *Zymomonas*, a unique genus of bacteria, p. 261–289. *In* A. L. Demain and N. A. Solomon (ed.), Biology of industrial microorganisms. The Benjamin Cummings Publishing Co., Inc., Menlo Park, Calif.
38. **Mueckler, M., C. Caruso, S. Baldwin, M. Panico, I. Blench, H. Morris, W. Allard, G. Lienhard, and H. Lodish.** 1985. Sequence and structure of a human glucose transporter. Science **229**:941–945.
39. **Osman, Y. A., T. Conway, S. J. Bonnetti, and L. O. Ingram.** 1987. Glycolytic flux in *Zymomonas mobilis*: enzyme and metabolite levels during batch fermentation. J. Bacteriol. **169**: 3726–3736.
40. **Persico, M. G., G. Vigiletto, G. Martini, D. Toniolo, G. Paonessa, C. Moscatelli, R. Dono, T. Vulliamy, L. Luzzatto, and M. D'Urso.** 1986. Isolation of human glucose-6-phosphate dehydrogenase (G6PD) cDNA clones: primary structure of the protein and unusual 5′ noncoding region. Nucleic Acids Res. **14**:2511–2522.
41. **Pette, D., W. Luh, and T. Bucher.** 1962. A constant-proportion group in the enzyme activity pattern of the Embden-Meyerhof chain. Biochem. Biophys. Res. Commun. **6**:419–424.
42. **Phibbs, P. V.** 1988. Genetic analysis of carbohydrate metabolism in *Pseudomonas*, p. 413–436. *In* S. R. Hagedorn, R. S. Hanson, and D. A. Kunz (ed.), Microbial metabolism and the carbon cycle. Harwood Academic Publishers, New York.
43. **Pond, J. L., K. Mackenzie, T. Conway, D. Borecky, and L. O. Ingram.** 1989. Cloning, sequencing and characterization of the principal acid phosphatase, the *phoC* product, from *Zymomonas mobilis*. J. Bacteriol. **171**:767–774.
44. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
45. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
46. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
47. **Scopes, R., and K. Griffiths-Smith.** 1984. Use of differential dye-ligand chromatography with affinity elution for enzyme purification: 6-phosphogluconate dehydratase from *Zymomonas mobilis*. Anal. Biochem. **136**:530–534.
48. **Scopes, R. K., V. Testolin, A. Stoter, K. Griffiths-Smith, and E. M. Algar.** 1985. Simultaneous purification and characterization of glucokinase, fructokinase and glucose-6-phosphate dehydrogenase from *Zymomonas mobilis*. Biochem. J. **228**:627–634.
49. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
50. **Stormo, G. D.** 1986. Translation initiation, p. 195–224. *In* W. Reznikoff and L. Gold (ed.), Maximizing gene expression. Butterworth Publishers, Stoneham, Mass.
51. **Swings, J., and J. De Ley.** 1977. The biology of *Zymomonas*. Bacteriol. Rev. **41**:1–46.
52. **Thomson, J., P. D. Gerstenberger, D. E. Goldberg, E. Gociar, A. O. deSilva, and D. G. Fraenkel.** 1979. ColE1 hybrid plasmids for *Escherichia coli* genes of glycolysis and the hexose monophosphate shunt. J. Bacteriol. **137**:502–506.
53. **Viikari, L.** 1988. Carbohydrate metabolism in *Zymomonas*. Crit. Rev. Biotechnol. **7**:237–261.
54. **Vinopal, R. T., J. D. Hillman, H. Schulman, W. S. Reznikoff, and D. G. Fraenkel.** 1975. New phosphoglucose isomerase mutants of *Escherichia coli*. J. Bacteriol. **122**:1172–1174.
55. **Wolf, R. E.** 1980. Integration of specialized transducing bacteriophage lambda *c*I857 *St*68 d*gnd his* by an unusual pathway promotes formation of deletions and generates a new translocatable element. J. Bacteriol. **142**:588–602.
56. **Wolf, R. E., D. M. Prather, and F. M. Shea.** 1979. Growth-rate-dependent alteration of 6-phosphogluconate dehydrogenase and glucose-6-phosphate dehydrogenase levels in *Escherichia coli* K-12. J. Bacteriol. **139**:1093–1096.
57. **Yager, T. D., and P. H. von Hippel.** 1987. Transcript elongation and termination in *Escherichia coli*, p. 1241–1275. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
58. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and M13mp19 vectors. Gene **33**:103–119.
59. **Yanofsky, C., and I. P. Crawford.** 1987. The tryptophan operon, p. 1453–1472. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.