JOURNAL OF BACTERIOLOGY, Feb. 1991, p. 968–977
0021-9193/91/030968-10$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 173, No. 3

Molecular Characterization of the *Escherichia coli* K-12 *zwf* Gene Encoding Glucose 6-Phosphate Dehydrogenase

DANIEL L. ROWLEY AND RICHARD E. WOLF, JR.*

Department of Biological Sciences, University of Maryland Baltimore County, Catonsville, Maryland 21228

Received 3 August 1990/Accepted 20 November 1990

In *Escherichia coli* K-12, expression of *zwf*, the gene for glucose 6-phosphate dehydrogenase, is coordinated with the cellular growth rate and induced by superoxide-generating agents. To initiate the study of the molecular mechanisms regulating its expression, the gene was cloned and its DNA sequence was determined. The 5′ ends of *zwf* mRNA isolated from cells growing in glucose and acetate minimal media were mapped. The map was complex in that transcripts mapped to −45, −52, and −62, with respect to the beginning of the coding sequence. Three analytical methods were used to search the DNA sequence for putative promoters. Only one sequence for a promoter recognized by the σ70 form of RNA polymerase was found by all three search routines that could be aligned with a mapped transcript, indicating that the other transcripts arise by processing of the mRNA. A computer-assisted search did not reveal a thermodynamically stable long-range mRNA secondary structure that is capable of sequestering the translation initiation region, which suggests that growth-rate-dependent regulation of glucose 6-phosphate dehydrogenase level may not be carried out by a mechanism similar to the one for the gene (*gnd*) for 6-phosphogluconate dehydrogenase. The DNA segment between the −10 hexamer and the start point of transcription resembles the discriminator sequence of stable RNA genes, which has been implicated in stringent control and growth-rate-dependent regulation.

The oxidative branch of the pentose phosphate pathway provides ribose for nucleoside biosynthesis and NADPH for reductive biosyntheses (15, 20). The two dehydrogenases of this pathway are glucose 6-phosphate dehydrogenase (G6PD; EC 1.1.1.49) and 6-phosphogluconate dehydrogenase (6PGD; EC 1.1.1.44). In *Escherichia coli*, the specific activities of these enzymes increase in proportion to growth rate during steady-state growth on different carbon sources (56). Although the growth rate dependence of the level of these two enzymes resembles that of the components of the translational apparatus, they are not part of the same regulatory network. After a nutritional shiftup, the accumulation rate of ribosomal components increases immediately, whereas the accumulation rate of G6PD and 6PGD has the same kinetics as that of total protein, i.e., increasing only after a lag (16). Thus, the mechanism(s) underlying the growth-rate-dependent regulation of *zwf* and *gnd*, which encode G6PD and 6PGD, respectively, may be common to other nonribosomal proteins whose levels also increase with increasing growth rate.

Studies of the control of *gnd* expression point to an interesting mechanism. Regulation is exerted at a posttranscriptional step and requires sequences within the 6PGD coding region (3, 4). The internal regulatory site, which lies between codons 71 and 74, is complementary to the translation initiation region on *gnd* mRNA (12). This internal complementary sequence (ICS) appears to function as a *cis*-acting antisense RNA by forming a long-range secondary structure that sequesters the ribosome binding site. A model for regulation has been proposed which suggests that the frequency of formation of this translation initiation-inhibiting structure is dependent on the concentration of free ribosomes (12).

As a way of determining whether this mechanism of growth rate control of nonribosomal genes is unique or global, we have initiated a molecular characterization of *zwf* expression. We report the cloning, nucleotide sequence, and transcript mapping of *zwf*. To facilitate the sequencing, we prepared a set of unidirectional deletions of an M13 phage carrying *zwf* that gave rise to in-frame *zwf-lacZα* fusions. In work to be reported elsewhere, we use the *zwf-lacZ* protein fusions as a means of searching genetically for a putative ICS in *zwf* and we go on to show that the mechanism for metabolic control of G6PD level is completely different from that of 6PGD.

MATERIALS AND METHODS

Chemicals and enzymes. Glucose-6-phosphate and isopropyl-β-D-thiogalactoside were purchased from Sigma, St. Louis, Mo., and 5-bromo-4-chloro-3-indolyl-β-D-galactoside was purchased from Boehringer Mannheim Biochemicals, Indianapolis, Ind. Restriction enzymes were purchased from Bethesda Research Laboratories, Gaithersburg, Md.; International Biotechnologies Inc., New Haven, Conn.; and New England BioLabs, Beverly, Mass. Phage T4 DNA ligase was from New England BioLabs. Phage T4 DNA polymerase was from either Amersham Corp., Arlington Heights, Ill., or Pharmacia P-L Biochemicals, Milwaukee, Wis. Reverse transcriptase was from Promega Biotec, Madison, Wis., and S1 nuclease was from Boehringer Mannheim Biochemicals. Deoxynucleoside triphosphates and [α-^{35}S]dATP were from Pharmacia P-L Biochemicals, and [α-^{35}S]dATP, [α-^{32}P]dATP, [α-^{32}P]dCTP, and [γ-^{32}P]ATP were from New England Nuclear, Boston, Mass. Sequencing kits were from New England BioLabs or Promega Biotec.

Media and growth conditions. Minimal medium 63, glucose tetrazolium agar medium, and other standard rich media were prepared as described by Miller (40). Bromthymol blue agar medium supplemented with 1% glucose or gluconate was prepared by the method of Wolf and Shea (57). Antibiotics were used at the following concentrations: ampicillin (Ap), 50 mg/liter, and tetracycline (Tc), 25 mg/liter.

* Corresponding author.

TABLE 1. Bacterial strains

Strain	Genotype and phenotype[a]	Source
DF100	HfrC *tonA22 garB10 ompF627 relA1 pit-10 spoT1* T2[r]	24
DR500	RW231(F[+] pLC3-33)	This study
DR612	HB351 Δ(*zeb*::Tn*10*) *pgi*::ΔTn*10* (Pgi[−])	This study
GB1815	*supF* Δ(*argF-lac*)U169 Δ(*edd-zwf*)22 Δ(*sbcB-his-gnd-rfb*) *trpR trpA9605 kdgR proC*[+]	7
HB351	Δ(*argF-lac*)*U169 zeb-1*::Tn*10* Δ(*edd-zwf*)22	3
JA200	Δ*trpE5 recA thr leu lacY* F[+]	14
JM109	*recA endA gyrA96 thi hsdR17 supE44 relA1* Δ(*lac-proAB*)(F′ *traD36 proAB*[+] *lacI*[q] *lacZ*Δ*M15*)	Lab stock
KK2186	Δ(*lac-proAB*) *spuE thi endA sbcB15 strA hsdR4*/F′ *proAB*[+] *lacI*[q] *lacZ*Δ*M15*	Lab stock
NF3079	*dam-3 araD139* Δ(*araAB01C-leu*)*7679 rpsL galK galU* Δ(*lac*)*X74*	Lab stock
RW231	*trpR kdgR lacZ*(Am) *trpA9605* Δ(*edd-zwf*)22 Δ(*sbcB-his-gnd-rfb*) *recA rpsL20*	55
SV107	*pgi*::Tn*10 araD139* Δ(*argF-lac*)*U169 flbB pstF relA rpsL thi deoC*	T. Silhavy

[a] All strains are *E. coli* K-12 derivatives. Strain designations are according to Bachmann (2).

Bacterial strains and scoring of *zwf* and *edd*. Table 1 shows the strains of *E. coli* K-12 used in this study. Genetic methods were standard procedures (40).

In *E. coli*, glucose 6-phosphate can be metabolized by phosphoglucose isomerase (encoded by *pgi*) and the Embden-Meyerhoff pathway or by G6PD and the pentose phosphate pathway (25). Since Zwf[−] mutants of an otherwise wild-type strain have no readily scorable phenotype (19), the presence of the *zwf* gene was determined in *pgi* mutants. Pgi was scored on glucose tetrazolium agar, on which Pgi[+] colonies are white and Pgi[−] colonies are red (23). Zwf was scored on glucose-bromthymol blue agar, on which a Zwf[+] Pgi[−] strain forms yellow colonies, while a Pgi[−] Zwf[−] colony is white. Similarly, Edd[+] (6-phosphogluconate dehydrase) strains were distinguished from Edd[−] mutants on gluconate-bromthymol blue plates (55).

To facilitate selection and scoring of plasmids carrying the *zwf-edd* region of the chromosome, we prepared strain DR612, which carries *pgi* and *zwf-edd* deletions. First, we selected a tetracycline-sensitive derivative of the Δ(*edd-zwf*) strain HB351 by the method described by Maloy and Nunn (37). Next, a *pgi* mutation was introduced by generalized transduction with a phage P1 lysate grown on the *pgi*::Tn*10* strain DR110 and selection for Tc[r] transductants; the desired derivative was identified by streaking clones on glucose tetrazolium indicator plates and by enzyme assay for phosphoglucose isomerase deficiency (25). Finally, a tetracycline-sensitive derivative was isolated by another round of selection for fusaric acid resistance (37). Strain DF612 does not revert to growth on glucose.

To transfer plasmid pLC3-33 from the *zwf*[+] *gnd*[+] *edd*[+] background of strain JA200 in which the Clarke-Carbon library was prepared (14), the plasmid-carrying strain was mated with the Δ(*zwf-edd*) Δ*gnd* strain RW231. Edd[+] transconjugants were selected on appropriately supplemented gluconate minimal agar, with the absence of leucine and threonine selecting against the donor. The presence of *zwf* was verified by assaying for G6PD activity. Subsequently, the presence of *zwf* or *edd* on plasmids was determined by complementation in strain DF612 and scoring for growth on glucose or gluconate, respectively.

Plasmids and phages. Table 2 shows the properties of the plasmids and M13 phages used in this study.

Assay and purification of G6PD. The activity of G6PD in sonic extracts was assayed spectrophotometrically as described by Wolf et al. (56); 1 enzyme unit was equivalent to 1 nmol of NADPH formed per min at 25°C per mg of protein. Protein concentration was determined by the method of Bradford (9), using immunoglobulin G as the standard. G6PD was purified as follows by the method of Wolf and Shea (57). A sonic extract of strain RW231(pDR17) was bound to a column of blue dextran-Sepharose; the column was washed with sonication buffer (10 mM Tris hydrochloride [pH 8.0], 1 mM dithiothreitol), and G6PD was eluted with 1 mM NADP in sonication buffer.

Recombinant DNA and M13 methods. Standard methods were used as described previously (1, 38), except for the following modifications. Plasmids were prepared by the alkaline lysis method modified for removal of chromosomal DNA by the addition of a phenol extraction at pH 4.0 (58). Strain DR612 was made competent for transformation by the low-pH procedure in the P-L Biochemicals nucleotide sequencing manual. Restriction digestions with *Bcl*I were done on DNA prepared from strain NF3079.

Single-stranded DNA of M13 phages was isolated as described by Messing (39). The replicative form of phage M13 was isolated as described above by the pH 4.0 phenol method, which also removes contaminating single-stranded DNA.

Southern hybridizations. Chromosomal DNA was isolated from strain DF100 as described before (50) and digested with various restriction enzymes. Cleaved DNA was electrophoresed through a 1% agarose gel, transferred to a Gene Screen Plus membrane (New England Nuclear), and hybridized with the 1.7-kb *Sal*I fragment labeled by the random priming method (17).

Preparation of unidirectional deletions. The replicative form of phage mDR11a, a derivative of M13mp18, was cleaved at the unique *Hin*dIII restriction site. The recessed 3′ ends were made flush with the Klenow fragment of *E. coli* DNA polymerase I and a mixture containing dGTP, dCTP, TTP, and 2′-deoxyadenosine-*O*-(1-thiotriphosphate); then the DNA was digested with *Pst*I, which cleaves in the multiple cloning site between the *Hin*dIII site and the target DNA. Unidirectional deletions were prepared as described previously (6), and the DNA was used to transfect strain JM109 or KK2186. To enrich for phages that carried deletions and to obtain *zwf-lacZ*α translational fusions at the same time, only phages with LacZα-complementing activity were chosen for further study (6). Deletion mutants were characterized by agarose gel electrophoresis with phages M13mp18 and mDR11a as size markers.

DNA sequence determination. DNA sequencing was done by the chain termination method of Sanger et al. (46) with the modifications of Biggin et al. (8). The primer for sequencing across the fusion joints of the 3′ deletions has the sequence CGGGCCTCTTCGCTA and hybridizes to codons 39 to 44 of *lacZ*α.

TABLE 2. Plasmids and M13 bacteriophages

Plasmid or phage	Relevant phenotype[a]	Source or reference
Plasmids		
pBR322		Lab stock
pZ152	Phage M13 ori at *Aha*III site in pBR322	R. Zagursky
pUC18		54
pLC3-33	≈16-kb fragment of *E. coli* chromosome containing *zwf* and *edd* genes in plasmid ColE1	14, 52
pDR2	6.4-kb *Pst*I fragment from pLC3-33 in pZ152	This study
pDR3	3.0-kb *Sal*I fragment from pDR2 in pZ152	This study
pDR4	pDR2 with 1.7-kb *Sal*I fragment deleted	This study
pDR5	1.7-kb *Sal*I fragment from pDR2 in pZ152	This study
pDR9	2.8-kb *Sst*I-*Kpn*I fragment from pDR2 in pUC18	This study
pDR11	1.7-kb *Sal*I fragment from pDR2 in pUC18	This study
pDR16	2.0-kb *Pvu*II fragment from pDR9 in *Sma*I site of pUC18	This study
pDR17	1.9-kb *Eco*RI(*Sst*I)-*Bam*HI(*Pvu*II *Sma*I) fragment from pDR16 in *Eco*RI-*Bam*HI-cut pBR322	This study
Phages		
M13mp18, 19		Lab stock
mDR9	2.8-kb *Kpn*I-*Sst*I from pDR2 in M13mp18 (antisense)	This study
mDR11a	1.7-kb *Sal*I fragment from mDR9 in M13mp18 (sense)	This study
mDR11b	1.7-kb *Sal*I fragment from mDR9 in M13mp18 (antisense)	This study
mDR12	0.7-kb *Sal*I-*Sst*I fragment from MDR9 in M13mp18 (antisense) (Δ*Sal*I)	This study
mDR14a	1.0-kb *Hin*cII-*Sal*I fragment from mDR11a in *Hin*cII site of M13mp18 (sense)	This study
mDR14b	1.0-kb *Hin*cII-*Sal*I fragment from mDR14 in *Hin*cII site of M13mp19 (antisense)	This study
mDR15a	0.7-kb *Sal*I-*Hin*cII fragment from mDR11a in *Hin*cII site of M13mp18 (sense)	This study
mDR15b	0.7-kb *Sal*I-*Hin*cII fragment from mDR11a in *Sma*I site of M13mp18 (sense)	This study
mDR15c	0.7-kb *Sal*I-*Hin*cII fragment from mDR11a in *Hin*cII site of M13mp18 (antisense)	This study
mDR16	2.0-kb *Pvu*II fragment from pDR9 in *Sma*I site of M13mp18 (antisense)	This study
mDR17a	1.9-kb *Eco*RI(*Sst*I)-*Bam*HI(*Pvu*II) fragment from pDR16 in *Sst*I-*Bam*HI site of M13mp18 (antisense)	This study
mDR17b	1.9-kb *Eco*RI(*Sst*I)-*Bam*HI(*Pvu*II) fragment from pDR16 in *Sst*I-*Bam*HI site of M13mp19 (sense)	This study
mDR26a	0.7-kb *Bcl*I fragment from pDR17 in *Bam*HI site of M13mp18 (antisense)	This study
mDR101	Φ(*zwf'*-*'lacZ*α$^+$)*19*(Hyb) deletion of mDR11a resulting in codon 19 of *zwf* fused in-frame to *lacZ*α	This study
mDR102	Φ(*zwf'*-*'lacZ*α$^+$)*302*(Hyb) deletion of mDR11a resulting in codon 302 of *zwf* fused in-frame to *lacZ*α	This study
mDR104	Φ(*zwf'*-*'lacZ*α$^+$)*215*(Hyb) deletion of mDR11a resulting in codon 215 of *zwf* fused in-frame to *lacZ*α	This study
mDR106	Φ(*zwf'*-*'lacZ*α$^+$)*136*(Hyb) deletion of mDR11a resulting in codon 136 of *zwf* fused in-frame to *lacZ*α	This study
mDR111	Φ(*zwf'*-*'lacZ*α$^+$)*115*(Hyb) deletion of mDR11a resulting in codon 115 of *zwf* fused in-frame to *lacZ*α	This study
mDR114	Φ(*zwf'*-*'lacZ*α$^+$)*263*(Hyb) deletion of mDR11a resulting in codon 263 of *zwf* fused in-frame to *lacZ*α	This study
mDR161	Φ(*zwf'*-*'lacZ*α$^+$)*76*(Hyb) deletion of mDR11a resulting in codon 76 of *zwf* fused in-frame to *lacZ*α	This study
mDR165	Φ(*zwf'*-*'lacZ*α$^+$)*125*(Hyb) deletion of mDR11a resulting in codon 125 of *zwf* fused in-frame to *lacZ*α	This study

[a] Restriction enzyme in parentheses indicates that the site is nearby in the *zwf* gene.

Transcript mapping. RNA was isolated by a variation of the hot phenol method (10). A 35-ml amount of a log-phase culture was poured over 15 ml of crushed frozen killing buffer (300 mM Tris [pH 7.5], 30 mM sodium azide, 5 mM magnesium chloride, 300 μg of chloramphenicol per ml) and centrifuged at 1,000 × *g* for 20 min at 4°C. The pellet was resuspended in 3 ml of extraction buffer (40 mM sodium acetate [pH 4.0], 150 mM sodium chloride, 0.5% sodium dodecyl sulfate [SDS], 10 mM sodium azide, 1 mM EDTA) and placed in a boiling water bath for 1 min. Two extractions were carried out with pH 4.0 phenol at 60°C. After a third extraction at 60°C with a mixture of phenol, chloroform, and isoamyl alcohol (25:24:1), the aqueous phase was precipitated twice with ethanol and resuspended in diethylpyrocarbonate-treated water. The quality of the RNA preparations was assessed by electrophoresis through a 1.5% agarose formaldehyde gel and subsequent staining with ethidium bromide (48). Only preparations with unsmeared rRNA bands were used for transcript mapping.

Transcript mapping by the primer extension and S1 nuclease protection methods was done as described previously (1). The oligonucleotide for the primer extension assays was labeled at its 5′ end with phage T4 polynucleotide kinase and [γ-^{32}P]ATP. The sequence of the oligonucleotide was GGCT GTTTGCGTTACCGCC, which is complementary to codons 1 to 7 of *zwf* mRNA. The probe for S1 nuclease mapping was prepared by annealing the same 5′-end-labeled oligonucleotide to DNA from phage mDR11a, synthesizing DNA com-

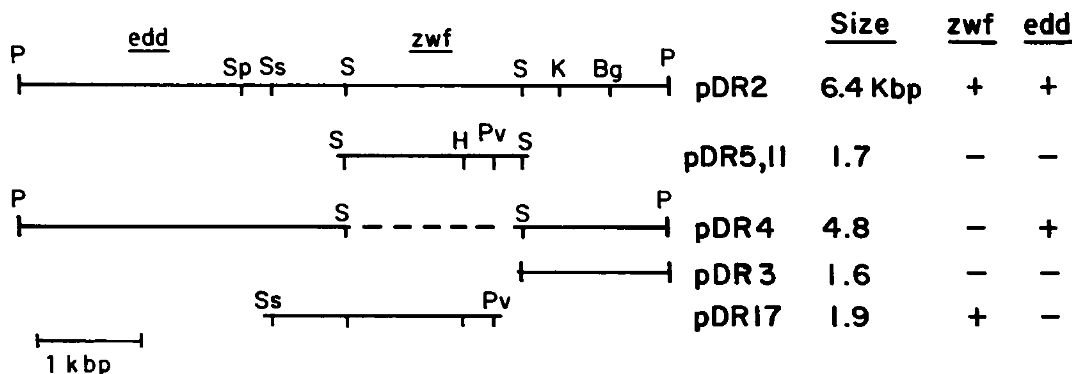

FIG. 1. Restriction maps and complementation analysis of plasmids carrying the *E. coli zwf* and *edd* genes. Plasmids were constructed as described in the text and introduced into strain DR612. The ability of the plasmids to complement the *zwf* and *edd* deletion mutations was scored by growth of the strains on glucose and gluconate indicator plates, respectively, as described in Materials and Methods. Symbol: ------, deletion. Bg, *Bgl*II; H, *Hinc*II; K, *Kpn*I; P, *Pst*I; Pv, *Pvu*II; S, *Sal*I; Sp, *Sph*I; Ss, *Sst*I.

plementary to the template strand in a reaction mixture containing Klenow fragment and deoxynucleoside triphosphates, digesting the DNA with *Eco*RI, which cleaves in the multiple cloning site at the distal end of the cloned DNA, and isolating the labeled DNA strand after electrophoresis through an alkaline agarose gel (38).

Sequence analysis. Three methods were used to search for promoterlike sequences recognized by the σ70 form of RNA polymerase. One was a computer program developed by O'Neill (43), which is based on an analysis of the information content of a set of known *E. coli* promoters subdivided into classes based on the number of base pairs between the −35 and −10 hexamers. A second computer program used a neural network trained by M. C. O'Neill on 39 known promoters and more than 4,000 non-promoter-containing sequences. The third method used the −10 and −35 sequences of a set of functional promoters isolated by Oliphant and Struhl (42).

Potential secondary structures of *zwf* mRNA were identified by the computer program of Zucker and Stiegler (59) using the energy values of Freier et al. (26).

Nucleotide sequence accession number. The accession number for the *E. coli zwf* gene is M55005.

RESULTS

Cloning of *zwf* and correlation of the genetic and restriction maps. In *E. coli*, the *zwf* gene is closely linked to two genes for the inducible Entner-Doudoroff pathway of gluconate metabolism, *edd* and *eda* (22). Thomson et al. (52) screened the Clarke-Carbon library for plasmids able to complement mutations in central metabolism genes and identified plasmid pLC3-33 as containing *edd*. The specific activity of G6PD was 10- to 15-fold higher in the plasmid-containing strain than in the wild-type strain, suggesting that the plasmid also carries *zwf*. To confirm these results, we mated the Clarke-Carbon clone carrying pLC3-33 with strain RW231, which carries an *edd-zwf* deletion, and selected Edd^+ transconjugants. They expressed elevated amounts of G6PD, which indicates that the plasmid carries *zwf* as well as *edd*.

Preliminary restriction digestions indicated that plasmid pLC3-33 contains about 16 kb of cloned DNA. A *Pst*I digestion of the plasmid was ligated with similarly cleaved plasmid pZ152, the mixture was transformed into strain DR612, which is unable to grow on glucose (Glu), and Glu^+ transformants were selected. Plasmid pDR2, isolated from one of the transformants, contained a single cloned *Pst*I fragment of 6.4 kb. The specific activity of G6PD in strain DR612(pDR2) was about 15 times that of a normal haploid strain. The ability to complement the *zwf* deletion mutation and to confer high-level expression of G6PD indicated that the entire *zwf* gene had been cloned.

To localize *zwf*, additional restriction mapping of pDR2 was carried out, selected fragments were subcloned into vector plasmids, and the resulting recombinant plasmids were tested for their ability to complement the *edd* and *zwf* mutations. Figure 1 shows that the 1.9-kb *Pvu*II-*Sst*I fragment of plasmid pDR17 contains the entire *zwf* gene, that the *Sal*I site of this fragment is within the gene, and that *edd* lies to the left of this *Sal*I site as the map is drawn. Based on genetic mapping that placed putative "up" promoter mutations of *zwf* on the *edd*-distal side of the gene (24), we inferred that *zwf* transcription is from the *Pvu*II side of the gene toward the *Sst*I site. As further evidence that *zwf* was present on pDR17, G6PD was purified in one step from a crude extract of strain RW231(pDR17) by the affinity chromatography method used previously for purification of 6PGD (57). The specific activity of the purest fraction, which was about 80% pure, was 31,500 U/mg, and the molecular weight of the enzyme in the purified preparation was estimated by SDS-polyacrylamide gel electrophoresis to be 52,000. Both values agree with those for G6PD purified by other methods (5). Determination of the N-terminal amino acid sequence was attempted by automated methods, but no amino acid residues were recovered, indicating that the N terminus was blocked.

To confirm the *E. coli* origin of the cloned DNA, the 1.7-kb *Sal*I fragment (Fig. 1) was labeled and used in Southern analysis (51) of chromosomal DNA isolated from strain DF100 and digested singly with *Sal*I or *Pvu*II, or doubly with *Pvu*II and *Hinc*II, or with *Pst*I and *Sst*I. In all cases, the size of the hybridizable restriction fragments was as expected from the restriction map of plasmid pDR2 (data not shown).

Construction of 3′ deletions for nucleotide sequencing and genetic studies. In addition to determining the nucleotide sequence of *zwf*, we wanted to determine genetically whether *zwf* contains a regulatory site within the coding sequence similar to that of *gnd* (4, 12). These two objectives were addressed simultaneously by the unidirectional deletion mutagenesis method of Barcak and Wolf (6). They had

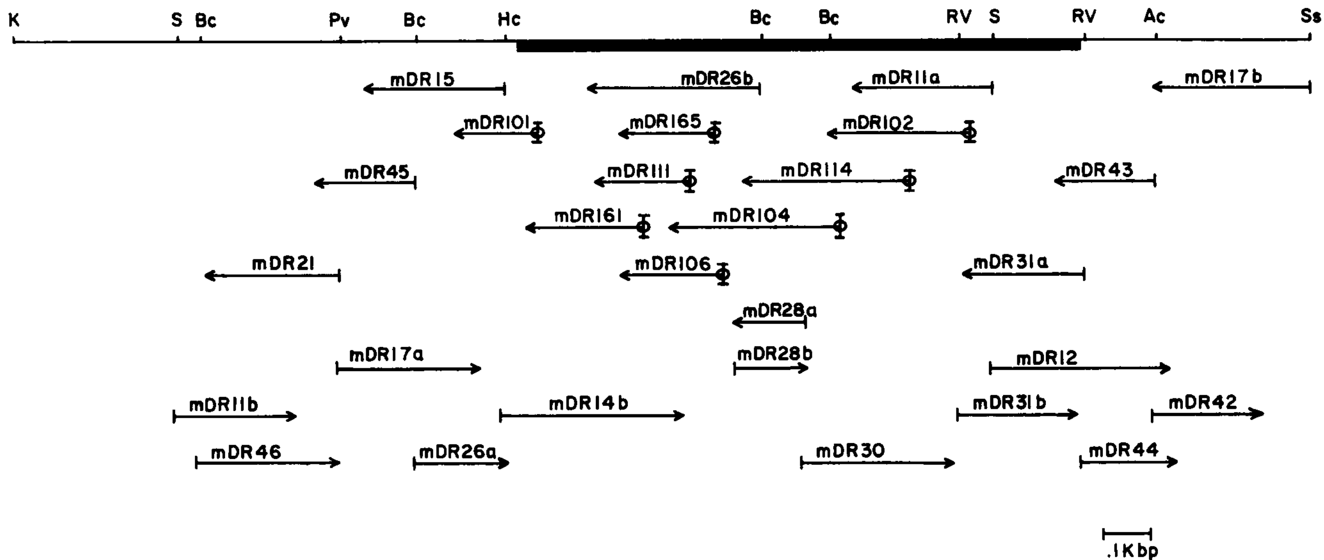

FIG. 2. Fusion joints of in-frame *zwf-lacZα* protein fusions and strategy for sequencing both strands of the *zwf* gene. The upper line shows a restriction map of the *zwf* region, with the coding sequence for G6PD indicated by the dark line. Arrows beneath the map represent the direction and extent of sequencing on the respective M13 phage templates. Deletions extending from the *Sal*I site in the G6PD coding sequence toward the promoter region were prepared in M13 phage mDR11a by a unidirectional deletion mutagenesis method (7), and phages with in-frame *zwf-lacZα* protein fusions were identified by their LacZα-complementing activity. The direction of *zwf* transcription is from left to right as the map is drawn. Symbol: Φ, phages with in-frame *zwf-lacZα* protein fusions. K, *Kpn*I; S, *Sal*I; Bc, *Bcl*I; Pv, *Pvu*II; Hc, *Hin*cII; RV, *Eco*RV; Ac, *Acc*I; Ss, *Sst*I.

reported that, when a gene is cloned into an M13 phage in the same orientation as *lacZα*, deletions entering the distal end of the gene and extending toward the promoter can produce in-frame fusions to *lacZα*. Since the parental recombinant phage does not provide LacZα-complementing activity, deletion clones can be identified directly by their ability to confer this activity. Thus, they suggested that deletion clones obtained by this method would facilitate sequencing of the target gene as well as providing a set of protein fusions for subsequent genetic studies. This approach was particularly useful for analysis of *zwf*, since a putative internal regulatory region could be defined by differences in the growth rate dependence of β-galactosidase activity in strains carrying the various fusions, as was done previously for *gnd* (4).

Unidirectional deletions extending from the *Sal*I site in *zwf* (Fig. 1) toward the promoter were prepared in phage mDR11a as described in Materials and Methods. About 10% of the clones recovered after mutagenesis conferred LacZα-complementing activity. About 90 of these $LacZ^+$ clones were characterized by agarose gel electrophoresis, and 8 with deletion endpoints spaced at relatively equal intervals across the region (mDR101, mDR102, mDR104, mDR106, mDR111, mDR114, mDR161, and mDR165) were chosen for sequencing and for subsequent genetic analysis.

Nucleotide sequence of *zwf* and derived amino acid sequence of G6PD. The M13 phages described above and others carrying selected restriction fragments were used as templates for DNA sequencing of the *zwf* region (Fig. 2). More than 90% of the sequence was determined for both DNA strands, with the remainder being verified by sequencing several overlapping in-frame fusion clones.

Figure 3 shows the DNA sequence of *zwf* and surrounding DNA and the derived amino acid sequence of G6PD. The *zwf* structural gene contains 490 codons. Since the N-terminal amino acid sequence of the purified enzyme could not be determined (see above), the beginning of the G6PD coding sequence was initially identified by a landmark, a Shine-Dalgarno sequence, AAGGAG, which is perfectly complementary to the 3′ end of 16S rRNA (49) and spaced an optimal six nucleotides upstream from an ATG start codon (27). The deduced molecular mass of the polypeptide encoded by this open reading frame (ORF) is 55,651 Da, which is close to the molecular weight for purified preparations of the *E. coli* K-12 enzyme (5; and above). The 1,473 bp of the *zwf* gene ends with a TAA stop codon. The codon usage pattern of the gene suggests that it is a weakly expressed gene, since it contains a relatively high percentage (9.7%) of rare codons as defined by Sharp and Li (47).

The primary amino acid sequence of the *E. coli* G6PD was aligned by computer to that of the G6PD of the gram-negative bacterium *Zymomonas mobilis* (7a). The sequences contain 44% identical amino acids, and 62% of the residues are chemically conserved. The 14 amino acid residues at positions 171 to 184 of the *E. coli* G6PD are perfectly conserved between the enzymes of these two species; within this segment is a stretch of 8 residues that is also perfectly conserved in the human (44), rat (30, 33, 36), and *Drosophila* (18) G6PDs. Modification of the lysine residue corresponding to position 181 in the *E. coli* enzyme has been shown to block binding of glucose 6-phosphate to the human and yeast enzymes (11, 32). The presence of the binding site for glucose 6-phosphate further confirms that the sequence encodes G6PD.

The amino acid sequence comparisons also provided evidence that the assignment made for the beginning of the coding sequence for the *E. coli* enzyme is probably correct. The assigned start codon is the first potential start codon upstream from an amino acid sequence that is conserved in another G6PD; specifically, the glycine-aspartic acid-leucine

peptide at positions 19 to 22 matches the residues at positions 16 to 18 of the *Z. mobilis* enzyme (data not shown).

The coding sequence of the *E. coli zwf* gene was used as a query sequence in a search of the GenBank data base. The search did not reveal any significant matches other than those mentioned above. Moreover, a direct comparison of the primary amino acid sequences of the *E. coli* G6PD and 6PGD enzymes yielded no striking similarities, even though both oxidize hexose phosphate substrates and use NADP as the coenzyme.

The 707-bp region upstream of the G6PD coding sequence contains only one significant ORF; it extends from −535 to −230. Although there are several potential start codons within this 101-codon ORF, none are preceded by a very strong Shine-Dalgarno sequence. Moreover, the full-length ORF as well as the shorter internal ORFs contains >30% rare codons. Thus, in agreement with the fact that the deletion endpoints of all 3′ deletions of mDR11a that conferred LacZα-complementing activity were within the *zwf* gene, this ORF is probably not expressed.

The 150-bp region downstream of *zwf* does not contain an ORF with a potential start codon, so *edd*, which is >95% cotransducible with *zwf* (24), probably lies further away. Furthermore, the region contains a sequence capable of forming in the mRNA a GC-rich stem-loop structure followed by a run of U residues, a putative factor-independent transcription terminator (45). These data, along with the transcript map presented below, suggest that *zwf*, like *gnd*, is monocistronic (41).

High-resolution mapping of the 5′ ends of *zwf* mRNA. RNA was isolated from the wild-type *E. coli* K-12 strain DF100 growing in glucose MOPS (morpholinepropanesulfonic acid) minimal medium, and the 5′ ends of *zwf* mRNA were mapped by both primer extension and S1 nuclease protection methods as described in Materials and Methods. The transcript maps were complex in that multiple 5′ ends were revealed by both methods. With primer extension, 5′ ends mapped to −62, −60, and several positions surrounding −52 and −45 (Fig. 4); in addition, minor transcripts were observed that mapped to the region beyond −150 (data not shown). Transcript families mapping to the −52 and −45 regions were also identified by S1 mapping (Fig. 4), as were minor transcripts mapping further upstream. Since the −52 and −45 transcripts were identified by both methods, we presume that they represent bona fide in vivo transcript species.

Noting that the sequence immediately downstream from −62 is AT rich, we considered the possibility that the basis for our inability to detect the −60/−62 transcripts by S1 nuclease protection was the propensity of the enzyme to nibble the ends of AT-rich DNA-RNA hybrids (1, 41). Accordingly, we mapped with both methods RNA isolated from strain GB1815 carrying plasmid pDR17 wherein the *zwf* mRNA fraction of total RNA should be considerably larger. In this case, 5′ ends mapping to −62 and −60 were revealed by the S1 nuclease technique as well as by primer extension, indicating that they too represent bona fide *zwf* transcripts (Fig. 4). In work to be reported elsewhere, we used the primer extension method to map the transcripts from a strain carrying a *zwf-lacZ* fusion whose *zwf* DNA extends upstream only to the *Bcl*I site at −202; the −62, −60, −52, and −45 transcript species were also present in this strain, which indicates that they originate from a promoter(s) downstream of this site (36a).

We also mapped by both methods the *zwf* transcripts in RNA prepared from strain DF100 growing in acetate minimal medium. The same species were found as when the strain was grown on glucose, and the ratios of the relative amounts of the various transcripts to one another were approximately the same in the RNAs prepared from the two growth conditions (data not shown).

Identification of the *zwf* promoter by computer-assisted promoter-search algorithms. The complexity of the map of the 5′ ends of *zwf* mRNA raised the possibilities that transcription initiates from multiple promoters, which might be recognized by RNA polymerases with different sigma factors, or that the mRNA leader is extensively processed, or both. As an initial approach to distinguish between these possibilities, we analyzed the DNA sequence upstream of the structural gene for promoterlike sequences recognized by the σ70 form of RNA polymerase. We used a computer search based on total information content (43) and a back-propagation neural network trained on the 17-bp spacing class of *E. coli* promoters (43a). We also visually inspected the sequence for matches to the consensus sequence matrix defined by random selection of functional σ70 promoters (42). Although each analytical approach found more than one promoterlike sequence, the only promoter common to all three methods that aligned with a mapped transcript species was the sequence TTTTCG. .N17. .TACAGT at positions −96 to −68. This promoter would initiate transcription at the purines located at −60 and −62. The neural network assigned a value to this promoter of >0.98 on a scale of 0.00 to 1.00, strongly predicting that it is a functional promoter, and the ranking function of O'Neill (43) placed this promoter at a moderate strength level. Aside from the caveat that the mapping methods used here cannot distinguish between primary transcription products and processed derivatives, we also note the following observations and additional points: (i) no promoterlike sequences were found by any of the methods to align with the −45 and −52 transcript families and thus they probably result from processing; (ii) no other promoterlike sequences were found by all three methods within the region bounded by the start codon and the *Pvu*II site at −370; and (iii) the region between the −10 hexamer and the start site of transcription contains the GC-rich sequence GCACC that matches four of five residues of the GCGCC discriminator sequence of the rRNA P1 and *tyrT* promoters (53), a sequence that has been associated with growth rate regulation and stringent control (reviewed in reference 13).

Computer-assisted search for potential long-range mRNA secondary structures sequestering the translation initiation region. To see whether an mRNA segment like the ICS of the *gnd* gene is present in *zwf*, we used the RNA folding algorithm of Zucker and Stiegler (59) to examine the secondary structure of *zwf* mRNA molecules, increasing in length by 10 nucleotides starting at −62 and extending to +328, with the total interval being the largest mRNA segment that can be folded by this computer program. No such structure was found. We also searched the entire structural gene for sequences complementary to the Shine-Dalgarno region of the *zwf* leader. The sequence TCTCC at positions +748 to +752 was the best match to this segment of the mRNA. Whether this structural gene segment or any other is involved in metabolic control of *zwf* expression will await genetic analysis, e.g., using the *zwf-lacZ* protein fusions described here.

We also examined in smaller increments the folded structure of mRNA segments containing the leader and initial codons. Unlike *gnd*, no segments within the leader were found that sequester the ribosome binding site (41). How-

-700 -690 -680 -670 -660 -650 -640 -630 -620 -610 -600
GTCGACGG CGCGGTTGAT GGCAGATTTA TCCAGTGAAT GACGGACATG ATCAAGCGTT GCCATTGCGG ACTCAAATAT TTTCCCTGTG TATGATTCAA CGCTGTCATC TTCATTGACA

-590 -580 -570 -560 -550 -540 -530 -520 -510 -500 -490 -480
TTGCGATTAA CATAGGGAGT GCCATTCGCC AGACTCTGTG CCAGATGAAG TTTAAAATCA GGAAAACCGC GCTGTCCATG CTGCGACAGA AACGATTCAC CGTCGGTTCG CTAACATTGG

-470 -460 -450 -440 -430 -420 -410 -400 -390 -380 -370 -360
CTTCCAGTGC CATAGCAGCA ATACTCGAAT GGATCGCGTT ATCGGGCGAA CGAGAATGAC CTCGGCAACT TTGCGCTCTG ATTTCGCTCA AATGTTCCAG CTGAGACTGG ATTTTTTCCA

-350 -340 -330 -320 -310 -300 -290 -280 -270 -260 -250 -240
GCATATTCAT GATGTAAAGA GACTCACACG GGTAATGACG AATTTCCGCA CTGAAAGAAA TCGAAATGCA GTTTTGTCAG ATATTACGCC TGTGTGCCGT GTTAATGACA AAAGCAGATA

-230 -220 -210 -200 -190 -180 -170 -160 -150 -140 -130 -120
AAAAAGTTGT TATTTTTTTT CATAACATGA TCAGTGTCAG ATTTTTACCC AATGGAAAAC GATGATTTTT TTATCAGTTT TGCCGCACTT TGCGCGCTTT TCCCGTAATC GCACGGGTGG

-110 -100 -90 -80 -70 • • •• -50 ••• -40 -30 -20 -10
ATAAGCGTTT ACAGTTTTCG CAAGCTCGTA AAAGCAGTAC AGTGCACCGT AAGAAAATTA CAAGTATACC CTGGCTTAAG TACCGGGTTA GTTAACTTAA GGAGAATGAC ATG GCG
Met Val 2

30 45 ⧫101 60 75 90
GTA ACG CAA ACA GCC CAG GCC TGT GAC CTG GTC ATT TTC GGC GCG AAA GGC GAC CTT GCG CGT CGT AAA TTG CTG CCT TCC CTG TAT CAA CTG GAA
Val Thr Gln Thr Ala Gln Ala Cys Asp Leu Val Ile Phe Gly Ala Lys Gly Asp Leu Ala Arg Arg Lys Leu Leu Pro Ser Leu Tyr Gln Leu Glu 34

105 120 135 150 165 180 195
AAA GCC GGT CAG CTC AAC CCG GAC ACC CGG ATT ATC GGC GTA GGG CGT GCT GAC TGG GAT AAA GCG GCA TAT ACC AAA GTT GTC CGC GAG GCG CTC
Lys Ala Gly Gln Leu Asn Pro Asp Thr Arg Ile Ile Gly Val Gly Arg Ala Asp Trp Asp Lys Ala Ala Tyr Thr Lys Val Val Arg Glu Ala Leu 66

210 225 ⧫161 240 255 270 285
ACT TTC ATG AAA GAA ACC ATT GAT GAA GGT TTA TGG GAC ACC CTG AGT GCA CGT CTG GAT TTT TGT AAT CTC GAT GTC AAT GAC ACT GCT GCA GAA
Glu Thr Phe Met Lys Glu Thr Ile Asp Glu Gly Leu Trp Asp Thr Leu Ser Ala Arg Leu Asp Phe Cys Asn Leu Asp Val Asn Asp Thr Ala Ala 98

300 315 330 ⧫111 345 360 ⧫165 390
TTC AGC CGT CTC GGC GCG ATG CTG GAT CAA AAA AAT CGT ATC ACC ATT AAC TAC TTT GCC ATG CCG CCC AGC ACT TTT GGC GCA ATT TGC AAA GGG
Phe Ser Arg Leu Gly Ala Met Leu Asp Gln Lys Asn Arg Ile Thr Ile Asn Tyr Phe Ala Met Pro Pro Ser Thr Phe Gly Ala Ile Cys Lys Gly 130

405 ⧫106 420 435 450 465 480
CTT GGC GAG GCA AAA CTG AAT GCT AAA CCG GCA CGC GTA GTC ATG GAG AAA CCG CTG GGG ACG TCG CTG GCG ACC TCG CAG GAA ATC AAT GAT CAG
Leu Gly Glu Ala Lys Leu Asn Ala Lys Pro Ala Arg Val Val Met Glu Lys Pro Leu Gly Thr Ser Leu Ala Thr Ser Gln Glu Ile Asn Asp Gln 162

495 510 525 540 555 570
GTT GGC GAA TAC TTC GAG GAG TGC CAG GTT TAC CGT ATC GAC CAC TAT CTT GGT AAA GAA ACG GTG CTG AAC CTG TTG GCG CTG CGT TTT GCT AAC
Val Gly Glu Tyr Phe Glu Glu Cys Gln Val Tyr Arg Ile Asp His Tyr Leu Gly Lys Glu Thr Val Leu Asn Leu Leu-Ala Leu Arg Phe Ala Asn 194

585 600 615 630 ⧫104 660 675
TCC CTG TTT GTG AAT AAC TGG GAC AAT CGC ACC ATT GAT CAT GTT GAG ATT ACC GTG GCA GAA GAA GTG GGG ATC GAA GGG CGC TGG GGC TAT TTT
Ser Leu Phe Val Asn Asn Trp Asp Asn Arg Thr Ile Asp His Val Glu Ile Thr Val Ala Glu Glu Val Gly Ile Glu Gly Arg Trp Gly Tyr Phe 226

690 705 720 735 750 765
GAT AAA GCC GGT CAG ATG CGC GAC ATG ATC CAG AAC CAC CTG CTG CAA ATT CTT TGC ATG ATT GCG ATG TCT CCG CCG TCT GAC CTG AGC GCA GAC
Asp Lys Ala Gly Gln Met Arg Asp Met Ile Gln Asn His Leu Leu Gln Ile Leu Cys Met Ile Ala Met Ser Pro Pro Ser Asp Leu Ser Ala Asp 258

780 ⧫114 810 825 840 855 870
AGC ATC CGC GAT GAA AAA GTG AAA GTA CCT GAA GTC TCG TCG CCG CAT CGA CCG CTC CAA CGT ACG CGA AAA AAC CGT ACG CGG GCA ATA TAC TGC
Ser Ile Arg Asp Glu Lys Val Lys Val Pro Glu Val Ser Ser Pro His Arg Pro Leu Gln Arg Thr Arg Lys Asn Arg Thr Arg Ala Ile Tyr Cys 290

885 900 ⧫102 915 930 945 960
GTT CCC CAG GGC AAA AAA GTG CCG GGA TAT CTG GAA GAA GAG GGC GCG AAC AAG AGC AGC AAT ACA GAA ACT TTC GTG GCG ATC CGC GTC GAC ATT
Val Pro Gln Gly Lys Lys Val Pro Gly Tyr Leu Glu Glu Glu Gly Ala Asn Lys Ser Ser Asn Thr Glu Thr Phe Val Ala Ile Arg Val Asp Ile 322

975 990 1005 1020 1035 1050
GAT AAC TGG CGC TGG GCC GGT GTG CCA TTC TAC CTG CGT ACT GGT AAA CGT CTG CCG ACC AAA TGT TCT GAA GTC GTG GTC TAT TTC AAA ACA CCT
Asp Asn Trp Arg Trp Ala Gly Val Pro Phe Tyr Leu Arg Thr Gly Lys Arg Leu Pro Thr Lys Cys Ser Glu Val Val Val Tyr Phe Lys Thr Pro 354

1065 1080 1095 1110 1125 1140 1155
GAA CTG AAT CTG TTT AAA GAA TCG TGG CAG GAT CTG CCG CAG AAT AAA CTG ACT ATC CGT CTG CAA CCT GAT GAA GGC GTG GAT ATC CAG GTA CTG
Glu Leu Asn Leu Phe Lys Glu Ser Trp Gln Asp Leu Pro Gln Asn Lys Leu Thr Ile Arg Leu Gln Pro Asp Glu Gly Val Asp Ile Gln Val Leu 386

1170 1185 1200 1215 1230 1245
AAT AAA GTT CCT GGC CTT GAC CAC AAA CAT AAC CTG CAA ATC ACC AAG CTG GAT CTG AGC TAT TCA GAA ACC TTT AAT CAG ACG CAT CTG GCG GAT
Asn Lys Val Pro Gly Leu Asp His Lys His Asn Leu Gln Ile Thr Lys Leu Asp Leu Ser Tyr Ser Glu Thr Phe Asn Gln Thr His Leu Ala Asp 418

1260 1275 1290 1305 1320 1335 1350
GCC TAT GAA CGT TTG CTG CTG GAA ACC ATG CGT GGT ATT CAG GCA CTG TTT GTA CGT CGC GAC GAA GTG GAA GAA GCC TGG AAA TGG GTA GAC TCC
Ala Tyr Glu Arg Leu Leu Leu Glu Thr Met Arg Gly Ile Gln Ala Leu Phe Val Arg Arg Asp Glu Val Glu Glu Ala Trp Lys Trp Val Asp Ser 450

1365 1380 1395 1410 1425 1440
ATT ACT GAG GCG TGG GCG ATG GAC AAT GAT GCG CCG AAA CCG TAT CAG GCC GGA ACC TGG GGA CCC GTT GCC TCG GTG GCG ATG ATT ACC CGT GAT
Ile Thr Glu Ala Trp Ala Met Asp Asn Asp Ala Pro Lys Pro Tyr Gln Ala Gly Thr Trp Gly Pro Val Ala Ser Val Ala Met Ile Thr Arg Asp 482

1455 1470 1480 1490 1500 1510 1520 1530 1540 1550 1560
GGT CGT TTC TGG AAT GAG TTT GAG TAA TATCTGC GCCTTATCCT TTATGGTTAT TTTACCGGTA ACATGATCTT GCGCAGATTG TAGAACAATT TTTACACTTT CAGGCCTCGG
Gly Arg Ser Trp Asn Glu Phe Glu *

1570 1580 1590 1600 1610 1620
TGCCGATTCA CCCACGAGCC TTTTTTTATT ACACTGACTG AAACGTTTTT GCCCTATGAG CTC

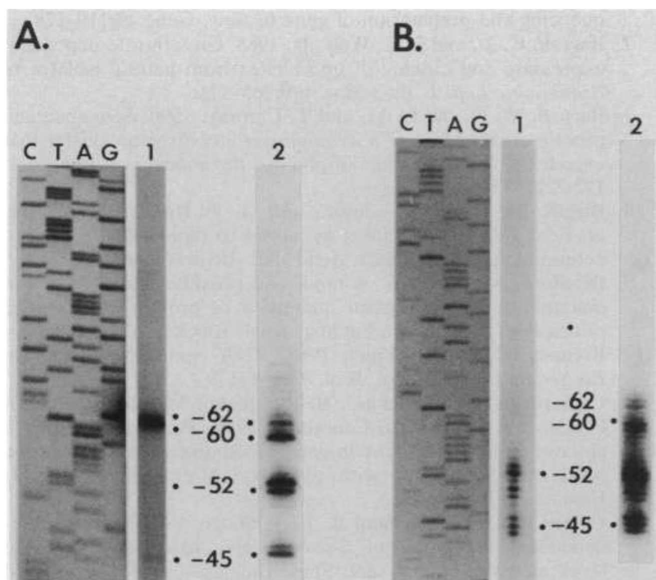

FIG. 4. Transcript mapping of the 5′ ends of *zwf* mRNA. RNA was isolated and the mapping was carried out as described in the text. (A) Mapping by the reverse transcriptase primer extension method. Lane 1, RNA from strain DF100; lane 2, RNA from strain GB1815(pDR17). (B) Mapping by S1 nuclease protection. Lane 1, RNA from strain DF100; lane 2, RNA from strain GB1815(pDR17). The products of sequencing reactions with reverse transcriptase of DNA from phage mDR17a are shown to the left of each panel. The numbers between lanes 1 and 2 in each panel refer to the location of the 5′ ends of the transcripts with respect to the sequence given in Fig. 3. No transcripts mapped to the interval between −40 and +1 (not shown).

ever, the segment from −61 to −12, which is just upstream from the Shine-Dalgarno core sequence, folded into an extended secondary structure with a free energy of −14.4 kcal (−60.25 kJ)/mol (data not shown). Possibly, this structure is the substrate for endonucleases that process the mRNA to the species whose 5′ ends are at −53 and −45.

DISCUSSION

The work described here initiates the molecular characterization of the *E. coli zwf* gene and the regulation of its expression. There are several reasons for studying this gene. First, few *E. coli* genes for enzymes of central intermediary metabolism have been subjected to molecular study. Thus, little is known about specific mechanisms governing their expression or even whether there are common themes (20). Second, growth-rate-dependent regulation of the level of 6PGD, another enzyme of the pentose phosphate pathway, depends on a *cis*-acting control sequence located in the structural gene that is complementary to the translation initiation region (12). Since it is important to determine whether this unusual genetic regulatory device is used by other genes, a logical place to look was at *zwf*, because both genes encode dehydrogenases of the same metabolic pathway and they elicit similar physiological responses to steady-state growth at different rates and to nutritional shiftups (16, 56). Third, Fraenkel and colleagues isolated three mutants that produce G6PD at levels 3- to 15-fold higher than wild-type levels (21, 24). Although they were the first regulatory mutants of *E. coli* central metabolism genes, the molecular basis for their phenotype has not been determined. However, genetic studies have shown that the mutations are *cis* dominant and map to the *edd*-distal end of the gene, beyond all known structural gene mutations (24); furthermore, we have found that the "up" mutations alter growth-rate-dependent regulation of the G6PD level (45a). Thus, identifying the responsible mutations might facilitate characterization of the mechanism(s) controlling the level and metabolic regulation of *zwf* expression. Finally, the level of G6PD is induced by the superoxide-generating agent paraquat (34), and as such *zwf* is one of about 35 members of the regulon that responds to redox-cycling agents (28); moreover, regulation by oxidative stress is at the transcriptional level (28, 55a). Thus, it will be interesting to determine the mechanism for induction by superoxides and its relationship, if any, to that of growth-rate-dependent regulation.

Several lines of evidence demonstrate that the gene characterized here is the *E. coli zwf* gene: the cloned DNA complements a *zwf* deletion; the restriction map of the *zwf* locus of strain DF100, as determined by Southern analysis, is the same as that of the cloned DNA; and the primary amino acid sequence of the polypeptide deduced from the DNA sequence is very similar from end to end to the G6PDs from *Z. mobilis*, *Drosophila melanogaster*, rats and humans, and it includes the peptide sequence shown for other G6PDs to be involved in substrate binding (11, 18, 30, 32, 33, 36, 44).

The strategy for sequencing *zwf* used an approach suggested by Barcak and Wolf (6), who had noted that a subset of 3′ unidirectional deletions of a gene cloned into phage M13 in the proper orientation confers LacZα-complementing activity, because the deletions produce in-frame protein fusions between the target gene and *lacZα*. Thus, the approach makes deletions easy to identify because they confer a new phenotype, it facilitates sequencing by bringing various segments of the target gene close to the annealing site for the sequencing primer, and it aids subsequent genetic study of the gene by giving rise to protein fusions. The approach was particularly useful for characterization of *zwf* because of our interest in using protein fusions to search for an ICS within the G6PD coding sequence, as was done previously with *gnd* (4).

The map of the 5′ ends of *zwf* mRNA molecules in RNA preparations from a wild-type strain growing on glucose was complex, with at least three transcript families being identifiable in the 100-bp region upstream of the start codon (Fig. 4). The same map was obtained with RNA from acetate-grown cells. Although it is possible that the various tran-

FIG. 3. Nucleotide sequence of the *E. coli zwf* locus and deduced amino acid sequence of G6PD. The sequence of the 2,331 bp of the region between the *Sal*I and *Sst*I sites is shown. Numbers above the sequence are nucleotide positions relative to the ATG start codon of the structural gene, and numbers at the right margin are the positions of amino acid residues in the polypeptide chain. The symbol Φ is placed over the last base of *zwf* present at the junction between *zwf* and *lacZα* in the in-frame *zwf-lacZα* protein fusions. Note that the fusion joint is between codons for all of the fusions except fusion mDR161, which is in-frame because one nucleotide was not filled in during the preparation of the deletion. The −35 and −10 hexamers of the *zwf* promoter are underlined, the Shine-Dalgarno (SD) sequence is in bold typeface, and the location of the oligonucleotide used for the transcript mapping is overlined. The dots above the sequence mark the location of the 5′ ends of transcripts mapped as described in the text.

scripts originated from different promoters, transcript mapping of RNA from plasmid and fusion strains combined with computer analysis of the nucleotide sequence strongly suggests that *zwf* has one promoter, as noted in Fig. 3. Thus, transcripts with ends mapping to −62 and −60 are probably the primary initiation products, and those mapping to −52 and −45 are probably due to processing. Accordingly, the *zwf* mRNA leader would be about the same length as that of *gnd* (41) and many other *E. coli* mRNAs (35). Also, since the transcript maps and the relative proportions of the various species were the same in RNA from cells growing on glucose and acetate, we can tentatively conclude that growth-rate-dependent regulation of G6PD level does not depend on alternative expression from two different promoters or on differential processing of *zwf* mRNA.

The nucleotide sequence of *zwf* has provided some information about possible mechanisms of metabolic control of G6PD level. Of particular interest was whether the G6PD coding sequence contains an ICS like that of *gnd*, i.e., a sequence capable of forming a thermodynamically stable mRNA secondary structure that sequesters the translation initiation region (12). Computer analysis did not reveal such a structure; however, we did notice a short mRNA segment complementary to the Shine-Dalgarno region that could potentially be involved in regulation. The *zwf-lacZ* protein fusions described here have been used in work to be reported elsewhere to determine genetically whether this or any other sequence within the structural gene is involved in *zwf* regulation. Also, we point out that the *zwf* leader does not contain a secondary structure resembling a factor-independent transcription terminator, so the mechanism proposed for growth rate control of *ampC* expression is not likely to be used for *zwf* regulation (29, 31). However, a secondary structure of the leader was found by computer analysis that could be responsible for the processing that putatively gives rise to the −52 and −45 transcript families. Finally, the region between the −10 hexamer and the start of transcription resembles the discriminator sequence that has been implicated in stringent control and growth rate regulation of stable RNA genes (13, 53) and thus could be involved in metabolic control of G6PD level.

ACKNOWLEDGMENTS

We thank M. C. O'Neill for his interest in the work and for help in the promoter searches. We also thank T. Conway for providing the sequence of the *Z. mobilis zwf* gene prior to publication and D. G. Fraenkel, T. Silhavy, and R. Zagursky for strains.

This research was supported by Public Health Service grant GM27113 from the National Institutes of Health.

REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1987. Current protocols in molecular biology. Wiley Interscience, New York.
2. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
3. **Baker, H. V., II, and R. E. Wolf, Jr.** 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.
4. **Baker, H. V., II, and R. E. Wolf, Jr.** 1984. Essential site for growth rate-dependent regulation within the *Escherichia coli gnd* structural gene. Proc. Natl. Acad. Sci. USA **81**:7669–7673.
5. **Banerjee, S., and D. G. Fraenkel.** 1972. Glucose-6-phosphate dehydrogenase from *Escherichia coli* and from a "high-level" mutant. J. Bacteriol. **110**:155–160.
6. **Barcak, G. J., and R. E. Wolf, Jr.** 1986. A method for unidirectional deletion mutagenesis with application to nucleotide sequencing and preparation of gene fusions. Gene **49**:119–128.
7. **Barcak, G. J., and R. E. Wolf, Jr.** 1988. Growth-rate-dependent expression and cloning of *gnd* alleles from natural isolates of *Escherichia coli*. J. Bacteriol. **170**:365–371.

7a. **Barnell, W. O., K. C. Yi, and T. Conway.** 1990. Sequence and genetic organization of a *Zymomonas mobilis* gene cluster that encodes several enzymes of glucose metabolism. J. Bacteriol. **172**:7227–7240.

8. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and 35S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
9. **Bradford, M. M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
10. **Bremer, H., and D. Yuan.** 1968. RNA chain growth rate in *Escherichia coli*. J. Mol. Biol. **75**:161–179.
11. **Camardella, L., C. Caruso, B. Rutigliano, M. Romano, G. Di Prisco, and F. Descalzi-Cancedda.** 1988. Human erythrocyte glucose-6-phosphate dehydrogenase: identification of a reactive lysyl residue labelled with pyridoxyl 5′-phosphate. Eur. J. Biochem. **171**:485–489.
12. **Carter-Muenchau, P., and R. E. Wolf, Jr.** 1989. Growth-rate-dependent regulation of 6-phosphogluconate dehydrogenase level mediated by an anti-Shine-Dalgarno sequence located within the *Escherichia coli gnd* structural gene. Proc. Natl. Acad. Sci. USA **86**:1138–1142.
13. **Cashel, M., and K. E. Rudd.** 1987. The stringent response, p. 672–704. *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.
14. **Clarke, L., and J. Carbon.** 1976. A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *E. coli* genome. Cell **9**:91–99.
15. **Csonka, L., and D. G. Fraenkel.** 1977. Pathways of NADPH formation in *Escherichia coli*. J. Biol. Chem. **152**:3382–3391.
16. **Farrish, E. E., H. V. Baker II, and R. E. Wolf, Jr.** 1982. Different control circuits for growth rate-dependent regulation of 6-phosphogluconate dehydrogenase and protein components of the translational machinery in *Escherichia coli*. J. Bacteriol. **152**:584–594.
17. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267.
18. **Fouts, D., R. Ganguly, A. G. Gutierrez, J. C. Lucchesi, and J. E. Manning.** 1988. Nucleotide sequence of the *Drosophila* glucose-6-phosphate dehydrogenase gene and comparison with the homologous human gene. Gene **63**:261–275.
19. **Fraenkel, D. G.** 1968. Selection of *Escherichia coli* mutants lacking glucose-6-phosphate dehydrogenase or gluconate-6-phosphate dehydrogenase. J. Bacteriol. **95**:1267–1271.
20. **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.
21. **Fraenkel, D. G., and S. Banerjee.** 1971. A mutation increasing the amount of a constitutive enzyme in *Escherichia coli*, glucose 6-phosphate dehydrogenase. J. Mol. Biol. **56**:183–194.
22. **Fraenkel, D. G., and S. Banerjee.** 1972. Deletion mapping of *zwf*, the gene for a constitutive enzyme, glucose 6-phosphate dehydrogenase in *Escherichia coli*. Genetics **71**:481–489.
23. **Fraenkel, D. G., and S. R. Levisohn.** 1967. Glucose and gluconate metabolism in an *Escherichia coli* mutant lacking phosphoglucose isomerase. J. Bacteriol. **93**:1571–1578.
24. **Fraenkel, D. G., and A. Parola.** 1972. "Up-promoter" mutations of glucose 6-phosphate dehydrogenase in *Escherichia coli*. J. Mol. Biol. **71**:107–111.
25. **Fraenkel, D. G., and R. T. Vinopal.** 1973. Carbohydrate metabolism in bacteria. Annu. Rev. Microbiol. **27**:69–100.
26. **Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner.** 1986. Improved

free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83**:9373–9377.
27. **Gold, L., D. Pribnow, T. Schneider, S. Shinedling, B. S. Singer, and G. Stormo.** 1981. Translational initiation in prokaryotes. Annu. Rev. Microbiol. **35**:365–403.
28. **Greenberg, J. T., and B. Demple.** 1989. A global response induced in *Escherichia coli* by redox-cycling agents overlaps with that induced by peroxide stress. J. Bacteriol. **171**:3933–3939.
29. **Grundström, T., and S. Normark.** 1985. Initiation of translation makes attenuation of *ampC* in *E. coli* dependent on growth rate. Mol. Gen. Genet. **198**:411–415.
30. **Ho, Y.-S., A. J. Howard, and J. D. Crapo.** 1988. Cloning and sequence of a cDNA encoding rat glucose-6-phosphate dehydrogenase. Nucleic Acids Res. **16**:7746.
31. **Jaurin, B., T. Grundström, T. Edlund, and S. Normark.** 1981. The *Escherichia coli* β-lactamase attenuator mediates growth-rate-dependent regulation. Nature (London) **290**:221–225.
32. **Jeffery, J., L. Hobbs, and H. Jörnvall.** 1985. Glucose-6-phosphate dehydrogenase from Saccharomyces cerevisiae: characterization of a reactive lysine residue labeled with acetylsalicylic acid. Biochemistry **24**:666–671.
33. **Jeffery, J., J. Söderling-Barros, L. A. Murray, R. J. Hansen, B. Szepesi, and H. Jörnvall.** 1988. Molecular diversity of glucose-6-phosphate dehydrogenase: rat enzyme structure identifies NH2-terminal segment, shows initiation from sites nonequivalent in different organisms, and establishes otherwise extensive sequence conservation. Proc. Natl. Acad. Sci. USA **85**:7840–7843.
34. **Kao, S. M., and H. M. Hassan.** 1985. Biochemical characterization of a paraquat-tolerant mutant of *Escherichia coli*. J. Biol. Chem. **260**:10478–10481.
35. **Kennell, D. E.** 1986. The instability of messenger RNA in bacteria, p. 101–142. *In* W. Reznikoff and L. Gold (ed.), Maximizing gene expression. Butterworths, Stoneham, Mass.
36. **Kletzein, R. F., C. R. Prosko, D. J. Stumpo, J. K. McClung, and K. L. Dreher.** 1985. Molecular cloning of DNA sequences complementary to rat liver glucose-6-phosphate dehydrogenase. Nutritional regulation of mRNA levels. J. Biol. Chem. **260**:5621–5624.
36a. **Luethy, M., and R. E. Wolf, Jr.** Unpublished data.
37. **Maloy, S. R., and W. D. Nunn.** 1981. Selection for loss of tetracycline resistance by *Escherichia coli*. J. Bacteriol. **145**:1110–1112.
38. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **Messing, J.** 1983. New M13 cloning vectors for cloning. Methods Enzymol. **101**:20–78.
40. **Miller, J.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. **Nasoff, M. S., H. V. Baker II, and R. E. Wolf, Jr.** 1984. Nucleotide sequence of the *Escherichia coli* K-12 gene, *gnd*, for 6-phosphogluconate dehydrogenase. Gene **27**:253–264.
42. **Oliphant, A. R., and K. Struhl.** 1988. Defining the consensus sequences of *Escherichia coli* promoter elements by random selection. Nucleic Acids Res. **16**:7673–7683.
43. **O'Neill, M. C.** 1989. Consensus methods for finding and ranking DNA binding sites. Application to *Escherichia coli* promoters. J. Mol. Biol. **207**:301–310.
43a. **O'Neill, M. C.** Nucleic Acids Res., in press.
44. **Persico, M. G., G. Viglietto, 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′ non-coding region. Nucleic Acids Res. **14**:2511–2522.
45. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
45a. **Rowley, D. L., and R. E. Wolf, Jr.** Unpublished data.
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. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15**:1281–1295.
48. **Sheiness, D., and D. Sullivan.** 1986. Staining of RNA in formaldehyde gels. Focus **8**:11.
49. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16s ribosomal RNA; complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **72**:1342–1346.
50. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
51. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
52. **Thomson, J., P. D. Gerstenberger, D. E. Goldberg, E. Gociar, A. Orozco de Silva, 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. **Travers, A.** 1984. Conserved features of coordinately regulated *E. coli* promoters. Nucleic Acids Res. **12**:2605–2618.
54. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertional mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
55. **Wolf, R. E., Jr.** 1980. Integration of specialized transducing bacteriophage λ *c*I857 *St*68 *h*80 d*gnd his* by an unusual pathway promotes formation of deletions and generates a new translocatable element. J. Bacteriol. **142**:588–602.
55a. **Wolf, R. E., Jr.** Unpublished data.
56. **Wolf, R. E., Jr., 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. **Wolf, R. E., Jr., and F. M. Shea.** 1979. Combined use of strain construction and affinity chromatography in the rapid, high-yield purification of 6-phosphogluconate dehydrogenase from *Escherichia coli*. J. Bacteriol. **138**:171–175.
58. **Zasloff, M., G. D. Ginder, and G. Felsenfeld.** 1978. A new method for purification and identification of covalently closed circular DNA molecules. Nucleic Acids Res. **5**:1139–1152.
59. **Zucker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.