JOURNAL OF BACTERIOLOGY, Feb. 1989, p. 767–774
0021-9193/89/020767-08$02.00/0

Vol. 171, No. 2

Cloning, Sequencing, and Characterization of the Principal Acid Phosphatase, the *phoC*$^+$ Product, from *Zymomonas mobilis*†

J. L. POND, C. K. EDDY, KYLIE F. MACKENZIE, T. CONWAY, D. J. BORECKY, AND L. O. INGRAM*

Department of Microbiology and Cell Science, University of Florida, Gainesville, Florida 32611

Received 12 July 1988/Accepted 1 November 1988

The *Zymomonas mobilis* gene encoding acid phosphatase, *phoC*, has been cloned and sequenced. The gene spans 792 base pairs and encodes an M_r 28,988 polypeptide. This protein was identified as the principal acid phosphatase activity in *Z. mobilis* by using zymograms and was more active with magnesium ions than with zinc ions. Its promoter region was similar to the −35 "*pho* box" region of the *Escherichia coli pho* genes as well as the regulatory sequences for *Saccharomyces cerevisiae* acid phosphatase (*PHO5*). A comparison of the gene structure of *phoC* with that of highly expressed *Z. mobilis* genes revealed that promoters for all genes were similar in degree of conservation of spacing and identity with the proposed *Z. mobilis* consensus sequence in the −10 region. The *phoC* gene contained a 5′ transcribed terminus which was AT rich, a weak ribosome-binding site, and less biased codon usage than the highly expressed *Z. mobilis* genes.

Zymomonas mobilis is obligately fermentative and utilizes the Entner-Doudoroff pathway for glycolysis (25). Since this organism derives only 1 mol of ATP, net, per mol of glucose consumed, it makes large amounts of the glycolytic and ethanologenic enzymes, which account for 30 to 50% of the soluble cell protein (28). The DNA sequences for four highly expressed *Z. mobilis* genes are now known (5, 7–9). A comparison of these genes with genes not involved in central metabolism provides an opportunity to identify the structural features involved in high-level gene expression in *Z. mobilis*.

In this study, we report the cloning and sequencing of the *phoC* gene for the principal acid phosphatase isozyme from *Z. mobilis*.

Phosphatase genes have been important tools for the study of membrane proteins and protein secretion (15). Multiple phosphatases are typical in microbial systems (11, 27). These are divided into two groups based on pH optima, each represented by isozymes. Alkaline phosphatases exhibit a pH optimum above 8, while the optimal pH for acid phosphatases is pH 5 or below. An unusually low-pH phosphatase (pH 2.5) has been reported for *Escherichia coli*, which may constitute a third type (39).

Alkaline phosphatases from *E. coli* (4), *Saccharomyces cerevisiae* (19), animals (38), and humans (18) have considerable amino acid similarity. In contrast, the acid phosphatase genes which have been described for different organisms (*PHO3*, *PHO5*, and *PHO11* from *S. cerevisiae* [3, 34], *PHO1* from *Schizosaccharomyces pombe* [12], and the partial sequence for a pH 2.5 acid phosphatase from *E. coli* [39]) have no amino acid similarity.

The regulation of *E. coli* alkaline phosphatase (40, 41) and yeast acid phosphatases (3, 12) is complex and involves both regulatory proteins and cyclic AMP. Sequences involved in regulation have been identified for the *pho* regulon of *E. coli* (34) and for the cyclic AMP-repressible acid phosphatase from *S. cerevisiae* (3).

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. Strains and plasmids have been described previously (5) except for those constructed in this study. *E. coli* was grown at 37°C in Luria broth or agar (1.5%) without added carbohydrate (22). Growth conditions and antibiotic supplements were as previously described (5). Luria agar containing 5-bromo-4-chloro-3-indolyl phosphate, *p*-toluidine salt (200 μg/ml), potassium phosphate (20 mM), and appropriate antibiotics was used for the direct selection of transformants expressing phosphatase activity. For in vitro examination of phosphatase activity in recombinants, cells were grown at 37°C in a shaking water bath and harvested in exponential phase (OD_{550}, 0.5 to 0.6).

Z. mobilis CP4 was maintained on complex medium as described previously (26). To examine phosphatase activity under conditions of excess and limited phosphate, a semisynthetic medium was devised which contained (per liter): 0.5 g of NaCl, 1.0 g of $(NH_4)_2SO_4$, 0.5 g of $MgSO_4 \cdot 7H_2O$, 2 g of MES (morpholineethanesulfonic acid)-KOH (pH 6.0), 0.05 g of yeast extract, 100 g of glucose, 5 mg of calcium pantothenate, 1 mg of biotin, 1 mg of thiamine hydrochloride, 1 mg of pyridoxine hydrochloride, and 1 mg of nicotinic acid. This medium was prepared from four separately autoclaved stock solutions: inorganic salts, glucose, yeast extract, and MES. Vitamins were sterilized by filtration. This medium was supplemented with 3 mM KH_2PO_4 (pH 6.0) to provide excess phosphate, which supported vigorous cell growth. Growth was limited to a final OD_{550} of 0.16 in basal medium lacking phosphate.

Phosphatase assay. Cells were harvested by centrifugation (5,000 × *g*) and washed once in an equal volume of 50 mM bicine hydrochloride buffer (pH 8.0). Cell pellets were suspended in bicine buffer, dispensed into 1.5-ml centrifuge tubes (1.0 ml per tube), harvested by centrifugation (10,000 × *g*, 1 min), drained, and stored on ice. Cells were permeabilized for enzyme assays by resuspension in bicine buffer containing lysozyme (1 mg/ml, freshly prepared) and 0.5% Triton X-100, followed by incubation on ice for 15 min. Enzymatic activity was measured at pH 5.0 (100 mM sodium acetate buffer, 0.5 mM $ZnSO_4$) and pH 9.0 (100 mM bicine buffer, 0.5 mM $MgSO_4$) containing 10 μM *p*-nitro phenylphosphate. Incubations were performed at 30°C, and reac-

* Corresponding author.
† Florida Agricultural Experiment Station publication no. 9122.

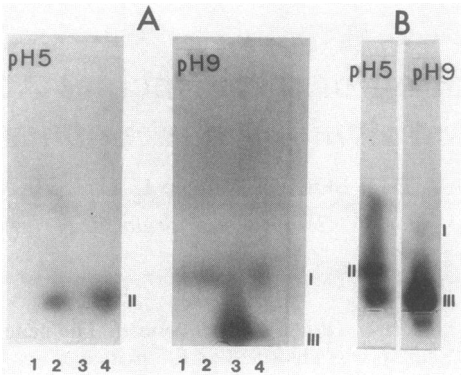

FIG. 1. Zymograms developed for phosphatase activity. (A) Comparison of recombinant and native phosphatase activity. Duplicate gels were developed at pH 5 and pH 9. The three principal phosphatase activities in extracts from cells grown with added phosphate are labeled I, II, and III. All lanes except phosphate-starved *Z. mobilis* cells (lane 3, 5 μg of cell protein) contained approximately 75 to 100 μg of cell protein. Lanes: 1, *E. coli* HB101(pUC18); 2, HB101(pLOI334); 3, phosphate-starved *Z. mobilis* CP4; 4, CP4 with added phosphate. (B) Comparison of phosphatase activity in phosphate-starved *Z. mobilis* cells. Both lanes contained 50 μg of cell protein. Gels were developed at pH 5 and pH 9.

tions were terminated by the addition of 4 volumes of 2 N NaOH. Absorbance was measured at 410 nm (10).

Zymograms of phosphatase activity. Cell pellets were resuspended in bicine buffer containing lysozyme and disrupted by the freeze-thaw method (6), then mixed with an equal volume of 50% sucrose containing 0.1% bromophenol blue, and separated in a slab gel. Electrophoresis was done in Tris-borate buffer containing 5.4 g of Tris-free base and 2.7 g of boric acid (pH 8.5) per liter at ambient temperature for approximately 3 h at 75 V in a Bio-Rad (Richmond, Calif.) horizontal submarine gel apparatus. Gels (10 by 7 cm) were prepared in Tris-borate buffer containing 1% agarose and 5% clinical-grade dextran (M_r 60,000 to 90,000 fraction; Sigma Chemical Co., St. Louis, Mo.). Gels were stained with Coomassie blue (23) or developed for enzymatic activity as zymograms.

TABLE 1. Effect of phosphate starvation on phosphatase activity in *Z. mobilis*

Growth condition	Metal ion	Activity (IU/mg of protein)	
		pH 5	pH 9
Phosphate (3 mM)	Mg	0.11	0.02
	Zn	0.08	ND[a]
Phosphate starved	Mg	1.0	0.40
	Zn	0.57	0.62

[a] ND, Not determined.

Before being developed for phosphatase activity, gels were soaked for 1 h at room temperature in four changes of either 100 mM potassium acetate buffer (pH 5.0) or 100 mM bicine buffer (pH 9.0). Gels were stained for phosphatase activity by soaking overnight at 30°C in 50 ml of either the low-pH buffer with 5 mM magnesium sulfate or high-pH buffer with zinc chloride, both containing 9.2 mg of 5-bromo-4-chloro-3-indolyl-phosphate, *p*-toluidine salt. Cleavage of the chromogenic substrate resulted in blue spots in the translucent gel, which were best viewed over an orange background.

SDS-PAGE. For sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), whole-cell proteins were denatured and examined on 10% acrylamide gels as described previously (5).

Cloning the gene encoding acid phosphatase. A *Z. mobilis* chromosomal DNA library was constructed previously by inserting partial *Sau*3A-digested fragments of 4 to 6 kilobases (kb) into the *Bam*HI site of pLOI193 (9). *E. coli* TC4 transformants were screened on phosphatase indicator plates.

Subclones were prepared with the pUC18 and pUC19 vectors by transformation into *E. coli* HB101, by standard methods (23).

Determination of DNA sequence. The gene encoding acid phophatase was sequenced in both directions by using the dideoxy method of Sanger et al. (33). An overlapping set of Bal31-generated deletions was obtained from pLOI333 and subcloned into bacteriophages M13mp18 and M13mp19 (24).

The universal M13 primer was used for most of the sequencing. In addition, the 15-mer GATGAACCGCGGAC

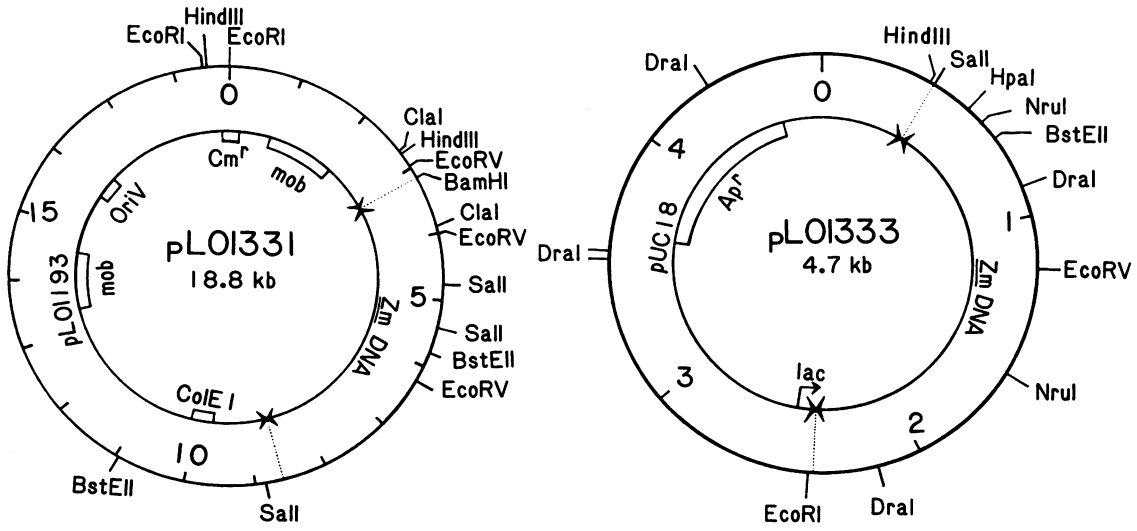

FIG. 2. Circular restriction maps of pLOI331 and pLOI333. Double arrows indicate junctions of *Z. mobilis* (Zm) and vector DNAs.

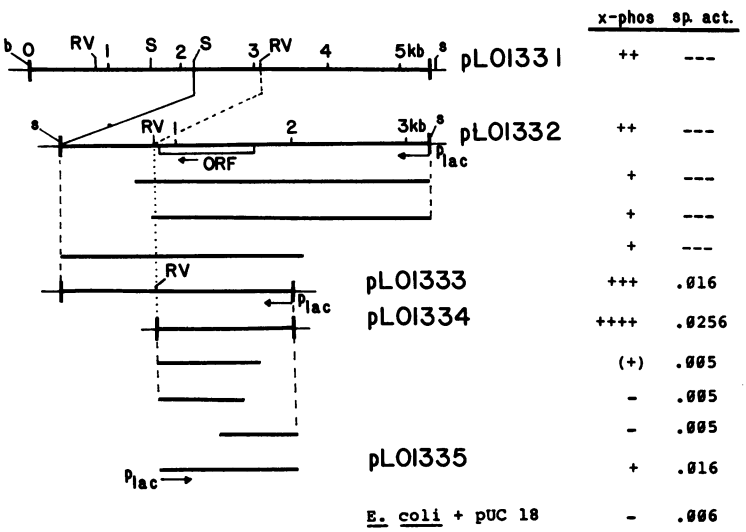

FIG. 3. Linear restriction maps of plasmids. Activity of clones on X-phos plates and their specific activity in permeabilized cells are also indicated. The direction of transcription from the *lac* promoter (p_{lac}) is shown by arrows. Uppercase letters show restriction sites in the *Z. mobilis* DNA, and lowercase letters show sites in the vector or at the site of vector-insert fusion. Abbreviations: B, *Bam*HI; RV, *Eco*RV; S, *Sal*I; ORF, open reading frame. Symbols under X-phos indicate relative blueness of colonies on indicator plates, ranging from ++++ to − (white). Specific activity is given in units per milligram per minute.

was synthesized and used to sequence a portion of the amino terminus and upstream region and also for primer extension analysis.

Sequence data were analyzed with the programs of Pustell and Kafatos (30; International Biotechnologies, Inc., New Haven, Conn.) and with programs provided by the Protein Identification Resource, National Biomedical Research Foundation (Washington, D.C.). Amino acid homology searches were performed with the FASTP program described by Lipman and Pearson (21). Predicted RNA folding was examined by using PCFOLD, developed by Zucker and Steigler (42) and distributed by the Molecular Biology Computer Research Resource (Harvard School of Public Health, Boston, Mass.).

Analysis of transcription initiation site. The 5′ terminus of transcription was mapped by primer extension analysis essentially as described before (5). A 15-base oligonucleotide complementary to the message spanning codons 4 through 8 was used as the primer.

RESULTS AND DISCUSSION

Phosphatase activities in *Z. mobilis*. Phosphatase activities capable of hydrolyzing *o*-nitrophenylphosphate were examined in permeabilized cells (Table 1). Permeabilization of cells with chloroform, as previously used for glycolytic enzyme assays in *Z. mobilis* (26), caused substantial inactivation of phosphatase activity. Treatment with a combination of Triton X-100 and lysozyme gave a reproducible and efficient method for cell permeabilization while retaining high enzyme activities. Preliminary experiments revealed both acid and alkaline phosphatase activities with optima at pH 5 and 9, respectively. Acid phosphatase activity was highest in the presence of magnesium ions, while alkaline phosphatase activity was highest with zinc. Both activities increased during growth in phosphate-limited medium (Table 1).

Cell lysates were fractionated by electrophoresis through agarose-dextran gels and developed for phosphatase activity (Fig. 1). Three activity bands which migrated anodally in Tris-borate (pH 8.5) were resolved for cells grown with added phosphate. No cathodally migrating activities were detected in cells grown with added phosphate. The phosphatase migrating most rapidly toward the anode (PHOSI) and the slowest phosphatase activity (PHOSIII) were visualized only in gels assayed at pH 9. The single acid phosphatase (PHOSII) evident in the zymogram developed at pH 5 exhibited intermediate mobility.

By comparing the zymograms for cells grown with abundant phosphate with ones for cells grown under phosphate-limited conditions, seven phosphatase activities were identified. Three were most active at pH 5, and four were most active at pH 9.

Cloning the gene encoding the major acid phosphatase. Three of 2,000 transformants were phosphatase positive on indicator plates. These three may be siblings because each had a 5.4-kb DNA insert and their preliminary restriction patterns were similar. One plasmid, pLOI331, was studied further (Fig. 2).

Phosphatase activity was retained on a 3.2-kb *Sal*I fragment which was subcloned into pUC18 to produce pLOI332 (Fig. 2 and 3). Deletion derivatives of this clone were generated by using Bal31 and reinserted into pUC18 to identify the coding region. Approximately 1.2 kb was deleted from the right side without affecting activity. The resulting

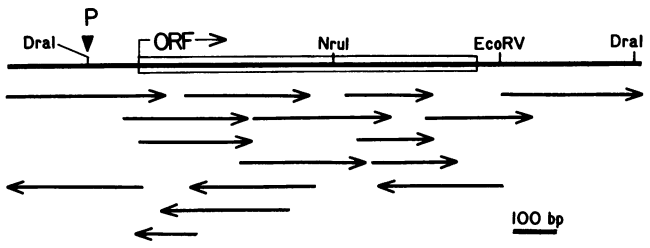

FIG. 4. Sequencing strategy for the acid phosphatase gene. P, Promoter; ORF, open reading frame.

```
                                                -150                                                -100
        *         *         *         *         *         *         *         *         *         *
GTGCCTTATATCACGGGGGGATCAGTCTGCGGCAATTGGTGGCATGGCACGCGCTACGGCACGCCGGAAGGCTTCATGGTCGTCAAAGTCGAAAAGGGTA

                                                -50          "-35"                          "-10"   0
        *         *         *         *         *         *         *         *         *         *
AAGTCATTCCGCATTACGAAAGCTATGGCTTCCACACGATAGACCCGCGCAACACATAATTGTCTTATTATAGCCACATGATATTTTTATATTACAATTT

+1                                              50                                                  100
:       *         *         *         *         *         *         *         *         *         *
TAAACTAAAATTAAGAATTAAATTCTTGAAATAAAGGTTTTTTTATTAAAAGGATAGGAAATGTCGTGAAATCGGCATTTTCTATCCATATTATATAACA

                                                150
   S.D.  *         *         *         *         *         *         *         *
AGGGAAGACTGACGAC ATG ATA AAA GTC CCG CGG TTC ATC TGT ATG ATC GCG CTT ACA TCC GGC GTT CTG GCA AGC GGC
              fMet Ile Lys Val Pro Arg Phe Ile Cys Met Ile Ala Leu Thr Ser Gly Val Leu Ala Ser Gly

                        200                                                                 250
*         *         *         *         *         *         *         *         *         *
CTT TCT CAA AGC GTT TCA GCT CAT ACA GAA AAA AGT GAA CCC TCC TCG ACT TAT CAT TTC CAC AGC GAT CCC CTT
Leu Ser Gln Ser Val Ser Ala His Thr Glu Lys Ser Glu Pro Ser Ser Thr Tyr His Phe His Ser Asp Pro Leu

                                                300
      *         *         *         *         *         *         *         *
CTT TAC CTT GCG CCC CCA CCC ACT TCC GGC AGT CCA TTA CAG GCG CAT GAT GAT CAA ACC TTT AAC AGC ACC AGA
Leu Tyr Leu Ala Pro Pro Pro Thr Ser Gly Ser Pro Leu Gln Ala His Asp Asp Gln Thr Phe Asn Ser Thr Arg

                        350                                                                 400
*         *         *         *         *         *         *         *         *         *
CAA TTA AAA GGT AGC ACG CGC TGG GCA TTG GCA ACT CAA GAT GCC GAT CTT CAT CTC GCT TCA GTT CTC AAA GAC
Gln Leu Lys Gly Ser Thr Arg Trp Ala Leu Ala Thr Gln Asp Ala Asp Leu His Leu Ala Ser Val Leu Lys Asp

                                                450
      *         *         *         *         *         *         *         *
TAT GCC TGC GCC GCA GGA ATG AAT CTC GAT ATT GCG CAA TTA CCG CAT CTT GCC AAT TTG ATT AAA CGC GCA CTT
Tyr Ala Cys Ala Ala Gly Met Asn Leu Asp Ile Ala Gln Leu Pro His Leu Ala Asn Leu Ile Lys Arg Ala Leu

                        500                                                                 550
*         *         *         *         *         *         *         *         *         *
CGC ACC GAA TAT GAC GAT ATT GGC AGA GCC AAA AAT AAC TGG AAT CGC AAA CGA CCT TTT GTG GAT ACC GAT CAA
Arg Thr Glu Tyr Asp Asp Ile Gly Arg Ala Lys Asn Asn Trp Asn Arg Lys Arg Pro Phe Val Asp Thr Asp Gln

                                                600
      *         *         *         *         *         *         *         *
CCC ATC TGC ACG GAA AAA GAT CGC GAA GGT CTG GGA AAA CAA GGC TCC TAT CCT TCA GGT CAT ACG ACT ATC GGT
Pro Ile Cys Thr Glu Lys Asp Arg Glu Gly Leu Gly Lys Gln Gly Ser Tyr Pro Ser Gly His Thr Thr Ile Gly

                        650                                                                 700
*         *         *         *         *         *         *         *         *         *
TGG AGC GTT GCG CTC ATT CTG GCT GAA TTG ATC CCC GAT CAT GCG GCG AAT ATT TTG CAG CGT GGC CAA ATT TTT
Trp Ser Val Ala Leu Ile Leu Ala Glu Leu Ile Pro Asp His Ala Ala Asn Ile Leu Gln Arg Gly Gln Ile Phe

                                                750
      *         *         *         *         *         *         *         *
GGA ACC AGC CGG ATT GTC TGC GGC GCC CAT TGG TTC AGC GAT GTG CAG GCA GGC TAT ATC ATG GCA TCG GGC GAA
Gly Thr Ser Arg Ile Val Cys Gly Ala His Trp Phe Ser Asp Val Gln Ala Gly Tyr Ile Met Ala Ser Gly Glu

                        800                                                                 850
*         *         *         *         *         *         *         *         *         *
ATT GCA GCT TTA CAT GGG GAT GCC GAT TTC CGC CGA GAT ATG GAA TTA GCT CGG AAA GAA TTA GAA AAG GCA CGC
Ile Ala Ala Leu His Gly Asp Ala Asp Phe Arg Arg Asp Met Glu Leu Ala Arg Lys Glu Leu Glu Lys Ala Arg

                                                900
      *         *         *         *         *         *         *         *         *
ACA TCA GCG CAC ACG CCA GAC GAT CTT CTA TGC AAG ATT GAA CAA AGC GCT CGC TAA ATTCAATCAAGTATTATTTCAACA
Thr Ser Ala His Thr Pro Asp Asp Leu Leu Cys Lys Ile Glu Gln Ser Ala Arg ---

              950                                                  1000
      *       *         *         *         *         *         *         *         *         *
AGGGGAAAGATTGCTTGCTGTAATTTTTGGATATCAAACAGGCGAAAAAATGAAAGAGCGCACGCTCTTTCAAAGGCAATTCGATTTAGTCCGGTGGCAT

              1050                                                 1100
      *       *         *         *         *         *         *         *         *         *
TCTCACGCCACAAACCAAATCATAAATAACCGCCTCTTTTCCGCCAGATAACTGCAAAATTATAGAATACCGACAGCTGGAATATCGTCACTTTTCCTAG
```

FIG. 5. Sequence and translation of the *Z. mobilis phoC* gene. The −35 region, −10 region, and Shine-Dalgarno (S.D.) region are underlined and labeled. A 29-base imperfect palindrome (positions 38 to 63 and 73 to 101) which exhibits homology with the *E. coli pho* box (34) is shown below the line of sequence.

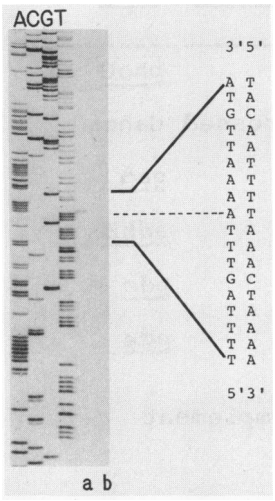

FIG. 6. Primer extension analysis of transcriptional initiation. Lanes: a, cDNA of transcript from *Z. mobilis* CP4; b, control lane (*E. coli* TC4). The lanes marked A, C, G, and T represent a sequencing ladder of the coding strand.

clone was designated pLOI333. The 2-kb fragment in pLOI333 was sequenced. Mutants carrying deletions from the left up to the *Eco*RV site retained activity.

Based on the deletions from both the left and right, the *Eco*RV to *Eco*RI (polylinker) fragment of pLOI333 was predicted to contain the coding sequence for phosphatase. This fragment was subcloned into pUC18 and pUC19 to provide both orientations with respect to the *lac* promoter (pLOI334 and pLOI335).

Expression of the *Z. mobilis* acid phosphatase gene in *E. coli*. Figure 3 shows the physical maps of the acid phosphatase clones, their appearance on indicator plates, and their acid phosphatase activities in vitro. The subclone which gave the highest activity (pLOI334) contained the 1.2-kb fragment in pUC18. This subclone formed dark blue colonies overnight on indicator plates, whereas the clone containing the 1.2-kb fragment in pUC19 (pLOI335) with the opposite orientation with respect to the *lac* promoter formed light blue colonies. Longer clones with additional DNA sequence on either side of the 1.2-kb fragment also formed lighter blue colonies, which were readily distinguished from the white colonies of strain HB101 containing the vector alone. These results are consistent with expression in *E. coli* being primarily from the *lac* promoter, with transcription occurring from right to left. The lower level of expression in the reverse orientation is attributed to cryptic promoter activity from pUC or from the *Z. mobilis* sequences. The proteins were compared in strain HB101(pUC18) and strain HB101(pLOI334). Although phosphatase activity was readily measured in pLOI334, no new protein bands were visible in SDS-polyacrylamide gels (data not shown).

Identification of the *phoC* product as the principal *Z. mobilis* acid phosphatase. Figure 1 shows a zymogram which compares the recombinant phosphatase activity with that for *Z. mobilis* grown with limiting and excess phosphate. The recombinant activity corresponded to the principal acid phosphatase (PHOSII) produced by *Z. mobilis* under both conditions. No acid phosphatase activity was observed in *E. coli* containing vector alone.

Sequence of the acid phosphatase gene. The sequencing strategy for the acid phosphatase (*phoC*) gene is shown in Fig. 4. Both strands were sequenced in the region of the gene except for a small region near the central *Nru*I site, and this sequence is shown in Fig. 5. A single, large open reading frame was found between the *Eco*RV site and the polylinker fusion, which was bordered by AT-rich sequences. Numerous stop codons were found in both other reading frames. The coding region contained 792 base pairs (bp), corresponding to 264 amino acids, with an aggregate molecular weight of 28,988. This reading frame was terminated by the codon TAA.

A single initiation site was identified by primer extension mapping in *Z. mobilis*, 116 bp upstream from the translational start codon (Fig. 6). This untranslated leader had the potential for considerable secondary structure. An imperfect palindromic sequence of 58 bases centered at bp 68 (Fig. 5) could form a stable stem-and-loop structure with a predicted energy of −25 kcal (42). Potential −10 and −35 sequences were also indicated by comparison with other *Z. mobilis* sequences (5, 7–9).

The most probable Shine-Dalgarno region (35, 37) in *phoC*, GGAAGAC, was located upstream from the start codon, which matches five of seven bases of the *E. coli* sequence (Fig. 7A). No good candidate for transcriptional termination (29) was apparent.

High codon bias has often been correlated with high-level expression (2, 13, 16, 17), and replacement of codons for which tRNA species are abundant with those specifying rare tRNAs has been shown to dramatically reduce expression in *S. cerevisiae* (14). Codon usage in the *phoC* gene was less biased than that of the glycolytic enzymes and alcohologenic enzymes of *Z. mobilis* (Table 2). Indeed, the patterns of codon usage of the highly expressed *Z. mobilis* proteins and the average values for *E. coli* proteins were more similar to each other than to *phoC*. If codon abundance in highly expressed genes reflects the relative abundance of tRNA species (16, 17), then the low codon bias of *phoC* may represent a major limitation for expression.

No protein sequences were found in the current Protein Identification Resource data base that had significant identity to the translated sequence of *phoC*. No amino acid homology was observed between *phoC* and the yeast acid phosphatases (3, 12), *E. coli* (4) or mammalian (18, 38) alkaline phosphatases, or the partial sequence for the *E. coli* pH 2.5 acid phosphatase (39). All acid phosphatase genes described thus far appear to be unique, lacking appreciable regions of amino acid homology.

A hydrophobicity plot of the translated amino acid sequence for *phoC* revealed a prominent hydrophobic segment near the N-terminus, preceded by two basic amino acids, arginine and lysine. This feature resembles a leader sequence for protein transport (20, 31) and may function to localize the *phoC* product in the periplasmic region.

Comparison of *phoC* promoter region. The sequence in the −35 region of *phoC* was remarkably similar to that of the "*pho* box" in *E. coli* (34) (Fig. 7B). In the best alignment, 12 of 18 bases were conserved in *Z. mobilis phoC*. Five conserved bases in the −10 region were also identical. The high degree of conservation of promoter structure between the *phoC* from *Z. mobilis* and the *pho* genes from *E. coli* suggests that the regulation of phosphatase expression may be similar.

Although not previously recognized, the *pho* box and the conserved regulatory sequences in *S. cerevisiae PHO5* share partial identity (CTGTCATAAAACTGTCAC and CTGCA CAAAG, respectively). The latter sequence has been shown

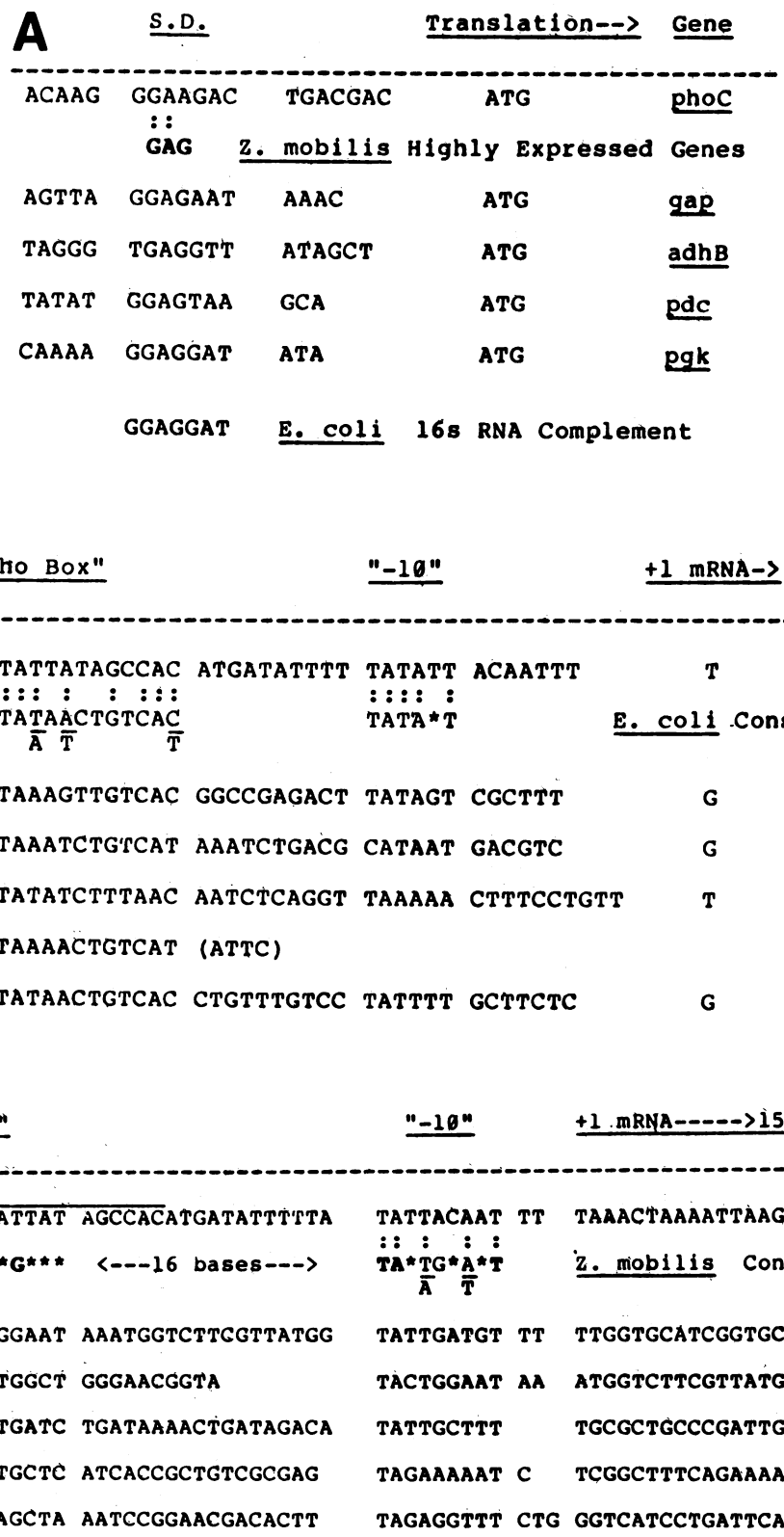

FIG. 7. Sequence comparisons. (A) Comparison of the *phoC* ribosome-binding region with that of highly expressed *Z. mobilis* genes. The probable Shine-Dalgarno (S.D.) region is labeled. The boldface sequence GAG is conserved in all highly expressed genes. Identity is denoted by a colon. The sequence complementary to *E. coli* 16s RNA is shown for comparison (35, 37). (B) Comparison of *Z. mobilis phoC* promoter region with that of the *E. coli pho* genes. Identity of aligned bases is indicated by a colon. Stacked bases indicate alternatives (34). Two sequences with identity to the *pho* box are present in *pstS* (shown in parentheses). (C) Comparison of the *phoC* promoter region with that of highly expressed *Z. mobilis* genes. In the consensus sequence, bases which were aligned in all highly expressed genes are shown in boldface, Uppercase bases are conserved in four of the five promoters from highly expressed genes. An asterisk denotes a variable base. The region of *phoC* which shows homology with the *E. coli pho* box is overlined.

TABLE 2. Comparison of translated codon usage

Amino acid	Codon	Frequency (mol% codon usage)		
		Z. mobilis combined[a]	*phoC*	*E. coli* combined[b]
Phe	TTT	0.5	1.1	1.3
	TTC	2.6	1.5	2.2
Leu	TTA	0.2	2.3	0.7
	TTG	1.5	1.5	0.9
	CTT	1.9	3.4	0.8
	CTC	1.7	1.5	0.8
	CTA	0	0.4	0.2
	CTG	3.9	1.1	6.8
Ile	ATT	1.3	3.4	2.2
	ATC	3.4	2.3	3.7
	ATA	0	0.4	0.2
Met	ATG	2.7	1.9	2.8
Val	GTT	5.6	1.5	2.9
	GTC	2.6	0.8	1.2
	GTA	<0.1	0	1.8
	GTG	0.5	0.8	2.2
Ser	TCT	1.1	0.4	1.3
	TCC	1.4	1.5	1.5
	TCA	0.3	1.5	0.4
	TCG	0.4	0.8	0.6
	AGT	0.2	0.8	0.3
	AGC	1.3	3.4	1.4
Pro	CCT	0.8	0.8	0.5
	CCC	0.2	2.3	0.3
	CCA	0.4	1.1	0.7
	CCG	2.9	0.8	2.5
Thr	ACT	0.6	1.5	1.1
	ACC	3.8	1.9	2.4
	ACA	<0.1	1.1	0.3
	ACG	1.3	1.5	0.8
Ala	GCT	8.2	2.3	2.6
	GCC	3.0	2.7	2.2
	GCA	2.4	3.4	2.3
	GCG	0.8	3.0	3.2
Tyr	TAT	1.6	1.9	1.0
	TAC	0.7	0.4	1.5
His	CAT	1.1	3.4	0.7
	CAC	1.3	0.8	1.2
Gln	CAA	0.1	3.4	1.0
	CAG	1.6	1.1	3.2
Asn	AAT	1.6	1.9	1.0
	AAC	3.1	0.8	2.8
Lys	AAA	3.9	3.8	4.1
	AAG	2.7	0.8	1.3
Asp	GAT	3.4	6.1	2.5
	GAC	2.6	1.1	3.0
Glu	GAA	5.1	4.2	4.9
	GAG	0.4	0	1.8
Cys	TGT	0.2	0.4	0.4
	TGC	0.9	1.5	0.5
Trp	TGG	0.6	1.5	0.7
Arg	CGT	2.0	0.4	3.1
	CGC	1.3	3.0	2.0
	CGA	0	0.8	0.2
	CGG	<0.1	1.1	0.2
	AGA	0	0.8	<0.1
	AGG	0	0	<0.1
Gly	GGT	5.8	1.5	3.8
	GGC	2.3	3.4	3.1
	GGA	0.1	1.1	0.4
	GGG	0	0.4	0.6

[a] Average of *Z. mobilis gap* (8), *pgk* (5), *adhB* (9), and *pdc* (7).
[b] Average of 52 proteins from *E. coli* (1).

to be sufficient to confer regulation by cyclic AMP in *S. cerevisiae* (3).

Figure 7C summarizes our knowledge of promoter structure in *Z. mobilis*. In the −10 region of *Z. mobilis*, three bases are conserved in *phoC* and the highly expressed genes. The −35 region of the highly expressed genes contains two conserved bases. Thus, the *phoC* −10 region and spacing appear similar to those of the highly expressed genes, with differences primarily in the −35 region, the region associated with regulation by virtue of homology to the *E. coli pho* box (34).

In four of the five known *Z. mobilis* genes, the coding region is preceded by a relatively long, untranslated leader sequence which is capable of forming numerous stem-and-loop structures. The length of this leader sequence varies from 277 bases in *adhB* (9) to 116 bases in *phoC* and 47 bases in *pdc* (7).

The 5′ end of *phoC* mRNA was quite different from that of the highly expressed genes in terms of AT content (Fig. 7C). In *phoC*, 13 of the first 15 bases were A or T (87%). The highly expressed genes contained 33 to 60% A+T in this region. The RNA polymerase initiation complex is thought to include this 5′ transcribed region of genes (32). The high A+T content in this region of *phoC* may reduce the efficiency of transcriptional initiation.

Comparative expression in *E. coli*. In this and previous studies, each of the sequenced *Z. mobilis* genes has been expressed in *E. coli* under the control of the *lac* promoter (5, 7–9). Recently, *Z. mobilis trpA* and *trpB* genes have also been expressed in *E. coli* under the control of the *lac* promoter (T. Conway, C. K. Eddy, K. F. Mackenzie, J. P. Mejia, J. L. Pond, E. A. Utt, and L. O. Ingram, Abstr. Annu. Meet. Am. Soc. Microbiol. 1988, H138, p. 168). In all cases, the insertion of highly expressed *Z. mobilis* genes into *E. coli* resulted in the production of high levels of new protein which stained as dense bands in SDS-polyacrylamide gels. In contrast, no new bands were evident in *E. coli* recombinants containing *Z. mobilis phoC*, *trpA*, and *trpB*. Thus, highly expressed *Z. mobilis* genes are also highly expressed in *E. coli*, even from a surrogate promoter.

ACKNOWLEDGMENTS

We appreciate the advice of B. J. Bachmann and B. L. Wanner in the selection of the acid phosphatase gene designation.

These studies were supported in part by grants from the National Science Foundation (DMB 8204928), the Department of Energy, Office of Basic Energy Sciences (DE-FG05-86ER13575), the Department of Agriculture Alcohol Fuels Program (86-CRCR-1-2134), and the Florida Agricultural Experiment Station.

LITERATURE CITED

1. **Allf-Steinberger, C.** 1984. Evidence for coding pattern on the non-coding strand of the *E. coli* genome. Nucleic Acids Res. **12**:2235–2241.
2. **Bennetzen, J. L., and B. D. Hall.** 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.
3. **Bergman, L. W., D. C. McClinton, S. L. Madden, and L. H. Preis.** 1986. Molecular analysis of the DNA sequences involved in the transcriptional regulation of the phosphate-repressible acid phosphatase gene (*PHO5*) of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **83**:6070–6074.
4. **Chang, C. N., W.-J. Kuang, and E. Y. Chen.** 1986. Nucleotide sequence of the alkaline phosphatase gene of *Escherichia coli*. Gene **44**:121–125.
5. **Conway, T., and L. O. Ingram.** 1988. Phosphoglycerate kinase from *Zymomonas mobilis*: cloning, sequencing, and localization within the *gap* operon. J. Bacteriol. **170**:1926–1933.
6. **Conway, T., Y. A. Osman, and L. O. Ingram.** 1987. Gene

expression in *Zymomonas mobilis*: promoter structure and identification of membrane anchor sequences forming functional LacZ' fusion proteins. J. Bacteriol. **169**:2327–2335.
7. **Conway, T., Y. A. Osman, J. I. Konnan, E. M. Hoffmann, and L. O. Ingram.** 1987. Promoter and nucleotide sequences of the *Zymomonas mobilis* pyruvate decarboxylase. J. Bacteriol. **169**: 949–954.
8. **Conway, T., G. W. Sewell, and L. O. Ingram.** 1987. Glyceraldehyde-3-phosphate dehydrogenase gene from *Zymomonas mobilis*: cloning, sequencing, and identification of promoter region. J. Bacteriol. **169**:5653–5662.
9. **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.
10. **Cox, G. B., H. Rosenberg, J. A. Downie, and S. Silver.** 1981. Genetic analysis of mutants affected in the *pst* inorganic phosphate transport system. J. Bacteriol. **148**:1–9.
11. **Dvorak, H. F., R. W. Brockman, and L. A. Heppel.** 1967. Purification and properties of two acid phosphatase fractions isolated from osmotic shock fluid of *E. coli*. Biochemistry **6**:1743–1751.
12. **Elliott, S., C. Chang, M. E. Schweingruber, J. Schaller, E. E. Rickli, and J. Carbon.** 1986. Isolation and characterization of the structural gene for secreted acid phosphatase from *Schizosaccharomyces pombe*. J. Biol. Chem. **261**:2936–2941.
13. **Grantham, R., P. Perrin, and D. Monchiroud.** 1986. Patterns in codon usage of different kinds of species. Oxford Surv. Evol. Biol. **3**:48–81.
14. **Hoekema, A., R. A. Kastelein, M. Vasser, and H. A. De Boer.** 1987. Codon replacement in the *PGK1* gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Mol. Cell. Biol. **7**:2914–2924.
15. **Hoffman, Y. F., and A. Wright.** 1988. Alkaline phosphatase as a tool for analysis of protein secretion, p. 78–82. *In* A. Torriani-Gorini, F. G. Rothman, S. Silver, A. Wright, and E. Yagil (ed.), Phosphate metabolism and cellular regulation in microorganisms. American Society for Microbiology, Washington, D.C.
16. **Ikemura, T.** 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151**:389–409.
17. **Ikemura, T.** 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol. **158**:573–597.
18. **Kam, W., E. Clauser, Y. S. Kim, Y. W. Kan, and W. J. Rutter.** 1985. Cloning, sequencing, and chromosomal localization of human term placental alkaline phosphatase cDNA. Proc. Natl. Acad. Sci. USA **82**:8715–8719.
19. **Kaneko, Y., N. Hayashi, A. Toh-e, I. Banno, and Y. Oshima.** 1987. Structural characteristics of the *PHO8* gene encoding repressible alkaline phosphatase in *Saccharomyces cerevisiae*. Gene **58**:137–148.
20. **Kendall, D. A., S. C. Bock, and E. T. Kaiser.** 1986. Idealization of the hydrophobic segment of the alkaline phosphatase signal peptide. Nature (London) **321**:706–708.
21. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
22. **Luria, S. E., and M. Delbrück.** 1943. Mutations of bacteria from virus sensitivity to virus resistance. Genetics **28**:491–511.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
25. **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. Benjamin-Cummings Publishing Co., Menlo Park, Calif.
26. **Osman, Y. A., T. Conway, S. J. Bonetti, and L. O. Ingram.** 1987. Glycolytic flux in *Zymomonas mobilis*: enzyme and metabolite levels during batch fermentation. J. Bacteriol. **169**:3726–3736.
27. **Parent, S. A., A. G. Tait-Kamradt, J. LeVitre, O. Lifanova, and K. A. Bostian.** 1987. Regulation of the phosphatase multigene family of *Saccharomyces cerevisiae*, p. 63–70. *In* A. Torriani-Gorini, F. G. Rothman, S. Silver, A. Wright, and E. Yagil (ed.), Phosphate metabolism and cellular regulation in microorganisms. American Society for Microbiology, Washington, D.C.
28. **Pawluk, A., R. K. Scopes, and K. Griffiths-Smith.** 1986. Isolation and properties of the glycolytic enzymes from *Zymomonas mobilis*. Biochem. J. **238**:275–281.
29. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
30. **Pustell, J., and F. C. Kafatos.** 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis, and homology determination. Nucleic Acids Res. **12**:643–655.
31. **Randall, L. L., and S. J. S. Hardy.** 1984. Export of protein in bacteria. Microbiol. Rev. **48**:290–298.
32. **Reznikoff, W. S., and W. R. McClure.** 1986. *E. coli* promoters, p. 1–34. *In* W. Reznikoff and L. Gold (ed.), Maximizing gene expression. Butterworths Publishers, Stoneham, Mass.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
34. **Shinagawa, H., K. Makino, M. Amemrua, and A. Kakata.** 1987. Structure and function of the regulatory genes for the phosphate response n *Escherichia coli*, p. 20–25. *In* A. Torriani-Gorini, F. G. Rothman, S. Silver, A. Wright, and E. Yagil (ed.), Phosphate metabolism and cellular regulation in microorganisms. American Society for Microbiology, Washington, D.C.
35. **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 **71**:1342–1346.
36. **Sidhu, R. S., and A. P. Bollon.** Analysis of α-factor secretion signals by fusing with acid phosphatase of yeast. Gene **54**: 175–184.
37. **Stormo, G. D.** 1986. Translational initiation, p. 195–224. *In* W. Reznikoff and L. Gold (ed.), Maximizing gene expression. Butterworths Publishers, Stoneham, Mass.
38. **Terao, M., and B. Mintz.** 1987. Cloning and characterization of a cDNA coding for mouse placental alkaline phosphatase. Proc. Natl. Acad. Sci. USA **84**:7051–7055.
39. **Touati, E., and A. Danchin.** 1987. The structure of the promoter and amino terminal region of the pH 2.5 acid phosphatase gene (*appA*) of *E. coli*: a negative control of transcription mediated by cyclic AMP. Biochimie **69**:215–221.
40. **Wanner, B. L.** 1987. Bacterial alkaline phosphatase gene regulation and the phosphate response in *Escherichia coli*, p. 12–19. *In* A. Torriani-Gorini, F. G. Rothman, S. Silver, A. Wright, and E. Yagil (ed.), Phosphate metabolism and cellular regulation in microorganisms. American Society for Microbiology, Washington, D.C.
41. **Wanner, B. L., M. R. Wilmes, and D. C. Young.** 1988. Control of bacterial alkaline phosphatase synthesis and variation in an *Escherichia coli* K-12 *phoR* mutant by adenyl cyclase, the cyclic AMP receptor protein, and the *phoM* operon. J. Bacteriol. **170**:1092–1102.
42. **Zucker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.