JOURNAL OF BACTERIOLOGY, Mar. 1984, p. 909–917
0021-9193/84/030909-09$02.00/0

Vol. 157, No. 3

Nucleotide Sequence of the *phoS* Gene, the Structural Gene for the Phosphate-Binding Protein of *Escherichia coli*

KOJI MAGOTA,[1] NOZOMU OTSUJI,[1] TAKEYOSHI MIKI,[1] TADAO HORIUCHI,[1] SUSUMU TSUNASAWA,[2] JUN KONDO,[2] FUMIO SAKIYAMA,[2] MITSUKO AMEMURA,[3] TAKASHI MORITA,[3] HIDEO SHINAGAWA,[3] AND ATSUO NAKATA[3]*

Faculty of Pharmaceutical Sciences, Kyushu University, Fukuoka, Japan 812,[1] and the Institutes for Protein Research[2] and Microbial Diseases,[3] Osaka University, Yamadaoka, Suita, Osaka, Japan 565

Received 28 November 1983/Accepted 28 November 1983

phoS is the structural gene for the phosphate-binding protein, which is localized in periplasm and involved in active transport of phosphate in *Escherichia coli*. It is also a negative regulatory gene for the *pho* regulon, and the gene expression is inducible by phosphate starvation. The complete nucleotide sequence of the *phoS* gene was determined by the method of Maxam and Gilbert (A. M. Maxam and W. Gilbert, Methods Enzymol. **65**:499–560, 1980). The amino acid sequences at the amino termini of the pre-PhoS and PhoS proteins and at the carboxy terminus of the PhoS protein were determined by using the purified proteins. Furthermore, the amino acid sequence of enzymatically digested peptide fragments of the PhoS protein was determined. The combined data established the nucleotide sequence of the coding region and the amino acid sequence of the pre-PhoS and the PhoS proteins. The pre-PhoS protein contains an extension of peptide composed of 25 amino acid residues at the amino terminus of the PhoS protein, which has the general characteristics of a signal peptide. The mature PhoS protein is composed of 321 amino acid residues, with a calculated molecular weight of 34,422, and lacks the disulfide bond and methionine. The regulatory region of *phoS* contains a characteristic Shine-Dalgarno sequence at an appropriate position preceding the translational initiation site, as well as three possible Pribnow boxes and one −35 sequence. The nucleotide sequence of the regulatory region of *phoS* was compared with those of *phoA* and *phoE*, the genes constituting the *pho* regulon.

Studies on the phosphate-binding protein (PBP) of *Escherichia coli* originated from the investigation of the regulatory mechanism of alkaline phosphatase (APase; EC 3.1.3.1) synthesis and of the transport system of inorganic phosphate. In the early genetic studies on the synthesis of APase, the regulatory mutations were mapped at two unlinked regions, R1 and R2, in addition to the P (*phoA*) gene, the structural gene for APase (11). Mutations in the R2 region, which produced APase in media containing high concentrations of inorganic phosphate, were subdivided into two classes: one class produced a periplasmic protein named R2a, and the other did not (16). The synthesis of R2a protein, later identified as PBP(17), was repressed under the same genetic and physiological conditions as those of APase. The mutations in R1 region were subdivided into three classes: $R1^a$, $R1^b$, and $R1^c$ (14, 15). The $R1^c$ mutant did not synthesize APase even under phosphate starvation, although the $R1^a$ and $R1^b$ mutants produced APase and R2a protein constitutively. Another APase-negative mutant was isolated (8). The *phoB* mutant, originally designated as *phoT*, did not produce APase and the R2a protein, whereas the R2a protein was produced in the $R1^c$ mutant (48). $R1^c$ and *phoB* were proved to be mutations in the same cistron (35), and both of them were designated as *phoB*. *phoB* ($R1^c$) and *phoR* ($R1^a$ and $R1^b$) were mapped at 9 min on the *E. coli* standard genetic map (5, 22). Thus, it is concluded that the synthesis of APase and PBP is subject to a common regulatory mechanism, and the *phoA* and *phoS* genes constitute the *pho* (phosphate) regulon.

The PBP was isolated from osmotic shock fluid, and a mutant lacking this protein was defective in uptake of inorganic phosphate (28, 29). The protein was later shown to be identical with the R2a protein (17). A *pst* (phosphate-specific transport) mutant with altered inorganic phosphate growth response was isolated (6). It was mapped at the same region as the R2 gene. The R2a and R2b mutants were tested for uptake of inorganic phosphate and shown to be defective in phosphate transport (47). The name of *phoS* (previously used for the R2 gene [3]) was retained for the R2a gene, and the new name *phoT* was proposed for the previous R2b gene. Mutants auxotrophic for an organic phosphate were isolated. One, *pst-2*, was mapped at the locus closely linked to *phoS* and *phoT*, and the other, *pit*, was mapped at an unlinked locus (42). It was shown that the phosphate inorganic transport and phosphate-specific transport systems are two major transport systems for inorganic phosphate in *E. coli* (37, 42). Mutations in any of the phosphate-specific transport genes resulted in more or less constitutive synthesis of APase. Another *pst* mutant, *pst-401*, was isolated (9). On the basis of a complementation test for the repression of APase synthesis, it was proposed to designate *pst-2* as *pstA* and the new *pst* mutant (*pst-401*) as *pstB*. Thus, the phosphate-specific transport system includes the genes *phoS*, *phoT*, *pstA*, and *pstB*, all of which were mapped at 84 min on the genetic map.

In previous work (1), we cloned an *E. coli* chromosomal DNA fragment containing the R2 region on plasmid pBR322 and constructed various deletion plasmids in vitro. By using these plasmids for complementation tests, we showed that the *phoS* and *phoT* genes are expressed independently, but *phoT* and *pstA* were unseparable. A mutation formerly designated *phoT35*, which does not cause defective phosphate uptake (50), is in an independent cistron located next to these genes. Therefore, we proposed a new designation,

* Corresponding author.

phoU, for the *phoT35* allele. The work also clarified the previous confusions (26) as to whether *phoS* and *phoT* are allelic or not and the location of the *phoS* promoter (9). Cells carrying only the *phoS* gene on a multiple-copy-number plasmid accumulated a large amount of the PBP in periplasm, as well as a large amount of its precursor protein (pre-PBP) in cytoplasm (30). We utilized this overproduction system for preparation of the PBP and pre-PBP to be used for the present studies on the primary structure of the proteins. Since *phoS* constitutes the *pho* regulon, together with *phoA* and *phoE*, which are under common positive regulation by the *phoB* gene product (4, 34, 39, 45, 46), it is of interest to compare the structures of the regulatory regions of these genes to find a common feature underlying the *pho* regulon.

MATERIALS AND METHODS

Bacterial strains and plasmids. *E. coli* ANCC75 (*leu purE trp his argG rpsL phoS64 metA* or *metB thi*) carrying plasmid pSN518 (for DNA sequencing) or pSN5182 (for preparation of PBP and pre-PBP) was used. Both plasmids carry the functional *phoS* gene of *E. coli* K-12 and were described previously (1, 30).

Enzymes and chemicals. Restriction enzymes used in this study were purchased from New England Biolabs, Inc., or Takara Shuzo Co., Ltd. T4 polynucleotide kinase was purchased from Takara Shuzo Co., Ltd. Each enzyme reaction was performed as recommended by the suppliers. Tolylsulfonyl phenylalanyl chloromethyl ketone-trypsin (3× recrystallized) was purchased from Worthington Diagnostics, and *Staphylococcus aureus* V8 protease was from Miles Laboratories. Chemicals used for sequencing proteins were analytical grade. [γ-^{32}P]ATP (specific activity, 3 Ci/nmol) was purchased from Amersham Japan Co., Ltd.

Preparation of PBP and pre-PBP. PBP was prepared as described previously (30). For preparation of pre-PBP, the conditions for growing cells and the cold osmotic shock procedure were described previously (30). Starting from 1.2 cell culture, the cellular fraction, from which the periplasmic proteins had been removed by cold osmotic shock, was washed once in 200 ml of membrane buffer (10 mM Tris-hydrochloride [pH 8.5], 5 mM EDTA, 5 mM 2-mercaptoethanol) (40) containing 0.2 M KCl. It was suspended in 30 ml of the same buffer containing 0.2 M KCl and 1% (vol/vol) Triton X-100 and sonicated. The cell debris and undisrupted cells were removed by centrifugation at 7,000 × *g* for 10 min. The supernatant was then centrifuged at 105,000 × *g* for 2 h. The precipitate was suspended in 30 ml of the same buffer and sonicated. The final suspension was solubilized by adding sodium dodecyl sulfate (SDS; final concentration of 1%), and then it was dialyzed against 0.1% SDS. It was applied on a Sephadex G-100 column (2 by 45 cm) previously equilibrated with 0.1% SDS. The proteins were eluted with 0.1% SDS. Each fraction was assayed by SDS-polyacrylamide gel electrophoresis, and the pre-PBP fractions were collected. Pre-PBP thus obtained (about 3.5 mg of protein in 7 ml of buffer) was lyophilized, dissolved in 70% formic acid (2 ml), and dialyzed against the formic acid several times to remove the SDS (10). The dialyzed protein gave a single band by electrophoresis on SDS-polyacrylamide gel and was lyophilized (3 mg).

Amino acid analyses. Amino acid analyses were performed on a Hitachi KLA-5 amino acid analyzer after hydrolysis of the protein samples in 4 M methanesulfonic acid with 0.2% indolethylamine (41) at 110°C for 24 h. For determination of amino acid composition of PBP, the analysis was also performed for 48- and 72-h hydrolysates. Peptides separated by high-pressure liquid chromatography were analyzed with a Hitachi 835S amino acid analyzer after they were hydrolyzed in 6 M HCl at 110°C for 24 h.

Protease digestion of PBP and separation of peptide mixture by reverse-phase high-pressure liquid chromatography. PBP (1.5 mg) was digested with tolylsulfonyl phenylalanyl chloromethyl ketone-treated trypsin (the ratio of substrate to enzyme was 100:1, by weight) in 1% NH_4HCO_3 (1 ml) at 37°C for 6 h. After being heated at 100°C for 10 min to inactivate the trypsin, the digest was further incubated with *S. aureus* V8 protease (the ratio of substrate to enzyme was 50:1, by weight) at 37°C overnight. Peptide mixture thus obtained was lyophilized, dissolved in 0.1% trifluoroacetic acid (100 μl), and chromatographed on a Toyo Soda LS-410K reverse-phase column (0.4 by 30 cm). Chromatography was performed by increasing linearly the concentration of acetonitrile from 0 to 60% in 0.1% trifluoroacetic acid for 1 h at 37°C with a Yanaco L-2000 liquid chromatograph, which was equipped with a M215 variable wavelength UV detector, a

TABLE 1. Amino acid composition of PBP and pre-PBP

Amino acid	PBP		pre-PBP[a]	
	Amino acid analysis[b]	From DNA sequence	Amino acid analysis[b]	From DNA sequence
Lys	31.0	31	32.0	32
His	1.0	1	3.2	1
Arg	4.2	4	10.2	5
Trp	8.0	8	7.1	8
Asp	36.6	20	37.7	20
Asn		17		17
Thr	20.4[c]	21	22.7	25
Ser	18.7[c]	19	19.2	22
Glu	28.8	15	33.3	15
Gln		14		14
Pro	17.1	15	15.6	15
Gly	33.5	34	36.4	34
Ala	35.0	35	40.0	40
Cys	0	0	0	0
Val	23.0[d]	22	26.8	27
Met	0	0	4.8	3
Ile	16.0[d]	17	17.6	17
Leu	24.2	24	29.1	25
Tyr	12.0	12	11.9	12
Phe	12.5	12	14.0	14

[a] Values listed are from one determination.
[b] Calculated by assuming the number of lysine residues as 31 for PBP and as 32 for pre-PBP.
[c] Corrected for destruction during hydrolysis.
[d] Values are from a 72-h hydrolysate.

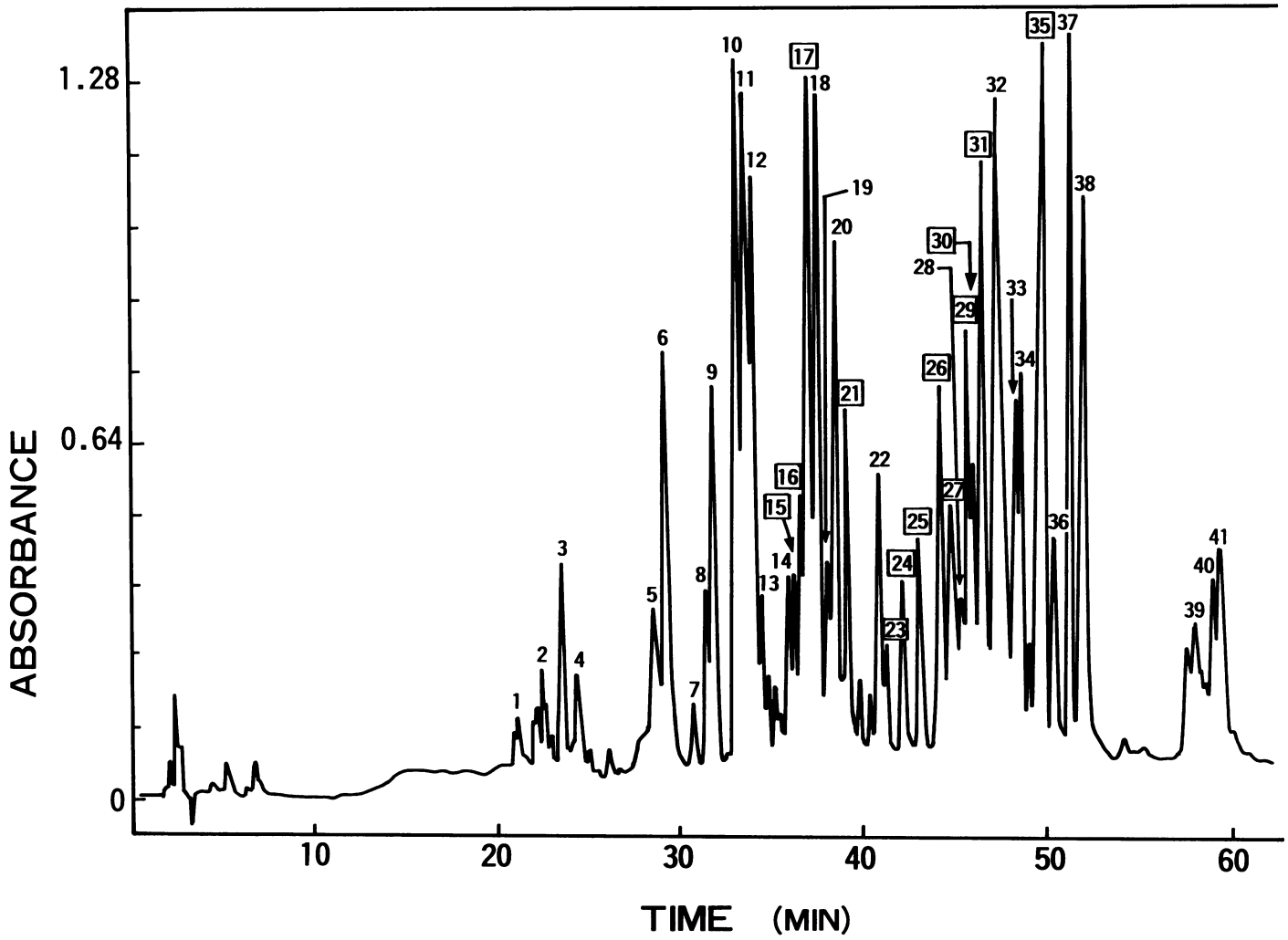

FIG. 1. Reverse-phase chromatography of protease digest of PBP. Peptides obtained by successive digestion of PBP with trypsin and staphylococcal protease were chromatographed on a Toyo Soda LS410K column (0.4 by 30 cm). Conditions for chromatography were described in the text. One-fifth of the eluent collected for each peak was submitted to amino acid analysis. The peaks with boxed numbers contained peptides whose amino acid sequence was analyzed by Edman degradation.

TABLE 2. Amino acid sequences of 13 peptides obtained from PBP by successive digestion with trypsin and *S. aureus* V8 protease

Peak no.	Amino acid sequence[a]	Yield (%)	No. of residues in the primary structure of PBP
15	Leu Asn Pro Gly Leu Lys	70.8	119–124
16	Val Asn Glu Glu Trp Lys	15.6	152–157
17	Trp Ala Asp Thr Tyr Gln Lys	71.8	19–25
21	Lue Ile Ser Ala Asp Gly Lys Pro Val Ser Pro Thr Glu Glu	91.0	209–223
23	Thr Phe Ala Gln Asp Leu Thr Asn Gln Lys	24.2	236–245
24	Leu Pro Ser Gln Asn Ile Ala Val Val Arg	55.6	125–134
25	Thr Val Asp Phe Gly Ala Ser Asp Ala Pro Leu Ser Asp Glu Lys Leu Ala Gln	34.6	49–66
26	Tyr Ala Ser Leu Pro Asp Ser Val Val Glu	50.8	293–302
27	Gly Asn Asp Gly Ile Ala Ala Phe Val Gln Arg	80.0	176–186
29	Thr Leu Gly Asp Ile Tyr Leu Gly Lys	62.0	99–107
30	Leu Pro Gly Ala Ile Gly Tyr Val Glu	30.4	187–195
31	Glu Ala Ser Leu Thr Gly Ala Gly Ala Thr Phe Pro Ala Pro Val Tyr Ala Lys	83.4	1–18
35	Phe Phe Asp Trp Ala Tyr Lys	61.2	276–282

[a] Amino acids determined by Edman degradation are shown by underlining.

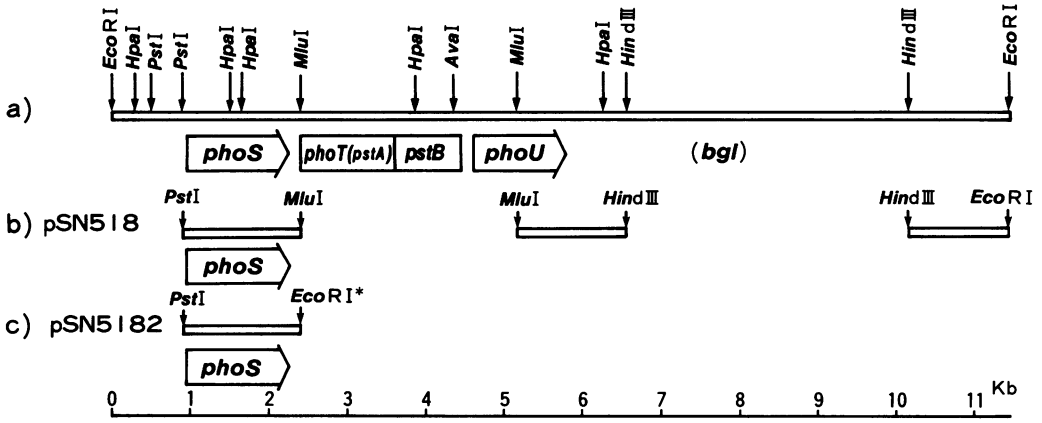

FIG. 2. (a) Restriction map of *E. coli* chromosomal DNA fragment carried on plasmid pSN401 (1). The restriction sites on the 3.6-kb *Hin*dIII fragment were not examined. The gene order located in this region was indicated (see the text). (b) Plasmid pSN518, in which the *Hin*dIII, *Mlu*I, and *Pst*I fragments of pSN401 were deleted. (c) Plasmid pSN5182, in which the *Mlu*I-*Eco*RI fragment of pSN518 was removed, and the protruding single-stranded ends were filed and ligated. The ligated sequence generated a new *Eco*RI site indicated by the asterisk.

2000G solvent programmer, and a Rhenohyne sample injector (100 μl). The flow rate as 1 ml/min. Peptides were detected by monitoring absorbance at 215 nm, and the eluent peaks were collected manually.

Determination of amino acid sequence. The N-terminal amino acid sequences of PBP (50 nmol) and pre-PBP (10 nmol) were determined by manual Edman degradation (12). A smaller amount of samples (5 to 13 nmol) was used for partial sequencing of peptides obtained by protease digestion. The C terminus of PBP (25 nmol) was determined by hydrazinolysis.

Amino acid phenylthiohydantoins were identified by a modification of the method of Zimmerman et al. (49) with a Zorbax CN column (0.46 by 25 cm), a buffer system of methanol-acetonitrile (85:15) and sodium acetate, and an eight-channel stepwise elution system.

Purification of plasmid DNA. Plasmid DNA was purified from strain ANCC75 (pSN518) by the cleared lysate procedure, followed by cesium chloride-ethidium bromide centrifugation (24).

DNA sequencing. Plasmid pSN518 DNA was purified and digested with the restriction endonucleases *Pst*I and *Mlu*I. The resultant fragments were separated by electrophoresis on a 3.5% (wt/vol) polyacrylamide gel, and the 1.5-kilobase (kb) DNA fragment was extracted from the gel by electroelution. Nucleotide sequences were determined according to the method of Maxam and Gilbert (27).

Computer analysis. Computer analysis of the DNA sequence was performed by using the programs developed by Staden (43), modified by Kuhara et al. (23).

RESULTS

Determination of partial amino acid sequences of PBP and pre-PBP by Edman degradation. Amino acid compositions of PBP and pre-PBP are presented in Table 1. The values obtained by amino acid analysis of PBP are mostly consistant with those calculated from DNA sequencing data. In pre-PBP, a discrepancy in the values obtained by the above two analyses is found for arginine, histidine, glutamic acid, methionine, and leucine. This may be due to contamination by a proteinacious material(s) which was difficult to remove by reverse-phase high-pressure liquid chromatography from pre-PBP solubilized with detergent. However, since the N-terminal eight amino acid residues were unequivocally determined by Edman degradation, the amount of contaminant appears to be very small. A comparison of the amino acid compositions of the mature and precursor proteins shows that all three methionine residues in pre-PBP are present in the signal peptide because PBP lacks this amino acid.

The amino acid sequence determined by direct Edman degradation of pre-PBP was Met-Lys-Val-Met-Arg-Thr-(X)-Val-Ala. Although the amino acid in parentheses was not identified, the elucidation of this sequence was critically important in determining the translational initiation site of the *phoS* gene. Accordingly, we concluded that the ATG which is five nucleotides downstream of a possible ribosome-binding sequence (AGGAG) is the translational initiation site.

Edman degradation determined the N-terminal 21 amino acids of PBP, except for the four unidentified residues: Glu-Ala-Ser-Leu-Thr-Gly-Ala-Gly-Ala-(X)-Phe-(X)-Ala-(X)-Val-Tyr-Ala-Lys-(X)-Ala-Asp. This amino acid sequence is identical to that deduced from the DNA sequence of the *phoS* gene.

The protein was successively digested with trypsin and *S. aureus* V8 protease, which cleave preferentially at arginine-lysine and glutamic acid-aspartic acid, respectively, and the digest was fractionated on a reverse-phase high-pressure liquid chromatography column (C18) (Fig. 1). Among about

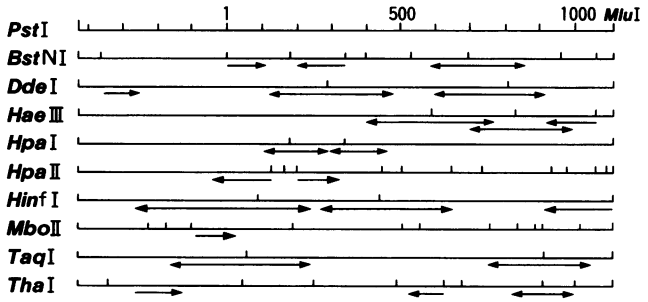

FIG. 3. Restriction nuclease map and sequencing strategy of the *Pst*I-*Mlu*I 1.5-kb fragment. The arrows indicate the extent of the nucleotide sequence determined and are aligned in 5′ → 3′ direction.

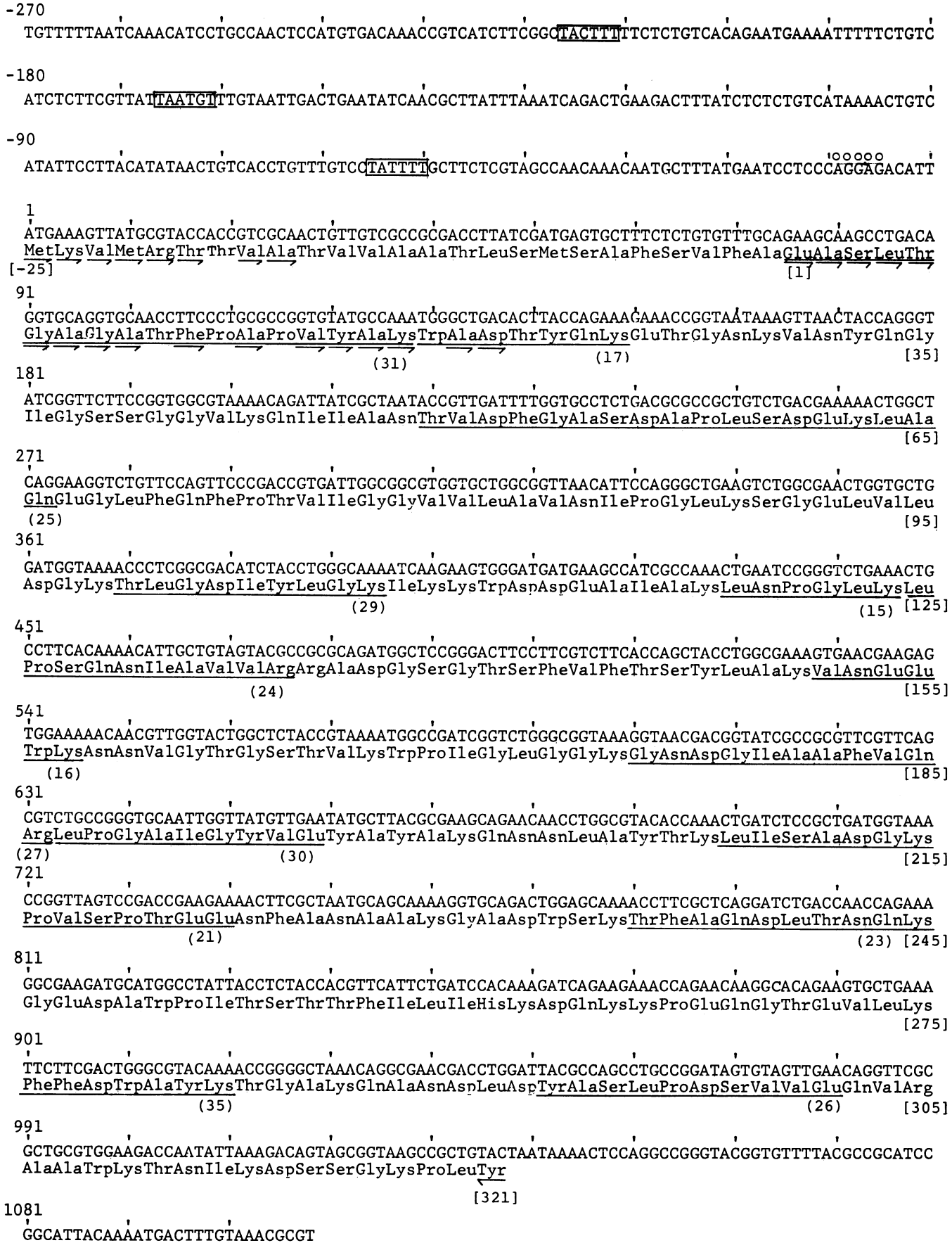

FIG. 4. DNA nucleotide sequence of the *phoS* gene and primary structure of its product, the PBP. The nucleotides are numbered taking the first nucleotide of the translational initiation codon as 1. The amino acids are numbered taking the N-terminal amino acid of the mature protein as [1]. Three possible promoters are boxed. Open circles represent the ribosome-binding site (Shine-Dalgarno sequence). The analyzed amino acid sequences of the N termini in the mature protein and its precursor protein are denoted by arrows. The analyzed 13 polypeptides that appear in Table 2 are underlined, and the peptide numbers are in parentheses.

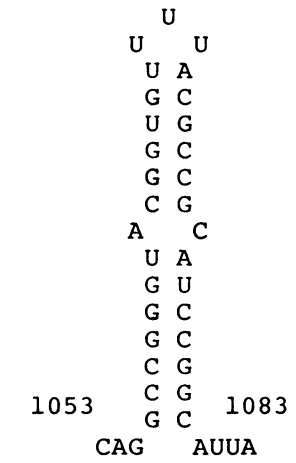

FIG. 5. Possible secondary structure of the mRNA in the 3′ end noncoding region of the *phoS* gene. Numbers of the corresponding nucleotides in the DNA sequences in Fig. 4 are indicated.

40 peaks produced, 13 were shown to be pure by amino acid analysis, and their amino acid sequences were examined. The result of the sequence analysis is summarized in Table 2. The structures of all 13 peptides deduced from the data obtained by Edman degradation and amino acid composition analysis are found in various regions of the primary structure of PBP deduced from the nucleotide sequence of the *phoS* gene. Additionally, hydrazinolysis (32) gave tyrosine as the C terminus of PBP. The results obtained by protein sequencing are consistent with those obtained by DNA sequencing in all respects.

Physical localization of the *phoS* gene. A chromosomal DNA fragment which carries the *phoS* gene has been cloned on pBR322, and localization of the gene on the cloned 11.5-kb *Eco*RI DNA fragment was determined by constructing various deletion derivatives of pSN401 (1, 30). The restriction nuclease map of the *phoS* gene and the neighboring genes related to the *pho* region is shown in Fig. 2. The functional *phoS* region was located within the 1.5-kb *Pst*I-*Mlu*I DNA fragment, since both plasmids pSN518 and pSN5182 complemented *phoS* mutants and the DNA fragment carried on them directed synthesis of both pre-PBP and PBP which was under phosphate regulation (30). The transcriptional and translational direction of the *phoS* gene was determined by constructing *phoS′-′lacZ* gene fusions and was found to be counterclockwise on the standard circular *E. coli* genetic map (Fig. 2). (M. Amemura et al., unpublished data).

DNA nucleotide sequence. Using the chemical method of Maxam and Gilbert (27), we determined the complete nucleotide sequence of the regulatory and protein-coding regions of the *phoS* gene. The *Pst*I-*Mlu*I 1.5-kb fragment carried on pSN518 was used for the nucleotide sequencing. The restriction nuclease map and the sequencing strategy are shown in Fig. 3. In this figure, the adenine of the first methionine codon (ATG) of the pre-PBP is numbered as 1 (see below). The sequence of the regulatory region (nucleotides −1 to −300) was determined for both strands or for only one of the strands but with different overlapping fragments so that they can cover each region at least twice. All the restriction nuclease sites used for sequencing were determined by using overlapping fragments so that no information was lost as a result of small restriction fragments being unnoticed. The established DNA nucleotide sequence is shown in Fig. 4.

The DNA sequence contains two open translation reading frames whose directions are opposite. One of the open reading frames is initiated from the site in the downstream region not shown in Fig. 4 and ends at nucleotide 454. The second coding frame (nucleotides 1 through 1,038), which codes for a polypeptide with 346 amino acid residues and a molecular weight of 37,030, is preceded by a typical ribosome-binding sequence (AGGAG). From the reported molecular weight of pre-PBP (39,000 [30]) and the direction of transcription, the second open reading frame most likely codes for pre-PBP.

This is firmly established by the complete matching of the amino acid sequences in amino-termini of PBP and pre-PBP, and an amino acid in the carboxy end of PBP actually determined (see above).

Regulatory sequences and transcription of *phoS*. Three possible RNA polymerase binding sequences (TANNNT) (19), which are known as Pribnow boxes, can be found (nucleotides) −56 to −51, −167 to −162, and −217 to −212) in the region (nucleotides −1 to −250) preceding the translational start point (Fig. 4). Sixteen nucleotides upstream of the third possible Pribnow sequence (TACTTT, nucleotides −217 to −212), a possible RNA polymerase recognition site (GTGACA, nucleotides −239 to −234) is found. But at the appropriate positions preceding the other two Pribnow consensus sequences, no RNA polymerase recognition sequence is found.

When we looked for the possible transcription termination signal of *phoS* in the nucleotide sequence, a region of mRNA corresponding to the sequence of the 3′ end noncoding region was found to be able to form a rather stable stem-and-loop structure with $\Delta G = -17.4$ kcal (−72.8 kJ) (7, 44) (nucleotides 1,053 to 1,083) (Fig. 5). Since the *phoT* gene, which is located immediately downstream of the *phoS* gene, possesses its own promoter and is transcribed in the same direction as the *phoS* gene (see below), the *phoS* gene should have the transcription termination signal downstream of the coding region. It is likely that this stem-and-loop structure can serve as a termination signal of the *phoS* transcription (36). However, this RNA stem-and-loop structure is not followed by polyuridylate sequence, which is a well-known feature of rho-independent terminators.

DISCUSSION

Common features of the regulatory region of the *pho* genes. The expression of the *phoS* gene, as well as the *phoA* and *phoE* genes, is under the positive control of the *phoB* gene product (4, 34, 39, 45, 46). Therefore, these genes may have a common feature in their promoter-operator regions, with

```
      -106                          -56
phoS    CTGTCATAAAACTGTCA---33 bp---TATTTT
       -80                           -28
phoA    CTGTCATAAAGTTGTCA---35 bp---TAATGT
       -99                           -47
phoE    CTGTAATATATCTTTAA---35 bp---GACTCT

        CTGTcATAaA-cTgTcA            Pribnow
                                      sequence
```

FIG. 6. Homologies in the nucleotide sequences of the regulatory regions of *phoS*, *phoA* (20, 21), and *phoE* (33). The first nucleotide preceding the translational initiation codon is numbered as −1 for each gene. In the last line, common nucleotides shared by the three genes are shown by capital letters, and common nucleotides shared by the two genes are shown by lower-case letters.

```
phoA    GTGAAACAAAGCACTATTGCACTGGCACTCTTACCGTTACTGTTTACCCCTGTGACAAAAGCC
        MetLysGlnSerThrIleAlaLeuAlaLeuLeuProLeuLeuPheThrProValThrLysAla

phoE    ATGAAAAAGAGCACTCTGGCATTAGTGGTGATGGGCATTGTGGCATCTGCATCTGTACAGGCT
        MetLysLysSerThrLeuAlaLeuValValMetGlyIleValAlaSerAlaSerValGlnAla

phoS    ATGAAAGTTATGCGTACCACCGTCGCAACTGTTGTCGCCGCGACCTTATCGATGAGTGCTTTCTCTGTGTTTGCA
        MetLysValMetArgThrThrValAlaThrValValAlaAlaThrLeuSerMetSerAlaPheSerValPheAla
```

FIG. 7. Comparison of the signal sequences and the corresponding nucleotide sequences of the three *pho* regulon genes. Two N-terminal amino acids and one amino acid at the C end of the signal peptides are identical in the three gene products (double underlined). Dipeptides found in any of the two proteins are underlined. The data about *phoA* (20, 21) and *phoE* (33) were previously reported.

which the *phoB* gene product interacts. To obtain further information about sequences possibly involved in the regulation, the 200-nucleotide sequence of *phoS* before the translation starting point was compared with those of *phoA* and *phoE* (20, 21, 33). A remarkable common sequence (CTGTCATAAA—TGTCA, nucleotides −106 to −90 in *phoS* and −80 to −64 in *phoA*, respectively) was found in both the *phoS* and *phoA* genes (Fig. 6). In each gene, the sequence is followed by a sequence homologous to the Pribnow box (19, 36) at a distance of about 35 nucleotides (TATTTT, nucleotides −56 to −51 in *phoS*; TAATGT, nucleotides −28 to −23 in *phoA*). In the *phoE* gene, a similar sequence (CTGTAATATATCTTTAA, nucleotides −99 to −83) had also been detected. In addition, a sequence (GACTCT, nucleotides −57 to −52) weakly homologous to the Pribnow box exists 35 nucleotides downstream of this sequence (Fig. 6). It is possible that this homologous sequence interacts with the *phoB* gene product. It should be mentioned that in the regulatory region of the *phoS* gene, three repetitive sequences (CTGTCATA) were also detected (Fig. 4).

Overbeeke et al. (33) pointed out three regions of homologous nucleotide sequence in the regulatory regions of *phoA* and *phoE*. However, when we include the regulatory region of *phoS* for comparison, the regions either have less homology or their relative distance to the presumptive Pribnow boxes are quite different. Further work should be done before we can make significant conclusions about the common regulatory mechanism shared by the genes belonging to the *pho* regulon. To correlate the function and the structure in the regulatory region of the *pho* genes, isolation and sequence analyses of more regulatory mutants of *phoS*, *phoA*, and *phoE* may be required. We are preparing in vitro transcription experiments, using the *phoS* and the *phoA* genes cloned on the vectors as templates, to see the stimulatory effect of the PhoB protein. This line of experiments may contribute to the understanding of the regulatory mechanism of the *pho* regulon.

Signal peptide of PBP. The amino terminus of pre-PBP possesses an extra 25-amino-acid-long peptide, as compared with the mature PBP (Fig. 4 and 7). This amino acid sequence has common characteristics with those of other bacterial signal peptides (13, 18). The amino-terminal region is basic with lysine and arginine (residues −24 and −21). There is a stretch of hydrophobic amino acids (residues −15 to −10). The residue at the cleavage site (residue −1) is alanine. This sequence is consistent with the loop model proposed by Halegoua and Inouye (18). The signal sequence of PBP was compared with those of APase and the PhoE protein (Fig. 7). Although considerable homology between the signal peptides of APase and the PhoE protein is seen, the signal peptide of PBP has less homology with them.

Structure and function of PBP. Distribution of hydrophobic and hydrophylic portions in the primary structure of PBP

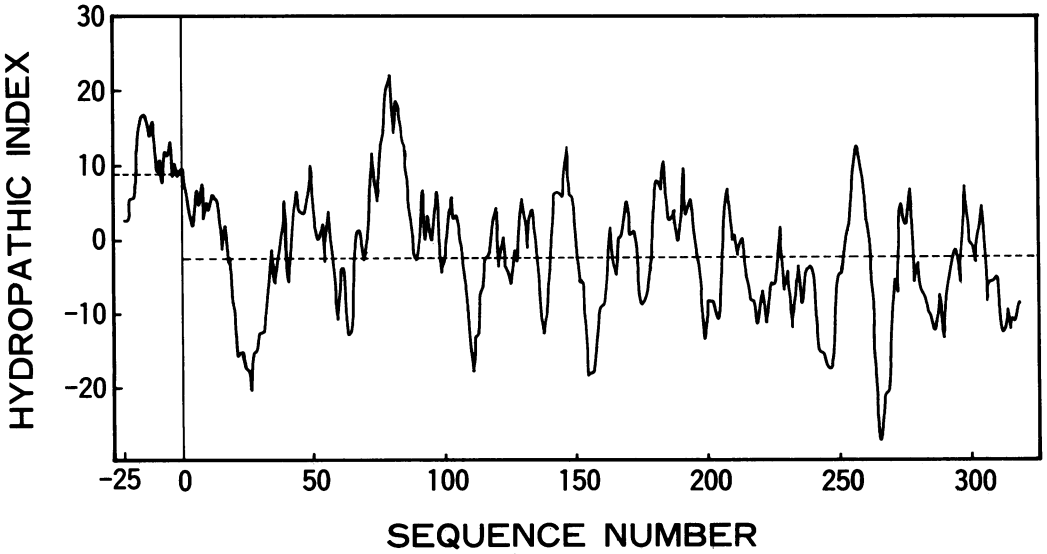

FIG. 8. Hydropathy profile of pre-PBP. Hydropathic values were calculated at a span setting of 8. The dashed line denotes the average hydropathic value.

and pre-PBP was examined based on the hydropathy index of amino acids estimated according to Kyte and Doolittle (25) (Fig. 8). A more hydrophobic portion than the signal peptide, which is also highly hydrophobic, is found in the internal portion around Val at position 80. Some structural roles in the formation and stability of active conformation are suggested for this hydrophobic core of PBP, which contains no disulfide bonds. It has been shown that arginine residues have an important role in enzymes that act on phosphate-containing substrates (2, 31, 38). Among the four arginine residues of PBP, two consecutive arginine residues are localized at positions 109 and 110. Therefore, they could play an important role in the binding of phosphate. Although at the present stage we can only speculate on the structure of PBP in terms of function in phosphate binding, the hyperproduction strain of PBP (30) provides us easy access to the preparation of a large amount of PBP. Furthermore, we are currently conducting physicochemical studies on PBP to elucidate its conformation.

A cluster of the *pho* genes in the vicinity of *phoS*. Several APase constitutive mutants (*phoS* and *phoT*) and phosphate-transport-defective (*pstA* and *pstB*) were isolated and mapped at 84 min on the *E. coli* genetic map in the vicinity of *phoS* (5). There was confusion about the cistrons and operons in this region and their relative orders (9, 26). We examined the *phoS21* strain, which was studied by Cox et al. (9). The results obtained from the complementation tests with hybrid plasmids carrying DNA fragments in this region suggest that the *phoS21* strain is a double mutant, with one mutation in the *phoS* gene and another either in *pstB* or *phoT* having a polar effect on *pstB* expression (data not shown). We speculate that the double-mutant nature of this strain misled Cox et al. (9) to conclude that *phoS* and *pstA* are parts of an operon, in which *pstA* is proximal to the promoter. Our previous work (1, 30) and the present work unequivocally show that the *phoS* gene possesses its own promoter and constitutes a single cistron operon. Evidence obtained by complementation tests with the hybrid plasmids carrying DNA segments of the *phoS-phoU* region and fine physical mapping by Tn*1000* insertion inactivation suggest that *phoT* and *pstA* are allelic, and *phoT* and *pstB* constitute an operon in which *phoT* is upstream of *pstB* (Nakata et al., manuscript in preparation). It is also suggested that the *phoU* gene comprises another operon and the direction of transcription is counterclockwise on the standard genetic map (Nakata et al., in preparation). The gene order and the transcriptional direction of some of the genes in this region are shown in Fig. 2.

ACKNOWLEDGMENTS

We thank T. Sugita for preparing PBP. Many thanks to G. B. Cox, H. Rosenberg, and B. P. Surin for the bacterial strains and for sharing their results with us before publication.

LITERATURE CITED

1. **Amemura, M., H. Shinagawa, K. Makino, N. Otsuji, and A. Nakata.** 1982. Cloning of and complementation tests with alkaline phosphatase regulatory genes (*phoS* and *phoT*) of *Escherichia coli*. J. Bacteriol. **152**:692–701.
2. **Akeroyd, R., L. G. Lange, J. Westerman, and K. W. A. Wirtz.** 1981. Modification of the phosphatidylcholine-transfer protein from bovine liver with butanedione and phenylglyoxal. Evidence for one essential arginine residue. Eur. J. Biochem. **121**:77–81.
3. **Aono, H., and N. Otsuji.** 1968. Genetic mapping of regulator gene *phoS* for alkaline phosphatase in *Escherichia coli*. J. Bacteriol. **95**:1182–1183.
4. **Argast, M., and W. Boos.** 1980. Co-regulation in *Escherichia coli* of a novel transport system for *sn*-glycerol-3-phosphate and outer membrane protein Ic (e, E) with alkaline phosphatase and phosphate-binding protein. J. Bacteriol. **143**:142–150.
5. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
6. **Bennett, R. L., and M. H. Malamy.** 1970. Arsenate resistant mutants of *Escherichia coli* and phosphate transport. Biochem. Biophys. Res. Commun. **40**:496–503.
7. **Borer, P. N., B. Dengler, I. Tinoco, Jr., and O. C. Uhlenbeck.** 1974. Stability of ribonucleic acid double-stranded helices. J. Mol. Biol. **86**:843–853.
8. **Bracha, M., and E. Yagil.** 1973. A new type of alkaline phosphatase-negative mutants in *Escherichia coli* K12. Mol. Gen. Genet. **122**:53–60.
9. **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.
10. **Dianoux, A.-C., P. V. Vignais, and A. Tsugita.** 1981. Partial proteolysis of the natural ATPase inhibitor from beef heart mitochondria. Isolation and characterization of an active cleavage product. FEBS Lett. **130**:119–123.
11. **Echols, H., A. Garen, S. Garen, and A. Torriani.** 1961. Genetic control of repression of alkaline phosphatase in *E. coli*. J. Mol. Biol. **3**:425–438.
12. **Edman, P., and A. Henschen.** 1975. Sequence determination, p. 232–279. *In* S. B. Needleman (ed.), Protein sequence determination, 2nd ed. Springer-Verlag, Berlin.
13. **Emer, S. D., M. N. Hall, and T. J. Silhavy.** 1980. A mechanism of protein localization: the signal hypothesis and bacteria. J. Cell Biol. **86**:701–711.
14. **Garen, A., and H. Echols.** 1962. Genetic control of induction of alkaline phosphatase synthesis in *E. coli*. Proc. Natl. Acad. Sci. U.S.A. **48**:1398–1402.
15. **Garen, A., and H. Echols.** 1962. Properties of two regulating genes for alkaline phosphatase. J. Bacteriol. **83**:297–300.
16. **Garen, A., and N. Otsuji.** 1964. Isolation of a protein specified by a regulatory gene. J. Mol. Biol. **8**:841–852.
17. **Gerdes, R. G., and H. Rosenberg.** 1974. The relationship between the phosphate-binding protein and a regulator gene product from *Escherichia coli*. Biochim. Biophys. Acta **351**:77–86.
18. **Halegoua, S., and M. Inouye.** 1979. Biosynthesis and assembly of the outer membrane protein, p. 67–114. *In* M. Inouye (ed.), Bacterial outer membranes: biogenesis and functions. John Wiley & Sons, Inc., New York.
19. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
20. **Inouye, H., W. Barnes, and J. Beckwith.** 1982. Signal sequence of alkaline phosphatase of *Escherichia coli*. J. Bacteriol. **149**:434–439.
21. **Kikuchi, Y., K. Yoda, M. Yamasaki, and G. Tamura.** 1981. The nucleotide sequence of the promoter and the amino-terminal region of alkaline phosphatase structural gene (*phoA*) of *Escherichia coli*. Nucleic Acids Res. **9**:5671–5678.
22. **Kreuzer, K., C. Pratt, and A. Torriani.** 1975. Genetic analysis of regulatory mutants of alkaline phosphatase of *E. coli*. Genetics **81**:459–468.
23. **Kuhara, S., F. Matsuno, S. Futamura, T. Shinohara, T. Takagi, A. Fujita, and Y. Sakaki.** 1984. GENAS: A data base system for nucleic acid sequence analysis. Nucleic Acids Res. **12**:(in press).
24. **Kupersztoch-Portnoy, Y. M., M. A. Lovett, and D. R. Helinski.** 1974. Strand and site specificity of the relaxation event for the relaxation complex of the antibiotic resistance plasmid R6K. Biochemistry **13**:5484–5490.
25. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
26. **Levitz, R., R. Bittan, and E. Yagil.** 1981. Complementation test

between alkaline phosphatase-constitutive mutants (*phoS* and *phoT*) of *Escherichia coli*. J. Bacteriol. **145**:1432–1435.
27. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavage. Methods Enzymol. **65**:499–560.
28. **Medveczky, N., and H. Rosenberg.** 1969. The binding and release of phosphate by a protein isolated from *Escherichia coli*. Biochim. Biophys. Acta **192**:369–371.
29. **Medveczky, N., and H. Rosenberg.** 1970. The phosphate-binding protein of *Escherichia coli*. Biochim. Biophys. Acta **211**:158–168.
30. **Morita, T., M. Amemura, K. Makino, H. Shinagawa, K. Magota, N. Otsuji, and A. Nakata.** 1983. Hyperproduction of phosphate-binding protein, *phoS*, and pre-*phos* proteins in *Escherichia coli* carrying a cloned *phoS* gene. Eur. J. Biochem. **130**:427–435.
31. **Morkin, E., I. L. Flink, and S. K. Banerjee.** 1979. Phenylglyoxal modification of cardiac myosin S-1. Evidence for essential arginine residues at the active site. J. Biol. Chem. **254**:12647–12652.
32. **Narita, K., H. Matsuo, and T. Nakajima.** 1975. End group determination, p. 70–79. *In* S. B. Needleman (ed.), Protein sequence determination, 2nd ed. Springer-Verlag, Berlin.
33. **Overbeeke, N., H. Bergmans, F. van Mansfeld, and B. Lugtenberg.** 1983. Complete nucleotide sequence of *phoE*, the structural gene for the phosphate limiation inducible outer membrane pore protein of *Escherichia coli* K12. J. Mol. Biol. **163**:513–532.
34. **Overbeeke, N., and B. Lugtenberg.** 1980. Expression of outer membrane protein e of *Escherichia coli* K12 by phosphate limitation. FEBS Lett. **112**:229–232.
35. **Pratt, C., and A. Torriani.** 1977. Complementation test between alkaline phosphatase regulatory mutations *phoB* and *phoRc* in *Escherichia coli*. Genetics **85**:203–208.
36. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
37. **Rosenberg, H., R. G. Gerdes, and K. Chegwidden.** 1977. Two systems for the uptake of phosphate in *Escherichia coli*. J. Bacteriol. **131**:505–511.
38. **Schrijen, J. J., W. A. H. M. Luyben, J. J. H. H. M. De Pont, and S. L. Bonting.** 1980. Studies on (K^+ + H^+)-ATPase. 1. Essential arginine residue in its substrate binding center. Biochim. Biophys. Acta **597**:331–344.
39. **Shinagawa, H., K. Makino, and A. Nakata.** 1983. Regulation of the *pho* regulon in *Escherichia coli* K-12. Genetic and physiological regulation of the positive regulatory gene *phoB*. J. Mol. Biol. **168**:477–488.
40. **Silhavy, T. J., M. J. Casadaban, H. A. Shuman, and J. Beckwith.** 1976. Conversion of β-galactosidase to a membrane-bound state by gene fusion. Proc. Natl. Acad. Sci. U.S.A. **73**:3423–3427.
41. **Simpson, R. J., M. R. Neuberger, and T.-Y. Liu.** 1976. Complete amino acid analysis of proteins from a single hydrolysate. J. Biol. Chem. **251**:1936–1940.
42. **Sprague, G. F., Jr., R. M. Bell, and J. E. Cronan, Jr.** 1975. A mutant of *Escherichia coli* auxotrophic for organic phosphates: evidence for two defects in inorganic phosphate transport. Mol. Gen. Genet. **143**:71–77.
43. **Staden, R.** 1978. Further procedures for sequence analysis by computer. Nucleic Acids Res. **5**:1013–1015.
44. **Tinoco, I., Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
45. **Tommassen, J., and B. Lugtenberg.** 1980. Outer membrane protein e of *Escherichia coli* K-12 is co-regulated with alkaline phosphatase. J. Bacteriol. **143**:151–157.
46. **Tommassen, J., and B. Lugtenberg.** 1982. *pho*-regulon of *Escherichia coli* K12: a minireview. Ann. Microbiol. (Paris) **133A**:243–249.
47. **Willsky, G. R., R. L. Bennett, and M. H. Malamy.** 1973. Inorganic phosphate transport in *Escherichia coli*: involvement of two genes which play a role in alkaline phosphatase regulation. J. Bacteriol. **113**:529–539.
48. **Willsky, G. R., and M. H. Malamy.** 1976. Control of the synthesis of alkaline phosphatase and the phosphate-binding protein in *Escherichia coli*. J. Bacteriol. **127**:595–609.
49. **Zimmerman, C. L., E. Appella, and J. J. Pisano.** 1977. Rapid analysis of amino acid phenylthiohydantoins by high-performance liquid chromatography. Anal. Biochem. **77**:569–573.
50. **Zuckier, G., and A. Torriani.** 1981. Genetic and physiological tests of three phosphate-specific transport mutants of *Escherichia coli*. J. Bacteriol. **145**:1249–1256.