JOURNAL OF BACTERIOLOGY, Oct. 1986, p. 294–302
0021-9193/86/100294-09$02.00/0

Vol. 168, No. 1

# Nucleotide Sequence of the *phoM* Region of *Escherichia coli*: Four Open Reading Frames May Constitute an Operon

MITSUKO AMEMURA, KOZO MAKINO, HIDEO SHINAGAWA, AND ATSUO NAKATA*

*Department of Experimental Chemotherapy, The Research Institute for Microbial Diseases, Osaka University, 3-1, Yamadaoka, Suita, Osaka 565, Japan*

Received 26 March 1986/Accepted 22 July 1986

**The *phoM* gene is one of the positive regulatory genes for the phosphate regulon of *Escherichia coli*. We analyzed the nucleotide sequence of a 4.7-kilobase chromosomal DNA segment that encompasses the *phoM* gene and its flanking regions. Four open reading frames (ORFs) were identified in the order $ORF_1$-$ORF_2$-$ORF_3$ (*phoM*)-$ORF_4$-*dye* clockwise on the standard *E. coli* genetic map. Since these ORFs are preceded by a putative promotor sequence upstream of $ORF_1$ and followed by a putative terminator distal to $ORF_4$, they seem to constitute an operon. The 157-amino-acid $ORF_1$ protein contains highly hydrophobic amino acids in the amino-terminal portion, which is a characteristic of a signal peptide. The 229-amino-acid $ORF_2$ protein is highly homologous to the PhoB protein, a positive regulatory protein for the phosphate regulon. The $ORF_3$ (*phoM* gene) protein contains two stretches of highly hydrophobic residues in the amino-terminal and central regions and, therefore, may be a membrane protein. The 450-amino-acid $ORF_4$ protein contains long hydrophobic regions and is likely to be a membrane protein.**

*Escherichia coli* has a very complex network for the regulation of the genes involved in the uptake and metabolism of phosphate (8, 9, 19, 37; see references 30 and 31 for recent reviews). Expression of the *phoA* and *phoE* genes and the *pst* and *phoB-phoR* operons is dependent on a transcriptional activator, the *phoB* gene product (10, 14, 26). In wild-type strains, expression of the *phoB-phoR* operon is positively regulated by the products of the *phoB* and *phoR* genes with limited phosphate and negatively regulated by the products of the *phoR* gene and *pst* operon with excess phosphate. Two types of *phoR* mutants were isolated; one, like *phoR68*, expresses *phoA* constitutively but at a low level, and the other, like *phoR69*, expresses *phoA* constitutively but at a high level. Only in the former type of *phoR* mutants does expression of the *phoB-phoR* operon and the genes of the phosphate regulon depend on *phoM* gene function with both high and low phosphate (10, 14, 26, 35). The PhoM protein is functionally analogous to the positive regulatory form of the PhoR protein, although *phoM*-dependent *phoA* expression is not as high as *phoR*-dependent *phoA* expression.

The *phoM* gene maps at 100 min on the *E. coli* genetic map (2, 34) and has been cloned on a vector plasmid (13, 15, 29). The gene product has been identified by the maxicell method as a protein with a molecular weight of 55,000 to 60,000, and the orientation of transcription has been shown to be clockwise on the genetic map (13, 15). To determine the primary structure of the PhoM protein and the structure of the *phoM* gene, we determined the nucleotide sequence of the *phoM* gene and its flanking regions. The immediate flanking regions of the *phoM* gene do not contain sequences typical for a promoter and terminator. We continued to sequence the flanking regions until a typical promoter upstream of an open reading frame (ORF) and a terminator downstream of an ORF were found.

* Corresponding author.

## MATERIALS AND METHODS

**Bacteria, bacteriophages, and plasmids.** The *E. coli* strains used were BW521 (F⁻ *lacZ524 phoR68 phoM451 rpsL thi*) (from B. L. Wanner) for selecting recombinant plasmids containing the *phoM* gene and JM103 [Δ(*pro-lac*) *supE thi*/F' *traD36 proAB lacI*$^q$ *lacZΔM15*] (18) as a host for bacteriophage M13. Phage M13mp9 was purchased from Pharmacia Japan, Tokyo, and used for cloning and sequencing the DNA fragment. Plasmid pTHR34, carrying the chromosomal fragment of *E. coli* KLF125/KL181 (CGSC 4320) containing the *phoM* gene, has been described (15).

**Media.** L broth, T broth, other liquid media, and agar plates used for experiments with M13 phage have been described (1).

**DNA manipulation.** Plasmid and bacteriophage M13 replicative-form DNAs were prepared by a method previously described (3). Restriction endonuclease digestion, agarose and polyacrylamide gel electrophoreses, ligation of DNA fragments with T4 DNA ligase, transformation with plasmid DNA, and transfection with phage DNA were done as previously described (1).

**Nucleotide sequencing.** The M13 phage was manipulated as described previously (18). The chromosomal DNA fragments to be sequenced were isolated after digestion of pTHR34 plasmid DNA with restriction enzymes. The 1.5-kilobase-pair (kbp) *Eco*RI-*Pvu*II fragment, 2.6-kbp *Kpn*I-*Kpn*I fragment, and 1.9-kbp *Hpa*I-*Hpa*I fragment were thus isolated (Fig. 1). Protruding single-stranded regions were either filled in or cut with T4 DNA polymerase as described previously (17). The blunt-end DNA fragments were ligated into the *Hin*cII site of M13mp9 replicative-form DNA in both orientations. A series of phage clones with deletions from one end of the cloned DNA fragments were prepared by the method of Hong (11). Details of the procedure were described previously (1). The nucleotide sequences of the cloned DNA fragments were analyzed by the dideoxynucleotide chain termination method of Sanger et al. (23, 24).

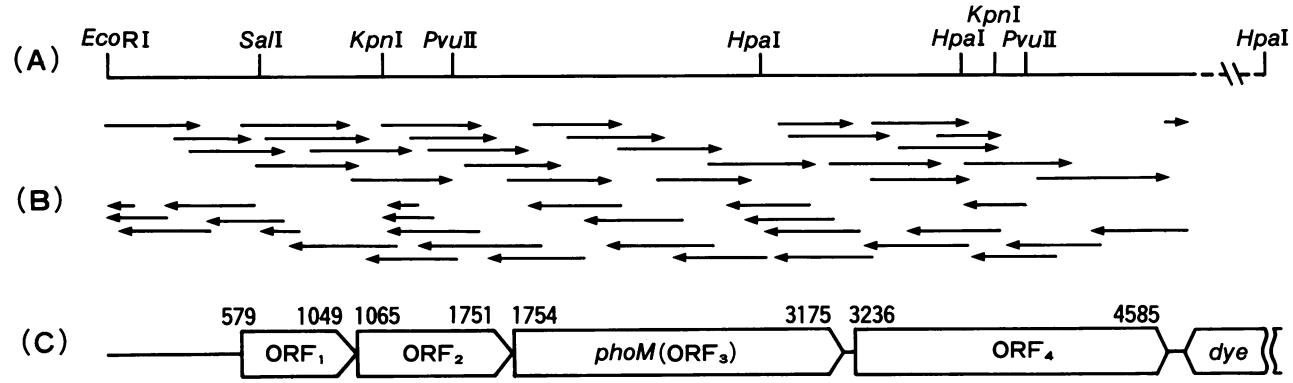


FIG. 1. Summary of DNA sequencing strategy. (A) Restriction map of DNA fragment from plasmid pTHR34 (15) containing the *phoM* gene and flanking regions. (B) Directions and extents of sequences analyzed shown by arrows. (C) Open arrows indicating putative translational ORFs.

**Enzymes and radioisotopes.** Restriction endonucleases, T4 DNA ligase, and an M13 nucleotide sequencing kit including DNA polymerase (Klenow fragment) were obtained from Takara Shuzo Co., Ltd., Kyoto, Japan. All enzymes were used as directed by the supplier. [$\alpha$-$^{32}$P]dCTP was purchased from Amersham Japan, Tokyo.

## RESULTS

**Nucleotide sequence of *phoM* gene and its flanking region.** We had located the *phoM* gene by Tn*1000* insertional inactivation in the 1.5-kbp region between the *Kpn*I and *Hpa*I sites (15; Fig. 1). This region of DNA containing the *phoM* gene was sequenced, but no promoter or terminator was found at the 5′ or 3′ end of the translational ORF for *phoM*, $ORF_3$. Therefore, the DNA flanking this region was analyzed until sequences corresponding to these regulatory structures were found. We analyzed the nucleotide sequence of a 4,657-base-pair region encompassing the *phoM* gene. The sequencing strategy is shown in Fig. 1, and the nucleotide sequence is presented in Fig. 2. Sequence data for both strands were obtained for all of the 4,657 base pairs, with overlaps between the junctions.

**Identification of translational ORFs in *phoM* region.** Since the *phoM* gene was located within a 1.5-kbp region on the DNA fragment and the gene product is a protein with a molecular weight of 55,000 to 60,000 (13, 15, 29), the ORF consisting of 1,422 bases from nucleotides 1754 to 3175 (Fig. 1 and 2) corresponds to the coding region of the gene. This ORF can code for a protein of 474 amino acids with a molecular weight of 52,116, and the orientation of the gene is the same as that of the *phoM* gene previously identified (13, 15). The first ATG codon of $ORF_3$ (*phoM*) is preceded by a typical ribosome-binding-site sequence (12, 27), GAGG, located at nucleotides 1742 to 1745.

Drury and Buxton (7) identified an ORF between the *phoM* and *dye* genes and established part of the nucleotide sequence. $ORF_4$, consisting of 1,350 bases from nucleotides 3236 to 4585, which is distal to the *phoM* gene (Fig. 1 and 2), was identified, and the nucleotide sequence of the distal half of $ORF_4$ corresponded to the nucleotide sequence reported by Drury and Buxton.

Two ORFs were identified proximal to the *phoM* gene. $ORF_1$, from nucleotides 579 to 1049, can code for a protein of 157 amino acids with a molecular weight of 17,107, and $ORF_2$, from nucleotides 1065 to 1751, can code for a protein of 229 amino acids with a molecular weight of 26,124 (Fig. 1 and 2).

Ludtke et al. (13) detected three proteins with molecular weights of 17,000, 28,000, and 55,000 coded for by a *phoM*$^+$ plasmid. The molecular weights of the first two proteins are consistent with those deduced from the DNA sequences for $ORF_1$ and $ORF_2$, respectively.

Each of the four ORFs is preceded by a putative ribosome-binding-site sequence (12, 27) (Fig. 2). It is likely that these four ORFs code for proteins, although the function of each ORF has not been identified nor has the mutant strain corresponding to each ORF been isolated (except for the *phoM* mutant).

Preceding $ORF_1$, a well-conserved putative promoter consisting of the Pribnow box (−10 region) TATGTT from nucleotides 522 to 527 and the −35 region TTGAGA from nucleotides 500 to 505 was found (Fig. 2). Although two other putative promoters, one consisting of the −10 region TTTTGT from nucleotides 506 to 511 and the −35 region TTGACG from nucleotides 482 to 487 and the other consisting of the −10 region TATGTT from nucleotides 469 to 474 and the −35 region TTCAGT from nucleotides 445 to 450, were found (Fig. 2), they deviate more from the canonical sequences for promoters (22) than does the one found in the nucleotide 522-to-527 and 500-to-505 regions.

A nucleotide sequence with an inverted repeat that may form a stable stem-and-loop structure in the transcript was found distal to $ORF_4$ from nucleotides 4596 to 4630 (Fig. 2), and this may be a transcriptional terminator for this operon, as well as the terminator for the *dye* gene, as suggested by Drury and Buxton (7).

Since the *chlG* gene also maps in this region (34), we were interested in determining whether *chlG* corresponded to one of the unassigned ORFs in the region sequenced. However, no complementation was found (data not shown). Since we used plasmid pTHR32, which carries the four ORFs and the *dye* and *thrAB* genes on a mini-F vector (15), for the complementation test, it is likely that the *chlG* gene is not on this part of the chromosome, i.e., between $ORF_1$ and the *thrA* gene.

**Comparison of PhoM and PhoR proteins.** Since the function of the *phoM* gene is analogous to the positive regulatory function of the *phoR* gene in the regulation of the phosphate regulon (15, 35), we examined whether the amino acid sequences of their products deduced from the DNA sequences shared any homology. A homology search with the Mutation Data Matrix computer program (6, 20, 25) revealed no significant homologies between the two protein sequences (data not shown).

```
1                                30                                60                                90
GAATTCACCCAGGCGTAGCGGCGGGCGAATA CCAAAGGCGCTCCATTCAGGAGAACGGCGG TAAAGTGCAGGAGTCTGGGCAAACTGCTTC

                                120                               150                               180
TTGAATGCGCGGGTAAATGTCTGTTGAGAG TCGAAGCGGTATTGCAGCGCGATGTCCAGA ATCGGACGCGCAGTCAGGCGTAGTGCGACC

                                210                               240                               270
GCCGATTTCGACAAACGACGAGCACGAATA TACGCGCCAATAGCATGGCCAGTGACATCT TTAAACATTCTCTGTAAGTGCCACTTGGAA

                                300                               330                               360
TAACCTGCTTTCGCCGCTACATTGTCGAGC GACAGGGGCTGATCCAGATGACCTTTCAGC CAGATTAAAAGGTCGCGAATAATGCCGGCC

                                390                               420                               450
TGATCCATAAAATATCCTCATCCTTTCAAC AACGAGCACCTGACATCAGGTAATTGGATA ATAGCATTTTTTGCTGTTTTAGCATTCAGT

                                480                               510                               540
GTTTTTTTCTTAGTAGAGTATGTTTTAGGG CTTGACGGAAATAAAAGTATTGAGATTTTG TTCTTAATCAATATGTTATTTACCGTGACG

                                570                               600                               630
AACTAATTGCTCGTGTAATAGATAAAAATG GTAACAATATGAAATACAAGCATTTGATCC TGTCTTTAAGCCTGATAATGCTGGGGCCAT
                                        MetLysTyrLysHisLeuIle LeuSerLeuSerLeuIleMetLeuGlyPro
                                          (1)                               (10)

                                660                               690                               720
TGGCTCATGCAGAAGAGATTGGTTCGGTCG ACACCGTATTTAAAATGATCGGCCCGGATC ACAAAATTGTTGTGGAAGCCTTTGATGATC
LeuAlaHisAlaGluGluIleGlySerVal AspThrValPheLysMetIleGlyProAsp HisLysIleValValGluAlaPheAspAsp
      (20)                              (30)                              (40)

                                750                               780                               810
CCGATGTGAAAAATGTCACCTGTTATGTGA GCCGGGCGAAAACCGGTGGTATTAAAGGGG GATTGGGTCTGGCGGAAGATACCTCCGATG
ProAspValLysAsnValThrCysTyrVal SerArgAlaLysThrGlyGlyIleLysGly GlyLeuGlyLeuAlaGluAspThrSerAsp
      (50)                              (60)                              (70)

                                840                               870                               900
CGGCCATTTCTTGTCAGCAAGTCGGGCCGA TTGAACTGTCGGATCGTATTAAAAACGGCA AAGCTCAGGGCGAGGTAGTATTCAAAAAAC
AlaAlaIleSerCysGlnGlnValGlyPro IleGluLeuSerAspArgIleLysAsnGly LysAlaGlnGlyGluValValPheLysLys
     (80)                              (90)                              (100)

                                930                               960                               990
GCACGTCCCTGGTCTTTAAGTCGTTACAGG TCGTGCGCTTTTATGATGCCAAACGCAACG CGCTCGCTTATCTGGCTTACTCCGACAAAG
ArgThrSerLeuValPheLysSerLeuGln ValValArgPheTyrAspAlaLysArgAsn AlaLeuAlaTyrLeuAlaTyrSerAspLys
     (110)                             (120)                             (130)

                               1020                              1050                              1080
TTGTAGAAGGTTCGCCGAAAAACGCGATTA GCGCGGTTCCTGTCATGCCGTGGCGGCAAT AACAGAGGCGATTTATGCAACGGGAAACGG
ValValGluGlySerProLysAsnAlaIle SerAlaValProValMetProTrpArgGln                MetGlnArgGluThr
     (140)                             (150)                (157)               (1)

                               1110                              1140                              1170
TCTGGTTAGTGGAAGATGAGCAAGGGATAG CCGACACGCTGGTCTACATGTTGCAGCAGG AAGGTTTTGCCGTCGAGGTCTTTGAGCGAG
ValTrpLeuValGluAspGluGlnGlyIle AlaAspThrLeuValTyrMetLeuGlnGln GluGlyPheAlaValGluValPheGluArg
          (10)                              (20)                              (30)

                               1200                              1230                              1260
GCTTGCCGGTGCTGGATAAAGCTCGCAAGC AGGTACCCGACGTCATGATTCTCGATGTTG GTCTGCCGGATATTAGCGGCTTTGAATTGT
GlyLeuProValLeuAspLysAlaArgLys GlnValProAspValMetIleLeuAspVal GlyLeuProAspIleSerGlyPheGluLeu
          (40)                              (50)                              (60)

                               1290                              1320                              1350
GCCGCCAGTTACTGGCGCTCCATCCGGCGT TACCTGTACTGTTCCTGACGGCCCGAAGTG AAGAGGTCGATCGCCTGCTTGGGCTGGAAA
CysArgGlnLeuLeuAlaLeuHisProAla LeuProValLeuPheLeuThrAlaArgSer GluGluValAspArgLeuLeuGlyLeuGlu
          (70)                              (80)                              (90)

                               1380                              1410                              1440
TTGGTGCTGACGACTACGTGGCTAAACCGT TTTCACCCCGCGAAGTGTGCGCCAGGGTGC GCACCTTACTGCGTCGGGTGAAGAAGTTCT
IleGlyAlaAspAspTyrValAlaLysPro PheSerProArgGluValCysAlaArgVal ArgThrLeuLeuArgArgValLysLysPhe
          (100)                             (110)                             (120)

                               1470                              1500                              1530
CGACGCCGTCTCCCGTCATCCGTATTGGAC ATTTTGAATTGAATGAACCCGCGGCGCAGA TCAGCTGGTTTGACACGCCATTAGCGCTGA
SerThrProSerProValIleArgIleGly HisPheGluLeuAsnGluProAlaAlaGln IleSerTrpPheAspThrProLeuAlaLeu
          (130)                             (140)                             (150)

                               1560                              1590                              1620
CTCGGTATGAGTTTTTATTGTTGAAGACGT TACTCAAGTCACCGGGCCGCGTCTGGTCCC GCCAGCAACTGATGGATAGCGTATGGGAAG
ThrArgTyrGluPheLeuLeuLeuLysThr LeuLeuLysSerProGlyArgValTrpSer ArgGlnGlnLeuMetAspSerValTrpGlu
          (160)                             (170)                             (180)
```

```
                          1650                          1680                          1710
ATGCGCAGGACACCTACGATCGCACCGTCG ATACCCACATTAAAACGCTGCGTGCCAAGC TGCGCGCCATCAACCCCGATCTTTCACCGA
AspAlaGlnAspThrTyrAspArgThrVal AspThrHisIleLysThrLeuArgAlaLys LeuArgAlaIleAsnProAspLeuSerPro
        (190)                         (200)                         (210)

                          1740                          1770                          1800
TTAATACTCATCGCGGCATGGGATATAGCC TGAGGGGCCTGTAATGCGTATCGGCATGCG GTTGTTGCTGGGCTATTTTTTACTGGTGGC
IleAsnThrHisArgGlyMetGlyTyrSer LeuArgGlyLeu  MetArgIleGlyMetArg LeuLeuLeuGlyTyrPheLeuLeuValAla
        (220)                       (229)  (1)                        (10)

                          1830                          1860                          1890
GGTGGCAGCCTGGTTCGTACTGGCCATTTT TGTCAAAGAAGTTAAACCGGGCGTGCGAAG AGCAACGGAGGGGACGTTGATCGACACCGC
 ValAlaAlaTrpPheValLeuAlaIlePhe ValLysGluValLysProGlyValArgArg AlaThrGluGlyThrLeuIleAspThrAla
        (20)                          (30)                          (40)

                          1920                          1950                          1980
AACGTTGCTGGCGGAGCTGGCGCGTCCCGA TTTGCTCTCTGGGGACCCAACGCATGGGCA ACTGGCGCAGGCGTTTAATCAGCTACAACA
 ThrLeuLeuAlaGluLeuAlaArgProAsp LeuLeuSerGlyAspProThrHisGlyGln LeuAlaGlnAlaPheAsnGlnLeuGlnHis
        (50)                          (60)                          (70)

                          2010                          2040                          2070
TCCCCCGTTTCGCGCCAATATCGGTGGCAT TAACAAAGTGCGCAATGAATATCATGTCTA TATGACCGATGCGCAGGGCAAAGTATTGTT
 ProProPheArgAlaAsnIleGlyGlyIle AsnLysValArgAsnGluTyrHisValTyr MetThrAspAlaGlnGlyLysValLeuPhe
        (80)                          (90)                          (100)

                          2100                          2130                          2160
CGATTCGGCAAATAAAGCCGTTGGACAGGA TTATTCGCGCTGGAATGACGTCTGGCTAAC GTTGCGTGGTCAGTATGGTGCGCGCAGCAC
 AspSerAlaAsnLysAlaValGlyGlnAsp TyrSerArgTrpAsnAspValTrpLeuThr LeuArgGlyGlnTyrGlyAlaArgSerThr
        (110)                         (120)                         (130)

                          2190                          2220                          2250
GTTGCAAAATCCTGCCGATCCCGAAAGTTC TGTGATGTATGTTGCCGCACCGATTATGGA CGGCTCGCGGCTTATTGGCGTTTTGAGCGT
 LeuGlnAsnProAlaAspProGluSerSer ValMetTyrValAlaAlaProIleMetAsp GlySerArgLeuIleGlyValLeuSerVal
        (140)                         (150)                         (160)

                          2280                          2310                          2340
AGGCAAACCGAACGCGGCGATGGCTCCGGT CATTAAGCGTAGCGAGCGGCGAATTTTATG GGCCAGCGCCATTTTGTTGGGGATTGCACT
 GlyLysProAsnAlaAlaMetAlaProVal IleLysArgSerGluArgArgIleLeuTrp AlaSerAlaIleLeuLeuGlyIleAlaLeu
        (170)                         (180)                         (190)

                          2370                          2400                          2430
GGTGATTGGCGCAGGCATGGTTTGGTGGAT CAACCGCTCTATTGCCCGGCTCACTCGCTA TGCTGATTCCGTCACTGACAATAAGCCCGT
 ValIleGlyAlaGlyMetValTrpTrpIle AsnArgSerIleAlaArgLeuThrArgTyr AlaAspSerValThrAspAsnLysProVal
        (200)                         (210)                         (220)

                          2460                          2490                          2520
TCCTCTCCCCGATCTCGGTAGTAGCGAGTT GCGTAAACTCGCGCAGGCGCTGGAAAGTAT GCGCGTGAAGCTGGAAGGGAAAAACTATAT
 ProLeuProAspLeuGlySerSerGluLeu ArgLysLeuAlaGlnAlaLeuGluSerMet ArgValLysLeuGluGlyLysAsnTyrIle
        (230)                         (240)                         (250)

                          2550                          2580                          2610
TGAGCAGTATGTTTACGCATTAACTCATGA GCTAAAAAGCCCACTGGCGGCGATTCGTGG AGCGGCGGAAATTTTACGCGAAGGTCCGCC
 GluGlnTyrValTyrAlaLeuThrHisGlu LeuLysSerProLeuAlaAlaIleArgGly AlaAlaGluIleLeuArgGluGlyProPro
        (260)                         (270)                         (280)

                          2640                          2670                          2700
GCCGGAAGTGGTGGCTCGTTTTACTGACAA CATTCTGACGCAAAATGCGCGTATGCAGGC ATTGGTAGAAACGTTACTACGCCAGGCAAG
 ProGluValValAlaArgPheThrAspAsn IleLeuThrGlnAsnAlaArgMetGlnAla LeuValGluThrLeuLeuArgGlnAlaArg
        (290)                         (300)                         (310)

                          2730                          2760                          2790
ACTGGAGAATCGTCAGGAAGTCGTTCTGAC TGCTGTTGATGTGGCGGCATTATTCCGCCG CGTCAGCGAAGCGCGCACCGTGCAGTTGGC
 LeuGluAsnArgGlnGluValValLeuThr AlaValAspValAlaAlaLeuPheArgArg ValSerGluAlaArgThrValGlnLeuAla
        (320)                         (330)                         (340)

                          2820                          2850                          2880
AGAAAAAAAAAATCACTCTGCATGTTACGCC CACCGAGGTTAACGTTGCTGCTGAACCGGC GTTACTGGAGCAGGCGCTGGGAAATTTACT
 GluLysLysIleThrLeuHisValThrPro ThrGluValAsnValAlaAlaGluProAla LeuLeuGluGlnAlaLeuGlyAsnLeuLeu
        (350)                         (360)                         (370)

                          2910                          2940                          2970
GGATAACGCCATCGATTTTACCCCCGAGAG CGGTTGCATCACGCTAAGCGCCGAAGTGGA TCAGGAACACGTCACGCTTAAGGTGCTGGA
 AspAsnAlaIleAspPheThrProGluSer GlyCysIleThrLeuSerAlaGluValAsp GlnGluHisValThrLeuLysValLeuAsp
        (380)                         (390)                         (400)
```

```
                          3000                                3030                                3060
TACCGGTAGTGGTATTCCTGACTACGCGCT TTCACGTATTTTTGAACGCTTTTACTCTTT GCCTCGTGCAAATGGGCAAAAAAGCAGCGG
 ThrGlySerGlyIleProAspTyrAlaLeu SerArgIlePheGluArgPheTyrSerLeu ProArgAlaAsnGlyGlnLysSerSerGly
        (410)                               (420)                               (430)

                          3090                                3120                                3150
TCTGGGGTTGGCGTTCGTCAGTGAGGTCGC CCGTTTGTTTAACGGCGAAGTCACGCTGCG CAACGTGCAGGAAGGTGGCGTGCTGGCCTC
 LeuGlyLeuAlaPheValSerGluValAla ArgLeuPheAsnGlyGluValThrLeuArg AsnValGlnGluGlyGlyValLeuAlaSer
        (440)                               (450)                               (460)

                          3180                                3210                                3240
GCTTCGACTTCACCGTCACTTCACA**TAG**CT TCAAATTCTTCCCACATAGTCTTCGTATCC TGCTGCCATTGCAA**AGGAG**AAGACT**ATG**TT
 LeuArgLeuHisArgHisPheThr                                                                      MetLeu
        (470)                                                                                   (1)

                          3270                                3300                                3330
GAAATCCCCCCTGTTCTGGAAAATGACTAG CCTGTTTGGTGCAGTATTGCTGTTGTTAAT TCCGATAATGCTGATTCGGCAGGTGATTGT
 LysSerProLeuPheTrpLysMetThrSer LeuPheGlyAlaValLeuLeuLeuIle ProIleMetLeuIleArgGlnValIleVal
                  (10)                                (20)                                (30)

                          3360                                3390                                3420
CGAACGTGCTGATTACCGTAGCGATGTGGA AGATGCGATTCGCCAAAGTACCAGCGGGCC GCAAAAACTCGTTGGGCCGCTCATCGCTAT
 GluArgAlaAspTyrArgSerAspValGlu AspAlaIleArgGlnSerThrSerGlyPro GlnLysLeuValGlyProLeuIleAlaIle
                  (40)                                (50)                                (60)

                          3450                                3480                                3510
TCCTGTGACCGAGCTTTATACGGTGCAGGA AGAGGATAAAACCGTGGAGCGGAAACGAAG TTTTATCCATTTTTGGTTACCTGAGTCATT
 ProValThrGluLeuTyrThrValGlnGlu GluAspLysThrValGluArgLysArgSer PheIleHisPheTrpLeuProGluSerLeu
                  (70)                                (80)                                (90)

                          3540                                3570                                3600
GATGGTTGATGGCAATCAGAACGTGGAAGA ACGCAAGATAGGGATTTATACCGGTCAGGT CTGGCACAGTGATTTAACGTTAAAAGCCGA
 MetValAspGlyAsnGlnAsnValGluGlu ArgLysIleGlyIleTyrThrGlyGlnVal TrpHisSerAspLeuThrLeuLysAlaAsp
                 (100)                               (110)                               (120)

                          3630                                3660                                3690
TTTCGATGTTTCGCGTCTTAGCGAACTCAA CGCGCCAAATATCACCTTAGGCAAGCCATT TATTGTGATTAGCGTCGGGGATGCGCGTGG
 PheAspValSerArgLeuSerGluLeuAsn AlaProAsnIleThrLeuGlyLysProPhe IleValIleSerValGlyAspAlaArgGly
                 (130)                               (140)                               (150)

                          3720                                3750                                3780
TATTGGTGTGGTGAAAGCGCCTGAAGTTAA CGGAACGGCGCTGACCATTGAACCCGGCAC CGGGTTAGAGCAAGGCGGGCAGGGCGTGCA
 IleGlyValValLysAlaProGluValAsn GlyThrAlaLeuThrIleGluProGlyThr GlyLeuGluGlnGlyGlyGlnGlyValHis
                 (160)                               (170)                               (180)

                          3810                                3840                                3870
TATCCCTTTACCTGAAGGGGACTGGCGGAA GCAGAACCTGAAGCTGAATATGGCCCTGAA TTTAAGCGGTACCGGCGATCTTTCTGTGGT
 IleProLeuProGluGlyAspTrpArgLys GlnAsnLeuLysLeuAsnMetAlaLeuAsn LeuSerGlyThrGlyAspLeuSerValVal
                 (190)                               (200)                               (210)

                          3900                                3930                                3960
GCCTGGCGGGCGTAATAGCGAAATGACCTT AACCAGCAACTGGCCGCATCCCAGTTTTTT AGGTGATTTTCTACCAGCCAAACGGGAAGT
 ProGlyGlyArgAsnSerGluMetThrLeu ThrSerAsnTrpProHisProSerPheLeu GlyAspPheLeuProAlaLysArgGluVal
                 (220)                               (230)                               (240)

                          3990                                4020                                4050
TAGCGAGTCAGGTTTTCAGGCGCACTGGCA AAGCAGCTGGTTTGCTAATAATCTCGGTGA GCGTTTTGCTTCAGGCAATGATACCGGCTG
 SerGluSerGlyPheGlnAlaHisTrpGln SerSerTrpPheAlaAsnAsnLeuGlyGlu ArgPheAlaSerGlyAsnAspThrGlyTrp
                 (250)                               (260)                               (270)

                          4080                                4110                                4140
GGAAAACTTCCCGGCGTTTAGCGTCGCAGT AACGACGCCAGCCGATCAATACCAATTAAC TGACCGGGCGACTAAGTACGCCATTCTGCT
 GluAsnPheProAlaPheSerValAlaVal ThrThrProAlaAspGlnTyrGlnLeuThr AspArgAlaThrLysTyrAlaIleLeuLeu
                 (280)                               (290)                               (300)

                          4170                                4200                                4230
GATTGCACTGACTTTTATGGCGTTCTTTGT TTTTGAAACGCTCACCGCGCAACGTTTACA CCCAATGCAATATTTGCTGGTGGGGCTTTC
 IleAlaLeuThrPheMetAlaPhePheVal PheGluThrLeuThrAlaGlnArgLeuHis ProMetGlnTyrLeuLeuValGlyLeuSer
                 (310)                               (320)                               (330)

                          4260                                4290                                4320
ATTGGTGATGTTTTATTTGCTCTTGCTGGC GCTTTCTGAACATACCGGTTTTACCGTGGC ATGGATAATCGCCAGTCTGATTGGGGCGAT
 LeuValMetPheTyrLeuLeuLeuLeuAla LeuSerGluHisThrGlyPheThrValAla TrpIleIleAlaSerLeuIleGlyAlaIle
                 (340)                               (350)                               (360)
```

```
                          4350                                4380                                4410
AATGAACGGTATTTATTTGCAAGCGGTATT GAAAGGTTGGTGCAACAGCATGTTGTTTAC CCTCGCGCTGTTGTTGCTGGATGGTGTGAT
 MetAsnGlyIleTyrLeuGlnAlaValLeu LysGlyTrpCysAsnSerMetLeuPheThr LeuAlaLeuLeuLeuLeuAspGlyValMet
                     (370)                               (380)                               (390)

                          4440                                4470                                4500
GTGGGGACTGCTCAACTCTGCCGATAGCGC GCTGTTGTTGGGAACCAGTGTGCTGGTGGT GGCGCTGGCCGGCATGATGTTTGTGACCCG
 TrpGlyLeuLeuAsnSerAlaAspSerAla LeuLeuLeuGlyThrSerValLeuValVal AlaLeuAlaGlyMetMetPheValThrArg
                     (400)                               (410)                               (420)

                          4530                                4560                                4590
TAATATCGACTGGTATGCGTTTTCACTGCC GAAAATGAAAGCCAGTAAAGAAGTTACAAC GGACGATGAGTTACGTATCTGGAAATAAGG
 AsnIleAspTrpTyrAlaPheSerLeuPro LysMetLysAlaSerLysGluValThrThr AspAspGluLeuArgIleTrpLys
                     (430)                               (440)                               (450)

                          4620                                4650
TTGAAAAATAAAAACGGCGCTAAAAAGCGC CGTTTTTTTTGACGGTGGTAAAGCCGATTA ATCTTCC
```

FIG. 2. Nucleotide sequence of *phoM* gene and flanking regions and amino acid sequences of putative proteins deduced form nucleotide sequence. Nucleotide numbers are indicated, with the second nucleotide of the *Eco*RI endonuclease recognition sequence designated as 1. The putative translational initiation and termination codons and putative ribosome-binding-site sequences are in boldface letters. The putative −10 and −35 regions and the nucleotide sequences whose transcripts may form a stable stem-and-loop structure in the 3′ flanking region of $ORF_4$ are underlined.

The amino acid sequence of the PhoR protein deduced from the DNA sequence contains a very long stretch ofhydrophobic residues in its amino-terminal segment, and thus, the protein may be a membrane protein (unpublished data). To compare the hydrophilicity profiles of the two proteins, we plotted a profile of the hydrophilicity of the PhoM protein with a computer program, the Genetyx Hydrophilicity Plot (SDC Software Co., Ltd., Tokyo, Ja-

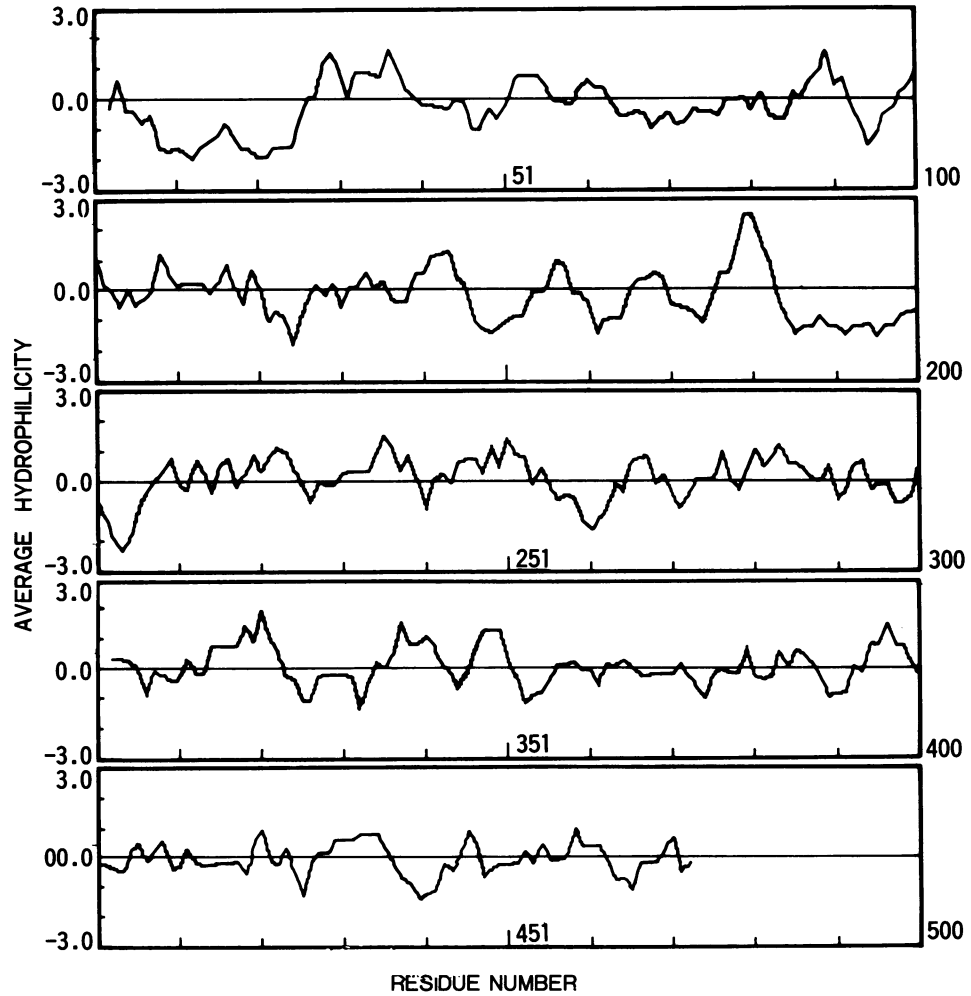


FIG. 3. Hydrophilicity profile of the *phoM* gene product. The profile was plotted with a computer program, the Genetyx Hydrophilicity Plot, by taking averages for five consecutive amino acids.

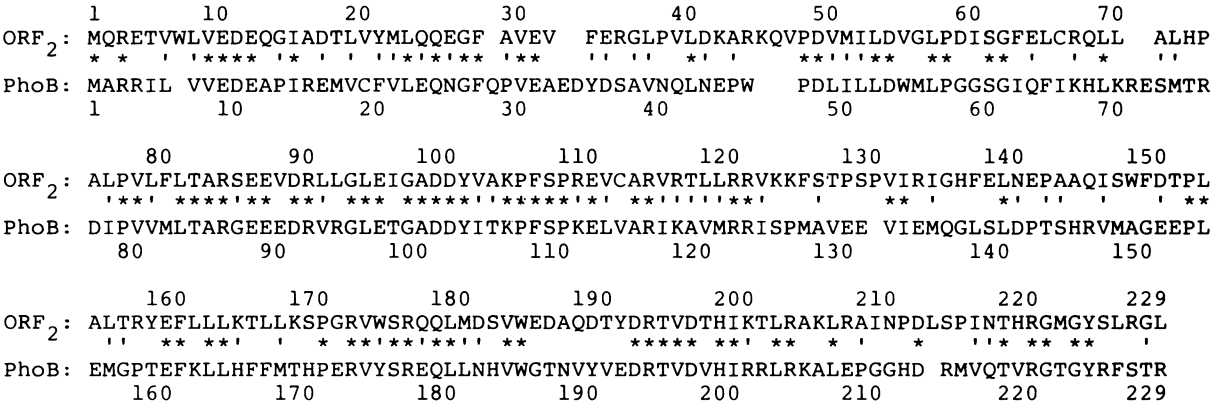


FIG. 4. Comparison of amino acid sequences of $ORF_2$ product and the PhoB protein deduced from nucleotide sequences. The result shown was obtained by computer analysis of maximum matching with the Mutation Data Matrix program (6, 20, 25). Identical and similar amino acids are indicated by asterisks and primes, respectively. Amino acid residue numbers are indicated, with the first methionine residues designated as 1.

pan), by taking averages for five consecutive amino acids. The results indicated that there are two regions with highly hydrophobic amino acid sequences in the PhoM protein, one in the amino-terminal region and the other in the middle region (Fig. 3).

**Homology of $ORF_2$ product with PhoB protein.** Since the *phoR* gene constitutes an operon with the proximal gene, *phoB* (16), and the function of the *phoR* gene can be partially replaced by the *phoM* gene, we examined whether any other products of the ORFs in the putative *phoM* operon showed homology with the PhoB protein (14). The amino acid sequence deduced from $ORF_2$ shows high homology with that deduced from the DNA sequence of the *phoB* gene (Fig. 4). The aligned sequences of the two proteins have 37% identical and 25% similar amino acids.

**The products of $ORF_1$ and $ORF_4$ may be envelope proteins.** Hydrophilicity profiles of the amino acid sequences deduced from the DNA sequences of $ORF_1$ and $ORF_4$ were plotted by a computer program to find clues to the nature of the proteins. The amino-terminal region of the $ORF_1$ product has several features characteristic of signal peptides involved in the secretion of proteins through the inner membrane (21). The putative $ORF_1$ product contains three basic amino acid residues next to the first Met residue, followed by a long stretch of hydrophobic residues that ends with an alanine residue at position 19 or 21 that may be a cleavage site for a signal peptidase (Fig. 2). Since the rest of the sequence does not contain highly hydrophobic regions, this putative protein may be a periplasmic protein.

Since the hydrophilicity profile of the putative $ORF_4$ product revealed similar features for the amino-terminal region (Fig. 2), the $ORF_4$ product may also contain a signal peptide. In addition, it also contains several long stretches with high hydrophobicity near its carboxyl terminus. Therefore, it may be an outer membrane protein or one that spans the inner membrane.

## DISCUSSION

The complete nucleotide sequence of the *phoM* region revealed four ORFs which appear to constitute an operon in which $ORF_3$ corresponds to the *phoM* gene. The direction of transcription is from $ORF_1$ toward $ORF_4$ clockwise on the circular *E. coli* genetic map (2). $ORF_4$ overlaps with the DNA sequence distal to the *dye* gene reported by Drury and Buxton (7). The terminator of this operon is likely to be shared with the *dye* gene, which is transcribed counterclockwise, as suggested by Drury and Buxton (7). Therefore, the organization of the putative genes in this region is *trpR*. . .$ORF_1$-$ORF_2$-*phoM* ($ORF_3$)-$ORF_4$-*dye-thrABC* clockwise on the standard *E. coli* genetic map.

In earlier studies, Makino et al. (15) studied the expression of a *phoM'-'lacZ* fusion. However, the fusion used did not contain the putative promoter for the *phoM* operon, as determined here. Presumably, the *phoM'-'lacZ* fusion studied by Ludtke et al. (13) did contain the putative promoter proximal to $ORF_1$. Their results suggest that the *phoM* gene is not regulated by phosphate. However, the restriction maps of the *phoM* region reported by our group (15) and by Tommassen et al. (29) do not agree with that reported by Ludtke et al., as noted earlier (13). Therefore, we constructed an $ORF_1$-$ORF_2$-*phoM'-'lacZ* fusion with the putative promoter and studied the regulation of the gene expression by phosphate. The results suggest that the expression of the *phoM* gene is not regulated by phosphate in the medium (unpublished results).

Since the expression of this putative operon is not regulated by phosphate and the function of the *phoM* gene could only be detected in the *phoR* mutants, the function of the operon in the regulation of the phosphate regulon may be auxiliary. The functions of the genes in the phosphate regulon determined thus far are all related to the uptake and use of phosphate compounds. Therefore, the phosphate regulon is considered to be a global regulatory system for the cell to adapt to a limited supply of phosphate. The major function of the putative *phoM* operon remains to be determined and is unlikely to be regulated by the phosphate regulon. However, Wanner and McSharry (33, 36) showed that expression of several (although unknown) phosphate-regulated promoters is induced by other physiological conditions for which the *phoB* and *phoM* products are required. The identification of a physiological trigger to induce this operon should be an important clue to the understanding of its biological significance.

Although the *phoM* gene can replace the positive regulatory function of the *phoR* gene, no homology in the primary structures of the gene products was detected. Their only common structural features are the long stretches of hydrophobic regions, and therefore, both may be associated with

the membrane and may function as components of a signal transducer that modifies the function of the PhoB protein.

The primary structure of the $ORF_2$ product deduced from the DNA sequence has extensive homology with that of the PhoB protein, and the $ORF_2$ and PhoB proteins consist of an identical number of amino acids. Since the PhoB protein is likely a transcriptional activator of the genes belonging to the phosphate regulon (14, 31; our unpublished data), the product of $ORF_2$ may also be a transcriptional regulator for the *phoM* operon or some unidentified gene(s) or both. Recently, we purified the PhoB protein and showed that it binds to the consensus sequence for the regulatory regions of the genes of the phosphate regulon and activates transcription of these genes in vitro (14; unpublished results). Since the products of the *phoB* and *phoR* genes are involved in the regulation of the phosphate regulon and these genes constitute an operon (16), a similar relationship can be considered for $ORF_2$ and the *phoM* gene. However, the combination of *phoM* and $ORF_2$ cannot substitute for the combined functions of *phoR* and *phoB* for the regulation of the phosphate regulon, since a strain with a deletion in *phoB-phoR* but with intact $ORF_2$-*phoM* did not synthesize alkaline phosphatase (4). Therefore, although *phoM* can substitute for the positive regulatory function of *phoR* (31, 35), $ORF_2$ cannot substitute for *phoB* in the regulation of the phosphate regulon.

Our recent work showed that both the PhoB and $ORF_2$ proteins have extensive homology with the OmpR (5) and Dye (7) proteins of *E. coli*, the VirG protein of *Agrobacterium tumefaciens* (P. R. Ebert, S. C. Winans, S. E. Stachel, and E. W. Nester, personal communication), and the Spo0A protein of *Bacillus subtilis* (32). All of these proteins may be transcriptional regulators for the genes coding for envelope proteins. The amino-terminal halves of the PhoB and $ORF_2$ proteins have extensive homology with the CheB and CheY proteins of *Salmonella typhimurium* (28) and the Spo0F protein of *B. subtilis* (32). The CheB and CheY proteins are components of a sensory transducer in chemotaxis, and the Spo0F protein is required for sporulation. Therefore, the amino-terminal domains of the $ORF_2$ and PhoB proteins may be involved in the reception of physiological signals and may affect the function in the carboxyl-terminal domains that interact with the regulatory regions of the genes to be regulated.

## ACKNOWLEDGMENTS

We thank B. L. Wanner for supplying the bacterial strain.

This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, and Culture, Japan.

## LITERATURE CITED

1. **Amemura, M., K. Makino, H. Shinagawa, and A. Nakata.** 1985. Nucleotide sequence of the genes involved in phosphate transport and regulation of the phosphate regulon in *Escherichia coli*. J. Mol. Biol. **184**:241–250.
2. **Bachmann, B. J.** 1983. Linkage map of *Escherichia coli* K-12, edition 7. Microbiol. Rev. **47**:180–230.
3. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
4. **Brickman, E., and J. Beckwith.** 1975. Analysis of the regulation of *Escherichia coli* alkaline phosphatase synthesis using deletions and f80 transducing phages. J. Mol. Biol. **96**:307–316.
5. **Comeau, D. E., K. Ikenaka, K. Tsung, and M. Inouye.** 1985. Primary characterization of the protein products of the *Escherichia coli ompB* locus: structure and regulation of synthesis of the OmpR and EnvZ proteins. J. Bacteriol. **164**:578–584.
6. **Dayhoff, M. O., R. M. Schartz, and B. C. Orcutt.** 1978. A model of evolutionary change in proteins, p. 345–352. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
7. **Drury, L. S., and R. S. Buxton.** 1985. DNA sequence analysis of the *dye* gene of *Escherichia coli* reveals amino acid homology between the Dye and OmpR proteins. J. Biol. Chem. **260**:4236–4242.
8. **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.
9. **Garen, A., and H. Echols.** 1962. Genetic control of induction of alkaline phosphatase synthesis in *E. coli*. Proc. Natl. Acad. Sci. USA **48**:1398–1402.
10. **Guan, C.-D., B. Wanner, and H. Inouye.** 1983. Analysis of regulation of *phoB* expression using a *phoB-cat* fusion. J. Bacteriol. **156**:710–717.
11. **Hong, G. F.** 1982. A systematic DNA sequencing strategy. J. Mol. Biol. **158**:539–549.
12. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
13. **Ludtke, D., J. Bernstein, C. Hamilton, and A. Torriani.** 1984. Identification of the *phoM* gene product and its regulation in *Escherichia coli* K-12. J. Bacteriol. **159**:19–25.
14. **Makino, K., H. Shinagawa, M. Amemura, and A. Nakata.** 1986. Nucleotide sequence of the *phoB* gene, the positive regulatory gene for the phosphate regulon of *Escherichia coli* K-12. J. Mol. Biol. **190**:37–44.
15. **Makino, K., H. Shinagawa, and A. Nakata.** 1984. Cloning and characterization of the alkaline phosphatase positive regulatory gene (*phoM*) of *Escherichia coli*. Mol. Gen. Genet. **195**:381–390.
16. **Makino, K., H. Shinagawa, and A. Nakata.** 1985. Regulation of the phosphate regulon of *Escherichia coli* K-12: regulation and role of the regulatory gene *phoR*. J. Mol. Biol. **184**:231–240.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
19. **Morris, H., M. J. Schlesinger, M. Bracha, and E. Yagil.** 1974. Pleiotropic effects of mutations involved in the regulation of *Escherichia coli* K-12 alkaline phosphatase. J. Bacteriol. **119**:583–592.
20. **Needleman, S. B., and C. D. Wunsch.** 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. **48**:443–453.
21. **Oliver, D.** 1985. Protein secretion in *Escherichia coli*. Annu. Rev. Microbiol. **39**:615–648.
22. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
23. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
25. **Schwartz, R. M., and M. O. Dayhoff.** 1978. Matrices for detecting distant relationships, p. 253–258. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
26. **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.
27. **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.
28. **Stock, A., D. E. Koshland, Jr., and J. Stock.** 1985. Homologies between the *Salmonella typhimurium* CheY protein and proteins involved in the regulation of chemotaxis, membrane pro-

tein synthesis and sporulation. Proc. Natl. Acad. Sci. USA **82**:7989–7993.

29. **Tommassen, J., P. Hiemstra, P. Overduin, and B. Lugtenberg.** 1984. Cloning of *phoM*, a gene involved in regulation of the synthesis of phosphate limitation inducible proteins in *Escherichia coli* K12. Mol. Gen. Genet. **195**:190–194.
30. **Tommassen, J., and B. Lugtenberg.** 1982. *pho*-regulon of *Escherichia coli* K12: a minireview. Ann. Microbiol. (Paris) **133A**:243–249.
31. **Torriani, A., and D. N. Ludtke.** 1985. The pho regulon of *Escherichia coli*, p. 224–242. *In* M. Schaechter, F. C. Neidhart, J. Ingraham, and N. O. Kjeldgaard (ed.), The molecular biology of bacterial growth. Jones and Bartlett Publishers, Inc., Boston.
32. **Trach, K. A., J. W. Capman, P. J. Piggot, and J. A. Hoch.** 1985. Deduced product of the stage 0 sporulation gene *spo0F* shares homology with the Spo0A, OmpR, and SfrA proteins. Proc. Natl. Acad. Sci. USA **82**:7260–7264.
33. **Wanner, B. L.** 1983. Overlapping and separate controls on the phosphate regulon in *Escherichia coli* K12. J. Mol. Biol. **166**:283–309.
34. **Wanner, B. L., and J. Bernstein.** 1982. Determining the *phoM* map location in *Escherichia coli* K-12 by using a nearby transposon Tn*10* insertion. J. Bacteriol. **150**:429–432.
35. **Wanner, B. L., and P. Latterell.** 1980. Mutants affected in alkaline phosphatase expression: evidence for multiple positive regulators of the phosphate regulon in *Escherichia coli*. Genetics **96**:353–366.
36. **Wanner, B. L., and R. McSharry.** 1982. Phosphate-controlled gene expression in *Escherichia coli* K12 using Mu*dI*-directed *lacZ* fusion. J. Mol. Biol. **158**:347–363.
37. **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.