JOURNAL OF BACTERIOLOGY, Apr. 1991, p. 2665–2672
0021-9193/91/082665-08$02.00/0

Vol. 173, No. 8

Molecular Analysis of the Cryptic and Functional *phn* Operons for Phosphonate Use in *Escherichia coli* K-12

KOZO MAKINO, SOO-KI KIM, HIDEO SHINAGAWA, MITSUKO AMEMURA, AND ATSUO NAKATA*

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

Received 18 October 1990/Accepted 7 February 1991

We cloned the cryptic *phn* operon of a K-12 strain, *phn*(EcoK), and analyzed the nucleotide sequence of the *phn* region (11,672 bp). An mRNA start site upstream of the *phnC* gene was identified by S1 nuclease mapping. The *pho* regulon activator PhoB protects a *pho* box region near the mRNA start in DNase I footprinting and methylation protection experiments. The sequence of the cryptic *phn*(EcoK) operon was very similar to that of the functional *phn* operon of an *Escherichia coli* B strain, *phn*(EcoB) (C.-M. Chen, Q.-Z. Ye, Z. Zhu, B. L. Wanner, and C. T. Walsh, J. Biol. Chem. 265:4461–4471, 1990). The *phnE*(EcoK) gene has an 8-bp insertion, absent from the *phnE*(EcoB) gene, which causes a frameshift mutation. The spontaneous activation of the cryptic *phn*(EcoK) operon is accompanied by loss of this additional 8-bp insertion. Studies of the structure, regulation, and function of the *phn* region suggest that the phosphate starvation-inducible *phn* operon consists of 14 cistrons from *phnC* to *phnP*.

The *Escherichia coli phn* locus contains the genes responsible for the use of phosphonate (Phn), including the gene for carbon-phosphorus (C-P) lyase. The *phn* locus in the K-12 strain was originally identified as a phosphate starvation-inducible gene, *psiD*, by constructing and analyzing Mud1 (*bla lacZ*) fusions, and it was renamed *phn* when it was found that the insertion of the Mu phage into the *psiD* locus made the strain unable to grow in a medium containing methylphosphonic acid (MPA) as the sole phosphorus source (26, 28). As expression of *phn* (*psiD*) is induced by P_i starvation and requires the *phoB* function, *phn* is a member of the phosphate (*pho*) regulon (26, 29). The *phn* function of the wild-type K-12 strains, *phn*(EcoK), is cryptic because K-12 strains in media that contain MPA as the sole phosphorus source begin to grow only after prolonged incubation. However, cells adapted to the MPA medium start to grow without the lag period. In a previous study, although the activating mutation near or in the *phn* genes enabled the cells to use MPA as a phosphorus source, no gross structural change between cryptic and functional *phn*(EcoK) was detected (26). The *phn* operon in *E. coli* B strains, *phn*(EcoB), is functional, and the complete nucleotide sequence of *phn*(EcoB) has been analyzed (7). *phn*(EcoB) includes the genes that encode a binding protein-dependent transport system and C-P lyase (7, 26, 27).

To analyze the primary structure of *phn*(EcoK), which is regulated by the *pho* system, and analyze the cryptic nature of *phn*(EcoK) by comparing the nucleotide sequence with that of *phn*(EcoB), we cloned *phn*(EcoK) into a plasmid and analyzed the DNA sequence. We found three tandem repeats of an 8-bp sequence in *phnE* of strain K-12, instead of the two copies of the sequence that are found in the functional *phnE* gene of B strains. Functional mutants that used MPA as the sole phosphorus source and were derived from the cryptic strain had a deletion of one copy of the repeated sequence in the *phnE* locus.

* Corresponding author.

MATERIALS AND METHODS

Bacterial strains, plasmids, and bacteriophage. *E. coli* SE5000 [F^- *araD139* Δ(*argF-lac*)U169 *rpsL150 relA1 deoC1 ptsF25 rbsR flbB5301 recA56*] was used for isolation of phage λp*lac*Mu53-inserted *phn* mutants (4). MV1184 (25) was used as the host strain for pUC118-derived plasmids. pUC9 (24) and pUC118 (25) were used as the cloning vectors. pMC1403 (6) and pKK232-8 (5), which carry the promoterless *lacZ* and *cat* genes, respectively, as reporter genes, were used for assaying promoter activities. An M13-derived phage, M13KO7 (25), was used for production of phage particles containing single-stranded DNA by superinfection of the cells carrying pUC118-derived recombinant plasmids. Bacteriophage λp*lac*Mu53 [*imm*λ *'trp'-'lacZ*$^+$*/lacY*$^+$*/lacA' 'uvrD' Xho*::Mu(*c*Its62*ner*$^+$*A*$^+$ *'S*)] was used for constructing operon fusions in vivo (4). SE5000 and λp*lac*Mu53 were kindly supplied by E. Bremer. The lambda phage clones of the ordered genomic library of *E. coli* were kindly supplied by Y. Kohara.

Media and enzyme assay. Luria-Bertani (LB) liquid medium, LB agar plates, and Tris-glucose (TG) medium supplemented with excess P_i (HP) or limiting P_i (LP) were described previously (2). The Phn phenotype of the bacteria was tested by examining their ability to grow on TG medium supplemented with 500 μM MPA. To detect β-galactosidase (β-Gal) activity of the colony, 40 μg of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) per ml was added to agar plates (20). The activities of β-Gal and chloramphenicol acetyltransferase (CAT) in the cultured cells were assayed as described previously (11, 20).

DNA manipulation. Standard recombinant DNA methods were used (8, 17). The DNA fragments with various deletions in one end of the cloned DNA were prepared by exonuclease III and mung bean nuclease treatment. The nucleotide sequence of DNA was analyzed by the dideoxy-chain termination method (22) using single-stranded M13 DNA as the template. Restriction enzymes, exonuclease III, mung bean nuclease, phage T4 DNA ligase and polymerase, and M13 DNA sequencing kits were purchased from Takara Shuzo Co., Ltd. (Kyoto, Japan).

DNase I footprinting and methylation protection and anal-

ysis of transcripts. DNase I footprinting, DNA methylation by dimethyl sulfate, and nuclease S1 mapping of the in vivo transcripts were done as described previously (14, 15). Purification of PhoB and PhoR1084 proteins and phosphorylation of PhoB by PhoR1084 in the presence of ATP were done as described previously (14).

Radioisotopes and DNA hybridization. [α-^{32}P]dCTP (>400 Ci/mmol) and [γ-^{32}P]ATP (>5,000 Ci/mmol) were purchased from Amersham, Tokyo, Japan. DNA probes for Southern hybridization were labeled by nick translation (21) and used for plaque and Southern hybridization experiments (23).

Nucleotide sequence accession number. The entire nucleotide sequence of the *phn* region (11,672 bp) has been submitted to the DDBJ, GenBank, and EMBL nucleotide sequence data bases under accession number D90227.

RESULTS AND DISCUSSION

We have cloned the entire *phn* operon of a K-12 strain into a plasmid in three steps. First, we isolated *phn* mutants in which the *phn* operon was inactivated by integration of a mutagenic phage, λp*lac*Mu53. Second, DNA fragments containing parts of the *phn* operon of the phage-lysogenized strains were identified by using the *lacZ* DNA fragment as a hybridization probe and cloned by ligation with the *'lacZY* gene and reconstruction of functional *lacZY*. Finally, the entire *phn* operon was cloned from a lambda phage clone of the *E. coli* genomic library by using the DNA fragments containing parts of the *phn* operon as probes.

Isolation of *phn*::λp*lac*Mu53 mutants. To isolate the *phn* mutant strains with *phn'-'lacZ* operon fusions, SE5000 cells were infected with λp*lac*Mu53, and kanamycin-resistant colonies were selected. About 30,000 colonies were replica plated on TGLP–X-Gal and TGHP–X-Gal plates, and 54 colonies that were Lac$^+$ on TGLP–X-Gal plates but Lac$^-$ on TGHP–X-Gal plates were isolated. They were examined for the Phn phenotype by testing their ability to grow on MPA as the sole phosphorus source after prolonged incubation (28). Three strains that did not show growth in TG-MPA medium after 60 h of incubation were isolated. The *phn* operon in these strains was likely inactivated by the insertion of the lambda phage. The β-Gal activities of SE5000, SE5001 (*phn-1*::λp*lac*Mu53), SE5010 (*phn-10*::λp*lac*Mu53), and SE5011 (*phn-11*::λp*lac*Mu53), were 0.6, 2.3, 0.5, and 1.2 Miller units, respectively, in an excess-P_i medium and 0.5, 57, 107, and 600 Miller units, respectively, in a limited-P_i medium.

Cloning of DNA fragments that contained parts of the *phn* operon. We intended to clone the DNA fragment containing the *phn'-'lacZ* operon fusions. Since the DNA segment of the MuR-*'trp'-'lacZY* region of λp*lac*Mu53 is identical to that of Mud1 (19) and the distance between the 5′ end of MuR, with which the phage is integrated into the chromosome, and the *Sst*I site in the *lacZ* gene is 3.4 kb (19; GenBank accession number M33723), the DNA fragments with the upstream *lacZ* larger than 3.4 kb obtained by digestion of the phage-lysogenized chromosomes with *Sst*I or with *Sst*I and *Eco*RI should contain some parts of the *phn* locus (Fig. 1). The chromosomal DNA was isolated from strains SE5001, SE5010, and SE5011 and digested with *Sst*I or with *Eco*RI and *Sst*I enzymes. Southern hybridization analysis was done with the *Eco*RI-*Sst*I segment of pMC1403 (6), which contained the 5′-end region of *lacZ*, as a probe. Only one DNA band hybridized with the probe DNA in each digest of the chromosomal DNA of each strain (data not shown), indicating that only one copy of the phage genome was integrated into each bacterial genome.

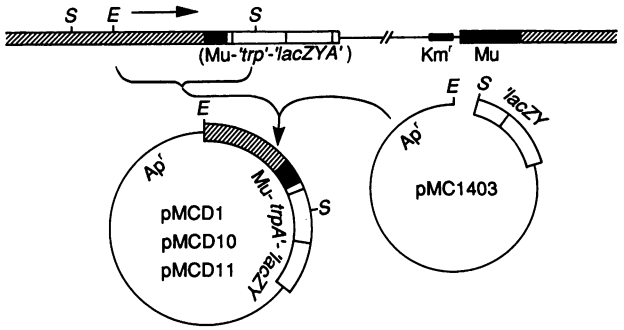

FIG. 1. Strategy for cloning of a chromosomal fragment containing the phage-integrated region into a plasmid. Hatched boxes represent chromosomal DNA; thin lines represent lambda or vector plasmid DNA; black boxes represent Mu sequences, and open boxes represent *trpA'-'lacZY*. The arrow above the map indicates the orientation of the transcription and translation of the gene with the λp*lac*M53 phage insertion. The *Eco*RI (*E*) and *Sst*I (*S*) cleavage sites upstream of the phage integrated site are hypothetical. Plasmids pMCD1, pMCD10, and pMCD11 contained the *phn'*-Mu-*'trp'-'lacZY* fusion genes that encode active β-Gal. Ap^r and Km^r represent ampicillin and kanamycin resistance genes, respectively.

The *Eco*RI-*Sst*I fragment that hybridized with the probe DNA was electroeluted and ligated with pMC1403 digested with *Eco*RI and *Sst*I to reconstruct a functional *lacZ* gene (Fig. 1). SE5000 was transformed with the ligated DNA, and ampicillin-resistant (Ap^r) and Lac$^+$ transformants were selected on LB medium–X-Gal–ampicillin plates. The Ap^r Lac$^+$ plasmids pMCD1, pMCD10, and pMCD11, reconstructed from the chromosomal DNA of SE5001, SE5010, and SE5011, respectively, contained 9.1-kb, 6.4-kb, and 5.4-kb *Eco*RI-*Sst*I fragments, respectively. The β-Gal activity of SE5000 cells carrying pMCD11 was induced by P_i starvation, but that of the cells carrying pMCD1 or pMCD10 was not (data not shown). These results suggest that the lambda phage was inserted at different sites in the *phn* locus in the three strains, and only pMCD11, but not pMCD1 or pMCD10, carried the functional *phn* promoter fused with *lacZ*.

Identification of the phage integration sites and cloning of the entire *phn*(EcoK) operon. To clone the *phn* operon, we examined the phage integration sites on the chromosome in the three lysogenic strains. The *Eco*RI-*Sst*I fragments isolated from pMCD1, pMCD10, and pMCD11, which were expected to contain parts of the *phn* operon, were hybridized with the DNAs of lambda phage clones of the ordered genomic library of *E. coli* constructed by Kohara et al. (12). The DNA fragments derived from pMCD1 and pMCD10 hybridized with clones 10H5 and 12H2, and that derived from pMCD11 hybridized with clones 12H2 and 2C12 (Fig. 2).

Since the *Eco*RI-*Sst*I fragments used for the probe DNA contained a part of the *'trp'-'lacZ'* fusion gene, they also hybridized with clones 10A6 and 7H10, which contained the *lacZ* locus, and with clones 14C4 and 4F1, which contained the *trp* locus.

The hybridization patterns of the *Eco*RI-*Sst*I fragments of these plasmids with the clones of the genomic library and the sizes of the probe DNA fragments suggest that the lambda phage was integrated into the sites about 1.9 kb from $Eco\text{RI}_1$ on the 3.6-kb $Eco\text{RI}_{1\text{-}2}$ fragment in SE5011, 3.0 kb from

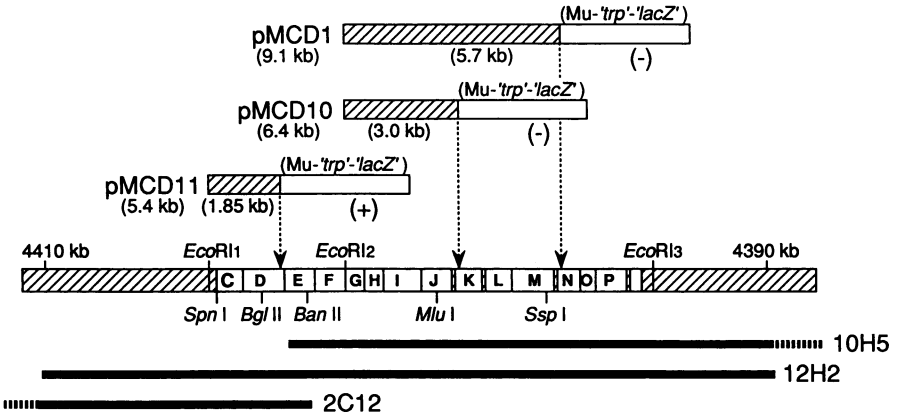

FIG. 2. λp*lac*Mu53 phage insertion sites and arrangement of the *phn* operon on the *E. coli* K-12 chromosome. Hatched boxes represent chromosomal DNA, and open boxes represent the phage-derived Mu-*'trp'-'lacZ'* segment. Arrowheads indicate λ phage insertion sites on the *E. coli* chromosome in the three strains identified by the distances between *Eco*RI and *Sst*I sites in pMCD1, pMCD10, and pMCD11. + or − represents the inducibility of β-Gal by P_i starvation when the *lacZ* gene was reconstructed by recloning the *Eco*RI-*Sst*I fragment of each plasmid into pMC1403 (6). The *phnE* gene in K-12 is cryptic (Fig. 3B). Thick black lines show the chromosomal segments in the lambda phage clones of the *E. coli* genomic library (12).

EcoRI_2 on the 8.1-kb *Eco*RI_{2-3} fragment in SE5010, and 5.7 kb from *Eco*RI_2 on the 8.1-kb *Eco*RI_{2-3} fragment in SE5001 (Fig. 2). The *phn* operon is likely to span the two *Eco*RI segments, the 3.6-kb (*Eco*RI_{1-2}) and 8.1-kb (*Eco*RI_{2-3}) fragments. This agrees with the report (28) that the *phn* operon is located around kb 4410 to 4400 on the *E. coli* restriction map (12) near 93 min in the counterclockwise orientation.

The 3.6-kb (*Eco*RI_{1-2}), 8.1-kb (*Eco*RI_{2-3}), and 11.7-kb (*Eco*RI_{1-3}) *Eco*RI fragments of clone 12H2 were subcloned into pUC9, yielding pKM1, pKM2, and pKM3, respectively, and the Phn phenotype of the plasmids was examined. The culture of SE5011 (*phn-11*::λp*lac*Mu53) carrying pKM3 in TG-MPA medium became turbid after 60 h of incubation, but those of SE5011 carrying pKM1 or pKM2 did not grow on the medium. Thus, the 11.7-kb *Eco*RI fragment in pKM3 was required for complementation of the defect in SE5011 for the Phn function.

Nucleotide sequence of the cryptic *phn*(EcoK) operon. The 3.6-kb and 8.1-kb *Eco*RI fragments in pKM3, as well as the fragment containing the *Eco*RI_2 region, were subcloned into pUC118 in both orientations, and the nucleotide sequence was analyzed for both strands by the dideoxy-chain termination method with overlapping junctions. In the sequenced region of 11,672 bp, 16 open reading frames (ORFs), which are preceded by the putative promoter sequence (Fig. 3A), were found.

The DNA sequence of the *phn*(EcoK) region was compared with that of the *phn*(EcoB) region (7). Fourteen ORFs of the *phn*(EcoK) region were very similar to the corresponding cistrons of the *phn*(EcoB) region, except for the two ORFs, *orf3* and *orf4*, which correspond to *phnE*(EcoB) (Fig. 3B). The sequence corresponding to *phnA* and *phnB* in *phn*(EcoB) were not found in the sequenced 5'-end region. These genes have been shown not to be necessary for the use of phosphonates as the sole phosphorus source (27).

In *orf3*, we found three nearly identical sequences of 8 bp that are tandemly repeated (nucleotides 2438 to 2461), but two 8-bp repeats are found in *phnE*(EcoB) (Fig. 3B). One extra copy of the sequence found in this region of K-12 alters the downstream coding frame and generates a translation stop codon at nucleotide 2601 which overlaps the next ORF, *orf4*. Deletion of any of the 8-bp sequences from nucleotide 2439 to 2461 makes the combined deduced amino acid sequences of the Orf3 and Orf4 regions completely identical to that of PhnE(EcoB). We suspect that the *phn*(EcoK) might have become functionally defective by the duplication or insertion of the 8-bp sequence in *phnE*.

orf16 lies downstream of the *phnP* gene and encodes 135 amino acids from nucleotides 11046 to 11450 (Fig. 3C). This ORF overlaps the *phnQ* open reading frame in *phn*(EcoB) (7) and extends beyond *phnQ* into downstream DNA in our clone.

We found two GC-rich regions between *phnP* and *orf16* which are very similar (Fig. 3C and 4). One 42-bp sequence (from nucleotides 10944 to 10985) is identical to 42 bp in the noncoding region between the *pstA* and *pstB* genes (1). A second 36-bp sequence which is similar to the 42-bp sequence is also similar to bases in the *melR* and *melA* intercistronic region (30). These GC-rich sequences are found in noncoding regions between two cistrons within an operon or between two operons, and their transcripts could form stem-and-loop structures. Although no function has been proved for these sequences, they might modulate gene expression by functioning as a transcriptional terminator or an attenuator. Transcription of the *phn* operon may be terminated around these GC-rich sequences.

Four repetitive-extracistronic-palindromic (REP) sequences (9) were found in the sequenced region, two between *phnD* and *phnE* (Fig. 3B) and two between *phnK* and *phnL*. Since, to our knowledge, none of the REP sequences has been shown to overlap coding regions, we assigned the beginning of the coding frame of *phnE* to nucleotide 2034. The *phnE* coding frame assigned by us is preceded by a highly conserved ribosome binding sequence, GGAG, at an appropriate position (Fig. 3B).

In total, 133 bases differ between the sequences of the *phn*(EcoK) region (11,257 bp) and the corresponding *phn*(EcoB) region, except for the extra 8 bp in *phnE*(EcoK). They are summarized in Table 1. There are 105 transition-type base substitutions and 28 transversion-type base substitutions. All the base differences are in the putative coding regions (9 bases in the first letter of the codon, 5 bases in the second letter, and 119 bases in the third letter). The two sequences differ mostly in the third letter, and this phenom-

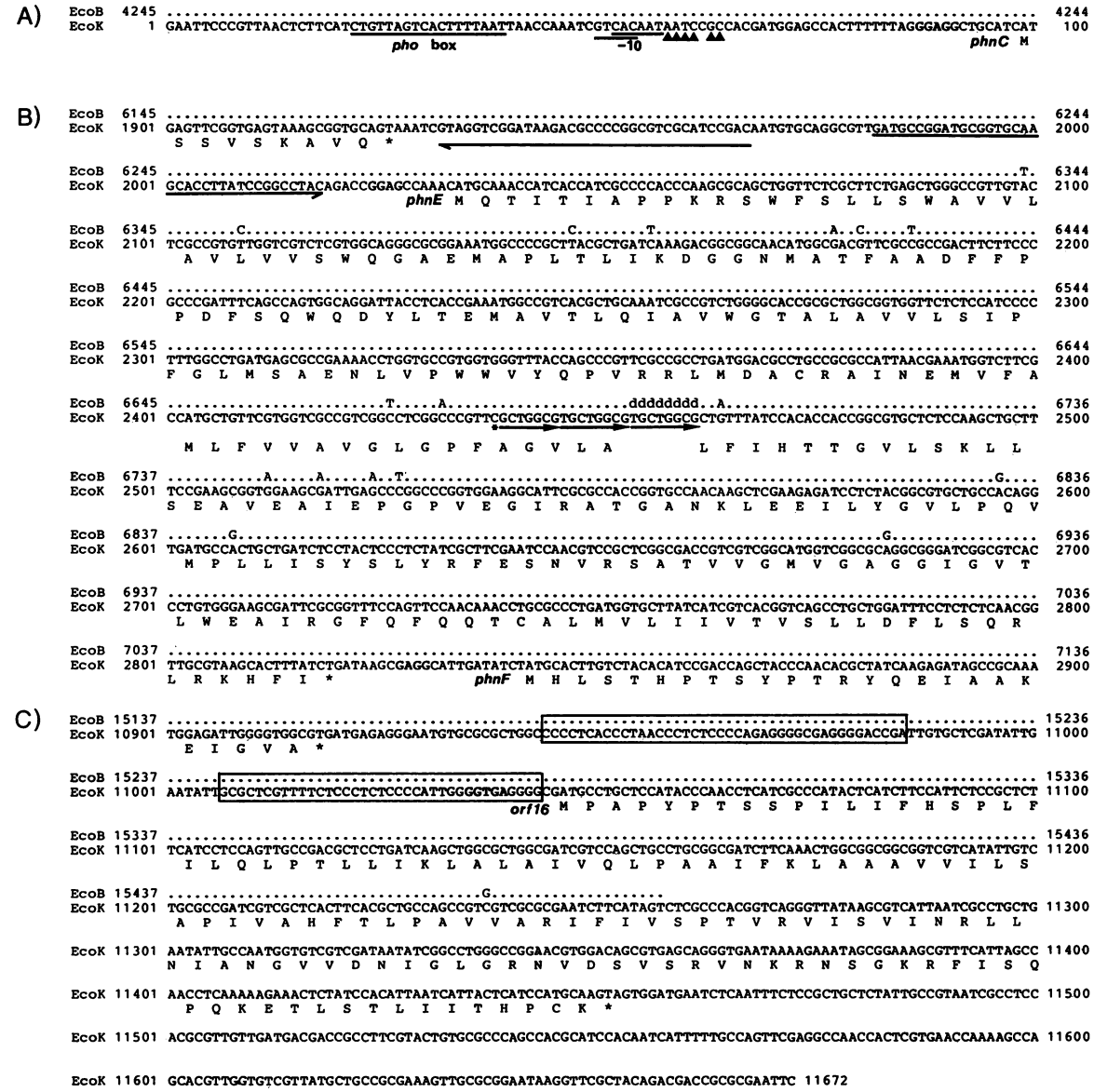

FIG. 3. DNA sequence of part of the *phn* operon of *E. coli* K-12 and B. The DNA sequence of the *phn*(EcoK) (middle line) is compared with that of *phn*(EcoB) (top line) (7). The nucleotide numbering for K-12 starts from the *Eco*RI$_1$ site. Dots represent nucleotides identical in B and K-12. The deduced amino acid sequence of *phn*(EcoK) is given on the bottom line in the single-letter code. (A) 5′-end noncoding region of the *phn* operon containing the promoter region. The *pho* box and the putative −10 regions are underlined. Triangles indicate transcription start sites. (B) *orf3* and *orf4* in *phn*(EcoK). In *orf3*, repetitions of the 8-bp sequence indicated by horizontal arrows were found. Strain B lacks one of the 8-bp units from nucleotides 2439 to 2461 (indicated with d's). Inverted half-arrows indicate REP sequences between *phnD* and *phnE*. The putative ribosome binding site for *phnE* is in boldface. (C) The 3′-end region of the *phn*(EcoK) operon. The GC-rich regions between *phnP* and *orf16* are boxed. The DNA sequence of strain B corresponding to the sequence from nucleotides 11256 to 11672 (*Eco*RI$_3$ site) of strain K-12 has not been analyzed (7).

enon minimizes differences in the amino acid sequences they code for (Table 1).

Upstream of *phnC*, an 18-bp sequence, CTGTtAgtcActTt TaAT (the *pho* box [16]), which is similar to the consensus sequence, CTGTCATA$^{A}_{T}$A$^{A}_{T}$CTGT$^{C}_{A}$A$^{C}_{T}$, shared by the promoters of the *pho* genes, was found (Fig. 3A). Downstream of the putative *pho* box with a 12-bp space, a sequence, cAcAAT, similar to the consensus −10 sequence of many promoters (10) is found. The space between the *pho* box and the −10 sequence was always 10 bp in the *pho* promoters

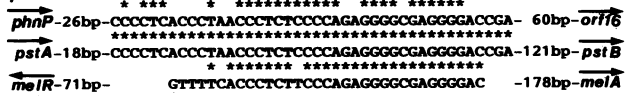

FIG. 4. High degrees of similarity of the GC-rich sequences in the noncoding region between *phnP* and *orf16* with those in the intercistronic regions between *pstA* and *pstB* and between *melR* and *melA*. The asterisks indicate identical nucleotides. The arrows indicate the direction of transcription and translation in each operon.

TABLE 1. Comparison of *phn*(EcoK) and *phn*(EcoB)

Locus	No. of[a]:	
	Nucleotides	Amino acids
phnC	787 (24)	262 (4)
phnD	1,014 (29)	338 (1)
phnE[b]	780 (17)	260 (0)
phnF	723 (1)	241 (0)
phnG	450 (1)	150 (1)
phnH	592 (10)	194 (0)
phnI	1,062 (32)	354 (3)
phnJ	843 (15)	281 (1)
phnK	756 (0)	252 (0)
phnL	678 (1)	226 (0)
phnM	1,134 (1)	348 (1)
phnN	555 (0)	184 (0)
phnO	432 (0)	144 (0)
phnP	756 (1)	252 (0)

[a] Numbers in parentheses are numbers of bases or amino acids which differ between the two strains.

[b] An insertion of 8 bp in *phnE*(EcoK) and the 5'-end 48 bp or amino-terminal 16 amino acids in *phnE*(EcoB) were not included.

examined so far (16). The sequence identified at the corresponding site is not similar to the consensus −10 sequence.

Identification of the *pho* promoter of the *phn* operon. The β-Gal activity encoded by a *phn'-'lacZ* operon fusion in pMCD11 was induced by P_i starvation, but when the 161-bp *Eco*RI$_1$-*Sph*I fragment was removed from pMCD11, β-Gal activity was no longer induced (Fig. 5). The *phn* promoter that responds to the P_i concentrations in the medium should be contained in this region.

To further analyze the regulation of the operon, the 3.6-kb *Eco*RI$_1$-*Eco*RI$_2$ fragment in pMK3 and the DNA fragments with deletions of various lengths from the 5' or 3' end of the 3.6-kb DNA were recloned into pKK232-8, a promoter-probing vector, and the promoter activities of the cloned fragment in the plasmids were assayed by measuring CAT activity. The CAT activities encoded by the plasmids containing the 3.6-kb *Eco*RI fragment or the 339-bp fragment from the 5' end were efficiently induced by P_i starvation, but no promoter activity was detected in the fragments lacking 276 bp from the 5' end (Fig. 5). These results indicate that the promoter of the *phn* operon is located in the segment spanning the *Eco*RI$_1$ site and nucleotide 339.

PhoB protein binding site. To identify the PhoB protein recognition sequence, the *pho* box, we examined the binding of PhoB to the regulatory region of the *phn* operon by

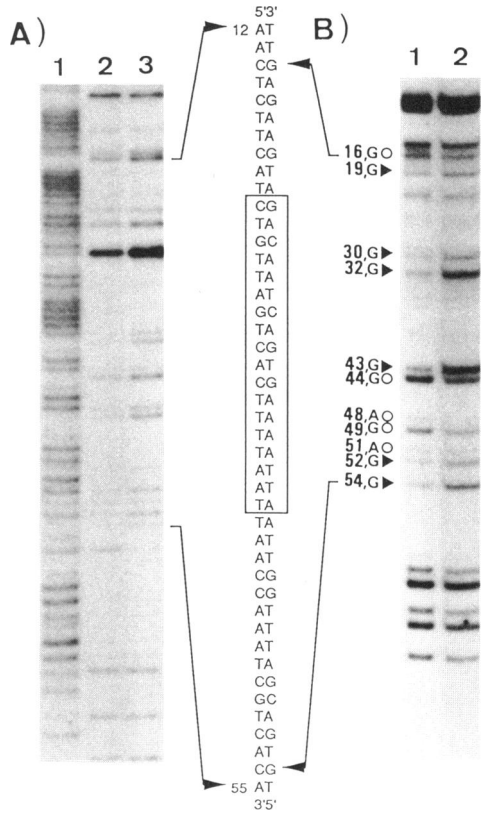

FIG. 6. Protection of the *phn* promoter region by PhoB from DNase I digestion or from methylation by dimethyl sulfate. The 281-bp *Bbe*I-*Taq*I fragment was labeled with ^{32}P at the 5' end of the *Taq*I-cleaved site for the footprinting of the transcribed strand of *phn*. (A) The labeled DNA was treated with DNase I in the presence (lane 2) or absence (lane 3) of PhoB. The ladder of the Maxam-Gilbert A+G reaction (18) of the probe DNA is shown in lane 1. The brackets flanked with arrows show the region that was protected by PhoB from DNase I digestion. (B) The labeled DNA was treated with dimethyl sulfate in the presence (lane 1) or absence (lane 2) of PhoB and then subjected to the Maxam-Gilbert G>A reaction. The bases protected from methylation and those with enhanced methylation are marked with arrowheads and open circles, respectively. The nucleotide sequence shown in the middle is the sequence of the *phn* regulatory region (Fig. 3A).

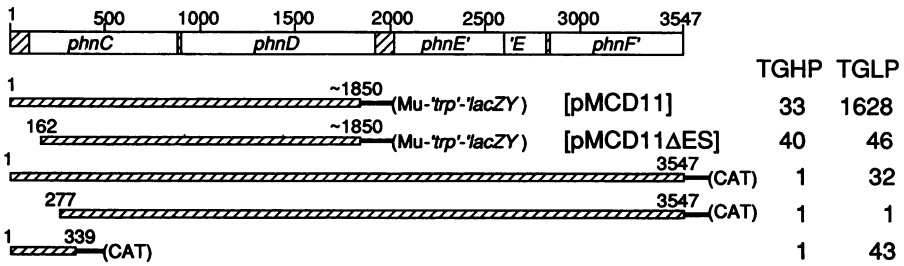

FIG. 5. Regulation of the *phn* genes by the P_i concentration in the medium. Hatched boxes represent the chromosomal fragment recloned into transcription probing vectors. The numbers above the map indicate the nucleotide numbers of the *phn*(EcoK) sequence. The activities of β-Gal and CAT in SE5011 (*phn11*::λ) carrying the plasmid containing the indicated DNA fragment grown in TGHP or TGLP medium are shown on the right. The activities of β-Gal and CAT are expressed in Miller units (20) and in nanomoles of 5-thio-2-nitrobenzoate liberated per minute per optical density unit of the cell culture at 540 nm (11), respectively.

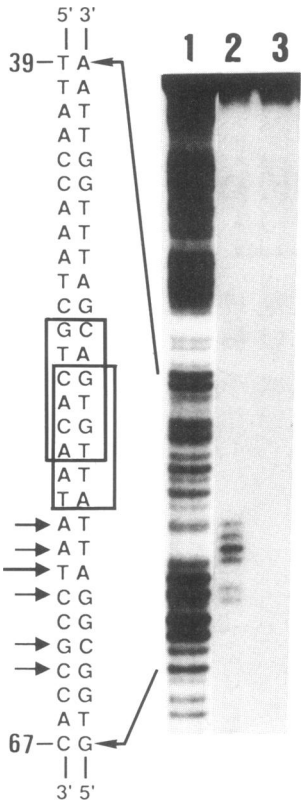

FIG. 7. S1 nuclease mapping for identification of the transcription initiation site of the *phn* operon. The end-labeled 120-bp *Eco*RI$_1$-*Taq*I fragment was hybridized with RNA extracted from the cells of strain SE5000 grown in TGLP medium (lane 2) or in TGHP medium (lane 3) and then digested with S1 nuclease. Lane 1 shows the ladder of the Maxam-Gilbert A+G reaction of the probe DNA. The nucleotide sequence of the probe is shown on the left. The putative −10 sequences are boxed. The arrows indicate the start sites of the transcripts estimated from the bands on the gel.

footprinting experiments (Fig. 6A). The region from nucleotide 12 to nucleotide 55, which contains the putative *pho* box, was protected by phosphorylated PhoB protein from DNase I digestion. The interaction of phosphorylated PhoB with this region was also examined by methylation protection experiments (Fig. 6B). Methylation of the purine bases in the *pho* box region was either inhibited or enhanced by the phosphorylated PhoB protein.

Analysis of an in vivo transcript. To identify the transcription start site of *phn*, RNA was extracted from SE5000 cells carrying pKM3 that were grown in either TGHP or TGLP medium. The S1 nuclease mapping of the transcripts was done with the 120-bp *Eco*RI$_1$-*Taq*I fragment (nucleotides 1 to 120) as a probe, and the sizes of the undigested DNA fragments were estimated by comparison with the products of the Maxam-Gilbert sequencing reaction (18) of the same DNA fragment (Fig. 7). The in vivo transcripts of *phn* were detected in RNA prepared from the cells grown in TGLP medium but not in RNA from those grown in TGHP medium. This result agrees with the fact that the transcription of *phn* is induced by P_i starvation. Several DNA bands apparently differing in length by one or two bases, which might have arisen from the formation of the secondary structure of the transcripts, were detected. A possible transcription initiation site is indicated in Fig. 3A and 7.

The *phn* operon. On the basis of the analyzed nucleotide sequence, we reexamined the lambda phage integration sites in SE5001, SE5010, and SE5011 more accurately. The 5.4-kb *Eco*RI-*Sst*I fragment in pMCD11 that was derived from SE5011 contained a *Bgl*II site at nucleotide 1648, and the distance between the *Bgl*II site and *Hin*dIII site that is 207 bp downstream of the 5'-end MuR segment in the MuR-'*trp*'-'*lacZY* region (19) was 410 bp. Therefore, the lambda phage integration site in SE5011 was estimated to be 203 bp downstream from the *Bgl*II site, or about 1.85 kb from *Eco*RI, the carboxyl-terminal region of *phnD* (Fig. 2). Similarly, pMCD10 contained the *Mlu*I site at nucleotide 5973 in *phnJ*, and pMCD1 contained the *Ssp*I site at nucleotide 8862 in *phnM* (Fig. 2). On the basis of the distance between the *Hin*dIII site in the fusion gene and the *Mlu*I or *Ssp*I site in these plasmids, the insertion sites of lambda phage were estimated to be in the amino-terminal regions of *phnK* in SE5010 and of *phnN* in SE5001, respectively (Fig. 2). Thus, the cistrons from *phnC* to *phnN* are likely to be involved in the Phn function and to compose the *phn* operon.

Since the coding region of *phnN* overlaps that of *phnO* and the noncoding space between *phnO* and *phnP* is very short, we suspected that the cistrons *phnO* and *phnP* are also involved in the *phn* operon. The SE5011 cells carrying a deletion plasmid lacking *phnP* did not grow in the TG-MPA medium, but those carrying a plasmid lacking *orf16*, which corresponds to *phnQ* in the B strain, grew in the same medium. Thus, we concluded that the *phn* operon consists of 14 cistrons from *phnC* to *phnP*. This conclusion supports the prediction that the *phn* operon probably encodes 14 ORFs (27, 28).

Mutation that activates the function of the cryptic *phnE* gene. A mutation near or in the *phn* locus was suggested to activate cryptic *phn*(EcoK) (28). We suspected that the deletion of one unit of the three tandemly repeated 8-bp sequences might activate the cryptic *phn*(EcoK) operon. To verify this hypothesis, eight overnight cultures of SE5011 (*recA56 phn-11*::λp*lac*Mu53) carrying pKM3 in T broth (2) were washed three times with TG medium to remove P_i and inoculated into fresh TG-MPA medium by diluting the cultures 100-fold. All the cultures became turbid after 60 h of incubation at 37°C with shaking. Single colonies were isolated on TG-MPA plates, and plasmid DNA was prepared from each colony. The cells of strain SE5011 transformed with the plasmids isolated from the eight adapted cultures started to grow immediately in TG-MPA medium.

Since the 8-bp repeats in the cryptic *phnE*(EcoK) gene are within the 2.4-kb *Spn*I-*Ban*II segment (Fig. 2), the *Spn*I-*Ban*II fragment in pKM3 was replaced with the corresponding fragment isolated from the adapted plasmids. The SE5011 cells carrying pKM3-derived plasmids that contained the *Spn*I-*Ban*II fragments of the eight adapted plasmids grew in TG-MPA medium without the lag time. This result indicates that the mutation(s) that activated the cryptic *phn* operon occurred within the *Spn*I-*Ban*II segment.

The *Spn*I-*Ban*II fragments were isolated from the eight adapted plasmids and recloned into M13mp18, and the nucleotide sequences of the segments were analyzed. We found a deletion of one unit of the 8-bp sequence in the region with three tandem repeats in all of the eight plasmids (Fig. 8). Therefore, we concluded that activation of the cryptic operon accompanied the deletion of the extra copy of the sequence and that the *phn*(EcoK) gene had become cryptic by duplication or insertion of the 8-bp sequence in *phnE*. The activation of the cryptic gene by the deletion of 8 bp took place in a *recA* mutant strain. Spontaneous deletion

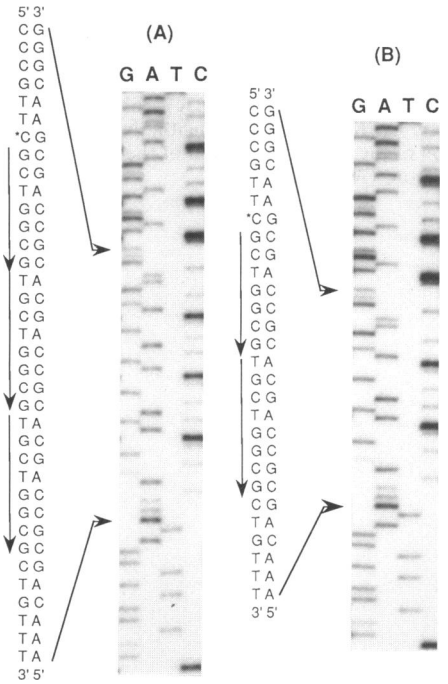

FIG. 8. Difference in the nucleotide sequences between the cryptic (A) and functional (B) *phnE* genes. Only the sequencing patterns of the regions containing the 8-bp repeated sequences in *phnE*, which are indicated by arrows, are shown. The base marked by an asterisk is different from the base at the corresponding position in the other 8-bp sequences.

or insertion of a unit of repeated sequences in the *recA* mutant strain might arise by "slipped-strand mispairing" during DNA synthesis (13). Deletion of the 8-bp sequence from the repeated region in the cryptic *phnE*(EcoK) strain should give a selective advantage in a medium containing MPA as the sole phosphorus source.

ACKNOWLEDGMENTS

We thank E. Bremer and G. M. Weinstock for kindly supplying *E. coli* and bacteriophage strains. We also thank Y. Kohara for providing the *E. coli* gene library.

This work was supported by a grant-in-aid for scientific research from the Ministry of Education, Science, and Culture of Japan. S.-K.K. is supported by a fellowship from the Rotary Yoneyama Memorial Foundation.

REFERENCES

1. **Amemura, M., K. Makino, H. Shinagawa, A. Kobayashi, 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. **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.
3. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli*, edition 8. Microbiol. Rev. **54**:130–197.
4. **Bremer, E., T. J. Silhavy, J. M. Weisemann, and G. M. Weinstock.** 1984. λ *plac*Mu: a transposable derivative of bacteriophage lambda for creating *lacZ* protein fusions in a single step. J. Bacteriol. **158**:1084–1093.
5. **Brosius, J.** 1984. Plasmid vectors for the selection of promoters. Gene **27**:151–160.
6. **Casadaban, M. J., J. Chou, and S. N. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the deletion and cloning of translational initiation signals. J. Bacteriol. **143**:971–980.
7. **Chen, C.-M., Q.-Z. Ye, Z. Zhu, B. L. Wanner, and C. T. Walsh.** 1990. Molecular biology of carbon-phosphorus bond cleavage: cloning and sequencing of the *phn*(*psiD*) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. J. Biol. Chem. **265**:4461–4471.
8. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Gilson, E., J. M. Clément, D. Perrin, and M. Hofnung.** 1987. Palindromic units: a case of highly repetitive DNA sequences in bacteria. Trends Genet. **3**:226–230.
10. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11**:2237–2255.
11. **Kimura, S., K. Makino, H. Shinagawa, M. Amemura, and A. Nakata.** 1989. Regulation of the phosphate regulon of *Escherichia coli*: characterization of the promoter of the *pstS* gene. Mol. Gen. Genet. **215**:374–380.
12. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
13. **Levinson, G., and G. A. Gutman.** 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4**:203–221.
14. **Makino, K., H. Shinagawa, M. Amemura, T. Kawamoto, M. Yamada, and A. Nakata.** 1989. Signal transduction in the phosphate regulon of *Escherichia coli* involves phosphotransfer between PhoR and PhoB proteins. J. Mol. Biol. **210**:551–559.
15. **Makino, K., H. Shinagawa, M. Amemura, S. Kimura, A. Nakata, and A. Ishihama.** 1988. Regulation of the phosphate regulon of *Escherichia coli*: activation of *pstS* transcription by PhoB protein *in vitro*. J. Mol. Biol. **203**:85–95.
16. **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.
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. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
19. **Metcalf, W. W., P. M. Steed, and B. L. Wanner.** 1990. Identification of phosphate starvation-inducible genes in *Escherichia coli* K-12 by DNA sequence analysis of *psi*::*lacZ*(Mu d1) transcriptional fusions. J. Bacteriol. **172**:3191–3200.
20. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Rigby, P. W. J., M. Diechmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
22. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
23. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
24. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
25. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**:3–11.
26. **Wackett, L. P., B. L. Wanner, C. P. Venditti, and C. T. Walsh.** 1987. Involvement of the phosphate regulon and the *psiD* locus

in carbon-phosphorus lyase activity of *Escherichia coli* K-12. J. Bacteriol. **169**:1753–1756.
27. **Wanner, B. L.** 1990. Phosphorus assimilation and its control of gene expression in *Escherichia coli*, p. 152–163. *In* G. Hauska and R. Thauer (ed.), 41st Mosbach Colloquium: the molecular basis of bacterial metabolism. Springer-Verlag, Heidelberg, Germany.
28. **Wanner, B. L., and J. A. Boline.** 1990. Mapping and molecular cloning of the *phn* (*psiD*) locus for phosphonate utilization in *Escherichia coli*. J. Bacteriol. **172**:1186–1196.
29. **Wanner, B. L., and R. McSharry.** 1982. Phosphate-controlled gene expression in *Escherichia coli* K12 using Mu*dI*-directed *lacZ* fusions. J. Mol. Biol. **158**:347–363.
30. **Webster, C., K. Kempsell, I. Booth, and S. Busby.** 1987. Organization of the regulatory region of the *Escherichia coli* melibiose operon. Gene **59**:253–263.