Nonessential Region of Bacteriophage P4: DNA Sequence, Transcription, Gene Products, and Functions

DANIELA GHISOTTI,¹ STEVEN FINKEL,^{2†} CONRAD HALLING,^{2‡} GIANNI DEHÒ,¹ GIANPIERO SIRONI,¹
AND RICHARD CALENDAR^{2*}

Dipartimento di Genetica e di Biologia dei Microrganismi, Università degli Studi di Milano, Via Celoria 26, 20133 Milan, Italy,¹ and Department of Molecular Biology, University of California, Berkeley, California 94720²

Received 12 June 1989/Accepted 11 September 1989

We sequenced the leftmost 2,640 base pairs of bacteriophage P4 DNA, thus completing the sequence of the 11,627-base-pair P4 genome. The newly sequenced region encodes three nonessential genes, which are called *gop*, β , and *cII* (in order, from left to right). The *gop* gene product kills *Escherichia coli* when the β protein is absent; the *gop* and β genes are transcribed rightward from the same promoter. The *cII* gene is transcribed leftward to a *rho*-independent terminator. Mutation of this terminator creates a temperature-sensitive phenotype, presumably owing to a defect in expression of the β gene.

Bacteriophage P4 depends for its lytic growth on a helper phage such as P2. P4 requires the helper late genes for production of capsids and tails as well as for lysis of the *Escherichia coli* host (56). The P4 genome consists of a double-stranded DNA molecule of about 11,600 base pairs (bp) (63) with 19-base 5' cohesive ends (34). The DNA sequence of the rightmost 9,084 bp of the P4 genome has been determined (11, 18, 32, 45). This DNA encodes the P4 functions required for lysogenization, DNA replication, and interactions with the helper phage (reviewed in reference 6).

The 2,600 bp of the P4 genome that remain to be sequenced lie in the P4 nonessential region (Fig. 1), which contains no genes that are essential for P4 lytic multiplication (58). This nonessential region, defined by substitutions and deletions in the left end of the P4 genome (27, 34, 48), is bounded by *cos* (which is essential for packaging P4 DNA into the phage head) and by *crr* (which is essential for P4 DNA replication [18]). The region includes *int* and *attP*, which are required for lysogenization and which have already been sequenced (45).

Mutations conferring three different phenotypes occur in the unsequenced portion of the P4 nonessential region (10, 12, 22), suggesting that this region contains at least three genes. In this study, we analyzed the DNA sequence, protein production, and transcription of this region as a first step toward a full understanding of the function of its genes.

MATERIALS AND METHODS

Media, enzymes, and chemicals. LB broth, which was used for growth of bacteria, contained 1% tryptone, 0.5% yeast extract, 0.5% NaCl, and 0.1% glucose. LC agar, which was used for plating P2 and P4 phage, contained LB broth supplemented to 2.5 mM CaCl₂ and 1% agar. M9 minimal salts medium is described by Lindahl et al. (33). Methionine assay medium was purchased from Difco Laboratories (Detroit, Mich.). Ampicillin (50 μ g/ml) or chloramphenicol (30 μ g/ml) was added to the media when indicated. Restriction enzymes were purchased from Boehringer Mannheim Bio-

chemicals (Indianapolis, Ind.), Promega Biotec (Madison, Wis.), or New England BioLabs, Inc. (Beverly, Mass.) and were used according to the instructions of the manufacturer. T4 DNA ligase, Klenow fragment, and the 17-mer –40 primer and 16-mer reverse primer were purchased from New England BioLabs. Modified T7 DNA polymerase (Sequenase) was obtained from United States Biochemical Corp. (Cleveland, Ohio). 5'-[α -³⁵S]thio-dATP (>600 Ci/mmol), L-[³⁵S]methionine (>800 Ci/mmol), and ¹⁴C-labeled protein molecular weight markers were obtained from Amersham Corp. (Arlington Heights, Ill.).

Bacteria, phage, and plasmids. Bacterial and phage strains and plasmids are described in Table 1. P4 *cIIam20*, P4 *cIIam21*, and P4 *cIIam22* were isolated after UV mutagenesis of P4 *wt* (55).

Preparation of P4 DNA. Solid CsCl was added to P4 phage in P4 storage buffer to a final density of 1.38 g/ml, and the phage were banded twice by centrifugation at 150,000 $\times$ g for 24 h. The phage band was removed and dialyzed for 1 h each against two changes of 1 liter of 10 mM Tris hydrochloride (pH 8.0)–1 mM EDTA. Linear DNA was purified from P4 phage particles by phenol extraction.

For the preparation of P4 closed circular DNA, a 400-ml culture of C-1a was grown at 37°C to a density of 4×10^8 cells per ml in LB broth supplemented with 1 mM CaCl₂ and 2 mM MgCl₂. The cells were infected with P4 phage at a multiplicity of infection of 5. After incubation with aeration for 90 min to allow replication of P4 DNA (29), closed circular P4 DNA was extracted and purified by conventional methods for the large-scale preparation of plasmid DNA (38).

Construction of P4 deletion phage. Closed circular DNAs from wild-type P4 and P4 *vir1 gop24* were digested with *Bgl*III (at positions 1706 and 2113) or with *Nhe*I (1244) and *Xba*I (1959). The fragments were self-ligated as described by Rusche and Howard-Flanders (51) and introduced into C-5204 by transfection as described by Dehò et al. (13). Phage lysates were prepared from single plaques, the DNA was extracted, and the deletions were analyzed by restriction mapping.

Insertions in P4. Closed circular DNAs from P4 *vir1* and P4 *gop24 vir1* were linearized by digestion with *Pst*I (1183), ligated to the *Pst*I fragment of pGM11 that encodes chloramphenicol acetyltransferase, and used to transform C-1a.

* Corresponding author.

† Present address: Department of Biological Chemistry, School of Medicine, University of California, Los Angeles, CA 90024.

‡ Present address: Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637.

FIG. 1. Physical map of the P4 nonessential region. (A) DNA sequencing strategy. The restriction fragments indicate the location of each mutation as determined by physical mapping, and the position of the mutation is denoted by the vertical arrow. The horizontal arrows indicate the sequence obtained in the 5' to 3' direction on each strand of DNA. The scale indicates the distance in kilobases (1,000 bp) from the left end of linear P4 DNA. (B) Strategy for mapping the 5' ends of transcripts. The probes used for S1 mapping are shown, and the 5'-labeled ends are represented by the solid diamonds. The deduced transcripts are indicated by the arrows. (C) Physical map of the *Hy1* substitution and of deletions of P4 and pCH5. The locations of P4 deletions *del1* to *del37* are from Raimondi et al. (48). The dotted lines indicate the maximum possible extent of the deletions. Only the restriction sites used for sequencing or those mentioned in the text are shown: A, *Apa*I; As, *Asa*I; B, *Bam*HI; Bg, *Bgl*II; Bs, *Bst*NI; C, *Cla*I; D, *Dra*I; E, *Eco*RI; H, *Hind*III; N, *Nde*I; Nh, *Nhe*I; P, *Pst*I; S, *Sal*I; Sc, *Scal*I; Sp, *Sph*I; X, *Xmn*I; Xb, *Xba*I.

Plasmid DNAs were extracted from chloramphenicol-resistant transformants, and their structures were determined by restriction analysis.

Construction of deletions in pCH5 by nuclease treatment. pCH5 was linearized with *Sa*I (3044), and DNA was deleted in both directions by sequential treatment with exonuclease III and S1 nuclease (24). The DNA was ligated and used to transform JM101 to ampicillin resistance.

Analysis of P4- or plasmid-coded proteins. Minicells were prepared from strain DS410 or from DS410 derivatives carrying the appropriate plasmid as described by Reeve (50). The purified minicells were suspended in M9 minimal medium containing 0.4% glucose, 10% Difco methionine assay medium, 2 μ g of methionine per ml, and 100 μ g of D-

cycloserine per ml, incubated at 37°C for 60 min, pelleted, and suspended at a concentration of 4×10^9 /ml in P4 buffer (3) supplemented with 5 mM CaCl_2 . The minicells were infected with P4 phage at a multiplicity of infection of 3, allowed to stand for 15 min at 37°C to permit adsorption by the phage, pelleted, and suspended at a concentration of 2×10^{10} /ml in M9 medium supplemented with glucose, Difco assay medium, and [35 S]methionine. The radioactively labeled proteins were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (30); the apparent M_r s of P4 *vir1*-coded proteins were determined by comparison with 14 C-labeled protein molecular weight markers (2,350 to 30,000 and 14,300 to 200,000), and these phage proteins were used as markers for further experiments.

TABLE 1. Bacterial strains, bacteriophages, and plasmids

Strain, phage, or plasmid	Characteristics	Source or reference
<i>Escherichia coli</i>		
C-1a	Prototrophic	53
C-117	Prototrophic (P2)	5
C-436	Polyauxotrophic	61
C-2322	Prototrophic (P2 <i>lg</i>)	From C-1a by lysogenization
C-2323	Polyauxotrophic (P2 <i>lg</i>)	From C-436 by lysogenization
C-5024	<i>pin-3</i> (P2)	21
C-5204	Polyauxotrophic (P2)	12
DS410	<i>min</i>	49
JM101	$\Delta(lac-proAB) thi supE F' traD36 proAB^+ lacI^a lacZ\Delta M15$	39
JM103	$\Delta(lac-proAB) rpsL thi supE endA sbcB15 hsdR^+ (P1) F' traD36 proAB^+ lacI^a lacZ\Delta M15$	17, 40
Bacteriophages		
P4 <i>wt</i>	Wild type	57
P4 <i>vir1</i>	Makes clear plaques, insensitive to P4 Immunity	35
P4 <i>org1 vir1</i>	Grows on <i>E. coli rpoA109</i> (P2)	M. Sunshine
P4 <i>Hy1 vir1</i>	Carries λ DNA in place of the DNA between the <i>EcoRI</i> sites at positions 220 and 3634	58
P4 <i>gop24 vir1</i>	Gop ⁻ (makes plaques on C-5024)	12
P4 <i>gop24 cIIam3</i>	Gop ⁻ ; makes clear plaques	10
P4 <i>cat gop24 vir1</i>	Insertion of the <i>cat</i> -containing <i>PstI</i> fragment of pGM11 into the <i>PstI</i> site (1183) of P4 <i>gop24 vir1</i>	This work
P4 <i>ts37 vir1</i>	Does not make plaques at 42°C	22
P4 <i>ts37 rev1 vir1</i>	Spontaneous Ts ⁺ Gop ⁺ revertant of P4 <i>ts37 vir1</i>	This work
P4 <i>ts37 rev2 vir1</i>	Spontaneous Ts ⁺ Gop ⁻ revertant of P4 <i>ts37 vir1</i>	This work
P4 β ts103 <i>vir1</i>	Does not make plaques at 42°C	12
P4 <i>cIIam3</i>	Makes clear plaques; suppressed by <i>supF</i>	10
P4 <i>cIIam20</i>	Makes clear plaques; suppressed by <i>supF</i>	This work
P4 <i>cIIam21</i>	Makes clear plaques; suppressed by <i>supF</i>	This work
P4 <i>cIIam22</i>	Makes clear plaques; suppressed by <i>supF</i>	This work
P4 <i>del1 vir1</i>	$\Delta(cII)$	58
P4 <i>del10 vir1</i>	$\Delta(gop-\beta)$; the <i>ClaI</i> (628) and <i>BstNI</i> (1473) sites are lost; the <i>EcoRI</i> (220) and <i>BglII</i> (1706) sites are present	27 and this work
P4 <i>del21 vir1</i>	$\Delta(int)$	48
P4 <i>del22 vir1</i>	$\Delta(cII-int)$	48
P4 <i>del23 vir1</i>	$\Delta(int)$	48
P4 <i>del26 vir1</i>	$\Delta(gop-\beta-cII)$; Gop ⁻	48
P4 <i>del34</i>	$\Delta(cII-int)$; makes clear plaques	48
P4 <i>del35</i>	$\Delta(cII)$; makes clear plaques	48
P4 <i>del37</i>	$\Delta(gop-\beta)$; Gop ⁻	48
P4 <i>del40 gop24 vir1</i>	$\Delta(BglII [1706]-BgIII [2113])$; Gop ⁻	This work
P4 <i>del41 gop24 vir1</i>	$\Delta(NheI [1244]-XbaI [1959])$; Gop ⁻	This work
M13mp18	<i>lacZ'</i>	43, 62
M13mp19	<i>lacZ'</i>	43
Plasmids		
pBR322	Ap ^r Tc ^r	8
pBR1824	Ap ^r <i>argU</i> ; a derivative of pBR322	19, 42
pCH5	Ap ^r ; <i>XmnI</i> (11272)-to- <i>XmnI</i> (4786) fragment of P4 <i>vir1</i> ligated to pUC9 <i>HincII</i>	This work
pCH51	Ap ^r ; <i>PstI</i> (1183)-to- <i>EcoRI</i> (3159) fragment of P4 <i>gop24 cIIam3</i> ligated to pUC9 <i>EcoRI PstI</i>	This work
pCH52	Ap ^r ; <i>EcoRI</i> (220)-to- <i>PstI</i> (1183) fragment of P4 <i>gop24 cIIam3</i> ligated to pUC9 <i>EcoRI PstI</i>	This work
pGM11	Ap ^r Cm ^r ; 1,870-bp <i>PstI</i> fragment containing the structural gene (<i>cat</i>) for chloramphenicol acetyltransferase from <i>lambda::Tn9</i> (37) ligated to pUC8 <i>PstI</i> ; the <i>cat</i> gene is inserted in the same direction as <i>lacZ</i>	This work
pGM42	Ap ^r ; pCH5 $\Delta(ClaI [11507] to ClaI [628])$	This work
pGM44	Ap ^r ; deletion derivative of pCH5; the deletion is 1 kilobase long and removes the <i>SphI</i> (2471) site, but the <i>HindIII</i> (2381) and <i>SphI</i> (3592) sites remain	This work
pGM45	Ap ^r ; deletion derivative of pCH5; the deletion is 500 bp long and removes the <i>EcoRI</i> (3159) and <i>SalI</i> (3044) sites but not the <i>HindIII</i> sites at positions 2381 and 3216	This work
pGM53	Ap ^r ; deletion derivative of pCH5; the deletion is 3.7 kilobases long and removes the <i>ClaI</i> (628) and <i>EcoRI</i> (3634) sites; the <i>EcoRI</i> (220) and <i>BamHI</i> (4250) sites remain	This work
pRC3	P4 <i>BssHII</i> (22, Klenow filled)- <i>AseI</i> (268) fragment ligated to pBR322 <i>SspI-AseI</i> large fragment	This work
pSF50	Ap ^r ; <i>ClaI</i> (11507)-to- <i>ClaI</i> (628) fragment of P4 <i>gop24 cIIam3</i> ligated to pUC19 <i>AccI</i>	This work
pUC8	Ap ^r <i>lacZ'</i>	60
pUC9	Ap ^r <i>lacZ'</i>	60
pUC18	Ap ^r <i>lacZ'</i>	43
pUC19	Ap ^r <i>lacZ'</i>	43, 62

Rescue of *gop24* allele. C-2322 containing a plasmid which carried a restriction fragment from P4 *gop24 cIIam3* was grown in LB broth plus ampicillin at 37°C with shaking to a density of 7×10^7 cells per ml. The cells were pelleted by centrifugation at 4°C and suspended in an equal volume of 100 mM NaCl–10 mM Tris hydrochloride (pH 7.4)–10 mM MgCl₂–5 mM CaCl₂. A 10-ml portion of the cells was transferred to a glass petri dish (100 by 20 mm) and irradiated with 10-J/m² UV light from a General Electric germicidal lamp. Irradiated cells (9 ml) were pelleted and suspended in 3 ml of LB broth containing 5 mM CaCl₂. A 2-ml portion of P4 *vir1* phage in P4 buffer at a concentration of 2×10^9 phage per ml was irradiated in a glass petri dish (60 by 15 mm) with 30-J/m² UV light. Irradiated cells (0.1 ml) were infected with 0.1 ml of phage (multiplicity of infection of 10) and incubated for 10 min at 37°C without shaking to permit adsorption. The infected cells were diluted 100-fold in LB broth and incubated at 37°C with aeration for 90 min. The progeny P4 phage were assayed on C-2323 to determine the total yield (55 to 185 per cell) and on C-5024 to determine the number of P4 *gop* mutants.

DNA sequence analysis. P4 *vir1* DNA was used to determine the wild-type sequence. P4 *gop24 cIIam3*, P4 *ts37 vir1*, and P4 *βts103 vir1* were used as the sources of DNA to sequence mutations in the *cII*, *gop*, and *β* genes. P4 restriction fragments were ligated into the polylinker region of M13mp18, M13mp19, pUC9, pUC18, or pUC19. JM103 was used as the bacterial host for recombinant M13 and pUC derivatives. Single-stranded DNA in the M13 vectors was sequenced by the method of Sanger et al. (52), and double-stranded DNA in the pUC vectors was sequenced by the method of Kraft et al. (28). For both methods we used 5'-[α-³⁵S]thio-dATP (7) and modified T7 DNA polymerase, Sequenase (59). We sequenced across all restriction sites used in cloning.

DNA sequence data were stored and analyzed by MacGene Plus (Applied Genetic Technology, Inc.) on a Macintosh Plus computer. The deduced amino acid sequences of the P4 *gop*, *β*, and *cII* proteins were compared with the protein sequences in the NBRF-PIR database (version 16.0) and with the translated DNA sequences in the NBRF-Nuc. (version 33.0) and GenBank (version 57.0) Bacteria, Phage, and Unannotated databases using the programs of Lipman and Pearson (36) on Sun Microsystems computers running the UNIX operating system. Proteins were searched for the helix-turn-helix DNA-binding structural motif of λ Cro-like proteins by the method of Dodd and Egan (15). This analysis was performed by the program MOTIF written in SBASIC for the Kaypro II computer (C. Halling, unpublished data).

S1 nuclease mapping. S1 nuclease protection mapping (4) was performed as described by Dehò et al. (14).

RESULTS AND DISCUSSION

We determined the DNA sequence of the nonessential region of P4 from *cos* to *int*, identified the proteins coded by the three genes in this region, and analyzed the transcription of these genes. Our results are summarized in Table 2.

DNA sequence. The 2,640 bp of P4 DNA from *cos* to the *int* gene were sequenced by the strategy presented in Fig. 1A. The sequence is presented in Fig. 2. This completes the DNA sequence of the 11,627-bp P4 genome. The newly sequenced region contains three open reading frames which, as described in detail below, we identified as the *gop*, *β*, and *cII* genes (Fig. 1). The *gop* and *β* genes read from left to right. The *gop* gene, which starts with an ATG codon at

TABLE 2. Features of the P4 nonessential region between *cos* and *int*

Site	Position ^a	Comments
<i>cos</i>	1–19 t	Left 19-base cohesive end ^b
<i>P_{gop}</i>	142–179 t	promoter for transcription of <i>gop</i> and <i>β</i> ; TTGAATG-N ₁₆ -TATGCA
<i>gop</i>	247–648 t	ATG-TAA; 133 amino acids; <i>M_r</i> = 14,874; product is protein H
<i>gop24</i>	541 t	Grows on <i>E. coli pin-3</i> (P2); GAG→TAG, Glu→amber
<i>β</i>	651–1718 t	ATG-TAA; 355 amino acids <i>M_r</i> = 39,864; product is protein B
<i>βts103</i>	763 t	TTA→TCA, Leu→Ser
<i>T_{cII}</i>	1749–1720 b	Proposed <i>rho</i> -independent terminator of transcription of <i>cII</i>
<i>ts37</i>	774 t	GGG→AGG, Gly→Arg in <i>β</i> ; does not confer the Ts phenotype
	1544 t	AGG→AGA, Arg→Arg in <i>β</i> ; does not confer the Ts phenotype
	1734 t	G:C→A:T transition in stem of proposed <i>T_{cII}</i> ; confers the Ts phenotype
<i>cII</i>	2540–1746 b	ATG-TAG; 264 amino acids; <i>M_r</i> = 30,134; product is protein C
<i>cIIam3</i>	1868 b	Clear plaque cistron II; AAG→TAG, Lys→amber
<i>P_{cII}</i>	2599–2571 b	Promoter for transcription of <i>cII</i> ; TGGAGT-N ₁₇ -TATAAT

^a t, Top strand; b, bottom strand.

^b Reference 34.

position 247 and ends with a TAA codon at position 648, codes for a 133-amino-acid-residue protein of predicted *M_r* 14,874. The *β* gene starts with an ATG codon at position 651, ends with a TAA codon at position 1718, and codes for a 355-residue protein of predicted *M_r* 39,864. The *cII* gene reads from right to left; thus, we cite the sequence from the bottom strand. The *cII* gene begins with an ATG codon at position 2540, ends with a TAG (amber) codon at position 1746, and codes for a 264-residue protein of predicted *M_r* 30,134.

No significant similarities were detected when the predicted amino acid residue sequences of the *gop*, *β*, and *cII* gene products were compared with sequences in three databases. According to analysis by the method of Dodd and Egan (15), none of the three proteins contains the helix-turn-helix motif characteristic of the λ Cro-like DNA-binding proteins (9, 44). The proteins also show no evidence of containing the zinc-finger structural motif (16, 41), or the leucine zipper structure (31) characteristic of other classes of DNA-binding proteins.

Proteins encoded by nonessential region of P4. The proteins synthesized after P4 infection of minicells were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (50). Six proteins were detected after infection by wild-type P4 (Fig. 3A). The electrophoretic pattern obtained with the minicell system was comparable to that observed in maxicells (1), and we indicate the proteins with the same letters. Protein A corresponds to protein 1, the product of the P4 α gene, and protein B, which has an apparent *M_r* of 45,000, corresponds to the early protein 2 (3). The genes coding for proteins C, F, H, and I (of apparent *M_r*s 31,000, 17,000, 15,000, and 12,000, respectively) have not been identified previously. None of these proteins is the product of the late *sid* (S.-J. Lee, Bachelor thesis, Harvard College, Boston, Mass., 1981) or *psu* gene (54) since the amber *sid1* and *psu1* mutations did not affect production of any of these proteins

cos

GGCGAGGCGGGGAAAGCACCACCGCGCAAAACCGACAAGTTAGTTAATTATTTGTGTAGTCAAAGTGCCTTCAGTACATACCTCGTTAATACATTGGAGCATAATGAAGAAAATCTATGGCC
 10 20 30 40 50 60 70 80 90 100 110 120
 GCGCGCGTTTGGCTGTTCAATCAATTAATAAACACATCAGTTTCACGGAAGTCATGTATGGAGCAATTATGTAACCTCGTATTACTCTTTTAGATACCGG

-35 **Pgop** -10 +1
 TATGGTCCAAACTGTCTTTT**TTGATG**GCACATATCCTGAAAAAT**ATGCA**AAAAATAGATTGATGTAAGTGGTCTTGTCTAGTGTGCGAAGATCCTTA**AGAA**TTCTGGCATGAGAGAT
 130 140 150 160 170 180 190 200 210 *EcoRI* 230 240
 ATACCAGGTTTTGACAGAAAAAACTACCGTGATAGGACTTTTTATACGTTTTTTATCTAACTACATTCCACCAAGAAGTCACAGCGTTCTAGGAATT**CTTAAG**CACCGTACTCTCTCA

gop...

M L N H V Y G L I G V V G T I I T I A T S F I R N Q D I S Q W L L V C S G W
 TAAAGGATGCTGAATCATGTATATGG**ATTAAT**TGGCGTTGTTGGGACCATAATCACTATTGCTACTAGCTTCATCCGCAACCAAGATATTTCACAATGGTTGCTCGTTTGTCTGGATGG
 250 260 *AseI* 280 290 300 310 320 330 340 350 360
 ATTTCTACGACTTAGTACATATACCT**TAATTA**ACCGCAACAACCTGGTATTAGTGATAACGATGATCGAAGTAGCGGTTGGTTCTATAAAGTGTTACCAACGAGCAACAAGACCTACC
 L A A L L I G W F T H R T I K A I S N N H T N V I K S N M E V I N S H N E S N Q
 TTAGCTGCTTTACTTATTGGATGGTTTACACACAGAAGCATTAAAGCAATTAGCAACAACACAGAAATGTTATAAAAGTAATATGGAGGTAATCAACAGTCATAATGAATCCAATCAA
 370 380 390 400 410 420 430 440 450 460 470 480
 AATCGACGAAATGAATAACCTACCAATGTGTCTTGTCTAATTCGTTAATCGTTGTTGGTGTGCTTACAATATTTTCATTATACCTCCATTAGTTGTCTAGTATTACTTAGGTTAGTT
 N L M A E N K E L I K E L A R T S E Q K E K M E S I A A Y L A T Q N P Q I N A M
 AACTTGATGGCTGAAACAAAGAATTGATTAAGAACTTGCAAGAAGCTTCAGAACAGAAAGAGAAGATGGAAAGTATAGCTGCTTATCTTGTCTACTCAGAACCCCAAAATTAACGCTATG
 490 500 510 520 530 540 550 560 570 580 590 600
 TTGAACACGACTTTTGTCTTAACTAATTTCTTGAAGCTTCTGAAGTCTTGTCTTCTCTACCTTTTCATATCGACGAATAGAACGATGAGTCTTGGGGGTTAATTGCGGATAC

...gop β...

P R T A S R P E N I D S E A N * M K V Y F E N Y S Y Y P A L R T R S A E M T G L
 CCAAGAACAGCAGCCGACCAGAAAAT**ATCGAT**TCGGAGGCTAATTAATATGAAGTATATTTTGAAATTAATCATATCTCCAGCCCTAAGAACACGATCAGCAGAAATGACCGGAC
 610 620 *ClaI* 640 650 660 670 680 690 700 710 720
 GGTTCTGTGCTCGGCTGGTCTTTT**ATAGCTA**AGCCTCCGATTAATTTATCTTTTCATATAAACTTTTAATGAGTATGATAGGTCGGGATTCTTGTCTAGTCGTCTTTACTGGCCTG

N N L S Y E N K K K I L P L I S L G K W P R S E E I Q V S L D K S L E V M S N L
 TGAATAATTTAT**CATATG**AGAATAAGAAAAATATTACCTTTAATTTCTTTGGGGAATGGCCTCGCTCGGAAGAAATACAGGTATCGTTAGATAAAAGTTTGAAGTATGAGTAATC
 730 *NdeI* 740 750 760 770 780 790 800 810 820 830 840
 ACTTATTAATAG**TATATC**CTTATTCTTTTTTATAATGGAATTAAGAAACCCCTTTACCGGAGCGAGCCTTCTTTATGTCCATAGCAATCTATTTTCAAATCTTCACTACTCATTAG

P F I L D V T K D N S H H C A S S F E L L S P E N G F K N W I E F C S R N D N I
 TCCCATTTATTTGGATGTCACAAAGGATAATTCTCATCATTCGCGCT**CAAGCTT**CGAACTACTCTCTCGTGAAGTGGTTTAAAACTGGATAGAGTTTGTTCAGAAATGATAATA
 850 860 870 880 *HindIII* 900 910 920 930 940 950 960
 AGGGTAAATAAAACCTACAGTGTCTTCTATTAAAGTAGTAACGCGCAGT**TCGAA**AGCTTGATGAGAGAGGACTCTTACCAAAATTTTACCTATCTCAAAACAAGGCTTTACTATTAT

I P V V Q M P D S A K L R D I S I Q A R V L E E L K G S I A F R I R N L N T D I
 TTATTCAGTAGTCCAGATGCCGATTACGCAAACTTAGAGATATATCTATCCAAAG**CCAGG**GTTTATAGAGAACTTAAAGGTTCAATTGCATTAGAATCTAAATACAGACA
 970 980 990 1000 1010 *BstNI* 1030 1040 1050 1060 1070 1080
 AATAAGGTATCAGGTCTACGGCTAAGTCGTTTGAATCTCTATATAGATAGGTT**CGGTCC**CAAAATCTCCTTGAATTTCCAAGTTAACGTAAATCTTAATCTTTAGATTTATGTCTGT

N K T L T S L V S M N S P E N A I V F I D L G Y I R G N V S A I T A A A I N S I
 TTAATAAAACCTGACGTCACTAGT**ATCGAT**GAACTCTCCAGAAAATGCCATAGTTTTCATAGATTGGGGCTACATTAGAGGGAATGTATCTGCTATAACAG**CTGCAG**CAATAAATTCGA
 1090 1100 *ClaI* 1120 1130 1140 1150 1160 1170 1180 *PstI* 1200
 AATTATTTGGGACTGCAGTGATCA**TAGCTA**CTTGAGAGGTCTTTTACGGTATCAAAAGTATCTAAACCGGATGTAATCTCCCTTACATAGACGATATTGT**CGACGT**CTTATTAAAGCT

N Q I R T E I P E A I I S V L A T S F P S S V T N F C R E N G Q S G Y I D V I E
 TTAATCAGATTAGAAGTGAATTCCTGAAGCCATAATTAGTGT**GTAGCAACAGCTT**CCCTAGTTCAGTTACAAATTTTGGCGTGAAATGGACAGTCCGGATATATTGATGTTATTG
 1210 1220 1230 1240 *NheI* *HindIII* 1270 1280 1290 1300 1310 1320
 AATTAGTCTAATCTTGACTCTAAGGACTTCGGTATTAATCAC**AGATCGT**TG**TCGAA**GGGATCAAGTCAATGTTTAAAAACGGCACTTTTACCTGTCAGGCTATATAACTACAATAAC

R E L H Q N I G G S D V A I Y G D H G S I H S V V Y D N I I G R Y V P R I D I A
 AGAGAGAATTGCATCAAAATATTGGTGAAGTATGTCGCTATTATGGCGATCATGGTTCAATTCATTAGTGTATGACAACATAATTGGGCGATATGTTCTAGAATTGATATAG
 1330 1340 1350 1360 1370 1380 1390 1400 1410 1420 1430 1440
 TCTCTCTTAACGTAGTTTTATAACACCTTCACTACAGCGATAAATACCGCTAGTACCAAGTTAAGTAAGTCATCACATACTGTTGTATTAACCCGCTATACAAGGATCTTAACATATAC

FIG. 2. DNA sequence of the P4 nonessential region between *cos* and *int*. Restriction sites used for sequencing or mentioned in the text are underlined on both strands. The deduced amino acid sequences of the *gop*, β , and *cII* genes are shown. The +1, -10, and -35 positions of the promoter for transcription of *gop* and β and of the promoter for transcription of *cII* are presented in bold face. The arrows indicate the stem of the proposed *rho*-independent terminator downstream of the *cII* gene. This DNA sequence has been submitted to GenBank under the accession number M27748.

FIG. 2—Continued.

(data not shown). P4 *Hyl vir1*, in which the entire nonessential region between the *Eco*RI sites at positions 220 and 3634 is replaced by a portion of the phage λ genome (Fig. 1) (58), did not direct the synthesis of proteins B, C, and H (Fig. 3A). Additional polypeptides produced after infection by this mutant may be encoded by the inserted λ DNA fragment. pCH5, which carries the 5,141-bp *Xmn*I (11272 to 4786)

fragment from P4 *vir1* that includes the entire nonessential region (Fig. 1), directed the synthesis of the B, C, and H proteins (Fig. 3B). The two additional proteins observed with molecular masses similar to that of protein C were also produced by cells containing the pUC9 vector (Fig. 3B). The results obtained with P4 *Hyl vir1* and pCH5 complement each other, indicating that proteins B, C, and H are encoded

FIG. 3. P4 proteins synthesized by the nonessential region. (A) Minicells of strain DS410. Lanes: 1, noninfected; 2, infected with P4 *vir1*; 3, P4 wild type; 4, P4 *Hyl vir1*; 5, P4 *cIIam3*; 6, P4 *gop24 cIIam3*; 7, P4 *cIIam20*; 8, P4 *cIIam21*; 9, P4 *cIIam22*. (B) Minicells of strain DS410 infected with P4 *vir1* (lane 1), DS410/pUC9 (lane 2), DS410/pCH5 (lane 3), DS410/pGM45 (lane 4), DS410/pGM42 (lane 5), DS410/pGM44 (lane 6), DS410/pGM53 (lane 7). (C) Minicells of strain DS410 infected with the following: lane 1, P4 *del10 vir1*; lane 2, P4 *del26 vir1*; lane 3, P4 *del22 vir1*; lane 4, P4 *del21 vir1*; lane 5, P4 *del23 vir1*; lane 6, P4 *del1*; lane 7, P4 *del34*; lane 8, P4 *del35*; lane 9, P4 *del37*; lane 10, P4 wild type; lane 11, P4 *vir1*; lane 12, P4 *gop24 vir1*; lane 13, P4 *del40 gop24 vir1*; lane 14, P4 *del41 gop24 vir1*.

by the nonessential region of P4. Cells containing pGM42, a derivative of pCH5 in which the *Cla*I restriction fragment from positions 11,507 to 628 had been deleted, synthesized neither protein B nor protein H, indicating that the genes coding for these proteins lie near the left end of P4. On the basis of the sizes of the open reading frames, we tentatively concluded that protein H is coded by *gop*, protein B by β , and protein C by *cII*. These conclusions were confirmed by the experiments described below.

The *cII* gene. P4 *cII* mutants have a clear-plaque phenotype and kill *E. coli* after infection (10). The *del1*, *del22*, *del34*, and *del35* mutations are believed to affect the *cII* gene since they confer a clear-plaque phenotype (48; unpublished data). The positions of these deletions indicated that *cII* is the open reading frame at the immediate left of *int* (Fig. 1). We confirmed that protein C, which has an apparent M_r of 31,000, is the product of the *cII* gene by analyzing the pattern of proteins produced after infection by P4 *cII* or P4 deletion mutants. None of the four P4 *cII* amber mutants tested could direct the production of protein C (Fig. 3A). We also found that protein C was not produced after infection by P4 mutants carrying the *del1*, *del22*, *del26*, *del34*, *del35*, *del40*,

or *del41* deletion, whereas protein C was produced after infection by P4 mutants carrying the *del10*, *del21*, *del23*, or *del37* deletion (Fig. 3C). Furthermore, pGM42 synthesized protein C, whereas pGM53 did not (Fig. 3B). The positions of these deletions are consistent with the deduced location of the *cII* gene (Fig. 1).

The amber fragment of protein C synthesized after infection by P4 *cIIam3* had an apparent M_r of 27,000 (Fig. 3A). This size suggested that the amber mutation was located in the small *Bgl*II restriction fragment between positions 1706 and 2113. We cloned and sequenced this fragment from P4 *gop24 cIIam3* and found an A-to-T transversion at position 1868 that creates an amber mutation (AAG to TAG) in *cII* (Table 2). We wished to confirm that this amber mutation caused the clear-plaque phenotype. Thus, we removed the *Bgl*II fragment (1706 to 2113) from P4 *gop24 cIIam3* and replaced it with the *Bgl*II fragment from P4 *org1 vir1*. After transfection of C-2323, we obtained four plaques, all of which were turbid (Table 3, experiment 1), in confirmation of our sequence analysis.

The M_r -20,000 and -27,000 polypeptides synthesized after infection by P4 *cIIam20* and P4 *cIIam22*, respectively, are

TABLE 3. Phenotypes of P4 recombinants after fragment exchange^a

Expt	Digest (site positions)	Fragment source and size (bp)		Pheno- type ^b
1	<i>Bgl</i> III (1706; 2113)	P4 <i>gop24 cIIam3</i> 11,220	P4 <i>org1 vir1</i> 407	C ⁺
2	<i>Apa</i> LI (10652) + <i>Pst</i> I (1183)	P4 <i>ts37 vir1</i> 9,469	P4 <i>org1 vir1</i> 2,158	Ts
		2,158	9,469	Ts ⁺
	<i>Pst</i> I (1183) + <i>Sca</i> I (1765)	11,045	582	Ts ⁺
		582	11,045	Ts
	<i>Bgl</i> III (1706; 2113)	11,220	407	Ts ⁺
3	<i>Bgl</i> III (1706; 2113)	P4 <i>ts37 vir1</i> 11,220	P4 <i>vir1</i> 407	Ts ⁺
		407	11,220	Ts
4	<i>Apa</i> LI (10652) + <i>Pst</i> I (1183)	P4 <i>βts103 vir1</i> 9,469	P4 <i>org1 vir1</i> 2,158	Ts ⁺
		2,158	9,469	Ts
	<i>Pst</i> I (1183) + <i>Sca</i> I (1765)	11,045	582	Ts
		582	11,045	Ts ⁺
	<i>Nde</i> I (733; 8638)	3,721	7,906	Ts ⁺
		7,906	3,721	Ts

^a P4 DNA was digested, and the fragments were isolated and ligated in low-melting-temperature agarose. Phage phenotypes were determined after transfection of C-2323 with the ligated DNA.

^b C⁺, Makes turbid plaques on C-2323; Ts, temperature sensitive, unable to make plaques on C-2323 at 42°C; Ts⁺, makes plaques on C-2323 at 42°C.

probably the protein C amber fragments (Fig. 3A). Thus, *cIIam22* maps very close to *cIIam3*, and *cIIam20* maps between the ATG codon and *cIIam3*. P4 *del40 gop24 vir1*, which was derived from P4 *gop24 vir1* by deletion of the *Bgl*III fragment from positions 1706 to 2113, was predicted to direct the synthesis of a hybrid *cII* protein of *M_r* 19,523 that resulted from fusion of the 5' portion of *cII* to an open reading frame that continues to an amber codon at positions 1617 to 1615. We found that cells infected by the P4 *del40* mutant did not produce protein C but did produce a protein of apparent *M_r* 20,000, confirming the prediction (Fig. 3C). Similarly, the hybrid *cII* protein synthesized from P4 *del41 gop24 vir1*, which was derived from P4 *gop24 vir1* by deletion of the *Nhe*I (1244)-to-*Xba*I (1959) fragment, was predicted to have an *M_r* of 23,284. Cells infected by the P4 *del41* mutant produced a protein of apparent *M_r* 25,000 in place of protein C, confirming the prediction (Fig. 3C). Cells infected by P4 *del26 vir1* did not produce protein C but did produce a protein of apparent *M_r* 32,000, which is larger than protein C (Fig. 3C). The endpoints of *del26* are not well characterized, but examination of the DNA sequence revealed that the *cII* open reading frame could be fused to any of three short open reading frames to create a hybrid gene capable of coding for a protein larger than protein C.

Transcription of *cII*. To determine the location of the promoter for the *cII* gene, we performed S1 nuclease protection mapping. A P4 *Bgl*III (2113)-to-*Eco*RI (3159) DNA restriction fragment that was 5' end labeled at the *Bgl*III end (Fig. 1B) was denatured, hybridized with RNA from P4-infected cells, P4 lysogenic cells, or cells carrying P4 as a plasmid, and treated with S1 nuclease. RNA from cells containing P4 as a plasmid protected the DNA to position

FIG. 4. S1 nuclease protection mapping of the promoter for gene *cII*. The first four lanes are Maxam and Gilbert sequencing reactions of the probe described in the text. The reactions are (in order) G, Pu (G + A), Py (T + C), and C. The remaining lanes are DNA fragments protected by hybridization to RNA from C-1a infected with P4, 20 min after infection (lane 5), C-1a infected with P4, 90 min after infection (lane 6); C-1a carrying a P4 prophage (lane 7); C-1a carrying P4 as a plasmid (lane 8); C-1a (lane 9).

2560 (Fig. 4). A longer autoradiographic exposure revealed that the same DNA fragment was protected by RNA isolated from cells at 20 and 90 min after P4 infection but that it was not protected by RNA from a P4 lysogenic strain. We therefore propose that the promoter for the *cII* gene is defined by the -35 sequence TGGAGT beginning at 2599 and the consensus -10 sequence TATAAT, separated by a

FIG. 5. Putative *rho*-independent terminator of transcription of the *cII* gene. The position of the *ts37* mutation is indicated.

17-nucleotide spacer (Fig. 2). The fainter protected signals observed in the plasmid condition may be due to additional upstream transcription initiation points. However, our experiments do not rule out the possibility that the 5' end of the RNA is the product of the processing of a longer transcript. The putative transcription terminator for *cII* is a hairpin loop at positions 1749 to 1726 followed by a run of six Ts at 1725 to 1720 (Fig. 2 and 5).

The *gop* gene. The *pin-3* mutation of *E. coli* blocks lytic multiplication of P4 in a P2 lysogen (21). In the absence of a prophage helper, infection by P4 greatly reduces the survival of *E. coli pin-3* mutants compared with *pin*⁺ strains (10). The P4 *gop* mutations allow P4 to multiply in a P2 lysogenic *pin-3* host (Gop[−] phenotype); in addition, they suppress the killing of *pin-3* cells by P4 in the absence of a P2 helper prophage (10, 21). The *Hy1* substitution and the *del10*, *del26*, and *del37* deletions all confer the Gop[−] phenotype (21, 48). The positions of these deletions suggest that the *gop* gene is located near the left end of the P4 genome (Fig. 1C). The *gop* point mutations also have been mapped by three-factor crosses to the same region (10, 12).

The *gop24* mutation eliminated the synthesis of protein H (compare lane 5 with lane 6 of Fig. 3A and lane 11 with lane 12 of Fig. 3C). The deletions that confer the Gop[−] phenotype (*del10*, *del26*, *del37*) also eliminated or reduced the signal corresponding to protein H (Fig. 3C). These results suggested that protein H is the *gop* gene product, and the size of protein H indicates that the *gop* gene is the open reading frame next to *cos* (Fig. 1).

The identification of the *gop* gene was confirmed by sequencing and mapping the *gop24* mutation. When we sequenced the entire *gop* gene from P4 *gop24 cIIam3* (Fig. 1A), we found that the *gop24* mutation was a G-to-T transversion at position 541, creating an amber mutation in *gop* (Table 2). To confirm the physical location of *gop24*, we cloned three contiguous or overlapping fragments of P4 *gop24 cIIam3* that spanned the region from *cos* to the middle of *int* into plasmid vectors. We then attempted to obtain P4 *gop* mutant phage through recombination of P4 *vir1* with the plasmids. The results demonstrated that the *gop24* mutation was located between the *EcoRI* (220) and the *Clal* (628) restriction sites (Table 4). Since the only change detected by DNA sequencing between these two sites was the transversion at position 541, this amber mutation was confirmed as the *gop24* mutation. However, the *gop24* mutation is not suppressed by the amber suppression *supE* (12, 21). This might be due to the amino acid substitution performed by

TABLE 4. Generation of P4 *gop24 vir1* phage by recombination of P4 *vir1* with plasmids carrying fragments of P4 *gop24 cIIam3*^a

Plasmid ^b	Fragment ^c	Frequency Gop [−] /Gop ⁺ (10 ^{−5})	Relative frequency ^d
None	None	1.1	1.0
pUC18	None	0.9	0.8
pSF50	<i>Clal</i> (11507)- <i>Clal</i> (628)	40	36
pCH52	<i>EcoRI</i> (220)- <i>PstI</i> (1183)	72	65
pCH51	<i>PstI</i> (1183)- <i>EcoRI</i> (3159)	1.2	1.1

^a Marker rescue experiments were done as described in Materials and Methods.

^b Plasmids were contained in C-2322.

^c See Fig. 1 for the positions of these fragments.

^d Frequency Gop[−]/Gop⁺ relative to frequency obtained in C-2322.

supE (Gln instead of Glu) and/or to its relatively low suppression efficiency (20). Even *gltT*, a suppressor inserting the correct amino acid Glu but at low efficiency (46, 47), does not suppress *gop24* (data not shown). Thus, we believe that a low level of functional *gop* gene product is not sufficient to confer the wild phenotype.

The *β* gene. The *ts37* and *ts103* mutations confer a lethal temperature-sensitive (Ts) phenotype on P4 (12, 22). At 42°C, P4 *ts37 vir1* kills *E. coli* strains that are not lysogenic for P2; the killing is accompanied by a reduction in the synthesis of DNA, RNA, and protein (22). Temperature-resistant revertants of P4 *ts37 vir1* occur at a frequency of about 5 × 10^{−5}, and most revertants acquire the Gop[−] phenotype, as discussed in greater detail below. Since the reversion frequency of these mutants is comparable to the frequency of recombination in P4, it has been difficult to map *ts37* and *ts103* mutations by multifactor crosses (12). Such crosses have shown only that the *ts37* and *ts103* mutations are located near each other outside the α-to-*vir1* region (12), but neither mutation has been mapped in relation to *sid*, *δ*, *psu*, *gop*, or *cII* mutations. The *ts37* and *ts103* mutations were believed to define a gene called β (12, 22).

P4 *ts37 vir1* was recessive to P4 *vir1* for cell killing (Table 5), suggesting that P4 *vir1* complements P4 *ts37 vir1* by supplying the wild-type β function. However, P4 *ts37 vir1* was not complemented by P4 *del10 vir1* (Table 5), indicating that the *del10* deletion inactivated the β gene. Cells infected by P4 *del10 vir1* produced neither *gop* protein (protein H) nor protein B (Fig. 3C). Since P4 *ts37 vir1* was complemented by P4 *gop24 vir1* (Table 5), the β gene is distinct from *gop* and probably codes for protein B. The open reading frame between *gop* and *cII* has a size appropriate for coding protein B, suggesting that this open reading frame is the β gene.

TABLE 5. Cell survival after P4 infection^a

Infecting phage	% Survival
P4 <i>ts37</i>	0.2
P4.....	91
P4 <i>gop24</i>	100
P4 <i>del10</i>	96
P4 <i>ts37 rev1</i>	100
P4 <i>ts37 rev2</i>	99
P4 <i>ts37</i> + P4.....	74
P4 <i>ts37</i> + P4 <i>gop24</i>	78
P4 <i>ts37</i> + P4 <i>del10</i>	1.4
P4 <i>ts37</i> + P4 <i>ts37 rev1</i>	89
P4 <i>ts37</i> + P4 <i>ts37 rev2</i>	5.5

^a All phage carried the *vir1* mutation. *E. coli* C-1a was infected at a multiplicity of infection 10 and plated at 42°C.

We attempted to map the *ts37* and *ts103* mutations by marker rescue. We were able to demonstrate rescue of both mutations by plasmid pCH5, which carries the 5,141-bp *XmnI* (11272 to 4786) fragment from P4 *vir1* (data not shown). However, we were unable to detect rescue by plasmids carrying smaller fragments of P4 because the frequency of reversion of the mutations was higher than the frequency of recombination of P4 *ts103 vir1* with the plasmids (data not shown).

Since we believed that the *ts37* mutation would define the β gene as that coding for protein B, we sequenced across the entire protein B open reading frame from P4 *ts37 vir1* and found three G-to-A transitions (Table 2). One mutation, at position 744, caused a Gly-to-Arg change; the second mutation, at position 1544, did not lead to an amino acid change; and the third mutation, at position 1734, occurred outside of the protein B open reading frame in the stem of the putative *rho*-independent terminator for transcription of the *cII* gene.

To determine which of the three transitions conferred the temperature-sensitive phenotype, we mapped the mutation that caused thermosensitivity by exchanging DNA restriction fragments between P4 *ts37 vir1* and P4 *org1 vir1* or P4 *vir1*. The results of these experiments are presented in Table 3, experiments 2 and 3. We first found that the mutation conferring the temperature-sensitive phenotype mapped between the *PstI* (1183) and *ScaI* (1765) sites and was therefore not the transition that changed the amino acid sequence of protein B. We then found that the 407-bp *BglII* (1706 to 2113) fragment from either P4 *org1 vir1* or P4 *vir1* was sufficient to confer the wild-type phenotype when ligated to the 11,220-bp *BglII* fragment from P4 *ts37 vir1*. We were also able to isolate temperature-sensitive recombinants in which the small *BglII* fragment from P4 *ts37 vir1* was ligated to the large *BglII* fragment from P4 *vir1*. We sequenced the β gene region of one such temperature-sensitive recombinant and confirmed that it is mutant only at position 1734. These results eliminated the possibility that the transition at 1544 in the protein B open reading frame was responsible for the temperature-sensitive phenotype, and we concluded that the *ts37* mutation lies outside of the protein B gene at position 1734 in the stem of the putative terminator for *cII* transcription (Fig. 5).

Since *ts37* was located outside of what we believed to be the β gene, we mapped and sequenced *ts103* (12). By exchanging restriction fragments between P4 *ts103 vir1* and P4 *org1 vir1*, we found that the *ts103* mutation was located between the *NdeI* (733) site and the *PstI* (1183) site (Table 3, experiment 4). We sequenced this region of P4 *ts103 vir1* and found a single change at position 763 that changed a Leu codon to a Ser codon in the protein B open reading frame. Hence, the β gene is defined by the *ts103* mutation, and β is the gene coding for protein B.

Common transcript of β and *gop* genes. The *gop* and β genes are transcribed rightward. To locate the origin of transcription, we prepared a probe for S1 mapping by cutting covalently closed circular P4 wild-type DNA with *NheI* at position 1244. The 5' ends were dephosphorylated and labeled with ³²P, and the DNA was cut with *BamHI* at position 10658. The fragment containing the *gop* and β genes was isolated and hybridized with RNA from P4-infected cells, P4 lysogenic cells, or P4 plasmid-carrying cells (data not shown). The major protected bands of about 1,000 nucleotides suggested a rightward transcription initiation point upstream of *gop* and close to the *EcoRI* (220) site. Weaker signals may indicate possible minor transcription initiation points at 250, 460, 580, and 750 nucleotides, but

any of the signals may represent processing of a longer transcript. From the length of the longest transcript, we are able to conclude that *gop* and β are cotranscribed.

To define the promoter more precisely, we performed short-range S1 mapping with a probe obtained from pRC3 digested at position 269 with *AseI*, 5' end labeled with ³²P, and further digested with *AatII*. The 362-bp fragment obtained contained, on the 5'-end-labeled strand, the P4 nucleotides 270 to 22 plus an additional 113 nucleotides from the vector at the 3' end. Hybridization with P4 RNA gave the strongest protection at nucleotide 179 of the P4 sequence, after correction for the difference in migration with Maxam and Gilbert sequencing reactions (Fig. 6). Sequence analysis of the upstream region showed a consensus with the *E. coli* σ^{70} promoter: TTGATG-N₁₆-TATGCA-N₉-start. Such a consensus appears to be sufficient to allow expression of the *gop*- β operon early after infection; in the plasmid condition, however, the protected signal was much stronger, although the P4 copy number is lower in the plasmid state than 20 min postinfection (1). This suggests that promoter *P*_{gop} might be positively regulated in the plasmid state. *P*_{gop} activity was also detectable in the lysogenic condition, although at a low level. Additional faint protection signals were observed at nucleotides 152 to 153 (in the lysogen and the infected cells, early and late, but not in the plasmid state) and at position 107 to 108 (in all conditions), indicating possible weak promoters. In addition, we observed protection of the complete P4 portion of the probe by the P4 plasmid RNA. This indicated that in the plasmid condition some P4 rightward transcription continues beyond the late genes on the right arm of P4 (11) across the P4 *cos* site and into the *gop* gene.

Since we have not shown that these transcripts carry 5' triphosphates, we may be observing processed RNA species.

Insertion mutants obtained by random cloning of heterologous DNA into the *EcoRI* (220) site have been selected for their *Gop*⁻ phenotype (48). The proteins synthesized in minicells infected by a number of these mutants were analyzed (data not shown). In some of the mutants protein H was missing; in others, protein H was synthesized at a very reduced level. These results suggested that insertion in *EcoRI* (220) affects transcription of the *gop*- β operon. However, in the insertion mutants, the level of synthesis of protein B was normal. This suggests that there is an alternative promoter or promoters for the β gene.

Relationship of *gop* and β . We found that a majority of spontaneous *Ts*⁺ revertants of P4 *ts37* had also acquired the *Gop*⁻ phenotype; in fact, we tested 763 independent *Ts*⁺ revertants and found that 758 (99.3%) were *Gop*⁻. We carried out complementation analysis of P4 *ts37* with two *Ts*⁺ revertants, one with a *Gop*⁺ phenotype (P4 *ts37 rev1*) and one with a *Gop*⁻ phenotype (P4 *ts37 rev2*). We found that P4 *ts37 rev1* could complement P4 *ts37*, whereas P4 *ts37 rev2* could not (Table 5). These results indicate that only the *Gop*⁺ *Ts*⁺ revertants are true revertants and that *Gop*⁻ *Ts*⁺ revertants carry extragenic suppressors of the *ts37* mutation. In addition, we analyzed 50 spontaneous *gop* mutants obtained by plating P4 *ts37* on *E. coli pin-3* (P2) at 30°C, and in every case the *gop* mutation suppressed the temperature-sensitive phenotype.

All spontaneous P4 deletion mutants that failed to synthesize the β gene product (protein B) also lacked the *gop* gene product (protein H) (*del10*, *del26*, and *del37*; Fig. 3C). To study the functional interaction between *gop* and β , we tried to delete the β gene from pCH5 and P4. Plasmid pCH5 was linearized with *SalI* (3044), which cuts to the right of gene *cII*

FIG. 6. Short-range S1 mapping of the promoter for the *gop* and β genes. The probe was prepared as described in the text. Lanes 1 to 4, Maxam-Gilbert sequencing reactions of the probe: G, Pu (G + A), Py (T + C), C. Lanes 5 to 8, Probe protected by RNA extracted from C-1a carrying P4 as a plasmid (lane 5), C-1a carrying a P4 prophage (lane 6), C-1a infected with P4, 20 min after infection (lane 7), C-1a infected with P4, 40 min after infection (lane 8).

(Fig. 1A), and the cut was extended by treatment with exonuclease III for different time periods to obtain deletions of different length, as described in Materials and Methods. We analyzed 30 deletion derivatives of pCH5 and found two classes: class 1 deletions conserved the *Hind*III (2381) site and thus did not cover β ; class 2 deletions were much longer and all covered the *Cla*I (628) site, thus deleting all of β and part of *gop*. Examples of class 1 (pGM44, pGM45) and class 2 (pGM53) deletions are shown in Fig. 1C. Analysis of the proteins synthesized by these plasmids demonstrated that class 1 deletions produced both protein B and protein H, whereas class 2 deletions failed to produce either protein B

or protein H (Fig. 3B). This suggests that the *gop* gene product is lethal in the absence of the β gene product.

According to this hypothesis, deletions which cover gene β should be obtained only in a *gop* mutant genetic background. To test this model, we digested circular P4 *vir*1 DNA with *Nhe*I and *Xba*I to remove nucleotides 1244 to 1959. After ligation of the compatible ends, we failed to obtain viable phage carrying this deletion. However, we were able to delete the *Nhe*I-*Xba*I fragment in P4 *gop*24 *vir*1 to make viable P4 *del*41 *gop*24 *vir*1 (Fig. 1C). This deletion removed the carboxyl half of the β gene; however, it also removed the transcription terminator for gene *cII*, as well as part of the *cII* gene. To target only the β gene, we tried to insert the *cat* gene (37) into the *Pst*I site at 1183. When P4 *vir*1 was used as recipient, the frequency of transformation was very low and all nine of the chloramphenicol-resistant transformants obtained carried plasmids that had suffered gross rearrangements. Starting with P4 *gop*24 *vir*1, however, we could isolate the desired insertion mutants (seven of nine plaques analyzed). These experiments supported the idea that the *gop* gene product is lethal in the absence of a functional β gene. Thus, the β protein acts as a negative controlling element for the *gop* protein.

The *ts*37 mutation disrupted a potential base pair in the stem of the proposed terminator for transcription of gene *cII* (Fig. 5). We suggest that at high temperature, the *ts*37 mutation causes transcription to read through this terminator, interfering with the expression of the β gene. According to this hypothesis, deletions which cover the *cII* terminator could lower the level of the β protein, thus being lethal in a *gop*⁺ background. To test this proposition, we digested circular P4 *vir*1 DNA with *Bgl*II, which cuts at 1706 and 2113, thus deleting the terminator but leaving the β gene practically intact (the β protein is shortened only for the last amino acid by the *Bgl*II deletion; Fig. 2). The DNA digested with *Bgl*II was religated, and P2 lysogenic *E. coli* was transfected. The efficiency of transfection was very low, and most phage produced did not carry any deletion. Only three deletion mutants were found; all three had acquired the *Gop*⁻ phenotype and the extent of the deletion was greater than expected. On the other hand, we obtained several *Bgl*II deletion mutants from P4 *gop*24 *vir*1. Analysis of the protein synthesized in minicells showed that protein B was produced less abundantly from the deleted phage (P4 *del*40 *gop*24 *vir*1; Fig. 3C) than from P4 *vir*1, confirming that the deletion of the *cII* terminator alters β expression.

The β protein may form a complex with the *gop* protein and reduce its activity, or the β protein may reduce transcription or translation of the *gop* gene.

Complementation of the *pin*-3 mutation by *argU*. Since the *pin*-3 mutation maps at about 12.2 min on the *E. coli* chromosome (21), we attempted to supply the wild-type allele from plasmids that carry parts of this region. The plasmid pBR1824, a derivative of pBR322 that carries only the *argU* gene (19, 42), completely restored the ability of P4 to make plaques on *E. coli pin*-3 (P2) strain C-5024. The efficiency of plating of P4 *vir*1 on C-5024 was less than 10^{-4} relative to the *pin*⁺ strain C-117. When pBR1824 was introduced into C-5024, the efficiency of plating of P4 *vir*1 increased to 1 relative to C-117; no increase was observed for pBR322 transformants of C-5024.

Thus, *pin*-3 is likely to be a defective allele of *argU*, which codes for an arginine tRNA that reads the rare codon AGA and possibly AGG (19). These codons represent only 2% of arginine codons in *E. coli* proteins (2). When more than one codon is used to code for a particular amino acid residue,

there is evidence that codon frequency is correlated with the abundance of charged tRNAs (23, 25, 26). The *pin-3* mutation may reduce or eliminate the pool of tRNA molecules available for translation of AGA codons. Since there are 10 AGA codons in the β gene, β may be expressed poorly in *E. coli pin-3*; this would allow the *gop* gene product to kill the cells and abort P4 multiplication. The *pin-3* mutation was first defined by its ability to prevent interference with λ phage growth by the P2 phage *old* gene (21). The P2 *old* gene is rich in AGA and AGG codons and is poorly expressed in *E. coli pin-3* (E. Haggård-Ljungquist, V. Barreiro, R. Calendar, D. M. Kurnit, and H. Cheng, Gene, in press).

ACKNOWLEDGMENTS

We thank Stefania Crotta, Nora Linderroth, Franco Tronconi, Sandro Zangrossi, and Rainer Ziermann, who performed some of the experiments; L. S. Pierson, who provided the first cloned P4 fragments for sequencing; James R. Walker for the *argU* plasmid; Michael Yarus for *gluT* plasmids; and Nora Linderroth for a careful reading of the manuscript.

This work was supported by grant N 86.00.327.04 from Consiglio Nazionale delle Ricerche, by a grant from Ministero della Pubblica Istruzione, Roma, and by Public Health Service grant AI08722 from the National Institutes of Health to R.C.

LITERATURE CITED

- Alano, P., G. Dehò, G. Sironi, and S. Zangrossi. 1986. Regulation of the plasmid state of the genetic element P4. *Mol. Gen. Genet.* 203:445-450.
- Aota, S.-I., T. Gojbori, F. Ishibashi, T. Maruyama, and T. Ikemura. 1988. Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res.* 16:r315-r402.
- Barrett, K. J., M. L. Marsh, and R. Calendar. 1976. Interactions between a satellite bacteriophage and its helper. *J. Mol. Biol.* 106:683-707.
- Berk, A. J., and P. A. Sharp. 1978. Spliced early mRNAs of simian virus 40. *Proc. Natl. Acad. Sci. USA* 7:1274-1278.
- Bertani, L. E. 1964. Lysogenic conversion by bacteriophage P2 resulting in an increased sensitivity of *Escherichia coli* to 5-fluorodeoxyuridine. *Biochim. Biophys. Acta* 87:631-640.
- Bertani, L. E., and E. W. Six. 1988. The P2-like phages and their parasite, P4. p. 73-143. In R. Calendar (ed.), *The bacteriophages*, vol. 2. Plenum Publishing Corp., New York.
- Biggen, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and ³⁵S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* 80:3963-3965.
- Bolivar, F., R. L. Rodriguez, P. J. Green, M. C. Betlach, H. L. Heyneker, H. W. Boyer, J. H. Croso, and S. Falkow. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2:95-113.
- Brennan, R. G., and B. W. Matthews. 1989. The helix-turn-helix DNA binding motif. *J. Biol. Chem.* 264:1903-1906.
- Calendar, R., E. Ljungquist, G. Dehò, D. C. Usher, R. Goldstein, P. Youderian, G. Sironi, and E. W. Six. 1981. Lysogenization by satellite phage P4. *Virology* 113:20-38.
- Dale, E. C., G. E. Christie, and R. Calendar. 1986. Organization and expression of the satellite bacteriophage P4 late gene cluster. *J. Mol. Biol.* 192:793-803.
- Dehò, G. 1983. Circular genetic map of satellite bacteriophage P4. *Virology* 126:267-278.
- Dehò, G., D. Ghisotti, P. Alano, S. Zangrossi, M. G. Borrello, and G. Sironi. 1984. Plasmid mode of propagation of the genetic element P4. *J. Mol. Biol.* 178:191-207.
- Dehò, G., S. Zangrossi, D. Ghisotti, and G. Sironi. 1988. Alternative promoters in the development of bacteriophage plasmid P4. *J. Virol.* 62:1697-1704.
- Dodd, I. B., and J. B. Egan. 1987. Systematic method for the detection of potential λ Cro-like DNA-binding regions in proteins. *J. Mol. Biol.* 194:557-564.
- Evans, R. M., and S. M. Hollenberg. 1988. Zinc fingers: guilt by association. *Cell* 52:1-3.
- Felton, J. 1983. M13 host strain JM103 contains two restriction systems. *Bio/Techniques* 1:42-43.
- Flensburg, J., and R. Calendar. 1987. Bacteriophage P4 DNA replication. Nucleotide sequence of the P4 primase gene and the *cis* replication region. *J. Mol. Biol.* 195:439-445.
- Garcia, G. M., P. K. Mar, D. A. Mullin, J. R. Walker, and N. E. Prather. 1986. The *E. coli dnaY* gene encodes an arginine transfer tRNA. *Cell* 45:453-459.
- Garen, A. 1968. Sense and nonsense in the genetic code. *Science* 160:149-159.
- Ghisotti, D., S. Zangrossi, and G. Sironi. 1983. An *Escherichia coli* gene required for bacteriophage P2- λ interference. *J. Virol.* 48:616-626.
- Gibbs, W., R. N. Goldstein, R. Wiener, B. Lindqvist, and R. Calendar. 1973. Satellite bacteriophage P4: characterization of mutants in two essential genes. *Virology* 53:24-39.
- Gouy, M., and C. Gautier. 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10:7055-7074.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28:351-359.
- Ikemura, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* 146:1-21.
- Ikemura, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* 151:389-409.
- Kahn, M., D. Ow, B. Sauer, A. Rabinowitz, and R. Calendar. 1980. Genetic analysis of bacteriophage P4 using P4-plasmid ColE1 hybrids. *Mol. Gen. Genet.* 177:399-412.
- Kraft, R., J. Tardiff, K. S. Krauter, and L. A. Leinwand. 1988. Using mini-prep plasmid DNA for sequencing double stranded templates with Sequenase. *BioTechniques* 6:544-549.
- Krevolin, M. D., and R. Calendar. 1985. The replication of bacteriophage P4 DNA *in vitro*. Partial purification of the P4 α gene product. *J. Mol. Biol.* 182:509-517.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* 227:680-685.
- Landschulz, W. H., P. F. Johnson, and S. L. McKnight. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 240:1759-1764.
- Lin, C.-S. 1984. Nucleotide sequence of the essential region of bacteriophage P4. *Nucleic Acids Res.* 12:8667-8684.
- Lindahl, G., G. Sironi, H. Bialy, and R. Calendar. 1970. Bacteriophage lambda; abortive infection of bacteria lysogenic for phage P2. *Proc. Natl. Acad. Sci. USA* 66:587-594.
- Lindqvist, B. H. 1981. Recombination between satellite phage P4 and its helper P2. I. *In vivo* and *in vitro* construction of P4::P2 hybrid satellite phage. *Gene* 14:231-241.
- Lindqvist, B. H., and E. W. Six. 1971. Replication of bacteriophage P4 DNA in a nonlysogenic host. *Virology* 43:1-7.
- Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* 227:1435-1441.
- MacHattie, L. A., and J. B. Jackowski. 1977. Physical structure and deletion effect of the chloramphenicol resistance element Tn9 in phage lambda, p. 219-228. In A. I. Bukhari, J. A. Shapiro, and S. L. Adhya (ed.), *DNA insertion elements, plasmids, and episomes*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Messing, J. 1979. A multipurpose cloning system based on the single-stranded DNA bacteriophage M13. *Recomb. DNA Technol. Bull.* 2:43-48.
- Messing, J., R. Crea, and P. H. Seeburg. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res.* 9:309-321.
- Miller, J., A. D. McLachlan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA

- from *Xenopus* oocytes. EMBO J. 4:1609–1614.
42. Mullin, D. A., G. M. Garcia, and J. R. Walker. 1984. An *E. coli* DNA fragment 118 base pairs in length provides *dnaY*⁺ complementing activity. Cell 37:669–674.
 43. Norrander, J., T. Kempe, and J. Messing. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26:101–106.
 44. Pabo, C. O., and R. T. Sauer. 1984. Protein-DNA recognition. Annu. Rev. Biochem. 53:293–321.
 45. Pierson, L. S., III, and M. L. Kahn. 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*. DNA sequences of the phage and host regions involved in site-specific recombination. J. Mol. Biol. 196:487–496.
 46. Raftery, L. A., and M. Yarus. 1985. Site-specific mutagenesis of *Escherichia coli* *gluT* yields a weak, glutamic acid-inserting ochre suppressor. J. Mol. Biol. 184:343–345.
 47. Raftery, L. A., and M. Yarus. 1987. Systematic alterations in the anticodon arm make tRNA^{Glu}-Su^{oc} a more efficient suppressor. EMBO J. 6:1499–1506.
 48. Raimondi, A., R. Donghi, A. Montaguti, A. Pessina, and G. Dehò. 1985. Analysis of spontaneous deletion mutants of satellite bacteriophage P4. J. Virol. 54:233–235.
 49. Reeve, J. N. 1977. Bacteriophage infection of minicells. A general method for identification of “in vivo” bacteriophage directed polypeptide biosynthesis. Mol. Gen. Genet. 158:73–79.
 50. Reeve, J. 1979. Use of minicells for bacteriophage-directed polypeptide synthesis. Methods Enzymol. 68:493–503.
 51. Rusche, J. R., and P. Howard-Flanders. 1985. Hexamine cobalt chloride promotes intermolecular ligation of blunt end DNA fragments by T4 DNA ligase. Nucleic Acids Res. 13:1997–2008.
 52. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.
 53. Sasaki, I., and G. Bertani. 1965. Growth abnormalities in Hfr derivatives of *Escherichia coli* strain C. J. Gen. Microbiol. 40:365–376.
 54. Sauer, B., D. Ow, L. Ling, and R. Calendar. 1981. Mutants of satellite bacteriophage P4 that are defective in the suppression of transcriptional polarity. J. Mol. Biol. 145:29–46.
 55. Shore, D., G. Dehò, J. Tsipis, and R. Goldstein. 1978. Determination of capsid size by satellite bacteriophage P4. Proc. Natl. Acad. Sci. USA 75:400–404.
 56. Six, E. W. 1975. The helper dependence of satellite bacteriophage P4: which gene functions of bacteriophage P2 are needed by P4? Virology 67:249–263.
 57. Six, E. W., and C. A. C. Klug. 1973. Bacteriophage P4: a satellite virus depending on a helper such as prophage P2. Virology 51:327–344.
 58. Souza, L., J. Geisselsoder, A. Hopkins, and R. Calendar. 1978. Physical mapping of the satellite phage P4 genome. Virology 85:335–342.
 59. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA 84:4767–4771.
 60. 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.
 61. Wiman, M., G. Bertani, B. Kelly, and I. Sasaki. 1970. Genetic map of *Escherichia coli* C. Mol. Gen. Genet. 107:1–31.
 62. Yanisch-Perron, C., J. Vieira, and J. Messing. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33:103–119.
 63. Younghusband, H. B., J. B. Egan, and R. B. Inman. 1975. Characterization of the DNA from bacteriophage P2-186 hybrids and physical mapping of the 186 chromosome. Mol. Gen. Genet. 140:101–110.