JOURNAL OF BACTERIOLOGY, Nov. 2004, p. 7032–7068 Vol. 186, No. 21
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.21.7032–7068.2004
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

Genome of Bacteriophage P1†

Małgorzata B. Łobocka,[1]* Debra J. Rose,[2] Guy Plunkett, III,[2] Marek Rusin,[3] Arkadiusz Samojedny,[3] Hansjörg Lehnherr,[4] Michael B. Yarmolinsky,[5] and Frederick R. Blattner[2]

Department of Microbial Biochemistry, Institute of Biochemistry and Biophysics of the Polish Academy of Sciences, Warsaw,[1] and Department of Tumor Biology, Centre of Oncology, M. Skłodowska-Curie Memorial Institute, Gliwice,[3] Poland; Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin[2]; Department of Genetics and Biochemistry, Institute of Microbiology, Ernst-Moritz-Arndt-University Greifswald, Greifswald, Germany[4]; and Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland[5]

Received 11 March 2004/Accepted 9 July 2004

P1 is a bacteriophage of *Escherichia coli* and other enteric bacteria. It lysogenizes its hosts as a circular, low-copy-number plasmid. We have determined the complete nucleotide sequences of two strains of a P1 thermoinducible mutant, P1 *c*1-100. The P1 genome (93,601 bp) contains at least 117 genes, of which almost two-thirds had not been sequenced previously and 49 have no homologs in other organisms. Protein-coding genes occupy 92% of the genome and are organized in 45 operons, of which four are decisive for the choice between lysis and lysogeny. Four others ensure plasmid maintenance. The majority of the remaining 37 operons are involved in lytic development. Seventeen operons are transcribed from σ^{70} promoters directly controlled by the master phage repressor C1. Late operons are transcribed from promoters recognized by the *E. coli* RNA polymerase holoenzyme in the presence of the Lpa protein, the product of a C1-controlled P1 gene. Three species of P1-encoded tRNAs provide differential controls of translation, and a P1-encoded DNA methyltransferase with putative bifunctionality influences transcription, replication, and DNA packaging. The genome is particularly rich in Chi recombinogenic sites. The base content and distribution in P1 DNA indicate that replication of P1 from its plasmid origin had more impact on the base compositional asymmetries of the P1 genome than replication from the lytic origin of replication.

Three bacteriophages of *Escherichia coli* with different life styles have been particularly instrumental in the development of the concepts and tools of molecular biology. They are the virulent phage T4 and the two temperate phages, λ and P1. The two temperate phages were first described in close succession in the middle of the last century, namely in 1951 (phage P1 [31]) and 1953 (λ [185] and informally even earlier, in the first issue of Evelyn Witkin's *Microbial Genetics Bulletin*, 1950 [184a]). Whereas λ prophage leads an essentially passive existence within the chromosome of its host, P1 prophage exists as an autonomous plasmid that is maintained at low copy number. The complete genome sequences of λ, T4, and *Escherichia coli* have been reported. We present here the complete nucleotide sequence of P1 and summarize its salient features.

P1 infects and lysogenizes *E. coli* and several other enteric bacteria. Its virion consists of an icosahedral head attached at one vertex to a tail that bears six kinked tail fibers (see cover figure). As in T4, the tail consists of a tail tube and a contractile sheath. A variable part of the tail fibers determines the specificity of P1 adsorption on different hosts. The variable part is encoded by an invertible segment of P1 DNA, similar to that of phage Mu (reviewed in reference 268). Infective particles of P1 contain cyclically permuted, linear, double-stranded molecules with a terminal redundancy of 10 to15 kb of DNA (293). Prior to this work, about 70 genes had been identified in P1 by genetic and physiological studies (346).

After injection into a host cell, viral DNA circularizes by recombination between redundant sequences to enter a lytic or lysogenic path. The choice between the two paths is dictated by the interplay of a number of environmental factors with the complex immunity circuitry that controls synthesis and activity of a master repressor, C1 (reviewed in reference 126). As a prophage, P1 is a stable plasmid maintained at about one copy per bacterial chromosome. Several P1 genes scattered throughout the genome are expressed in the lysogenic state. Those of primary importance are involved in plasmid maintenance functions and in inhibition of lytic development. P1 prophage replicates as a circular plasmid from an origin (*oriR*) resembling that of several other plasmids and, to a lesser extent, *oriC* of *E. coli*.

P1 genes expressed in the lytic pathway are those involved in the timing of phage development, in replication (from a "lytic" origin different from *oriR*), in the formation of phage particles, including headful packaging, and finally in cell lysis to release the phage progeny. Genes of the lytic pathway have been divided into early and late. Transcription of the late genes requires, in addition to bacterial RNA polymerase, a P1-encoded activator protein, Lpa (188), and an *E. coli* RNA polymerase-associated stringent starvation protein, SspA (111).

P1, like λ, made its mark early in molecular biology. The significant capacity of P1 for mediating generalized transduction (196) led promptly to P1 becoming a workhorse of genetic exchange among strains of *E. coli*, a role it is still playing today. Moreover, because P1 can package slightly more than twice as much DNA as can λ, and packaging can be efficiently carried

* Corresponding author. Mailing address: Department of Microbial Biochemistry, Institute of Biochemistry and Biophysics of the Polish Academy of Sciences, Ul. Pawinskiego 5A, 02-106 Warsaw, Poland. Phone: 011 4822 8237188. Fax: 011 4822 6584636. E-mail: lobocka@ibb.waw.pl.

† Supplemental material for this article may be found at http://jb.asm.org/

out in vitro, P1-based vectors are now in common use for cloning and in vitro packaging of eukaryotic DNA (246, 293). P1 also made its mark early because of the restriction-modification genes that it carries, the analysis of which in the 1960s heralded the age of genetic engineering (16). The recognition that P1 is maintained as a plasmid prophage led to the identification of its plasmid maintenance functions. A site-specific recombinase, Cre, which appears promptly after infection, is involved in phage DNA cyclization (137), assists plasmid maintenance by resolving plasmid multimers (21), and perhaps, by a separate mechanism, may stabilize plasmid copy number (10). Ever since the introduction of the enzyme and its recombination site, *lox*, into *Saccharomyces cerevisiae* (271), the Cre-*lox* system has been a major tool of genetic engineering in eukaryotic cells.

The active partitioning of P1 plasmids to daughter cells requires two P1-encoded proteins and a *cis*-acting site (the centromere) (20). A surprising homology between the structural genes of plasmid partitioning and the *soj* and *spoOJ* sporulation genes of *Bacillus subtilis* provided the motivation for examining the effect of *soj* and *spoOJ* mutations on the production of anucleate progeny. Consequently, a connection was established between plasmid and nucleoid partitioning (157, 284). The finding that a gene of P1 encodes a toxin that is activated on P1 loss (192), among other findings of such "plasmid addiction" genes, helped to establish the concept that programmed cell death has its place in the life of bacterial populations, as in metazoan development (348). Finally, the analysis of P1 immunity maintenance and establishment (reviewed in reference 126) enlarges our view of controls on phage development. The master repressor, C1, unlike that of lambda, is monomeric, and the operators to which it binds are many and dispersed. It is modulated by at least two antirepressors (Ant1/2 and Coi) and a corepressor protein (Lxc), the latter capable of inducing certain C1-bound operators to loop. A *trans*-acting RNA (the *c*4 product) regulates Ant1/2 synthesis in a novel way, being excised from a transcript on which lies its site of action. Regrettably, the study of P1 immunity halted prematurely with the death of its prime mover, Heinz Schuster.

The most recent effort to compile a P1 map (made when no more than a total of 40% of the genome had been sequenced) was reported solely as a computer file (346). A recent short review of P1 biology has been prepared (186). Reviews of the older P1 literature may be found in references 29, 301, 345, and 347 and successive editions of *Genetic Maps*, one of which (344) was reprinted in reference 225. A thoughtful perspective on early studies of lysogeny by one of the pioneers in this field and the discoverer of P1 has recently appeared (30).

The overview of P1 genome organization and developmental regulation that we present here emerged from results of previous studies combined with the analysis of the entire genomic sequences of two strains of P1 described below.

MATERIALS AND METHODS

Sources of DNA and sequencing strategy. DNA sequences described here were derived from two sources. One was DNA of P1 phage induced from N1706, an *E. coli* K-12 P1 *c*1-100 Tn*9* lysogen preserved frozen since the early 1970s (kindly provided by J. L. Rosner). The other was DNA of P1 prophage carried by an MG1655-derived strain that was used in the *E. coli* genome sequencing project (35). The prophage had been acquired in the course of a P1-mediated transduction with the aid of P1*c*1-100 Tn*9* r$^-$ m$^-$ *dam* (now *dmt*) ΔMB *rev*-6, a high-frequency transducing derivative of the P1 *c*1-100 Tn*9* lysogen (303).

Prophage induction in N1706 and isolation of phage particles were performed as described previously (112). DNA was isolated from phage particles by using the QIAGEN Lambda Midi kit (QIAGEN, Inc.) according to the supplier's protocol. DNA of the prophage was isolated from a gel as described in reference 35.

The shotgun sequencing strategy was used for acquiring most of the sequencing data. Fragments of P1 phage or prophage DNA obtained by sonication were cloned in the pUC18 vector (266) or in the M13 Janus vector (45) and served as templates for sequencing. The ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit with AmpliTaq DNA polymerase FS (Perkin-Elmer), was used to perform sequencing reactions. Separation of products of sequencing reactions and reading of results was done with ABI 377 automated sequencers (Perkin-Elmer). After collection of initial sets of sequencing data at a four- to fivefold redundancy from readings in one direction, selected templates were resequenced in the opposite direction. The remaining gaps in the sequences were filled in by primer walking using oligonucleotides complementary to the ends of the sequenced regions and cloned P1 fragments or the entire DNA of the respective P1 strain as a template. Sequences were assembled using the program Seqman II (DNAStar Inc.).

Sequence analysis and annotation. The majority of analyses were carried out using programs from either the GCG package version 9.0 (Genetics Computer Group, 1996) or the DNAStar package (DNAStar Inc.).

Open reading frame (ORF) identification was performed using programs based on a Markov model: the Internet version of GeneMark (http://genemark.biology.gatech.edu/GeneMark) (36, 208) and TIGR Glimmer 2.1 (75, 264). GeneMark was trained on templates for genes of *E. coli*, and its temperate bacteriophages and transposons, and bacteriophage T4. The model for Glimmer 2.1 was prepared with the use of Build-icm trained on P1 ORFs longer than 600 nucleotides. Additionally, the entire sequence was divided into 400-bp overlapping fragments and the predicted products of their translation were searched against the database for possible similarities with known protein sequences. Identification of a few genes, whose predicted codon usage pattern did not match that of any model organism, was based on homologies of their putative products to known proteins and on the presence in their upstream regions of sequences resembling promoters or likely to encode ribosome binding sites.

The assignment of previously identified genetic loci to newly sequenced ORFs was assisted by alignments to existing physical and genetic maps (230, 252, 280, 281, 294, 323, 346). Predicted sizes of putative P1 structural proteins were compared with sizes of head and tail components of P1 determined by polyacrylamide gel electrophoresis (PAGE) (325) to verify the identity of certain structural genes.

Searching for similarities between P1 proteins and known proteins in databases was performed using the Internet versions of programs PSI-BLAST and PHI-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) (12, 354). Macaw (278) served to create multiple protein alignments. The putative helix-turn-helix motifs in protein sequences were identified by the HTH program at its website (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html) (78). Putative signal peptides in proteins were predicted by using the SignalP program (235) at its website (http://www.cbs.dtu.dk/services/SignalP/#submission). Putative transmembrane helices in proteins were predicted by using the program TMHMM (175) at its website (http://www.cbs.dtu.dk/services/TMHMM).

Identification of putative σ^{70} promoters was carried out using Targsearch (118). Putative Rho-independent transcriptional terminators were identified with the GCG program Terminator (40) and TIGR program TransTerm (85). To avoid false positives, only terminators that met the following additional criteria (69, 198) were taken into consideration: a 4- to 18-bp stem and a 3- to 10 nucleotide (nt) loop of the terminator hairpin; a thymidine-rich region downstream of the terminator hairpin and separated by less than 3 nt; more than three GC/CG or GT/TG bp in the hairpin stem; at least three T residues, no more than one G, and no 5′-TVVTT stretches (V is A, C, or G) in the 5-nt-long proximal part of the thymidine-rich region; absence of four purine or cytosine residues in the 4-nt-long distal part of the thymidine-rich region and at least four T residues together in the proximal and distal parts of the thymidine-rich region. Putative integration host factor (IHF)-binding sites were identified with MacTargsearch at SEQSCAN (http://www.bmb.psu.edu/seqscan/seqform1.htm).

Oligonucleotide frequencies were determined with the program OCTAMER (L. Łobocka, unpublished). The DNA was scanned for putative tRNA genes by DNA homology searches and by using tRNAscan-SE (81, 204) at its website (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/).

Nucleotide sequence accession numbers and strain availability. The nucleotide sequences described here have been deposited in GenBank under accession

numbers AF234172 (phage P1 *mod749*::IS*5* *c*1-100) and AF234173 (prophage P1 *mod1902*::IS*5* *c*1-100 *rev-6* *dmt*ΔMB). A lysogen (N1706) from which the first of these phages was induced has been deposited with the American Type Culture Collection (Manassas, Va.) as ATCC BAA-1001.

RESULTS AND DISCUSSION

Comparison of the sequenced P1 isolates. Current P1 phages in circulation are derived from P1*kc*, a mutant selected in two steps by Lennox (196) as able to plate on *E. coli* K-12 with plaques of increased clarity. The induction of lysogens was facilitated by the introduction of mutations that render P1 thermoinducible (279), the most widely used being the *c*1-100 mutation, originally designated *clr100* (261). It is present in both of the P1 isolates that we used for sequencing. For simplicity in selecting P1 lysogens as drug-resistant bacteria, Rosner (261) made use of P1 CM (171), a strain that harbors the unstable Tn*9* transposon (encoding chloramphenicol acetyltransferase) acquired from an R factor. However, P1 *c*1-100 Tn*9* has been widely used and hence was chosen for sequencing. The phage induced from the supposed P1 *c*1-100 Tn*9* lysogen was found to have lost Tn*9* and to have acquired IS*5* inserted in the *mod* gene, which renders it r^- m^-. The sequence of this strain, P1 *c*1-100 *mod749*::IS*5* (94,800 bp), was determined with an average 7.2-fold redundancy. The assembly was verified by comparison of the restriction map obtained by digestion with EcoRI, HindIII, HincII, BamHI, SacI, and PstI with that deduced from the entire P1 sequence (without the IS*5*) (346).

The other strain of P1 used here, P1 *c*1-100 *mod1902*::IS*5* *dmt*ΔMB *rev*-6 (94,481 bp), is a prophage carried by an MG1655-derived strain of *E. coli* K-12 that served in sequencing the *E. coli* genome (35). Its nucleotide sequence was determined with an average 7.8-fold redundancy. The prophage appears to have been acquired in the course of a P1-mediated transduction. In addition to a 319-bp deletion in *dmt* that impairs the function of the viral DNA methyltransferase gene, it carries a mutation, *rev-6*, that improves growth of the mutant in a *dam* mutant host and augments transduction (303). In the course of its history, the phage lost Tn*9*, acquired IS*5*, and accumulated other mutations such that, relative to the first strain, there are 13 additional mutations (four of them silent and one in a noncoding region). Five of those mutations are scattered in the variable part of the *S* gene (*S*v), which is essential for P1 adsorption to *E. coli*. Which alleles are parental remains uncertain. The *rev-6* mutation has not been correlated with any of the sequence differences between the two strains.

In the two sequenced strains, the IS*5* insertions occurred at different TTAG sequences, known to belong to the most commonly used targets of IS*5* (C/TTAA/G; reviewed in reference 212). In each case the IS*5* insert, accompanied by a 4-bp duplication of its target, appears responsible for the r^- m^- phenotype since, in other respects, the *res mod* sequences are identical to those of r^+ m^+ P1 (145).

Genomewide features. The genome of P1, without IS*5* and its duplicated target, consists of 93,601 bp of double-stranded DNA, which in the prophage can be represented as a circle or line with the center of the site-specific recombination site *lox* assigned as the zero point such that the first nucleotide to its right on the strand written 5′ to 3′ in the direction of *cre* is assigned position 1 and numbering proceeds rightwards (346) (Fig. 1). The genome contains one insertion sequence, IS*1*, as an integral part of P1 (151). In both sequenced strains, IS*1* has a base substitution mutation (IS*1* G757T) in its right inverted repeat (IRR).

(i) Base content and distribution. The GC content of P1 DNA is 47.3%, slightly less than that of its *E. coli* host (50.8%) (35). The distribution of purines and pyrimidines between the two complementary strands is similar, with the upper strand (+; coding strand of genes transcribed clockwise) containing 49.5% (26.0% adenosine and 23.5% guanosine), and the lower strand (−) containing 50.5% (26.7% adenosine and 23.8% guanosine) purine nucleotides.

The location of extensive AT-rich regions along the genome map confirms the denaturation mapping evidence (222, 347) for the relatively recent acquisition by P1 of its restriction-modification genes (*res*, *mod*) and suggests that genes of the *sim* and *rlf* operons, as well as *hot* and *isaB*, were also incorporated in the genome of P1 late in evolution. Sharp borders of long AT-rich regions within certain genes (e.g., the 3′-moiety of *parB*) may be indicative of mosaic structure.

Unique restriction sites in P1 DNA are located preferentially in AT-rich regions or within the IS*1* sequence (Fig. 2), suggesting their recent acquisition by P1 and selection against these sites in a P1 ancestor. A few 6-bp palindromes, the targets of known restriction enzymes, ApaI (GGGCCC), NarI (GGCGCC), NaeI (GCCGGC), SacI (GAGCTC), SalI (GTCGAC), and AvrII (CCTAGG), are absent from the P1 genome. Short DNA palindromes rare in or absent from P1 presumably identify sequences that could confer vulnerability in hosts that P1 otherwise finds particularly congenial.

According to the pattern of C-G skew (the relative excess of cytosines versus guanosines in the top strand), the genome of P1 can be divided into two polarized arms with the average C-G skew below the mean predominating in one part and above the mean in the other part (Fig. 2). The region where the average sign of C-G skew below the mean predominates coincides with the region where transcription in the clockwise direction predominates, and vice versa. The C-G skew switches polarity in two regions: around kilobase 64 of the genome, close to the origin of plasmid replication (*oriR*), and around kilobase 3 of the genome. This pattern is consistent with the skew observed for bacterial and double-stranded DNA viral genomes whose replication proceeds bidirectionally from an origin (reviewed in references 101 and 216). Typically, the regions of polarity switch in the C-G skew are close to an origin and to the terminus of replication. P1 possesses two separate origins, *oriR*, and the origin of lytic replication, *oriL*, which are about 9 kb apart. Replication from both origins can be bidirectional (61, 241). As seen in Fig. 2, P1 prophage replication appears to have had more impact on the base compositional asymmetries and transcriptional organization of the genome than the theta (θ) or rolling-circle (σ) replication that occurs during lytic development. Although *oriL* is further away than *oriR* from the region of the C-G skew polarity switch at 64 kb, it is located diametrically opposite to the region of the second switch in the C-G skew polarity, which may be indicative of the terminus of P1 replication. We did not find in this region (or anywhere else in the P1 DNA) the TGTTGTAACTA se-

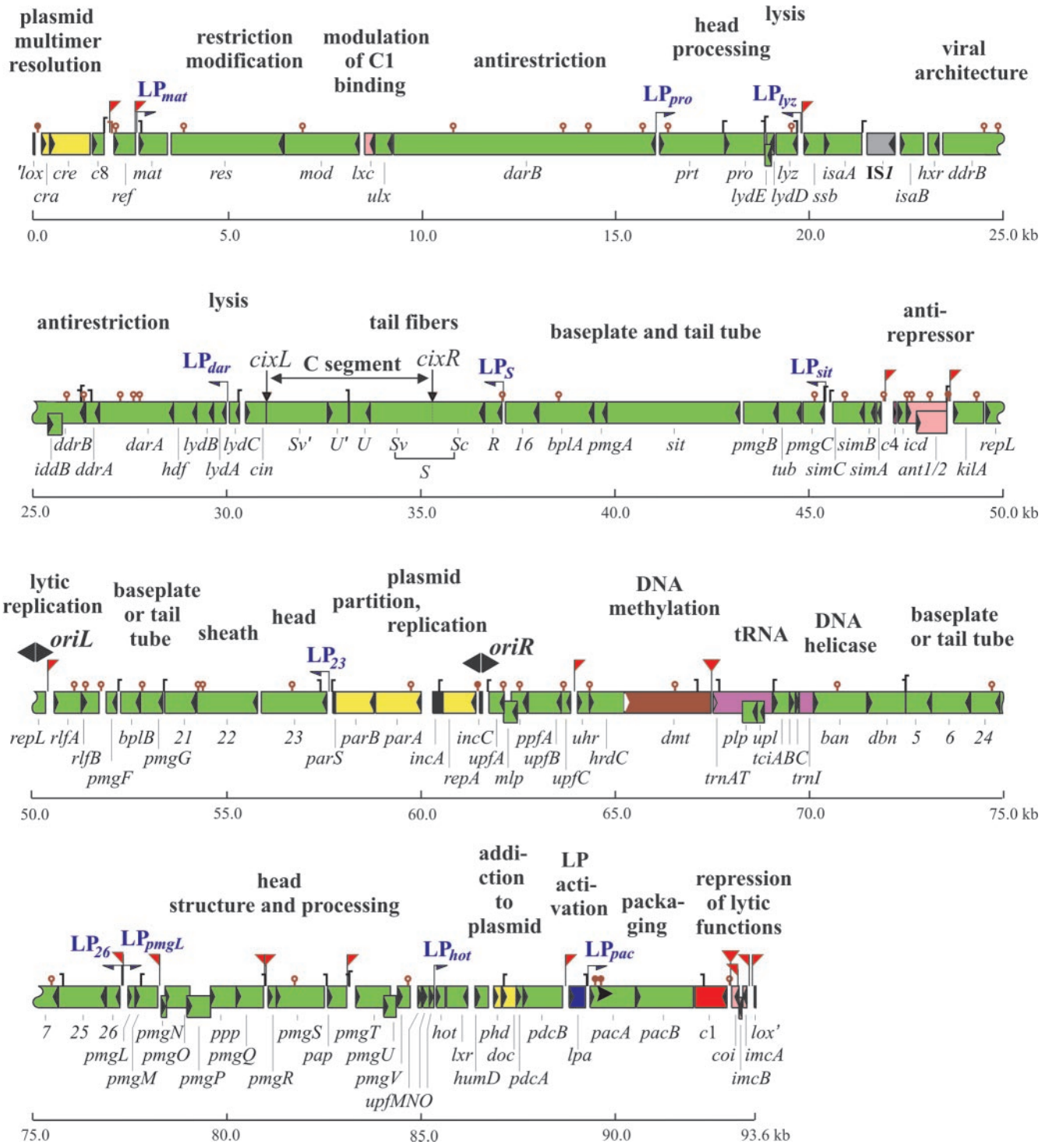

FIG. 1. Genetic and physical organization of the P1 genome. Boxes with internal triangles show positions and orientations of genes, color-coded by function: yellow, plasmid maintenance; red, repression of early functions; pink, immunity control, not *c*1 itself; magenta, source of tRNAs; brown, DNA methylation; deep blue, transcriptional activation of late genes; grey, defective IS*1*; green, all other. Black boxes are intergenic regions of defined function: recombination sites, iterons to which RepA binds, plasmid centromere, and origin of DNA packaging, the direction of packaging being indicated by an arrowhead at the *pac* site. Bidirectional replication determined at the phage (lytic) origin, *oriL*, and at the plasmid origin, *oriR*, are indicated by black arrowheads above the genome map. C1 operator sites are marked with red flags pointing to the left or right (see also Table 6). Thin lines with terminal deep blue half arrowheads indicate the start sites and directions of the transcripts from particular late promoters. GATC sequences that overlap transcriptional promoters and clustered 5′-GATC sequences (two or more sites with pairwise separation of not more than 50 bp), substrates for Dmt or Dam methylation, are marked above the gene map by brown lollipops that are filled in the case of sites shown to alter function upon methylation. Hooks indicate Rho-independent transcriptional terminators. They face the starts of transcripts that they terminate. The map refers to the genome of P1 *c*1-100 *mod749*::IS*5* without its nonintegral part, IS*5*.

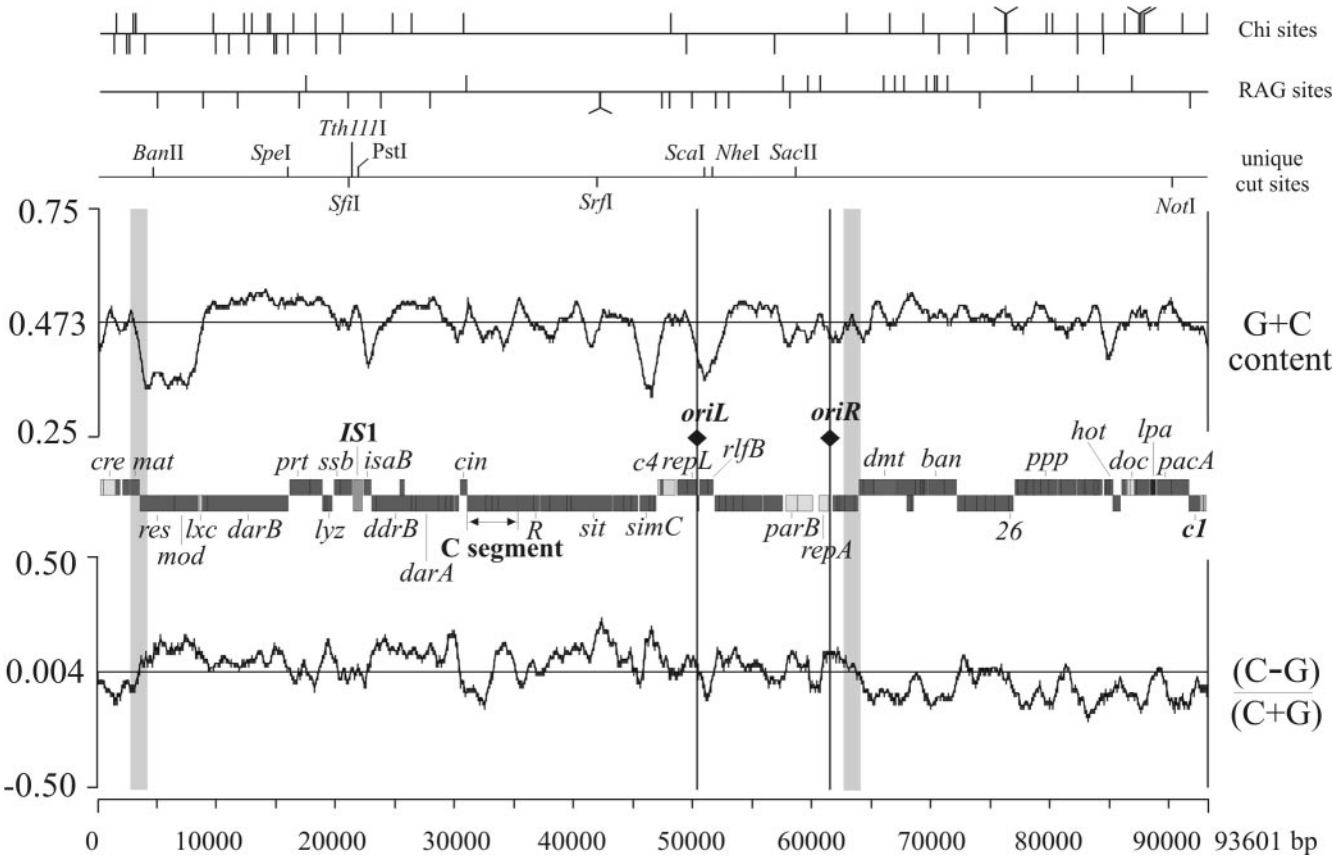

FIG. 2. Compositional organization of the P1 genome. The G+C content and (C−G)/(C+G) ratio (C-G skew) for one strand of P1 DNA are plotted as deviations from the mean. The plots were made with the GeneQuest program of DNAStar with a window size of 1,000 bp and a shift increment of 1 bp. For orientation, a simplified genome map is shown centrally. Upper-row genes are transcribed clockwise (left to right), and lower-row genes are transcribed counterclockwise (right to left). Two vertical black lines through the plots show the locations of the lytic (*oriL*) and plasmid (*oriR*) origins of replication. Two vertical grey bars show regions of polarity switch in the plot of the C-G skew. Unique recognition sites for restriction enzymes are indicated directly above the plot of G+C content; those below the line are for 8-bp cutters. See the text for absent 6-bp restriction sites. Chi sites and RAG sites (64) in the upper and lower strands are indicated by hash marks above and below the horizontal lines at the top of the figure. Where two hash marks are so close as to appear superimposed, a line at an angle to the vertical is added. The map refers to the genome of P1 *c*1-100 *mod749*::IS*5* without its nonintegral parts, IS*5*, and the associated 4-bp duplication.

quence that in *E. coli*, *Salmonella enterica* serovar Typhimurium, and several plasmids constitutes a conserved core of *Ter* sites (136) at which replication terminates. However, this region is within 3 kb of the P1 site-specific recombination site *lox*, which is involved in resolution of plasmid multimers into monomers, playing a role analogous to that of the *dif* site in the *E. coli* chromosome. The *lox* site of P1 was shown to suppress the phenotype of an *E. coli dif* deletion strain in the presence of the *lox*-specific recombinase, Cre, when *lox* was inserted in the region of the *E. coli* terminus of replication (197). This suppression requires the DNA translocase activity of the essential host protein FtsK (47), as does dimer resolution at *dif* (292).

The major family of strongly skewed sequences in *E. coli* with the motif 5′-RGNAGGGS (R = purine, N = any base, S = G or C) (265) represents putative sites of binding to FtsK, a protein involved in positioning of bacterial DNA within the cell (64). The strongly biased distribution of these so-called RAG sites in P1 DNA (Fig. 2) possibly argues for a comparable role for FtsK in the life of P1 prophage, although a connection between RAG sites and FtsK is still uncertain.

The most abundant octamer in the genome of P1 corresponds to the recombinational hot spot Chi (5′-**GCTGGTGG**; 31 and 19 sites in the upper and lower strands [31/19]). Three other octamers in descending order of abundance (5′-T**GCTG GTG**, 18/14; 5′-**CTGGTGG**A, 14/8; and 5′-**CTGGTGG**C, 14/6) contain heptamers (boldface) identical to a Chi heptamer. The canonical Chi sites are two and one-half times more frequent in the P1 genome (one site per 1,872 bp) than in the genome of its *E. coli* host (one site per 4,598 bp) (35, 291), where they are the third most abundant octamers and sevenfold more abundant than a random octamer would be. P1 could benefit in several ways from the accumulation of Chi sites in its genome. P1 DNA encapsidated in a virion is a terminally redundant linear molecule, which upon entering cells at infection, can serve as a substrate for loading of RecBCD nuclease at its ends and for RecBCD-mediated degradation. Each encounter of RecBCD with Chi modifies the enzyme from its destructive 3′-to-5′ nucleolytic form to its recombination-promoting form (reviewed in reference 173) and thus should both protect incoming P1 DNA from further degradation and facilitate, by homologous recombination, its cyclization. The involvement of

TABLE 1. Binding sites in P1 DNA of P1 proteins other than C1 and Lpa

Name[a]	Coordinates[b]	Strand	Sequence[c]	Function (protein bound)	Reference(s)
lox	93587–93601 and 1–15	+	AACTTCGTATAGCATACATTATACGAAGTT	Site of Cre-mediated recombination (Cre)	138
*cix*Pp	30379–30419	−	TTAACAAAAACCAGTATGTGTGGAAA	*cin* autoregulation (Cin)	150
*cix*L	31016–31035	+	CAATCCTAAACCTTGGTTTAAGAGAA	Site of Cin-mediated C-segment inversion	150
*cix*R	31016–31035	−	TTATCCAATACCTTGGTTTAAGAGAA	Site of Cin-mediated C-segment inversion	150
parS				Plasmid centromere	4, 92
boxA4	57713–57719	+	GTGAAAT	(ParB)	
boxB2	57722–57727	−	TCGCCA	(ParB)	
boxA3	57722–57733	−	GTGAAAT	(ParB)	
boxA2	57735–57741	+	TTGAAAT	(ParB)	
boxA1	57778–57784	+	GTGAAAT	(ParB)	
boxB1	57790–57795	−	TCGCCA	(ParB)	
parO	60027–60068	−	TTTATGCAGCATTTTTAATTAAATTCAAAA ATACAGCATAAA	*par* operon autoregulation (ParA)	73
incA				Plasmid replication control	Reviewed in reference 49
It1	60288–60306	+	GATGTGTGCTGGAGGGAGA	(RepA)	
It2	60319–60337	+	GATGTGCGATAGAGGGAAG	(RepA)	
It3	60349–60367	+	TATGTGCTGTGGAGG**GATC**	(RepA)	
It4	60380–60398	+	TATGTGTGCTGGAGGGAAA	(RepA)	
It5	60411–60419	+	CATGTGCACTGGAGGGAAA	(RepA)	
It6	60442–60460	+	GATGTGTGCTGGAGGGAAA	(RepA)	
It7	60482–60500	−	GATGTGCGCTGGAGGGGGA	(RepA)	
It8	60523–60541	−	GATGTGTGCTGCCGGGAAG	(RepA)	
It9	60554–60572	−	AGTGTGTGCTGGAGGGAAA	(RepA)	
incC				Part of plasmid replication initiation region	Reviewed in reference 51
It10	61489–61507	−	GATGTGTGTTGACGGGAAA	(RepA)	
It11	61511–61529	−	AATGTGTGCTGGCGGGATA	(RepA)	
It12	61532–61550	−	AATGTGTGCTGACGGGTTG	(RepA)	
It13	61553–61571	−	GATGTGTGCTGGAGGGATA	(RepA)	
It14	61574–61592	−	GATGTGTGCTGGAGGGAAA	(RepA)	
pacS				Sites required to initiate packaging	288
	89523–89528	+	**TGATC**A	(PacA)	
	89530–89535	+	**TGATC**A	(PacA)	
	89542–89547	+	**TGATC**A	(PacA)	
	89557–89562	+	**TGATC**A	(PacA)	
	89653–89658	+	**TGATC**G	(PacA)	
	89665–89570	+	**TGATC**A	(PacA)	
	89676–89681	+	**TGATC**A	(PacA)	
	89704–89709	+	**TGATC**G	(PacA)[d]	

[a] Names of functionally linked clusters of binding sites are followed by a list of particular sites within each cluster. It, iteron.
[b] The first number is the coordinate of the first nucleotide of upper-strand (+) binding sites or of the last nucleotide of the lower-strand (−) binding sites, based on arbitrary convention. Coordinates refer to positions of sites in the P1 *c*1-100 *mod749*::IS*5* genome without its nonintegral parts, IS*5*, and the associated 4-bp duplication.
[c] The 5′-GATC sequences that are targets for Dam- or Dmt-mediated N-6 adenine methylation are in boldface.
[d] Deletion of this site reduces the efficiency of cutting.

RecBCD in P1 cyclization could explain the 10-fold-reduced frequency of lysogenization by P1 of *E. coli recB* mutant cells observed by Rosner (261). Presumably, as in *E. coli* (181), Chi sites in P1, by their interaction with the RecBCD enzyme, could also facilitate the reassembly of collapsed replication forks. The burst size of P1 is reduced 20-fold in a *recBC* mutant of *E. coli* as compared to that in the $recB^+C^+$ strain (352).

The distribution of Chi sites in the P1 genome is unequal; the bias is away from the region that includes the invertible segment encoding tail fiber genes and at the right end favors the upper strand, where seven sites are within a 5-kb region around the *pac* site (Fig. 2). Of these, five are on the *lpa* side of *pac*. Their orientation is such that they offer protection against the nucleolytic activity of RecBCD loading onto DNA cut at the *pac* site and unprotected by packaging.

The two octamers that are the most frequent in the genome of *E. coli* (5′-CGCTGGCG and 5′-GGCGCTGG) (35) are not among the most frequent octamers in P1. However, their frequencies in both genomes are similar (approximately 1 per 6,000 nucleotides), suggesting that they have related functions. The majority of octamers that are frequent in P1 DNA, other than Chi sites and its variants, correspond to either sequences that are specific for P1 and represent fragments of 17-mer C1 binding sites or 19-mer RepA-binding iterons (Table 1) or fragments of AT-rich regions containing A or T tracts. The most infrequent tetramer in P1 DNA (5′-CTAG) is the same

TABLE 2. Alphabetical index of P1 genes

Gene	Location and direction or comment[a]	Gene	Location and direction or comment	Gene	Location and direction or comment[b]
1	See *mat*	*gta*	See *hdf*	*pmgM*	77.6 → 78.2
2	See *lydB*	*icd*	47.2 → 47.4	*pmgN*	78.2 → 78.4
3	See *bplA*	*iddB*	25.3 → 25.7	*pmgO*	78.3 → 79.0
4	See *pro*	*imcA*	93.2 ← 93.3	*pmgP*	78.9 → 79.5
5	72.5 ← 73.1	*imcB*	93.1 ← 93.2	*pmgQ*	80.2 → 80.9
6	73.1 ← 74.1	*insA*	21.8 ← 22.1	*pmgR*	81.0 → 81.2
7	74.9 ← 75.6	*insB*	21.4 ← 21.9	*pmgS*	81.2 → 82.5
8	See *pmgR*, *pmgS*	*isaA*	20.3 → 21.3	*pmgT*	83.3 → 84.2
9	See *pacA*	*isaB*	22.3 → 22.9	*pmgU*	84.0 → 84.3
10	See *lpa*	*hdf*	28.6 ← 29.2	*pmgV*	84.3 → 84.7
11	See *R*?	*hot*	85.3 → 85.6	*ppfA*	62.3 ← 62.7
12	See *R*?	*hrdC*	64.1 → 65.2	*pro*	17.8 → 18.8
13	See *prt*	*humD*	86.3 → 86.7	*ppp*	79.5 → 80.2
14	See *prt*, *pro*	*hxr*	23.0 ← 23.3	*prt*	16.1 → 17.8
15	See *22*, *23*	*kilA*	48.6 → 49.4	*R*	36.6 ← 37.0
16	37.1 ← 38.0	*lpa*	88.7 → 89.2	*rebA*	See *ant1*
17	See *lyz*	*lxc*	8.5 ← 8.7	*rebB*	See *ant2*
18	See *mat*	*lxr*	85.6 ← 86.2	*ref*	2.0 → 2.6
19	See *S*	*lydA*	29.6 ← 29.9	*repA*	60.0 ← 61.4
20	See *U*	*lydB*	29.2 ← 29.6	*repL*	49.5 → 50.3
21	53.3 ← 54.2	*lydC*	30.0 ← 30.3	*res*	3.5 ← 6.3
22	54.2 ← 55.8	*lydD*	18.8 ← 19.1	*rlfA*	50.5 → 51.2
23	55.8 ← 57.5	*lydE*	18.8 ← 19.0	*rlfB*	51.2 → 52.4
24	74.1 ← 74.9	*lyz*	19.1 ← 19.6	*sim*	See *simC*
25	75.6 ← 76.8	*mat*	2.7 → 3.4	*simA*	46.7 ← 46.8
26	76.8 ← 77.2	*mlp*	62.1 ← 62.8	*simB*	46.3 ← 46.6
ant1	47.4 → 48.5	*mod*	6.4 ← 8.3	*simC*	45.5 ← 46.3
ant2	47.7 → 48.5	ORF0	See *cre*	*sit*	39.7 ← 43.2
ban	70.0 → 71.4	ORF1	See *cra*	*ssb*	19.8 → 20.3
bof	See *lxc*	ORF2	See *imcA*	*S* (*Sc* + *Sv*)	33.6 ← 36.6
bplA	38.0 ← 39.4	ORF3	See *imcB*	*Sv′*	31.0 → 32.5
bplB	52.2 ← 52.7	ORF4	See *coi*	*tciA*	69.0 → 69.4
*c*1	92.0 ← 92.8	ORF50	See *simB*	*tciB*	69.4 → 69.6
*c*2	See *mod*	ORF93	See *pdcB*	*tciC*	69.6 → 69.7
*c*3	See *mod*	ORFVIII	See *hxr*	*trnA*	68.1 → 68.2
*c*4	47.1 → 47.2	ORFX	See *icd*	*trnI*	69.7 → 69.8
*c*5	See *6*?, *24*?	*pacA*	89.3 → 90.5	*trnT*	68.2 → 68.3
*c*6	See *lxc*	*pacB*	90.5 → 92.0	*tsu*	See *darA*
*c*7	See *22*?, *23*?	*pacC*	See *pacB*	*tub*	44.1 ← 44.8
*c*8	1.5 → 1.8	*pap*	82.5 → 83.0	*tsu*	See *darA*
cin	30.4 → 31.0	*parA*	58.8 ← 60.0	*U*	33.1 ← 33.6
cra	0.2 → 0.4	*parB*	57.8 ← 58.8	*U′*	32.5 → 33.1
cre	0.4 → 1.4	*pdcA*	87.4 → 87.6	*uhr*	64.0 → 64.3
coi	92.9 ← 93.2	*pdcB*	87.6 → 88.7	*ulx*	8.7 ← 9.2
dam	See *dmt*	*phd*	86.8 → 87.0	*upfA*	61.7 ← 62.1
darA	26.7 ← 28.6	*plp*	68.2 ← 68.6	*upfB*	62.7 ← 63.6
darB	9.2 ← 16.0	*pmgA*	39.4 ← 39.7	*upfC*	63.5 ← 63.8
dbn	71.4 → 72.4	*pmgB*	43.9 ← 44.1	*upfM*	84.9 → 85.0
ddrA	26.3 ← 26.7	*pmgC*	44.8 ← 45.3	*upfN*	85.0 → 85.2
ddrB	23.3 ← 26.3	*pmgF*	51.8 ← 52.1	*upfO*	85.2 → 85.3
dmt	65.2 → 67.4	*pmgG*	52.7 ← 53.3	*upl*	68.5 ← 68.8
doc	87.0 → 87.4	*pmgL*	77.4 → 77.6	*vad*	See *ddrA*, *darA*

[a] References to previously mapped genes whose assignment to ORFs in the P1 genome is unclear are summarized in reference 346.
[b] Directions of the tail fiber genes are for C(+) phage.

as the most infrequent tetramer in *E. coli* (35). Seven of its 27 copies are in intergenic regions. Of the remaining 20 that are within 15 protein-coding genes, 9 are within five genes of especially low GC content (three in *res*, two in *mod*, one in *isaB*, two in *rlfA*, and one in *pmgU*), which may represent the most recent acquisitions in the genome of P1.

(ii)Coding sequences. Analysis of the complete sequence of P1 has revealed 112 protein-coding genes and five genes encoding untranslated RNA (Tables 2 and 3 see also Table S1 in the supplemental material). Of these, complete sequences of 42 protein-coding genes and two RNA genes had been determined previously (149, 187, 190, 346). The remaining 73 genes were identified as described in Materials and Methods. Of the previously published sequences of P1 genes, some differ from those presented here by single-nucleotide changes or their overall length, due either to mutations or to previous sequencing errors.

Protein-coding genes account for 92% of the P1 genome. Translation of most of the ORFs is initiated independently from their own ribosome binding and translation start sites (AUG

TABLE 3. Gene products (listed by gene position)[a]

Name[b] (former name[s])	Gene coordinates[c]	Coding strand	Mol mass (Da) (nucleotides in RNA)	Basis for name; known or proposed function	Reference(s)[d]
Cra (gpORF0)	207–428	+	8,704	Putative *cre*-associated function	302[S]
Cre*	436–1467	+	38,539	Cyclization recombinase acting on P1 DNA at redundant *lox* sites on infection; resolves dimers, etc.; possibly modulates copy no.	7,* 10, 21, 106, 182, 302[S]
C8	1518–1829	+	11,313	Establishment of lysogeny?	280, 281, tw[S]
Ref	2077–2637	+	21,328	Recombination enhancement function; aids microhomology recognition, maximal function requires *recBCD* activity	184, 207, 337, 338[S]
Mat (gp1; gp18?)	2716–3468	+	28,933	Particle maturation	191,[S] 346
Res* (EcoP1; R)	3544–6456	−	111,457	Restriction component (EcoP1) of type III restriction-modification (ResMod) system	68, 107,* 145,[S] 152
Mod* (M, C2, C3)	6459–8399	−	73,485	Modification and site recognition component of type III ResMod system; modifies pairs of 5′-AGACC sequences	41, 42, 107,* 145,[S] 152, 262
Lxc* (Bof, C6)	8551–8799	−	9,645	Lowers expression of *c*1, enhances binding of C1 to its operators	164, 272, 311, 312, 318*[S]
Ulx	8796–9236	−	17,079	Unknown; product of gene upstream of *lxc*	tw[S]
DarB*	9270–16037	−	251,542	Defense (in *cis*) against a subset of type I Res enzymes, e.g. EcoB, EcoK; putative DNA methyltransferase and DNA helicase; may methylate DNA during injection	155,* tw[S]
Prt (gp13, gp14?)	16113–17822	+	62,731	Portal protein (by similarity)	252, tw[S]
Pro (gp4, gp14?)	17815–18834	+	36,882	Putative head processing protease and ? kinase; required for head morphogenesis and maturation of DarA precursor protein	321, 324, tw[S]
LydE	18817–19029	−	7,914	Lysis determinant, putative antiholin	tw[S]
LydD	18879–19133	−	9,936	Lysis determinant, putative holin	tw[S]
Lyz (gp17)	19126–19683	−	20,255	Lysozyme	277[S]
Ssb	19852–20340	+	18,023	Single-stranded DNA binding protein; can suppress *E. coli ssb*	187[S]
IsaA	20370–21333	+	35,969	Unknown; product of *IS1* associated gene	tw[S]
InsB	21442–21945	−	19,742	Transposition protein InsB of integral P1 IS*1*	tw[S]
InsA	21864–22139	−	9,901	Transposition protein InsA of integral P1 IS*1*	tw[S]
IsaB	22328–22933	+	23,633	Unknown; product of IS*1*-associated gene	149[S]
Hxr (ORFVIII)	23045–23353	−	11,033	Homolog of *xre;* possible repressor protein	149[S]
DdrB	23343–26330	−	108,752	Unknown; product of the second gene downstream of *darA;* possible protease or response regulator	149[SD]
IddB	25378–25776	+	13,540	Unknown; product of gene internal to ddrB	tw
DdrA (Vad?)	26343–26708	−	13,013	Uncertain; product of gene downstream of *darA;* defect possibly alters P1B:P1S:P1M ratios; *vad* (viral architecture determinant)	149[S]
DarA* (Vad?, Tsu?)	26705–28624	−	69,479	Defends against restriction by type I Res endonuclease and enables DarB to function; internal head protein processed by Pro	149,[S] 155,* 304*
Hdf (Gta?)	28626–29228	−	22,150	Uncertain; homolog to DarA fragment; defect possibly causes *gta* (generalized transduction affected) phenotype	149[S]
LydB (gp2)	29215–29658	−	17,097	Lysis determinant; prevents premature lysis, LydA antagonist	149,[S] 277[S]
LydA	29655–29984	−	11,433	Lysis determinant; holin; promotes cell lysis	149,[S] 277[S]
LydC	30052–30333	−	10,449	Lysis determinant; putative holin (by homology)	tw[S]
Cin*	30461–31021	+	21,230	C-segment inversion; *cix* site-specific recombinase providing for alternate fiber gene expression and hence host range enlargement	109, 135,[S] 142, 150*
gp Sv′	31016–32583	+	55,254	Variable part of S′ tail fiber; gene expressed in (−) orientation	tw[S]
gpU′	32586–33119	+	20,676	Tail fiber assembly chaperone (by similarity to Mu homolog); variable gene of tail fiber operon, expressed in (−) orientation	146, tw[S]
gpU* (gp20)	33148–33675	−	20,258	As gpU′, but expressed in (+) orientation; both gpU′ and gpU may be virion proteins	142,* 146, tw[S]
gpS:Sc + Sv (gp19)	33679–36642	−	104,837	Tail fiber with a constant (Sc) and variable (Sv) segment as in Mu; Sv is exchanged for shorter Sv′ upon C-segment inversion	104, 146, tw[S]
gpR (gp11? gp12?)	36654–37088	−	15,973	Tail fiber structure or assembly	104[S]
gp16	37167–38003	−	31,357	Baseplate or tail tube	104,[SP] 190,[SP] tw[S]
BplA (gp3?)	38003–39436	−	53,577	Putative baseplate structural protein (by homology), may mediate contact between hub and wedges, as does its homolog in T4	tw[S]
PmgA	39433–39789	−	13,239	Putative morphogenetic function	tw[S]
Sit	39789–43211	−	120,710	Structural injection transglycosylase; putative tail tube "ruler"	190[S]
PmgB	43293–44174	−	34,138	Putative morphogenetic function	tw[S]
Tub	44189–44800	−	22,316	Major tail tube protein	tw[S]
PmgC	44811–45377	−	21,017	Putative morphogenetic function	tw[S]
SimC* (Sim)	45587–46399	−	30,800	Confers superimmunity when in high copy number by blocking P1 at entry; requires removal of leader sequence	77, 170, 208*[S]
SimB (gpORF50)	46374–46691	−	12,033	Unknown; superimmunity-linked function	213,[SD] tw[S]
SimA	46713–46844	−	4,842	Unknown; superimmunity-linked function	tw[S]
C4*	47144–47220	+	(77)	Antisense RNA that inhibits Icd and Ant1/Ant2 synthesis by acting on target *ant* RNA; processed to active form by RNase P	26,[S] 33,[S] 57,* 58, 112,[S] 115, 130[S]
Icd* (gpORFX)	47271–47492	+	8,397	Reversible inhibition of cell division, apparently required for *ant* expression	123, 131, 257[S]*

Continued on following page

TABLE 3—*Continued*

Name[b] (former name[s])	Gene coordinates[c]	Coding strand	Mol mass (Da) (nucleotides in RNA)	Basis for name; known or proposed function	Reference(s)[d]
Ant1* (RebA)	47489–48532	+	38,667	Antagonism of C1 repression; forms complex with Ant2	33,[S] 112,[SD] 122, 131,[S] 256*
Ant2* (RebB)	47738–48532	+	29,039	Antagonism of C1 repression, forms complex with Ant1	33,[S] 112,[SD] 131,[S] 256*
KilA	48697–49497	+	29,575	Unknown, expression can kill host	112,[S] 295[S]
RepL	49527–50372	+	30,857	Lytic replication, initiates replication at *oriL* (within *repL*)	112,[S] 295[S]
RlfA	50541–51245	+	27,716	Unknown; possibly replication-linked function	tw[S]
RlfB	51245–52444	+	18,632	Unknown; possibly replication-linked function	tw[S]
PmgF	51875–52171	−	11,423	Putative morphogenetic function	tw[S]
BplB	52255–52764	−	18,804	Baseplate	tw[S]
PmgG	52776–53357	−	20,696	Putative morphogenetic function	tw[S]
gp21	53393–54208	−	29,329	Baseplate or tail tube	323, tw[S]
gp22 (gp15?, C7?)	54218–55807	−	56,935	Sheath protein	323, 325,* tw[S]
gp23 (gp15?, C7?)	55868–57574	−	62,248	Major head protein, present in P1 heads in full-length (62 kDa) and truncated (44 kDa) forms	323, 325,* tw[S]
ParB*	57800–58801	−	37,440	Plasmid partitioning; binds to *parS*, enhances ParA-mediated repression of *par* operon and ATPase of ParA; can spread over DNA flanking *parS*, silencing gene expression; pairs *parS* loci	4,[SD] 70, 73, 83, 90,* 114, 203,[SP] 258, 343
ParA*	58818–60014	−	44,269	Plasmid partitioning; weak ATPase, binds to *parO* repressing transcription, binds to ParB-*parS* partition complex	4,[S] 37, 70, 73,* 343
RepA*	60572–61432	−	32,219	Plasmid replication initiated by chaperone-activated monomers at *oriR;* represses own synthesis; binds at iterated sites (*incC* and *incA*) and can handcuff them, controling plasmid copy number	2,* 3, 6,[S] 49, 52
UpfA	61751–62143	−	14,747	Unknown protein function	tw[S]
Mlp	62155–62487	−	12,597	Membrane lipoprotein precursor (by homology); ? lysis control	tw[S]
PpfA	62321–62770	−	15,744	Possible periplasmic function	tw[S]
UpfB	62783–63628	−	29,310	Unknown protein function	tw[S]
UpfC	63580–63864	−	9,478	Unknown protein function	tw[S]
Uhr	64034–64318	+	10,781	Unknown; product of gene upstream of *hrdC*	tw[S]
HrdC	64311–65216	+	33,638	Homolog of *RdgC* of *E. coli;* possibly involved in replication, recombination, and/or Dmt function	tw[S]
Dmt (Dam)	65216–67477	+	83,595	Predicted bifunctional DNA methyltransferase; methylates A's in GATC sequences, and (by homology) probably C's at unknown sites; affects replication control, late gene expression, packaging	59, 66, tw[S]
tRNA1	68149–68224	+	(76)	Proposed tRNA-Asn specific for the AAC codon	tw[S]
tRNA2	68227–68302	+	(76)	Proposed tRNA-Thr specific for the ACA codon	tw[S]
Plp	68273–68650	−	13,800	Unknown; putative lipoprotein	tw[S]
Upl	68651–68842	−	7,092	Unknown; product of gene upstream of *plp*	tw[S]
TciA	69034–69459	+	15,187	Tellurite or colicin resistance or inhibition of cell division (by homology)	tw[S]
TciB	69459–69623	+	5,752	Tci accessory protein (by homology)	tw[S]
TciC	69623–69730	+	3,719	Tci accessory protein	tw[S]
tRNA3	69734–69809	+	(76)	Proposed tRNA-Ile specific for ATA codon	tw[S]
Ban*	70096–71460	+	50,479	DnaB (of *E. coli*) analog and homolog; replicative DNA helicase	67, 195*[S] 237, 254*
Dbn	71460–72458	+	37,379	Unknown; product of gene downstream of *ban*	tw[S]
gp5	72505–73137	−	21,669	Baseplate	323, tw[S]
gp6 (C5?)	73130–74146	−	37,222	Tail length determination	323, tw[S]
gp24 (C5?)	74148–74933	−	28,880	Baseplate or tail stability	323, tw[S]
gp7	74920–75648	−	27,145	Tail stability	323, tw[S]
gp25	75652–76869	−	45,834	Tail stability	323, tw[S]
gp26	76879–77256	−	14,456	Baseplate	323, tw[S]
PmgL	77436–77648	+	9,258	Putative morphogenetic function	tw[S]
PmgM	77651–78229	+	21,888	Putative morphogenetic function	tw[S]
PmgN	78296–78451	+	5,593	Putative morphogenetic function	tw[S]
PmgO	78393–79055	+	24,217	Putative morphogenetic function	tw[S]
PmgP	78953–79579	+	23,198	Putative morphogenetic function	tw[S]
Ppp	79579–80253	+	25,568	P1 protein phosphatase, possible role in head morphogenesis	tw[S]
PmgQ	80250–80951	+	26,912	Putative morphogenetic function	tw[S]
PmgR (gp8?)	81033–81251	+	8,318	Putative morphogenetic function	tw[S]
PmgS (gp8?)	81253–82515	+	48,345	Putative morphogenetic function; ? prohead scaffolding protein	tw[S]
Pap	82588–83094	+	19,200	P1 acid phosphatase (by homology); ? head morphogenesis	tw[S]
PmgT	83289–84218	+	35,155	Putative morphogenetic function	tw[S]
PmgU	84023–84358	+	12,163	Putative morphogenetic function	tw[S]
PmgV	84355–84717	+	13,975	Putative morphogenetic function	tw[S]

Continued on following page

TABLE 3—*Continued*

Name[b] (former name[s])	Gene coordinates[c]	Coding strand	Mol mass (Da) (nucleotides in RNA)	Basis for name; known or proposed function	Reference(s)[d]
UpfM	84900–85013	+	4,343	Unknown protein function	tw^{S}
UpfN	85029–85226	+	7,153	Unknown protein function	tw^{S}
UpfO	85204–85344	+	5,238	Unknown protein function	tw^{S}
Hot	85368–85631	+	9,694	Homolog of theta subunit of DNA polymerase III; ? replication	tw^{S}
Lxr	85638–86210	−	22,042	LexA-regulated function	tw^{S}
HumD*	86385–86774	+	14,313	Homolog of UmuD′ subunit of *E. coli* repair protein DNA polymerase V (UmuD′-UmuC); complements mutant UmuD′	200, 217,* 276^{S}
Phd*	86847–87068	+	8,132	Prevention of host death by Doc toxicity; confers, with Doc, addiction to P1; represses transcription of addiction operon	95,* 192,^{S} 194, 210
Doc*	87068–87448	+	13,587	Death on curing; toxin of P1 addiction system; reversible inhibitor of protein synthesis; corepressor of addiction operon	95, 121, 192,^{S} 210, 211*
PdcA	87432–87632	+	7,630	Unknown; post-doc	192,^{S} tw
PdcB (gpORF93)	87660–88703	+	39,877	Unknown; post-doc	192,^{SD} tw
Lpa* (gp10)	88792–89244	+	18,117	Late promoter activator	111,* 188^{S}
PacA* (gp9)	89330–90523	+	45,246	DNA packaging; cuts at *pac* together with PacB	286,^{S} 288*
PacB*	90523–92007	+	55,604	DNA packaging; cuts at *pac* together with PacA	286,^{S} 287*
C1*	92032–92883	−	32,544	Controls plaque clarity; primary repressor of lytic functions	79,* 84, 124, 238,^{S} 316, 319
Coi* (gpORF4)	92994–93203	−	7,741	C one (reversible) inactivator; forms 1:1 complex with C1	25,^{S} 27, 84,^{S} 127, 130*
ImcB (gpORF3)	93169–93264	−	3,659	Immunity C function; may regulate *c*1	27, 84^{S}
ImcA (gpORF2)	93283–93396	−	4,340	Immunity C function; may regulate *c*1	27, 84^{S}

[a] A more complete table, which includes pI, number of amino acid residues (or nucleotides in RNA), and features of the amino acid sequence (closest homologs and amino acid sequence motifs) is to be found as Table S1 in the supplemental material.

[b] Gene products marked by an asterisk were either purified or identified in PAGE gels.

[c] Coordinates of genes refer to their positions in the P1 *c*1-100 *mod749*::IS*5* genome without its nonintegral parts, IS*5*, and the associated 4-bp duplication. Coordinates of protein-coding genes are from the initial codon through the first stop codon.

[d] References marked by an asterisk concern identification of a protein or RNA product of a given gene; those marked by a superscript S concern previously published sequences. Superscript SD, published sequence differs internally from sequences presented here; superscript SP, published sequence is partial; tw, this work.

[86]≫GUG[16]>UUG[10]). Translation stops predominantly at UAA and less frequently at UGA and UAG.

Utilization of UUG as the translation initiation codon appears to be more frequent in P1 than in other coliphages (224) and in *E. coli* itself (35). It was confirmed experimentally that translation of the P1 *lxc* gene, which encodes a modulator of C1-mediated repression, initiates at UUG (272).

Whereas the translation initiation codons of P1 are generally 6 to 10 nt downstream of typical *E. coli* ribosome binding sites (RBS), genes *plp*, *pmgO*, and *pdcA* lack nearby RBSs. They are probably translationally coupled to genes immediately upstream (*plp*) or overlapping (*pmgO* and *pdcA*). The gene *ref* (recombination enhancement function) in a monocistronic operon is more unusual. The sequence that most closely resembles a canonical RBS is separated by 16 bp, including a run of eight T's, from a presumptive *ref* initiation codon (207, 338). What purpose the extra RNA might serve or how it is accommodated remains to be determined.

Promoter-proximal regions of four genes, *pro*, *hrdC*, *pmfA*, and *pmgL*, contain, in addition to ATG codons preceded by sequences with a good match to RBSs in mRNA, TTG codons similarly preceded by canonical RBS sequences and further upstream. Possibly, each of these genes encodes two proteins that differ in the length of their N-terminal domains. Alternative translational starts in mRNA at UUG or AUG could control expression of these genes by modulating the intracellular concentration or activity of their protein products.

Although genes of P1 are tightly packed within the genome, overlaps are rare and occur mostly at coding termini. Spaces between genes of adjacent operons are in most cases limited to short regions containing promoter and regulatory sequences. Of those that exceed 250 bp, one contains *oriR* and another *incA*, both involved in P1 prophage replication (Table 1). A 272-bp region between the first gene of the *sim* operon, *simA*, and *c*4, was shown to encode the 5′ region of the *c*4 transcript, which is cleaved to yield C4 RNA (115). Two regions within the *ban* operon (531 bp preceding the *trnA* gene and 359 bp between the *trnI* gene and *ban*, described in detail later in this paper) may also encode species of RNA that are processed.

Of previously identified or predicted products of P1 genes, 65 exhibit significant homology to known proteins of other organisms (Table 3; see also Table S1 in the supplemental material). Only 29 are homologous to proteins encoded exclusively by other phages. Although they include 11 homologs of proteins of T4 or T4-like phages, there is no clear preference in these homologies for proteins of a particular phage, suggesting that P1 has had a long, separate history. Homologs of known genes of different provenance are scattered throughout the P1 DNA. Whether their distribution results from the horizontal transfer of modules during the evolution of the P1 genome or from convergent evolution remains to be determined.

(iii)Organization of P1 genes and transcription units. P1 genes can be grouped into 45 operons, of which 15 appear to be monocistronic. Their organization resembles that of T4 (224) in that only some genes of related function are in clusters.

Regulatory regions of 38 operons contain one or more sequences that resemble strong σ^{70} promoters (Table 4). Of 26 σ^{70} promoters identified previously by transcriptional fusion or primer extension experiments, only 9 had a match to the consensus sequence below 50% and required relaxing the stringency of the prediction program for their detection. All nine

TABLE 4. Known and putative σ^{70} promoters of P1 genes

Name (former name)	Position[a]	Strand	Sequence[b] −35	Sequence[b] −10	Basis for identification[c] *tsp*	*pp*	*cp*	Reference(s)
P2*coi*	41	−	TGCTAAGGTA TTGAAC TGTATGGATTTACAGG	TAAATT **GATC**ATATTC			56.8	
P3*cre*	104	+	ACTTGATCAT TT**GATC** AAGGTTGCGCTACG	TAAAAT CTGTGAAAAA	√	√	42.6	302
P2*cre*	283	+	GCAAAAGGGT TT**GATC** GTGATAGTTGCCAACTGT	CATTAT CGCGCGGTGA	√	√	45.6	302
P1*cre*	400	+	GAGCC**GATC**C TGTACA CTTTACTTAAAAC	CATTAT CTGAGTGTTA	√	√	<40.0	302
P*c8*	1394	+	AATGTAAATA TTGTCA TGAACTATATCCGTAACCTG	GATAGT GAAACAGGGG			54.4	
P2*ref* (P*ref*-2)	1957	+	AACAAAATAT CTATCA TTGCTCTAATTGATTGC	TATAAT TGAGCCGCAG			59.8	338
P1*ref*	2017	+	TACGAAGACG TTGCCA TTACTTCACTCCTTGA	CATCAT TGGCGGCCAT			52.7	
P*mat*	2666	+	TGATAAATGA TTGACA ACTGACAAGTGACTTCAGT	CAGAAT CATCACACGC			56.8	
P1*res*	6558	−	GATGAAGATT TTGCCC CAAACAAAGTCGTGTTT	TATGGT TGTAACTTTG			60.4	
P2*res* (PR1)	6891	−	CAGGAGGTAT TTAAAA AAAGCACAACCACAAGT	GATAAT GCGGCTAAAA	S1	√	52.1	285
P3*res*	6926	−	A**GATC**TATTT TTGAAA TAACAAAAAAAAGGA	TACAGG AGGTATTTAA	S1	√	47.3	285
P1*mod* (PM1)	8450	−	CATATAGATA GTGATA GTGCCACAACTTCTGGC	TCTAAC GGGCTGGGGA		√	45.6	285
P2*mod* (PM2)	8548	−	ATCATAGATA ATGCAT TGCTTCTCAATGCGGAT	TAAACT TCCCCAAATT	√	√	50.9	285
P1*lxc*	8844	−	CTGCCAGTCC CTGAAG AAGTTGACGACCAGGAC	TATGAG TTTGAGTCTT	S1		46.7	285, 318
P2*lxc*	9022	−	AATTAAAAAA TTATTA CTGGAGCGCGGTTATGG	TAATTT CGGTCGTCTC	S1		51.5	285
P3*lxc*	9167	−	AAAACATATC TTGAGG CGGTACTGCGTACGG	CATTGT TAGGAAAGAC	S1		51.5	285
P*darB*	16093	−	CGTGATTTTA TTGTCT GCCAAGATTGCACCTTAAT	TAGAAT AATTCACATC			56.8	
P2*ssb* (Pr21)	19787	+	TGCAAATCAC TTGTTA GCTACGTTTCAAAGATA	TACATT ATTGCTCTAA	√		52.7	59
P1*ssb*	19821	+	ACATTATTGC TCTAAT TAATTTATTTTATTAGG	TAAGAT AAGTGGCACA			55.6	
P*lyz*	19845	−	GATGACTTTG TTTACA CCGCGTTGTGCCACT	TATCTT ACCTAATAAA			53.3	
P*isaB*	22220	+	CAATGATGGT TAGACA AGCAAGATAGGGCGAAT	TAAAAT TAATCAAGAG			61.5	
P*insA*	22229	−	AAAGTGACTC TTGATT AATTTTAATTCGCCC	TATCTT GCTTGTCTAA			57.4	
P2*cin*	30339	+	CACATAACTA TTTCCT TAATAGTGAATTAAA	TATCAT TGGGAAACGG			55.6	
P1*cin*	30390	+	TACTTTGTGA TTTCCA CACATACTGGTTTTTGTTAAT	TAAAAT CCGCAGCTTG			53.3	
P*lydC*	30461	−	TGTTCATTTG TTGATA CGCGTACATAGCCTAT	TAGCAT ATTTTCTGCT			59.2	
P1*pmgC*	45555	−	AAAATAGTGT TTGAAA GGATAGTCAATTTAAGT	TTTAAC ACAACGCCCA			59.8	
P2*pmgC*	45636	−	CAACAATTCG TTGATG TAATGGAAGGATTTGA	TAGCAT CATGTATAAA			59.8	
P3*pmgC*	45771	−	CAAATATATA TTCACA AATCAA**GATC**CTGTTCA	GATGAT GAATTTTATA			58.0	
P*simB*	46843	−	TGTACCCTAA ATGAGT TATAAGGCAGGTGAGGT	TATAAT GAGAAAACTA			56.8	
P*simA*	46909	−	GCATTATTTG TTTAAT AAATACACAGTTG**GA**	**TC**TAAT AACCTCTTTT			51.5	
P2*c4* (P51a)	46956	+	TCTAATAAAT TTGTAT TTTTAAGTCGCGAATGC	TATCTT TTCGCATCAT	√*		56.8	131
P1*c4* (P51b)	46996	+	TCGCATCATA TTGACC TTTTAATCGTTCAGGCT	TATAGT TCCACCGTCG	√*		60.4	131
P*kilA* (P53)	48593	+	CAACCC**GATC** TG**GATC** GGGTCAGAAAAATTTGC	TCTAAT AAATTTCGTT	√	√	50.3	295
P*askilA* (P53as)	48724	−	GGTTGAAGGA TCGACA TTTTGATGAAGGTTTGA	TATATT CATATCCGCA	√	√	62.7	125
P*repL*	49267	+	AAACTATCTC TTCACC TGGTTGCGTGATAACGG	AATTCT **GATC**GCAACC			52.1	
P*rlfA*	50486	+	CGCCCCATTA TTGCAA TTAATAAACAACTAACG	GACAAT TCTACCTAAC			52.7	
P1*rlfB*	50663	+	AAAAAATTGC TAGAGA CATCAAGGAATTAAAAT	TAAAAC ACAATCTAAA			55.0	
P2*rlfB*	50704	+	AATCTAAAAC CTGATT TCGAAATCAAGTGGAC	TAAAGT ATCTGCGTCC			56.2	
P3*rlfB*	50746	+	GCGTCCAAAG TTGAAT TTTACTTAGATGTCG	TAGATT ACTTCTTCAG			55.0	
P4*rlfB*	50971	+	ATTGAAAAAA TTGAGA AATTAAGAGGGGTTCTTCA	CAAT**GA** **TC**GCTATGAT			57.4	
P*asrlf*	51784	−	TAGCTAGCTA TTGAAA AGCTAAAGTA**GATC**GGT	TAGATT TGCATTACTC			69.2	
P1*pmgF*	52236	−	TACAACTGGA TTGAGT GGGATTAAGAGTCATCCCTTG	TATTTT AAAGCTCCTT			53.3	
P2*pmgF*	52342	−	GCTACAACAA TTGAAA TGTTGGACTGCAAAATC	TACAGT GATGCAATCG			60.9	
P1*23*	57631	−	AAAAAATTGC TTGTCA CGAGAAAGTCAACAAGTGACT	TTCAAT AAAATCTCTT			57.4	
P2*23*	57740	−	TATTAACTGA CTGTTT TTAAAGTAAATTACTC	TAAAAT TTCAAGGTGA			56.2	
P*parB*	59274	−	ATTGCCGCGG AAGATT TTGCTAAAGCAGTTTTT	GACCGT ATTGAATTTA	√	√	<40.0	A. Dobruk, unpublished
P*par*	60069	−	GAGAATGCTA TGTACA AGCATCTACGCATACAT	TATTAT TTTATGCAGC	√	√	54.4	73, 120
P*repA*	61505	−	TGCTGACGGG TTGCTA ATGTGTGCTGGCGGGATA	TAGGAT GTGTGTTGAC	√	√	<40.0	52
P1*ppfA*	62785	−	GATTCTGTTC TTGATG AATTAAACAAAGCGG	CACAGT AACAAAGGAC			50.3	
P2*ppfA*	63001	−	GAACAACAAG TTGAAG CACTGAAGGAAGCTACTG	AACAAT TTGGTGCTTT			52.7	
P*dmt*	63962	+	CTAATAAATA TTGTTT TTTATGTCGTGTTTTCGG	TACCAT TCAGCCATCG	√		63.3	59
P2*ban*	67452	+	GGTGTAGCTC ATGACG TTTTGGGTATATTGCTCTA	TAAATT TATTAGTGTA			52.7	
P1*ban*	67469	+	TTTGGGTATA TTGCTC TAATAAATTTATTAGTG	TAATAT CGCCTCAATG	√*		50.3	209
P1*upl*	69024	−	GAATTTCTTT TTGAGA ATGCCGAGCATTTAT	TAACCT CATTACTGGT			56.2	
P2*upl*	69055	−	TTCCATCTTC TTGACT CCGCCAGCCGCTTTGCG	GAATTT CTTTTTGAGA			51.5	
P2*dbn*	70823	+	GTTTTTAGCC TTGAGA TGCCGAGCCACCAGCT	GATGAT GCGCTCACTG			53.8	
P1*dbn*	71207	+	AATCGATCGC TTGAAC AGCGTGCGGACAAACGACCGG	TAAATT CAGATTTACG			52.1	
P2*pmgL*	77306	+	TAGTCATTTG TTGAAT ATTTAACTCAATAAAA	GAAAAT TATTAGTGCA			56.8	
P1*pmgL*	77339	+	AAAATTATTA GTGCAA TTTTGATTGTGAAATG	TATCAT TCTGCCCTTA			56.8	
P*26*	77617	−	AACATAATCA GTCAAA TCCTGCCGGTCGCATGC	CATATT TACGCTGTTT			50.3	
P*pmgN*	78197	+	TGAGCATGTA TTGAAC GGCGATAAGTGGTTG	TATGCT GCTTAAAATA			49.1	

Continued on following page

TABLE 4—*Continued*

Name (former name)	Position[a]	Strand	Sequence[b] −35	Sequence[b] −10	Basis for identification[c] *tsp*	*pp*	*cp*	Reference(s)
P*pmgR*	80954	+	CAGAATAATA TTGCTC TAATAATTCCATATTTT	TAAAAC GTGATGTACA			52.7	
P1*pmgT*	82778	+	ACTCCAGTTA TTGATA TGGTGCGCCAGTTATT	TAACGT TTACACGGTC			56.8	
P2*pmgT*	83095	+	TCAACAATAA TTGCTC TAATAAATCTTGATTTT	TAAAAC AGAGAAAGTG			52.7	
P3*pmgT*	83110	+	CTAATAAATC TTGATT TTTAAAACAGAGAAAGT	GAAAAT AAAAACATGC	√		66.3	59
P1*upfM*	84752	+	CAGAAAATGA TTGTTT CCACATCAAGGAGATTT	TAATGT TTCACTGAAA			55.6	
P2*upfM*	84785	+	TAATGTTTCA CTGAAA CATTAAGTAAGCCAGTG	CATAAT TCCATTTTTT			55.6	
P*upfN*	84973	+	CTTGCTCACG TTGACG TAGAAACCCAACCCTTA	TATAGT TGGATTCGGT			59.8	
P1*lxr*	86267	−	TCAACAACCG CTGATG ATTTTGTGCGCTTTGC	TACTAT TCATCACCAA			49.7	
P2*lxr*	86347	−	CCATAGCGTT TTCACC TCAATAATACTGTTCATTTA	TACAGT ACACATTAAA			48.5	
P1*humD*	86318	+	ACGTAACGCG TTTGCA CCAAAGGTGTCTCTT	TAATGT GTACTGTATA			<40.0	200
P*phd*	86787	+	TGCAAAGTGC TGGTGC TTTATGCCTGTGAAGTT	TATAAT TGTGTACACA		√	52.1	210
P*lpa* (Pr94)	88719	+	CTCTAATGTA TTGCTA TTTCTTTAATCGAGGG	TATTAT ATTCCACGTT	√	√	57.4*	193
P*pacB*	90384	+	GAAGAAATTG CCGATA TCGTCGATACAGGTGGT	TATGGT GATGTCGATG			50.3	
P*c1* (P99a)	92951	−	CACTAATAAA TCTATT ATTTTCGTTG**GATC**CTTC	TATAAT GGTGGCCAAC	√	√	46.7*	84
P1*coi* (P99d)	93438	−	TCATAGTTGT TTGACA ATTGCTCTAATAAATTAT	AGTTTT GCCGCCGTTT	√	√	60.9*	127
P4*cre* (P99e)	93541	+	CTAATAATTC TTGATT TTTATGCGCAGCTGGACG	TAAACT CCTCTTCAGA	√	√	63.9*	84

[a] Coordinate of the first nucleotide of the −35 sequence for promoters from which transcription proceeds on the + strand (clockwise) or of the last nucleotide of the −10 sequence for promoters from which transcription proceeds on the − strand (counterclockwise). Coordinates refer to positions of promoters in the P1 *c*1-100 *mod749*::IS*5* genome without its nonintegral parts, IS*5* and the associated 4-bp duplication. The locations of P3*res*, P1*mod*, P1*lxc*, P2*lxc*, and P3*lxc* promoters proposed in reference 285 are corrected.

[b] Conserved −10 and −35 promoter hexamers as well as the "extended −10" motifs (T and G at positions −15 and −14, respectively) and known transcription start sites are underlined. The 5′-GATC sequences that are targets for Dam- or Dmt-mediated N-6 adenine methylation are in boldface.

[c] Identification of promoters is based on results of transcription start point determination (*tsp*) or RNA polymerase binding assay (*tsp*∗), detection with the promoter probe vector (*pp*), or similarity of the sequence to the consensus sequence for σ^{70} promoters calculated with the use of the program Targsearch (*cp*) (229). References that indicate existence of a promoter in a given region based on S1 nuclease mapping of 5′ ends of transcripts are marked with an S1 in the *tsp* column. Putative promoters which have not been verified experimentally are shown only if their homology to the consensus exceeds 50%, as calculated by the method of Hawley and McClure (118) and if their location suggests a function in initiation of transcription of genes or regulatory RNA. Exceptionally, putative promoters that have a homology score to the consensus below 50% but overlap putative regulatory sites are also shown. For comparison, homology scores are shown for both known and putative promoters, the former marked with an asterisk.

promoters are associated with genes expressed during lysogeny: three with *cre*, three with *res* or *mod* (including the P1*lxc* promoter, which can also drive transcription of *lxc*), one with *parB*, one with *repA*, and one with *c*1. A relatively high predicted strength of known and putative promoters of genes expressed during P1 lytic development implies that as soon as these promoters become accessible to the host RNA polymerase, they can effectively compete with *E. coli* promoters for the enzyme. Multiple promoters within regulatory regions of several operons indicate a requirement for different controls on the expression of genes at different stages of phage development or plasmid maintenance or a requirement to adapt to alternative hosts or environmental conditions.

Ten promoters, including a promoter of the essential phage replication gene *repL*, contain the so-called "extended −10 region" characterized by the presence of T and G at positions −15 and −14, respectively. In *E. coli*, transcription from extended −10 promoters can initiate in the absence of typical −35 hexamers (178).

Sixteen operons end in sequences characteristic of typical Rho-independent transcriptional terminators (Fig. 1 and Table 5). Either termination of some operons of P1 occurs by a Rho-dependent mechanism or certain Rho-independent transcription termination sites are too weak to be detected. Perhaps in the latter case a terminator further downstream is alternatively used. In the case of certain P1 operons expressed during lysogeny (for instance, *res* and *phd doc pdcA pdcB*), C1 repressor-binding sites of downstream operons may possibly be used as roadblocks to transcript elongation.

Five predicted Rho-independent transcription terminators (T*c*8, T*ref*, T*ddrA*, Tp*lydC*, and T*dmt*) follow closely upon apparently functional promoters. The locations of those terminators (in a leader sequence or early in a gene) suggest control by antitermination. Indeed, the T*c*8 terminator was shown to participate in the regulation of transcription of a gene downstream of *c*8, *ref*, by prematurely terminating transcription from one of the *ref* promoters (338). Of the five terminators, four are positioned in such a way that, in addition to premature termination of *ref*, *mat*, *ddrB*, and *dmt* genes, they can function to terminate transcription of operons immediately upstream. It is likely that P1 encodes an antitermination mechanism. Whereas no P1 protein has been implicated in the process, antitermination might not require such a protein (330).

As many as 12 predicted Rho-independent terminators (Tas*ref*, Tas*pro*, Tas*lyz*, Tas*ddrB*, Tas*ddrA*, Tas*kilA*, Tas*23*, Tas*tciA*, Tas*7*, Tas*pmgM*, Tas*pmgT*, and Tas*doc*) are at the beginning of operons but opposite in orientation to the operon. This arrangement suggests that these terminators could terminate transcription of regulatory antisense RNAs. Antisense RNA has been shown previously to regulate the expression of *icd ant1/2* and *kilA repL* (125, 126). The location of the predicted Tas*kilA* terminator is consistent with that expected from the length of the antisense RNA transcribed from the Pas*kilA* promoter (125). The antisense RNA terminated at Tas*23* may

TABLE 5. Known and putative Rho-independent transcriptional terminators in P1 DNA

Name[a]	Position[b]	Strand	Sequence[c]	Relative location[d]	Basis for identification[e]
Tas*ref*	1836	−	TTTGTGCAGCCTGGCTCCTTGCCAGGCTTTTTTTTATTTCATCATGG	5′-*c*8-3′ ⌈5′-*ref*-3′	TT, TR
T*c*8	2046	+	TCCTTGACATCATTGGCGGCCATTAGGCCGCCTTTTTTTTGCCATAT	5′-*c*8-3′ ⌉5′-*ref*-3′	TT, TR
T*ref*	2788	+	CTTCAGTTGTATTGCTAAGCCGCCGCTGGTGGCTTTTCTTTTTTGTA	5′-*mat*⌋*mat*-3′	TR (338)
Tas*pro*	17761	−	CCGATTCAAGGATTTGCGCCAGTTCCTGTGGCCGGTGTTTTGATGATG	5′-*prt* ⌈*prt*-3′	TR
T*pro*	18848	+	CCACCAAAAATAACCCCGGCAGCTGCCGGGGTTCTCGTTAACTATTA	5′-*pro*-3′⌉3′-*lyz*-5′	TR
T*lyz*	18848	−	ATAGTTAACGAGAACCCCGGCAGCTGCCGGGGTTATTTTTGGTGGTT	5′-*pro*-3′ ⌈3′-*lyz*-5′	TT, TR
Tas*lyz*	19655	+	GATAGCGCAAATTGCACCGCCTCCTGCGGCTGTTTTTCCCTTCATAA	3′-*lyz*⌉*lyz*-5′	TR
T*insB*	21351	−	TTTCGTATTATCAGGGTGCTCTCTGGCACCCCTTTTTATTCTAGTAA	5′-*isaA*-3′ ⌈3′-*insB*-5′	TR
Tas*ddrB*	26246	+	CACCACATACGAAACCCGCCCGAAATCGGGCAGGTTCTTTAGCAGCT	3′-*ddrB*⌉*ddrB*-5′	TR
T*ddrA*	26246	−	TGCTAAAGAACCTGCCCGATTTCGGGCGGGTTTCGTATGTGGTGACA	3′-*ddrB*⌈*ddrB*-5′	TT, TR
Tas*ddrA*	26508	+	ACCAGTTTTGGCACACCATCATGTTCCGAAGGTTGTTTCTCCTCGCC	3′-*ddrA*⌉*ddrA*-5′	TR
Tp*lydC*	30287	−	ACAATAATGCTCCTCTCCATCCTTGGCGGGGTGCATTCGTTTCTGAA	3′-*lydC* ⌈*lydC*-5′	TR
T*U′*	33125	+	AACCTGAATGAGACAAGGCCCGATAGCGGGCCTTAATTTTTATTCAG	5′-*U′*-3′⌉3′-*U*-5′	TT, TR
T2*simC*	45384	−	ACTCGTAGAATCGGTTAACACACCAGATTCTACGAGGTTTCAATGAC	3′-*pmgC*-5′ ⌈3′-*simC*-5′	TT
T1*simC*	45535	−	GTTTTAACACAACGCCCATTAAAGGGCGTTTTATTGTTTTACTCAAA	3′-*pmgC*-5′ ⌈3′-*simC*-5′	TT, TR
Tas*kilA*	48546	−	CTGCTCAGAAACCTGCCAGTTTGCTGGCAGGTTTTTTTCTTTTGTTA	5′-*ant*-3′ ⌈5′-*kilA*-3′	TR
T*pmgG*	52219	−	CATCCCTTGTATTTTAAAGCTCCTTCGGGAGCTTTTTTGTGCTTAAA	3′-*pmgG*-5′ ⌈3′-*pmgF*-5′	TT, TR
T*22*	53368	−	GGCTAAAGGAGGAGGCCGGGGAACTCCGGCCTTTAACTTGAATGGCT	3′-*21*-5′ ⌈3′-*22*-5′	TT, TR
Tas*23*	57399	+	CATTATCAGAGATACCCGCAAAAACCGGGTCTTTACGTGCAGCTTCA	3′-*23*⌉*23*-5′	TR
T*parB*	57707	−	TTTCAAGGTGAAATCGCCACGATTTCACCTTGGATTTTACCTTCCTC	3′-*23*-5′ ⌈3′-*parB*-5′	TT, TR
T*repA*	60474	−	TTTGATGTGCGCTGGAGGGGGACGCCCCTCAGTTTGCCCAGACTTTC	3′-*parA*-5′ ⌈3′-*repA*-5′	TT
T*upfA*	61718	−	AATACGACTCCCTTCCAACCGGCTACGTTGGCCGGTTTTTCACTTAT	3′-*repA*-5′ ⌈3′-*upfA*-5′	TT, TR
T*hrdC*	67115	+	TGGAAGCCTTCCTCGGTGCGGAAGGGCGGTCTGAGTTTGTATGCGGT	5′-*dmt*⌉*dmt*-3′	TT
T*dmt*	67659	+	GAGCTTCCAGTTTGCCCATCTTCGGGTGGGCGTTTTTTCAGGGTTTT	5′-*dmt*-3′⌉5′-*trnA*-3′	TR
Tas*tciA*	69059	−	TCTTCTTGACTCCGCCAGCCGCTTTGCGGAATTTCTTTTTGAGAATG	5′-*tciA*⌉*tciA*-3′	TR
T*dbn*	72485	+	CCGTCATAACAGAAAAGCCCGAACGCCGGGCTTTTCTTAAGCCTTGT	5′-*dbn*-3′⌉3′-*5*-5′	TT, TR
T*5*	72485	−	TGACAAGGCTTAAGAAAAGCCCGGCGTTCGGGCTTTTCTGTTATGAC	5′-*dbn*-3′ ⌈3′-*5*-5′	TT, TR
Tas*7*	75768	+	ACTTTATGAGGTTGTTCGCCTGCGTTTGCAGGGCTTTTTTGAGAGCG	3′-*25*⌉*25*-5′	TR
Tas*pmgM*	77765	−	GTGCTGCAAGTGCCTTCGCATGTGGAGAAACTTCTTTTACCGGCGCA	5′-*pmgM* ⌈*pmgM*-3′	TR
T*pmgQ*	80982	+	TTTTAAAACGTGATGTACACTCATCACGTTTTTTATTAGAGCAATCT	5′-*pmgQ*-3′⌉5′-*pmgR*-3′	TT
T*pmgS*	82537	+	GTCACTTTTAATGCTGGTGGAGTGCGCCCACCAGCATTTTTTTCGTC	5′-*pmgS*-3′⌉5′-*pap*-3′	TT
T*pap*	83146	+	AGTGAAAATAAAAACATGCCGCAAGGCGCGGCATGTTTCCAATCAAT	5′-*pap*-3′⌉5′-*pmgT*-3′	TT, TR
Tas*pmgT*	83144	−	GATTGGAAACATGCCGCGCCTTGCGGCATGTTTTTATTTTCACTTTC	5′-*pap*-3′ ⌈5′-*pmgT*-3′	TR
Tas*doc*	87120	−	TCAGACATTCCCGGCAGGCCGCCGTAGCGGCTTATATTCGCATCATG	5′-*doc* ⌈*doc*-3′	TR
T*lpa*	89255	+	ATTGATTGCTTGCCCGTTCCGGGCCTTTTGACATGTGACTTTCGTTA	5′-*lpa*-3′⌉5′-*pacA*-3′	TT
T*pacB*	92399	+	TTGTTTCGCTAAAGCCGTGTACGCGCATACGGAATTTTTCATCCTCC	3′-*c*1⌉*c*1-5′	TR

[a] Names of the terminators (T) that could terminate transcription of RNA antisense to mRNA of a given gene contain "as" and the name of the gene. The terminator in *lydC* is marked T*lydC* to indicate that it could serve primarily to terminate *lydC* transcription prematurely.

[b] Coordinate of the first nucleotide of the terminator hairpin for terminators that terminate transcription on the + strand (clockwise) or the last nucleotide of the terminator hairpin for terminators that terminate transcription on the − strand (counterclockwise). Coordinates refer to positions of terminators in the P1 *c*1-100 *mod749*::IS*5* genome without its nonintegral parts, IS*5*, and the associated 4-bp duplication.

[c] Nucleotides that form the stem of the terminator hairpin are underlined.

[d] Terminators are represented schematically by symbols ⌉ and ⌈, which face toward the DNA region whose transcription they terminate.

[e] Identification of terminators is based on their similarity to known transcription terminators as determined by the program TransTerm (TT) or Terminator (TR) and verified by additional criteria (see Materials and Methods).

regulate translation of the *23* gene from an alternative, upstream start site.

Genes expressed during establishment or maintenance of P1 prophage. The alternative life-styles of P1 require two sets of genes with little functional overlap. In general, the genes associated with lysogeny can be distinguished by the absence of regulatory regions resembling those known to control P1 lytic functions, the C1-controlled operators or the Lpa-activated promoters, which will be described later in this paper.

Of 45 P1 operons, 14 are not directly controlled by C1 or Lpa (Fig. 1). Their expression is probably not up-regulated during lytic growth, other than by gene dosage, an effect that is damped in those operons that are subject to autoregulation (ImmC and ImmI genes, *repA*, *parAB*, *phd doc pdcAB*, *cin*, and probably others).

As many as 20 genes of operons that appear to be independent of C1 and Lpa are either known or predicted to be associated with lysogeny (Fig. 1 and Table 3; see also Table S1 in the supplemental material). We are unable to predict the function of 12 other genes in this class (*isaB*, *upfA*, *mlp*, *ppfA*, *upfB*, *upfC*, *plp*, *upl*, *upfM, upfN*, *upfO*, and *upfQ*). However, three of them, *mlp*, *ppfA* and *plp*, encode products that have putative periplasmic transport or lipoprotein attachment signals likely to contribute to lysogenic conversion.

Only six structural genes are essential for stable inheritance of P1 as a prophage. Other genes associated with lysogeny may protect the lysogen from entry of foreign DNA (*res* and *mod*; reviewed in references 32 and 38), from infection by another P1 or homologous phages (*sim*, superinfection exclusion) (170), and from DNA damage (*humD*, homolog of *umuD*) (217). Possibly for the *sim* product(s) to be effective, the gene(s) must be amplified, as superimmunity was observed only with high-copy-number clones of *sim*. The C-segment inversion recombinase *cin* is expressed by the prophage, but the known benefit it confers is to produce phages that differ in host specificity from their parent.

TABLE 6. C1-controlled operators of P1

Operator[a] (former name[s])	Position[b]	Strand	DNA sequence[c]	Genes controlled	Evidence for:		Reference(s)
					C1 binding	Regulatory function	
O*ref* (Op1, Op2a)	1962	+	ATTGCTCTAATTGATTC	*ref*	√	√	84, 337
O*mat* (Op2, Op2b)	2628	+	CATGCACTAATAAATAT	*mat*	√	√	84
O*ssb* (Op21)	19816	+	ATTGCTCTAATTAATTT	*ssb*, *isaA*	√	√	59
O*c4* (Op51)	46941	+	AATGCTCTAATAAATTT	*c4*, *icd*, *ant*	√	√	26, 84, 124, 316
O*kilA* (Op53)	48611	+	TTTGCTCTAATAAATTT	*kilA*, *repL*	√		27, 84, 316
O*rlf* (Op55)	50376	+	ATTGCTCTAATTATAAC	*rlfA*, *rlfB*	√		112; reviewed in reference 346
O*dmt* (Op68)	63946	+	ATTGCTCTAATAAATAT	*uhr*, *hrdC*, *dmt*	√	√	59
O*ban*ab (Op72, Op72ab)	67468	+/−		*trnA*, *trnT*, *tciA*, *tciB*, *tciC*, *trnI*, *ban*, *dbn*	√	√	84, 129, 315, 317
a	67468	+	ATTGCTCTAATAAATTT		√		132, 316
b	67479	−	ATTACACTAATAAATTT		√		
O*26* (Op82)	77329	−	ATTGCACTAATAATTTT	*26*, *25*, *7*, *24*, *6*, *5*		√	Reviewed in 346
O*ppp* (Op83ab)	78236	−	CATGCACTAATTAATTT	*pmgN*, *pmgO*, *pmgP*, *ppp*, *pmgQ*		√	Reviewed in reference 346
O*pmgR*a (Op86a)	80953	+	ATTGCTCTAATAATTCC	*pmgR*, *pmgS*	√		315, 346
O*pmgR*b (Op86, Op86b)	81002	−	ATTGCTCTAATAAAAAA	*pmgR*, *pmgS*	√		84, 316
O*pmgT* (Op88)	83094	+	ATTGCTCTAATAAATCT	*pmgT*, *pmgU*, *pmgV*	√	√	59
O*lpa* (Op94)	88705	+	ATTGCTCTAATGTATTG	*lpa*	√	√	193
O*c*lab$^{\Lambda}$ (Op99ab, Op99a)	92967	−/+		*c*1	√	√	27, 84, 128, 316, 317
a	92978	−	AATGCACTAATAAATCT		√		84, 316
b[d]	92967	+	ACGAAAATAATAGATTT		√		317
O*c*1c$^{\Lambda}$ (Op99c)	93150	−	ATTGCTCTAACGCTTTA	*c*1	√	√	84, 128
O*coi*$^{\Lambda}$ (Op99d)	93442	−	ATTGCTCTAATAAATTA	*c*1, *coi*	√	√	27, 128, 316
O*cre*$^{\Lambda}$ (Op99e)	93525	+	ATTGCTCTAATAATTCT	*cra*, *cre*	√		27, 128, 316
Consensus			**ATTGCTCTAATAAATTT**				

[a] The name of each operator refers to the name of a selected gene under its control. Bivalent operators names are underlined; constituent monovalent operators are listed below the name of each bivalent operator. Operators with a superscript Λ have been shown to engage in looping to nearby operators in vitro.

[b] Position of the first nucleotide if the operator is assigned to the + strand or to the first nucleotide of its complement if the operator is assigned to the − strand. Coordinates refer to positions of operators in the P1 *c*1-100 *mod749*::IS*5* genome without its nonintegral parts, IS*5*, and the associated 4-bp duplication.

[c] The underlined nucleotides are conserved in monovalent operators and in at least one constituent operator of each bivalent operator.

[d] The sequence of O*c*1b corresponds to that identified by DNase protection experiments (317) and differs from the sequence proposed earlier (84).

(i) Immunity. Whether P1 is maintained in cells as a plasmid or enters the lytic pathway is dictated by the interplay of environmental factors with the components of the immunity circuit encoded at loci designated ImmC, ImmI, and ImmT (Fig. 1 of the accompanying guest commentary [343a]; reviewed in reference 126). Genes of the immunity loci were sequenced and characterized prior to the present work. Analysis of the entire P1 genome allows us to extend this characterization.

(a) ImmC. The key repressor of P1 lytic functions is the C1 protein, encoded by *c*1 within ImmC (84, 238). Inactivation of the C1 protein or a decrease in C1 synthesis triggers lytic growth of the phage. A commonly used allele of *c*1, *c*1-100, encodes a thermosensitive C1 protein (261). In both sequenced strains, *c*1-100 contained two mutations, T569C and G577T, as compared to what has been taken as wild-type *c*1 (238). The second of these mutations, which causes the amino acid substitution G193C in C1, probably confers thermosensitivity to the C1-100 protein. The thermolabile C1-100 protein of P1 and the almost identical thermostable C1 protein of P7 do not differ at the locus of the T569C substitution (238).

C1 protein exists as a monomer in solution and binds to a score of widely dispersed operators (27, 59, 84, 193). The 17-bp operator consensus sequence is asymmetric and hence has directionality. Monovalent operators have a single repressor-binding site, whereas bivalent operators consist of two overlapping repressor-binding sites oriented in opposite directions and forming an incomplete palindrome (129) (Table 6).

Transcription of the *c*1 gene itself is autorepressed at operators that precede *c*1 and are part of ImmC (317, 128). One of these operators, O*coi*, additionally represses transcription of three small ORFs preceding *c*1 (27, 127). The product of the longest ORF, the 7.7-kDa Coi protein, forms a 1:1 complex with C1 and blocks its ability to bind to operators (25, 127, 130). A combination of negative effects of C1 on the synthesis of Coi and of Coi on the activity of C1 creates a sensitive regulatory system that is crucial for a choice between lysis and lysogeny. If C1 synthesis prevails in P1-infected cells, Coi synthesis is shut down, leading to the establishment of lysogeny. If Coi synthesis prevails, C1 becomes inactivated, leading to lytic growth.

The *E. coli* SOS response regulator LexA may be a major factor in this epigenetic switch. A LexA binding site identified previously in the P1 genome in the region upstream of the *imcAB* and *coi* genes (200) overlaps a predicted strong promoter that we designated P2*coi* (Tables 4 and 7; reference 343a, Fig. 1). Most likely, this promoter can drive transcription of *imcAB* and *coi* following inactivation of LexA. In support of this view, inducibility of P1 lysogens by UV light has been repeatedly reported, although not always reproduced (reviewed in reference 347).

TABLE 7. Known and potential binding sites in P1 DNA for *E. coli* proteins IHF, DnaA, and LexA

Coordinates[a]	Strand	Sequence	Similarity score[b]	Known or potential function or relative location (reference[s])
IHF				
1478–1526	−	TATGTCACCATAAATATCAAATAATTATA	52.0	*c*8 regulation
6985–7132	−	TTATATGCATACAAAAAGATGAAGTTATA	53.9	Internal to *mod*
19809–19857	−	ATCTTACCTAATAAAATAAATTAATTAGA	55.7	*ssb* regulation, *lyz* regulation
21424–21472	−	TGAACATAAAACACTATCAATAAGTTTGA	50.1	Overlaps terminal inverted repeat of defective IS*1* (93, 94)
30319–30367	−	TTTAATTCACTATTAAGGAAATAGTTATG	53.7	*cin* regulation
38415–38463	−	TATAATGTTCAGAATATCAATAAGATATT	51.8	*16* regulation
38419–38466	+	AATATCTTATTGATATTCTGAACATTATA	51.4	*16* regulation
46372–46419	+	ATTAATCGGTTTAAACTCACTAATTAATT	51.4	*sim* regulation
46498–46545	−	TGGTTGATTCTTATAATCAAAAAACTATT	50.3	*sim* regulation
49930–49977	−	AAAAAATTAAATAAAGCCAATGTCTTAGC	54.2	Internal to *repL*
50647–50694	+	AAAATTGCTAGAGACATCAAGGAATTAAA	57.0	*rlfB* regulation or part of alternative replication origin
50672–50719	+	TAAAATTAAAACACAATCTAAAACCTGAT	50.5	*rlfB* regulation or part of alternative replication origin
51151–51198	+	ATAAAATCACATACAAACAGGGAGTTACT	58.7	*rlfB* regulation or part of alternative replication origin
57740–57787	+	GTAATTTACTTTAAAAACAGTCAGTTAAT	57.2	*parS*-ParB-IHF centromeric complex formation (91)
61404–61451	−	GAAAAAGTAATATGAATCAATCATTTATC	57.3	*repA* regulation (reviewed in reference 49)
89562–89610	+	CTGAAATAGCGGAAAAA**C**AAAGAGTTAAT[c]	40.8	Involved in PacA- and PacB-mediated cleavage at *pac* site (287)
DnaA				
1076–1090	+	TAGTTACCCCCAGGC	54.3	Internal to *cre*
5894–5908	+	GTGTTAATAACAAGA	56.0	Internal to *res*
5892–5906	−	TTGTTATTAACACGG	55.1	Internal to *res*
6716–6730	+	GTGTTGTTAACAAGC	54.3	Internal to *mod*
9564–9578	−	GAGTTATCATCAATA	57.7	Internal to *darB*
41563–41577	+	AGGTTATCGACAGTA	55.2	Internal to *sit*
47322–47336	−	GTGTTTTTTACAGGA	54.3	Internal to *icd*
47945–47959	+	GAATTTTACACAAAA	54.3	Internal to *ant1/2*
49965–49979	+	TTTTTATCCACACCC	56.9	Internal to *repL*, presumably part of *oriL*
50148–50162	+	CGGTTATCCAAAAAA	56.0	Internal to *repL*, presumably part of *oriL*
50313–50327	+	GTTTTAGCCAAAAAT	54.4	Internal to *repL*, presumably part of *oriL*
50986–51000	+	GGGTTCTTCACAATG	54.3	Internal to *rlfA*, overlaps predicted promoter of *rlfB*, presumably regulates expression of *rlfB*
52202–52216	+	GATTTAAGCACAAAA	56.9	Internal to *pmgF*
61448–61462	−	AAACTATCCACACAA	51.8	Replication initiator complex formation at *oriR* (reviewed in reference 51)
61457–61471	−	CACTTATAAAAACTA	<50.0	Replication initiator complex formation at *oriR* (reviewed in reference 51)
61467–61481	−	TAGTTATCACCATTA	51.9	Replication initiator complex formation at *oriR* (reviewed in reference 51)
61690–61703	−	ACATTATCCACTGGA	51.8	Replication initiator complex formation at *oriR* (reviewed in reference 51)
61698–61712	−	CACTTATCCACATTA	56.0	Replication initiator complex formation at *oriR* (reviewed in reference 51)
68953–68967	−	TGGTTATCCACCGTG	54.3	Between *upl* and its predicted promoter; regulation of *upl*
72890–72904	−	CGGTTATGCACAGGA	56.9	Internal to *5*
73755–73769	+	GGCTTATCAACAAAA	56.9	Internal to *6*
80593–80607	−	TTGGTATCCACAGGG	57.7	Internal to *pmgQ*
85215–85229	−	AAGTTATGCACTTGC	<50.0	*hot* regulation
85227–85241	−	GATTTATCCACAAAG	63.5	*hot* regulation
85851–85865	−	GAATTTTCCTCAAGG	55.1	Vicinity of P2*humD* promoter internal to *upfQ;* probable *humD* regulation
89848–89862	−	GACTTTTCCAGAAGA	56.0	Vicinity of *pac* site internal to *pacA*, packaging regulation
LexA				
46–65	+/−	ACCTGTAAATCCATACAGTT	62.7/59.8	Regulation of ImmC genes
86346–86365	+/−	TACTGTATAAATGAACAGTA	67.7/66.1	*humD* regulation and possibly *lxr* regulation

[a] The first number is the coordinate of the first nucleotide of upper-strand (+) binding sites or of the last nucleotide of the lower-strand (−) binding sites, based on arbitrary convention. Coordinates refer to positions of sites in the P1 *c*1-100 *mod749*::IS*5* genome without its nonintegral parts, IS*5*, and the associated 4-bp duplication.

[b] Of predicted IHF, DnaA, or LexA binding sites, only those whose similarity score to the consensus exceeded 50.0, 54.0, or 60.0, respectively (as calculated by the GCG Fitconsensus program) are shown. Exceptions are made for tandem sites.

[c] For PP1PAC in GenBank, boldface C is a G.

An additional component of ImmC or a separate immunity locus may be located downstream of *cre*. This gene, tentatively named *c*8, is probably the locus (*c*8) of a clear plaque mutation (280, 281). Temperature shift experiments with a thermosensitive allele of *c*8 suggested that the gene is probably not expressed in a lysogen and hence is most likely involved uniquely in the establishment of lysogeny.

(b) ImmI. The ImmI region (antipodal to *immC* in the circular representation of the plasmid genome) contains a C1-controlled operon of three genes: *c*4, *icd*, and *ant1/2*, which confer a separate specificity to immunity. It permits P1 to plate on lysogens of the closely related phage P7, and vice versa, despite functional interchangeability of their C1 repressors (54, 131, 280, 326). The *ant1/2* and *icd* genes determine an antirepressor protein and an inhibitor of cell division, respectively, whereas the *c*4 gene determines an antisense RNA that acts as a secondary repressor of Icd and antirepressor synthesis (58, 131, 257). Maintenance of lysogeny requires, in addition to the expression of *c*1, the expression of *c*4 to prevent synthesis of the antirepressor protein (reviewed in reference 126). A 77-bp antisense RNA encoded by *c*4 regulates expression of downstream genes in its own message (57; diagrammed in Fig. 1 of reference 343a). Two short single-stranded regions, b′ and a′, exposed in its cloverleaf-like structure, block translation of *icd* (and of translationally coupled *ant1*) by interaction with the complementary regions, a2 and b2, in the *c*4 *icd ant1/2* mRNA. The interaction occludes the RBS in front of the *icd* gene (58). The resulting translational block additionally permits premature termination of *ant* transcription via a Rho-dependent terminator (33). Slight differences between P1 and P7 in these short regions are responsible for the heteroimmunity of the two phages. An additional specificity determinant, *sas* (site of Ant specificity), is involved in determining the capacity of Ant proteins to induce prophage on heteroimmune superinfection. The site resides centrally within the *ant* genes of P1 and P7 (J. Heinrich and H. Schuster, unpublished results). It was suggested that Ant proteins must normally load onto their phage-specific *sas* sites to perform their antirepressor function (294).

The action of C4 appears limited to the mRNA of its own operon, as we did not find any pair of sequences complementary to b′ and a′ in the P1 genome, other than those designated a1 and b1, located upstream of *c*4, and presumed to participate in unmasking of a2 and b2 by competitive binding of b′ and a′.

(c) ImmT. The ImmT region, distant from ImmC and ImmI, contains a small gene, *lxc*, whose product modulates the function of C1 by formation of a ternary complex with C1 and operator DNA (272, 317, 318). Participation of Lxc in the C1-operator complexes that control *c*1 expression increases the affinity of the repressor for the operator and down-regulates the synthesis of C1 itself (318). Lxc can thus relieve repression at weak operators by decreasing the intracellular concentration of C1 below that critical for their binding and simultaneously enhance repression at strong operators by increasing their affinity for C1. Consequently, *lxc* mutations can have divergent effects on transcription from different C1-controlled promoters (272, 273, 311). Part of the modulating effect of Lxc may be mediated by looping between nearby operators complexed with C1, as looping is dramatically increased by Lxc (128, 343a). As Lxc has been reported to strongly inhibit the ability of Coi to dissociate operator-C1 complexes (317), it may assist in establishing P1 as a prophage.

The initiation codon of *lxc* appears to overlap with the termination codon of the preceding *ulx* gene of unknown function, indicating transcriptional and translational coupling of the two genes (Table 3). The coupling of *lxc* and *ulx* implies that *lxc*, in addition to being transcribed from its own promoters, may be transcribed from P*darB*, which presumably drives transcription of *ulx* and the preceding gene, *darB* (Table 4). None of the promoters that might drive transcription of *lxc* is under the control of C1, suggesting that Lxc-mediated modulation of C1 function is independent of ImmC and ImmI.

(ii) Plasmid maintenance. Despite its low copy number per bacterial origin (0.7 to 1.4) (248), the P1 plasmid is lost with a frequency of only about 10^{-5} per cell per generation (261). Like other plasmids that have accepted the tradeoff of large size for low copy number, P1 must counteract the increased risk of loss at cell division. Four functions have been identified that allow it to accomplish this task: a regulated plasmid replicon, a partitioning mechanism, a site-specific recombination system, and a plasmid addiction (postsegregational killing or growth inhibition) mechanism. These are discussed below. Analysis of the genome of P1 indicates that whereas the lytic origin may contribute to plasmid replication (349), there is no second addiction or partition module in P1 as there are in certain other plasmids (80, 140).

(a) Plasmid replication and partition. The P1 plasmid replication and partition genes are encoded within one region of P1 DNA, where they form two operons transcribed in the same direction. The P1 replicon has been intensively studied (reviewed in references 49 and 51). It belongs to the large family of plasmid replicons that encode an initiator protein, called RepA in P1, and contain multiple binding sites for that initiator (19-bp iterons in P1). One set of these sites, *incC*, precedes the *repA* gene and is part of the replication origin (*oriR*). A second set, *incA*, follows *repA* and is a regulatory locus (Table 1). Replication from the P1 plasmid origin, *oriR*, proceeds in both directions (241). P1 plasmid and *E. coli* chromosome replication have much in common, including sequestration of the origin when its 5′-GATC sequences are hemimethylated (1) and a requirement for DnaA (Table 7) (113). One difference is that the ADP form of DnaA suffices for P1 replication, whereas the ATP form is essential for bacterial replication (333).

The P1 *repA* promoter is nested among the *incC* iterons and consequently is subject to autorepression by RepA (5). Replication control is exerted by *incC* and the additional iterons of *incA* in more than one way. By sequestering RepA, the *incA* iterons can prevent the concentration of RepA at the origin from attaining the threshold value required for firing (reviewed in reference 49). By becoming "handcuffed" to the iterons of *incC* via bound RepA, the additional iterons can interfere with both replisome assembly (239) and the release of autorepression that would otherwise replenish sequestered RepA (50). The availability of RepA for replication is limited not only by sequestration and handcuffing but also by the formation of inactive RepA dimers which require bacterial chaperones for conversion to active protein monomers (334). We suggest an additional role of *incA* in the control of *repA* expression. The putative transcriptional terminator of *repA* is located within the

incA region (Table 5), and thus occupation of *incA* iterons by RepA may act as a roadblock to the completion of *repA* transcripts. Abbreviated *repA* transcripts might undergo more rapid degradation than intact transcripts.

The partition module of P1, which is situated downstream of the *repA* gene and the *incA* iterons, is a major contributor to stability of the plasmid. It ensures partition of plasmid molecules to daughter cells and has been, with its homologs, the subject of numerous detailed studies and reviews (reviewed most recently in references 97 and 306). The module consists of an operon of two genes, *parA* and *parB*, which ends at a site, *parS*, a centromere analog (4). ParA is an ATPase whose activity is stimulated by ParB (73). Both ATP and ADP forms of ParA act as autorepressors, the ADP form being the more effective (70, 71). A stimulation of repression by ParB (87) is assisted by *parS* (114). The *parS* site contains two kinds of recognition sequences for ParB, four heptameric and two hexameric boxes, and a binding site for the host protein IHF (Tables 1 and 7) (249). ParB and IHF form a complex at *parS* (72, 89) to which ParA, in the ATP form, can bind via ParB (37, 90). Binding of ParB to *parS* can permit *parS* sites to pair (83) and can nucleate spreading of ParB to the flanking DNA and silence transcription of genes as far as several kilobases away (258). Whether ParA and ParB translocate plasmid molecules to their target positions within a cell during partitioning by themselves or attach them to unknown host components of partition machinery is unclear.

We did not find in P1 DNA any additional sequences similar to *parS* that could serve as potential binding sites for ParB. This suggests that the function of ParB is limited to partitioning in P1, whereas in certain other plasmids, e.g., RK2 and N15, the homologous proteins appear to have additional functions (102, 251, 335). Two putative σ^{70} promoters of P1 gene *23*, immediately downstream of *parB*, of which one overlaps the IHF binding region of *parS*, appeared weak when tested in vivo in the absence of ParB (259; A. Dobruk and M. Łobocka, unpublished results). The possibility that conditions might exist leading to their significant expression and regulation by ParB has not been studied.

(b) DNA cyclization and multimer resolution. Faithful partitioning of newly replicated plasmid molecules is assisted by prior resolution to monomers of plasmid multimers that are inevitably formed in *recA*$^+$ bacteria. The resolution is usually accomplished by a site-specific recombination. In P1 it is ensured by the recombinase Cre, encoded near its site of action, *lox* (8, 299, 300), far from the genes of replication and partition (Fig. 1). The resolution of plasmid multimers into monomers increases the number of partitionable P1 molecules and hence plasmid stability (21). The same accessory proteins that are required for the resolution of ColE1 dimers by the bacterial recombinase XerCD are also required for the stable maintenance of P1 prophage (244). These proteins (ArgR, for which a putative binding site to the left of *lox* has been located, and PepA) are proposed to constrain the directionality of the recombination in vivo. The possibility that the *lox-cre* module might assist plasmid stability in another way has also been suggested (10). Interplasmid recombination involving a replicating plasmid could generate a concatemeric replication substrate that, following amplification as a rolling circle, could be resolved into plasmid monomers by Cre-mediated recombination. The plasmid yield per replication initiation event would be increased, providing protection against RepA insufficiency.

The *lox* site is the unique target for Cre action in P1 DNA, as we did not find any other similar sites. In the P1 genome, the *cre* gene is downstream of *cra*, a gene of unknown function. In the prophage, *cre* is weakly expressed, either from promoters P1*cre* and P2*cre*, which lie within *cra*, or together with *cra* from a third promoter, P3*cre* (302). The −35 regions of P3*cre* and P2*cre* promoters overlap GATC sequences (Table 4), and P3*cre* was shown to be down-regulated in *E. coli* by Dam methylation, as P2*cre* probably also is. Whether methylation of one strand is sufficient for the repression is unknown. If this was not so, the expression of *cre* in a prophage would increase immediately following each round of replication and be attenuated later as a result of methylation of the newly replicated strand in the region of P3*cre* and P2*cre* promoters.

A putative fourth *cre* promoter, P4*cre*, predicted to be the strongest, lies almost 500 bp upstream of *cre* and on the opposite side of *lox* (Table 4). Its −35 region is partially overlapped by the C1 operator, O*cre* (Table 6), indicating that the promoter is inactive in the P1 prophage. Apparently, the *cre* gene is expressed most efficiently at the time of infection and during P1 lytic development. Following infection, the Cre recombinase can cyclize the DNA of phages that carry a *lox* site on each redundant end (302). The first phages to be packaged belong to this privileged minority and hence are to be found in every productive burst. The other phage DNAs must rely on homologous recombination to cyclize.

(c) Plasmid addiction. A fail-safe P1 stabilization mechanism is encoded by the addiction operon, *phd doc*, located in a region far from other plasmid maintenance genes (Fig. 1). The *phd doc* operon programs the inhibition of growth (and the eventual death) of any daughter cells emerging plasmid free. Two small genes of this operon encode a stable protein toxin, Doc, and an unstable cognate antidote, Phd (95, 96, 192, 211). Doc appears to block protein synthesis reversibly (R. Magnuson and M. Yarmolinsky, unpublished results). Phd and Doc can form a complex. This interaction prevents Doc from killing plasmid-containing cells and enhances the Phd-mediated repression of the autoregulated *phd doc* promoter (210, 211). Phd is slowly degraded by the host protease ClpXP (194), allowing for the release of the toxin protein upon plasmid loss. Recently documented connections among prokaryotic toxin-antitoxin systems, Phd-Doc among them, and their relationship to the eukaryotic nonsense-mediated RNA decay system places Phd-Doc in a grander evolutionary context than anticipated (13).

The nucleotide sequence downstream of *doc* contains two ORFs that could encode proteins of 66 and 347 residues, designated here *pdcA* and *pdcB* (post-doc). The 5′ end of *pdcA* overlaps 17 terminal nucleotides of the *doc* gene. The *phd-doc* transcript appears to include the transcripts of *pdcA* and *pdcB*, as suggested by the lack of predicted Rho-independent terminators of the *phd doc* operon within *pdcA* or *pdcB*, and the lack of a separate recognizable σ^{70} promoter sequence that could drive their transcription. Whether the functions of *pdcA* and *pdcB* are accessory or unrelated to the function of *phd doc* remains to be seen.

(iii) Restriction-modification. P1 *mod* and *res* are tandem genes that encode the subunits of a type III restriction-modi-

fication enzyme, EcoPI. This bifunctional enzyme has been the subject of extensive studies (reviewed in references 32 and 38; see also reference 161).

The P1 Mod subunit recognizes the DNA sequence 5′-AGACC and catalyzes methylation of the central adenine residue at the N-6 position (23), using adenosyl methionine as the methyl donor. The Res subunit catalyzes the double-strand cleavage of DNA about 25 bp to the 3′ side of a recognition sequence, but it does so only when bound to the Mod subunit in a Res_2Mod_2 complex (161).

Whereas methylation by Mod can occur at any recognition sequence, each double-strand scission by Res-Mod requires a pair of unmethylated recognition sequences, in head-to-head configuration (32, 219). The cut appears to be triggered by collision between two converging Res-Mod enzymes that, powered by ATP hydrolysis, have been translocated along the DNA (220). The requirement for collision, in combination with the asymmetry of enzyme recognition sites, provides an efficient protection from cleavage of newly replicated DNA in which only one strand is unmethylated, because the unmodified sites in one orientation are always paired with modified sites in the opposite orientation.

Host killing by restriction is avoided during the process of establishing a P1 prophage in a bacterial cell with an unmodified chromosome. This is achieved by complex regulatory mechanisms that delay restriction long enough to allow complete methylation of host DNA (253). In P1 lysogens, transcripts containing either *mod* or *res* messages are detected. This indicates that *mod* and *res* form separate operons, although the beginning of *res* immediately follows the end of *mod* (285). The σ^{70} promoters for *res*, separate from those for *mod*, have been identified previously or are predicted here (Table 4). Additional regulation occurs at the translational and posttranslational levels (253).

(iv) Superinfection exclusion. During attempts to clone the *c*1 gene of P1, Devlin et al. isolated a fragment of DNA that, although not carrying *c*1, protected cells from superinfection by wild-type P1, its *c*1 and *vir*^{s} mutants, and the heteroimmune P7 phage (77). This extended immunity phenotype, designated superimmunity and subsequently attributed to superinfection exclusion, was associated with a gene upstream of the *c*4 *icd ant* operon and transcribed in the opposite direction (170, 213). We find that this gene, designated previously *sim* and here *simC*, is the last of three genes of the *simABC* operon. The *sim* genes appear to encode precursors of periplasmic proteins, suggesting that their functions are related. The location of putative promoter sequences in the region preceding *simC* indicates that *simC* is cotranscribed with *simB*, or with *simA* and *simB* (Table 4). A plasmid carrying the *sim* operon was found by minicell analysis to specify a processed, as well as full-length, form of the SimC protein (213). Two additional proteins apparently expressed from this clone were seen following sodium dodecyl sulfate (SDS)-PAGE (170). One is probably SimB (12.0 kDa). The other migrated as a very small protein and, although not mentioned by the authors, is probably SimA (predicted molecular weight, 4.8 kDa).

The superimmunity (or superinfection exclusion) phenotype was observed in the presence of SimC alone (213). The SimC protein appears to be an analog of superinfection exclusion proteins that act to prevent injection of superinfecting phage DNA into the cytoplasm of the infected cell. Many phages, both temperate and virulent, encode such proteins. SimC appears to act in the periplasmic space, where the processed form is found, or in the cytoplasmic membrane, similarly to the SieA protein of P22 (139) and the Imm protein of T4 (205), either by helping to destroy the injected DNA or by preventing its entry into the cytoplasm (139, 170, 206, 213). The processing of SimC requires SecA (213), as probably does the processing of SimA and SimB, which, like SimC, have putative signal peptides.

Products of the *sim* operon, like the product of *sieA* of P22 (139, 307), exclude both phage and transducing particle DNA. However, exclusion of the latter by Sim is much less efficient, indicating that the Sim system can discriminate between P1 and foreign DNA. Conceivably, SimC or another Sim protein can interact with a P1 protein that is bound specifically with phage DNA as it enters a cell during infection.

Cells carrying low-copy-number *sim*$^+$ plasmids display a Sim$^-$ phenotype (77, 170), which appears inconsistent with a role of the *sim* genes in lysogeny. It is possible that in P1 lysogens, the *sim* functions are induced only under certain circumstances or that the main role of Sim proteins is to protect from superinfection those cells in which lytic development has been initiated and the *sim* gene dosage is high. The physiological role of *sim* functions could be analogous to that of superinfection exclusion functions encoded by the *immT* and *sp* genes of the lytic phage T4. It has been proposed that superinfection exclusion mechanisms may preserve genetic diversity among phages, protecting phage populations from dominance by a phage that would otherwise outcompete any related phage when given a possibility to propagate in a superinfected cell (205).

(v) HumD and Lxr. The monocistronic operon immediately upstream of the P1 addiction module encodes HumD protein, a homolog of *E. coli* UmuD′ (200). The P1 *humD* gene, like *E. coli umuD*, is transcribed from a LexA-regulated promoter (Tables 4 and 7). P1 lacks a homolog of *umuC*, which in *E. coli* is functionally associated and cotranscribed with *umuD*.

Aside from *recA*, *umuD* and *umuC* of *E. coli* are the only LexA-regulated genes of the SOS response that are required for DNA damage-induced mutagenesis (290). They encode a low-fidelity and low-processivity DNA polymerase, PolV, whose main function is to bypass DNA lesions (309). A RecA-mediated autocleavage of the *umuD* gene product (215) is required to form active PolV, which is a complex of UmuD′ and UmuC in a 2:1 ratio (43). P1 HumD corresponds to the processed form of UmuD, UmuD′, and can functionally replace it (217).

The regulatory region of *humD* overlaps the regulatory region of a divergently transcribed gene of unknown function. This gene could encode a 190-residue, slightly acidic protein that has no homologs in databases. We surmise that its transcription, like that of *humD*, is negatively regulated by LexA. One of two putative σ^{70} promoters of this gene overlaps the LexA binding site that controls *humD* (Tables 4 and 7), and thus we name the gene *lxr* (LexA regulated).

Genes expressed in lytic development. Expression of the majority of P1 genes during lytic growth follows a strict temporal pattern that can lead to the production of mature phage particles within less than 1 h. For many bacteriophages such as T4 (180), Mu (18, 55), P2 (339), and P4 (98), the regulatory

cascade has been subdivided into early, middle, and late stages. In P1, the cascade appears to be simpler; early transcription switches directly to late transcription, without a well-defined intermediate stage (188, 193). C1 operator sequences, which bind the primary phage repressor C1, act as a switch of early genes. Lpa (late promoter activator) binding sequences, which are in the −22 region of late promoters and enable RNA polymerase to initiate transcription upon binding of Lpa, act as a switch of late genes.

(i) C1-controlled operons. Analysis of the entire P1 genome has revealed 17 operons that have C1 operators in the region of their σ^{70} promoter sequences or, in some cases, between promoters and proximal genes (Table 6). One of these operons (*pmgR*) is preceded by two nonoverlapping operators, and two others (*ban* and *c*1) are preceded by operators that are bivalent, bringing the total number of monovalent C1-binding sites to 20, all of which were identified previously. Although we did not find any additional sequences with strong similarity to known C1-binding sites, sequences that resemble the C1-binding sites but are missing the highly conserved C at position 7, are present in the *parS* site and in the promoter region close to the *darB* gene. The sequence in *parS* did not interact with C1 in vitro (N. Sternberg, personal communication). Whether these sequences can act cooperatively with other C1 operator sequences to bind C1 in vivo, as does the O*c*1b operator, which also has the C at position 7 replaced by another base, remains to be seen. Looping between C1-bound operators located close to each other has been demonstrated (128), but looping between well-separated operator sequences has not.

The 17 C1-controlled operons, transcribed from σ^{70} promoters, contain 49 genes that are derepressed early in lytic development. Transcription does not start simultaneously, since the operators exhibit different affinities for C1 and are differently influenced by the Lxc protein, which modulates the C1-operator interaction. Lytic replication genes start to be transcribed within 5 min of infection, allowing a prompt initiation of phage DNA replication. Increased amounts of P1 DNA can be detected about 15 min after infection, at about the same time as cleavage of the packaging site, *pac*, in P1 DNA (297). Detection of DNA synthesis within 5 min has also been reported (282). Expression of other early functions was reported to start 10 to 15 min after induction of P1 lysogens (187). The intense transcription of P1 early functions is attenuated later in lytic development, prior to transcriptional activation of late genes (111). The attenuation depends on the *E. coli* RNA polymerase-associated protein SspA. Whether SspA acts directly on complexes of RNA polymerase with P1 σ^{70} promoter sequences or requires the action of a P1 protein is unknown. We did not find any conserved sequences, other than promoter sequences and C1-binding sites, in the regulatory regions of P1 early genes.

(a) P1-encoded tRNAs. Several phages encode their own tRNAs (e.g., see references 88, 177, and 247); the T4-like vibriophage KVP40 encodes at least 25 tRNAs at a single locus (223). Some of these tRNAs supplement the host pool of rare tRNAs to facilitate efficient expression of selected phage genes during lysogeny or at certain stages of lytic development (179, 247). Additionally, it has been proposed that phage tRNAs can provide selective pressure to keep base composition and codon usage of phage DNA significantly different from those of its host (summarized in reference 227). We find that the P1 genome contains three sequences characteristic of tRNA genes. All three are located between the *ban* gene and its promoter, in the apparently untranslated regions of the *ban* operon. Two are adjacent; one is further downstream, separated from the two by three protein-coding genes. Control of the P1 tRNA genes from a promoter repressed by C1 implies that they are not expressed during lysogeny.

The proposed cloverleaf structures of all predicted P1 tRNAs are shown in Fig. 3. The first tRNA, tRNA1, has the anticodon GUU, characteristic of *E. coli* $tRNA^{Asn}$. It is 90% identical to the predicted $tRNA^{Asn}$ of the *Salmonella enterica* serovar Typhi plasmid pHCM2 (242) (GenBank accession no. AL513384) and contains all bases essential for the aminoacylation of $tRNA^{Asn}$ in *E. coli*: the anticodon bases G34, U35, and U36 and the discriminator base G73 (201).

The second tRNA, tRNA2, has the anticodon UGU, characteristic of *E. coli* $tRNA^{Thr}$. Although it is highly homologous to alanine tRNA-UGC of a cyanobacterium, *Synechocystis* sp. (SYCSLRB; 83% identity), and to $tRNA^{Ala}$ of numerous other organisms, it contains neither the G3-U70 wobble pair, which is the primary determinant of the acceptor identity of the *E. coli* $tRNA^{Ala}$, nor the discriminator base G20, which is also an important characteristic of this tRNA (250, 308). Instead, in addition to the anticodon bases G35 and U36, which are identity determinants of all *E. coli* $tRNA^{Thr}$s, it contains base pairs G1-C72 and C2-G71, which in *E. coli* $tRNA^{Thr}$, are crucial for threonine charging activity (116).

The third tRNA, tRNA3, has the anticodon CAU, which should correspond to the Met codon AUG. However, in its sequence this tRNA appears to be 89% identical to the *E. coli* $tRNA^{Ile}$-2, encoded by the *ileX* gene. None of the differences (outside the anticodon) between tRNA3 and $tRNA^{Ile}$-2 affect residues known to be essential for the function of *E. coli* $tRNA^{Ile}$-2. In the $tRNA^{Ile}$-2 of *E. coli* the wobble position C34 is modified by lysinylation to lysidine (i.e., 2-lysylcytidine). The modified anticodon acquires the AUA (Ile)-decoding capacity and is required for the recognition of this tRNA by isoleucyl-tRNA synthetase (231, 232, 283). The other identifying characteristics of *E. coli* $tRNA^{Ile}$ are the anticodon loop bases A37 and A38, the discriminator base A73, and the base pairs C4-G69, U12-A23, and C29-G41 (236, 240). All of them are present in P1 tRNA3. Certain other phages, including 933W (247), T4 (224), and the T4-like phages RB69 (see http://phage.bioc.tulane.edu/) and KVP40 (223) also encode homologs of *E. coli* $tRNA^{Ile}$-2, predicted to contain the lysylcytidine modification at C34. The T4 homolog of *E. coli* $tRNA^{Ile}$-2 was confirmed to have the capacity to decode ATA (274).

Of the codons that are apparently recognized by P1 tRNAs, two, ACA and ATA, are rare in *E. coli* genes but overrepresented in certain P1 genes (Fig. 3). The ACA codon is the rarest threonine codon in *E. coli*, but its effect on translation has not been studied. The ATA codon, which is the fifth rarest codon in *E. coli*, is known to dramatically decrease translation of those *E. coli* mRNAs that contain it, especially when present in multiple copies, or in tandem, or in a single copy in the anterior part of a gene (99, 166, 353). The ATA codons are overrepresented more than twofold in 59 of 113 protein-coding genes of P1 as compared to their representation in *E. coli* genes. Four of these 59 genes (*isaB*, *rlfA*, *7*, and *pacB*) contain

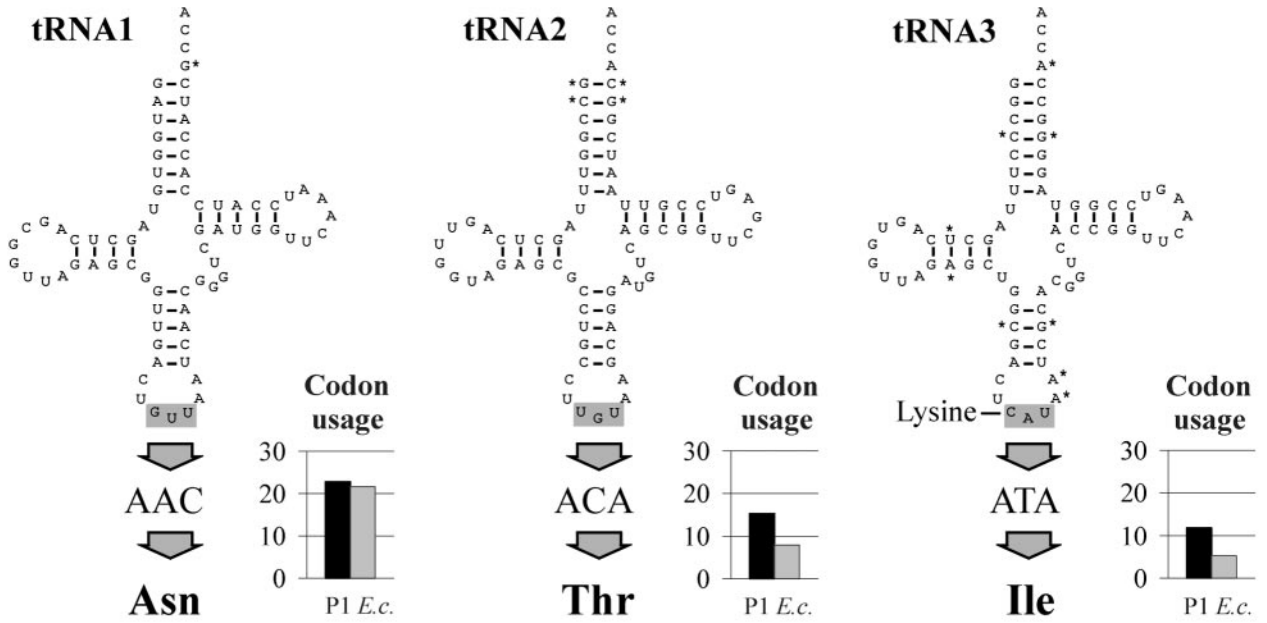

FIG. 3. P1-encoded tRNAs and codons presumably recognized by them. The identity determinants of tRNAs are marked by asterisks. The putative modification of tRNA3 by lysinylation at the 2 position of cytosine alters the recognition specificity of the anticodon. The bar graphs compare the usage frequencies of codons recognized by P1 tRNAs in protein-coding genes of P1 (black bars) and *E. coli* K-12 (grey bars) per 1,000 codons. The usage frequency of a particular codon in P1 is its relative abundance among codons in genes that encode proteins. The usage frequency of a particular codon in *E. coli* K-12 is according to the March 2004 edition of the Codon Usage Database found at its website (http://www.kazusa.or.jp/codon/).

both single ATAs and tandem ATAs. In eight, ATAs are among the first five codons of the proximal region (at position 5 of *lydC*, 4 of *21*, 2 of *ppfA*, 3 of *tciB*, 2 of *ppp*, 4 of *doc*, 4 of *pdcB*, and 2 of *c*1). At least two of these genes, *doc* and *c*1, are known to be expressed, albeit not efficiently, during lysogeny, indicating that the low levels of *ileX* tRNA that are normally present in *E. coli* cells are sufficient for their translation, since tRNA3 is not expressed. However, the known translation-limiting role of ATA codons in *E. coli* implies that the efficiency of translation of *doc*, *c*1, and other genes rich in the ATA codons increases under conditions of abundance of P1 tRNA3. In the case of *doc* and *c*1, the physiological significance of such an increase, if it does occur, is unclear. In the case of certain other genes, the requirement for abundant tRNA3 to permit their efficient expression may provide an additional control on the timing of P1 lytic development at the translational level. All P1 tRNA genes appear to be transcribed from the C1-controlled promoter of the *ban* operon and so become available early during lytic development.

It is noteworthy that *lpa*, whose product activates transcription of P1 late genes, has in its proximal region three rare *E. coli* codons, recognized by P1 tRNAs, namely ATA at position 13 and an ATA ACA pair at positions 18 and 19. Whether the abundance of P1 tRNA3, and perhaps tRNA2, early in P1 lytic development normally contributes to efficient translation of *lpa* mRNA, which in turn ensures efficient transcription of late genes, remains to be tested.

Insertions of numerous phages into bacterial chromosomes occur at or near tRNA genes (reviewed in reference 46). Conceivably, tRNA genes could have been acquired by a P1 ancestor either through its integration and then imperfect excision from the DNA of its host or via recombination from other phages whose prophages can integrate into a bacterial chromosome.

(b) C1-controlled replication functions. P1 DNA synthesis has been reported to start as early as 5 min after infection of *E. coli* (282). Lytic replication initiates bidirectionally in the theta mode from an origin, *oriL*, distinct from the plasmid origin and located within the essential replication gene, *repL* (61, 112, 295). Later, lytic replication switches into the rolling-circle mode (60). The shift is accompanied by a gradual decrease in accumulation of θ-shaped replication intermediates as σ-shaped intermediates accumulate.

Most likely *oriL* is within the second half of the *repL* gene, where we find putative binding sites for DnaA and IHF proteins and two 5′-GATC sequences (Table 7). P1 phage replication in the rolling-circle (σ) mode, predominant in the later stages of phage development, might initiate from an origin other than *oriL*. One possible location for an additional P1 origin of lytic replication is within the *rlf* operon, which directly follows the *repL* gene. It contains multiple sets of GATC sequences and several IHF binding sites, which are internal to its two protein-coding regions (Table 7). An alternative proposal is that rolling-circle replication starts at *pac* from a nick introduced in DNA by pacase (296).

In addition to the essential replication protein, RepL, P1 encodes homologs of certain *E. coli* replication-associated proteins: DnaB helicase (67, 237), single-stranded DNA-binding protein, Ssb (164, 187), Dam methyltransferase (66), and the theta subunit of DNA polymerase III. Genes encoding these homologs (*ban*, *ssb*, *dmt*, and *hot*, respectively), are in separate operons and, with the exception of the last mentioned, are controlled by C1. Four host proteins have been found to be essential for P1 lytic replication: DNA primase (DnaG), the

host helicase-loading factor (DnaC), and the DnaE and DnaX subunits of the DNA polymerase III holoenzyme (119, 234). This list of essential host proteins is probably incomplete.

The *repL* gene is in an operon with the preceding gene, *kilA*, whose function is unclear (112, 295). When expressed, *kilA* is lethal to *E. coli*. Its product is dispensable for phage replication. The 180-nt antisense RNA transcribed from the constitutive Pas*kilA* promoter internal to *kilA* interferes with transcription of the *kilA repL* operon initiated from P*kilA* (125). It may counteract any leakiness of the C1-mediated repression of P*kilA* in P1 lysogens. The activity of Pas*kilA* is much stronger on linear than on circular templates. Thus, it was proposed that transcription from Pas*kilA* during lytic development serves to down-regulate (the initiation of) P1 DNA replication at two stages: immediately after infection before the P1 DNA is circularized and late after infection when rolling-circle replication predominates and linear P1 DNA concatemers accumulate (125).

Presumably the origin region is transcriptionally activated (61, 295). A candidate promoter providing the activation, as well as the initiator message, is the putative P*repL* promoter, which we find directly upstream of *repL* (Table 4).

The *ssb* gene of P1 is in an operon with a downstream gene of unknown function, *isaA*, which does not have a separate promoter sequence (Fig. 1). The *isaA* gene probably has no essential function as it is absent from the *ssb* operon of the closely related P7 phage (M. Łobocka, unpublished results). The cloned P1 *ssb* gene can substitute for *E. coli ssb* (28).

A large P1 gene able to complement the defect in *dam* mutants of *E. coli* (originally *dam*, renamed here *dmt*) was shown to be directly under C1 control (59). We find it to be the last gene in a three-gene operon. The first gene of this operon (*uhr*) encodes a small protein without homology to proteins of known function. The second gene encodes a 34-kDa protein that appears to be a homolog of the *E. coli* nucleoid-associated protein RdgC (233), and thus we designate it *hrdC*. Proteins of the RdgC family are conserved in several species of bacteria. RdgC was proposed to be an exonuclease that is involved in the removal of stalled replication forks, based on its being required for proper replication in cells deficient in the recombination enzymes RecABC and SbcCD (263). The gonococcal homolog of RdgC plays a role in pilin antigenic variation, suggesting that it may have a role in site-specific DNA recombination (218). The amino acid sequence of the P1 RdgC homolog contains a motif characteristic of the active site of *E. coli* DNA polymerase I and related polymerases (see Table S1 in the supplemental material) (39, 74, 160; reviewed in reference 243). The function of HrdC, which is one-third the size of DNA polymerase I (PolI), remains to be elucidated.

The P1 homolog of *E. coli* Dam methyltransferase is a 754-residue protein. Its homologies to Dam are contained solely within the C-terminal region of 266 residues. The major part of the protein is highly homologous to C-5 cytosine methyltransferases of various bacteria and bacteriophages. Most likely the protein is able to methylate DNA at both the N-6 positions of adenines and the C-5 positions of cytosines. Accordingly, we designate it Dmt (DNA methyltransferase).

In *E. coli*, the Dam methyltransferase has multiple functions. It is involved in postreplicative methyl-directed mismatch repair, contributes to the control of initiation of DNA replication, and regulates the transcription of certain genes (reviewed in reference 255). Functions of the adenine methyltransferase activity of the P1-encoded Dmt protein are apparent only in a *dam* host. In such a host, P1 *dmt* mutants that are defective in adenine methylation of 5′-GATC sequences show a diminished burst size of 5 to 10 viable phage per cell (303). One of the two sequenced P1 strains contains a 319-bp deletion (formerly designated *dam*ΔMB) in the region of *dmt* that codes for the N-6 methyltransferase domain. Decreased specificity of the packaging endonuclease, resulting in impaired cleavage of P1 DNA at the *pac* site and increased cleavage of host DNA, could account for the low phage yield as well as the increased yield of transducing particles that has led to widespread use of the mutant phage.

In order to ascertain possible targets of Dmt-mediated adenine methylation, we searched for 5′-GATC sequences in the entire sequence of P1. About 20 clusters of two or three 5′-GATC sites were found, including the few that had been described previously (Fig. 1). One cluster is in the region of the lytic origin (30 nt from a predicted DnaA box). The location of another suggests that it regulates the *kilA repL* genes and controls transcriptional activation of the origin region. Two other clusters are internal to genes of the *rlf* operon, downstream of *repL*, where an additional origin of P1 lytic replication may be located. One that is internal to *rlfA* is flanked by putative IHF binding sites. The predicted strongest promoter in the genome of P1 (Pas*rlf*), which is located immediately downstream of the *rlfAB* genes and could drive transcription of RNA antisense to the mRNA of the *rlfAB* operon, overlaps another cluster of 5′-GATC sites and thus also appears to be controlled by DNA methylation. Several additional clusters within coding sequences, such as the 5′-GATC sequences that are internal to *pacA*, may be associated with particular stages of phage development. Of single 5′-GATC sites, 10 overlap known or putative promoters of P1 genes (Table 4) and presumably affect their function, as was shown for the 5′-GATC sites that overlap certain promoters of *cre* (302). The list of genes that are most likely controlled by methylation of 5′-GATC sites may thus be extended to include *c*1, *coi*, *rlfB*, and genes of the *sim* and *sit* operons.

The *ban* gene (67, 237), which can complement the replicative helicase defect of *E. coli dnaB* mutants, was mapped to a 1.6-kb region located 2.5 kb downstream of the bivalent C1 operator O*ban*ab (formerly Op72ab; Table 6) (132, 316). It encodes a 455-residue protein (195) identical in 78% of its amino acid sequence to *E. coli* DnaB. The Ban protein cross-reacts with anti-DnaB serum (129) and forms active thermally stabilized heteromultimers with the DnaB protein of *E. coli dnaB*(Ts) mutants (82, 183, 310). Although *ban* suppresses a number of conditionally lethal *dnaB* mutations in *E. coli* (67), the functional replacement of DnaB by Ban is somewhat incomplete, as indicated by the thermosensitivity of a *dnaB* deletion strain that carries the *ban* gene (195). Mutations exist which permit growth of this strain at the nonpermissive temperature (E. Lanka, personal communication). Whether they reside in the chromosome or in the *ban* gene remains to be determined. No phenotype was associated with a *ban* mutation of P1 in a *dnaB*$^+$ *E. coli* strain. Presumably *ban* can provide a selective advantage to P1 in other hosts or under other conditions.

TABLE 8. Known and putative late promoters of P1

Name	Position[a]	Strand	Sequence −22	Sequence −10	Identification[b] *tsp*	Identification[b] *pp*	Identification[b] *cmp*	Reference
LP*mat*	2678	+	TTGACAACTGAC AAGTGACTT CAGT	CAGAAT CATCACACGC		√		189
LP*pro*	16052	+	GTAATCCTTAAC AAGTGACTA GTGT	TAAATT CCGTTCAAAC			√	This work
LP*lyz*	19770	−	AACGTAGCTAAC AAGTGATTT GCGT	TATCCT GTGTCTTCTA	√	√		277
LP*darA*	30013	−	TGACCGGTTCAA AAGTTACTT TGCA	TACCAT TACCTCCTGA	√	√		105
LP*S*	37119	−	GGGCGGGGTGAC AAGTTACTT ATCT	TACAAT GAGGCTTCAC	√	√		105
LP*sit*	45409	−	TACTCAATGAAC AAATGACTA CTCG	TAGAAT CGGTTAACAC	√			H. Lehnherr, unpublished
LP*23*	57626	−	GAGAAAGTCAAC AAGTGACTT TCAA	TAAAAT CTCTTCCGAA			√	This work
LP*26*	77286	−	TAAATATTCAAC AAATGACTA GCGG	TAGAAT CACCATCATC	√			275
LP*pmgL*	77296	+	TCATTCTACCGC TAGTCATTT GTTG	AATATT TAACTCAATA		√		H. Lehnherr, unpublished
LP*hot*	85322	+	AAAAAAAATTAA AAGTTACTT TGCTGGTTA	AATAAT AGTCGTTACT	√	√		275
LP*pac*	89276	+	GGGCCTTTTGAC ATGTGACTT TCGT	TACCCT CGCGTCAAAA		√		H. Lehnherr, unpublished

[a] The coordinate is that of the first nucleotide of the promoter, that is, the first nucleotide of the −22 sequence for promoters read from the + strand (clockwise) or the last nucleotide of the −10 sequence for promoters read from the − strand (counterclockwise). Coordinates refer to positions of promoters in the P1 *c*1-100 *mod749*::IS*5* genome without its nonintegral parts, IS*5*, and the associated 4-bp duplication. Note that LP*mat* shares its −10 sequence with P*mat*.

[b] Promoter identification is based on a transcription start point determination (*tsp*), detection with the promoter probe vector (*pp*), or computer analysis of the sequence (*cmp*). The promoters distinguished by √ in the *cmp* column were identified on the basis of the similarity of their sequences to the sequences of the other nine late promoters, using the GCG Fitconsensus program and the consensus table of −22 sequences of the known late promoters made by the GCG program Consensus. Only the sequences with homology scores similar to the known promoters and located 4 to 14 nt (189) upstream from sequences similar to -10 hexamers of RNA polymerase promoters of *E. coli* (as estimated by the program Targsearch [229]) were considered.

The *ban* gene is the penultimate gene in a 5-kb operon of unusual structure (Fig. 1). Immediately downstream of the promoter (designated P1*ban*) is a 1.5-kb region that apparently specifies untranslated RNA, including tRNA1 and tRNA2. This region is followed by two clusters of ORFs separated by another region (0.3 kb) that specifies untranslated RNA, including tRNA3. The *ban* gene is the first of two ORFs in the second ORF cluster.

The first ORF cluster upstream of *ban*, named by us *tciABC*, and the ORF downstream of *ban*, named by us *dbn*, encode proteins that have no homologies with known replication-associated proteins. The predicted products of *tciA* and *tciB* have homologs of unknown function in various prophages (see Table S1 in the supplemental material). P1 TciA has significant similarities to the TerB proteins of R478 and other plasmids of the IncHI2 incompatibility group. The *terB* genes in those plasmids are internal genes of multigenic operons that confer upon *E. coli* resistance to tellurite and to channel-forming colicins and cause inhibition of cell division and of the propagation of certain phages (163, 260, 331, 332).

(c) Recombination enhancement. The product of the P1 gene *ref* (recombination enhancement function) can stimulate RecA-dependent recombination between regions of DNA homology shorter than the minimum needed for recombination mediated by RecA alone (207, 337, 338). The proximity of *ref* to *cre* might suggest that Ref plays a role in the cyclization of P1 DNA molecules that, lacking two copies of *lox*, are not substrates of the cyclization recombinase, Cre. However, within the limits of the assays, no enhancement of P1 *cre*$^-$ lysogenization by Ref could be detected. The substrates of Ref appear confined to regions of microhomology (337, 338).

Regulation of *ref* is complex (338). One presumptive promoter, 91 bp upstream of the first ATG of *ref*, is overlapped by a C1-recognition sequence. A second presumptive promoter (which we locate closer to *ref* than originally proposed) lies within a region that could fold into a long stem-loop and prevent formation of a terminator stem-loop further downstream. It is likely that transcripts from the C1-controlled promoter form the antiterminator structure, whereas transcripts from the C1-independent promoter would be blocked by the terminator (338). In addition, translation of *ref* message can be from either of two AUG codons 46 amino acid residues apart, although the shorter product has low activity. The benefit to P1 that *ref* confers remains to be determined.

(d) Late transcription activation. With a few exceptions, genes involved in late functions are preceded by late promoter sequences (Table 8) and are partly or entirely dependent on the product of *lpa* (late promoter activator) for their transcription (188, 193). The *lpa* gene is under the control of a C1-controlled promoter, P*lpa*. Early in lytic development, it is cotranscribed with two genes, *pacA* and *pacB*, that directly follow *lpa* and that are additionally transcribed from the late promoter LP*pac* (111). The sites recognized by the Lpa protein and requirements for its function are considered in the section "Lpa-controlled operons."

(e) Morphogenetic functions that start to be expressed early in lytic development. Among 17 P1 operons controlled by C1, five (*mat*, *26*, *ppp*, *pap*, and *pmgT*) encode morphogenetic functions (Fig. 1 and Table 3; see also Table S1 in the supplemental material). Although early expression of these operons may seem surprising, it is noteworthy that, in addition to being controlled by C1, they all appear to be activated by Lpa. Two of them, *mat* and *26*, contain in their immediate upstream regulatory regions, in addition to a C1-controlled σ^{70} promoter, an Lpa-activated late promoter. Three others, *pmgN*, *pmgR*, and *pmgT*, are likely to be transcribed from an Lpa-activated promoter of the upstream *pmgN* gene by readthrough transcription.

An Ssp-dependent mechanism was found to turn down transcription from C1-controlled promoters before the Lpa-controlled promoters are activated (28). This finding indicates that, for those genes transcribed from a C1-controlled promoter, additional control from a late promoter is required to ensure the continuity of their expression throughout lytic development and to satisfy the requirement late in development for large amounts of certain proteins. Dual regulatory signals

in control regions of certain operons are common in T4 and other phages, where they can serve to extend expression of a given gene to more than one developmental stage and to modulate it differently at different stages (224, 228).

P1 *mat* mutants produce viral particles with empty heads and unstable tails with tail sheaths in various stages of contraction (323). The pattern of *mat* expression following induction of a P1 lysogen is consistent with its being subject to dual control (191). Expression was seen to start within 10 to15 min of inception of lytic development, to continue at a low level for about 10 min, and then to increase sharply, exactly when the expression of other genes activated by Lpa begins. Expression profiles of the remaining genes that are controlled directly by both C1 and Lpa may be similar. As some products of the remaining C1- and Lpa-regulated operons are structural virion proteins, they are described later. Mat was proposed to have a regulatory role in morphogenesis (323), although with scant justification.

The *pacAB* operon, which encodes two subunits of the P1 DNA packaging enzyme, is also subject to dual control. Although it is directly preceded only by the late promoter LP*pac* (Table 8), it is expressed early from the C1-controlled promoter of the upstream *lpa* gene (105). Moreover, some PacAB-mediated cleavage of *pac* sites occurred in the absence of Lpa, indicating that Lpa is not essential for *pacAB* operon expression during P1 lytic development (297).

(ii) Lpa-controlled operons. P1 late promoters resemble typical *E. coli* σ^{70} promoters at their −10 regions but lack the −35 hexamer (188). Instead, they have a conserved 9-bp inverted repeat (previously termed late operator) that is centered about position −22 of the transcription start site and interacts with Lpa (111, 189). Results of in vivo footprinting experiments confirm that RNA polymerase binds to late promoter sequences in the presence, but not in the absence, of Lpa (111). In addition to Lpa, the *E. coli* RNA polymerase-associated protein, SspA (158), is required both in vivo and in vitro for late promoter activation (111, 336). The activation requires the σ^{70} RNA polymerase holoenzyme and does not occur when σ^{70} is replaced by σ^{S}. How Lpa and SspA cooperate in redirecting the host RNA polymerase towards P1 late promoter sequences is unknown.

Nine P1 late promoters were identified previously by empirical studies (Table 8). The optimal spacing between −10 and −22 regions of late promoters is 4 bp, and it is conserved in all but one of these promoters, LP*hot*. Although in the LP*hot* promoter the spacing (9 bp) is suboptimal, a promoter probe assay confirmed in vivo LP*hot* transcriptional activity to be dependent upon Lpa (275).

Here we identified two additional putative late promoters, based on the strong similarity of their predicted −22 regions to the −22 regions of known late promoters and on the presence of appropriately positioned putative −10 hexamers (Table 8). Both promoters have the optimal spacing between their predicted −10 and −22 regions, and both are located in the regulatory regions of genes known to encode phage morphogenetic functions. The 11 late promoters control at least 12 and as many as 14 P1 operons. We suspect that transcription from LP*pmgL*, before encountering the closest terminator, T*pmgQ* (Table 5), can read through two genes that are immediately downstream, *pmgL* and *pmgM*, and a C1-controlled operon further downstream that encodes morphogenetic functions (Fig. 1). It is likely that this transcript includes an additional two operons downstream from T*pmgQ* that also encode morphogenetic functions. Expression of these operons late during lytic development implies that P1 uses an efficient antitermination mechanism to attenuate the function of T*pmgQ*, T*pmgS*, and T*pap* terminators. Attenuation of terminator functions may also occur at T*22*, T*pmgG*, and T*ddrA* terminators, since these would otherwise prevent late expression of genes *22*, *pmgG*, *pmgF*, *ddrB*, and *hxr*, encoding known or predicted late functions. Alternatively, these terminators may be inefficient and serve to decrease the expression of certain late genes distant from the LP*pmgL*, LP*23*, and LP*dar* late promoters. One other late promoter, LP*sit*, could drive transcription of both the *sit* operon and the following *R S U* operon. However, the *R S U* operon is also transcribed from its own late promoter, LP*S*. The remaining nine late promoters drive transcription of single operons that adjoin operons transcribed in the opposite direction. In total, 52 genes of P1 can be transcribed from late promoter sequences.

In *E. coli* at 37°C, late gene products start to be synthesized 20 to 30 min after either transient thermal induction of a P1*c*1ts lysogen or infection of sensitive cells by P1. The completion of phage morphogenesis at about min 40 is accompanied by the onset of cell lysis. The late genes of P1 are involved in DNA replication, phage tail and head morphogenesis, packaging of DNA, synthesis and incorporation of injectable proteins, and cell lysis. Some of them, as explained above, are expressed both early and late in lytic development. Some that are apparently expressed in late developmental stages cannot be associated with any late promoter sequences. The *darB* gene, encoding an antirestriction protein and a P1 head component, is an example. Transcription of these genes may be controlled either indirectly by Lpa through another regulatory protein or by an as-yet-unidentified mechanism.

(a) Head morphogenesis. The morphogenesis of P1 resembles that of other tailed phages in that heads and tails have independent assembly pathways (323). At least some of the head proteins appear to be proteolytically processed (304). DNA is packaged into proheads from concatemers (60) by a headful mechanism, as in T4, but the packaging of each concatemer, unlike that in T4, starts from an initial cut at a specific *pac* site (286, 297, 298). Infective particles of P1 contain cyclically permuted, linear, double-stranded molecules with a terminal redundancy of about 10 kb of DNA (296). P1 tail fibers, unlike those of T4, can attach to the tail before head-tail joining occurs (323).

P1 virions resemble in many respects those of T4, although they are less complex, especially in the structures of the head and the baseplate (323). Mature P1 virions contain at least 28 proteins (325), as compared to at least 49 proteins in mature virions of T4 (summarized in reference 224). In T4, a further five proteins assist in morphogenesis and are removed during virion maturation. Proteins of similar function in P1 await identification.

Analysis of P1 virion proteins by SDS-PAGE indicates that P1 heads are composed of 15 to 19 protein components, possibly an underestimate as small proteins may have escaped detection (325). Although little is known about the assembly pathway of the P1 head, it is likely that, as in many other tailed

phages, a scaffold of the head is assembled on an initiator complex, which is then degraded after being surrounded by structural head proteins, allowing the resultant prohead to expand to a mature head in a process coupled to entry of DNA (34).

Mutations affecting head morphogenesis were mapped to five unlinked regions of P1 DNA (323; reviewed in references 346 and 347). Each of these regions contains an operon that is under the control of a late promoter, LP*pro*, LP*dar*, LP23, and LP*pmgL*, or LP*pac* (Fig. 1). Together these operons contain 21 genes. Two genes of the *pac* operon encode subunits of the P1 DNA packaging enzyme and are described here separately.

The assignment of the majority of head components to genes that encode them cannot be made based on simple size correlations. Certain P1 virion polypeptides are truncated by proteolytic processing; one or more are increased in size by the addition of multimers of an amino acid chain of 1 to 2 kDa, most likely through disulfide bonds (325). With two exceptions, the lack of significant homologies of P1 head proteins to characterized head proteins of other phages excludes sequence-based predictions of their function.

At least four head proteins (with molecular masses of 220, 76, 47.5, and 10 kDa) are dispensable for morphogenesis, as they were absent from heads of phages missing the 9-kb DNA segment containing the *darA* operon (325). The missing 220-kDa protein must correspond to DarB, the only P1-encoded protein larger than 120 kDa (predicted molecular mass, 251 kDa). Although the *darB* gene itself is outside the missing segment, it may require a protein encoded by this segment for its expression or the incorporation of its product into phage heads.

The other three dispensable proteins are probably products of the *darA* operon. The 47.5-kDa protein may represent DarA, which is proteolytically cleaved from a 69.6-kDa precursor. The 76-kDa polypeptide may be a processed form of the 108-kDa product of the second large gene of the *darA* operon, *ddrB*. The 10-kDa polypeptide could represent the product of *hdf*, a protein that is homologous to a fragment of DarA and may be a head component, as is DarA itself.

An unusual feature of P1 is the variety of head sizes. In addition to normal, infectious particles with big heads (P1B), P1 also produces small (P1S) and, more infrequently, minute (P1M) particles. S and M particles lack the capacity to hold the complete P1 genome and are nonproductive in single infections (14, 156, 322). A gene that regulates the proportion of phages of different head sizes in the progeny, designated *vad* (viral architecture determinant) was mapped to the late operon, under LP*dar*, but it is unclear whether it corresponds to the *dar* gene of this operon or to a gene close to *dar* (147, 149, 155).

The major head component is a 44-kDa protein that was present in different stoichiometric amounts in heads of big-headed (P1B) and small-headed (P1S) phages (325). It appears to represent a truncated product of gene *23* (H. Lehnherr, unpublished results). The full-length product, a 62.2-kDa protein, is also one of the major bands of P1 head proteins separated by SDS-PAGE. A mutation in gene *23* resulted in the production of complete tails and only a few heads following induction of the mutant lysogen (323). Whereas the neck and head connector are normally associated with the head, the tails of the mutant contained structures resembling head-neck connectors or aberrant tail extensions. The product of gene *23* has no homologies to known capsid proteins of other phages. Either its evolutionary distance from these proteins is so great that only similarities at the structural level have been retained or gp23 is representative of a separate class of phage capsid proteins characteristic of P1.

Gene *23* is the first of six genes transcribed from the LP*23* late promoter and perhaps the only head morphogenetic gene of the *23* operon, since other mutations mapped to this operon cause phenotypes not associated with head morphogenesis (281, 323). Homologies of the predicted gene products of the genes downstream of *23* to known phage proteins suggest that they have roles in sheath, tail tube, and baseplate morphogenesis.

The major set of head structure and assembly genes must be contained within the array of 13 genes transcribed from the LP*pmgL* promoter, as other late operons of P1 could not encode all the unassigned head proteins. An amber mutation in a gene of this cluster, previously designated *8*, conferred a phenotype similar to that of a *23* mutation, except that there were half as many empty heads as tails (323). Apparently, gene *8* encodes an essential head structure or assembly protein. Which of the 13 LP*pmgL*-controlled genes corresponds to *8* is unclear. One candidate is the largest gene of this cluster, *pmgS*, whose predicted product is a 48-kDa protein of almost entirely alpha-helical structure, characteristic of prohead scaffolding proteins of other tailed phages (221, 314). The gene that directly precedes *pmgS*, *pmgR*, is another candidate, as its predicted product contains a putative transmembrane domain. In T4, the assembly of the major head protein into proheads is initiated on membrane-bound connectors built of other T4-encoded proteins (34).

Only one protein encoded by genes transcribed from LP*pmgL*, the product of the fourth gene in this array, *pmgO*, appears to have homologies to putative phage head morphogenesis proteins, namely to the products of two directly linked ORFs of similar sequence (ORF7 and ORF8) of the *P. aeruginosa* D3 phage, which in the D3 genome directly follow a gene for the major capsid protein (176).

The predicted product of the sixth gene of this array, designated by us *ppp* (P1 protein phosphatase) appears to be 64% identical to phage λ serine/threonine protein phosphatase (lambdaPP), which is a representative of a group of phage-encoded protein phosphatases and has structural similarity to mammalian Ser/Thr phosphoprotein phosphatases, including calcineurin (62, 172, 320). LambdaPP is a product of the 221-codon ORF (ORF221) in the region of eight nonessential delayed-early genes (called the *ninR* region) of the λ genome. An ORF221 mutation, known as *byp*, which abolishes the phosphatase activity of lambdaPP, caused a defect in the establishment of lysogeny, unlike a *nin* deletion (62, 65). The defect was proposed to depend either on the inability of altered lambdaPP to dephosphorylate a host protein that normally participates in termination of transcription of λ early genes or on constitutive transcription, from a promoter created by the mutation, of the λ *Q* gene, encoding the antiterminator of late gene transcription (24, 53). Whether the product of P1 *ppp* serves P1 by dephosphorylating P1-encoded proteins involved in virion morphogenesis or a host protein or both remains to be determined.

The tenth gene in the array, *pap*, like *ppp*, can be transcribed both early and late in P1 lytic development. The gene encodes a protein that is homologous to the C-terminal moiety of the polynucleotide kinase (Pnk) of the T4 bacteriophage. By itself, this acid phosphatase domain of Pnk can remove 3′-PO_4 termini from DNA or RNA polynucleotides (328), hence our assignment of the name *pap* (P1 acid phosphatase). The aspartate residue at position 167 of T4 Pnk, which is an essential and presumably catalytic constituent of the Pnk 3′-phosphatase domain, is conserved in P1 Pap, as are the surrounding amino acid residues. A gene whose product would be a homolog of the N-terminal kinase domain of T4 Pnk could not be found in the genome of P1. Thus P1 Pap appears to have a different role from that of Pnk in the lytic development of T4; determination of that role will require further study.

As noted above, some structural head proteins of P1 are synthesized as precursors that are proteolytically cleaved to the form found in mature phage particles (304). One such cleavage, of the product of late gene *darA* to 9-kDa and 68-kDa proteins, was associated with a locus initially designated *4* and later *pro* (summarized in reference 346). The *pro* function is essential for P1 development as indicated by the conditionally nonproductive phenotype of all the mutants isolated. It was proposed that the product of *pro* must be involved in processing other structural proteins in addition to DarA, as *darA* is not an essential gene (304).

The *pro* gene belongs to an operon of two genes of previously unknown sequence which appear to be controlled by the putative late promoter, LP*pro*. They were originally defined as two complementation groups to which mutations with similar pleiotropic phenotypes could be assigned (230, 252, 294, 321, 324). In addition to being defective in the proteolytic processing of DarA, amber mutations in both complementation groups were found to cause production of complete phage tails and unattached empty heads, suggesting a defect in a protein involved in head-to-tail connection (304, 323). In support of this conclusion, the predicted product of the first gene of the *pro* operon is a protein that has similar length and some sequence similarities to the portal protein of bacteriophage T4, gp20, and its homologs in several T4-like cyanophages (110). We conclude that this gene encodes the P1 portal protein and designate it *prt*.

The predicted product of the *pro* gene is a protein of complex structure. Its N-terminal region is homologous to the N-terminal moiety of the ATP-dependent Hpr kinase/phosphatase of *Enterococcus faecalis*, HprK (174). Its central part has homologies to the C-terminal region of the ClpX protein of *E. coli*, which is an alternative ATP-binding subunit of the Clp protease (100, 350). Although it seems likely that Pro activates, and acts in a complex with a host protease, ClpP is excluded from that role since P1 is able to plate on a *clpP* null mutant of *E. coli* and transduce the mutation.

The Prt protein may be a substrate for Pro-mediated proteolytic processing. Thus, a likely explanation for similar pleiotropic phenotypes of amber mutations in either of the genes of the *prt pro* operon, each of which encodes a different function, is that mutations in *prt* have a polar effect on *pro* expression, whereas mutations in *pro* may lead to accumulation of an unprocessed product of *prt* that can not be incorporated into P1 virions.

(b) DNA packaging into proheads. The processive headful packaging from a specific cut in a concatemer that produces cyclically permuted, terminally redundant DNA in T4 phage populations was originally proposed to start from a specific cut in a concatemeric molecule. T4 was later shown not to be specifically cut for packaging, but certain other phages were, including P1 (22).

The site, *pac*, at which each concatemeric P1 DNA is cut to be packaged into a prohead lies just within the *pacA* gene, encoding one of the two subunits of the pacase enzyme (286). Packaging into the first prohead proceeds from *pac* towards the site-specific recombination site *lox* (22), 4 kb away (Fig. 1). It continues through the next *pac* and *lox* and roughly 4 kb further until the prohead is filled and packaging into a second prohead can begin. Consequently, the first DNA that is packaged from each concatemer contains two *lox* sites, one at the beginning and the second at its redundant end. Upon entering a cell, this DNA has the advantage of being able to cyclize either by homologous or site-specific recombination. The processivity of packaging appears limited to three or four headfuls, as determined by the proportion of nonstoichiometric fragments (generated by cuts at *pac*) in restriction digests of DNA from phage particles (22). The limiting factor may not be the processivity of pacase itself but the extent to which the substrate of packaging is subject to RecBCD-mediated degradation from the end unprotected by bound pacase (297).

Cleavages within the *pac* site are distributed over 13 bp, which are flanked on each side by four similarly oriented hexanucleotide elements (5′-TGATCA/G) (298). These sites are substrates for adenine methylation by the bacterial and viral methyltransferases Dam and Dmt, respectively. If the sites are unmethylated or hemimethylated, cleavage does not occur, although hemimethylated (as well as fully methylated) DNA forms complexes with the pacase (296).

These observations have suggested that pacase complexed with hemimethylated *pac* sites prevents their full methylation and hence their cleavage. The late expression of *dmt* could switch pacase function from nonselective sequestration of *pac* sites to selective cutting of those *pac* sites that are newly generated and able to become fully methylated. Presumably only late in lytic development, when Lpa has activated the transcription of *dmt* and accelerated the transcription of *pacA* and *pacB*, does cleavage at *pac* sites occur and then only at some nascent sites as they are generated by rolling-circle replication. This proposal implies that DNA packaging proceeds from the rolling circle in the direction of the first DNA to roll out. As suggested earlier, the nick responsible for initiating rolling-circle replication may be made by the same enzyme that cleaves both DNA strands to initiate packaging (296). The proposed roles of pacase in successively preventing and promoting *pac* site cleavage could explain the finding that *pacA* and *pacB* start to be transcribed from an early promoter well before packaging begins.

Cleavage of the P1 packaging site requires the product of both proteins PacA and PacB, presumably as heterooligomers. An additional protein (PacC) encoded within the distal third of *pacB* and translated in the same reading frame as PacB has been identified, but its role in packaging, if any, remains undetermined (286). The PacA subunit recognizes and binds to methylated *pac* sites independently of PacB but requires a host

factor to do so (287). The requirement can be satisfied by IHF, for which a binding site lies immediately adjacent to the cleavage region. HU can also satisfy this requirement, and these two host proteins work in synergy, apparently cooperating in bending the DNA at *pac* to bring together the two sets of hexameric repeats (287).

PacA is likely to interact with the P1 portal protein, since an analogous interaction of a DNA pacase complex with a portal protein in T4 has been demonstrated (34, 202). Homologies between P1 and T4 portal proteins strongly support this hypothesis. P1 PacB bears an ATP binding motif (see Table S1 in the supplemental material) and presumably transduces the energy of ATP hydrolysis into the energy of DNA translocation into the capsid, as does the analogous T4 gp17 protein.

(c) Tail and tail fiber morphogenesis. The P1 tail consists of a long cylindrical tube surrounded by a contractile sheath and attached at one end to a baseplate containing six tail fibers. Nine virion proteins (with apparent molecular masses of 117, 105, 72, 33.5, 32.7, 27.2, 24.0, 21.4, and 15.4 kDa) were identified as P1 tail structural components on the basis of their absence in preparations of P1 heads (325). The assignment of four additional proteins to a tail structure was uncertain. Small proteins that could not be detected may extend this list.

Mutations affecting tail assembly were assigned to regions where four P1 late operons, transcribed from the LP*S*, LP*sit*, LP*23*, and LP*26* promoters, are located. Head and tail assembly pathways appear independent, as each tail-defective mutant produced complete heads (323). Homologies of putative products of certain P1 tail genes to tail proteins of T4 or T4-like phages indicate a common ancestry of P1 and T4 tail morphogenetic functions. The clearest difference is in the appearance of the baseplate, which in P1 is thinner and smaller than in T4 (323).

A major tube component is a product of the second gene in the *sit* operon (designated here *tub*). Its C-terminal moiety is homologous to C-terminal moieties of tail tube proteins, products of gene *19* of T4-related phages RB49 and nt-1, and, to a lesser extent, to gp19 of T4 itself. In addition, the predicted molecular mass of the P1 Tub protein (22.3 kDa) roughly corresponds to the estimated molecular mass of one of the two most abundant tail components (21.4 kDa) (325), which by analogy to other phages, must correspond to either the tail or sheath subunits. In T4, the tail tube consists of gp19 subunits, arranged in stacked hexameric rings, which form a tunnel through which the phage DNA can pass (reviewed in reference 63). The organization of Tub protein molecules in the tail tube of P1 seems likely to be similar.

In the genomes of T4 and related phages, a gene encoding the tube protein is typically located between a gene encoding the sheath protein and a gene encoding a portal vertex (110). In the genome of P1, this organization is not conserved. The *tub* gene is flanked by genes whose products resemble an essential baseplate hub subunit of T4-related phage RB49 and a protein of unknown function of another T4-related phage, Aeh1.

A single P1 amber mutation that caused production of complete heads and tail tubes without sheaths was mapped to a gene designated *22* (323). This gene is contained within the late operon *23* and distant from the *sit* operon, in which the *tub* gene is located. Gene *22* was proposed to encode the sheath protein. This assignment is consistent with the presence in gp22 of a motif characteristic of conserved domains of bacteriophage tail sheath proteins (see Table S1 in the supplemental material) and regions similar to a nucleotide-binding motif (data not shown) that is present in the sheath protein of T4 (63). Although the calculated molecular mass of gp22 is only 57 kDa, smaller by 15 kDa than that of the P1 sheath protein estimated by SDS-PAGE, there is no other unassigned P1 ORF transcribed from a late promoter that could encode an essential protein of about 72 kDa. It is possible that the mature sheath protein of 72 kDa found in virions of P1 is formed by cross-linking of gp22 with another protein. Although translational frameshifting was found to occur during synthesis of λ and P2 phage tail components (56, 199), it is unlikely to be responsible for the extension of P1 gp*22*. No single frameshift within the *22* sequence could generate a significantly longer protein.

All four operons that encode P1 tail proteins contain genes for baseplate components. Products of five of these genes have homologies to baseplate proteins of T4 or T4-related phages. In T4, the baseplate consists of a hub surrounded by six wedges. Baseplate morphogenesis begins with assembly of wedges that polymerize around the independently assembled hub (reviewed in references 63 and 86).

Two P1 genes, belonging to well-separated operons, encode proteins that are likely to participate early in the assembly of baseplate wedges. One of them, which we name *bplA*, is identified by the extensive homology of its product to gp6, the largest baseplate protein and major wedge component of T4 (329). The *bplA* gene is located downstream of the *tubA* gene in the *sit* operon. The other gene, *26*, in the *26* operon, encodes a protein with homology to gp53 of T4 and related phages Aeh1 and RB49. The products of both P1 genes are somewhat smaller than their T4 homologs. In T4, the order of addition is gp6 before gp53 and the two proteins may interact (reviewed in reference 224). T4 phages mutant in either gp6 or gp53 produce polysheaths in addition to complete heads, but no visible tail structures, consistent with the requirement of gp6 and gp53 for the assembly of the T4 tail. P1 proteins BplA and gp26 may play analogous roles in baseplate and tail assembly. In support of this view, gene *26* mutant phages produced complete heads and no tail structures other than polysheaths (323). A similar phenotype was attributed to P1 mutants in a gene previously named *3*, which was mapped to the *pmgC tub pmgB sit pmgA bplA 16* operon region. Whether *bplA* and *3* are identical remains to be seen. Mutations in three other P1 genes, *R, 16*, and *21*, cause phenotypes identical to those of gene *26* mutants (323). The products of these genes lack homology to proteins of other phages, indicating that they form those parts of the P1 baseplate whose evolutionary history is different from baseplate components of T4, and they are perhaps unique to P1.

The largest tail component is the product of the *sit* gene (121 kDa). The Sit protein has a mosaic structure. Between amino acid residues 757 and 885, close to the C-terminal end of this 1,140-amino-acid protein, are motifs characteristic of soluble lytic transglycosylases (190). Sit was proposed to contact the outer surface of the bacterial cell wall at infection and facilitate passage of the viral genome through the cell wall by introducing small gaps into its peptidoglycan layer, as does the soluble lytic transglycosylase of *E. coli*. The structure of Sit has simi-

larities to that of T4 gp5, a 575-residue protein that contains a separable domain with lysozyme activity (17, 165). Posttranslational cleavage of gp5 allows this domain to join in the assembly of the baseplate and form a cell-puncturing device that penetrates the outer membrane and locally dissolves the cell wall at infection. Sit does not undergo major proteolytic processing; a protein of Sit's length can be isolated from virions of P1 (323). The maintenance of the integrity of Sit protein may be necessary to fulfill a putative second Sit function, that of a "ruler. " Sit is the only protein long enough to be a candidate for the P1 tail length determinant. Moreover, its central region (residues 450 to 691) has homologies to the C-terminal moiety of gp29 protein of T4 and the T4-related phage RB69. In T4, gp29 is the largest baseplate hub protein (63). Additionally, gp29 serves as a tail tube ruler which fills the core of the tube during assembly and determines the tail length (9). The modular organization of P1 Sit is not the same as T4 gp29 but resembles that of tail tube rulers of some mycobacteriophages, which have additional domains embedded near their C-terminal ends (245).

Genes of P1 that determine tail fibers and the specificity of P1 adsorption to different hosts were previously localized to two operons. One of them, the tail fiber operon, consists of three genes, *R*, *S*, and *U*, transcribed from the late promoter LP*S* (Tables 3 and 8) (104, 105). *S* specifies a structural protein that extends along the tail fiber with its constant N-terminal part at the baseplate, while its variable C-terminal part is available to bind a specific receptor on the host cell surface (reviewed in reference 347). The host range specificity of P1 is subject to alteration by inversion of the C segment of P1 DNA (Fig. 1). The C segment adjoins the anterior part of the *S* gene (*Sc*) and contains two alternative variants of the posterior part of the *S* gene (*Sv* and *Sv′*) and two alternative variants of the entire *U* gene (*U* and *U′*). The alternative variants are organized in oppositely oriented *Sv′U′* and *SvU* segments so that the inversion connects either of the segments, in frame, to the anterior part of the *S* gene (135, 146, 153). In the so-called C+ phage, expression of *Sv* and *U* enables infection of *E. coli* strains K-12 and C (153).

Analysis of the predicted amino acid sequence of *S* and *U* gene products and their alternative forms in the C− phage (*Sc*+*Sv′* and *U′*) confirms extensive homologies of these proteins to the products of the corresponding Mu genes (see Table S1 in the supplemental material), consistent with the results of early complementation and DNA hybridization experiments.

The gpS protein of P1 appears to be extended at its N terminus by 200 residues relative to its Mu-encoded counterpart. This extra region is partly identical to the N-terminal region of the tail fiber structural protein encoded by plasmid p15B, which has extensive regions of homology to P1 (268, 269). The tail fiber genes of P1, p15B, and Mu appear to have evolved by exchange of segments within a common pool of tail fiber genes.

P1 gpU and gpU′ appear to be 95% identical to their Mu-encoded counterparts and have informative homologies to certain tail fiber assembly proteins of other phages, including gp38 of T4, λ Tfa, and the Tfa protein of a wild strain of λ with tail fibers (Ur-λ) (133). In T4, gp38 acts as a chaperone essential for the assembly of the distal part of long tail fibers and is absent from virions (48; reviewed in reference 134). In contrast to T4 gp38, the Tfa protein of Ur-λ is a virion component (133). Although there is little homology between the two, each can replace the chaperone function of the other (117, 226). This function of Tfa does not require virion attachment. Whereas the C-terminal region of λ Tfa is not homologous to that of T4 gp38, it is homologous to those of P1 gpU and P1 gpU′. It is thus likely that the gpU and gpU′ proteins of P1, like λ Tfa, can function as both tail fiber assembly chaperones and virion components.

Comparison of variable parts of alternative P1 gpS proteins and variable P1 gpU and gpU′ proteins reveals that the main sequence differences between them reside within the C-terminal regions (Fig. 4). Further analysis of these differences may help identify subregions of gpS and gpU proteins essential for adsorption of P1 on its particular hosts.

Inversion of the C segment is accomplished by the Cin site-specific recombinase, the only product of an operon that adjoins the C segment on the opposite side from *Sc* (Fig. 1). Unlike the tail fiber operon, which is expressed from the late promoter, the *cin* operon is expressed from a σ^{70} promoter, allowing inversions of the C segment to occur in the prophage (Table 4).

Cin acts on a pair of inverted 26-bp crossover sites, *cix*L and *cix*R, at the C-segment borders (Table 1). The *cix*L site overlaps with the end of the *cin* gene. Inversions within the analog of the C segment in phages with similar invertible sequences have been reported and may well occur in P1 as well (267). In the p15B plasmid, whose *S* gene is homologous to that of P1 in its posterior region, inversions can occur at six crossover sites within the corresponding invertible segment and may result in 240 isomeric configurations of this region (269, 270). The inversions depend on the p15B invertase Min, which can functionally replace the Cin invertase of P1 (154).

A pseudo-*cix* site, *cix*Pp, overlaps one of the two predicted *cin* promoters, possibly allowing autoregulation of Cin synthesis (Tables 1 and 4) (135, 150). The *cin* segment, flanked by *cix*Pp and *cix*L, does not invert at a noticeable frequency, perhaps because the *cix*Pp site does not contain the conserved core sequence, TT, important for efficient Cin-mediated recombination (148).

Inversion is an infrequent process and, as a result, the progeny of an infecting phage and the infecting phage itself are generally of the same host range, whereas an induced lysogen is likely to produce a burst of phages of mixed genotypes, which are able to plate either on *E. coli* strains K-12 and C or on other P1 hosts. Control of inversion appears to be mediated at a palindromic recombinational enhancer located 18 bp downstream of the translational start signal of the *cin* gene (141). The host architectural protein Fis (factor for inversion stimulation) binds at the enhancer, stimulates (by more than 500-fold) recombination in *cis* to the recombination sites (109, 142–144), and determines the topological specificity of DNA inversion (108). An additional factor that influences the rate of C-segment inversion is probably IHF protein. We find a putative IHF binding site overlapping one of the two σ^{70} promoters that can drive *cin* transcription (Tables 4 and 7). IHF protein is three times more abundant in early stationary phase and two times more abundant in late stationary phase than in exponential phase (11). Fis levels in turn are drastically reduced in stationary phase. These considerations suggest that C-segment

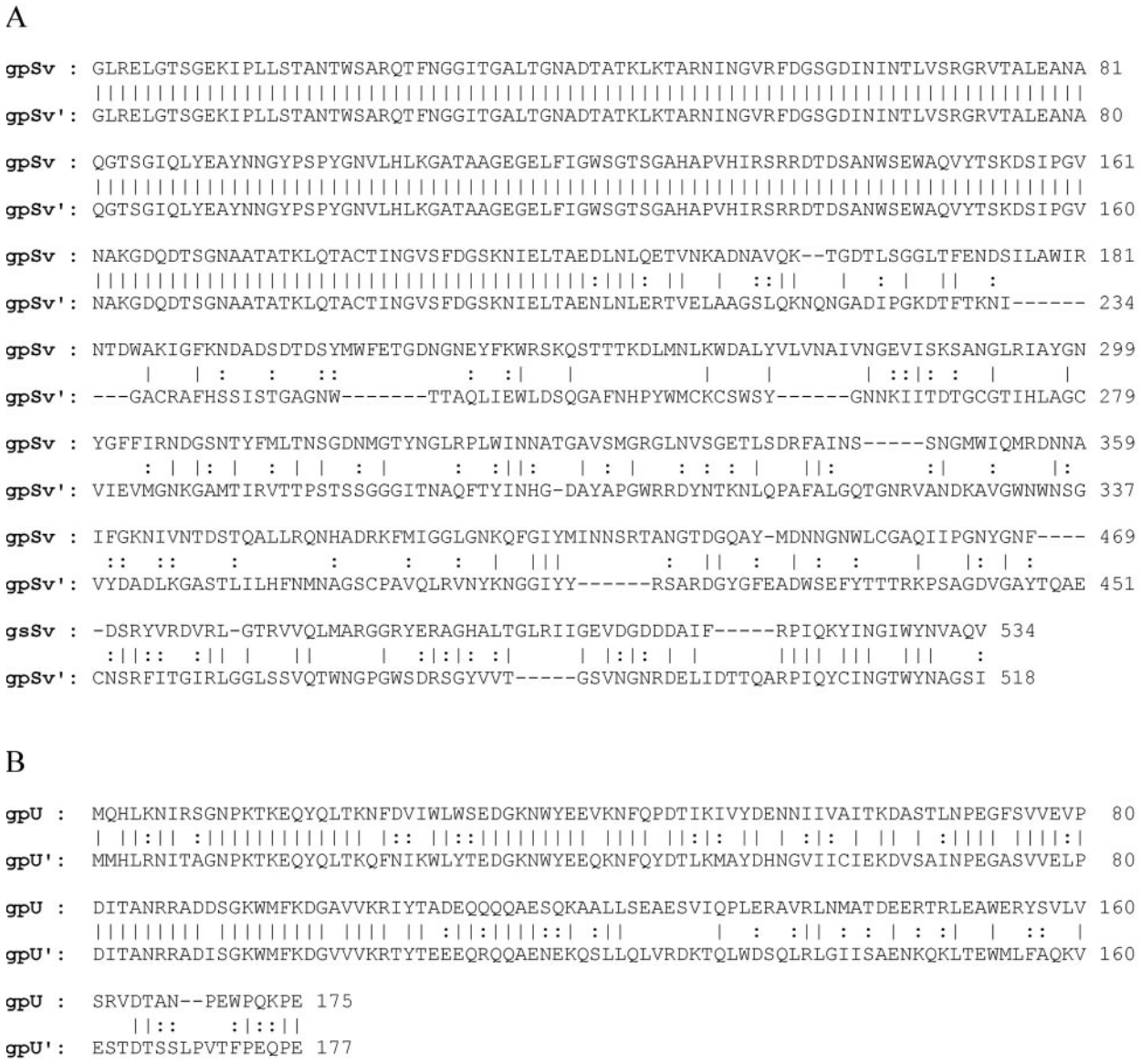

FIG. 4. Alignment of alternative tail fiber proteins and protein segments.

inversion will be especially infrequent during the stationary phase due to the predicted IHF-mediated down-regulation of *cin* expression and the lack of Fis-mediated enhancer function.

Arber has suggested that the ability of Cin to mediate occasional deletions and intermolecular reactions (e.g., under conditions of relaxed topological specificity) could lead to exchange of tail fiber segments and could have been selected for during evolution (15). Tail fiber genes from unrelated phages of enteric bacteria exhibit a mosaic of homologous segments suggestive of extensive horizontal transfer (267). On the other hand, conservation of the order of the various motifs has been taken to imply that independent divergence from a common ancestor also played a major role in tail fiber evolution (342).

(d) Antirestriction. Two P1 head components, DarA and DarB, protect infecting P1 DNA against degradation by different subsets of enterobacterial type I restriction enzymes (155, 304, 325). Both Dar (defense against restriction) proteins are synthesized late during lytic development. The corresponding *dar* genes were mapped to two unlinked operons (Fig. 1).

P1 *darA* is the fourth gene in a multigene operon, whose nucleotide sequence was determined previously (149) and corrected in this study. The operon is transcribed from the late promoter LP*dar* (Table 8) (105) and, in addition to *darA*, contains two lysis-determining genes that will be described separately and four other genes (Fig. 1). None of these genes are essential for P1 development, as cells infected with a P1 deletion mutant lacking the entire *darA* operon can produce infective phages.

The 5′ region of *darA* and the entire preceding gene appear to have evolved through internal duplication, as suggested by the corresponding homology of the predicted products of both genes (see Table S1 in the supplemental material). To reflect this property, we give the name *hdf* (homolog of *darA* fragment) to the gene that is directly upstream of *darA*. It is

Protein	Motif X	Motif I	Motif II	Motif III
M.*Bsu*BI	26 GqFFTPssIsiFMacLF	52 VLDAGaGiGsLtsAF	79 dlhlLEID	98 aLfkDYIeI
M.*Eco*57I	14 GgYYTPqnLadYVtkWV	39 ILEPscGdGvFIqAI	65 cfelFDtE	90 ItegDFLVW
M.*Hinc*II	10 GqFFTPthIvkYMiglM	34 ILEPssGnGvFLdsL	55 nltsYEID	71 VInsSFIts
M.*Pst*I	36 GqFMSssaVseLManLF	61 ILDAGaGvGdFLlPI	110 KiRAMEFE	124 IIesDFIqa
M.*Taq*I	19 GrVeTPpeVvdFMvsLa	43 VLEPAcAhGpFLrAF	66 RfvGVEID	85 gIlaDFLLV
DarB	89 seYYTPkpIaegVweIM	115 tLEPsaGtGvFnetk	134 vmtatEIs	157 VqispFeqL

Protein	Motif IV	Motif V	MotifVI	MotifVII
M.*Bsu*BI	137 FThaIlNPPYkKiksnsK	168 NLYsaFValtV-dlMSdGGeIvFIIPrsFc	204 FRqhLlnktsIkhMh
M.*Eco*57I	110 FDgaLGNPPFiRyqfleR	149 NaWvpFLlssL-alLKqGGrIgMVIPseIs	185 LRsyLgHvcskivII
M.*Hinc*II	85 YDsIIGNPPYvRwlnlse	123 DYFyiFIiKsI-lqLKvGGeLiFICPdyFF	159 LRkfLinngsFekII
M.*Pst*I	146 YNkaIlNPPYlKiaakgR	177 NLYsaFValaI-kqLKsGGeLvaItPrsFc	213 FRkqMldecsLnkIh
M.*Taq*I	99 FDlILGNPPYgivgeasK	141 NLYgaFLeKaV-rlLKpGGvLvFVVPatWL	177 LRefLaregktsvYY
DarB	174 FDhVVGNvPFggrdntrn	201 DMgsyFMlRmL-dkIKpGGfMcVIVPpsIV	237 LRlrLsrkaeflgah

Protein	Motif VIII	Number of amino acid residues	Target	Accession number ([S]Swissprot, [G]GenBank)
M.*Bsu*BI	223 rdkLFsaav	501	CTGCAG	P33563[S]
M.*Eco*57I	202 keiWFedtl	540	GTGAAG	P25240[S]
M.*Hinc*II	177 eskVFhgvs	502	GTYRAC	P17744[S]
M.*Pst*I	232 rksaFkasd	507	CTGCAG	P00474[S]
M.*Taq*I	192 lgeVFpqkk	421	TCGA	P14385[S]
DarB	254 ptgTFdang	2255		AAQ13985[G]

Consensus target: ..... TNNA

FIG. 5. Conserved sequence motifs in the predicted amino acid sequences of the N-terminal domain of DarB and of selected γ-type DNA N6-adenine methyltransferases. The motifs are indicated by Roman numerals according to reference 214. Methylated nucleotides in target sequences of DarB homologs are underlined (32, 159, 162, 167, 340).

unclear whether the Hdf protein contributes to P1 antirestriction or has evolved to serve a different function.

In addition to antirestriction and cell lysis, the *darA* operon determines three previously identified functions designated Vad, Gta, and Tsu (reviewed in reference 347, see also reference 149). The corresponding mutants have been identified on the basis of their variable proportions of heads of different sizes, altered generalized transduction, and transduction stimulation by UV, respectively. Assigning these functions to particular genes of the *darA* operon will require further study.

The DarA protein has no homologies to known proteins, other than Hdf. It is synthesized as a precursor and packed into phage particles following proteolytic processing. DarA and DarB are injected into the host cell with the phage DNA, where they act exclusively in *cis*. The Dar proteins do not directly inactivate type I restriction-modification enzymes (155). Moreover P1 DNA isolated from phage particles was not protected against type I restriction in vitro, suggesting that the protection requires entry of phage DNA into cells at infection.

DarB appears to be the largest P1-encoded protein (252 kDa). Analysis of the predicted amino acid sequence of DarB (see Table S1 in the supplemental material) reveals that this gigantic protein consists of at least two domains. The N-terminal domain contains nine appropriately ordered amino acid sequence motifs characteristic of DNA N6-adenine methyltransferases of type "gamma" (Fig. 5) (reviewed in references 44 and 214) and like the majority of them, may methylate adenine residues in a TNNA motif. DarB contains within motif IV the sequence NPPY/F (with one permitted mismatch) known to be essential for adenine DNA methyltransferase catalytic activity. The C-terminal moiety of DarB contains sequence motifs characteristic of DNA or RNA helicases (see Table S1 in the supplemental material). Most likely DarB methylates DNA on infection. The DNA methylase function of DarB is not specific for P1 DNA. Any DNA packed into P1 heads is protected from a DarB-counteracted host restriction system following its injection (155).

Although expressed late in P1 lytic development, *darB* is not preceded by any sequence resembling a late promoter. As the function of *darB* depends on functional *darA*, it is possible that the product of *darA* (or a downstream gene in the same operon) is required for transcriptional activation of *darB*.

(e) Replicative functions expressed late in lytic development. A late promoter that we designate LP*hot* can drive transcription of a gene whose predicted product, an 83-residue, basic polypeptide, appears to be a homolog of the θ subunit of the major replicative enzyme of *E. coli*, DNA polymerase III, whence our gene designation, *hot* (homolog of theta). In the regions of homology, which include 65 amino acid residues of both proteins, 62% of the amino acid residues are identical.

E. coli DNA polymerase III is a complex of at least 10 different protein subunits, which cooperate to ensure the rapid synthesis of DNA with high fidelity (reviewed in reference 168). The θ subunit binds tightly to another subunit, ε (proofreading 3′-5′exonuclease), which, in turn, binds to α (5′-3′ polymerase) in a linear array that forms the catalytic core of PolIII (76, 305). A loop region in θ was identified that directly interacts with the ε subunit (169). It consists of nine amino acid residues (AAAGVAFKE) at position 29 of ε. In the Hot protein of P1 this region is highly conserved (seven identical and

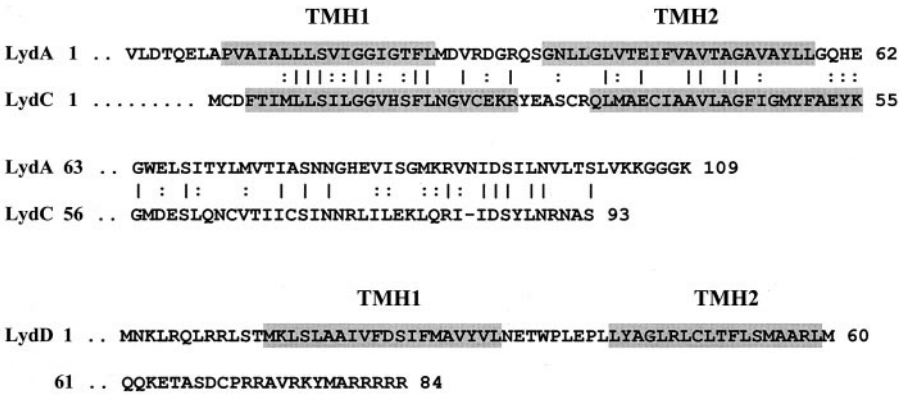

FIG. 6. Known and putative holins encoded by P1. (Top) Alignment of the amino acid sequences of the LydA holin and the putative holin, LydC. Two putative transmembrane helices (TMH) are highlighted in grey. (Bottom) The amino acid sequence of the predicted holin LydD. Putative transmembrane helices are highlighted.

two similar residues), suggesting that Hot can replace *E. coli* θ in the assembly of the *E. coli* core of PolIII, a conclusion recently confirmed (R. Schaaper, personal communication).

Whereas ϵ is known to play several roles, affecting the processivity, thermal stability, and fidelity of PolIII, the function of θ is elusive. There is some evidence to suggest a role in maintaining fidelity of the enzyme via stabilization of the ϵ subunit (76, 305). No mutant phenotype has been associated with a null mutation in *holE*, the structural gene for θ (289).

Upstream of the *hot* gene, in addition to the late promoter LP*hot*, there is a 69-bp region that is 91 to 96% identical to the complement of the regulatory region of the ribonucleotide diphosphate reductase encoded in the *nrd* operon of certain *E. coli* and *Shigella flexneri* strains. The homologous regions include two sequence motifs characteristic of *E. coli* DnaA protein recognition sites (Table 7). Transcription of the *nrd* operon was shown to be regulated by the cooperative binding of DnaA to the upstream sites (19, 313). Whether *hot* may be subjected to regulation by DnaA bound to oppositely oriented DnaA boxes is unknown. Expression of an LP*hot*Φ *(hot'-'lacZ)* in-frame fusion from an insert that, in addition to LP*hot*, contained about 200 bp of upstream P1 DNA, was initiated coordinately with expression of other Lpa-controlled operons upon induction of a P1 lysogen (275). Apparently, Hot synthesis starts late in P1 lytic development, much later than the synthesis of other known P1-encoded replication proteins, which are expressed from genes controlled directly by C1. Hot may possibly compensate for a limitation that appears only well after phage replication has begun or, as a virion protein that is injected into host cells together with P1 DNA, may facilitate early stages of P1 replication.

(f) Lysis. Lysis of the host cell to release phage progeny is the last stage of P1 lytic development. Lysis starts 60 to 70 min after infection under normal laboratory conditions and releases a burst of 100 to 200 infective phage particles. Studies with lysis-defective mutants of P1 have revealed three lytic functions typical of tailed phages: a primary cell-wall-degrading enzyme (endolysin), a helper protein (holin), and its antagonist (antiholin). The function of holins is to form hole-like lesions in the cytoplasmic membrane that provide endolysins access to the peptidoglycan layer of the cell wall (327, 351). Antiholins regulate the timing of lysis. They block function of their cognate holins until a complete set of mature phage particles is ready for release.

Genes for the P1 lytic functions were located previously in two unlinked operons (277, 324) (Fig. 1), both transcriptionally activated by Lpa. Mutations in only one of these genes, *lyz* (formerly *17*), located separately from the two others, prevent lysis. The lytic development of *lyz* mutant phages is normal until the expected lysis time. Then the production of phages ceases and even chloroform treatment of the phage-laden cells cannot induce lysis (324). This phenotype is characteristic of phage mutants with mutations in genes for primary cell-wall-degrading enzymes (reviewed in reference 351). The product of the *lyz* gene was found to be a 185-residue protein homologous to several proteins of the T4 lysozyme family (277), indicating that it is a "true" lysozyme capable of hydrolyzing peptidoglycan N-acetylmuramyl-β(1,4) linkages. However, unlike the lysozyme of T4, P1 Lyz does not require holin for its passage to the periplasm, at least in *E. coli* (341). Its N-terminal domain, which resembles a signal sequence, can mediate export of Lyz to the membrane and its release into the periplasm in a *sec*-dependent manner that is not associated with cleavage of Lyz by SecA.

Two other lysis-associated genes, *lydA* and *lydB* (formerly gene *2*), encode a holin and antiholin, respectively (277, 324). They are the first genes in the *darA* operon (Fig. 1), which also encodes functions not associated with cell lysis. LydA possesses two putative transmembrane domains characteristic of class II holins (327). LydB has no homologs among known proteins. Phages carrying an amber mutation in *lydB* lyse their host cells prematurely, indicating that LydB antagonizes a P1 holin (324).

Analysis of the entire genome sequence of P1 reveals two additional putative holin genes (Fig. 6). One of them, designated by us *lydC*, immediately precedes the promoter-operator region of the *dar* operon (Fig. 1). It can encode a 93-residue protein whose predicted amino acid sequence is 30% identical to that of LydA and features two putative transmembrane helices, a hallmark of class II holins (327). Recent studies confirm the holin function of LydC (R. Young, personal communication). It is likely that either *lydC* evolved from a duplication of the neighboring *lydA* or vice versa.

Although the end of *lydC* is separated from the beginning of *lydA* by a 67-nt noncoding region that contains the late pro-

moter of *lydA* and *lydB* (LP*dar*), there is no predicted Rho-independent terminator in this interval. Conceivably *lydA* and *lydB* and perhaps other genes of the *darA* operon, in addition to being transcribed from the late promoter in front of *lydA*, can also be transcribed from a putative strong promoter that drives transcription of *lydC*.

The regulatory region of *lydC* overlaps the regulatory region of the *cin* gene, which is transcribed in the opposite direction. Some coregulation of the two genes is expected. The −35 and −10 regions of a putative *lydC* promoter (Table 4) overlap the sequence that acts as an enhancer of Cin-mediated C-segment inversion and bind the Fis protein. The region between the predicted *lydC* promoter and the beginning of *lydC* also contains a binding site for the Cin protein (*cix*Pp; Table 1) and a putative binding site for the host protein IHF (Table 7). Occupation of these sites by Cin and IHF, respectively, could possibly repress the expression of *lydC* (as well as *cin*). Another factor that could limit the expression of *lydC* is a putative strong Rho-independent terminator at the beginning of the *lydC* coding sequence (Table 5). It is likely that transcription of *lydC* terminates prematurely in the absence of an efficient antitermination mechanism. Since *lydC* appears subject to multiple levels of repression, it is probably expressed at low levels. Expression of *lydA* and *lydB* from the late promoter LP*dar*, downstream of *lydC*, is expected to make LydA predominant over LydC late in lytic development. It is likely that LydB antiholin can control the time of lysis mediated by the related LydA and LydC proteins.

A second predicted holin gene, which we designate *lydD*, is apparently transcribed from the same promoter, P*lyz*, as the upstream gene *lyz*, which partially overlaps it. It can encode an 84-residue protein, which contains two putative transmembrane helices, like LydA and LydC. LydD contains a cluster of five highly basic (arginine) residues at its C-terminal end, another hallmark of phage holins (327). LydD may be the primary holin involved in the P1-mediated lysis of *E. coli* cells, as suggested by the inability of a *lyz* gene clone to complement fully a *lyz* amber mutant, unless *lydD* was part of the clone (59). Polarity of the amber *lyz* mutation, expected from an overlap of *lyz* and *lydD*, could explain this result. It is likely that, in addition to regulation by Lpa, *lyz* and *lydD* expression depends on RNA antisense to the proximal region of the *lyz* gene and is terminated in that region by a predicted Rho-independent terminator (Table 5).

In the central region of *lydD* starts yet another gene that is oriented in the same direction and is read in a different frame (Fig. 1). It can encode a 71-residue protein, which is likely to be an antiholin antagonistic to LydD and which we designate *lydE*. As a general rule, a phage holin gene is associated with a gene encoding its specific antagonist (327). The most frequent organization involves a "dual start motif"; the holin genes encode two polypeptides, read in the same frame, that differ in length and have antagonistic functions (holin and antiholin) (103). In P1, genes for known and predicted holins and their specific antiholins are either closely linked (*lydA* and *lydB*) or overlapping (*lydD* and *lydE*) but their products differ significantly. Thus, the P1 lysis machinery appears atypical in its complexity and organization.

ACKNOWLEDGMENTS

We are deeply indebted to the many members of phage community for their helpful comments on a rough draft of the manuscript and for their encouragement. We owe particular thanks to colleagues who have reviewed the initial draft in its entirety— Hans W. Ackermann, Richard D'Ari, David Lane, and Erich Miller—and to Sherwood Casjens and Roger Hendrix for suggestions concerning phage morphogenesis. We thank Judah L. Rosner for providing the N1706 strain from which P1 was induced, Włodzimierz Zagórski-Ostoja for crucial support of the project at the IBB PAS, and Sankar Adhya and Wacław Szybalski for their help in initiating this collaborative effort.

The main part of this work, including sequencing of P1 *c*1-100 *mod749*::IS*5*, was supported by grant Nr6P04B01116 from the Polish State Committee for Scientific Research (KBN). Sequencing of the second P1 strain, P1 *c*1-100 *mod1902*::IS*5* *dmt*ΔMB *rev*-6, was supported by NIH grants P01HG01428 and S10RR10379.

REFERENCES

1. **Abeles, A., T. Brendler, and S. Austin.** 1993. Evidence of two levels of control of P1 *oriR* and host *oriC* replication origins by DNA adenine methylation. J. Bacteriol. **175:**7801–7807.
2. **Abeles, A. L.** 1986. P1 plasmid replication. Purification and DNA-binding activity of the replication protein RepA. J. Biol. Chem. **261:**3548–3555.
3. **Abeles, A. L., and S. J. Austin.** 1991. Antiparallel plasmid-plasmid pairing may control P1 plasmid replication. Proc. Natl. Acad. Sci. USA **88:**9011–9015.
4. **Abeles, A. L., S. A. Friedman, and S. J. Austin.** 1985. Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. J. Mol. Biol. **185:**261–272.
5. **Abeles, A. L., L. D. Reaves, and S. J. Austin.** 1989. Protein-DNA interactions in regulation of P1 plasmid replication. J. Bacteriol. **171:**43–52.
6. **Abeles, A. L., K. M. Snyder, and D. K. Chattoraj.** 1984. P1 plasmid replication: replicon structure. J. Mol. Biol. **173:**307–324.
7. **Abremski, K., and R. Hoess.** 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. J. Biol. Chem. **259:**1509–1514.
8. **Abremski, K., R. Hoess, and N. Sternberg.** 1983. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. Cell **32:**1301–1311.
9. **Abuladze, N. K., M. Gingery, J. Tsai, and F. A. Eiserling.** 1994. Tail length determination in bacteriophage T4. Virology **199:**301–310.
10. **Adams, D. E., J. B. Bliska, and N. Cozzarelli.** 1992. Cre-*lox* recombination in *Escherichia coli* cells: mechanistic differences from the *in vitro* reaction. J. Mol. Biol. **226:**661–673.
11. **Ali Azam, T., A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama.** 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. J. Bacteriol. **181:**6361–6370.
12. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
13. **Anantharaman, V., and L. Aravind.** 2003. New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. Genome Biol. **4:**R81.
14. **Anderson, T. F., and D. H. J. Walker.** 1960. Morphological variants of the bacteriophage P1. Science **132:**1488.
15. **Arber, W.** 1991. Elements in microbial evolution. J. Mol. Evol. **33:**4–12.
16. **Arber, W., and D. Dussoix.** 1962. Host specificity of DNA produced by *Escherichia coli.* I. Host controlled modification of bacteriophage λ. J. Mol. Biol. **5:**18–36.
17. **Arisaka, F., S. Kanamaru, P. Leiman, and M. G. Rossmann.** 2003. The tail lysozyme complex of bacteriophage T4. Int. J. Biochem. Cell Biol. **35:**16–21.
18. **Artsimovitch, I., K. Murakami, A. Ishihama, and M. M. Howe.** 1996. Transcription activation by the bacteriophage Mu Mor protein requires the C-terminal regions of both alpha and sigma70 subunits of *Escherichia coli* RNA polymerase. J. Biol. Chem. **271:**32343–32348.
19. **Augustin, L. B., B. A. Jacobson, and J. A. Fuchs.** 1994. *Escherichia coli* Fis and DnaA proteins bind specifically to the *nrd* promoter region and affect expression of an *nrd-lac* fusion. J. Bacteriol. **176:**378–387.
20. **Austin, S., and A. Abeles.** 1983. Partition of unit-copy miniplasmids to daughter cells. II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. J. Mol. Biol. **169:**373–387.
21. **Austin, S., M. Ziese, and N. Sternberg.** 1981. A novel role for site-specific recombination in maintenance of bacterial replicons. Cell **25:**729–736.
22. **Bächi, B., and W. Arber.** 1977. Physical mapping of *Bgl*II, *Bam*HI, *Eco*RI, *Hin*dIII and *Pst*I restriction fragments of bacteriophage P1 DNA. Mol. Gen. Genet. **153:**311–324.
23. **Bächi, B., J. Reiser, and V. Pirrotta.** 1979. Methylation and cleavage se-

quences of the *Eco*P1 restriction-modification enzyme. J. Mol. Biol. **128:** 143–163.
24. **Barik, S.** 1993. Expression and biochemical properties of a protein serine/threonine phosphatase encoded by bacteriophage lambda. Proc. Natl. Acad. Sci. USA **90:**10633–10637.
25. **Baumstark, B. R., S. R. Stovall, and P. Bralley.** 1990. The ImmC region of phage Pl codes for a gene whose product promotes lytic growth. Virology **179:**217–227.
26. **Baumstark, B. R., and J. R. Scott.** 1987. The *c4* gene of phage P1. Virology **156:**197–203.
27. **Baumstark, B. R., S. R. Stovall, and S. Ashkar.** 1987. Interaction of the Plcl repressor with Pl DNA. Localization of repressor binding sites near the *c*l gene. Virology **156:**404–413.
28. **Bendtsen, J. D., A. S. Nilsson, and H. Lehnherr.** 2002. Phylogenetic and functional analysis of the bacteriophage P1 single-stranded DNA-binding protein. J. Virol. **76:**9695–9701.
29. **Bertani, G.** 1958. Lysogeny. Adv. Virus Res. **5:**151–193.
30. **Bertani, G.** 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J. Bacteriol. **186:**595–600.
31. **Bertani, G.** 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J. Bacteriol. **62:**293–300.
32. **Bickle, T. A., and D. H. Krüger.** 1993. Biology of DNA restriction. Microbiol. Rev. **57:**434–450.
33. **Biere, A. L., M. Citron, and H. Schuster.** 1992. Transcriptional control via translational repression by c4 antisense RNA of bacteriophages P1 and P7. Genes Dev. **6:**2409–2416.
34. **Black, L., M. Showe, and A. Steven.** 1994. Morphogenesis of the T4 head., p. 218–258. *In* J. D. Karam (ed.), Molecular biology of bacteriophage T4. ASM Press, Washington, D.C.
35. **Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1474.
36. **Borodovsky, M., and J. McIninch.** 1993. Recognition of genes in DNA sequence with ambiguities. Biosystems **30:**161–171.
37. **Bouet, J. Y., and B. E. Funnell.** 1999. P1 ParA interacts with the P1 partition complex at *parS* and an ATP-ADP switch controls ParA activities. EMBO J. **18:**1415–1424.
38. **Bourniquel, A. A., and T. A. Bickle.** 2002. Complex restriction enzymes: NTP-driven molecular motors. Biochimie **84:**1047–1059.
39. **Braithwaite, D. K., and J. Ito.** 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. **21:**787–802.
40. **Brendel, V., and E. N. Trifonov.** 1984. A computer algorithm testing potential prokaryotic terminators. Nucleic Acids Res. **12:**4411–4427.
41. **Brockes, J. P., P. R. Brown, and K. Murray.** 1972. The deoxyribonucleic acid modification enzyme of bacteriophage P1. Biochem. J. **127:**1–10.
42. **Brockes, J. P., P. R. Brown, and K. Murray.** 1974. Nucleotide sequences at the sites of action of the deoxyribonucleic acid modification enzyme of bacteriophage P1. J. Mol. Biol. **88:**437–443.
43. **Bruck, I., R. Woodgate, K. McEntee, and M. F. Goodman.** 1996. Purification of a soluble UmuD′C complex from *Escherichia coli.* Cooperative binding of UmuD′C to single-stranded DNA. J. Biol. Chem. **271:**10767–10774.
44. **Bujnicki, J. M.** 2001. Understanding the evolution of restriction-modification systems: clues from sequence and structure comparisons. Acta Biochim. Pol. **48:**935–967.
45. **Burland, V., D. L. Daniels, G. Plunkett III, and F. R. Blattner.** 1993. Genome sequencing on both strands: the Janus strategy. Nucleic Acids Res. **21:**3385–3390.
46. **Campbell, A.** 2003. Prophage insertion sites. Res. Microbiol. **154:**277–282.
47. **Capiaux, H., C. Lesterlin, K. Perals, J. M. Louarn, and F. Cornet.** 2002. A dual role for the FtsK protein in *Escherichia coli* chromosome segregation. EMBO Rep. **3:**532–536.
48. **Cerritelli, M. E., J. S. Wall, M. N. Simon, J. F. Conway, and A. C. Steven.** 1996. Stoichiometry and domainal organization of the long tail-fiber of bacteriophage T4: a hinged viral adhesin. J. Mol. Biol. **260:**767–780.
49. **Chattoraj, D. K.** 2000. Control of plasmid DNA replication by iterons: no longer paradoxical. Mol. Microbiol. **37:**467–476.
50. **Chattoraj, D. K., R. J. Mason, and S. H. Wickner.** 1988. Mini-P1 plasmid replication: the autoregulation-sequestration paradox. Cell **52:**551–557.
51. **Chattoraj, D. K., and T. D. Schneider.** 1997. Replication control of plasmid P1 and its host chromosome: the common ground. Prog. Nucleic Acid Res. Mol. Biol. **57:**145–186.
52. **Chattoraj, D. K., K. M. Snyder, and A. L. Abeles.** 1985. P1 plasmid replication: multiple functions of RepA protein at the origin. Proc. Natl. Acad. Sci. USA **82:**2588–2592.
53. **Cheng, S. W., D. L. Court, and D. I. Friedman.** 1995. Transcription termination signals in the nin region of bacteriophage lambda: identification of Rho-dependent termination regions. Genetics **140:**875–887.
54. **Chesney, R. H., and J. R. Scott.** 1975. Superinfection immunity and prophage repression in phage P1. II. Mapping of the immunity-difference and ampicillin-resistance loci of P1 and phi amp. Virology **67:**375–384.
55. **Chiang, L. W., and M. M. Howe.** 1993. Mutational analysis of a C-dependent late promoter of bacteriophage Mu. Genetics **135:**619–629.
56. **Christie, G. E., L. M. Temple, B. A. Bartlett, and T. S. Goodwin.** 2002. Programmed translational frameshift in the bacteriophage P2 FETUD tail gene operon. J. Bacteriol. **184:**6522–6531.
57. **Citron, M., and H. Schuster.** 1992. The c4 repressor of bacteriophage P1 is a processed 77 base antisense RNA. Nucleic Acids Res. **20:**3085–3090.
58. **Citron, M., and H. Schuster.** 1990. The c4 repressors of bacteriophages Pl and P7 are antisense RNAs. Cell **62:**591–598.
59. **Citron, M., M. Velleman, and H. Schuster.** 1989. Three additional operators, Op21, Op68, and Op88, of bacteriophage P1. Evidence for the control of the P1 dam methylase by Op68. J. Biol. Chem. **264:**3611–3617.
60. **Cohen, G.** 1983. Electron microscopy study of early lytic replication forms of bacteriophage P1 DNA. Virology **131:**159–170.
61. **Cohen, G., E. Or, W. Minas, and N. L. Sternberg.** 1996. The bacteriophage P1 lytic replicon: directionality of replication and *cis*-acting elements. Gene **175:**151–155.
62. **Cohen, P. T., and P. Cohen.** 1989. Discovery of a protein phosphatase activity encoded in the genome of bacteriophage lambda. Probable identity with open reading frame 221. Biochem. J. **260:**931–934.
63. **Coombs, D. H., and F. Arisaka.** 1994. T4 tail structure and function., p. 259–281. *In* J. D. Karam (ed.), Molecular biology of bacteriophage T4. ASM Press, Washington, D.C.
64. **Corre, J., and J. M. Louarn.** 2002. Evidence from terminal recombination gradients that FtsK uses replichore polarity to control chromosome terminus positioning at division in *Escherichia coli*. J. Bacteriol. **184:**3801–3807.
65. **Costantino, N., M. Zuber, and D. Court.** 1990. Analysis of mutations in the *ninR* region of bacteriophage lambda that bypass a requirement for lambda N antitermination. J. Bacteriol. **172:**4610–4615.
66. **Coulby, J. N., and N. L. Sternberg.** 1988. Characterization of the phage P1 *dam* gene. Gene **74:**191.
67. **D'Ari, R., A. Jaffé-Brachet, D. Touati-Schwartz, and M. B. Yarmolinsky.** 1975. A *dnaB* analog specified by bacteriophage P1. J. Mol. Biol. **94:**341–366.
68. **Dartois, V., O. De Backer, and C. Colson.** 1993. Sequence of the *Salmonella typhimurium* StyLT1 restriction-modification genes: homologies with *Eco*P1 and *Eco*P15 type-III R-M systems and presence of helicase domains. Gene **127:**105–110.
69. **d'Aubenton Carafa, Y., E. Brody, and C. Thermes.** 1990. Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. J. Mol. Biol. **216:**835–858.
70. **Davey, M. J., and B. E. Funnell.** 1997. Modulation of the P1 plasmid partition protein ParA by ATP, ADP, and P1 ParB. J. Biol. Chem. **272:** 15286–15292.
71. **Davey, M. J., and B. E. Funnell.** 1994. The P1 plasmid partition protein ParA. A role for ATP in site-specific DNA binding. J. Biol. Chem. **269:** 29908–29913.
72. **Davis, M. A., and S. J. Austin.** 1988. Recognition of the P1 plasmid centromere analog involves binding of the ParB protein and is modified by a specific host factor. EMBO J. **7:**1881–1888.
73. **Davis, M. A., K. A. Martin, and S. J. Austin.** 1992. Biochemical activities of the ParA partition protein of the P1 plasmid. Mol. Microbiol. **6:**1141–1147.
74. **Delarue, M., O. Poch, N. Tordo, D. Moras, and P. Argos.** 1990. An attempt to unify the structure of polymerases. Protein Eng. **3:**461–467.
75. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. **27:**4636–4641.
76. **DeRose, E. F., T. Darden, S. Harvey, S. Gabel, F. W. Perrino, R. M. Schaaper, and R. E. London.** 2003. Elucidation of the epsilon-theta subunit interface of *Escherichia coli* DNA polymerase III by NMR spectroscopy. Biochemistry **42:**3635–3644.
77. **Devlin, B. H., B. R. Baumstark, and J. R. Scott.** 1982. Superimmunity: characterization of a new gene in the immunity region of P1. Virology **120:** 360–375.
78. **Dodd, I. B., and J. B. Egan.** 1990. Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res. **18:**5019–5026.
79. **Dreiseikelmann, B., M. Velleman, and H. Schuster.** 1988. The c1 repressor of bacteriophage P1. Isolation and characterization of the repressor protein. J. Biol. Chem. **263:**1391–1397.
80. **Ebersbach, G., and K. Gerdes.** 2001. The double *par* locus of virulence factor pB171: DNA segregation is correlated with oscillation of ParA. Proc. Natl. Acad. Sci. USA **98:**15078–15083.
81. **Eddy, S. R., and R. Durbin.** 1994. RNA sequence analysis using covariance models. Nucleic Acids Res. **22:**2079–2088.
82. **Edelbluth, C., E. Lanka, W. von der Hude, M. Mikolajczyk, and H. Schuster.** 1979. Association of the prophage P1ban protein with the dnaB protein of *Escherichia coli.* Overproduction of ban protein by a P1*bac crr* mutant. Eur. J. Biochem. **94:**427–435.

83. **Edgar, R., D. K. Chattoraj, and M. Yarmolinsky.** 2001. Pairing of P1 plasmid partition sites by ParB. Mol. Microbiol. **42:**1363–1370.
84. **Eliason, J. L., and N. Sternberg.** 1987. Characterization of the binding sites of cl repressor of bacteriophage Pl: evidence for multiple asymmetric sites. J. Mol. Biol. **198:**281–293.
85. **Ermolaeva, M. D., H. G. Khalak, O. White, H. O. Smith, and S. L. Salzberg.** 2000. Prediction of transcription terminators in bacterial genomes. J. Mol. Biol. **301:**27–33.
86. **Ferguson, P. L., and D. H. Coombs.** 2000. Pulse-chase analysis of the in vivo assembly of the bacteriophage T4 tail. J. Mol. Biol. **297:**99–117.
87. **Friedman, S. A., and S. J. Austin.** 1988. The P1 plasmid-partition system synthesizes two essential proteins from an autoregulated operon. Plasmid **19:**103–112.
88. **Fukada, K., and J. Abelson.** 1980. DNA sequence of a T4 transfer RNA gene cluster. J. Mol. Biol. **139:**377–391.
89. **Funnell, B. E.** 1988. Mini-P1 plasmid partitioning: excess ParB protein destabilizes plasmids containing the centromere *parS*. J. Bacteriol. **170:**954–960.
90. **Funnell, B. E.** 1991. The P1 plasmid partition complex at *parS.* The influence of *Escherichia coli* integration host factor and of substrate topology. J. Biol. Chem. **266:**14328–14337.
91. **Funnell, B. E.** 1988. Participation of *Escherichia coli* integration host factor in the P1 plasmid partition system. Proc. Natl. Acad. Sci. USA **85:**6657–6661.
92. **Funnell, B. E., and L. Gagnier.** 1993. The P1 plasmid partition complex at *parS.* II. Analysis of ParB protein binding activity and specificity. J. Biol. Chem. **268:**3616–3624.
93. **Gamas, P., M. G. Chandler, P. Prentki, and D. J. Galas.** 1987. *Escherichia coli* integration host factor binds specifically to the ends of the insertion sequence IS*1* and to its major insertion hot-spot in pBR322. J. Mol. Biol. **195:**261–272.
94. **Gamas, P., D. Galas, and M. Chandler.** 1985. DNA sequence at the end of IS*1* required for transposition. Nature **317:**458–460.
95. **Gazit, E., and R. T. Sauer.** 1999. The Doc toxin and Phd antidote proteins of the bacteriophage P1 plasmid addiction system form a heterotrimeric complex. J. Biol. Chem. **274:**16813–16818.
96. **Gazit, E., and R. T. Sauer.** 1999. Stability and DNA binding of the phd protein of the phage P1 plasmid addiction system. J. Biol. Chem. **274:**2652–2657.
97. **Gerdes, K., J. Møller-Jensen, G. Ebersbach, T. Kruse, and K. Nordström.** 2004. Bacterial mitotic machineries. Cell **116:**359–366.
98. **Ghisotti, D., F. Briani, F. Forti, F. Piazza, S. Polo, P. Sabbattini, T. Sturniolo, S. Terzano, S. Zangrossi, M. Zappone, G. Sironi, and G. Dehò.** 1995. Multiple regulatory mechanisms controlling phage-plasmid P4 propagation. FEMS Microbiol. Rev. **17:**127–134.
99. **Goldman, E., A. H. Rosenberg, G. Zubay, and F. W. Studier.** 1995. Consecutive low-usage leucine codons block translation only when near the 5′ end of a message in *Escherichia coli*. J. Mol. Biol. **245:**467–473.
100. **Gottesman, S., W. P. Clark, V. de Crecy-Lagard, and M. R. Maurizi.** 1993. ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli.* Sequence and *in vivo* activities. J. Biol. Chem. **268:**22618–22626.
101. **Grigoriev, A.** 1999. Strand-specific compositional asymmetries in double-stranded DNA viruses. Virus Res. **60:**1–19.
102. **Grigoriev, P., and M. B. Łobocka.** 2001. Determinants of segregational stability of the linear plasmid prophage N15 of *Escherichia coli*. Mol. Microbiol. **42:**355–368.
103. **Gründling, A., M. D. Manson, and R. Young.** 2001. Holins kill without warning. Proc. Natl. Acad. Sci. USA **98:**9348–9352.
104. **Guidolin, A., J. M. Zingg, and W. Arber.** 1989. Organization of the bacteriophage P1 tail-fibre operon. Gene **76:**239–243.
105. **Guidolin, A., J. M. Zingg, H. Lehnherr, and W. Arber.** 1989. Bacteriophage P1 tail-fibre and *dar* operons are expressed from homologous phage-specific late promoter sequences. J. Mol. Biol. **208:**615–622.
106. **Guo, F., D. N. Gopaul, and G. D. van Duyne.** 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. Nature **389:**40–46.
107. **Hadi, S. M., B. Bächi, S. Iida, and T. A. Bickle.** 1983. DNA restriction-modification enzymes of phage P1 and plasmid p15B. Subunit functions and structural homologies. J. Mol. Biol. **165:**19–34.
108. **Haffter, P., and T. A. Bickle.** 1988. Enhancer-independent mutants of the Cin recombinase have a relaxed topological specificity. EMBO J. **7:**3991–3996.
109. **Haffter, P., and T. A. Bickle.** 1987. Purification and DNA-binding properties of FIS and Cin, two proteins required for the bacteriophage P1 site-specific recombination system, cin. J. Mol. Biol. **198:**579–587.
110. **Hambly, E., F. Tetart, C. Desplats, W. H. Wilson, H. M. Krisch, and N. H. Mann.** 2001. A conserved genetic module that encodes the major virion components in both the coliphage T4 and the marine cyanophage S-PM2. Proc. Natl. Acad. Sci. USA **98:**11411–11416.
111. **Hansen, A.-M., H. Lehnherr, X. Wang, V. Mobley, and D. J. Jin.** 2003. *Escherichia coli* SspA is a transcription activator for bacteriophage P1 late genes. Mol. Microbiol. **48:**1621–1631.
112. **Hansen, E. B.** 1989. Structure and regulation of the lytic replicon of phage Pl. J. Mol. Biol. **207:**135–149.
113. **Hansen, E. B., and M. B. Yarmolinsky.** 1986. Host participation in plasmid maintenance: dependence upon DnaA of replicons derived from P1 and F. Proc. Natl. Acad. Sci. USA **83:**4423–4427.
114. **Hao, J. J., and M. Yarmolinsky.** 2002. Effects of the P1 plasmid centromere on expression of P1 partition genes. J. Bacteriol. **184:**4857–4867.
115. **Hartmann, R. K., J. Heinrich, J. Schlegl, and H. Schuster.** 1995. Precursor of C4 antisense RNA of bacteriophages P1 and P7 is a substrate for RNase P of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **92:**5822–5826.
116. **Hasegawa, T., M. Miyano, H. Himeno, Y. Sano, K. Kimura, and M. Shimizu.** 1992. Identity determinants of *E. coli* threonine tRNA. Biochem. Biophys. Res. Commun. **184:**478–484.
117. **Hashemolhosseini, S., Y. D. Stierhof, I. Hindennach, and U. Henning.** 1996. Characterization of the helper proteins for the assembly of tail fibers of coliphages T4 and lambda. J. Bacteriol. **178:**6258–6265.
118. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237–2255.
119. **Hay, N., and G. Cohen.** 1983. Requirement of *E. coli* DNA synthesis functions for the lytic replication of bacteriophage P1. Virology **131:**193–206.
120. **Hayes, F., L. Radnedge, M. A. Davis, and S. J. Austin.** 1994. The homologous operons for P1 and P7 plasmid partition are autoregulated from dissimilar operator sites. Mol. Microbiol. **11:**249–260.
121. **Hazan, R., B. Sat, M. Reches, and H. Engelberg-Kulka.** 2001. Postsegregational killing mediated by the P1 phage "addiction module" *phd-doc* requires the *Escherichia coli* programmed cell death system *mazEF*. J. Bacteriol. **183:**2046–2050.
122. **Heilmann, H., J. N. Reeve, and A. Pühler.** 1980. Identification of the repressor and repressor bypass (antirepressor) polypeptides of bacteriophage P1 synthesized in infected minicells. Mol. Gen. Genet. **178:**149–154.
123. **Heinrich, J., M. Citron, A.Günther, and H. Schuster.** 1994. Second-site suppressors of the bacteriophage P1 *vir*^{s} mutant reveal the interdependence of the *c*4, *icd*, and *ant* genes in the P1 *imm*I operon. J. Bacteriol. **176:**4931–4936.
124. **Heinrich, J., H. D. Riedel, B. R. Baumstark, M. Kimura, and H. Schuster.** 1989. The c1 repressor of bacteriophage P1 operator-repressor interaction of wild-type and mutant repressor proteins. Nucleic Acids Res. **17:**7681–7692.
125. **Heinrich, J., H. D. Riedel, B. Rückert, R. Lurz, and H. Schuster.** 1995. The lytic replicon of bacteriophage P1 is controlled by an antisense RNA. Nucleic Acids Res. **23:**1468–1474.
126. **Heinrich, J., M. Velleman, and H. Schuster.** 1995. The tripartite immunity system of phages P1 and P7. FEMS Microbiol. Rev. **17:**121–126.
127. **Heinzel, T., M. Velleman, and H. Schuster.** 1990. The cl repressor inactivator protein coi of bacteriophage Pl. Cloning and expression of coi and its interference with cl repressor function. J. Biol. Chem. **265:**17928–17934.
128. **Heinzel, T., R. Lurz, B. Dobrinsky, M. Velleman, and H. Schuster.** 1994. C1 repressor-mediated looping is involved in C1 autoregulation of bacteriophage P1. J. Biol. Chem. **269:**31885–31890.
129. **Heinzel, T., M. Velleman, and H. Schuster.** 1989. *ban* operon of bacteriophage P1. Mutational analysis of the c1 repressor-controlled operator. J. Mol. Biol. **205:**127–135.
130. **Heinzel, T., M. Velleman, and H. Schuster.** 1992. Cl repressor of phage Pl is inactivated by non-covalent binding of Pl Coi protein. J. Biol. Chem. **267:**4183–4188.
131. **Heisig, A., H.-D. Riedel, B. Dobrinski, R. Lurz, and H. Schuster.** 1989. Organization of the immunity region immI of bacteriophage Pl and synthesis of the Pl antirepressor. J. Mol. Biol. **209:**525–538.
132. **Heisig, A., I. Severin, A. K. Seefluth, and H. Schuster.** 1987. Regulation of the *ban* gene containing operon of prophage P1. Mol. Gen. Genet. **206:**368–376.
133. **Hendrix, R. W., and R. L. Duda.** 1992. Bacteriophage lambda PaPa: not the mother of all lambda phages. Science **258:**1145–1148.
134. **Henning, U., and S. Hashemolhosseini.** 1994. Receptor recognition by T-even-type coliphages, p. 182–185. *In* J. D. Karam (ed.), Molecular biology of bacteriophage T4. ASM Press, Washington, D.C.
135. **Hiestand-Nauer, R., and S. Iida.** 1983. Sequence of the site-specific recombinase gene *cin* and of its substrates serving in the inversion of the C segment of bacteriophage P1. EMBO J. **2:**1733–1740.
136. **Hill, T. M.** 1996. Features of the chromosomal terminus region., p. 1602–1614. *In* F. C. Neidhardt et al. (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 2. ASM Press, Washington, D.C.
137. **Hochman, L., N. Segev, N. Sternberg, and G. Cohen.** 1983. Site-specific recombinational circularization of bacteriophage P1 DNA. Virology **131:**11–17.
138. **Hoess, R. H., M. Ziese, and N. Sternberg.** 1982. P1 site-specific recombination: nucleotide sequence of the recombining sites. Proc. Natl. Acad. Sci. USA **79:**3398–3402.
139. **Hofer, B., M. Ruge, and B. Dreiseikelmann.** 1995. The superinfection ex-

clusion gene (*sieA*) of bacteriophage P22: identification and overexpression of the gene and localization of the gene product. J. Bacteriol. **177:**3080–3086.
140. **Holcik, M., and V. N. Iyer.** 1997. Conditionally lethal genes associated with bacterial plasmids. Microbiology **143:**3403–3416.
141. **Huber, H. E., S. Iida, W. Arber, and T. A. Bickle.** 1985. Site specific DNA inversion is enhanced by a specific DNA element in *cis*. Proc. Natl. Acad. Sci. USA. **82:**3776–3780.
142. **Huber, H. E., S. Iida, and T. A. Bickle.** 1985. Expression of the bacteriophage P1 *cin* recombinase gene from its own and heterologous promoters. Gene **34:**63–72.
143. **Hübner, P., and W. Arber.** 1989. Mutational analysis of a prokaryotic recombinational enhancer element with two functions. EMBO J. **8:**577–585.
144. **Hübner, P., P. Haffter, S. Iida, and W. Arber.** 1989. Bent DNA is needed for recombinational enhancer activity in the site-specific recombination system Cin of bacteriophage P1. The role of FIS protein. J. Mol. Biol. **205:**493–500.
145. **Hümbelin, M., B. Suri, D. N. Rao, D. P. Hornby, H. Eberle, T. Pripfl, S. Kenel, and T. A. Bickle.** 1988. Type III DNA restriction and modification systems *Eco*P1 and *Eco*P15. Nucleotide sequence of the *Eco*P1 operon, the *Eco*P15 *mod* gene and some *Eco*P1 *mod* mutants. J. Mol. Biol. **200:**23–29.
146. **Iida, S.** 1984. Bacteriophage P1 carries two related sets of genes determining its host range in the invertible C segment of its genome. Virology **134:**421–434.
147. **Iida, S., and W. Arber.** 1977. Plaque forming specialized transducing phage P1: isolation of P1CmSmSu, a precursor of P1Cm. Mol. Gen. Genet. **153:**259–269.
148. **Iida, S., and R. Hiestand-Nauer.** 1987. Role of the central dinucleotide at the crossover sites for the selection of quasi sites in DNA inversion mediated by the site-specific Cin recombinase of phage P1. Mol. Gen. Genet. **208:**464–468.
149. **Iida, S., R. Hiestand-Nauer, H. Sandmeier, H. Lehnherr, and W. Arber.** 1998. Accessory genes in the *darA* operon of bacteriophage P1 affect antirestriction function, generalized transduction, head morphogenesis, and host cell lysis. Virology **251:**49–58.
150. **Iida, S., H. Huber, R. Hiestand-Nauer, J. Meyer, T. A. Bickle, and W. Arber.** 1984. The bacteriophage P1 site-specific recombinase cin: recombination events and DNA recognition sequences. Cold Spring Harbor Symp. Quant. Biol. **49:**769–777.
151. **Iida, S., J. Meyer, and W. Arber.** 1978. The insertion element IS*1* is a natural constituent of coliphage P1 DNA. Plasmid **1:**357–365.
152. **Iida, S., J. Meyer, B. Bächi, M. Stålhammar-Carlemalm, S. Schrickel, T. A. Bickle, and W. Arber.** 1983. DNA restriction-modification genes of phage P1 and plasmid p15B. Structure and *in vitro* transcription. J. Mol. Biol. **165:**1–18.
153. **Iida, S., J. Meyer, K. E. Kennedy, and W. Arber.** 1982. A site-specific, conservative recombination system carried by bacteriophage P1. Mapping the recombinase gene *cin* and the cross-over sites *cix* for the inversion of the C segment. EMBO J. **1:**1445–1453.
154. **Iida, S., H. Sandmeier, P. Hübner, R. Hiestand-Nauer, K. Schneitz, and W. Arber.** 1990. The Min DNA inversion enzyme of plasmid p15B of *Escherichia coli* 15T-: a new member of the Din family of site-specific recombinases. Mol. Microbiol. **4:**991–997.
155. **Iida, S., M. B. Streiff, T. A. Bickle, and W. Arber.** 1987. Two DNA antirestriction systems of bacteriophage P1, *darA* and *darB*: characterization of *darA*− phages. Virology **157:**156–166.
156. **Ikeda, H., and J.-I. Tomizawa.** 1965. Transducing fragments in generalized transduction by phage P1. III. Studies with small phage particles. J. Mol. Biol. **14:**120–129.
157. **Ireton, K., N. W. Günther, and A. D. Grossman.** 1994. Spo0J is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*. J. Bacteriol. **176:**5320–5329.
158. **Ishihama, A., and T. Saitoh.** 1979. Subunits of RNA polymerase in function and structure. IX. Regulation of RNA polymerase activity by stringent starvation protein (SSP). J. Mol. Biol. **129:**517–530.
159. **Ito, H., A. Sadaoka, H. Kotani, N. Hiraoka, and T. Nakamura.** 1990. Cloning, nucleotide sequence, and expression of the HincII restriction-modification system. Nucleic Acids Res. **18:**3903–3911.
160. **Ito, J., and D. K. Braithwaite.** 1991. Compilation and alignment of DNA polymerase sequences. Nucleic Acids Res. **19:**4045–4057.
161. **Janscak, P., U. Sandmeier, M. D. Szczelkun, and T. A. Bickle.** 2001. Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases *Eco*P1I and *Eco*P15I. J. Mol. Biol. **306:**417–431.
162. **Janulaitis, A., M. Petrusyte, Z. Maneliene, S. Klimasauskas, and V. Butkus.** 1992. Purification and properties of the *Eco*57I restriction endonuclease and methylase-prototypes of a new class (type IV). Nucleic Acids Res. **20:**6043–6049.
163. **Jobling, M. G., and D. A. Ritchie.** 1988. The nucleotide sequence of a plasmid determinant for resistance to tellurium anions. Gene **66:**245–258.
164. **Johnson, B. F.** 1982. Suppression of the *lexC* (*ssbA*) mutation of *Escherichia coli* by a mutant of bacteriophage P1. Mol. Gen. Genet. **186:**122–126.
165. **Kanamaru, S., P. G. Leiman, V. A. Kostyuchenko, P. R. Chipman, V. V. Mesyanzhinov, F. Arisaka, and M. G. Rossmann.** 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature **415:**553–557.
166. **Kane, J. F.** 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. Curr. Opin. Biotechnol. **6:**494–500.
167. **Kawakami, B., H. Christophe, M. Nagatomo, and M. Oka.** 1991. Cloning and nucleotide sequences of the *Acc*I restriction-modification genes in *Acinetobacter calcoaceticus*. Agric. Biol. Chem. **55:**1553–1559.
168. **Kelman, Z., and M. O'Donnell.** 1995. DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. Annu. Rev. Biochem. **64:**171–200.
169. **Keniry, M. A., H. A. Berthon, J. Y. Yang, C. S. Miles, and N. E. Dixon.** 2000. NMR solution structure of the theta subunit of DNA polymerase III from *Escherichia coli*. Protein Sci. **9:**721–733.
170. **Kliem, M., and B. Dreiseikelmann.** 1989. The superimmunity gene *sim* of bacteriophage P1 causes superinfection exclusion. Virology **171:**350–355.
171. **Kondo, E., and S. Mitsuhashi.** 1964. Drug resistance of enteric bacteria. IV. Active transducing bacteriophage P1CM produced by the combination of R-factor with bacteriophage P1. J. Bacteriol. **88:**1266–1276.
172. **Koonin, E. V.** 1993. Bacterial and bacteriophage protein phosphatases. Mol. Microbiol. **8:**785–786.
173. **Kowalczykowski, S. C.** 2000. Initiation of genetic recombination and recombination-dependent replication. Trends Biochem. Sci. **25:**156–165.
174. **Kravanja, M., R. Engelmann, V. Dossonnet, M. Blüggel, H. E. Meyer, R. Frank, A. Galinier, J. Deutscher, N. Schnell, and W. Hengstenberg.** 1999. The *hprK* gene of *Enterococcus faecalis* encodes a novel bifunctional enzyme: the HPr kinase/phosphatase. Mol. Microbiol. **31:**59–66.
175. **Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:**567–580.
176. **Kropinski, A. M.** 2000. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. J. Bacteriol. **182:**6066–6074.
177. **Kropinski, A. M., and M. J. Sibbald.** 1999. Transfer RNA genes and their significance to codon usage in the *Pseudomonas aeruginosa* lamboid bacteriophage D3. Can J. Microbiol. **45:**791–796.
178. **Kumar, A., R. A. Malloch, N. Fujita, D. A. Smillie, A. Ishihama, and R. S. Hayward.** 1993. The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. J. Mol. Biol. **232:**406–418.
179. **Kunisawa, T.** 1992. Synonymous codon preferences in bacteriophage T4: a distinctive use of transfer RNAs from T4 and from its host *Escherichia coli*. J. Theor. Biol. **159:**287–298.
180. **Kutter, E., K. Gachechiladze, A. Poglazov, E. Marusich, M. Shneider, P. Aronsson, A. Napuli, D. Porter, and V. Mesyanzhinov.** 1995. Evolution of T4-related phages. Virus Genes **11:**285–297.
181. **Kuzminov, A.** 1995. Collapse and repair of replication forks in *Escherichia coli*. Mol. Microbiol. **16:**373–384.
182. **Kwon, H. J., R. Tirumalai, A. Landy, and T. Ellenberger.** 1997. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. Science **276:**126–131.
183. **Lanka, E., B. Geschke, and H. Schuster.** 1978. *Escherichia coli dnaB* mutant defective in DNA initiation: isolation and properties of the dnaB protein. Proc. Natl. Acad. Sci. USA **75:**799–803.
184. **Laufer, C. S., J. B. Hays, B. E. Windle, T. S. Schaefer, E. H. Lee, S. L. Hays, and M. R. McClure.** 1989. Enhancement of *Escherichia coli* plasmid and chromosomal recombination by the Ref function of bacteriophage P1. Genetics **123:**465–476.
184a. **Lederberg, E. M.** 1950. Lysogenicity of *Escherichia coli* strain K-12. Microb. Genet. Bull. **1:**5–7.
185. **Lederberg, E. M., and J. Lederberg.** 1953. Genetic studies of lysogenicity in *E. coli.* Genetics **38:**51–64.
186. **Lehnherr, H.** Bacteriophage P1. *In* R. Calendar (ed.), The bacteriophages, in press. Oxford University Press, New York, N.Y.
187. **Lehnherr, H., J. D. Bendtsen, F. Preuss, and T. V. Ilyina.** 1999. Identification and characterization of the single-stranded DNA-binding protein of bacteriophage P1. J. Bacteriol. **181:**6463–6468.
188. **Lehnherr, H., A. Guidolin, and W. Arber.** 1991. Bacteriophage P1 gene 10 encodes a *trans*-activating factor required for late gene expression. J. Bacteriol. **173:**6438–6445.
189. **Lehnherr, H., A. Guidolin, and W. Arber.** 1992. Mutational analysis of the bacteriophage P1 late promoter sequence P*S*. J. Mol. Biol. **228:**101–107.
190. **Lehnherr, H., A.-M. Hansen, and T. V. Ilyina.** 1998. Penetration of the bacterial cell wall: a family of lytic transglycosylases in bacteriophages and conjugative plasmids. Mol. Microbiol. **30:**454–457.
191. **Lehnherr, H., C. D. Jensen, A. R. Stenholm, and A. Dueholm.** 2001. Dual regulatory control of a particle maturation function of bacteriophage P1. J. Bacteriol. **183:**4105–4109.
192. **Lehnherr, H., E. Maguin, S. Jafri, and M. Yarmolinsky.** 1993. Plasmid addiction genes of bacteriophage P1: *doc*, which causes cell death on curing of prophage, and *phd*, which prevents host death when prophage is retained. J. Mol. Biol. **233:**414–428.

193. **Lehnherr, H., M. Velleman, A. Guidolin, and W. Arber.** 1992. Bacteriophage P1 gene *10* is expressed from a promoter-operator sequence controlled by C1 and Bof proteins. J. Bacteriol. **174:**6138–6144.
194. **Lehnherr, H., and M. B. Yarmolinsky.** 1995. Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **92:**3274–3277.
195. **Lemonnier, M., G. Ziegelin, T. Reick, A. M. Gomez, R. Diaz-Orejas, and E. Lanka.** 2003. Bacteriophage P1 Ban protein is a hexameric DNA helicase that interacts with and substitutes for *Escherichia coli* DnaB. Nucleic Acids Res. **31:**3918–3928.
196. **Lennox, E. S.** 1955. Transduction of linked genetic characters of the host by bacteriophage P1. Virology **1:**190–206.
197. **Leslie, N. R., and D. J. Sherratt.** 1995. Site-specific recombination in the replication terminus region of *Escherichia coli*: functional replacement of *dif*. EMBO J. **14:**1561–1570.
198. **Lesnik, E. A., R. Sampath, H. B. Levene, T. J. Henderson, J. A. McNeil, and D. J. Ecker.** 2001. Prediction of rho-independent transcriptional terminators in *Escherichia coli*. Nucleic Acids Res. **29:**3583–3594.
199. **Levin, M. E., R. W. Hendrix, and S. R. Casjens.** 1993. A programmed translational frameshift is required for the synthesis of a bacteriophage lambda tail assembly protein. J. Mol. Biol. **234:**124–139.
200. **Lewis, L. K., G. R. Harlow, L. A. Gregg-Jolly, and D. W. Mount.** 1994. Identification of high affinity binding sites for LexA which define new DNA damage-inducible genes in *Escherichia coli*. J. Mol. Biol. **241:**507–523.
201. **Li, S., H. Pelka, and L. H. Schulman.** 1993. The anticodon and discriminator base are important for aminoacylation of *Escherichia coli* tRNA(Asn). J. Biol. Chem. **268:**18335–18339.
202. **Lin, H., V. B. Rao, and L. W. Black.** 1999. Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4. J. Mol. Biol. **289:**249–260.
203. **Łobocka, M., and M. Yarmolinsky.** 1996. P1 plasmid partition: a mutational analysis of ParB. J. Mol. Biol. **259:**366–382.
204. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25:**955–964.
205. **Lu, M. J., and U. Henning.** 1994. Superinfection exclusion by T-even-type coliphages. Trends Microbiol. **2:**137–139.
206. **Lu, M. J., Y. D. Stierhof, and U. Henning.** 1993. Location and unusual membrane topology of the immunity protein of the *Escherichia coli* phage T4. J. Virol. **67:**4905–4913.
207. **Lu, S. D., D. Lu, and M. Gottesman.** 1989. Stimulation of IS*1* excision by bacteriophage P1 *ref* function. J. Bacteriol. **171:**3427–3432.
208. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res. **26:**1107–1115.
209. **Lurz, R., A. Heisig, M. Velleman, B. Dobrinski, and H. Schuster.** 1987. The *ban* operon of bacteriophage P1. Localization of the promoter controlled by P1 repressor. J. Biol. Chem. **262:**16575–16579.
210. **Magnuson, R., H. Lehnherr, G. Mukhopadhyay, and M. B. Yarmolinsky.** 1996. Autoregulation of the plasmid addiction operon of bacteriophage P1. J. Biol. Chem. **271:**18705–18710.
211. **Magnuson, R., and M. B. Yarmolinsky.** 1998. Corepression of the P1 addiction operon by Phd and Doc. J. Bacteriol. **180:**6342–6351.
212. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
213. **Maillou, J., and B. Dreiseikelmann.** 1990. The *sim* gene of *Escherichia coli* phage P1: nucleotide sequence and purification of the processed protein. Virology **175:**500–507.
214. **Malone, T., R. M. Blumenthal, and X. Cheng.** 1995. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. J. Mol. Biol. **253:**618–632.
215. **McDonald, J. P., T. S. Peat, A. S. Levine, and R. Woodgate.** 1999. Intermolecular cleavage by UmuD-like enzymes: identification of residues required for cleavage and substrate specificity. J. Mol. Biol. **285:**2199–2209.
216. **McLean, M. J., K. H. Wolfe, and K. M. Devine.** 1998. Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. J. Mol. Evol. **47:**691–696.
217. **McLenigan, M. P., O. I. Kulaeva, D. G. Ennis, A. S. Levine, and R. Woodgate.** 1999. The bacteriophage P1 HumD protein is a functional homolog of the prokaryotic UmuD′-like proteins and facilitates SOS mutagenesis in *Escherichia coli*. J. Bacteriol. **181:**7005–7013.
218. **Mehr, I. J., C. D. Long, C. D. Serkin, and H. S. Seifert.** 2000. A homologue of the recombination-dependent growth gene, rdgC, is involved in gonococcal pilin antigenic variation. Genetics **154:**523–532.
219. **Meisel, A., T. A. Bickle, D. H. Krüger, and C. Schroeder.** 1992. Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. Nature **355:**467–469.
220. **Meisel, A., P. Mackeldanz, T. A. Bickle, D. H. Krüger, and C. Schroeder.** 1995. Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. EMBO J. **14:**2958–2966.
221. **Mesyanzhinov, V. V., B. N. Sobolev, E. I. Marusich, A. G. Prilipov, and V. P. Efimov.** 1990. A proposed structure of bacteriophage T4 gene product 22, a major prohead scaffolding core protein. J. Struct. Biol. **104:**24–31.
222. **Meyer, J., M. Stalhammar-Carlemalm, and S. Iida.** 1981. Denaturation map of bacteriophage P1 DNA. Virology **110:**167–175.
223. **Miller, E. S., J. F. Heidelberg, J. A. Eisen, W. C. Nelson, A. S. Durkin, A. Ciecko, T. V. Feldblyum, O. White, I. T. Paulsen, W. C. Nierman, J. Lee, B. Szczypinski, and C. M. Fraser.** 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J. Bacteriol. **185:**5220–5233.
224. **Miller, E. S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Rüger.** 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. **67:**86–156.
225. **Miller, J. H.** 1992. A short course in bacterial genetics. A laboratory handbook for *Escherichia coli* and related bacteria, p. 15.3–15.10. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
226. **Montag, D., and U. Henning.** 1987. An open reading frame in the *Escherichia coli* bacteriophage lambda genome encodes a protein that functions in assembly of the long tail fibers of bacteriophage T4. J. Bacteriol. **169:**5884–5886.
227. **Mosig, G.** 1994. Synthesis and maturation of T4-encoded tRNAs, p. 182–185. *In* J. D. Karam (ed.), Molecular biology of bacteriophage T4. ASM Press, Washington, D.C.
228. **Mosig, G., and D. H. Hall.** 1994. Gene expression: a paradigm of integrated circuits., p. 127–131. *In* J. D. Karam (ed.), Molecular biology of bacteriophage T4. ASM Press, Washington, D.C.
229. **Mulligan, M. E., D. K. Hawley, E. Entriken, and W. R. McClure.** 1984. *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. Nucleic Acids Res. **12:**789–800.
230. **Mural, R. J., R. H. Chesney, D. Vapnek, M. M. Kropf, and J. R. Scott.** 1979. Isolation and characterization of cloned fragments of bacteriophage P1 DNA. Virology **93:**387–397.
231. **Muramatsu, T., K. Nishikawa, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, and S. Yokoyama.** 1988. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature **336:**179–181.
232. **Muramatsu, T., S. Yokoyama, N. Horie, A. Matsuda, T. Ueda, Z. Yamaizumi, Y. Kuchino, S. Nishimura, and T. Miyazawa.** 1988. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. J. Biol. Chem. **263:**9261–9267.
233. **Murphy, L. D., J. L. Rosner, S. B. Zimmerman, and D. Esposito.** 1999. Identification of two new proteins in spermidine nucleoids isolated from *Escherichia coli*. J. Bacteriol. **181:**3842–3844.
234. **Nainen, O.** 1975. Ph.D. thesis. University of Georgia, Athens.
235. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
236. **Nureki, O., T. Niimi, T. Muramatsu, H. Kanno, T. Kohno, C. Florentz, R. Giege, and S. Yokoyama.** 1994. Molecular recognition of the identity-determinant set of isoleucine transfer RNA from *Escherichia coli*. J. Mol. Biol. **236:**710–724.
237. **Ogawa, T.** 1975. Analysis of *dnaB* function of *Escherichia coli* K12 and the *dnaB*-like function of P1 prophage. J. Mol. Biol. **94:**327–340.
238. **Osborne, F. A., S. R. Stovall, and B. R. Baumstark.** 1989. The *c*1 genes of P1 and P7. Nucleic Acids Res. **17:**7671–7680.
239. **Pal, S. K., and D. K. Chattoraj.** 1988. P1 plasmid replication: initiator sequestration is inadequate to explain control by initiator-binding sites. J. Bacteriol. **170:**3554–3560.
240. **Pallanck, L., M. Pak, and H. Schulman.** 1995. tRNA discrimination in aminoacylation, p. 371–394. *In* D. Soll and U. I. RajBhandary (ed.), tRNA: structure, biosynthesis, and function. ASM Press, Washington, D.C.
241. **Park, K., and D. K. Chattoraj.** 2001. DnaA boxes in the P1 plasmid origin: the effect of their position on the directionality of replication and plasmid copy number. J. Mol. Biol. **310:**69–81.
242. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848–852.
243. **Patel, P. H., M. Suzuki, E. Adman, A. Shinkai, and L. A. Loeb.** 2001. Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. J. Mol. Biol. **308:**823–837.
244. **Paul, S., and D. Summers.** 2004. ArgR and PepA, accessory proteins for XerCD-mediated resolution of ColE1 dimers, are also required for stable maintenance of the P1 prophage. Plasmid **52:**63–68.
245. **Pedulla, M. L., M. E. Ford, J. M. Houtz, T. Karthikeyan, C. Wadsworth, J. A. Lewis, D. Jacobs-Sera, J. Falbo, J. Gross, N. R. Pannunzio, W. Brucker, V. Kumar, J. Kandasamy, L. Keenan, S. Bardarov, J. Kriakov, J. G. Lawrence, W. R. Jacobs, Jr., R. W. Hendrix, and G. F. Hatfull.** 2003. Origins of highly mosaic mycobacteriophage genomes. Cell **113:**171–182.

246. **Pierce, J. C., and N. L. Sternberg.** 1992. Using bacteriophage P1 system to clone high molecular weight genomic DNA. Methods Enzymol. **216:**549–574.
247. **Plunkett, G., III, D. J. Rose, T. J. Durfee, and F. R. Blattner.** 1999. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J. Bacteriol. **181:**1767–1778.
248. **Prentki, P., M. Chandler, and L. Caro.** 1977. Replication of prophage P1 during the cell cycle of *Escherichia coli*. Mol. Gen. Genet. **152:**71–76.
249. **Radnedge, L., M. A. Davis, and S. J. Austin.** 1996. P1 and P7 plasmid partition: ParB protein bound to its partition site makes a separate discriminator contact with the DNA that determines species specificity. EMBO J. **15:**1155–1162.
250. **Ramos, A., and G. Varani.** 1997. Structure of the acceptor stem of *Escherichia coli* tRNA Ala: role of the G3.U70 base pair in synthetase recognition. Nucleic Acids Res. **25:**2083–2090.
251. **Ravin, N. V., J. Rech, and D. Lane.** 2003. Mapping of functional domains in F plasmid partition proteins reveals a bipartite SopB-recognition domain in SopA. J. Mol. Biol. **329:**875–889.
252. **Razza, J. B., C. A. Watkins, and J. R. Scott.** 1980. Phage P1 temperature-sensitive mutants with defects in the lytic pathway. Virology **105:**52–59.
253. **Redaschi, N., and T. A. Bickle.** 1996. Posttranscriptional regulation of *Eco*P1I and *Eco*P15I restriction activity. J. Mol. Biol. **257:**790–803.
254. **Reeve, J. N., E. Lanka, and H. Schuster.** 1980. Synthesis of P1 ban protein in minicells infected by P1 mutants. Mol. Gen. Genet. **177:**193–197.
255. **Reisenauer, A., L. S. Kahng, S. McCollum, and L. Shapiro.** 1999. Bacterial DNA methylation: a cell cycle regulator? J. Bacteriol. **181:**5135–5139.
256. **Riedel, H. D., J. Heinrich, A. Heisig, T. Choli, and H. Schuster.** 1993. The antirepressor of phage P1. Isolation and interaction with the C1 repressor of P1 and P7. FEBS Lett. **334:**165–169.
257. **Riedel, H. D., J. Heinrich, and H. Schuster.** 1993. Cloning, expression, and characterization of the *icd* gene in the *immI* operon of bacteriophage P1. J. Bacteriol. **175:**2833–2838.
258. **Rodionov, O., M. Łobocka, and M. Yarmolinsky.** 1999. Silencing of genes flanking the P1 plasmid centromere. Science **283:**546–549.
259. **Rodionov, O., and M. Yarmolinsky.** 2004. Plasmid partitioning and the spreading of P1 partition protein ParB. Mol. Microbiol. **52:**1215–1223.
260. **Rodriguez-Lemoine, V.** 1992. PacB, a plasmid-encoded property which confers to *E. coli* K12 resistance to channel-forming colicins. Zentralbl. Bakteriol. **276:**374–379.
261. **Rosner, J. L.** 1972. Formation, induction and curing of bacteriophage Pl lysogens. Virology **48:**679–689.
262. **Rosner, J. L.** 1973. Modification-deficient mutants of bacteriophage P1. I. Restriction by P1 cryptic lysogens. Virology **52:**213–222.
263. **Ryder, L., G. J. Sharples, and R. G. Lloyd.** 1996. Recombination-dependent growth in exonuclease-depleted *recBC sbcBC* strains of *Escherichia coli* K-12. Genetics **143:**1101–1114.
264. **Salzberg, S. L., A. L. Delcher, S. Kasif, and O. White.** 1998. Microbial gene identification using interpolated Markov models. Nucleic Acids Res. **26:**544–548.
265. **Salzberg, S. L., A. J. Salzberg, A. R. Kerlavage, and J. F. Tomb.** 1998. Skewed oligomers and origins of replication. Gene **217:**57–67.
266. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning. a laboratory manual. Cold Spring Harber Laboratory Press, Cold Spring Harbor, N.Y.
267. **Sandmeier, H.** 1994. Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibres. Mol. Microbiol. **12:**343–350.
268. **Sandmeier, H., S. Iida, and W. Arber.** 1992. DNA inversion regions Min of plasmid pl5B and Cin of bacteriophage Pl: evolution of bacteriophage tail fiber genes. J. Bacteriol. **174:**3936–3944.
269. **Sandmeier, H., S. Iida, P. Hübner, R. Hiestand-Nauer, and W. Arber.** 1991. Gene organization in the multiple DNA inversion region *min* of plasmid p15B of *E.coli* 15T-: assemblage of a variable gene. Nucleic Acids Res. **19:**5831–5838.
270. **Sandmeier, H., S. Iida, J. Meyer, R. Hiestand-Nauer, and W. Arber.** 1990. Site-specific DNA recombination system Min of plasmid p15B: a cluster of overlapping invertible DNA segments. Proc. Natl. Acad. Sci. USA **87:**1109–1113.
271. **Sauer, B.** 1987. Functional expression of the *cre-lox* site-specific recombination system in the yeast *Saccharomyces cerevisiae*. Mol. Cell. Biol. **7:**2087–2096.
272. **Schaefer, T., and J. Hays.** 1990. The *bof* gene of bacteriophage P1: DNA sequence and evidence for roles in regulation of phage *c*1 and *ref* genes. J. Bacteriol. **172:**3269–3277.
273. **Schaefer, T. S., and J. B. Hays.** 1991. Bacteriophage Pl Bof protein is an indirect positive effector of transcription of the Pl *bac-1 ban* gene in some circumstances and a direct negative effector in other circumstances. J. Bacteriol. **173:**6469–6474.
274. **Scherberg, N. H., and S. B. Weiss.** 1972. T4 transfer RNAs: codon recognition and translational properties. Proc. Natl. Acad. Sci. USA **69:**1114–1118.
275. **Schmidt, C.** 1991. Ph.D thesis. University of Basel, Basel, Switzerland.
276. **Schmidt, C., H. Lehnherr, A. Guidolin, and W. Arber.** 1994. GenBank accession no. M95666.
277. **Schmidt, C., M. Velleman, and W. Arber.** 1996. Three functions of bacteriophage P1 involved in cell lysis. J. Bacteriol. **178:**1099–1104.
278. **Schuler, G. D., S. F. Altschul, and D. J. Lipman.** 1991. A workbench for multiple alignment construction and analysis. Proteins **9:**180–190.
279. **Scott, J. R.** 1970. Clear plaque mutants of phage Pl. Virology **41:**66–71.
280. **Scott, J. R., M. Kropf, and L. Mendelson.** 1977. Clear plaque mutants of phage P7. Virology **76:**39–46.
281. **Scott, J. R., and M. M. Kropf.** 1977. Location of new clear plaque genes on the P1 map. Virology **82:**362–368.
282. **Segev, N., A. Laub, and G. Cohen.** 1980. A circular form of bacteriophage P1 DNA made in lytically infected cells of *Escherichia coli.* Characterization and kinetics of formation. Virology **101:**261–271.
283. **Senger, B., S. Auxilien, U. Englisch, F. Cramer, and F. Fasiolo.** 1997. The modified wobble base inosine in yeast tRNA-Ile is a positive determinant for aminoacylation by isoleucyl-tRNA synthetase. Biohemistry **36:**8269–8275.
284. **Sharpe, M. E., and J. Errington.** 1996. The *Bacillus subtilis soj-spo0J* locus is required for a centromere-like function involved in prespore chromosome partitioning. Mol. Microbiol. **21:**501–509.
285. **Sharrocks, A. D., and D. P. Hornby.** 1991. Transcriptional analysis of the restriction and modification genes of bacteriophage P1. Mol. Microbiol. **5:**685–694.
286. **Skorupski, K., J. C. Pierce, B. Sauer, and N. Sternberg.** 1992. Bacteriophage Pl genes involved in the recognition and cleavage of the phage packaging site (*pac*). J. Mol. Biol. **223:**977–989.
287. **Skorupski, K., B. Sauer, and N. Sternberg.** 1994. Faithful cleavage of the P1 packaging site (*pac*) requires two phage proteins, PacA and PacB, and two *Escherichia coli* proteins, IHF and HU. J. Mol. Biol. **243:**268–282.
288. **Skorupski, K., N. Sternberg, and B. Sauer.** 1994. Purification and DNA-binding activity of the PacA subunit of the bacteriophage P1 pacase enzyme. J. Mol. Biol. **243:**258–267.
289. **Slater, S. C., M. R. Lifsics, M. O'Donnell, and R. Maurer.** 1994. *holE*, the gene coding for the theta subunit of DNA polymerase III of *Escherichia coli*: characterization of a *holE* mutant and comparison with a *dnaQ* (epsilon-subunit) mutant. J. Bacteriol. **176:**815–821.
290. **Sommer, S., J. Knezevic, A. Bailone, and R. Devoret.** 1993. Induction of only one SOS operon, umuDC, is required for SOS mutagenesis in *Escherichia coli*. Mol. Gen. Genet. **239:**137–144.
291. **Sourice, S., V. Biaudet, M. El Karoui, S. D. Ehrlich, and A. Gruss.** 1998. Identification of the Chi site of *Haemophilus influenzae* as several sequences related to the *Escherichia coli* Chi site. Mol. Microbiol. **27:**1021–1029.
292. **Steiner, W., G. Liu, W. D. Donachie, and P. Kuempel.** 1999. The cytoplasmic domain of FtsK protein is required for resolution of chromosome dimers. Mol. Microbiol. **31:**579–583.
293. **Sternberg, N.** 1990. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc. Natl. Acad. Sci. USA **87:**103–107.
294. **Sternberg, N.** 1979. A characterization of bacteriophage P1 DNA fragments cloned in a lambda vector. Virology **96:**129–142.
295. **Sternberg, N., and G. Cohen.** 1989. Genetic analysis of the lytic replicon of bacteriophage Pl. II. Organization of replicon elements. J. Mol. Biol. **207:**111–133.
296. **Sternberg, N., and J. Coulby.** 1990. Cleavage of the bacteriophage P1 packaging site (*pac*) is regulated by adenine methylation. Proc. Natl. Acad. Sci. USA **87:**8070–8074.
297. **Sternberg, N., and J. Coulby.** 1987. Recognition and cleavage of the bacteriophage Pl packaging site (*pac*) I. Differential processing of the cleaved ends *in vivo*. J. Mol. Biol. **194:**453–468.
298. **Sternberg, N., and J. Coulby.** 1987. Recognition and cleavage of the bacteriophage Pl packaging site (*pac*) II. Functional limits of *pac* and location of *pac* cleavage termini. J. Mol. Biol. **194:**469–480.
299. **Sternberg, N., D. Hamilton, S. Austin, M. Yarmolinsky, and R. Hoess.** 1981. Site-specific recombination and its role in the life cycle of bacteriophage P1. Cold Spring Harbor Symp. Quant. Biol. **45:**297–309.
300. **Sternberg, N., D. Hamilton, and R. Hoess.** 1981. Bacteriophage P1 site-specific recombination. II. Recombination between *loxP* and the bacterial chromosome. J. Mol. Biol. **150:**487–507.
301. **Sternberg, N., and R. Hoess.** 1983. The molecular genetics of bacteriophage P1. Annu. Rev. Genet. **17:**123–154.
302. **Sternberg, N., B. Sauer, R. Hoess, and K. Abremski.** 1986. Bacteriophage Pl *cre* structural gene and its regulatory region. Evidence for multiple promoters and for regulation by DNA methylation. J. Mol. Biol. **187:**197–212.
303. **Sternberg, N. L., and R. Maurer.** 1991. Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella typhimurium*. Methods Enzymol. **204:**18–43.
304. **Streiff, M. B., S. Iida, and T. A. Bickle.** 1987. Expression and proteolytic processing of the *darA* anti-restriction gene product of bacteriophage P1. Virology **157:**167–171.
305. **Studwell-Vaughan, P. S., and M. O'Donnell.** 1993. DNA polymerase III

accessory proteins. V. Theta encoded by *holE*. J. Biol. Chem. **268:**11785–11791.
306. **Surtees, J. A., and B. E. Funnell.** 2003. Plasmid and chromosome traffic control: how ParA and ParB drive partition. Curr. Top. Dev. Biol. **56:** 145–180.
307. **Susskind, M. M., A. Wright, and D. Botstein.** 1971. Superinfection exclusion by P22 prophage in lysogens of *Salmonella typhimurium.* II. Genetic evidence for two exclusion systems. Virology **45:**638–652.
308. **Tamura, K., H. Asahara, H. Himeno, T. Hasegawa, and M. Shimizu.** 1991. Identity elements of *Escherichia coli* tRNA(Ala). J. Mol. Recognit. **4:**129–132.
309. **Tang, M., X. Shen, E. G. Frank, M. O'Donnell, R. Woodgate, and M. F. Goodman.** 1999. UmuD′(2)C is an error-prone DNA polymerase, Escherichia coli pol V. Proc. Natl. Acad. Sci. USA **96:**8919–8924.
310. **Touati-Schwartz, D.** 1979. A *dnaB* analog *ban*, specified by bacteriophage P1: genetic and physiological evidence for functional analogy and interactions between the two products. Mol. Gen. Genet. **174:**173–188.
311. **Touati-Schwartz, D.** 1979. A new pleiotropic bacteriophage P1 mutation, bof, affecting *c*1 repression activity, the expression of plasmid incompatibility and the expression of certain constitutive prophage genes. Mol. Gen. Genet. **174:**189–202.
312. **Touati-Schwartz, D.** 1978. Two replication functions in phage P1: *ban*, an analog of *dnaB*, and *bof*, involved in the control of replication., p. 683–691. *In* I. Molineaux, and M. Kohiyama (ed.), DNA synthesis: present and future, vol. 17. Plenum Press, New York, N.Y.
313. **Tuggle, C. K., and J. A. Fuchs.** 1986. Regulation of the synthesis of ribonucleotide reductase in *Escherichia coli.*, p. 227–232. *In* C.-I. Brandon, A. Holmgren, J.-O. Hoag, H. Jornvall, and B.-M. Sjoberg (ed.), Thioredoxin and glutaredoxin systems: structure and function. Raven Press, New York, N.Y.
314. **Tuma, R., and G. J. Thomas, Jr.** 1997. Mechanisms of virus assembly probed by Raman spectroscopy: the icosahedral bacteriophage P22. Biophys. Chem. **68:**17–31.
315. **Velleman, M.** 1989. PhD thesis. Freie Universitat Berlin, Berlin, Germany.
316. **Velleman, M., B. Dreisekelmann, and H. Schuster.** 1987. Multiple repressor binding sites in the genome of bacteriophage Pl. Proc. Natl. Acad. Sci. USA **84:**5570–5574.
317. **Velleman, M., T. Heinzel, and H. Schuster.** 1992. The Bof protein of bacteriophage Pl exerts its modulating function by formation of a ternary complex with operator DNA and Cl repressor. J. Biol. Chem. **267:**12174–12181.
318. **Velleman, M., M. Heirich, A. Günther, and H. Schuster.** 1990. A bacteriophage P1-encoded modulator protein affects the P1 *c*1 repression system. J. Mol. Biol. **265:**18511–18517.
319. **Velleman, M., and S. Parbus.** 1992. Purification of the C1 repressor of bacteriophage P1 by fast protein liquid chromatography. J. Chromatogr. **625:**41–46.
320. **Voegtli, W. C., D. J. White, N. J. Reiter, F. Rusnak, and A. C. Rosenzweig.** 2000. Structure of the bacteriophage lambda Ser/Thr protein phosphatase with sulfate ion bound in two coordination modes. Biochemistry **39:**15365–15374.
321. **Walker, D. H., Jr., and J. T. Walker.** 1976. Genetic studies of coliphage P1. III. Extended genetic map. J. Virol. **20:**177–187.
322. **Walker, D. H. J., and T. F. Anderson.** 1970. Morphological variants of coliphage P1. J. Virol. **5:**765–782.
323. **Walker, J. T., and D. H. J. Walker.** 1983. Coliphage P1 morphogenesis: analysis of mutants by electron microscopy. J. Bacteriol. **45:**1118–1139.
324. **Walker, J. T., and D. H. J. Walker.** 1980. Mutations in coliphage P1 affecting host cell lysis. J. Virol. **35:**519–530.
325. **Walker, J. T., and D. H. J. Walker.** 1981. Structural proteins of coliphage P1. Prog. Clin. Biol. Res. **64:**69–77.
326. **Wandersman, C., and M. Yarmolinsky.** 1977. Bipartite control of immunity conferred by the related heteroimmune plasmid prophages, P1 and P7 (formerly phi Amp). Virology **77:**386–400.
327. **Wang, I. N., D. L. Smith, and R. Young.** 2000. Holins: the protein clocks of bacteriophage infections. Annu. Rev. Microbiol. **54:**799–825.
328. **Wang, L. K., and S. Shuman.** 2001. Domain structure and mutational analysis of T4 polynucleotide kinase. J. Biol. Chem. **276:**26868–26874.
329. **Watts, N. R., and D. H. Coombs.** 1990. Structure of the bacteriophage T4 baseplate as determined by chemical cross-linking. J. Virol. **64:**143–154.
330. **Weisberg, R. A., and M. E. Gottesman.** 1999. Processive antitermination. J. Bacteriol. **181:**359–367.
331. **Whelan, K. F., E. Colleran, and D. E. Taylor.** 1995. Phage inhibition, colicin resistance, and tellurite resistance are encoded by a single cluster of genes on the IncHI2 plasmid R478. J. Bacteriol. **177:**5016–5027.
332. **Whelan, K. F., R. K. Sherburne, and D. E. Taylor.** 1997. Characterization of a region of the IncHI2 plasmid R478 which protects *Escherichia coli* from toxic effects specified by components of the tellurite, phage, and colicin resistance cluster. J. Bacteriol. **179:**63–71.
333. **Wickner, S., J. Hoskins, D. Chattoraj, and K. McKenney.** 1990. Deletion analysis of the mini-P1 plasmid origin of replication and the role of *Escherichia coli* DnaA protein. J. Biol. Chem. **265:**11622–11627.
334. **Wickner, S., J. Hoskins, and K. McKenney.** 1991. Monomerization of RepA dimers by heat shock proteins activates binding to DNA replication origin. Proc Natl Acad Sci USA **88:**7903–7907.
335. **Williams, D. R., D. P. Macartney, and C. M. Thomas.** 1998. The partitioning activity of the RK2 central control region requires only *incC*, *korB* and KorB-binding site O(B)3 but other KorB-binding sites form destabilizing complexes in the absence of O(B)3. Microbiology **144:**3369–3378.
336. **Williams, M. D., J. A. Fuchs, and M. C. Flickinger.** 1991. Null mutation in the stringent starvation protein of *Escherichia coli* disrupts lytic development of bacteriophage P1. Gene **109:**21–30.
337. **Windle, B. E., and J. B. Hays.** 1986. A phage P1 function that stimulates homologous recombination of the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA **83:**3885–3889.
338. **Windle, B. E., C. S. Laufer, and J. B. Hays.** 1988. Sequence and deletion analysis of the recombination enhancement gene (*ref*) of bacteriophage P1: evidence for promoter-operator and attenuator-antiterminator control. J. Bacteriol. **170:**4881–4889.
339. **Wood, L. F., N. Y. Tszine, and G. E. Christie.** 1997. Activation of P2 late transcription by P2 Ogr protein requires a discrete contact site on the C terminus of the alpha subunit of *Escherichia coli* RNA polymerase. J. Mol. Biol. **274:**1–7.
340. **Xu, G. L., W. Kapfer, J. Walter, and T. A. Trautner.** 1992. *Bsu*BI-an isospecific restriction and modification system of *Pst*I: characterization of the *Bsu*BI genes and enzymes. Nucleic Acids Res. **20:**6517–6523.
341. **Xu, M., D. K. Struck, J. Deaton, I. N. Wang, and R. Young.** 2004. A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. Proc. Natl. Acad. Sci. USA **101:**6415–6420.
342. **Xue, Q., and J. B. Egan.** 1995. DNA sequence of tail fiber genes of coliphage 186 and evidence for a common ancestor shared by dsDNA phage fiber genes. Virology **212:**128–133.
343. **Yamaichi, Y., and H. Niki.** 2000. Active segregation by the *Bacillus subtilis* partitioning system in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **97:**14656–14661.
343a. **Yarmolinsky, M. B.** 2004. Bacteriophage P1 in retrospect and in prospect. J. Bacteriol. **186:**7032–7068.
344. **Yarmolinsky, M.** 1990. Bacteriophage P1, p. 1.52–1.58. *In* S. J. O'Brien (ed.), Genetic maps. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
345. **Yarmolinsky, M.** 1977. Genetic and physical structure of bacteriophage P1 DNA, p. 721–732. *In* I. A. Bukhari, J. A. Shapiro, and S. L. Adhya (ed.), DNA insertion elements, plasmids, and episomes. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
346. **Yarmolinsky, M., and M. £obocka.** 1993. Bacteriophage Pl, p. 42–54. *In* S. J. O'Brien (ed.), Locus maps of complex genomes., vol. 3. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
347. **Yarmolinsky, M., and N. Sternberg.** 1988. Bacteriophage Pl, p. 291–438. *In* R. Calendar (ed.), The Bacteriophages, vol. 1. Plenum Press, New York, N.Y.
348. **Yarmolinsky, M. B.** 1995. Programmed cell death in bacterial populations. Science **267:**836–837.
349. **Yarmolinsky, M. B., E. B. Hansen, S. Jafri, and D. K. Chattoraj.** 1989. Participation of the lytic replicon in bacteriophage P1 plasmid maintenance. J. Bacteriol. **171:**4785–4791.
350. **Yoo, S. J., J. H. Seol, M. S. Kang, D. B. Ha, and C. H. Chung.** 1994. *clpX* encoding an alternative ATP-binding subunit of protease Ti (Clp) can be expressed independently from *clpP* in *Escherichia coli*. Biochem. Biophys. Res. Commun. **203:**798–804.
351. **Young, R.** 1992. Bacteriophage lysis: mechanism and regulation. Microbiol. Rev. **56:**430–481.
352. **Zabrovitz, S., N. Segev, and G. Cohen.** 1977. Growth of bacteriophage P1 in recombination-deficient hosts of *Escherichia coli*. Virology **80:**233–248.
353. **Zdanovsky, A. G., and M. V. Zdanovskaia.** 2000. Simple and efficient method for heterologous expression of clostridial proteins. Appl. Environ. Microbiol. **66:**3166–3173.
354. **Zhang, Z., A. A. Schaffer, W. Miller, T. L. Madden, D. J. Lipman, E. V. Koonin, and S. F. Altschul.** 1998. Protein sequence similarity searches using patterns as seeds. Nucleic Acids Res. **26:**3986–3990.