JOURNAL OF BACTERIOLOGY, June 1992, p. 4086–4093
0021-9193/92/124086-08$02.00/0
Copyright © 1992, American Society for Microbiology
Vol. 174, No. 12

Attachment Sites for Bacteriophage P2 on the *Escherichia coli* Chromosome: DNA Sequences, Localization on the Physical Map, and Detection of a P2-Like Remnant in *E. coli* K-12 Derivatives†

VIRGINIA BARREIRO AND ELISABETH HAGGÅRD-LJUNGQUIST*

Department of Microbial Genetics, Karolinska Institutet, Box 60400, S-10401 Stockholm, Sweden

Received 11 November 1991/Accepted 5 April 1992

Integration of bacteriophage P2 into the *Escherichia coli* genome involves recombination between two attachment sites, *attP* and *attB*, one on the phage and one on the host genome, respectively. At least 10 different *attB* sites have been identified over the years. In *E. coli* C, one site, called *locI*, is preferred, being occupied before any of the others. In *E. coli* K-12, no such preference is seen (reviewed in L. E. Bertani and E. W. Six, p. 73–143, *in* R. Calendar, ed., *The Bacteriophages*, vol. 2, 1988). The DNA sequence of *locI* has been determined, and it shows a core sequence of 27 nucleotides identical to *attP* (A. Yu, L. E. Bertani, and E. Haggård-Ljungquist, Gene 80:1–12, 1989). By inverse polymerase chain reactions, the prophage-host junctions of DNA extracted from P2 lysogenic strains have been amplified, cloned, and sequenced. By combining the *attL* and *attR* sequences, the *attB* sequences of locations II, III, and H have been deduced. The core sequence of location II had 20 matches to the 27-nucleotide core sequence of *attP*; the sequences of locations III and H had 17 matches. Thus, the P2 integrase accepts at least up to 37% mismatches within the core sequence. The *E. coli* K-12 strains examined all contain a 639-nucleotide-long cryptic remnant of P2 at a site with a sequence similar to that of *locI* but that may have a different map position. The P2 remnant consists of the C-terminal part of gene *D*, all of gene *ogr*, and *attR*. Locations II, III, and H have been located on Kohara's physical map to positions 3670, 1570 to 1575, and 2085, respectively.

P2 is a temperate phage that forms stable lysogens in several enterobacteria, including *Escherichia coli* C and K-12 (for a review, see reference 10). In the lysogenic stage, P2 has always been found as an integrated prophage. The integration occurs according to Campell's model by site-specific recombination between a bacterial attachment site, *attB*, and the attachment site of P2, *attP*, giving rise to two phage-host junctions, *attL* and *attR*. At least 10 different *attB* sites have been defined (Table 1). In *E. coli* C, one site, *locI*, is preferred, being occupied before any of the others (6). However, in *E. coli* K-12, no such preference has been noted, and locations H and II are occupied with about equal probability (19). The DNA sequence of *locI* has been determined. It consists of 27 nucleotides (nt) identical to the so-called core sequence of P2 *attP* (40). If *locI* is replaced with the homologous chromosomal segment of K-12 by transduction, the resulting strain, although mostly C, becomes like K-12 in respect to site preference; i.e., the phage will attach to one of several sites with roughly equal probabilities (35). There is also evidence for genetic changes in the phage which modify the site preference (33). Such variant phages, called *saf*, are produced from a prophage established in location II. DNA sequence analysis has shown that one *saf* mutant has a base replacement within the *attP* core sequence (40). In order to clarify the site preference pattern and sequence requirements for the site-specific recombination, we have determined the core sequences of alternative *attB* sites.

MATERIALS AND METHODS

Biological materials. For bacteria and plasmids used in this work, see Table 2.

Biochemical and recombinant DNA procedures. Generally, standard techniques according to reference 23 were used. The enzymes were obtained from Promega unless otherwise stated, and they were used as recommended by the manufacturer. The oligonucleotides were obtained from Scandinavian Gene Synthesis AB (Köping, Sweden).

(i) Extraction of bacterial DNA. Bacterial DNA was extracted as previously described (22).

TABLE 1. Known P2 attachment sites in *E. coli*[a]

Attachment site(s)	Map location
I	Between histidine operon and *metG* at about 48/100[b]; found only in *E. coli* C
II	Between *metE* and *rha* at about 88/100; found in *E. coli* C and K-12
H	Between *shiA* and *his* at about 44/100; found in *E. coli* K-12
III	Between *trp* and TerC at about 32/100; found in *E. coli* C and possibly K-12
IV	Weakly linked to *trp* and *metE* at about 85/100
V–IX	Not precisely located
E	Not precisely located

[a] The data are taken from references 9, 18, and 39.
[b] One-hundred-minute map of *E. coli* C (39) or *E. coli* K-12 (4).

* Corresponding author.

† This work is dedicated to our friend and colleague Berit Lundqvist, who recently died of cancer at the age of 45. She followed this work with great interest, and we deeply feel her loss.

TABLE 2. Bacterial strains and plasmids used

Strain or plasmid	Pertinent features	Origin or reference
Bacteria		
Derivatives of *E. coli* C		
C-1a	F^-, prototrophic	31
C-77	F^+ (P2)III	33
C-1094	F^- T1 *trp-1* (P2)II	11
C-1215	F^- *rha-1 metE4 his-4* (P2 *l c ts*4)II	11
Derivatives of *E. coli* K-12		
C-600r^-m^+	F^- *thi-1 thr-1 leuB6 lacY1 tonA21 supE44*	2, 29
K-207	F^- *his shiA aroD pro arg xyl* (P2 *l c ts*37)H	11
LG102	HfrH *thi* (P2)II	11
Plasmid pUC18	General cloning and sequencing vector	26

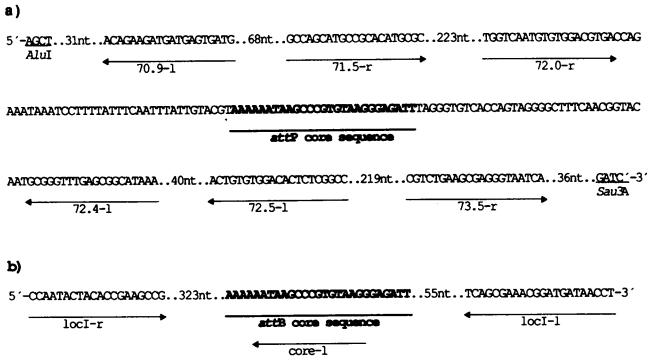

FIG. 2. DNA sequence of the P2 *attP* region (note that this represents the phage structure in which the *ogr* gene is located to the left of *attP* and the *int* gene is located to the right of *attP*) (a) and the *E. coli* C *attB* region of *locI* (b). The locations and directions of primers and pertinent restriction sites are indicated. The 27-nt-long core sequences are indicated by bold letters.

(ii) Enzymatic amplification of phage-host junctions by inverse PCR. The inverse polymerase chain reaction (PCR) was performed essentially as described in reference 32. DNA isolated from P2 lysogenic strains was cleaved with *Sau*3AI (for amplification of *attL*) or *Alu*I (for *attR*), and the fragments were ligated at a concentration of about 10 ng/μl to circularize them (Fig. 1). About 1 ng of ligated DNA was then amplified by using the GeneAmp kit from Perkin-Elmer. The reaction mixture, including 20 pmol of the pertinent primers [i.e., 72.5-l and 73.5-r for *attL* and 70.9-l and 71.5-r for *attR* (Fig. 2a)], was subjected to 30 repeated cycles of 30 s at 94°C, 30 s at 50°C, and 2 min at 70°C in a Perkin-Elmer Cetus Thermocycler (model PCR1000).

(iii) Enzymatic amplification by PCR of the cryptic P2 integrated into *locI* of *E. coli* K-12 (*locI*-K). Bacteria were grown, and the DNA was extracted as described before (22). About 25 μg of bacterial DNA was amplified by using 20 pmol of the primers locI-l and locI-r (Fig. 2b) as described above.

(iv) Enzymatic amplification by PCR of the *attB* region of *locII*, *locIII*, and *locH*. Strain C-600 r^-m^+ was grown, and the DNA was extracted as described before (22). About 25 μg of bacterial DNA was amplified by using 20 pmol of the respective l and r primers of *locII*, *locIII*, and *locH* (Fig. 3), as described above.

(v) Cloning of PCR-amplified DNA. Unless otherwise stated, the PCR reactions gave only one detectable DNA

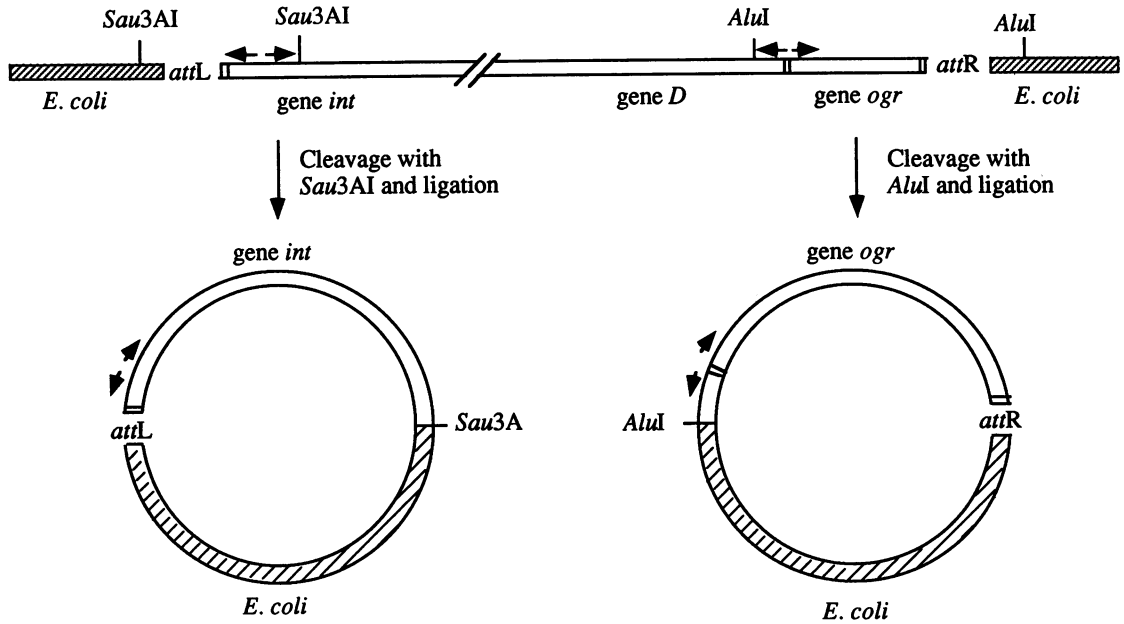

FIG. 1. Amplification of phage-host junctions by inverse PCR. The top line shows the ends of the prophage map and the junctions to the host chromosome. Note that the map is not drawn in scale. Only the restriction sites used for generation of the circles containing the *attL* and *attR* fragments are indicated. P2 DNA is symbolized by open boxes, and host DNA is symbolized by hatched boxes. The locations and orientations of the primers are indicated by short arrows. For a complete P2 map, see reference 16.

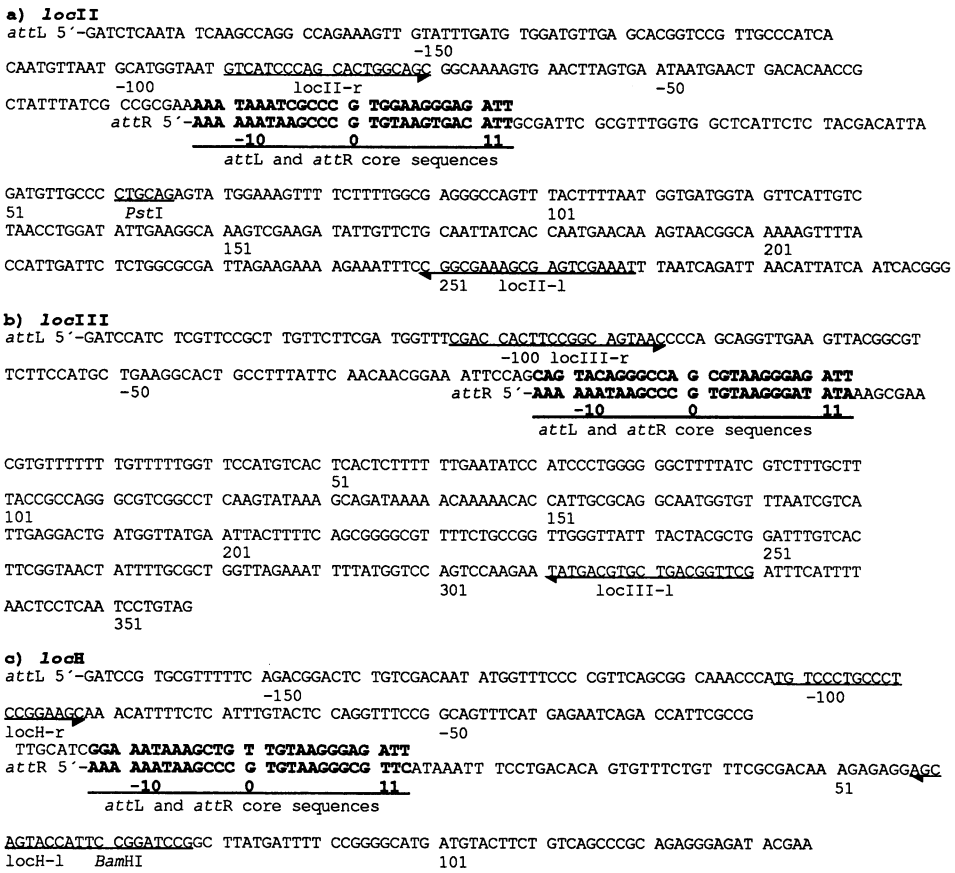

FIG. 3. DNA sequences of *attL* and *attR* sites and flanking *E. coli* sequences from locations II (a), III (b), and H (c). The core sequences are underlined and indicated by bold letters. The central nucleotide of the core sequence has been numbered 0, and the numbers below the sequences indicate distances in nucleotides from the centers of the core sequences. The primers used for PCR reactions are indicated by the arrows below the sequences.

fragment when products were analyzed by agarose gel electrophoresis. After phenol extraction and ethanol precipitation, the 5′ ends of the amplified DNA were phosphorylated with T4 polynucleotide kinase. After heat inactivation of the kinase and ethanol precipitation, the DNA was ligated with plasmid pUC18 (26), which had been cleaved with *Sma*I and treated with alkaline phosphatase. The ligated DNA was transformed into strain C-1a, and ampicillin-resistant colonies were isolated. To isolate clones containing P2 DNA, the colonies were hybridized against ^{32}P-labeled probes which differed from those used in the PCR reaction, i.e., 72.4-l for *attL*, 72.0-r for *attR* and core-l, and 70.9-l and 72.0-r for the isolation of the cryptic P2 in *locI*-K (Fig. 2). Several clones from each PCR reaction were kept for the sequence analysis.

(vi) DNA sequence analysis. The DNA sequence was determined by using the dideoxynucleotide method (30) on the fragments cloned in pUC18 with [α-^{35}S]dATP (Amersham), T7 Sequenase, the forward or reverse primer of M13, or P2 primer 72.0-r or 72.4-l. After dideoxy sequencing, the products were separated on urea-acrylamide sequencing gels. The gels were subsequently dried and autoradiographed.

(vii) Localization of attachment sites on the physical map of *E. coli*. Specialized transducing λ phages carrying identified, overlapping *E. coli* chromosomal fragments have been obtained from Y. Kohara (20). Pertinent clones were first tested for the presence of sequences from the respective location by PCR, using primers located to the right and left of the respective core region (Fig. 3). The conditions were as described above except that only 20 repeated cycles were used. The λ clones showing a positive reaction with PCR and the clones on either side on the physical map were grown up, the phage particles were isolated with Lambda Sorb (Promega), and the phage DNA was extracted with phenol. To verify that the positive clones contained sequences from the appropriate site, they were analyzed by PCR as described above and by dot blot hybridization using ^{32}P-labeled probes from the appropriate location. To screen the whole collection of transducing λ phages for the presence of *locI*, The *Escherichia coli* Gene Mapping Membrane was obtained from Kakara Shuzo Co., Ltd., and the membrane was hybridized to the ^{32}P-labeled probes as recommended by the manufacturer.

Nucleotide sequence accession numbers. The EMBL and GenBank accession numbers for the sequences in this paper are Z11491 (*attB* of *locH*), Z11492 (*attB* of *locII*), and Z11493 (*attB* of *locIII*).

RESULTS

Isolation and determination of host DNA sequences flanking *attL* in lysogens having the P2 prophage integrated in location II, III, or H. The amplification of DNA flanking a region of known sequence is possible by inverse PCR (27, 32, 37). The

attP wt	TTTTATTTCAATTTATTGTACGT	**AAAAAATAAGCCCGTGTAAGGGAGATT** TAGGGTGTCACCAGTAGGGGCTTTCAACGG
attB, *locI*	accagagatAgggctTatgcata	**AAAAAATAAGCCCGTGTAAGGGAGATT** acacagGctAaggaggtGGttccTggtaca
attB, *locII*	caaccgcTattTaTcgccgcgaa	**AAAtAAatcGCCCGTGgAAGtGAcATT** gcGatTcgCgttt ggtGGctcaTtctCtac
attB, *locIII*	TTaTtcaaCAAcggAaatTcCag	**cAgtAcaggGCCaGcGTAAGGGAtATa** aAGcGaacgtgtttTttGttTTTggttcc
attB, *locH*	aTcagaccattcgccgTtgcatc	**ggAAAtaAAGCtgtTGTAAGGGcGtTc** ataaaTtTCctgAcacaGtGtTTctgtttc

FIG. 4. Comparison of phage and bacterial *att* sequences. The DNA sequences of *attP* and *attB* of *locI* are from reference 40, and the others are deduced from the sequences in Fig. 3. The core region is indicated by bold letters, and nucleotides identical to those in the *attP* sequence are indicated by capital letters. Nucleotides outside the core region of *attB* from location II, III, or H that are identical to *attB* of *locI* are underlined. Note that a shift of 1 nt had to be introduced on the right side of the core region of *locII* to get the homology of the TG-rich region.

principle is that the DNA is digested with a restriction enzyme and the fragments obtained are circularized before amplification by using primers synthesized from the known DNA sequence in the opposite orientations to those normally employed for PCR. DNA was extracted from P2 lysogenic strains known to carry P2 in location II [C-1094, referred to here as C(P2)II; C-1215, referred to here as C(P2*c*)II; and LG102, referred to here as K(P2)II], location III [C-77, referred to here as C(P2)III], and location H [K-207, referred to here as K(P2)H]. For a full description of the strains, see Table 2. To amplify the prophage host junction containing *attL*, the DNA was first cleaved with *Sau*3AI, which cuts 393 nt to the right of the *attP* core sequence, i.e., within the P2 *int* gene, and after ligation, the pertinent region was amplified by using primers 72.5-l and 73.5-r, which are located in a back-to-back orientation within the coding part of the *int* gene (Fig. 1 and 2a). The PCR reaction gave several bands, which were cloned into pUC18 without separation of the individual fragments. After colony hybridization against primer 72.4-l, which was not used for the PCR reaction, the clones containing P2 DNA could be isolated and sequenced.

In all three strains containing P2 integrated in location II, a fragment which contained P2 DNA, the *attL* core sequence, and 167 nt of flanking host DNA was cloned (Fig. 3a). The only difference found between the strains was that C(P2*c*)II, as opposed to C(P2)II and K(P2)II, had a T instead of a G at position +3 of the *attL* core sequence. Whether this difference in strain C(P2*c*)II is due to a strain difference or a mismatch repair during integration is not known. It should be noted that the *saf* mutation has a G residue replacing the T at position +3 in the *attP* core sequence (40). Amplification of the prophage-host junction fragment from C(P2)III gave a fragment containing P2 DNA, *attL*, and 125 nt of flanking host DNA (Fig. 3b), and amplification of the prophage-host junction from K(P2)H gave a fragment containing P2 DNA, *attL*, and 153 nt of flanking host DNA (Fig. 3c).

Isolation and sequence determination of host DNA sequences flanking *attR* in lysogens having the P2 prophage integrated in location II, III, or H. As there is no *Sau*3AI site in the known DNA sequence to the left of *attP*, the DNA from the respective lysogen was cleaved instead with *Alu*I before ligation (Fig. 1). *Alu*I cleaves 417 nt to the left of the *attP* core sequence (Fig. 2a). After ligation, the prophage-host junction circles containing *attR* were amplified by using primers 70.9-l and 71.5-r (Fig. 2a). The PCR reaction yielded only one fragment each from strains C(P2)II, C(P2*c*)II, and C(P2)III, while strain K(P2)II gave two different fragments. Strain K(P2)H also gave two different fragments, one of which was the same size as a fragment obtained from strain K(P2)II. The amplified DNA was cloned, hybridized against primer 72.0-r (which was not used for the PCR amplification) (Fig. 2), and sequenced.

Strain C(P2)II gave only one type of insert in all isolates. This contained P2 DNA, *attR*, and more than 360 nt of flanking host sequences (Fig. 3a). Strain K(P2)II, however, gave isolates which had the flanking host sequences found in *locI* of *E. coli* C. Strain C(P2)III gave only one type of insert in all isolates, which contained P2 DNA, *attR*, and about 440 nt of flanking host sequences (Fig. 3b). Strain K(P2)H gave two types of inserts, one type containing P2 DNA, *attR*, and the 120 nt of flanking host sequences shown in Fig. 3c and the other containing the same flanking host sequences as *locI* of *E. coli* C, i.e., like strain K(P2)II.

Sequence comparison of the *attB* core sequences with the *attP* core sequence. By aligning the *attL* and *attR* sequences, as shown in Fig. 3, the host *attB* sequences of locations II, III, and H can be constructed and compared with those of *locI* and *attP* (40), presuming that the mismatches originate from the *attB* core region. As can be seen in Fig. 4, *locII* has 20 nt identical to the *attP* core sequence, and *locIII* and *locH* have 17 nt identical to the *attP* core sequence. Thus, the P2 integrase will accept at least up to 37% mismatches within the 27-nt-long core sequence. There are no detectable sequence similarities outside the core regions between the *attP* and *attB* sequences, but a comparison of the *attB* regions shows that *locI* and *locH* have a region of 23 nt with 50% identity on the left side of the core region and that locations I, II, III, and H seem to have a homologous TG-rich region (Fig. 4) on the right side of the core region.

To ensure that no mismatch repair had occurred during integrative recombination, the *attB* regions of *locII*, *locIII*, and *locH* were amplified from the nonlysogenic strain C-600r$^-$m$^+$, a commonly used K-12 strain (2, 29), by using the primers shown in Fig. 3. The amplified fragments were cloned into pUC18 and sequenced. In *locII* of strain C-600r$^-$m$^+$, the mismatched T at position +7 was found to be replaced by a C, which also is a mismatch compared with *attP* and probably reflects a difference between the strains used. In *locII* and *locIII*, the *attB* core sequences were identical to those deduced from the *attL* and *attR* sequences. Thus, no mismatch repair has occurred in the lysogens studied, besides possibly in strain C(P2*c*)II, as described above.

P2 has been found integrated in location H in *E. coli* K-12 strains but so far never in *E. coli* C (Table 1). To test whether *locH* is present in *E. coli* C, primers on either side of the core sequence of *locH*, i.e., locH-r and locH-l (Fig. 3), were used to amplify DNA from an *E. coli* C strain (C-1a) and a K-12 strain (C-600r$^-$m$^+$). The expected DNA fragment of 190 nt was found only with DNA from the K-12 strain. Thus, *E. coli* C seems to lack *locH*. P2 has been found in location III in *E.*

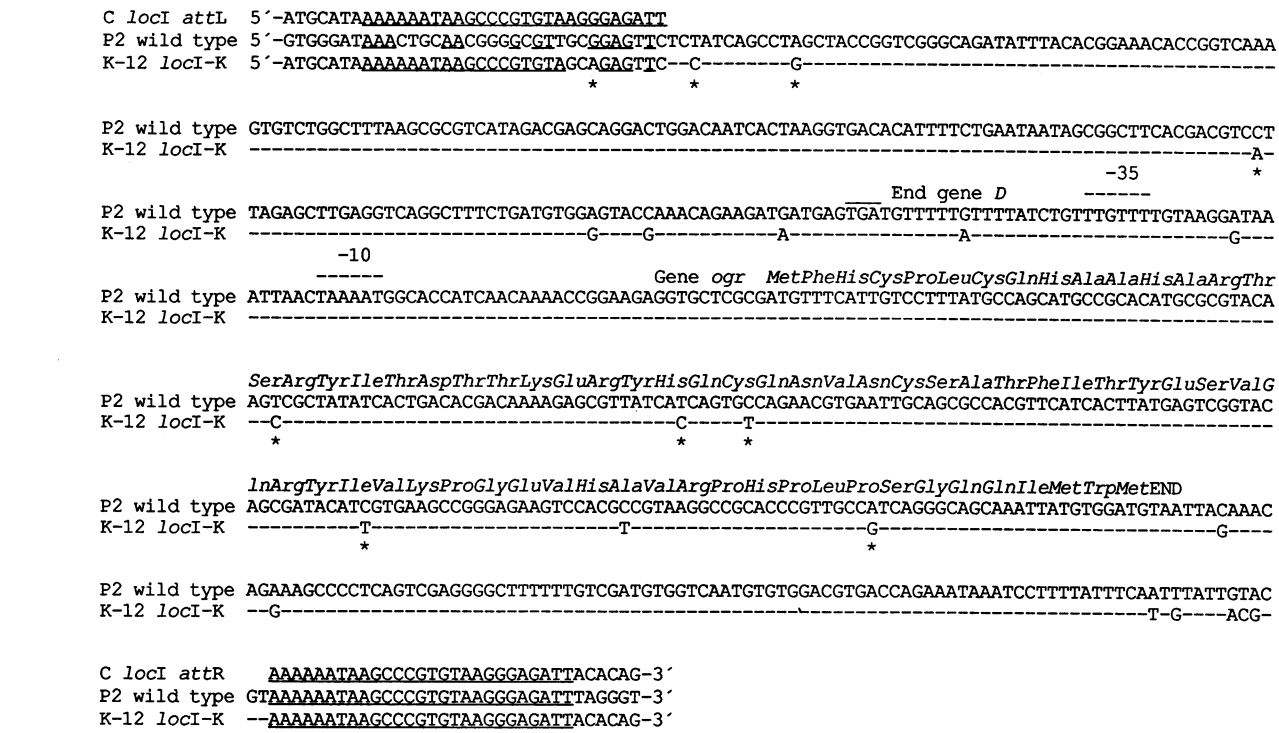

FIG. 5. DNA sequence of the cryptic P2 integrated into *locI* compared with the P2 wild-type sequence and *attL* and *attR* of *locI* in *E. coli* C. The *attL* and *attR* sequences of *locI* are given on the top row at the start and the end of the sequence. The P2 wild-type sequence (containing part of gene *D* [10], all of gene *ogr* [14], and *attP*) is shown above the cryptic P2 sequence found in *locI* of *E. coli* K-12 (indicated as K-12 *locI*-K). Only nucleotides that differ from the P2 wild-type sequence are shown. The stars indicate that the base substitutions are silent mutations. The promoter and amino acid sequence of the *ogr* gene are also indicated above the sequence. Nucleotides identical to the core sequence of *locI* are underlined.

coli C strains and possibly also in K-12 strains (Table 1). When primers on either side of *locIII*, i.e., locIII-r and locIII-l, were used to amplify DNA from a K-12 and a C strain, the expected fragment of 435 nt was found with DNA from both strains. Hence, *locIII* is present in *E. coli* K-12 as well as in C strains (data not shown).

***E. coli* K-12 has a cryptic P2 integrated into *locI*-K.** One possible interpretation of the fact that the right prophage-host junction of *locI* is found in strains K(P2)II and K(P2)H is that these strains carry a defective P2 prophage integrated into a site similar in DNA sequence to *locI* of *E. coli* C. Since it is not known whether these sites are at the same map location, this *E. coli* K-12 site is referred to as *locI*-K. To test this possibility, two primers from either side of *attB* of *locI*, i.e., locI-r and locI-l (Fig. 2b), were used to amplify location I from the control strain C-1a, which should give a fragment of 445 nt containing *attB*, and from strains K-40 [the nonlysogenic parent of strain K(P2)H], K(P2)II, and C-600r$^-$m$^+$. Strain C-1a gave a fragment of about 450 nt as expected, while the K-12 strains gave a fragment of about 1,000 nt (data not shown). The PCR-amplified fragments from the strains were inserted into pUC18 and sequenced. The amplified C-1a fragment contained the expected *attB* region, while the fragments from strains K-40, K(P2)II, and C-600r$^-$m$^+$ contained a piece of P2 DNA, including the C-terminal part of the *D* gene, all of the *ogr* gene, and *attR* of *locI* (Fig. 5). As can be seen in Fig. 5, it seems as if there has been a recombination event between *attL* and a region within the *D* gene which has 14 nt identical to the 27-nt core sequence of *attL*. There are some differences between the P2 wild-type sequence and the sequence of the cryptic P2 in *locI*-K. There are 21 base substitutions, of which 6 are within the coding part of the *ogr* gene. However, only one of these will give rise to a new amino acid in the Ogr polypeptide (Fig. 5).

Location of the attachment sites in the physical map of *E. coli*. Kohara et al. (20) have compiled a physical map of the *E. coli* genome by assembling a collection of specialized transducing λ phages carrying identified, overlapping *E. coli* chromosomal fragments.

Location II has been mapped between *metE* and *rha*, i.e., 86.3 and 88.6 min, respectively (4), which should correspond to λ clones 541 to 552 in Kohara's miniset. An initial screening of these clones by PCR with primers on either side of the core sequence (Fig. 2a) gave the expected 358-nt-long fragment only with clones 547 and 548 (data not shown). Therefore, phage stocks of λ clones 546 to 549 were made, and the DNA was extracted and analyzed by dot blot hybridization against ^{32}P-labeled probe locII-r. As can be seen in Fig. 6, only clones 547 and 548 hybridized against this probe. Thus, the P2 attachment site II should be located within the overlap of clones 547 and 548, i.e., between positions 3654 and 3672. This region contains several *Pst*I sites, all located around position 3670. As a *Pst*I site is located only 50 nt from the core sequence of *locII*, its location must be around position 3670.

Location III has been located between *trp* and the terminus on the map of *E. coli* C (18, 39). On the genetic map of *E. coli* K-12 (4), this should be between 28 and 34.25 min, which should correspond to λ clones 254 to 305 in the miniset. The initial screening of these clones by PCR with

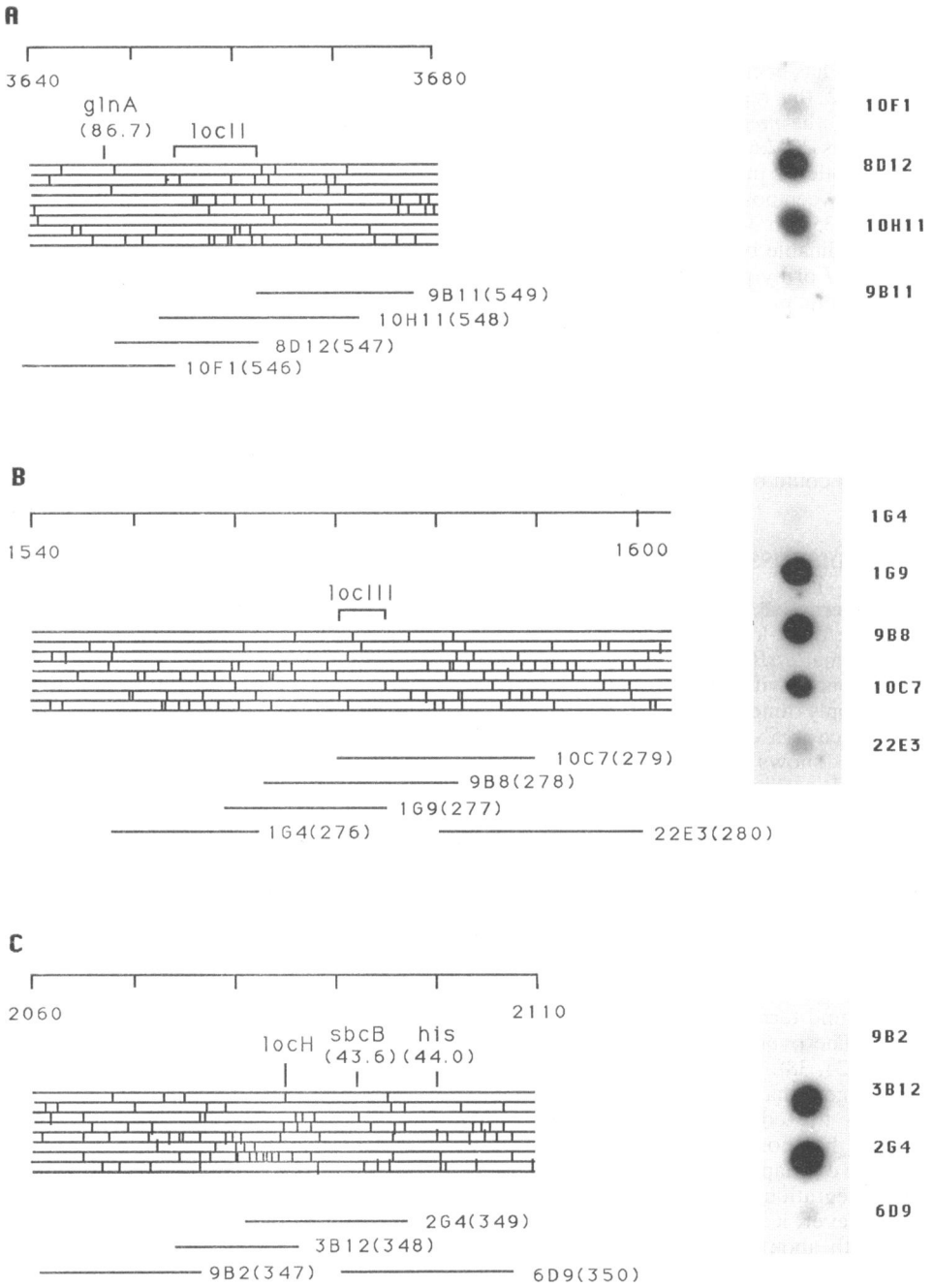

FIG. 6. Locations of *locII* (a), *locIII* (b), and *locH* (c) on the physical map of *E. coli* (20). The format of the restriction map shows, from the top row, *Bam*HI, *Hin*dIII, *Eco*RI, *Eco*RV, *Bgl*I, *Kpn*I, *Pst*I, and *Pvu*II. The results of dot blot analyses are shown to the right of the maps.

primers on either side of the core sequence (Fig. 3a) gave the expected 435-nt-long fragment only with clones 277, 278, and 279 (data not shown). Therefore, phage stocks of λ clones 276 to 280 were made, and the DNA was extracted and analyzed by dot blot hybridization against ^{32}P-labeled probe locIII-r. As can be seen in Fig. 6, only clones 277, 278, and 279 hybridized against this probe. Thus, the P2 attachment site II should be located within the overlap of clones 277, 278, and 279, i.e., between position 1570 and 1575.

Location H has been located between *shiA* and *his*, which are at about 43.4 and 44 min, respectively (4), and this should correspond to a region covered by λ clones 344 to 350. The initial screening of these clones by PCR with primers on either side of the core sequence (Fig. 3a) gave the expected 190-nt-long fragment only with clones 348 and 349 (data not shown). Phage stocks were therefore made of λ clones 347 to 350, and the extracted DNA was analyzed by dot blot hybridization against the ^{32}P-labeled probe locH-r. As can be seen in Fig. 6, only clones 348 and 349 hybridized against this probe. Thus, *locH* should be located within the overlap of these clones, i.e., between positions 2080 and 2090. Since this region contains only one *Bam*HI site and since *locH* has

a *Bam*HI site 60 nt from the core sequence, *locH* must be at about 2085.

Location I in *E. coli* C has been located between genes *his* and *metG*. In *E. coli* K-12, this region has given controversial mapping results (4), but the *metG* gene has recently been located between positions 2202.6 and 2204.8 on the physical map (15). Thus, the cryptic P2 prophage located in *locI*-K is expected to be found between positions 2100 and 2202, which is covered by λ clones 350 to 360. However, these clones contain no material amplifiable by primers on either side of the core sequence of *locI* or by primers from the cryptic P2 and either side of the cryptic prophage. Plaque hybridizations against ^{32}P-labeled locI-r and locI-l primers were also negative. To test whether *locI* has been translocated to another part of the genome in *E. coli* W3110, from which the collection of transducing λ phages were obtained (20), a membrane containing the whole collection of transducing λ phages was hybridized to the ^{32}P-labeled locI-r, locI-l, and core-l. However, no positive signals could be detected.

DISCUSSION

Core sequence in alternative attachment sites. The core sequence defined as the 27-nt-long sequence identical in P2 *attP* and the chromosomal *attB* in *locI* shows a number of mismatches when compared with alternative *attB* sites of *E. coli* K-12. There are only nine conserved bases, and the consensus core sequence is 5′----A----GC----G-AAG-G---T-3′. The core sequence shows no symmetry, and as the sequenced *attL* and *attR* regions show variations up to the conserved AAG sequence, the point of crossover is likely to be located here. The P2 integrase belongs to the Int family of site-specific recombinases (3). This group of proteins show a large diversity of sequences, but they seem to share the same features in strand cleavage and exchange reactions. Three members of this family have been studied in detail, i.e., the λ Int, P1 Cre, and yeast 2μm circle FLP proteins, and they have been shown to nick their substrates at specific positions to create 3′ protein-bound termini and 5′ overhangs with a free hydroxyl group. The overlap regions are at least 6 nt long, i.e., 7 nt for Int (24, 38), 6 nt for Cre (17), and 8 nt for FLP (1). Our data suggest that one of the strand exchange points is within or at the edges of the conserved AAG sequence, but as *locII* has nonhomologies at both sides of the AAG sequence, the overlap region in P2 might be shorter than 6 nt, or else integration into *locII* involves overlap region mismatch. However, it should be noted that *locII* and *locH* are occupied with about equal probability in *E. coli* K-12 (19), and in the case of phage λ, sequence homology between recombining *att* sites within the overlap region is required for effective recombination (5).

Presence of a cryptic P2 in *locI*-K of *E. coli* K-12. The *E. coli* K-12 strains used in these studies contain a cryptic P2 in *locI*-K. This suggests that an ancestor of *E. coli* K-12 was lysogenized by P2 and that an imprecise excision event which removed most of the phage DNA then occurred, leaving only part of gene *D*, all of gene *ogr*, and *attR* in the chromosome (Fig. 5). This unequal recombination event would have involved *attL* and a region within the *D* gene. Indeed, gene *D* offers 14 matches to the 27-nt core sequence in the region where the recombination is presumed to have occurred. Even though the *ogr* gene has accumulated six base substitutions, only one changes an amino acid in the polypeptide (Ala-57 to a Val). Since the *ogr* gene has a promoter located in the spacer region between genes *D* and *ogr* (14, 28), the *ogr* gene is probably expressed. In fact, an *ogr* activity has been observed in *E. coli* B and K-12 strains (34).

That a cryptic P2 is present in *locI*-K explains why P2 cannot stably integrate into this site. The P2 *int* transcript which is initiated at the Pc promoter, located to the right of the *C* repressor gene, proceeds across the *attP* core sequence. In the phage structure, it is terminated after the *attP* sequence, while in the prophage structure, it proceeds into the host chromosome. Therefore, it has been hypothesized that a shorter half-life of the *int* transcript of the prophage leads to a very low *int* gene expression in the prophage compared with the unintegrated phage (40). Integration of P2 into *locI*-K, however, generates a normal *int* transcriptional stop signal, as in the phage structure, because of the presence of the cryptic P2, and the expression of the *int* gene will be as high as with an unintegrated phage. As a consequence, the structure would be unstable, which would explain why it has been impossible to establish a stable P2 prophage in *locI*-K. It is known that double lysogens with two P2 prophages in tandem at the same bacterial attachment site segregate single lysogens at a high frequency unless the *int* gene in the center of the structure is inactivated by a mutation (7, 13).

When prophage P2 occupies site H in *E. coli* K-12, it occasionally generates P2 *int*-dependent deletions that include genes *sbc*, *his*, *gnd*, *rfb*, *udk*, and *dcd* and reach into the *mglP* locus (25, 35, 36). The *mglP* locus is located at kb 2247 on the physical map (20), and *metG* has recently been located between kb 2202.6 and 2204.8 (15). Location I in *E. coli* C has been located between *his* and *metG*, but as the gene order on the physical map is *his-metG-mglP*, it is hard to explain the phenotype of the eductants, presuming that they have arisen from a recombination event between P2 in *locH* and the cryptic P2 in *locI*-K. One possibility is that the gene orders in the strains used were different. This region has given controversial mapping results, and it is possible that the length and orientation of part of this region are not the same in all strains (4). Alternatively, the *mglP* locus is not physically lost in the eductants, but its expression is altered. In fact, it has been observed that the expression of *metG* is altered in the P2 eductants (12). Our attempt to locate *locIJK* in this region on the physical map of K-12 has failed. In fact, the restriction map of *E. coli* C around *locI* does not match the K-12 map in the region between *his* (kb 2100) and *metG* (kb 2203) (8). Thus, it is possible that some *E. coli* K-12 strains lack the region corresponding to *locI*. Differences in this region in *E. coli* C and K-12 strains have been noted. Strain C contains the ribitol-arabitol genes between *metG* and *his*, while K-12 contains the galacticol genes, and these genes are mutually exclusive when exchanged between *E. coli* C and K-12 (21).

ACKNOWLEDGMENTS

This work was supported by grant 72 from the Swedish Medical Research Council.

We are grateful to Elizabeth Bertani, Nils-Kåre Birkeland, Kirsti Gebhardt, Ewy Kristiansen, Björn Lindqvist, and Audun Slettan for helpful discussions and for communication of unpublished results and to Giuseppe Bertani and Mel Sunshine for reading the manuscript and offering many helpful suggestions.

REFERENCES

1. **Andrews, B. J., G. A. Proteau, L. G. Beatty, and P. D. Sadowski.** 1985. The FLP recombinase of the 2μ circle DNA of yeast: interaction with its target sequences. Cell **40**:795–803.
2. **Appleyard, R. K.** 1954. Segregation of lambda lysogenicity during bacterial recombination in *E. coli* K-12. Genetics **39**:429–437.

3. **Argos, P., A. Landy, A. K. Abremski, J. B. Egan, E. Haggård-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. L. Narayana, L. S. Pierson III, N. Sternberg, and J. M. Leong.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. **5**:433–440.
4. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
5. **Bauer, C. E., J. F. Gardner, and R. I. Gumport.** 1985. Extent of sequence homology required for bacteriophage lambda site-specific recombination. J. Mol. Biol. **181**:187–197.
6. **Bertani, G., and E. Six.** 1958. Inheritance of prophage P2 in bacterial crosses. Virology **6**:357–381.
7. **Bertani, L. E.** 1971. Stabilization of P2 tandem double lysogens by *int* mutations in the prophage. Virology **46**:426–436.
8. **Bertani, L. E.** Personal communication.
9. **Bertani, L. E., and G. Bertani.** 1971. Genetics of P2 and related phages. Adv. Genet. **16**:199–237.
10. **Bertani, L. E., and E. W. Six.** 1988. The P2-like phages and their parasite, P4, p. 73–143. *In* R. Calendar (ed.), The bacteriophages, vol. 2. Plenum Publishing Corp., New York.
11. **Calendar, R., and G. Lindahl.** 1969. Attachment of prophage P2: gene order at different host chromosomal sites. Virology **39**: 867–881.
12. **Cassio, D., Y. Mathien, and J. P. Waller.** 1975. Enhanced level and metabolic regulation of methionyl-transfer ribonucleic acid synthetase in different strains of *Escherichia coli* K-12. J. Bacteriol. **123**:580–588.
13. **Choe, B. K.** 1969. Integration defective mutants of bacteriophage P2. Mol. Gen. Genet. **105**:275–284.
14. **Christie, G. E., E. Haggård-Ljungquist, R. Feiwell, and R. Calendar.** 1986. Regulation of the P2 late-gene expression: the *ogr* gene. Proc. Natl. Acad. Sci. USA **83**:3238–3242.
15. **Dardel, F., M. Panvert, S. Blanquet, and G. Fayat.** 1991. Locations of the *metG* and *mrp* genes on the physical map of *Escherichia coli*. J. Bacteriol. **173**:3273.
16. **Haggård-Ljungquist, E.** 1990. Bacteriophage P2, p. 1.63–1.69. *In* S. O'Brien (ed.), Genetic maps, 5th ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. **Hoess, R. H., and K. Abremski.** 1984. Interaction of the bacteriophage P1 recombinase Cre with the recombining site *loxP*. Proc. Natl. Acad. Sci. USA **81**:1026–1029.
18. **Jonasson, J.** 1973. Evidence for bidirectional chromosome replication in *Escherichia coli* C based on marker-frequency analysis by DNA/DNA hybridization with P2 and λ prophages. Mol. Gen. Genet. **120**:69–90.
19. **Kelly, B.** 1963. Localization of P2 prophage in two strains of *Escherichia coli*. Virology **19**:32–39.
20. **Kohara, Y., K. Akiyama, and K. Isoni.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
21. **Link, C. D., and M. A. Reiner.** 1983. Genotypic exclusion: a novel relationship between the ribitol-arabitol and galacticol genes of *E. coli*. Mol. Gen. Genet. **189**:337–339.
22. **Ljungquist, E., and A. I. Bukhari.** 1977. State of prophage Mu DNA upon induction. Proc. Natl. Acad. Sci. USA **74**:3143–3147.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Mizuuchi, K., R. Weisberg, L. Engquist, M. Mizuuchi, M. Buraczynska, C. Foeller, P.-L. Hsu, W. Ross, and A. Landy.** 1980. Structure and function of the phage λ *att* site; size, int-binding sites and location of the crossover point. Cold Spring Harbor Symp. Quant. Biol. **45**:429–437.
25. **Neuhard, J., and E. Thomassen.** 1976. Altered deoxyribonucleotide pools in P2 eductants of *Escherichia coli* K-12 due to deletion of the *dcd* gene. J. Bacteriol. **126**:999–1001.
26. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene **26**:101–106.
27. **Ochmann, H., A. S. Gerber, and D. L. Hartl.** 1988. Genetic applications of an inverse polymerase chain reaction. Genetics **120**:621–625.
28. **Pritchard, M., and J. B. Egan.** 1985. Control of gene expression in P2-related coliphages: the in vitro transcription pattern of coliphage 186. EMBO J. **4**:3599–3604.
29. **Raleigh, E., and G. Wilson.** 1986. *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA **83**:9070–9074.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
31. **Sasaki, I., and G. Bertani.** 1965. Growth abnormalities in Hfr derivatives of *Escherichia coli* strain C. J. Gen. Microbiol. **40**:365–376.
32. **Silver, J., and V. Keerikatte.** 1989. Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J. Virol. **63**:1924–1928.
33. **Six, E.** 1966. Specificity of P2 for prophage site I on the chromosome of *Escherichia coli* strain C. Virology **29**:106–125.
34. **Slettan, A., K. Gebhardt, E. Kristiansen, N.-K. Birkeland, and B. H. Lindqvist.** 1992. *Escherichia coli* K-12 and B contain functional bacteriophage P2 *ogr* genes. J. Bacteriol. **174**:4094–4100.
35. **Sunshine, M. G., and B. L. Kelly.** 1967. Studies on P2 prophage-host relationships. I. Alteration of P2 prophage localization patterns in *Escherichia coli* by interstrain transduction. Virology **32**:644–653.
36. **Templin, A., S. R. Kushner, and A. J. Clark.** 1972. Genetic analysis of mutants indirectly suppressing *rec*B and *rec*C mutations. Genetics **72**:205–215.
37. **Triglia, T., M. G. Peterson, and D. J. Kemp.** 1988. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. **16**:8186.
38. **Weisberg, R., L. W. Engquist, C. Foeller, and A. Landy.** 1983. Role for DNA homology in site specific recombination. The isolation and characterization of a site affinity mutant of coliphage λ. J. Mol. Biol. **170**:319–342.
39. **Wiman, M., G. Bertani, B. Kelly, and S. Sasaki.** 1970. Genetic map of *Escherichia coli* strain C. Mol. Gen. Genet. **107**:1–31.
40. **Yu, A., L. E. Bertani, and E. Haggård-Ljungquist.** 1989. Control of prophage integration and excision in bacteriophage P2: nucleotide sequences of the *int* gene and *att* sites. Gene **80**:1–12.