Control and Regulation of KplE1 Prophage Site-specific Recombination

A NEW RECOMBINATION MODULE ANALYZED

Received for publication, March 2, 2007, and in revised form, May 22, 2007 Published, JBC Papers in Press, May 31, 2007, DOI 10.1074/jbc.M701827200

Gaël Panis1, Vincent Méjean, and Mireille Ansaldi2

From the Laboratoire de Chimie Bactérienne, Institut de Biologie Structurale et Microbiologie, CNRS, 31 chemin Joseph Aiguier, Marseille 13402, Cedex 20, France

KplE1 is one of the 10 prophage regions of Escherichia coli K12, located at 2464 kb on the chromosome. KplE1 is defective for lysis, but it is fully competent for excisive recombination. In this study, we have mapped the binding sites of the recombination proteins, namely IntS, TorI, and IHF on attL and attR, and the organization of these sites suggests that the intasome is architecturally different from the λ canonical form. We also measured the relative contribution of these proteins to both excisive and integrative recombination by using a quantitative in vitro assay. These experiments show a requirement of the TorI excisionase for excisive recombination and of the IntS integrase for both integration and excision. Moreover, we observed a strong influence of the supercoiled state of the substrates. The KplE1 recombination module, composed of the integrated and excisionase genes together with the attL and attR DNA regions, is highly similar to that of several phages infecting various E. coli strains as well as Shigella flexneri and Shigella sonnei. The in vitro recombination data reveal that HK620 and KplE1 att sequences are exchangeable. This study thus defines a new site-specific recombination module, and implications for the mechanism and regulation of recombination are discussed.

Phage λ has long served as a model system for studies of regulated site-specific recombination (1). Indeed, bacteriophages such as λ may choose between the lytic and the lysogenic cycle for their propagation in the bacterial host (2). In conditions favorable for bacterial growth the phage genome is inserted into the host genome by an integrative recombination reaction, which takes place between DNA attachment sites called attP and attB, in the phage and the bacterial genome, respectively. As a result, the integrated λ DNA (or prophage DNA) is bounded by hybrid attachment sites, termed attL and attR. In response to a change in the physiological state of the bacterium, mainly in response to stress conditions such as DNA damage, λ phage DNA is excised from the host chromosome. This excisive recombination recombines attL and attR to restore the attP and attB sites on the circular λ and Escherichia coli genomes (recently reviewed in Ref. 3). Although the phage-encoded integrase (Int)3 protein catalyzes both integrative and excisive reactions, it requires the assistance of several accessory proteins depending on the direction of the reaction (4). The host-encoded integration host factor (IHF) is required for both integration and excision, whereas the phage-encoded excisionase (Xis) is necessary for excision only and prevents re-integration (5–7). Excisionase binds and bends DNA, assists the formation of the intasome, and controls directionality of the reaction toward excision (8, 9). Moreover, directionality of the excisive reaction is conferred by the irreversibility of multiple reaction steps (10). The phage-encoded integrase is a heterogeneous DNA-binding protein. It consists of a large C-terminal domain that binds to core-type DNA sequences and carries out the enzymatic steps of recombination (11, 12). Additionally, Int contains a small N-terminal domain that binds to arm-type DNA sites that are distant from the sites of DNA strand exchange (13). Recombination initiates with the pairing of two specific DNA segments by a tetramer of recombinase molecules. A four-way DNA junction (Holliday junction) is formed by the cleavage, exchange, and ligation of one pair of strands (14). The switch in DNA cleavage activity from one pair of DNA strands to the other, which allows the Holliday junction to be resolved, results from isomerization of the recombinase tetramer (15, 16).

Temperate phages have some importance in mediating gene transfer and inactivation (17–19). Gene disruption can occur by prophage integration into the bacterial genome. Prophage genes can confer selective advantage through a superinfection mechanism, and they often increase the pathogenic properties of some strains for example through the acquisition of toxin genes. In the lysogenic state, prophages occasionally mutate to some of the functions essential for lytic growth, in which case the strain is no longer able to liberate infectious particles (20). Such prophages are termed defective, but nevertheless they can inactivate host genes or conserve genes important for host pathogenicity or fitness.

The KplE1 prophage (also named CPS-53) is one of the 10 prophage regions present in E. coli K12 MG1655 (21, 22). The boundaries of KplE1 prophage are 16-bp-duplicated core sequences (CTGCAGGGGACACCAT) separated by 10.216 kb outlined by the argW tRNA gene and the dsdC gene, and com-

1 A fellowship recipient from the French Research Ministry.
2 To whom correspondence should be addressed: Tel.: 33-4-91-16-45-85; Fax: 33-4-91-71-89-14; E-mail: ansaldi@ibsm.cnrs-mrs.fr.
3 The abbreviations used are: Int, integrase; IHF, integration host factor; RDF, recombination directionality factor; AU, arbitrary unit(s); Xis, excisionase; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.

© 2007 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.
posed of 16 open reading frames (Fig. 1). The torl gene is the last gene of the KplE1 prophage genome, originally identified using a genetic multiplicity approach as a negative regulator of the torCAD operon that encodes the trimethylamine oxide reductase respiratory system in E. coli (23, 24). The negative effect was due to a previously unidentified small open reading frame (66 amino acids) that we called torl for Tor inhibition. The torl gene is part of the KplE1 prophage genome, because it is located upstream of the 3′ core sequence that marks the end of this prophage. Several lines of evidence lead us to propose Torl as an excisionase protein: (i) it shares several properties with other recombination directionality factor (RDF) proteins such as a small size and a basic pI (25, 26), and (ii) despite no primary sequence homology, the Torl three-dimensional structure is highly homologous to that of ΛXis and Twi16Xis proteins (27). To demonstrate the role of Torl in KplE1 excision, we developed an in vivo excision assay and observed that overexpression of the torl gene rapidly led to the complete excision of KplE1 (27).

In this study, we analyze more precisely the site-specific recombination process of the KplE1 prophage. We show that the IntS integrase promotes KplE1 excision together with the Torl excisionase. We present a detailed analysis of the recombination sites of KplE1 that are highly homologous to that of Sf6 and HK620 phages and of strains K1 RS218, UT189, APEC-O1, and SsO46 prophages. These are the closest relatives of KplE1 considering the site-specific recombination module composed of the integrase and excisionase genes and the recombination DNA sequences attL and attR.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Phage, Plasmids, Primers, Media, and Growth Conditions—Bacterial strains, phage, plasmids, and primers are listed in Table 1. Strains were grown in LB medium and, when necessary, ampicillin (50 μg/ml), chloramphenicol (25 μg/ml), kanamycin (50 μg/ml), or isopropyl 1-thio-β-D-galactopyranoside (1 mM) were added.

Strain Construction—Strain LCB1007 (intS::CmR) is a derivative of strain MC4100 and was constructed by insertion of the chloramphenicol acetyl transferase (cat) gene in the intS gene leading to the deletion of the last 239 codons, according to Datsenko and Wanner (28). Briefly, the cat gene was PCR amplified using the pKD3 as a template with the primer pair ΔintS1/ΔintS2, and the product was recombined into strain MC4100 containing plasmid pKD20. The cat gene was then removed by using the pCP20 plasmid to generate strain LCB1024 (intS, CmR). The truncated intS gene was verified by PCR amplification of the chromosomal region (primer pair Nde-intS/AintS2).

Strain LCB1031 is a derivative of strain ENZ1734 (MG1655 ΔlacZ) where the intS::CmR construction (LCB1007 source) was P1-transduced. The cat gene was then removed to generate strain LCB1032 (intS, CmR). Strains LCB1033 and LCB1034 are derivatives of strains LCB1032 and ENZ1734, respectively, obtained by P1 transduction of the pcnB::TcR marker (LCB506 source).

Strain LCB1019 was constructed by overexpressing the torl gene in the wild-type strain MC4100 (pJFi plus isopropyl 1-thio-β-D-galactopyranoside, 1 mM). The pJFi was then cured by several isolations on LB medium without ampicillin. The presence of attB, and thus the absence of the KplE1 prophage region, was verified by PCR amplification of the chromosomal region (primer pair attL-Spel/attR-XbaI).

Plasmid Construction—To construct plasmid pETintS, the intS coding sequence was PCR-amplified using MC4100 chromosomal DNA as a template. We used the primer Nde-intS that contains an NdeI site followed by the 5′ of the intS coding sequence and the primer Hind3-intS that contains a HindIII site followed by the 3′ of the intS coding sequence. After enzymatic hydrolysis, the PCR product was cloned into corresponding sites of the pET22(b) vector.

To construct plasmid pBrattS, the 5′ coding sequence of intS, as well as the 84 nucleotides upstream, were PCR-amplified using MC4100 chromosomal DNA as a template. We used the primer pr + intS1 that contains a KpnI site followed by a sequence complementary to the putative promoter of intS and the primer Hind3-intS that contains a HindIII site followed by the 3′ end of the intS coding sequence. After enzymatic hydrolysis, the PCR product was cloned into the corresponding sites of pBAD33 vector.

The different plasmid constructions were performed by insertion of att vector products into the pCR2.1 vector (Invitrogen). attL KplE1 and attR KplE1 were PCR-amplified using MC4100 chromosomal DNA as a template with the primer pairs attL-Spel/attL-KpnI and attR-XbaI/attR-IHF2, respectively. attLHK620 and attRHK620 were PCR-amplified with the same primer sites using DNA of a HK620 in vitro integration assay as a template. attPK620 was PCR-amplified using HK620 phage suspension as a template with the primer pair attR-XbaI/attL-Spel. attPKpolE1 was PCR-amplified with the same primer pair using the DNA of an in vitro excision assay as a template. attB was PCR-amplified using LCB1019 chromosomal DNA as a template with the primer pair attR-Xbal/attL-Spel.

Plasmid pG eatL was constructed by insertion of the attL region (region covering the core sequence to the start codon of the intS gene) upstream of a lacZ gene. The attL region was PCR-amplified using the primer pair attL-pro/attL-ter and MC4100 chromosomal DNA as a template. The PCR product was directly cloned into a blunt site (Smal) of the pGE593 vector. Sequence accuracy of the cloned inserts was checked by sequencing. All primer sequences are listed in Table 1.

Protein Purifications—IHF and Torl proteins were overproduced and purified as described (23, 29). IntS was produced from C41(DE3) harboring plasmid pETintS. Cells were grown in LB medium with ampicillin until the A600 reached 0.8 unit, and isopropyl 1-thio-β-D-galactopyranoside (1 mM) was added for 2 h at 37°C. French-pressed extract was equilibrated with 40 mM Tris-HCl buffer (pH 7.4) and loaded onto a HiTrap heparin HP column (Amersham Biosciences). The protein was eluted with a step gradient of KCl and was found in the 0.6 and 0.7 M KCl containing fractions. These fractions were pooled and dialyzed to reach 0.1 M KCl before loading onto a HiTrap SP FF column (Amersham Biosciences) equilibrated with 40 mM Tris-HCl buffer (pH 7.4). The protein was eluted with a step gradient of KCl and was found in the fraction containing 0.4 M KCl. The
protein was then dialyzed against 40 mM Tris-HCl buffer (pH 7.4) containing 10% glycerol.

The purity of TorI, IntS, and IHF was estimated by running the purified fractions onto a 20% SDS-Tris/Tricine-PAGE followed by Coomassie Blue staining. The absence of contamination of TorI and IntS by IHF was checked by Western blot using IHF antisera. The protein concentrations were estimated by the Lowry method.

**DNA Fragment Labeling**—Primers were labeled with [γ-32p]ATP (4000 Ci.mmol⁻¹) using T4 polynucleotide kinase (Promega, Madison, WI). The labeled primers were separated from unincorporated [γ-32p]ATP by using the Nucleotide Removal Kit (Qiagen). attL and attR DNA fragment were PCR-amplified by using MC4100 chromosomal DNA as a template with appropriate labeled and unlabeled primers (attL-pro/attL-later primers for the attL region and attR-Xbal/attR-IHF2 for the attR region).

**DNase I Footprinting**—The footprint assays were performed as follows: the labeled DNA fragment was diluted to a concentration of ~1 nm in 50 μl of binding mix (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 2.5 mM MgCl₂, 0.5 mM dithiothreitol, 4% glycerol, and 40 ng μl⁻¹ poly(dI-dC)-poly(dI-dC)) to which different amounts of the purified IHF, IntS, or TorI (0.03–0.5 μl) were added. After 30 min of incubation at room temperature, DNase I was added (0.07 unit, Invitrogen), and the reaction was conducted for 1 min, then stopped by the addition of 140 μl of DNase Stop Solution (192 mM sodium acetate, 32 mM EDTA, 0.14% SDS, and 64 μg ml⁻¹ yeast RNA). After DNA ethanol-precipitation in the presence of DNA carrier (Pellet Paint co-precipitant, Novagen), the pellets were resuspended in a loading dye solution (95% formamide, 10 mM EDTA, 0.3% bromphenol blue, 0.3% xylene cyanol) and loaded onto a 8% polyacylamide/6% urea electrophoresis gel. The locations of the protected nucleotides were deduced by running a ladder of TorI and IntS by IHF was checked by Western blot using IHF antisera. The protein concentrations were estimated by the Lowry method.

**Quantitative Reverse Transcription-PCR**—The relative abundance of intS mRNA was assessed by real-time reverse transcription-PCR. Two specific primer pairs intS-L/intS-R and intS-L2/intS-R2 with ~50 guanine-cytosine percent and 55 °C melting temperature were designed using the Primer3 software. The first pair (intS-L/intS-R) amplifies a region of 202 bp starting from intS codon 215. The second pair (intS-L2/intS-R2) amplifies a region of 155 bp that starts before the intS start codon (~10 bp) and goes up to the 48th codon; this pair thus allows amplification of intS mRNA in an intS-defective background. Total RNAs from ~10⁹ cells of MC4100 and LCB1024 strains grown aerobically up to mid-log phase were extracted using the PureYield™ RNA Midiprep System (Promega). RNAs were then retrotranscribed with random nonamers using the Superscript II™ RNase H⁻ (Invitrogen) at 42 °C for 50 min, and then treated at 70 °C for 15 min to inactivate the enzyme. The real-time PCR quantifications were performed as described above. The real-time PCR experiments were performed twice, with both independent total RNA and cDNA preparations by the comparative threshold cycle method and results are expressed in arbitrary units (AU).

**β-Galactosidase Assay**—β-Galactosidase activities were measured on whole cells by the Miller method (52) after aerobiotic growth at 37 °C for 6 h. Values represent the averages of at least three independent experiments with no more than 15% variation from the mean.

**RESULTS**

**In Vivo Requirement of the intS Gene for KplE1 Excision**—The IntS integrase is encoded at the 5’ extremity of the KplE1 genome (Fig. 1). To check if intS was essential for KplE1 excision, we constructed an IntS-defective strain (LCB1024). The

---

**KplE1 Site-specific Recombination Module**

---

**In Vivo Excision Assay and intS Complementation**—Strain MC4100 carrying plasmids pFLi and pBAD33 and strain LCB1024 carrying plasmids pFLi and pBpr + intS were grown in LB medium until the A₆₀₀ reached 0.5 unit (0.5 × 10⁹ cells ml⁻¹). Isopropyl-β-D-galactoside (1 mM) was then added for 2 h at 37 °C under agitation. Culture dilutions were prepared and plated onto rich medium containing ampicillin. A basic PCR amplification was performed on 10 random chosen colonies (30-s denaturation at 94 °C, 30-s annealing at 55 °C, and 30-s elongation at 72 °C) using the GoTaq® DNA polymerase (Promega) and specific primer pairs to test the presence of attL, attR, and attB, respectively.

**In Vitro Integration and Excision Assays**—Purified IHF, IntS, and TorI were used in all experiments. All reaction mixtures (25 μl) included linear and/or supercoiled att sites (32 nm) in buffer containing 20 mM Tris-HCl (pH 7.4), 10 mM spermidine, 5 mM EDTA, 1 mg.ml⁻¹ bovine serum albumin, and 5% glycerol. IntS (0.8 μM), IHF (0.8 μM), and TorI (1.6 μM) were added as indicated in the figure legends. The reactions were carried out in optimized conditions: 37 °C for 1 h at an IntS:IHF:TorI protein ratio of 1:1:2. In vitro excision assays, with linear attL and attR products, were separated onto a 2.5% agarose gel after a DNA purification step with the PCR Clean-Up System (Promega).

**Real-time PCR Analysis**—The abundance of attP formed during in vitro excision assays (attL and attR substrates) and the abundance of attR formed during in vitro integration assays (attB and attP substrates) were quantified by real-time PCR. For each reaction, a known concentration of PCR-amplified attP and attR was used as a reference standard for excision and integration assays, respectively. The real-time PCR quantifications were performed at the IBSM genomic facility (Marseille, France) on diluted in vitro reaction products, using the LightCycler instrument and the LightCycler Fast Start DNA masterplus SYBR Green I kit (Roche Applied Science) with external standards as described in Roche Molecular Biochemical technical note LC 11/2000. Different dilutions of each reaction of the in vitro integration and excision assays were then mixed with 0.5 μM of each primer and 2 μl of master mix in a 10-μl final volume. The primer pairs used to quantify attP and attR were attR-XbaI/attL-Spel and attR-XbaI/attR-IHF2, respectively (Table 1). PCR assay parameters were one cycle at 95 °C for 8 min followed by 35 cycles at 95 °C for 10 s, 55 °C for 6 s, and 72 °C for 12 s. Excision and integration percentages were calculated using the initial substrate concentration (32 nm) as the 100% reference (Tables 2 and 3). The real-time PCR results represent the average of at least three independent measurements with no more than 12% variation from the mean.

---

![Image](374x26 to 401x38)
This work JOURNAL OF BIOLOGICAL CHEMISTRY

The 16 open reading frames of the KplE1 prophage are because they belong to the same recombinase family. Only overexpression of the tyrosine recombinase family (30, 31), and in our E. coli FIGURE 1. Strains, phage, plasmids, and primers

TABLE 1

| Characteristic or sequence | Reference |
|---------------------------|-----------|
| araD139 Δ(lacP02ZYa-argF) UI69 rpsL thi | M. J. Casadaban |
| pr + intS | M. Casdel |
| pACYC origin vector containing the pprom promoter | 51 |
| pAC953 containing the promoter (position −4 to +4 relative to the ATG of intS) and the intS coding sequence | This work |
| Vector containing the T7 promoter | This work |
| pAC953 containing attL ′R region (position −22 to +6 relative to the ATG of intS) | This work |
| pET22(b) | Novagen |
| pET22(b) containing torl coding sequence | 23 |
| pET119EHI | This work |
| Vector containing the pprom promoter with a colE1 origin | This work |
| pET119EHI containing torl coding sequence | 23 |
| pKD3 | 28 |
| pKD46 | 28 |
| pTrc-IFH | 28 |
| Plasmid with temperature-sensitive replication and thermal induction of FLP synthesis | 28 |
| pTrc containing hlmA and dhlB in operon | 29 |
| pCR2.1 | Invitrogen |
| pCR2.1 containing attL ′R region (position −28 to +192 relative to the center of attR ′L region) | This work |
| pCR2.1 containing attR ′L region (position −105 to +30 relative to the center of attR ′L region) | This work |
| pCR2.1 containing attR ′L region (position −105 to +30 relative to the center of attR ′L region) | This work |
| pCR2.1 containing attR ′L region (position −105 to +30 relative to the center of attR ′L region) | This work |
| pCR2.1 containing attL ′R region (position −28 to +192 relative to the center of attR ′L region) | This work |
| Primers | This work |
| pr + intS | M. J. Casdel |
| attl-KplI | This work |
| attl-Spel | This work |
| attR-XbaI | This work |
| attl-RHI2 | This work |
| attl-pro | This work |
| attl-ter | This work |
| SmaI | This work |
| intS-L | This work |
| intS-R | This work |
| intS-L2 | This work |
| Nde-intS | This work |
| Hind3-intS | This work |

FIGURE 1. Representation of the KplE1 prophage. The KplE1 prophage is bordered by the argW RNA and the dsdC genes on the E. coli chromosome (2464–2476 kb). The open reading frames of the KplE1 prophage are shown; the arrows indicate the relative position on the prophage as well as the direction of transcription.

E. coli K12 genome contains 10 putative integrase genes of the tyrosine recombinase family (30, 31), and in our in vivo assay only overexpression of the torl gene was sufficient to promote KplE1 excision (27). It could therefore be argued that either one of the integrase gene products could be involved in the excision reaction, because they belong to the same recombinase family. However, in an intS-defective background the expression of the Torl RDF protein did not lead to the excision of KplE1 as was the case in a wild-type background (data not shown). To confirm this result, we performed colony-PCR amplification across the att sequences using specific primers for the recombination substrates attl and attR, and the attB product (Fig. 2). In a wild-type background, no excision of KplE1 could be observed unless the torl gene was expressed from a multicopy plasmid, which may explain how the KplE1 prophage was maintained in the host genome. Yet, in an intS-defective background, no attB sequence could be amplified even if the torl gene was overexpressed and it was only in the presence of the intS gene added in trans that the attB sequence could be detected (Fig. 2, compare lanes 3 and 4). These results thus confirmed the in vivo requirement of both Torl and IntS for proper KplE1 excision. These
sequence (WATCAAN4TTR) was found on the sites. The closest one relative to the canonical IHF binding DNA sequence analysis predicted three putative IHF binding gene and the IHF, we analyzed the involved in site-specific recombination, namely IntS, TorI, and IntS was found apart on the prophage genome, permitting the delimitation of the KplE1 gene and downstream of the tRNA site, and is bordered by a reconstituted plain arrows — represent primers on KplE1 prophage. B, DNA amplification of attB, attL, and attR in MC4100 and LCB1024 (intS) strains transformed with pJFi (TorI; lines 2–4) and pBPi + intS (intS; line 4). + and – indicate the presence or the absence of IntS and TorI proteins, respectively.

experiments also suggested that the intS gene was expressed in a wild-type background and at a sufficient level to allow the excisive recombination to occur. To confirm that the intS gene was expressed in the wild-type strain we performed quantitative reverse transcription-PCR using RNAs extracted from the wild-type and the attL region. The IntS protein covered two different regions, one overlapping the core sequence and spanning positions −33 to +14 relative to the center of the core sequence, and a large region encompassing positions +116 to +175 that overlapped with the four sequences (two direct repeats and two inverted repeats) identified upstream of the intS start codon. These sequences are thus thought to represent integrase arm-type binding sequences. The TorI protein also protected two different regions, one located just downstream of the core sequence covering positions +11 to +43 relative to the center of the core sequence, and another one, with multiple binding upstream of the IntS side-arm binding sites encompassing positions +75 to +107. What was striking about these latter binding sites, was the presence of several DNase I-hypersensitive sites that appeared upon TorI binding in between the protected DNA regions. DNase I-hypersensitive sites are characteristic of a bending of the DNA template, therefore suggesting that the attL DNA was wrapped around TorI. Regarding IHF binding, only a slight protection was observed at the level of one of the degenerate sites identified on attL. In contrast, a strong IHF-protected region was observed on attR at the level of the predicted IHF-binding sequence (positions −53 to −17), but it was much larger than the consensus sequence itself, which is often the case with IHF because of its extreme bending capacity (32, 33). Regarding IntS, once again the attR core
sequence was fully protected, and the covered region extended from position -24 to position +18 relative to the center of the attR core sequence. A second protected region covered the two repeated sequences identified above (positions -93 to -60). No DNA protection appeared upon incubation with TorI, thus indicating the absence of TorI binding on attR. Through these footprinting experiments on both strands, we defined a complete picture of the binding sites for the proteins involved in site-specific recombination, namely IntS, TorI, and IHF (Figs. 4C and 5C). Consensus sequences for arm-type integrase (IntS) and excisionase (TorI) binding sites can thus be defined as GGGTAAAAW and CRGTTCGY, respectively. Although these sequences are found elsewhere on the E. coli chromosome, they do not appear associated with any other prophage.

Conservation of a Site-specific Recombination Module in Various Phages—The KplE1 attP region shows a different site organization than the one found in λ and other related phages such as HK97 and P22. Whereas the attP site of λ (and of most lambdoid phages) lies upstream of the int gene leading to contiguous int and xis genes in the integrated prophage, the attP site of the ancestral KplE1 phage genome actually lies in-between intS and torI genes, leading to a separation of these latter genes and a location at both extremities of the prophage once integrated (Figs. 1 and 6). The central part of the KplE1 prophage genome was found to be highly related to several parts of the phage SfV genome (34, 35). However, the site-specific recombination module, composed of the integrase and excisionase genes, and the attL and attR sequences, is unrelated to SfV. In fact, this module shows a high degree of homology with different phages and prophages of various E. coli strains (TD2158, APEC-O1, UTI89, and RS218), S. flexneri and S. sonnei Ss046. However, no or little homology was observed outside of the recombination module between KplE1 and these latter phages and prophages (36–38). The core sequences located at both extremities of KplE1 in the E. coli K12 genome sequence are 100% identical in phages Sf6 and HK620 and the APEC-O1 prophage, and all of them integrate at the argW site in the respective host genomes. The core sequence is slightly different in the attR region of the RS218, UTI89 (one nucleotide deletion), and Ss046 (one nucleotide substitution) prophage regions but identical in the corre-
Using the ClustalW software, we performed a sequence comparison of the recombination modules on the attP form of these seven (pro)phages, and we observed that these modules are at least 95% identical if the sequences downstream of the integrase gene and upstream of the excisionase gene were excluded (Fig. 6). It is important to note that, among the differences in the attP sequences, only two changes arose inside a protein binding site (HK620, box I2, and Ss046, box I3) but at positions that did not affect the consensus. Nucleotide sequences of the intS and torI genes were at least 98% identical, leading to 98% identity for IntS proteins and 100% identity for TorI proteins. We therefore propose that all features of KplE1 site-specific recombination can be extended to the other phages and prophages mentioned above.

In Vitro Excision Assay—To further characterize KplE1 site-specific recombination we designed an in vitro assay in which we used various DNA substrates and the purified IntS, TorI, and IHF proteins. The attL and attR sequences were either PCR-amplified and used directly as linear substrates or cloned into a cloning vector to obtain circular and supercoiled substrates. Equimolar DNA substrates were incubated with various combinations of recombination proteins (IntS, TorI, and IHF) for 1 h at 37 °C, which constitute the optimal conditions for this reaction in vitro. DNA molecules were then purified to remove the proteins and separated onto an agarose gel to identify the products obtained. Fig. 7 shows the results obtained with linear substrates. In agreement with our in vivo assays, in the absence of either IntS or TorI no attP was produced (Fig. 7, lanes 2 and 3). In contrast, in the presence of both proteins a band corresponding to the attP expected size was observed, indicating that IntS and TorI are sufficient in vitro to promote recombination between KplE1 attL and attR. However, the reaction appeared to reach a maximum efficiency in the presence of IHF, illustrating the positive effect of DNA bending in this process. This experiment also indicated that supercoiling of the DNA substrate was not absolutely required, because efficient recombination occurred between linear molecules as short as 220 bp (attL) and 135 bp (attR).

To evaluate the relative efficiency of the excisive recombination on different substrates, as well as the contribution of the accessory proteins TorI and IHF, we ran the same kind of experiments as described above, except that they were followed by...
real-time PCR analysis as described under “Experimental Procedures.” We calculated the reaction efficiency as the amount of attP produced divided by the amount of substrate provided and expressed in percentage (Table 2). In the presence of linear attL and attR, ~17% of the substrates were converted into attP product, indicating as above that the reaction was possible between linear substrates. However, the amount of attP produced was about five times higher in the presence of at least one circular substrate, and ~3.9 times higher when both substrates were provided as circular molecules. These results clearly indicate that, although linear attL and attR are competent for recombination, the maximum efficiency was reached with at least one circular substrate, regardless of which one was provided in circular form. Surprisingly, when both substrates were provided as circular DNA, the excisive recombination was less efficient than with only one circular substrate.

In all cases, we observed an absolute requirement for the IntS integrase, but the relative contribution of the accessory proteins TorI and IHF were very different. Indeed, whatever the form of the DNA substrates, IHF contribution was minor compared with that of TorI. Excision can be done in the absence of IHF, and the presence of IHF led to a 1.2- to 2.9-fold increase in

---

**TABLE 2**

Efficiency of excision reactions performed with either KplE1 or HK620, attL, and attR substrates

The numbers indicate the amount of substrate transformed into attP product expressed in percentage.

|                      | attL + attR | attL + attR | attL + attR | attL + attR |
|----------------------|------------|------------|------------|------------|
| **KplE1 substrates** |            |            |            |            |
| No protein           | <0.1       | <0.1       | <0.1       | <0.1       |
| IntS                 | <0.1       | <0.1       | <0.1       | <0.1       |
| IntS + TorI          | 9 ± 3       | 70 ± 7     | 67 ± 8     | 2 ± 2      |
| IntS + IHF           | <0.1       | <0.1       | <0.1       | <0.1       |
| IntS + TorI + IHF    | 17 ± 9     | 85 ± 11    | 85 ± 12    | 67 ± 9     |
| **HK620 substrates** |            |            |            |            |
| No protein           | <0.1       | <0.1       | <0.1       | <0.1       |
| IntS                 | <0.1       | <0.1       | <0.1       | <0.1       |
| IntS + TorI          | 13 ± 3     | 47 ± 6     | 69 ± 8     | 10 ± 2     |
| IntS + IHF           | 0.5 ± 0.4  | 0.5 ± 0.2  | <0.1       | <0.1       |
| IntS + TorI + IHF    | 22 ± 8     | 74 ± 8     | 86 ± 12    | 57 ± 9     |

*Slash symbols indicate linear substrates, whereas circles indicate plasmid-cloned substrates.*
KplE1 Site-specific Recombination Module

### TABLE 3
Efficiency of integration reactions performed with either KplE1 or HK620, attB, and attP substrates

The numbers indicate the amount of substrate transformed into attR product expressed in percentage.

| attB + attP | attB + attP O | attBO + attP | attBO + attP O |
|------------|---------------|--------------|---------------|
| **KplE1 substrates** | | | |
| No protein | <0.1 | <0.1 | <0.1 | <0.1 |
| IntS | -0.1 | 6.2 ± 1 | 0.5 ± 0.1 | 11 ± 1 |
| IntS + TorI | -0.1 | 1 ± 0.1 | 1 ± 0.5 | 1 ± 0.5 |
| IntS + IHF | <0.1 | 24 ± 2 | 9 ± 2 | 41 ± 1 |
| IntS + TorI + IHF | 0.5 ± 0.4 | 1 ± 0.3 | 1 ± 0.3 | 1 ± 0.2 |

| HK620 substrates | | | |
| No protein | <0.1 | <0.1 | <0.1 | <0.1 |
| IntS | -0.1 | 6.2 ± 1 | 0.5 ± 0.3 | 5 ± 3 |
| IntS + TorI | -0.1 | 1 ± 0.5 | 1 ± 0.4 | 1 ± 0.5 |
| IntS + IHF | <0.1 | 35 ± 12 | 6 ± 2 | 34 ± 10 |
| IntS + TorI + IHF | 0.5 ± 0.2 | 1 ± 0.6 | 1 ± 0.2 | 1 ± 0.5 |

*Slash symbols indicate linear substrates, whereas circles indicate plasmid-cloned substrates.

Given the homology consistency between the recombination modules of the (pro)phages analyzed (Fig. 6), we can predict that the results would be identical with any substrate origin.

**intS Gene Is Negatively Autoregulated**—In KplE1-excised DNA the attP sequence lies in between the integrase and excisionase genes leading to a separation of these genes at both extremities of the integrated prophage (Fig. 1). The orientation of the intS gene in the integrated KplE1 prophage genome suggested that transcription of intS occurred from a dedicated promoter located downstream of the core sequence and independently of the torI gene, which is located at the other extremity of the integrated prophage. As recently proposed for the P4 int gene (39) we wanted to investigate if the presence of the attL region overlapping with the intS promoter region would interfere with intS transcription. In front of the intS start codon we identified putative −35 and −10 sequences separated by 18 nucleotides and close to the consensus sequences recognized by the σ70 RNA polymerase holoenzyme. These sequences are located from position −66 to −61 (TTGACA) and −42 to −37 (TAAaAaa) relative to the A of the intS start codon. Interestingly, one of the IntS arm-type binding sites overlaps with the putative −10 sequence, and three others are located between the −10 and the ATG of intS (Fig. 4B). To investigate the intS gene autoregulation, we cloned the attL region into the pGE593 vector, which permitted the formation of a transcriptional fusion between the intS promoter region (positions −223 to +65 relative to the A of the intS start codon) and the lacZ gene devoid of promoter. We measured the β-galactosidase activity in the presence or absence of IntS (LCB1034 versus LCB1033) in a pcnB negative background to reduce the copy number of the plasmid carrying the intS-lacZ fusion. Expression of the fusion in the presence of an intact intS gene was low (−4 Miller units). In contrast, in an intS-defective background the measured β-galactosidase activity increased −10 times and up to 44 Miller units, suggesting a negative autoregulation of the intS gene. To confirm this result, we performed real-time PCR on cDNA retrotranscribed from total RNAs isolated from an intS-defective strain and from a wild-type strain. In this experiment the primer pair used for the PCR amplification hybridizes at the beginning of the intS gene, in a region that is not affected by the deletion (see “Experimental Procedures”). Using these primers, we measured a 5-fold increase of the intS mRNA level extracted from the strain devoid of an intact intS gene (96 ± 8 AU) compared with the wild-type strain (19 ± 2.6 AU). All together, these results indicate that the intS gene is negatively autoregulated by IntS, which is consistent with the presence of IntS binding sites on the intS promoter region.

**DISCUSSION**

The KplE1 prophage of E. coli K12 is a defective prophage that cannot form infectious particles, yet it is competent for genome excision (27). Its genome is composed of 16 putative genes (Fig. 1). Part of this prophage is dedicated to the integration and excision process and contains the torI and intS genes, encoding the TorI RDF protein (27) and the KplE1-dedicated integrase, respectively.

It is noteworthy that the KplE1 recombination module, composed of the integrase and excisionase genes as well as the att
regions, is highly homologous to several (pro)phages integrating at the \textit{argW} locus in \textit{E. coli} strains K1 RS218, APEC—O1, UTI89, TD2158, \textit{S. flexneri}, and \textit{S. sonnei} Ss046 (36–38). Indeed, sequence analysis of the recombination module of these phages showed a very high conservation (at least 95% identity) with the corresponding module of KplE1 (Fig. 6). Moreover, the data provided in this study show that HK620 and KplE1 \textit{att} sequences are exchangeable in both integrative and excisive recombination reactions \textit{in vitro} (Tables 2 and 3). It is therefore remarkable that the reactions performed with KplE1 IntS protein and KplE1 or HK620 DNA substrates were in the same range of efficiency. It is thus very likely that similar results can be obtained with DNA substrates originating from the other phages having a similar recombination module. Consequently, we propose that all characteristic features relative to KplE1 site-specific recombination can be extended to the other phages of this subfamily. This situation is contrasting with the case of \textit{\lambda} and HK022, which shows virtually identical \textit{attL} and \textit{attR} sequences, with the exception of a few nucleotides in the extended core sequence. This latter feature confers a high specificity to the reaction, because in this case the DNA substrates are not exchangeable (40).

In all temperate phages, site-specific recombination is believed to take place according to the same mechanism. However, the architecture of the intasome, the nucleoprotein complex competent for the recombination reaction, must differ according to the specific organization of the \textit{att} sequences. Indeed, different organizations are found in various prophages and conjugative transposons (Fig. 8). In this study we provide a complete picture of the binding sites of recombination proteins in the \textit{attL} and \textit{attR} regions of KplE1. We identified four IntS arm-type binding sites and no less than five binding sites for TorI in \textit{attL}. In the \textit{attR} region, two IntS arm-type and one IHF binding sites were mapped (Figs. 4 and 5). Notably, multiple excisionase binding sites seemed to replace the IHF site located on the \textit{P} arm in \textit{\lambda}. Because RDF proteins are known to bend DNA, multiple sites may overcome the lack of IHF to bend the DNA to promote the formation of the intasome. This is consistent with our quantitative \textit{in vitro} experiments that show a weak IHF dependence as well as an absolute requirement of TorI for excisive recombination (Table 2). DNA sequence analysis as well as footprinting experiments led us to identify the IntS and TorI consensus binding sequences. Interestingly, the IntS arm-type consensus sequence contains several adenine bases in a row, which provides a bent DNA substrate in the absence of any bound accessory protein. It is thus likely that intrinsic DNA bending is of particular importance for arm-type binding of the integrase. We also observed an important influ-
KplE1 Site-specific Recombination Module

ence of the supercoiling state of the substrates, however not equivalent, in both excisive and integrative recombination (Tables 2 and 3). Indeed, supercoiled substrates, and in particular attP, were required for integration. In contrast, supercoiling was not mandatory for excision but strongly enhanced the reaction efficiency.

Recombination protein binding sites have been described for a variety of phages and summarized in Fig. 8 (39, 41–43). Considering the RDF proteins, it is noteworthy that multiple binding sites ranging from three in λ and up to seven in P4 are present on the att sites. This suggests, as recently described for λ, that RDF proteins form a nucleofilament important for the intasome function (42). In contrast, the number of P-arm binding sites for the integrase proteins is more constant. A recent study describing the architecture of the 99-bp DNA–six-protein regulatory complex of the λ att site using fluorescence resonance energy transfer technology (44) suggests a mechanism for λ P-arm folding. To better characterize the formation of the particular intasome of the KplE1 subfamily, we need to investigate the chronology of recombination protein binding to their specific sites as well as the interactions that take place between these proteins.

A characteristic, which is unique to the KplE1 recombination module, is the localization of the integrase and excisionase genes in the integrated prophage. In most prophages, the integrase and excisionase genes are either adjacent and co-transcribed under lytic conditions (λ) or located on the same side of the prophage (transcribed in the same direction or not) and generally separated by a few hundred base pairs (Fig. 8). Control of the integrase gene expression is part of the genetic switch that leads to the completion of the lytic cycle. In the λ prophage, the int gene is under the control of two promoters P1 and P2 and is differentially regulated by a retroregulation process that involves mRNA processing (2, 45, 46). In λ prophage, the P2 promoter is controlled by antitermination and allows the expression of xis and int genes in lytic conditions; in contrast the P1 promoter is under positive regulation by λ CII protein and is located inside the xis gene. Therefore, Int but not Xis production is stimulated by λ CII in lysogenic conditions. In the case of the KplE1 prophage family the attP sequence lies in between the integrase and excisionase genes, leading to a separation of these genes at both extremities of the integrated prophage (Fig. 1). This situation implies that the phage-encoded integrase and excisionase are not co-regulated. Considering the intS gene, its orientation suggested that it was transcribed from a dedicated promoter overlapping with the attL region; therefore, the integrase gene is not under the control of a general lytic promoter such as P1. The promoter we seek to define is clearly obstructed by arm-type integrase binding sites (Fig. 4). As a consequence, the intS promoter is under a negative autoregulation control. The physiological relevance of these controls is so far poorly understood in P4 (39) as well as in KplE1 and necessitates further investigation. However, we observed that the effect of overexpression of the KplE1 integrase gene may be dramatically negative on cell growth (data not shown), and negative autoregulation is a very potent way to diminish a toxic effect due to an endonuclease protein. Moreover, under lytic conditions, an excess of integrase could lead to the titration of the RDF protein by protein-protein interaction resulting in an aberrant re-integration of the phage genome into the bacterial chromosome. It has recently been proposed that directionality of site-specific recombination is driven by the irreversibility of multiple steps in the excisive reaction (10). We propose that the down-regulation of the integrase gene may also play a role in directionality by lowering the amount of integrase and avoiding excisionase titration.

Acknowledgments—We thank Y. Denis at the IBSM genomic facility for helpful advice with real-time PCR and J. Hugues for technical assistance with in vitro assays. We are grateful to A. J. Clark and S. Casjens for the kind gift of phage HK620 and E. coli strain TD2158, to S. Goodman for express sending of IFH antisera, and to P. Moreau and F. Bocard for strains, plasmids, and discussion. We also thank A. Manvell for critical reading of the manuscript.

REFERENCES

1. Nash, H. (1996) in Escherichia coli and Salmonella: Cellular and Molecular Biology (Neidhardt, F. C., ed) pp. 2363–2376, ASM Press, Washington, D. C.
2. Court, D. L., Oppenheim, A. B., and Adhya, S. (2007) J. Bacteriol. 189, 298–304
3. Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., and Adhya, S. (2005) Annu. Rev. Genet. 39, 409–429
4. Nash, H. A., and Robertson, C. A. (1981) J. Biol. Chem. 256, 9246–9253
5. Abremski, K., and Gottesman, S. (1982) J. Biol. Chem. 257, 9658–9662
6. Miller, H. L., Kikuchi, A., Nash, H. A., Weisberg, R. A., and Friedman, D. I. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 1121–1126
7. Miller, H. I., and Friedman, D. I. (1980) Cell 20, 711–719
8. Numirych, T. E., Gumport, R. L., and Gardner, J. F. (1992) EMBO J. 11, 3797–3806
9. Swalla, B. M., Cho, E. H., Gumport, R. L., and Gardner, J. F. (2003) Mol. Microbiol. 50, 89–99
10. Mumm, J. P., Landy, A., and Gelies, J. (2006) EMBO J. 25, 4586–4595
11. Tirumalai, R. S., Kwon, H. I., Cardente, E. H., Ellenberger, T., and Landy, A. (1998) J. Mol. Biol. 279, 513–527
12. Kazmierczak, R. A., Swalla, B. M., Burgin, A. B., Gumport, R. I., and Gardner, J. F. (2002) Nucleic Acids Res. 30, 5193–5204
13. Wojcicki, J. M., Sarkar, D., Landy, A., and Chubb, R. T. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 3434–3439
14. Kitts, P. A., and Nash, H. A. (1988) Nucleic Acids Res. 16, 6839–6856
15. Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005) Nature 435, 1059–1066
16. Aihara, H., Kwon, H. J., Nunes-Duby, S. E., Landy, A., and Ellenberger, T. (2003) Mol. Cell 12, 187–198
17. Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M. L., and Brussow, H. (2003) Curr. Opin. Microbiol. 6, 417–424
18. Brussow, H., Canchaya, C., and Hardt, W. D. (2004) Microbiol. Mol. Biol. Rev. 68, 550–602
19. Campbell, A. M. (1996) in Escherichia coli and Salmonella: Cellular and Molecular Biology (Neidhardt, F. C., ed) pp. 2041–2046, ASM Press, Washington, D. C.
20. Rudd, K. E. (1999) Res. Microbiol. 150, 653–664
21. Caisens, S. (2003) Mol. Microbiol. 49, 277–300
22. Ansaldi, M., Théraulaz, L., and Méjean, V. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 9423–9428
23. Méjean, V., Iobbi-Nivol, C., Lepelletier, M., Giordano, G., Chippaux, M., and Pascal, M. C. (1994) Mol. Microbiol. 11, 1169–1179
24. Caisens, S. R. (2005) Curr. Opin. Microbiol. 8, 451–458
25. Lewis, J. A., and Hatfull, G. F. (2001) Nucleic Acids Res. 29, 2205–2216
26. El-Antak, L., Ansaldi, M., Guerlesquin, F., Méjean, V., and Morelli, X. (2005) J. Biol. Chem. 280, 36802–36808
28. Datsenko, K. A., and Wanner, B. L. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 6640–6645
29. Murtin, C., Engelhorn, M., Geiselmann, J., and Bocard, F. (1998) J. Mol. Biol. 284, 949–961
30. Nunes-Duby, S. E., Kwon, H. I., Tirumalai, R. S., Ellenberger, T., and Landy, A. (1998) Nucleic Acids Res. 26, 391–406
31. Swalla, B. M., Gumport, R. I., and Gardner, J. F. (2003) Nucleic Acids Res. 31, 805–818
32. Rice, P. A., Yang, S., Mizuuchi, K., and Nash, H. A. (1996) Cell 87, 1295–1306
33. Ellenberger, T., and Landy, A. (1997) Structure 5, 153–157
34. Allison, G. E., Angeles, D., Tran-Dinh, N., and Verma, N. K. (2002) J. Bacteriol. 184, 1974–1987
35. Huan, P. T., Whittle, B. L., Bastin, D. A., Lindberg, A. A., and Verma, N. K. (1997) Gene (Amst.) 195, 207–216
36. Casjens, S., Winn-Stapley, D. A., Gilcrease, E. B., Morona, R., Kuhlewein, C., Chua, J. E., Manning, P. A., Inwood, W., and Clark, A. J. (2004) J. Mol. Biol. 339, 379–394
37. Clark, A. J., Inwood, W., Cloutier, T., and Dhillon, T. S. (2001) J. Mol. Biol. 311, 657–679
38. Stummeyer, K., Schwarzer, D., Claus, H., Vogel, U., Gerardy-Schahn, R., and Muhlenhoff, M. (2006) Mol. Microbiol. 60, 1123–1135
39. Piazzolla, D., Cali, S., Spoldi, E., Forti, F., Sala, C., Magnoni, F., Deho, G., and Ghisotti, D. (2006) J. Gen. Virol. 87, 2423–2431
40. Dorgai, L., Sloan, S., and Weisberg, R. A. (1998) J. Mol. Biol. 277, 1059–1070
41. Smith-Mungo, L., Chan, I. T., and Landy, A. (1994) J. Biol. Chem. 269, 20798–20805
42. Abbani, M. A., Papagiannis, C. V., Sam, M. D., Cascio, D., Johnson, R. C., and Clubb, R. T. (2007) Proc. Natl. Acad. Sci. U. S. A. 104, 2109–2114
43. Calendar, R. (2006) The Bacteriophages, 2nd Ed., Oxford University Press, New York
44. Sun, X., Mierke, D. F., Biswas, T., Lee, S. Y., Landy, A., and Radman-Livaja, M. (2006) Mol. Cell 24, 569–580
45. Schindler, D., and Echols, H. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 4475–4479
46. Guarnieros, G., Montanez, C., Hernandez, T., and Court, D. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 238–242
47. Eraso, J. M., and Weinstock, G. M. (1992) J. Bacteriol. 174, 5101–5109
48. Moreau, P. L. (2007) J. Bacteriol. 189, 2249–2261
49. Ansaldi, M., Simon, G., Lepelletier, M., and Méjean, V. (2000) J. Bacteriol. 182, 961–966
50. Dhillon, T. S., Poon, A. P., Chan, D., and Clark, A. J. (1998) FEMS Microbiol. Lett. 161, 129–133
51. Guzman, L. M., Belin, D., Carson, M. J., and Beckwith, J. (1995) J. Bacteriol. 177, 4121–4130
52. Miller, J. H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
53. Furste, J. P., Pansegrau, W., Frank, R., Blocker, H., Scholz, P., Bagdasarian, M., and Lanka, E. (1986) Gene (Amst.) 48, 119–131