JOURNAL OF BACTERIOLOGY, Sept. 1997, p. 5705–5711
0021-9193/97/$04.00+0
Copyright © 1997, American Society for Microbiology
Vol. 179, No. 18

Integration Specificities of Two Lambdoid Phages (21 and e14) That Insert at the Same *attB* Site

HUI WANG, CHUNG-HUI YANG, GRACE LEE,† FELICIA CHANG,‡ HILARY WILSON,§ ALICE DEL CAMPILLO-CAMPBELL, AND ALLAN CAMPBELL*

Department of Biological Sciences, Stanford University, Stanford, California 94305

Received 10 February 1997/Accepted 8 July 1997

It was shown previously that phage 21 and the defective element e14 integrate at the same site within the *icd* gene of *Escherichia coli* K-12 but that 21 integrase and excisionase excise e14 in vivo very infrequently compared to excision of 21. We show here that the reverse is also true: e14 excises itself much better than it excises an adjacent 21 prophage. In vitro integrase assays with various *attP* substrates delimit the minimal *attP* site as somewhere between 366 and 418 bp, where the outer limits would include the outermost repeated dodecamers suggested as arm recognition sites by S. J. Schneider (Ph.D. dissertation, Stanford University, Stanford, Calif., 1992). We speculate that the reason 21 *attP* is larger than λ *attP* (240 bp) is because it must include a 209-bp sequence homologous to the 3′ end of the *icd* transcript in order to allow *icd* expression in lysogens. Alteration of portions of 21 *attP* to their e14 counterparts shows that 21 requires both the arm site and core site sequences of 21 but that replacements by e14 sequences function in some positions. Consistent with Schneider's in vivo results, and like all other known integrases from lambdoid phages, 21 requires integration host factor for activity.

The lambdoid phages are a group of natural temperate phages whose best-known member is coliphage λ. They have a common genetic map and are related to one another by frequent natural recombination but exhibit extensive variation in function and nucleotide sequence (6). One example of such variation is seen in their integrase genes, which mediate insertion into the chromosome by site-specific recombination. Each lambdoid phage has a specific preferred site of insertion on the chromosome. Whereas some phages, like λ and 434, insert at the same chromosomal site and closely resemble each other in their integrase and excisionase genes and phage attachment sites, many distinct insertion specificities are found within the group. Between phages that insert at different sites, like λ and 21, the protein components are not interchangeable. Thus, each integrase protein has its own specificity of site recognition. The integrase genes of all lambdoid phages have sufficient sequence similarity to imply origin from a common ancestor, from which the whole spectrum of integration specificities must have evolved.

The biochemical pathway of integrase-mediated site-specific recombination is well known (Fig. 1). For λ, the required extent of specific sequence is about 21 bp for the bacterial partner (*attB*) and 240 bp for the phage partner (*attP*). The actual crossover event takes place between *attB* and a similar sequence (core sequence) within *attP* and entails an exchange between the top two strands, forming a Holliday junction intermediate, followed by a homology-dependent process equivalent to branch migration through a 7-bp overlap (O) segment (identical in *attB* and the *attP* core) and resolution by exchange in the lower two strands at a position displaced 7 bp from the initiating exchange. The overlap segment is flanked by oppositely oriented sequences that are similar to each other, so that *attB* and the *attP* core can be represented as BOB′ and COC′, respectively. B, B′, C, and C′ all approach a consensus sequence and to a first approximation can be considered identical, although the deviations from identity may have functional significance. These sites are recognized by the C terminus of the integrase protein, where the catalytic activity resides.

The N terminus of integrase also recognizes specific DNA sequences present in *attP* (Fig. 1). There are five such "arm" sequences in λ (P1, P2, P1′, P2′, and P3′), all variants of the same consensus. Binding of integrase to these sites is essential for full integrase activity, which also requires supercoiled *attP* DNA and the DNA-bending integration host factor (IHF) protein, which binds at specific locations between core and arm sites. The tight binding of integrase to arm sites presumably positions several integrase molecules to form an appropriate complex with one another and with the *attP* core (intasome), poised to accept an *attB* molecule juxtaposed to the core (13).

Regeneration of *attP* and *attB* from the junction sites *attL* and *attR* requires, in addition to integrase and IHF, a second phage-coded protein, Xis, which binds to a specific sequence in the P arm.

Thus, the specificity of a phage's integration system is determined through interactions between specific protein domains and their cognate binding sites in *attP* and/or *attB* and through base pairing of reaction intermediates in the overlap segment. These are the elements that must undergo coordinated change to evolve a new specificity. For example, an alteration in site preference from one *attB* to another requires that the new *attB* have some symmetry suitable for B and B′ sites, that the phage core become similar to the new *attB* (identical in the O segment), and that the C-terminal integrase domain change to acquire affinity for the new B and B′. Definition of these specificities in terms of base or amino acid change is prerequisite to understanding how the changes may have taken place. Looking just at *attB* sites, a correlation between recognition sequences and integrase phylogeny can be

* Corresponding author. Phone: (650) 723-1170. Fax: (650) 723-6132.

† Present address: Cornell University Medical College, New York, NY 10021.

‡ Present address: University of North Carolina School of Medicine, Chapel Hill, NC 27599.

§ Present address: 5242 Veronica St., Los Angeles, CA 90008.

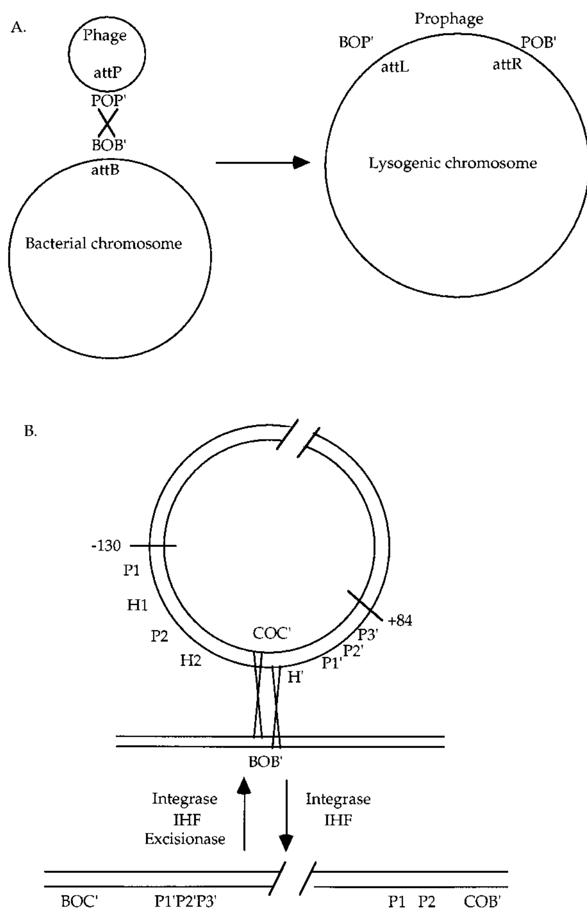


FIG. 1. Mechanism of λ integration. (A) Overall process of insertion during lysogenization. Circular phage and host DNAs (each represented by a single line) cross over at a specific site on each to give a lysogenic chromosome with an inserted prophage. (B) Detail of reaction at the crossover point. About 240 bp of specific sequence are needed on the supercoiled *attP* partner. The *attB* partner requires 21 bp of specific sequence and can be linear. Crossover occurs by exchange of the upper two strands, followed by migration through the 7-bp overlap region and resolution by exchange of the lower two strands. *attP*, phage attachment site; *attB*, bacterial attachment site; *attL* and *attR*, sites at the left and right boundaries, respectively, of the inserted prophage; P, left arm of *attP*; O, overlap region; P′, right arm of *attP*; C, C′, B, and B′, flanking sequences of *attP* core and *attB* core, respectively (core recognition sites); H1, H2, and H′, binding sites for host protein IHF.

discerned; but in general, integrase specificities are either the same (e.g., λ and 434) or so different that evolutionary pathways are difficult to infer (6).

For at least two phage pairs, recognition specificities are similar but not identical. The best studied of these pairs consists of λ and HK022. These phages insert at different *attB* sites whose B and B′ sequences are similar but distinct, while their arm recognition and excisionase specificities seem to be identical. From a study of λ-HK022 hybrid integrase proteins, followed by directed mutagenesis, Weisberg and collaborators (8, 20) switched the recognition specificity by five base substitutions, two of which broaden the range of substrates acceptable by the integrase and three others that narrow it down to the new specificity.

Another phage pair of interest is phage 21 and the excisable 14-kb element e14 that is present in *Escherichia coli* K-12. Excision of e14 is induced by the SOS system (2), indicating a repressed integrase-excisionase system within the element (which presumably originated through deletion of large segments from the prophage form of a complete lambdoid phage). The e14 sequence, now available as part of the complete *E. coli* genome (1), includes homologs to 21 *int* and *xis*, slightly overlapping as in 21, located about 3 kb from the left end and oriented leftward, as well as a homolog of the repressor gene (*c*2) of phage P22 located 1.2 kb to the right of *xis*. Thus, the order and orientation of these genes are the same as in 21, although in 21 and most other lambdoid phages there are no genes between *int* and *attP*. No mutations have been made in e14 *int* and *xis*, nor is there any information about the promoter(s) they use. Schneider and colleagues (7, 17) found that phage 21 inserts at the same site as e14 and that phage 21 Int and Xis can efficiently cure cells of a 21 prophage, but not of e14, even when 21 and e14 are inserted next to each other in the chromosome. Taken together with the ability of e14 to cure itself on SOS induction, this finding suggested that e14 and 21, although using the same *attB*, have different specificities of site recognition.

Candidate arm and core site consensus sequences have been proposed (6, 17). The suggested core consensus was based on local dyad sequence symmetry around the crossover point (which was localized to 10 bp of sequence identity between *attP* and *attB* [17]). In analogy to other lambdoid phages, a 7-bp overlap region was suggested but has not been directly tested. Choice of candidate arm sequences was based on the facts that (i) oligonucleotide sequences approximating a common consensus occur in both arms of *attP* and *attE*, at identical locations in both, and (ii) as in λ, there is a cluster of three such sites in P′. For 21, the sequence YYTTCCCCAAAA occurs in direct orientation at coordinates −134 to −123, +224 to +235, +236 to +247, and +249 to +260, assigning coordinate 0 to the center of the proposed overlap. For e14, the consensus is KTTATACAAAA (where Y = C or T and K = G or T). By analogy to λ, we call the site at −134 to −123 P1 and the cluster of three sites at the right P1′, P2′, and P3′. We do not know whether 21 has a site equivalent to the P2 site of λ (Fig. 1).

We show here, by a combination of in vivo and in vitro results, that 21 and e14 indeed have different recognition specificities, that 21 has a strong preference both for its own arm sequences and for its own core sequences, and that the proposed site assignments are probably correct.

We have tried to use a nomenclature that is sufficiently precise but not too cumbersome. We use *attP* to refer to the 21 *att* site, including both core and arms. It can also be written as POP′ (Fig. 1). Following Brody and Hill (3), we call the analogous site of e14 *attE* or EOE′. We use *attB* to denote the bacterial insertion site for phage 21. Because e14 is endogenous to the K-12 host, the *attB* sequence is BOE′. When *E. coli* K-12 is cured of e14, a BOB′ site is generated. Because BOE′ and BOB′ differ at only one position near the crossover point and behave equivalently in our experiments, we treat BOB′ as a variant *attB*. The core regions of *attP* and *attE* are denoted CpOCp′ and CeOCe′, respectively.

MATERIALS AND METHODS

Preparation of digoxin-labeled substrates (*attB*, *attL*, and *attR*). The 0.5-kb *attB* fragment used for in vitro assay was cut from the multicloning site of pBS::*attB* with *Hin*dIII and *Eco*RI then end labeled by filling the ends with DNA polymerase (Klenow fragment), using digoxin-labeled dUTP (Boehringer Mannheim). The source of the *attB* insert was PCR amplification from the BOE′ plasmid p21*attB*7-35 (17) of DNA from −264 to +277. Primers were GGCAAAGAGATCTTCATTAAAGACG and TGATAACCGGTAGCAGCCTAACAAAAA. The PCR product was cut with *Bgl*II and *Age*I (sites underlined) and then cloned into pBS cut with *Bam*HI and *Xma*I.

	attP substrate	Product
1	P1 C P1′P2′P3′	+
2		-
3	e	-
4	e	+
5	e	+
6	e e	-
7		-
8		-
9	e	-
	-100 0 +100 +200	

FIG. 2. In vitro integrase reaction. All reactions were done with crude extracts and repeated at least four times on different days. Reaction tubes contained 20 μg of crude extract in 20 μl. Substrates used for constructs 1 and 2 were also checked with purified Int (0.75 μg in 20 μl). Constructs: 1, complete substrate with arm sequences P1, C, P1′, P2′, and P3′; 2, deletion of P1; 3, replacement of P1 by e14 consensus arm site; 4, replacement of P3′ by e14 consensus; 5, replacement of P2′ by e14 consensus; 6, replacement of both P2′ and P3′ by e14 consensus; 7, deletion of P3′; 8, deletion of P2′ and P3′; 9, replacement of core by e14 consensus. A coordinate scale is given at the bottom (where 0 is the crossover point). Presence (+) or absence (−) of reaction product when these *attP* substrates were reacted with end-labeled *attB* DNA (500-bp fragment containing BOE′) is shown at the right.

To test the sufficiency of the 25-bp *attB* sequence centered on the crossover point, pBS with the BOB′ 25-mer cloned into the *Hin*dIII site (see Fig. 6, line 2) was cut with *Pvu*II, and the 312-bp blunt-ended fragment was gel purified. This was cleaved with *Eco*RI to give a 267-bp fragment that was digoxin labeled with Klenow enzyme.

Digoxin-labeled linear *attR* and *attL* substrates for excision assays were prepared as follows. For *attR*, the 1-kb *Sma*I-*Sma*I insert from Schneider's plasmid p21*attR*I-44-4 (17) was gel purified and then cut with *Acc*I, giving a 676-bp fragment that was then digoxin labeled by Klenow enzyme. For *attL*, Schneider's plasmid p21*attL*6-1, containing a 1.7-kb *Acc*I-*Sph*I fragment, was first cut with *Acc*I and *Eco*RV to give a 1,233-bp fragment. This fragment was gel purified and then cut with *Ava*II to give a 1,078-bp *Ava*II-*Eco*RV *attL* fragment, which was digoxin labeled at the *Ava*II end.

Construction of *attP* substrate. Plasmids p21*attP*6-1 and p21*attP*1-2-4 contained a 2.9-kb *Bgl*II fragment and a 1.1-kb *Sph*I fragment, respectively, from phage 21 cloned into pBS^+ (15). Other *attP* constructs were made by PCR with primers P-1 (TAAAGAATTCATTATCGTAGATATCAAAAC), P-2 (GAGGAATTCTTTCCCCAAAATAAAAAC), P-6 (CCAAAATGAATTCGAACGTCAA), P-7 (GAGGAATTCTTTTATCAAAATAAAAACGAACG), P-8 (ATAGAATTCAGCCAGTTTTGATAAAG), P-9 (TGAATTCCAGTTTTGGGGAAGGGTTTTGATAAA), P-10 (GGGGAATTCTTTTGGGGAAAGTTTTGG), P-11 (TTTGAATTCAGTTTTGGGGAAGATTTTACATCA), P-12 (TGAATCCAGTTTTGATAAAGGGTTTTGATAAAA), P-12 (TGAATTCCAGTTTTGATAAAGGGTTTTGATAAAA), PEC-1 (TTTTTAACTAATGCTGCCCACATGGGCTG), and PEC-2 (CAGCCCATGTGGCGCAGCATTAGTTAAAAA).

For constructs 1 to 8 (Fig. 2), the appropriate primer combination was used for PCR with a p21*attP*6-1 template. PCR products were cut with *Eco*RI (sites underlined) and cloned into the *Eco*RI site of pBS+ (construct 1, P-1 × P-2; construct 2, P-1 × P-6; construct 3, P-1 × P-7; construct 4, P-8 × P-2; construct 5, P-9 × P-2; construct 6, P-12 × P-2; construct 7, P-10 × P-2; construct 8, P-11 × P-2). For construct 9, two separate PCRs were run: P-1 × PEC-1 and P-2 × PEC-2. Gel-purified PCR products were heated to 94°C for 30 s to separate the strands, mixed for 10 min at 72°C with added Vent polymerase so that priming from overlapping sites could generate full-length product, and then amplified by PCR using primers P-1 and P-2. The final product was cut with *Eco*RI and cloned into pBS+.

Cloning of p21 *int* and *xis* into pET-11 vectors. The *int* and *xis* genes were PCR amplified from the *int xis* expression vector pUC21*intxis*4 (17). For cloning into pET-11d, the upstream primer was TCGCCCATGGCTGCTAGACCCCGATCT (*Nco*I site underlined); for pET-11c, it was TCGCATATGGCTGCTAGACCCCGATCT (*Nde*I site underlined). The downstream primer was CGGGATCCAACGCCAAGACC-GTAACC (*Bam*HI site underlined). When cut with the appropriate enzymes, these PCR fragments extend from the 5′ end of *int* to 146 bp downstream of the 3′ end (which is between *int* and the crossover point). When the pET-11 vectors are cut with the same enzymes and ligated to the enzyme-trimmed PCR fragments, the initiating ATG of *int* is placed under control of the T7 promoter and ribosome binding site and upstream of the T7 termination site (16). DNA was used to transform *E. coli* XL1-blue (Stratagene), selecting for ampicillin resistance, and plasmids were checked by restriction cutting (15).

To clone *xis* into pET-11c, the primers were GAGGAGGACATATGTCTCGACTAATC and TTGGGATCCCGGGGTCTAGCAGCCATCAG, which places the *Bam*HI site 9 bp downstream from the 3′ end of *xis.*

The crude extracts of Int used in this study were prepared from pET-11d. A pET-11c clone was used for recloning into the His tag vector. Several isolates of each type were checked, with no obvious difference in activity. The inserts were not resequenced after PCR.

Cloning of *int* into the His tag vector pET-15b. From a pET-11c clone of *int* which had been verified to produce active integrase, the insert was cut out with *Nde*I and *Bam*HI, gel purified, and cloned into pET-15b.

Preparation of enzyme extracts. Plasmid preparations made from XL1-blue transformants were used to transform the T7 expression strain BL21(DES)pLysS (Novagen). To obtain active integrase or excisionase preparations without massive sequestration into inclusion bodies, the following protocol was developed. Transformants were grown overnight in LB broth, diluted 1:100 into fresh broth, and grown to an A_{600} of 0.6. Isopropylthiogalactopyranoside (IPTG) was added to 1 mM, and growth was continued for 3.5 h at 37°C (A_{600} of ~1.25). Cells were spun for 15 min at 8,000 rpm at 4°C, washed twice with ice-cold 10 mM $MgSO_4$, resuspended in ice-cold buffer A (21) (3 mM EDTA, 0.3 M KH_2PO_4 [pH 7.6], 0.5 M KCl), and sonicated. Typical extracts contained 20 to 30 g of protein per liter.

These crude extracts from pET-11d were used for enzyme assay. Extracts of His-tagged integrase were purified on a Ni^{2+} resin as described in the protocol for the Novagen His Bind buffer kit except that DNase was not added. The bound protein was eluted from the column in 15 ml of elute buffer, which was collected in 1-ml samples and run on a protein gel. Positive samples were probed and concentrated with a Centricon 10 model 4241 concentrator (Amicon, Inc., Beverly, Mass.) to a volume of 1 ml. Two successive dilutions to 2 ml with buffer A and reconcentrations to 1 ml replaced most of the elution buffer with buffer A. All operations were performed at 4°C. Typical extracts contained 0.75 g of integrase protein per liter.

The His tag could be removed by thrombin cleavage as described in the protocol for the Novagen His Bind buffer kit (Fig. 3, lane 3). However, the uncleaved His-tagged enzyme was used in the experiments reported here.

Extracts were stored at −70°C.

Integrase and excisionase assay. Integrase reactions were run for 90 min at room temperature in 20-μl volumes, using conditions based on those described in reference 21. The assay tube contained 1 g of bovine serum albumin (New England Biolabs) per liter, 5 mM spermidine, 50 mM Tris hydrochloride (pH 7.8), 5 mM NaCl, 180 mM KCl, 5 mM EDTA, and 1% (vol/vol) glycerol, plus DNA substrates and proteins as indicated. The reaction was terminated by dilution to 100 μl in Tris-EDTA (pH 8.0) and addition of 100 μl of phenol with 0.1% 8-hydroxyquinoline equilibrated with 1 M Tris (pH 8.0). Following phenol extraction, ethanol precipitation, and electrophoresis on 0.8% agarose in TAE buffer (15), the digoxin label was developed by using a Genius 1 Dig DNA labeling and detection kit (Boehringer Mannheim).

Cointegration assays. The *attB* plasmids for cointegration assays were prepared by annealing synthetic oligonucleotides with overhanging 5′ AGCT ends and cloning into the *Hin*dIII site of pBS+. Control plasmids p21*attB*7-35 and p21*attB*7-70 (17) contained a 2.0-kb *Acc*I BOE′ fragment cloned into the *Acc*I site of pBS. Strain W3350 (4) was transformed with these plasmids. An overnight culture in tryptone broth plus 100 mg of ampicillin per liter was grown at 37°C, diluted 1:1,000, and grown to a Klett reading of 30 (green filter), which is about 10^8 cells/ml. The culture was centrifuged and concentrated sevenfold in 10 mM $MgSO_4$. Then 0.1 ml of the cell suspension was mixed with 10^6 PFU of phage 21, reincubated for 15 min at 37°C, and then plated with soft agar onto tryptone plates. After 5 to 7 h at 37°C (when plaques were barely visible), plates were eluted with 5 ml of 10 mM $MgSO_4$ for 2 h at 4°C. The eluate was heated at 47°C for 30 min over $CHCl_3$ to sterilize and assayed for plaque titer on W3350.

For transduction, an overnight culture of W3350 was diluted 1:100 in tryptone broth, grown to a Klett reading of 50, and resuspended at 25× in 10 mM $MgSO_4$. Then 0.1 ml of cells was added to 0.1 ml of phage lysate at various dilutions and incubated at 37°C for 40 min; 1 ml of L broth containing 1.25 mM sodium pyrophosphate (to prevent reinfection) was added for 2 h to allow expression of ampicillin resistance. Transductants were enumerated on L agar with 100 mg of ampicillin per liter, 75 mg of methicillin per liter, and 1.25 mM sodium pyrophosphate. Transduction efficiencies were calculated as ampicillin-resistant colonies per PFU.

Some transductions (not shown) were performed with W3350(21) as the recipient, to avoid loss of potential transductants through lysis; however, the efficiency was not significantly different from that on W3350 (data not shown).

Curing by e14. To demonstrate specific curing by e14, we needed a lysogen where (i) e14 and prophage 21 are present and easily scorable, (ii) the e14 excision system can be induced without inducing 21, and (iii) the results are not biased by selection against the *sfiC* gene present in e14. To achieve these goals, we constructed HW5, with genotype *recA442* (Tif) *sfiA11 sfiB103 sfiB* thr pro his arg ilv*(Ts) *gal str* (21*xis-12*::*kan*Δ*cI-B*)(e14-1272::Tn*10*). The chromosomal background of HW5 was derived from GC3217 (9), the *sfiB103 sfiB** mutation came from GC4780 (14), the *xis-12*::*kan* mutation came from 21*xis*12 (17), and the e14-1272::Tn*10* was from CH1494 (2). Details of strain construction and of the construction of plasmids and mutants used as source materials in this work are available on request.

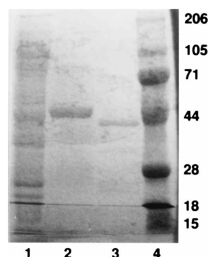


FIG. 3. Purification of His-tagged Int protein. Lane 1, crude extract from pET-15b clone of 21 *int*; lane 2, after Ni^{+} resin (predicted molecular mass, 45.7 kDa); lane 3, after thrombin cleavage (predicted molecular mass, 44.1 kDa); lane 4, molecular mass standards (high range); Gibco BRL). Shown is a 10% acrylamide gel (Tris-glycine-sodium dodecyl sulfate buffer [13]) stained with Coomassie blue. Sizes are indicated in kilodaltons.

Nucleotide sequence accession number. The sequence of a 2.9-kb *Bgl*II fragment including *attP*, *int*, and *xis* (17) was entered in GenBank under accession no. M61865. The *attB* sequence is available as part of the *E. coli* genome (1).

RESULTS

In vitro integrase assays. Following many failed attempts to demonstrate in vitro recombination with extracts made either from infected cells or from cells expressing *int* from a *lac* promoter on a multicopy plasmid, we cloned the 21 *int* gene behind a T7 promoter and ribosome binding site, using the pET-11d vector. Purified protein was made by using the His tag vector pET-15b and eluting the protein from a nickel column (Fig. 3). As expected from the in vivo effects of *himA* mutations (17), the purified protein required IHF for activity (Fig. 4). For most other experiments, we used a crude extract with protein concentration of about 20 g/liter. We do not have a precise comparison of the specific activities of the His-tagged enzyme and the native enzyme (available only in crude extracts). Judging from the intensity of the protein band in the crude extract, they appear to be similar.

To determine the sequence requirements of *attP*, the proposed arm site sequences were YYTTCCCCAAAA (phage 21) and KTTATACAAAA (e14) (where Y = C or T and K = G or T). In our replacements, we used TTTTATACAAAA as an e14 arm sequence.

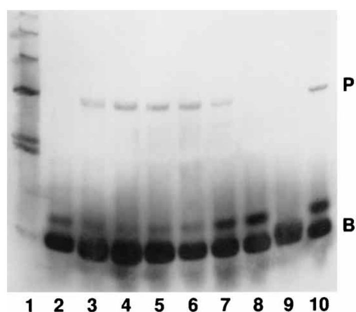


FIG. 4. Requirement for purified His-tagged integrase and IHF on integration reaction and inhibition by crude excisionase. Assays were done as described in Materials and Methods. Lane 1, digoxin-labeled standards (end labeled with Klenow enzyme, using a λ *Hin*dIII digest from New England Biolabs). Lanes 2 to 10 all contain 0.16 pmol of digoxin-labeled 0.5-kb *attB* fragment (B) and 0.30 pmol of 3.7-kb *attP* plasmid DNA (P-1 × P-2, cloned into pBS [Fig. 2, construct 1]). Lane 2, 6.7 pmol of Int; lane 3, 6.7 pmol of Int and 0.47 pmol of IHF; lane 4, 6.7 pmol of Int and 0.94 pmol of IHF; lane 5, 6.7 pmol of Int and 1.9 pmol of IHF; lane 6, 3.3 pmol of Int and 1.9 pmol of IHF; lane 7, 1.7 pmol of Int and 1.9 pmol of IHF; lane 8, 1.9 pmol of IHF; lane 9, 6.7 pmol of Int, 1.9 pmol IHF, and 20 μg of crude Xis extract; lane 10, 6.7 pmol of Int and 3.8 pmol of IHF. The linear 4.2-kb reaction product (P) is generated by recombination between the 0.5-kb linear *attB* fragment and the 3.7-kb supercoiled *attP* plasmid.

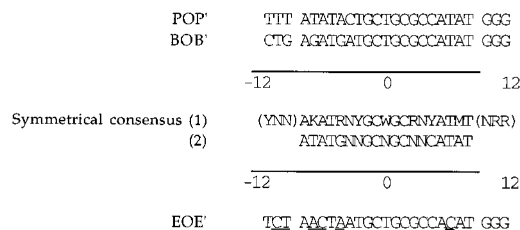


FIG. 5. Sequences of *attP* core, *attB* (2), and *attE* core (17). Consensus sequence 1 is given for 25 bases, but the outermost three on each site are in parentheses, because a 19-bp core works fairly well (Fig. 6). Y = C or T; K = G or T; M = A or C; R = A or G; W = A or T. Specified bases are those seen at all four positions (+*n* and the complementary strand at −*n*, for both POP′ and BOB′), so that the consensus is a palindrome. Consensus sequence 2 agrees with consensus sequence 1 but specifies only those bases found at three of four of the relevant positions, with no ambiguity. Those bases in EOE′ that differ from BOB′ are underlined. The proposed 7-bp overlap region is in boldface.

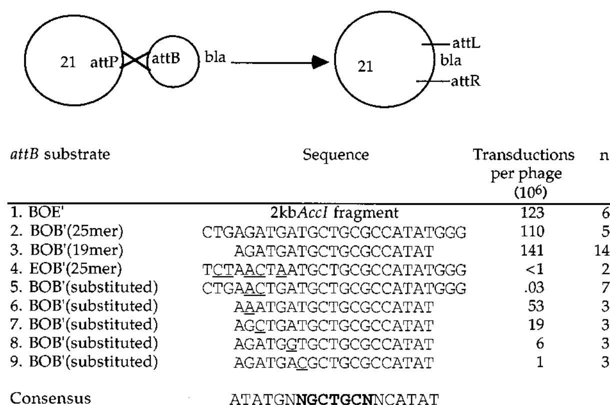
The results (Fig. 2) allow two conclusions. (i) The functional limits of *attP* are between −138 and −117 on the left and +248 and +279 on the right, i.e., enough sequence to include the proposed arm sequences (−135 to +261) and not much beyond. (ii) e14 sequences are accepted by the 21 enzyme in some contexts but not in others. For example, the *attP* sequence on the left is not replaceable by the e14 consensus, but P2′ or P3′ on the right can be replaced; however, replacement of both P2′ and P3′ can lead to inactivity. Also, replacement of the core sequence with the e14 core destroys activity.

In the experiments represented in Fig. 2, we used a 0.5-kb BOE′ fragment derived from the *E. coli* chromosome (see Materials and Methods). A 267-bp fragment that includes 25 bp of BOB′ cloned into pBS also worked (data not shown).

attB **specificity.** e14 and 21 use the same *attB* site. However, Schneider (17) observed that in vivo, 21 always inserts to the left, rather than to the right, of e14, as though *attL* and *attR* of e14 did not serve as equivalent *attB* sites for phage 21. In hosts without e14, 21 inserts well at the bacterial site. Denoting the *attE* core by CeOCe′, this implies that BOCe′ and BOB′ are good *attB* sites for 21, but CeOB′ is not. Sequence comparison of CeOCe′, BOB′, and CpOCp′ shows that Ce′, B′, and Cp′ are almost identical but that Ce differs from the others (and from the dyad-symmetrical proposed consensus sequence) at two adjacent bases, −8 and −7 (Fig. 5).

To study the influence of these substitutions on *attB* function, we used a cointegration assay. Synthetic oligonucleotides with the CeOB′ core sequence or with various replacements of BOB′ bases with those from CeOB′ were cloned into pBS, cells harboring these plasmids were infected by phage 21, and we looked for the ability of the output phage to render tester cells resistant to ampicillin. (This assay may require two site-specific recombinations, one to generate the cointegrate in the donor strain and one to release the plasmid from it in the recipient.) The results (Fig. 6, lines 3 and 5) verify that the changes in Ce (especially the simultaneous substitution of bases at −7 and −8) explain the inability of CeOB′ to act as an effective *attB*.

We assume that what matters here is deviation at two adjacent positions from the consensus, not just from BOB′. Thus, *attP* and *attB* also differ at two adjacent positions (−5 and −4), but both follow symmetrical consensus 1, and *attP* deviates at only −5 from consensus 2 (Fig. 5). Although *attP* and *attB* were used to construct the consensus, we would nonetheless con-



attB substrate	Sequence	Transductions per phage (10⁶)	n
1. BOE′	2kb*AccI* fragment	123	6
2. BOB′(25mer)	CTGAGATGATGCTGCGCCATATGGG	110	5
3. BOB′(19mer)	AGATGATGCTGCGCCATAT	141	14
4. EOB′(25mer)	TCTAACTAATGCTGCGCCATATGGG	<1	2
5. BOB′(substituted)	CTGAACTGATGCTGCGCCATATGGG	.03	7
6. BOB′(substituted)	AAATGATGCTGCGCCATAT	53	3
7. BOB′(substituted)	AGCTGATGCTGCGCCATAT	19	3
8. BOB′(substituted)	AGATGGTGCTGCGCCATAT	6	3
9. BOB′(substituted)	AGATGACGCTGCGCCATAT	1	3
Consensus	ATATGN**NGCTGCN**NCATAT		

FIG. 6. Cointegration assay for *attB* function. The *attB* sequences shown were cloned into pBS. Cells carrying these plasmids were infected with phage 21, and the phage produced were used to infect an ampicillin-sensitive recipient. Ampicillin-resistant transductants per output phage were scored. In the consensus sequences, the 7-bp overlap segment is in boldface. n, number of determinations (each with a different lysate).

sider the result to corroborate the importance of the specified bases.

Several other conclusions may be drawn from Fig. 6: (i) 19 bp around the center of symmetry (line 3) work about as well as the whole 2 kb of BOE′ DNA from the bacteria; (ii) the individual substitutions at −7 and −8 (lines 6 and 7) have fairly small effects; and (iii) changing an unspecified base just outside the proposed 7-bp overlap region (line 8) has less effect than changing a base within the overlap region (line 9). This is expected because the latter change creates a heterology within the proposed overlap region, which is known to depress λ integration (19). The substitution in line 8 was not expected to depress recombination at all (especially because it is symmetrical with the normal base at the right), but it appears to have some effect.

Specificity of e14. The above-described findings indicate that 21 integrase does not accept e14 sequences as equivalent to 21 sequences. They do not formally prove that e14 has a different specificity from 21, rather than just comprising a much poorer substrate (perhaps for its own integrase as well). Having thus far failed to demonstrate e14 integrase activity in vitro, we cannot do the directly reciprocal experiments. However, in vivo results strongly indicate that e14 is specific for its own *attP* counterpart, *attE*.

The basic protocol is to construct a strain that harbors e14 and 21 inserted next to each other on the chromosome and to then induce e14 excision by the SOS system and see whether one or both plasmids are excised.

As with λ and e14, 21 integrase and excisionase are themselves induced by SOS (as judged by the production of high-titer lysates following UV induction of a lysogen). Therefore, SOS induction of a strain with 21 inserted next to e14 should excise both elements. So a strain with a Tn*10* insertion in e14 was lysogenized by a 21 derivative marked by an *xis::kan* insertion, and an internal deletion that removed genes *c*I-*B* was isolated (Fig. 7). To study the effect of SOS induction on prophage excision, the construct was transduced into a *recA*442 *sfiB103 sfiB** strain, selecting for resistance to kanamycin and tetracycline. The *sfiB103 sfiB** double mutation renders the cell insensitive to growth inhibition by the e14-encoded SfiC (14).

Cells of this strain were grown 5 h at 39 to 42°C, returned to 37°C for 3 h of additional growth and segregation of excised

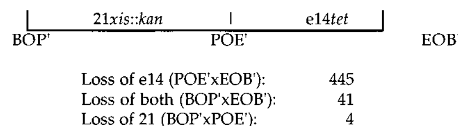


FIG. 7. Losses of e14 and 21 from a tandem double lysogen in which e14 gene expression was induced by heating a *tif sfiB** host.

elements, and then plated at 30°C. Individual colonies were plated and scored for 21 (kanamycin resistant) and e14 (tetracycline resistant). Among those colonies that were not resistant to both antibiotics, a total (in all experiments) of 445 had lost only e14, 41 had lost both e14 and 21, and 4 had lost only 21. Because reconstruction experiments (data not shown) showed that the double excisants and the parental culture survived the treatment equally well (thanks to the *sfiB** background), we conclude that e14 strongly prefers its own sites to those of 21. Some of the rare excisants from 21 sites might result from a low level of expression of the 21 *int* gene (which is present but not known to be induced).

As a control, strain HW5 was transformed with a plasmid, p21*intxis*4 (17), which expresses 21 *int* and *xis* from *plac*. After IPTG induction, about 80% of the cells had lost kanamycin resistance but retained tetracycline resistance. This finding verifies that the 21 sites in HW5 were intact and that the system responded to the 21 enzymes like a normal lysogen with 21 adjacent to e14 (17).

21 excisionase. Schneider (17) showed that 21 has an excisionase gene, located (like λ *xis*) upstream of and overlapping *int* and needed for excision but not for insertion. Crude extracts of 21 Xis are required in vitro for excision (Fig. 8) and, as in λ, Xis inhibits the insertion reaction. The excision reaction was clearly demonstrable with labeled linear *attL* and supercoiled *attR* (Fig. 8). The reaction between labeled *attR* and supercoiled *attB* was much weaker (invisible in Fig. 8 but detectable in other runs). The inhibition of the integrase reaction by crude extracts of Xis (Fig. 4, lane 9) was not seen with a control crude extract of *E. coli*[pET-11d] (data not shown).

DISCUSSION

Size and structure of *attP*. Table 1 shows that as in λ, 21 *attP* function requires the presence of the putative arm sites and little if any DNA beyond them. Furthermore, the inactivity of

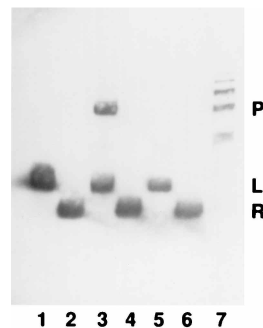


FIG. 8. Excisionase assay, performed as described in Materials and Methods. Lanes 1, 3, and 5, 73 fmol of digoxin-labeled *attL* (1.1 kb; L) plus 960 fmol of supercoiled *attR* (4.2 kb); lanes 2, 4, and 6, 115 fmol of digoxin-labeled *attR* (0.7 kb; R) plus 806 fmol of supercoiled *attL* (4.9 kb). Crude enzyme extracts (20 μg of each) were added as follows: lanes 1 and 2, Xis only; lanes 3 and 4, Xis plus Int; lanes 5 and 6, Int only. Linear product (P) in lane 3 has an expected length of 5.3 kb.

constructs where these sequences are replaced by e14 counterparts supports Schneider's (17) proposal that they are in fact arm sites. One noteworthy difference between 21 *attP* and λ *attP* is that the P1′-P2′-P3′ cluster in λ is at +54 to +84, whereas in 21 it is +225 to +261.

The significance of this fact can be appreciated by considering the group of phages λ, 434, and HK022, all of which have the same arm specificities. Despite many single-base substitutions between the crossover point and the arm sites, the spacings are strictly conserved. Likewise, 21 and e14 show identical spacing between core and arm sites, although all four arm sites have diverged coordinately.

Conservation of spacing is expected from current models for intasome formation (13). If integrase molecules bind to DNA to form intramolecular bridges, then a change in spacing of even one nucleotide should cause some strain, given the rigidity of the supercoiled DNA substrate. But if spacing is generally conserved, it must have diverged radically at some point to generate completely new types like 21 and λ, starting from their common ancestor.

We suggest that the selective forces acting to conserve spacing are sufficiently strong to prevent random genetic drift during evolution. Radically new types survive only because they experience new selective pressures. The most obvious pressure results from the fact that 21 and e14 insert into the *icd* gene of *E. coli* and carry within the phage a duplication of the distal 165 bp of the gene and an additional less conserved 44 bp through the transcriptional terminator (17). This duplication allows lysogens to exhibit *icd* function equivalent to that of nonlysogens (10). Because bacterial and phage sequences are so similar within these 209 bp, it is impractical for the phage to recognize a cluster of arm sequences (P1′, P2′, and P3′) within the duplicated segment. Even if a potential arm site cluster were to exist in the duplicated segment, presence of the cluster in both *attP* and *attB* would disallow discrimination between excision and insertion, because *attB* and *attP* would be equivalent to *attL* and *attR*, respectively. As phage 21, like λ, has an *xis* gene, we presume that such discrimination is advantageous. The obvious solution is to place the P1′-P2′-P3′ cluster beyond the *icd* duplication, which is where we find it.

Other phages fit the same pattern. *Haemophilus* phage HP1 has a 182-bp duplication of the 3′ end of the tRNA operon into which it inserts, and the minimal 418-bp *attP* site extends beyond that duplication (10, 11). Phage P22 has a shorter duplication (46 bp, plus an additional 17 bp of partial homology, which includes an IHF binding site). As the IHF binding site is further to the right than the λ H′ site, the conformation of the intasome cannot be the same as λ's in any case; in fact, a cluster of two arm sites (P1′ and P2′) is found at +90 to +108 (18).

Thus the spacing between core and arm sites can change drastically under rationalizable selective pressure. Because the three-dimensional structure of the intasome is as yet unknown, it is not clear that any profound alteration in intasome structure is needed; perhaps the macromolecular contacts are about the same, but larger loops of uninvolved DNA project from 21's intasome than from λ's.

Core sequences. Integration of λ proceeds by an initiating exchange at a specific nucleotide position, followed by a resolving exchange 7 bp to the right (Fig. 1). Direct evidence for such a mechanism in phage 21 is unavailable. Because the *icd* duplication, though similar, is not identical between phage and host, the crossover point(s) can be localized within 10 bp by sequencing junction fragments from lysogens (17). Assuming a mechanism similar to λ's, Campbell (5) proposed core recognition sequences with the consensus (−10) ATATGNNGC NCGNNCATAT (+10). Figure 6 shows that simultaneous alteration of bases at −7 and −8 greatly depressed the recombination frequency, implying that these bases are important recognition elements. As the altered sequence is present in the *attE* core, the e14 enzyme apparently accepts the alteration.

Does e14 actually prefer this sequence to that of the 21 core? Whereas we have not confronted the system with all possible combinations of core sites, that possibility seems unlikely (and certainly not required by the data). There is no symmetrical consensus sequence that fits e14 better than the 21 consensus. We suggest instead that e14 (which, remember, uses the same *attB* site as 21 does) is simply less stringent than 21 in its core sequence requirements. From their work on λ and HK022, Dorgai et al. (8) suggested that evolutionary transitions from one integrase specificity to another may proceed through mutations that loosen specificity followed by mutations that tighten it in a different direction. In our view, e14 integrase would correspond to their postulated less specific evolutionary intermediate.

Arm sequences. Our in vitro assays on *attP* substrates with altered arm sites clearly indicate that phage 21 prefers its own arm sequences to those of e14. Does e14 likewise prefer its own arm sequences to those of 21? Our in vivo excision experiments suggest that it does. The scoring of excisants from the strain diagrammed in Fig. 7 shows the substrate preferences (EOB′ × POE′) >> (EOB′ × BOP′) >> (POB′ × POE′). In principle, this could result from differences in core sites, arm sites, or excisionase binding sites. The cores seem unlikely to be responsible. To the right of the crossover point, B′ and Cp′ are identical, and Ce′ differs by only one nucleotide. To the left, a preference of e14 integrase for Ce over B could explain the second inequality but not the first. To explain the first inequality on the basis of core sites, one would have to postulate a 10-fold preference for Cp over B. As to the Xis binding site(s), if, as in λ, they are to the left of the crossover point and required only in one of the two partners, then again the first inequality is unexplained. If they happened instead to lie to the right of the crossover point, the first inequality, but not the second, could be explained. Certainly, the specificity could depend on some more complex interactions, involving perhaps other unidentified sites outside of the core, but the most straightforward interpretation is that 21 and e14 integrases each evolved to prefer their cognate recognition sites. Minimally, we can say that in the experiment represented in Fig. 7, the e14 system shows a site preference very different from that seen when the 21 system acts on the same substrate. Our experimental design does not allow us to exclude a preferential *cis* action of integrase (which is encoded near the POE′ site). Qualitatively, the e14 system is active in *trans* (3).

Has this alteration in specificity come about through random drift, or is selection behind it? As suggested elsewhere (6), the best candidate for a selective force is frequency-dependent selection for new types that are resistant to heteroimmune curing by previously existing phages. The explanatory value of this hypothesis is equally strong whether or not the new system causes integration at the same site as the ancestral phage.

ACKNOWLEDGMENTS

We thank Stephanie Schneider for biological materials, Howard Nash for purified IHF, and Michael Caine and Len Penacchio for participation in experiments related to those shown in Fig. 6.

This work was supported by grant GM51117 from the National Institute of General Medical Sciences.

REFERENCES

1. **Blattner, F. R., G. Plunkett III, G. F. Mayhew, N. T. Perna, and F. D. Glasner.** 1997. *Escherichia coli* from bases 1195814 to 1209543 (section 104

of 400) of the complete genome. Accession no. AE000214 U00096.
2. **Brody, H., A. Greener, and C. W. Hill.** 1985. Excision and reintegration of the *Escherichia coli* K-12 chromosomal element e14. J. Bacteriol. **161:**1112–1117.
3. **Brody, H., and C. W. Hill.** 1988. Attachment site of the genetic element e14. J. Bacteriol. **170:**2040–2044.
4. **Campbell, A.** 1961. Sensitive mutants of bacteriophage λ. Virology **14:**22–32.
5. **Campbell, A.** 1992. Chromosomal insertion sites for phages and plasmids. J. Bacteriol. **174:**7495–7499.
6. **Campbell, A.** 1994. Comparative molecular biology of lambdoid phages. Annu. Rev. Microbiol. **48:**193–222.
7. **Campbell, A., S. J. Schneider, and B. Song.** 1992. Lambdoid phages as elements of bacterial genomes. Genetica **86:**259–267.
8. **Dorgai, L., E. Yagil, and R. A. Weisberg.** 1995. Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases. J. Mol. Biol. **252:**18–188.
9. **George, J., M. Castellazzi, and G. Buttin.** 1975. Prophage induction and cell division in *E. coli*. III. Mutations of *sfiA* and *sfiB* restore division in *tif* and *lon* strains and permit the expression of mutator properties of *tif*. Mol. Genet. **140:**309–332.
10. **Hakimi, J. M., and J. J. Scocca.** 1994. Binding sites for bacteriophage HP1 integrase on its DNA substrates. J. Biol. Chem. **269:**21340–21345.
11. **Hauser, M. A., and J. J. Scocca.** 1992. Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1: location of the boundaries of the phage attachment site. J. Bacteriol. **174:**6674–6677.
12. **Hill, C. W., J. A. Gray, and H. Brody.** 1989. Use of the isocitrate dehydrogenase structural gene for attachment of e14 in *Escherichia coli* K-12. J. Bacteriol. **171:**4083–4084.
13. **Landy, A.** 1993. Mechanistic and structural complexity in the site-specific recombination pathways of Int and Flp. Curr. Opin. Genet. Dev. **3:**699–707.
14. **Maguin, E., H. Brody, C. W. Hill, and R. D. d'Ari.** 1986. SOS-associated division inhibition gene *sfiC* is part of excisable element e14 in *Escherichia coli*. J. Bacteriol. **168:**464–466.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Novagen.** 1995. pET system manual no. 55, 5th ed. Novagen, Madison, Wis.
17. **Schneider, S. J.** 1992. Site-specific recombination of lambdoid phage 21 into the *icd* gene of *Escherichia coli*. Ph.D. dissertation. Stanford University, Stanford, Calif.
18. **Smith-Mungo, L., I. T. Chan, and A. Landy.** 1994. Structure of the P22 *att* site. Conservation and divergence in the lambda motif of recombinogenic complexes. J. Biol. Chem. **269:**20798–20805.
19. **Weisberg, R., and A. Landy.** 1983. Site-specific recombination in phage lambda, p. 211–250. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Yagil, E., L. Dorgai, and R. A. Weisberg.** 1995. Identifying determinants of recombination specificity: construction and characterization of chimeric bacteriophage integrases. J. Mol. Biol. **252:**163–177.
21. **Yu, A., and E. Haggard-Ljungquist.** 1993. Characterization of the binding sites of two proteins involved in the bacteriophage P2 site-specific recombination system. J. Bacteriol. **175:**1239–1249.