doi:10.1016/S0022-2836(03)00442-X J. Mol. Biol. (2003) 329, 983–996

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Analysis of Insertion into Secondary Attachment Sites by Phage λ and by *int* Mutants with Altered Recombination Specificity

Edit Rutkai[1], László Dorgai[1], Regina Sirot[2], Ezra Yagil and Robert A. Weisberg[3]*

[1]*Bay Zoltán Institute for Biotechnology, Derkovits Faser 2, H-6726 Szeged, Hungary*

[2]*Department of Biochemistry Tel Aviv University, Tel Aviv 69978, Israel*

[3]*Section on Molecular Genetics Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda MD 20892-2785, USA*

**Corresponding author*

When phage λ lysogenizes a cell that lacks the primary bacterial attachment site, integrase catalyzes insertion of the phage chromosome into one of many secondary sites. Here, we characterize the secondary sites that are preferred by wild-type λ and by λ *int* mutants with altered insertion specificity. The sequences of these secondary sites resembled that of the primary site: they contained two imperfect inverted repeats flanking a short spacer. The imperfect inverted repeats of the primary site bind integrase, while the 7 bp spacer, or overlap region, swaps strands with a complementary sequence in the phage attachment site during recombination. We found substantial sequence conservation in the imperfect inverted repeats of secondary sites, and nearly perfect conservation in the leftmost three bases of the overlap region. By contrast, the rightmost bases of the overlap region were much more variable. A phage with an altered overlap region preferred to insert into secondary sites with the corresponding bases. We suggest that this difference between the left and right segments is a result of the defined order of strand exchanges during integrase-promoted recombination. This suggestion accounts for the unexpected segregation pattern of the overlap region observed after insertion into several secondary sites. Some of the altered specificity *int* mutants differed from wild-type in secondary site preference, but we were unable to identify simple sequence motifs that account for these differences. We propose that insertion into secondary sites is a step in the evolutionary change of phage insertion specificity and present a model of how this might occur.

Published by Elsevier Science Ltd

Keywords: site-specific recombination; lysogeny; phage HK022; tyrosine recombinase; evolution

Introduction

Insertion of phage λ DNA into the bacterial chromosome to form lysogens is catalyzed by integrase (or Int), a phage-encoded site-specific recombinase.[1] Int cleaves, exchanges, and rejoins specific phosphodiester bonds within the bacterial attachment site, *attB*, and the phage attachment site, *attP*, to produce an inserted prophage that is flanked by two reciprocally recombinant sites, *attL* and *attR* (Figure 1). Int belongs to the tyrosine recombinase family. Family members are characterized by an invariant tyrosine residue that becomes transiently joined to the DNA backbone during strand cleavage and by several other highly conserved amino acids that activate strand cleavage, exchange, and rejoining.[2] Comparison of the three-dimensional structures of four widely diverged family members reveals considerable conservation of structure around the catalytic center.[3–6]

Phage λ Int (Int-λ) recognizes two distinct sequence motifs, the arm-type and the core-type binding sites (Figure 1).[7] The amino-terminal 64 residues form a domain that recognizes the arm-type sites, and the remainder of the protein, which includes the catalytic center, recognizes the core-type sites.[8,9] *attB* consists of two core-type sites

Abbreviations used: Int, integrase; Int-λ, phage λ integrase.

E-mail address of the corresponding author: rweisberg@nih.gov

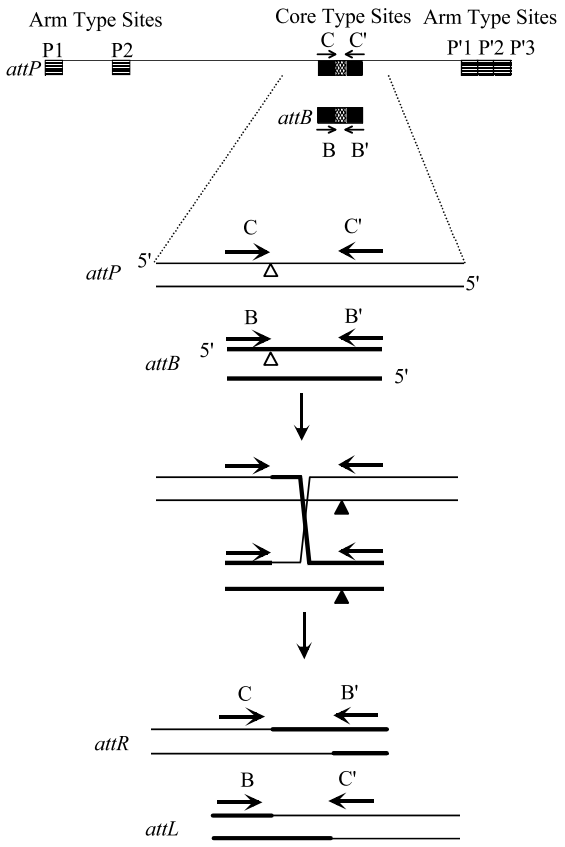

Figure 1. Arrangement of Int-binding sites in *attP* and *attB* and strand exchange in site-specific recombination. The five hatched boxes labeled P1, P2, P′1, P′2 and P′3 represent arm-type Int-binding sites in *attP*, and the arrows labeled C, C′ and B, B′ represent the core-type Int-binding sites in *attP* and *attB*, respectively. C is separated from C′, and B from B′, by the 7 bp overlap region (cross hatched). The initial strand cleavages occur in the two top strands at the left edge of the overlap region (open triangles). Three base-pairs are swapped and the strands are resealed to produce a Holliday structure (middle).[34] The two bottom strands are cleaved at the right edges of the overlap region (filled triangles), the remaining base-pairs are swapped, and the strands are resealed to produced *attR* and *attL* recombinants (bottom).

that are arranged as imperfect inverted repeats separated by a 7 bp spacer called the overlap region. *attP* contains an identically arranged pair of core-type sites that are flanked by multiple copies of the arm-type sites. Recombination specificity is dictated by Int recognition of the arm-type and the core-type sites, and by matching of the overlap region sequences. During recombination, each strand of the overlap region pairs with the complementary strand of the partner site (Figure 1), and therefore two attachment sites recombine efficiently only if their overlap regions are identical.[10–13]

Different temperate bacteriophages of *Escherichia coli* generally insert at sites that differ from *attB* of λ.[14] Many use tyrosine recombinases that are related to Int-λ, but these proteins are usually not functionally interchangeable with each other because each is highly specific to its cognate sites. It is not obvious how a new recognition specificity evolves from an existing one, especially if it does so within the constraints of a temperate bacteriophage lifestyle. A complete specificity change, defined as acquisition of a preference for a new *attB* and loss of the ability to recombine the original attachment sites, requires alteration of both Int and *attP*, and thus cannot occur in a single step. The problem is aggravated by the multiplicity of specificity elements in *attP*: changes in the overlap region and core-type binding sites are likely to be required, and differences in the arm-type sites can contribute.[15,16] The multiple alterations that are required for a complete change in specificity are probably constrained by selective pressures. Since insertion and excision doubtless contribute to the survival of temperate bacteriophage in nature, we expect that mutants that insert or excise poorly will be at a disadvantage.

We have shown that replacement of a small number of residues suffices to change the recombination specificity of Int-λ to that of HK022, a related phage. The two Ints recognize the same arm-type but different core-type sites, and the two phages insert at different *attB* sites in the bacterial chromosome.[17–21] The aligned sequences of the two proteins differ at 91 of 356 positions, nearly all of which lie outside of the amino-terminal arm-binding domain. Each protein recombines the non-cognate attachment sites with low but detectable efficiency relative to the cognate sites. Replacement of five λ-specific residues by their HK022 counterparts alters specificity completely; that is, the quintuple mutant recombines HK022 sites well and λ sites poorly.[22,23] Two of the mutations, *intN99D* and *intE319R*, relax specificity: they increase recombination of HK022 sites without changing that of λ sites significantly. The other three mutations, *intS282P*, *intG283K*, and *intR287K*, restrict specificity: they decrease recombination of wild-type λ sites without increasing recombination of HK022 sites significantly.

Although the recombination specificity of an Int can be altered by a relatively small number of amino acid replacements, this is not sufficient to change the insertion specificity of a phage. As noted above, changes in *attP* are required also, and it is likely that Int and *attP* co-evolve. Here, we explore the possibility that insertion at secondary attachment sites contributes to alteration of *attP* specificity. Insertion of λ into secondary sites occurs much more frequently than mutation and can be detected easily after infection of cells that lack *attB*.[24] Known secondary sites resemble *attB* in sequence and function,[25] and are therefore comparable to a collection of mutant *attB* sites. We characterized the most frequently used secondary sites and measured how often they are used by wild-type and by phage carrying relaxed, restricted, or both types of *int* mutations. Although many secondary sites have been sequenced, this is

the first systematic characterization of those that are used most frequently. We have characterized the products of insertion at secondary sites whose overlap region differs from that of the primary site and propose a model for evolutionary change in insertion specificity.

Results

Frequency of insertion into cells lacking *attB*

We estimated the overall frequency of insertion into secondary attachment sites by determining the proportion of stable lysogens relative to surviving cells after infection of a host that lacks *attB* (Δ*attB*).[24] We measured this frequency for λ *int*$^+$ and for three relaxed specificity mutants, λ *intN99D*, λ *intE319R*, and the double mutant. The two single mutants recombine HK022 sites better than wild-type λ Int, and the double mutant recombines HK022 sites almost as well as wild-type HK022 Int.[22] We found that all three mutants formed lysogens about as frequently as wild-type in both Δ*attB* and *attB*$^+$ hosts (Table 1). Therefore, the relaxing mutations do not change the overall efficiency of insertion into secondary sites.

The restricted specificity mutant, λ *int*S282P-G283K-R287K, lysogenized a Δ*attB* host very poorly and therefore inserts inefficiently into secondary sites (Table 1). It lysogenized a wild-type host poorly, as expected from the previously documented effect of these mutations on *attB* × *attP* recombination.[22] A quintuple mutant containing both relaxing and restricting mutations lysogenized the Δ*attB* and *attB*$^+$ hosts with similar efficiency, and this efficiency was well above that found for the restricted specificity mutant (Table 1). Therefore, as expected of an *int* mutant with HK022 specificity, it has weak, if any, affinity for *attB* of λ, and prefers to insert elsewhere in the bacterial chromosome.

Patterns of secondary site usage

To see if the different *int* mutants and wild-type prefer different secondary sites, we isolated DNA from a mixed population consisting of the progeny of about 3000 independent lysogens of a Δ*attB* host, cleaved the DNA with a restriction enzyme, and visualized the fragments that contained the left prophage-bacterial chromosome junctions by Southern hybridization to a probe containing the right arm of *attP*. Previous genetic analysis of individual secondary site lysogens revealed that wild-type λ has a relatively small number of preferred secondary attachment sites.[24] We therefore expected that the junction fragments would have a discrete size distribution, and that *int* mutants that alter the specificity of insertion could give a size distribution different from that of the wild-type. These expectations were largely confirmed by the results (Figure 2). All of the phages tested had a discrete distribution of junction fragments, and some of the specificity mutants had distributions that differed from that of the wild-type. Those of the relaxed mutants *intN99D* and *intN99D-E319R* differed from that of the wild-type, but resembled each other. *intE319R* gave a pattern that was very similar to although not completely identical with that of wild-type. None of the relaxed mutants inserted into the primary *attB* site of phage HK022, although they are known to promote recombination of HK022 attachment sites. (Insertion at *attB*-HK022 should give a fragment of about 6.5 kb, which was not seen.) This result is unsurprising because efficient Int-promoted recombination is favored by matching overlap regions, and the overlap regions of *attB*-HK022 and *attP*-λ differ at six of the seven positions.[26]

We were able to determine the size distribution of junction fragments from only a small number (12) of independent lysogens of the restrictive mutant, λ *intS282P-G283K-R287K*, because it integrated very poorly. The distribution differed from those of wild-type *int* and *intN99D*, and it therefore appears that this mutant has a distinct secondary site preference, although the small sample size limits our confidence in this conclusion. The integration pattern of the quintuple mutant does not reflect a simple sum of the sites used by λ *intN99D-E319R* and λ *intS282P-G283K-R287K* because a new and very efficient target for integration was detected (*yahM*; Figure 2).

Identification of individual secondary att sites

To identify the most frequently used secondary

Table 1. Frequency of λ insertion into secondary sites

Host	*int*$^+$	*intN99D*	*intE319R*	*intN99D-E319R*	*intS282P-G283K-R287K*	*intN99D-S282P-G283K-R287K-E319R*
attB$^+$	63 ± 13	47 ± 7	52 ± 5	48 ± 6	0.03 ± 0.01	1.8 ± 0.3
Δ*attB*	1.1 ± 0.3	0.6 ± 0.2	1.7 ± 0.5	0.5 ± 0.1	<0.04	2.2 ± 0.2
(Δ*attB*)/(*attB*$^+$)	0.017 ± 0.009	0.013 ± 0.007	0.033 ± 0.013	0.01 ± 0.003	–	1.2 ± 0.4

Frequencies of lysogeny were measured as described and are presented as the percentage of lysogens per surviving cell. The experiments were repeated at least three times, and the indicated errors are standard deviations of the mean. The hosts were RW495 (*attB*$^+$) and RW1401 (Δ*attB*). The phage stocks used were Y1 (column 2), Y1206 (column 3), Y1213 (column 4), Y1221 (column 5), Y1225 (column 6), and Y1228 (column 7).

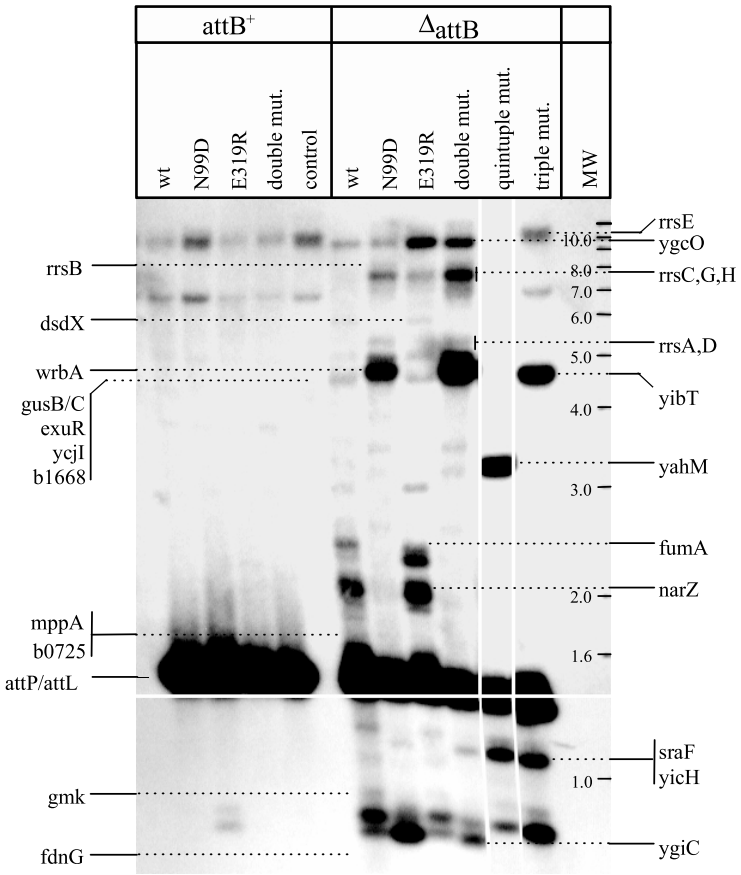

Figure 2. Southern blot analysis of the pattern of integration at secondary sites in the *E. coli* chromosome. *attB*$^+$ (left) or Δ*attB* (right) cultures were infected with λ carrying the *int* genes indicated on the second line, and independently arising lysogenic colonies were selected. The lane marked control contains DNA extracted from the same number of uninfected nonlysogenic cells. The double mutant is *intN99D-E319R*, the triple mutant is *intS282P-G283K-R287K*, and the quintuple mutant contains all five mutations. About 3000 such colonies were resuspended in liquid, DNA was extracted and digested with *Eco* RV, and the fragments were analyzed by Southern hybridization to a probe extending from positions −9 to +239 of the λ *attL* site (position 0 is the center of the overlap region). Bands whose sizes correspond to those of identified secondary sites are labeled. Bands expected from insertion at sites *mppA*, *b0725*, and *fdnG* were not visible (see the text). The very intense band in each lane below 1.6 kb has the size of *attL* and *attP*. *attL* is the result of insertion at *attB*-λ, while *attP* is found at the junction between multiple, tandem lysogens. To compensate for the difference in intensity of the bands, the top and bottom part of the Figure are taken from different exposures.

sites, we transferred chromosomal fragments from individual lysogens into a plasmid, amplified a segment containing a prophage–host junction by PCR, and determined the sequences of the junction regions (see Materials and Methods). We identified 19 different junction sequences in a collection of 471 lysogens. The phage chromosome was fused to the host chromosome at the overlap region in every lysogen tested, as expected if insertion occurred by Int-promoted recombination between attP and the secondary attachment site. Comparison of the junction sequences to the *E. coli* genome sequence allowed us to determine the chromosomal locations of the secondary sites, and these sites were named after the open reading frames or genes they were located in or were closest to (Figure 3). We saw no strong tendency of these sites to cluster in particular chromosomal regions. We then calculated the expected sizes of *attL*-containing restriction fragments and looked for fragments of these sizes in the Southern blots of DNA obtained from mixed cultures of lysogens (Figure 2). The sizes of many, but not all, of these fragments corresponded to those expected from the identified sites. Conversely, most, but not all, of the identified sites corresponded to bands found on the blot. There are three exceptions: bands with the expected mobility of *mppA*, *b0725*, and *fdnG* were not visible on the blot.

To measure the frequency of insertion at particular secondary sites, we isolated a large number of independent lysogenic colonies and asked if primers derived from any of the 19 sequenced sites could be used to amplify a prophage–host junction from each lysogen. The sequence of the second primer corresponded to phage DNA close to *attP*. The estimated frequencies are presented in Figure 3. We found that sites *ygiC*, *wrbA*, *rrsX*, *ygcO*, *narZ*, *fumA*, *dsdX*, and *exuR* were collectively used in 54%, 82%, 61% and 77% of the lysogens formed by λ *int*$^+$, λ *intN99D*, λ *intE319R*, and λ *intN99D-E319R*, respectively.

As noted above, some prophage–host junction fragments seen in the Southern blot of mixed populations of secondary site lysogens, including several that appear to be quite abundant, did not correspond to any of those expected from the secondary sites that were identified by PCR analysis of clones of individual lysogens (unlabelled bands in Figure 2). The apparent inconsitency could be due to insertions that alter the rate of bacterial growth. Less outgrowth after infection is needed to isolate the mixed populations that were used in Southern blotting than to isolate

A. Secondary Sites for λ *int*$^+$ and λ *intE319R*:

Site	Left Core	Overlap	Right Core	% Lysogens WT	% Lysogens *E319R*	Location (min)
λattP	CGTT CAGCTTT	TTTATAC-	TAAGTTG GCAT	—	—	—
λattB	AAGC CTGCTTT	TTTATAC-	TAACTTG AGCG	—	—	17
rrsX[a]	AAGT ACGCTTC	TTTAAGG-	TAAGGAG GTGA	24	1	—
ygcO	TATC CTTATTA	TTTATCC-	AAAGCGG TAGG	14	20	62
ygiC	CTGC CCGCTTA	TTTTGCG-	GAAGATG ATCA	3	21	69
narZ	ACGC CCACTTC	TTTAC C-	GAAGTAC GCTA	11	5	33
fumA	GTAT TCACTTC	TTTATACA	TATCCAG CGGC	1	7	36
dsdX	TGAC CCAGAAA	TTTAAGGA	AAAGTGG GCCA	1	6	53
exuR	AGAT CATGTTA	TTTAACG-	AAACCAG CGAC	0	1	70
gusB/C	ATGG CCACTAT	TTTTCTC-	ATAGTTG CACC	1	10	36
ycjI	TGTA TCACTTT	TTTGCCG-	GAAGTTG TTTG	3	3	30
gmk	GCTC ACACTT-	TTTGCGC-	AAAGTTG TTGC	1	0	82
b0725	TAAG CATCTTT	TTTATGC-	TTACTTC GCCG	0	1	16
b1668	CGCA CGGGTAT	TTTACCC-	GAAGCCG ATTT	0	1	38
yicH	AGAC CATGAAA	TTTATTGG	GAAGCTG CTTC	0	1	83
Not Identified				41	23	

B. Secondary Sites for λ *intN99D* and λ *intN99D-E319R*:

Site	Left Core	Overlap	Right Core	*N99D*	*N99D-E319R*	
λattP	CGTT CAGCTTT	TTTATAC-	TAAGTTG GCAT	—	—	—
λattB	AAGC CTGCTTT	TTTATAC-	TAACTTG AGCG	—	—	17
ygiC	CTGC CCGCTTA	TTTTGCG-	GAAGATG ATCA	49	54	69
wrbA	GCTA CCACTTA	TTTAAAGA	TAAGACG TCCT	21	9	23
ygcO	TATC CTTATTA	TTTATCC-	AAAGCGG TAGG	3	10	62
rrsX[a]	AAGT ACGCTTC	TTTAAGG-	TAAGGAG GTGA	9	3	—
ycjI	TGTA TCACTTT	TTTGCCG-	GAAGTTG TTTG	1	1	30
fumA	GTAT TCACTTC	TTTATACA	TATCCAG CGG-	0	1	36
mppA	TTTG TCACTTT	TTTGATG-	GAAGTTG TTTG	0	1	30
Not Identified				37	21	

C. Secondary Sites for λ *intS282P-G283K-R287K*:

Site	Left Core	Overlap	Right Core			
λattP	CGTT CAGCTTT	TTTATAC	TAAGTTG GCAT			—
yibT[b]	GGAG GAAGTTT	TTTCAGA	TACTCCC GGAA			81
rrsX[a]	AAGT ACGCTTC	TTTAAGG	TAAGGAG GTGA			—
sraF	CGTA TTCCTTT	TTTGTTG	TAAGTGA GACC			70

D. Secondary Sites for λ *intN99D-S282P-G283K-R287K-E319R*:

Site	Left Core	Overlap	Right Core			
λattP	CGTT CAGCTTT	TTTATAC	TAAGTTG GCAT	—		—
HK022*attP*	TGTC ATCCTTT	AGGTGAA	TAAGTTG TATA	—		—
HK022*attB*	CGGT GCACTTT	AGGTGAA	AAAGGTT GAGT	—		23
yahM	CGTG GTCCTTA	TTTCCGA	TAAGTTT ACTT	38		7
sraF[c]	CGTA TTCCTTT	TTTGTTG	TAAGTGA GACC	31		70
rrsX[a]	AAGT ACGCTTC	TTTAAGG	TAAGGAG GTGA	21		—
fdnG	ACTG ACCTTTC	TTCGCAA	TAAGCAC GCCA	1		33
Not Identified				9		

Figure 3. The sequences of secondary attachment sites obtained after infection with λ *int*$^+$ and various *int* mutants were grouped and aligned as described in the text. The sequences of λ and HK022 *attP* and *attB* were included in each alignment as appropriate for the known specificities of the phages. The group of six to eight bases in the center of each alignment is considered the overlap region, and the flanking bases to the left and right are considered the core-type Int-binding sites. The most highly conserved residues are indicated by shading and summarized in Table 1. The columns after the sequences show the frequency of insertions at each site for each of the infecting phages as judged by PCR amplification of junction fragments from individual lysogens. This frequency is not shown for λ *intN99D-G283K-R287K* because of the small number of lysogens analyzed. The rightmost column shows the location of each site on the *E. coli* chromosome. In all, 66 λ *int*$^+$, 102 λ *intN99D*, 78 λ *intE319R*, 69 λ *intN99D-E319R*, 12 λ *intS282P-G283K-R287K*, and 144 λ *intN99D-S282P-G283K-R287K-E319R* lysogens were analyzed. Not Identified means that no productive amplification was achieved with any of the primer pairs. [a]*rrsX* indicates one of the seven loci that encode 16 S rRNA. The sizes of the PCR fragments indicate that at least three of the seven loci were used. [b]*yibT* encodes a small peptide.[55] [c]*sraF* encodes a small RNA.[56]

individual lysogens, and strains that grow poorly because the insertion has disrupted important genes are likely to be more under-represented in the second population. Another source of variability, especially for sites that are used relatively infrequently, is statistical fluctuation due to limited population size. To obtain another estimate of site preference, we measured relative insertion

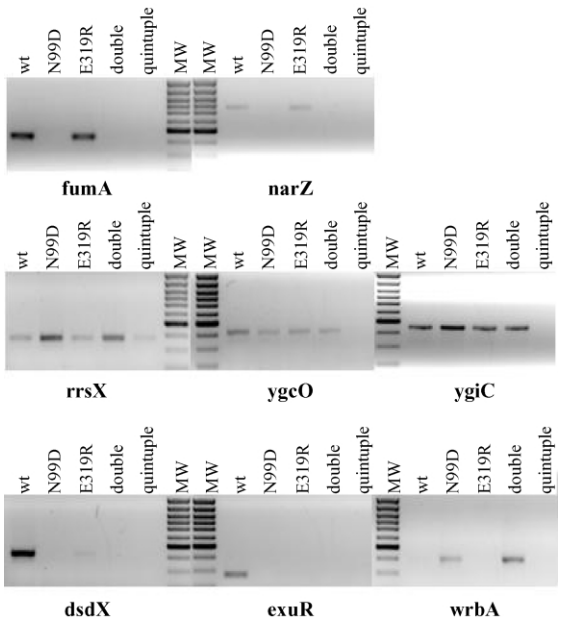

Figure 4. PCR analysis of the pattern of integration at secondary sites in the *E. coli* chromosome. Genomic DNA was prepared from a population of about 3000 independent lysogens formed by wild-type and mutant phages (see the legend to Figure 2). PCR reactions for detecting *rrsX*, *ygcO*, *wrbA* and *narZ* contained 10^4 genome equivalents of DNA while the reactions for *fumA*, *ygiC*, *dsdX* and *exuR* 10^5 genome equivalents of DNA. These conditions were determined by preliminary experiments that ensured that the reactions shown in the Figure were substrate-limited (data not shown). The site detected is indicated below each panel, and the Int proteins are indicated above each lane.

frequencies at eight frequently used sites by PCR amplification of DNAs prepared from mixed populations of lysogens that were prepared in the same way and were about the same size as those used for Southern blots. The results (Figure 4) were largely but not completely consistent with those obtained by amplifying DNA from many individual lysogens. One difference is that λ *int*$^+$ inserted into *dsdX* and *exuR* more frequently than did λ *intE319R* in the mixed population shown in Figure 4 but not in the collection of individual lysogens. Another is that differential usage of sites *rrsX*, *ygcO* and *ygiC* by λ *int*$^+$ and λ *int-N99D* appeared less extreme in the mixed population than in the collection of individual clones. We conclude that λ *intN99D* and λ *intN99D-E319R* differ from λ *int*$^+$ in their pattern of secondary site utilization. λ *intE319R* is similar to λ *int*$^+$ in this respect, but is probably not identical. Our estimates of frequencies of integration at specific sites are approximate because of the reduced viability of some lysogens, sampling errors, or both.

The data of Figures 2, 3, and 4 suggest the following classification of site usage. *rrsX* and *ygcO* are used about equally by *int*$^+$, *intN99D*, *intE319R*, and the double mutant. *wrbA* and *ygiC* are used more frequently by *intN99D* and *intN99D-E319R* than by *int*$^+$ and *intE319R*. Conversely, *narZ*, *fumA*, *dsdX* and *exuR* are used more frequently by *int*$^+$ and *intE319R* than by *intN99D* and *intN99D-E319R*. The secondary site preferences of the quintuple mutant that alters specificity from λ to HK022, *intN99D-E319R-S282P-G283K-R287K*, differ markedly from those of *int*$^+$, *intN99D*, and *intE319R*. Among the secondary sites used frequently by the two single mutants, only *rrsX* and another, unidentified, site were used by the quintuple mutant. Instead, two new sites, *yahM* and *sraF*, were highly preferred. *yahM* was not detected by PCR in mixed populations of λ *int*$^+$ or λ *intE319R* lysogens, even if 100-fold more template was used in the reactions. However, in mixed populations of λ *intN99D* or λ *intN99D-E319R* lysogens, *yahM* was detectable when tenfold more template was used, suggesting that substitution N99D is involved in the recognition of this site (data not shown). *sraF* was found in the small population of lysogens analyzed after infection by the restrictive mutant, λ *S282P-G283K-R287K*, suggesting that one or more of these mutations is involved in recognition of this site.

Sequence variability in secondary attachment sites

To analyze sequence variability in the secondary sites, we first divided them into four groups according to the patterns of site usage: those used by λ *int*$^+$ and λ *intE319R*; those used by λ *intN99D* and λ *intN99D-E319R*; those used by the restrictive mutant (λ *intS282P-G283K-R287K*); and those used by the altered specificity mutant (λ *intN99D-S282P-G283K-R287K-E319R*). We then aligned each group of sequences with the aid of clustal W, a multiple sequence alignment program (Figure 3).[27] The parameters used for the alignments imposed a small penalty (−1) for a 1 bp gap and a much larger one (−40) for extending the gap. We chose different penalties because Int-promoted recombination between *attP*, which has a 7 bp overlap region, and secondary sites with 6 bp or 8 bp overlap regions has been reported, but recombination involving sites with overlap regions of less than 6 bp or greater than 8 bp has not.[11,28,29] With these parameters, the program aligned the sites so that all of the experimentally determined prophage–host junctions obtained within each Int group aligned with each other, and no sequence had a gap larger than 1 bp (Figure 3). Two sites (*narZ* and *gmk*) were realigned by eye in such a way as to introduce an additional 1 bp gap in and adjacent to the overlap region to increase the amount of sequence identity in the flanking regions. A summary of sequence variability within the aligned secondary sites is shown in Table 2. The leftmost 3 bp of the overlap region are nearly invariant. They were T:A in all but one (*fdnG*) of the 19 we sequenced. The fourth base-pair of the overlap region was frequently A:T. These 4 bp match those at the corresponding positions of λ *attP*. The

Table 2. Summary of sequence conservation at positions corresponding to the left and right core-type binding sites and the overlap region of frequently used secondary attachment sites (Figure 3)

		Core site position											Overlap region position						
Int	Core site location	−4	−3	−2	−1	1	2	3	4	5	6	7	1	2	3	4	5	6	7
WT and E319R	Left	r	R	R	Y	C	c	R	C	T	T	H	T	T	T	A	N	N	c
	Right	N	R	R	N	C	R	R	C	T	T	H							
N99D and N99D-E319R	Left	N	r	t	Y	c	C	R	C	T	T	H	T	T	T	A	N	N	g
	Right	N	N	r	r	C	a	R	C	T	T	H							
N99D-S282P-G283K-R287K-E319R	Left	c	G	T	N	R	Y	N	C	T	T	t	T	T	T	N	N	N	N
	Right	r	N	N	C	N	r	r	C	T	T	A							

All sequences are written in a 5′ to 3′ direction, and sequences corresponding to the right core-type binding sites have been reversed and complemented compared to those presented in Figure 3. R indicates A or G, Y indicates T or C, H indicates A, C, or T, and N indicates no apparent preference. Upper case letters indicate more apparent conservation than lower case. Positions located 5′ of the "canonical" core-type sites are given negative numbers.[33] Secondary sites used by the restrictive mutant, *intS282P-G283K-R287K*, are omitted from this Table because only a small number of lysogens was analyzed.

rightmost three positions of the overlap region were much more variable. We present a model that accounts for these findings. The sequences to each side (numbered −4 to 7 in Table 2) were also non-random, although more variable than the leftmost 3 bp of the overlap region. In *attP* and *attB*, these segments contain a pair of core Int-binding sites, arranged as imperfect inverted repeats (Figure 1), and the alignment of Figure 3 suggests that this is true of the secondary sites as well. This can be seen more clearly in Table 2, where the right flanking sequences are reversed and complemented. The least variable positions in each of the four groups of sites were 4, 5, and 6, whose sequence is frequently 5′-CTT. This sequence is found at positions 4, 5, and 6 of most of the λ and HK022 *attP* and *attB* core sites as well. The base C was frequent at position 1 and, in the putative left core binding site, at position 2. G was never found at position 7, suggesting strongly that its presence would inhibit recombination. The sequence was more variable at positions 3 and −1 through −4.

Some secondary sites were clearly preferred to others by wild-type phage, but the degree of similarity of a secondary site to the consensus pattern (5′C-CTT-TTTA- - - -AAG- -G) and to wild-type *attB* or *attP* showed little correlation with the relative frequency with which it was used. We were unable to find sequence motifs that account for differential site usage by λ *int*$^+$ and λ *intN99D* by inspection of the aligned sequences or by a more rigorous analysis (X. Ma and T. Schneider, personal communication).[30] The sequences do, however, offer some clues about the basis for differential site utilization by the quintuple *int* mutant. This mutant is less particular than wild-type *int* or the relaxed mutants about the identity of the bases at position 1 of the core site. It appears to prefer sites with G and T at positions −3 and −2 of the left core site and C at position −1 of the right core site.

Table 3. Frequency of lysogeny after infection by phage with λ or HK022 overlap regions

	Host	
Phage overlap region	WT	Δ[*attB*-HK022]
λ	86	85
HK022	1.3	0.15
None	0.06	NT

The hosts were RW495 (WT) and EY1259 (Δ[*attB*-HK022]). The phage stocks used were W316 (λ overlap region), W314 (HK022 overlap region) and W307 (lacks overlap region and core sites). All three phage had *int* of λ, and W314 and W316 had wild-type λ core binding sites. NT, not tested.

The role of the overlap region sequence in secondary site selection

The overlap regions of *attP* and *attB* can be changed to many other sequences without affecting recombination if the changes are the same in the two recombining partners.[10–13] A requirement for sequence identity could, therefore, explain the finding of wild-type *attP* sequence at the leftmost positions of the overlap region in secondary sites.

Table 4. Secondary sites used by phages with the HK022 overlap region

Site	Left flank	Overlap	Right flank
A. *Primary HK022* att *sites*			
attP	ATCCTTT	**AGGTGAA**	TAAGTTG
attB	GCACTTT	**AGGTGAA**	AAAGGTT
B. *Secondary sites used by* attP-*overlap HK022 (W314)*			
ytfK	CCACTTC	**AGGTGgc**	GAAGTAC
yhiV	AGGCTTA	**AGGgGAc**	TTTCATG
pqqL	CAGGTTC	**AGGgttA**	TTAGTCC
C. *Secondary sites used by* int-*321h*			
yihL	CAACTTC	**AGGattA**	ATATATG
relE	GGGTTAT	**AGacGAg**	AAAGTTG
frvR	CCAGTTG	**AGGTGAA**	TAACGAA

The overlap nucleotides are in bold type. Matches of the overlap region to the wild-type HK022 sequence are indicated by upper case letters, and mismatches by lower case letters.

To test this hypothesis, we determined the sequences of sites used by a λ derivative with an HK022 overlap region. The HK022 overlap region differs from that of λ at six of seven positions (5′-TTTATAC for λ and 5′-AGGTGAA for HK022).[19,26] We found that the mutant phage integrated inefficiently into a wild-type host, but even less efficiently into a host that lacks the *attB* site of HK022 (Table 3).[19] Therefore, the preferred insertion site for this phage appears to be *attB* of HK022, and this conclusion was confirmed by PCR: eight of ten lysogens tested carried a prophage at this site. We identified and sequenced three secondary sites from lysogens of the host lacking *attB* of HK022 (Table 4). All contained 5′-AGG at the left end of the overlap region, as do the primary attachment sites of HK022. In addition, we characterized three secondary sites used by a λ-HK022 hybrid, λ321h. This phage has the *attP* site of HK022 and a chimeric *int* gene with the amino-terminal 321 residues from λ and the remaining residues from HK022.[23] Two of the sites had 5′-AGG and the third had 5′-AGA at the left end of the overlap region (Table 4). The rightmost four positions showed less sequence conservation among these six secondary sites. We conclude that the sequence conservation in the left segment of the overlap region arises from stimulation of recombination by sequence match and is not a consequence of Int preference. We propose an explanation for the much weaker sequence conservation observed in the right segment of the overlap region in Discussion.

Directional segregation of overlap region sequence

Recombination between *attP* and *attB* produces two sites that have each inherited one strand from each parent in the overlap region (Figure 1).[31] The *attL* recombinant has the top strand from *attP* and the bottom strand from *attB*, while the *attR* recombinant has the reciprocal arrangement. If recombination between *attP* and a secondary site with a different overlap region produces two heteroduplex recombinants, both overlap regions will have mismatched bases. If mismatch correction does not occur or has no strand or base preference, the *attL* and *attR* progeny of the original recombinants should have an equal chance of inheriting the phage or bacterial overlap region sequence. However, previous studies of λ insertion into a secondary site whose overlap region differed from that of λ at three rightward positions (4, 5, and 7) showed that the progeny of the *attL* recombinants preferentially inherited the phage sequence, and those of the *attR* recombinants preferentially inherited the bacterial sequence.[10] To explain such "directional segregation", it was proposed that both strand exchanges usually occur to the left of the three heterologous base-pairs, and therefore heteroduplex DNA with mismatched bases is formed only rarely.

Table 5. Directional segregation of the phage and secondary-site overlap sequences into the *attL* and *attR* sites of prophages inserted at the secondary *att* sites indicated in column 1

	attL overlap		*attR* overlap	
	λ	*E. coli*	λ	*E. coli*
ygiC	110	0	0	110
wrbA	33	0	0	33
yahM	54	0	0	54
Total	197	0	0	197

The sequences of the overlap regions of these sites are TTT ATAC (λ), TTTTGCG (*ygiC*), TTTAAAGA (*wrbA*), and TTT CCGA (*yahM*). PCR amplification of individual lysogens with primers containing overlap region sequences was used to decide if the overlap regions were of λ or *E. coli* origin, and the numbers show the number of sites of each type. The lysogens were from the collections described in Figure 3.

To see if directional segregation is more general and occurs with the same directionality at other sites, we sequenced the *attR* overlap regions of prophages inserted at 17 different secondary sites in which the overlap region differs from that of *attP*. We found that 15 *attR*s had the sequence of the secondary site, and two had the λ sequence. The two exceptional sites, *b0725* and *ygcO*, both have a single mismatch at position 6. When we examined 197 independent insertions at three frequently used secondary sites that have multiple overlap region mismatches, we found that all of the *attL* overlap regions had the λ sequence and all of the *attR* overlap regions had the secondary site sequence (Table 5). We conclude that unidirectional segregation is the rule when the right segments of the two overlap regions are multiply mismatched.

Discussion

We have shown that the identity of the bases at about 12 positions is constrained strongly in secondary attachment sites that are used frequently by wild-type λ. Site preference is probably determined mainly or entirely by sequence because we detected no strong preference for particular chromosomal regions. Eight of the constrained positions correspond to the left and right core-type Int-binding sites in *attB*, and they presumably serve the same function in the secondary sites. The remaining four correspond to the left segment of the overlap region, and we suggest below that these are constrained because they must undergo complementary base-pairing with the overlap region of *attP* during recombination.

One *int* mutation, *N99D*, altered the pattern of usage of individual sites. This mutation also relaxes Int specificity so that it recombines the attachment sites of phage HK022 more efficiently. We suggest that *intN99D* relaxes specificity by altering specific contacts that occur between Int and *att* sites at some stage of the reaction. It

probably does not increase the rate of a sequence-independent step in the reaction, since this should have caused an increase in overall insertion frequency and should not by itself have changed the pattern of usage of individual sites. It is likely that residue 99 is at or close to a DNA-binding surface of Int. Cheng *et al*. showed that mutation of D99 in HK022 Int promotes binding to the λ B′ core site, and proposed that D99 has an unfavorable interaction with core residue G4.[20] Tirumalai *et al*. showed that nearby residue K103 of λ Int can be cross-linked to core site DNA.[32] We were unable to identify a simple sequence motif that accounts for differential site utilization by *intN99D*. Our inability to find such a motif could be due to the overriding importance of the overlap region sequence in secondary site selection. However, differential site utilization could be a result of sequence-dependent changes in DNA properties such as twist or stiffness that do not have a unique sequence signature.

A second specificity relaxing mutation, *intE319R*, did not appreciably change either the pattern or the overall frequency of secondary site usage. Comparison of λ Int to the related Cre recombinase suggests that E319 corresponds to R301 of Cre, which is not located at the protein–DNA interface in a co-crystal.[2,4] Replacement of E319 by any of several amino acids (including glycine), whose only common element is loss of a negative charge, increases recombination of HK022 sites.[22] These considerations suggest that E319 does not directly contact core site DNA, but instead affects site recognition indirectly, perhaps through interaction with residues that themselves contact DNA. We speculate that the potential new secondary sites made accessible by the relaxed specificity of *intE319R* are represented poorly in the *E. coli* chromosome.

The restrictive mutant, λ *intS282P-G283K-R287K*, inserts much less efficiently than wild-type into secondary attachment sites. We suggest that this mutant has acquired a new DNA contact or contacts that limit or prevent insertion into *attB* of λ and most potential secondary sites. This suggestion is consistent with previous work showing that some or all of these three mutations sensitize λ Int to the presence of certain nucleotides in the λ attachment sites.[33] The very low frequency of insertion of the mutant phage at secondary attachment sites could be the result of inhibitory contacts of the mutant Int with λ *attP* if we assume that relaxing mutations can compensate for the inhibition. It is interesting that introduction of the two relaxing mutations into the restrictive triple mutant allows recombination of HK022 *att* sites, and increases the overall frequency of insertion at secondary *att* sites. Only four secondary sites were used frequently by the quintuple mutant, and their pattern does not appear to be a simple sum of the patterns of the component mutations. The frequently used *yahM* site was not detected among sites used by the relaxed or the restrictive mutants, and *yibT*, which was used by λ *intS282P-G283K-R287K*, was not detected among sites used by the quintuple mutant. We suggest that there is some interaction among the surfaces altered by these mutations in the protein–DNA complex, and that these interactions contribute to the specificity alteration.

The sequence of the left part of the overlap region is the most constrained segment of this collection of secondary sites. A previous analyses had hinted at the same trend.[25] Although a few secondary sites in which the leftmost positions differed from this pattern had been found, all of them were recognized as rare Int-promoted deletions of λ phage rather than as insertion sites in the host chromosome. We show here that this sequence is nearly invariant because it tends to match the sequence at the corresponding positions in the attachment site of the inserting phage. The positive effect of overlap region identity on Int-promoted recombination has been known for some time,[10] but we had not anticipated its dominant effect on secondary site selection or the left-right asymmetry of sequence conservation.

Why are the rightmost bases of the overlap region less well conserved than those on the left? Mismatch between bases at the left is known to reduce recombination between primary attachment sites much more than mismatch at the right.[13] This observation can be understood as follows. Int catalyzes two successive strand swaps involving first the left and then the right segments of the participating overlap regions, and mismatches inhibit either the first or second swap according to their locations.[34–36] The first swap produces and the second resolves the Holliday structure intermediate (Figure 1). Inhibition of the first swap by sequence heterology has a more severe effect on the overall reaction than inhibition of the second because the Holliday structure can probably be resolved to recombinant products by Int-independent mechanisms *in vivo*.[10] For example, resolution could occur by DNA replication through the

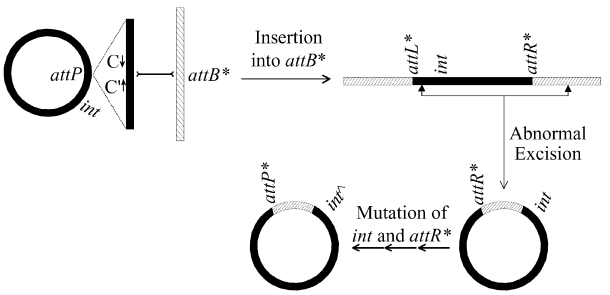

Figure 5. Chromosome jumping model for the evolution of insertion specificity. Prophage insertion into a secondary attachment site (*attB**) is followed by abnormal excision at the points indicated by the filled arrowheads to give a transducing phage that carries *attR** and adjacent bacterial DNA. Mutations of *attR** and *int* confer *attP* function on *attR** (*attP**) and alter Int specificity (*int*^) so that it recombines *attP** with *attB** but no longer recombines the original sites.

Holliday structure or by cleavage of the unexchanged strand by nucleases such as RuvC, Rap, or Rus.[37,38] If we assume that recombination of *attP* with secondary sites that differ at the right of the overlap region proceeds to the stage of the top strand swap, resolution of the resulting Holliday structure by an Int-independent mechanism will produce the observed directional segregation of the sequence differences. A similar model can account for the sequences of the recombinants produced by the integron-encoded integrases, as pointed out by Recchia & Sherratt.[39]

This work was motivated, in part, by our lack of understanding of how new prophage insertion specificities arise in nature (see Introduction). The findings presented here led us to modify an earlier model proposed to explain the distribution of prophage attachment sites in the *E. coli* chromosome.[40] The new model, "chromosome jumping", is illustrated by Figure 5. First, a phage inserts at a secondary bacterial attachment site, *attB**, which will become a new *attB*. This step could be preceded by an *int* mutation that relaxes specificity. Next, abnormal excision of this prophage produces a specialized transducing phage in which *attP* and flanking phage DNA to its right has been replaced with a secondary *attR* site, *attR**, and bacterial DNA to its right. The point of abnormal excision within the prophage is close enough to the host–phage junction that the transducing phage retains some or all of the arm-type Int-binding sites that are normally located to the right of the overlap region (the P′ sites in λ) and, of course, the entire *int* gene. *attB** and *attR** will become new *attB* and *attP* sites, respectively, by accumulation of appropriate mutations.

Co-adaptation of Int and the new *att* sites plausibly occurs by *int* mutations that are analogous to those required to change the specificity of Int-λ to that of Int-HK022. The initial mutations relax the stringency with which Int protein recognizes core type binding sites, thereby improving recognition of the new *attP* and *attB*. Then, a second class of *int* mutations restricts specificity by preventing recombination of the original *attP* and *attB* sites.

Chromosome jumping has the attractive feature that new *attP* and *attB* sites evolve from sequences that already have some activity as Int substrates, a feature not found in the earlier model.[40] This is especially important for generation of a new *attP* because it is large and complicated. *attR**, the proposed ancestor of a new *attP*, should contain the overlap region of *attB** because of directional segregation. Our present results confirm the importance of a changed overlap region to a changed insertion specificity. *attR** will have core-type and a partial set of arm-type Int-binding sites. The remaining arm-type sites will be separated from the core-type sites by a segment of bacterial DNA whose length will depend on the location of the sequence used for abnormal excision. The corresponding segment of wild-type λ *attP* contains a binding site for IHF, a DNA bending protein, and it is has been shown that the requirement for at least some of the IHF-binding sites in the attachment sites can be satisfied by naturally bent DNA segments of variable length.[41,42] These considerations suggest that the newly born *attR** transducing phage might already possess substantial *attP* function or be able to acquire it by deletion or rearrangement of the bacterial DNA. If the *attR** site of the newborn phage recombines poorly with *attB**, the bacterial DNA will allow inefficient insertion into the bacterial chromosome by homologous recombination, a property that could increase phage survival during the period of adaptation of *attR** to *attB** and of Int to the new attachment sites. In principle, abnormal prophage excision that included *attL** and bacterial DNA to the left could give rise to a phage with *attP* function, but such a phage would probably not contain the overlap region of the secondary site.

It is noteworthy that the *attP* sites of four known phages, P22, HP1, L5, and 21, have the structure predicted by the model. Each overlap region is at one end of a duplication of bacterial DNA that is used to reconstruct the bacterial genes into which they insert. Int-binding sites, which are probably analogs of the arm-type sites of *attP*-λ, are near the other end of the duplication.[15,43–46]

Nearly all of the secondary sites we characterized contained 5′-TTT at the left of the overlap region, although the bases at these positions do vary among phages with closely related Int proteins, such as λ, HK022, and 21.[14] How can chromosome jumping generate such variability? One possibility is that the order of strand swapping in the new *attP* site differs from that in the old, since in λ this order is dictated by unknown elements in the arms of *attP*.[35] In this scenario, reiteration of chromosome jumping with an appropriate choice of excision end-points provides a mechanism for changing the bases in the left segment of the overlap region. Reiteration of chromosome jumping could account for the observation that none of the frequently used λ secondary sites corresponds to the insertion site of a known phage. Campbell has proposed that the orientation of prophages in the bacterial chromosome is subject to weak selection arising from the direction of bacterial chromosome replication.[47] Thus, the secondary sites we have found might represent unstable intermediates on the pathway to a more favorable *attB* location. In addition, some known phage insertion sites could be missing from our collection of secondary sites because they are located within bacterial genes encoding vital functions (e.g. genes encoding tRNAs, which often contain *attB* sites). Our screen would not isolate such insertions as stable λ lysogens. Insertion at these sites by phages that are adapted to them does not disrupt gene function because duplicated sequences carried by the phage reconstruct the

Table 6. Bacterial and phage strains used in this work

	Relevant genotype	Origin/reference
A. *Bacteria*		
EY1546	LE392 Δ[*attB*]:: neo	This work (see Materials and Methods)
RW495	*attB*$^+$ *supE44*	54
RW1401	ΔSA269[*galT-attB*λ-*bioB*] *supE44*	10
B. *Phage*		
Y1	λ *cI857*	NIH collection
Y1206	λ *intN99D cI857*	22
Y1213	λ *intE319R cI857*	22
Y1221	λ *intN99D-E319R cI857*	22
Y1225	λ *intS282P-G283K-R287K cI857*	22
Y1228	λ *intN99D-S282P-G283K-R287K-E319R cI857*	22
W248	λ *hφ80 Δ248*[*attP-int*] *cI248*; used to select λ lysogens	24
W307	λ *Δ307*[*attP*] (core sites and overlap deleted)	19
W314	λ *attP* with overlap of HK022	This work; *attP* derived from pNR161[19]
W316	λ *attP*$^+$ (isogenic to W307 except for Δ307)	This work; *attP* derived from pHN955[35]
λ321h	λ-HK022 hybrid with chimeric Int and *attP* of HK022	22

gene. Duplications of segments of vital genes could arise by chromosome jumping if abnormal prophage excision and lytic phage growth occurred within a few generations of insertion, or if the lysogen contained a pre-existing duplication of the entire gene.

Campbell has proposed that the primordial *attB* of phages whose insertion depends on a tyrosine recombinase is in a tRNA gene, and that subsequent diversification produced integrases and *attP* sites capable of using *attB*s located in genes encoding proteins, in intergenic regions, and in other tRNA genes.[48] Among the possible mechanisms of diversification are recombination with *int* and *att* homologs that have evolved in a different biological context, such as a plasmid, the host chromosome, or a mobile genetic element such as an integron. Such homologs are not subject to the same selective pressures as are phages and might be freer to change their specificity or be less specific to begin with. In this view, the insertion specificity of more highly evolved phages like λ and HK022, both of which insert in intergenic regions, might be unable to change without recombinational transfers from other genetic elements. We are attempting to challenge this idea by asking if phages with altered specificity can arise after insertion into a secondary site and, if so, whether they have the structure predicted by the chromosome jumping model.

Materials and Methods

Strains and media

Phages and bacterial strains used in this work are listed in Table 6. Bacterial strain EY1546, which lacks *attB* of HK022, was constructed as follows. The overlap region of *attB*-HK022 carried by a plasmid (pSD5[18]) was replaced by an *Nru*I restriction site using the following oligomer: 5′-GGGTTAAATTCACGGTCGGT GCATCGCGAAAAGGTTGAGTCGC (the *Nru*I site is underlined). The sequence of the mutant plasmid was confirmed. A blunt-ended KmR cassette was cloned into the *Nru*I site, and the resulting KmR plasmid was linearized and transformed into strain JC7623 in order to incorporate the *attB*:: KmR mutation into the chromosome of the host.[49] Strain EY1546 is a derivative of LE392 that carries the *attB*:: KmR mutation and is unable to support efficient lysogenization by HK022. Cells were grown in tryptone broth (1% (w/v) Difco tryptone, 0.5% (w/v) NaCl) or Luria-Bertani broth (tryptone broth plus 0.5% (w/v) yeast extract). LBMM (Luria-Bertani broth containing 0.2% (w/v) maltose and 10 mM $MgSO_4$) was used when needed to promote phage absorption. Phage were grown and assayed as described.[50,51]

Measuring integration frequency at secondary sites and selecting lysogens

Overnight cultures of the appropriate hosts were diluted 100-fold in LBMM, grown until late logarithmic phase, collected by centrifugation and resuspended in 10 mM $MgSO_4$ to $A_{600} = 1 \pm 0.05$. The cells were starved at 32 °C for one hour and chilled on ice. A portion (100 μl of this suspension; approximately 10^8 cells) was mixed with the appropriate phage at a multiplicity of infection of 10. Phages were absorbed on ice for ten minutes followed by adjusting the volume to 1 ml with LBMM at 32 °C. After one hour of incubation at 32 °C cells were plated in the presence of λ *c* (strain W248) to select lysogens, and without W248 to measure cells surviving phage infection. Integration frequency is expressed as 100 × (No. lysogens)/(No. surviving cells). Individual lysogens were purified from plates containing W248, and mixed populations of independently isolated lysogens were prepared by resuspending all of the colonies on one or more such plates in a few millilitres of medium.

Identification of secondary sites

Genomic DNA prepared from an individual lysogen was digested with a restriction enzyme, and the fragment population was ligated to similarly cut pBC SK vector DNA (Stratagene). The ligation mixture (1 μl) was used for template in PCR reactions, in which vector-specific and phage-specific primers were used in pairs to amplify either *attL* or *attR*-containing DNA segments. To ensure specificity, the products of the first

PCR reactions were further amplified in a second reaction using nested vector- and phage-specific primers.

For successful amplification, several trials were performed: different restriction enzymes (*Eco*RI, *Eco*RV, *Bam*HI, *Hpa*II and *Hae*III) were used, and several ligation reactions were carried out with a fixed amount of vector and variable amounts of digested genomic DNA. The product of the ligation was used in the first PCR reactions either directly or after a 10–100-fold dilution. The specificity of the product of the nested PCR was further tested in a semi-nested reaction. The amplicons that passed the specificity check were isolated from agarose gel and sequenced.

Measuring integration frequencies at selected secondary sites

A collection of individual lysogens or a mixed population of lysogens, formed by either int^+ or *int* mutant phage, was screened by PCR, using primers complementary to the P or P′ arms of *attP* paired with a primer complementary to sequences flanking a particular secondary *attR* or *attL*, respectively. For template, 1 μl of overnight cultures of the lysogens, lysed by 1% (v/v) Triton X-100 at 95 °C, was used. Amplification of a product of the expected size signaled integration at the particular site. Lysogens without productive amplification were further screened with a primer pair specific to another secondary site. The sequences of the primers are available from the authors upon request.

Checking the segregation of the overlap region

Primers to sites *ygiC*, *wrbA* and *yahM* were paired with phage-specific primers. A primer specific to the wild-type overlap region and another with the overlap region of the secondary site was designed for each secondary *attL* and *attR*. Productive amplification with one of the primer pairs and lack of the amplification with the other signaled the sequence of the overlap.

Nucleic acid techniques

Hybridization

The genomic DNAs were digested with *Eco*RV, fractionated on a 1% (w/v) agarose gel, blotted onto GeneScreen Plus membranes (NEN Research Products) by capillary transfer, fixed to the membrane by UV irradiation (CL-1000 UC Crosslinker, UVP), and hybridized to the radioactive probe as described.[52,53] The probe was a segment of λ P′ (from −9 to +229) that was labeled by incorporation of [α-^{32}P]dCTP in a PCR reaction. Restriction enzymes and phage T4 DNA ligase were purchased from Fermentas and used as recommended by the manufacturer.

Amplification by PCR

PCR reactions were performed in a PTC 200 thermal cycler (MJ Research). Oligonucleotides were purchased from BioServe Biotechnologies (Laurel, MD, USA) or Biocenter Kft. (Szeged, Hungary). A typical PCR amplification consisted of 30 cycles of denaturation at 95 °C, annealing at 55 °C and synthesis at 72 °C, 30 seconds each. For amplification of larger products, 60 seconds synthesis time was used. The amplification was done in a 50 μl volume using 50–100 pmol of primers. The other conditions were as recommended by the supplier of *Taq* DNA polymerase (Perkin Elmer or ZenonBio Kft., Szeged, Hungary). dNTPs were purchased from Fermentas. The genomic DNAs used for PCR templates and for Southern hybridizations were prepared from populations or individual lysogens using the Wizard Genomic DNA Purification Kit (Promega).

Acknowledgements

We are grateful to Xiao Ma and Tom Schneider for helping us to analyze sequence conservation in secondary sites, to Art Landy for his comments on the manuscript, and to Allan Campbell and Howard Nash for helpful discussions and the communication of unpublished information. This work was supported, in part, by an NIH FIRCA grant. E.R. received a PhD fellowship grant from the Bay Zoltán Foundation for Applied Research, E.Y. was supported by the Israel Science Foundation (grant no. 405/98), and the Hungarian Ministry of Education and the Israel Ministry of Science supported travel grants to E.R., L.D., and E.Y.

References

1. Azaro, M. A. & Landy, A. (2002). λ integrase and the λ Int family. In *Mobile DNA II* (Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M., eds), pp. 118–148, ASM Press, Washington, DC.
2. Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucl. Acids Res.* **26**, 391–406.
3. Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. & Rice, P. A. (2000). Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell*, **6**, 885–897.
4. Guo, F., Gopaul, D. N. & Van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.
5. Hickman, A. B., Waninger, S., Scocca, J. J. & Dyda, F. (1997). Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. *Cell*, **89**, 227–237.
6. Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J. & Wigley, D. B. (1997). Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **16**, 5178–5187.
7. Ross, W. & Landy, A. (1982). Bacteriophage lambda int protein recognizes two classes of sequence in the phage att site: characterization of arm-type sites. *Proc. Natl Acad. Sci. USA*, **79**, 7724–7728.
8. Wojciak, J. M., Sarkar, D., Landy, A. & Clubb, R. T. (2002). Arm-site binding by lambda-integrase: solution structure and functional characterization of its amino-terminal domain. *Proc. Natl Acad. Sci. USA*, **99**, 3434–3439.
9. Moitoso de Vargas, L., Pargellis, C. A., Hasan, N. M., Bushman, E. W. & Landy, A. (1988). Autonomous DNA binding domains of lambda integrase

recognize two different sequence families. *Cell*, **54**, 923–929.
10. Weisberg, R. A., Enquist, L. W., Foeller, C. & Landy, A. (1983). Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. *J. Mol. Biol.* **170**, 319–342.
11. de Massy, B., Studier, F. W., Dorgai, L., Appelbaum, E. & Weisberg, R. A. (1984). Enzymes and sites of genetic recombination: studies with gene-3 endonuclease of phage T7 and with site-affinity mutants of phage lambda. *Cold Spring Harbor Symp. Quant. Biol.* **49**, 715–726.
12. Bauer, C. E., Gardner, J. F. & Gumport, R. I. (1985). Extent of sequence homology required for bacteriophage lambda site-specific recombination. *J. Mol. Biol.* **181**, 187–197.
13. Bauer, C. E., Gardner, J. F., Gumport, R. I. & Weisberg, R. A. (1989). The effect of attachment site mutations on strand exchange in bacteriophage lambda site-specific recombination. *Genetics*, **122**, 727–736.
14. Campbell, A. M. (1992). Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* **174**, 7495–7499.
15. Smith-Mungo, L., Chan, I. T. & Landy, A. (1994). Structure of the P22 att site. Conservation and divergence in the lambda motif of recombinogenic complexes. *J. Biol. Chem.* **269**, 20798–20805.
16. PeZa, C. E., Stoner, J. & Hatfull, G. F. (1998). Mycobacteriophage D29 integrase-mediated recombination: specificity of mycobacteriophage integration. *Gene*, **225**, 143–151.
17. Dhillon, T. S. (1981). Temperate coliphage HK022: virions, DNA, one-step growth, attachment site, the prophage genetic map. *J. Gen. Virol.* **55**, 487–492.
18. Yagil, E., Dolev, S., Oberto, J., Kislev, N., Ramaiah, N. & Weisberg, R. A. (1989). Determinants of site-specific recombination in the lambdoid coliphage HK022. An evolutionary change in specificity. *J. Mol. Biol.* **207**, 695–717.
19. Nagaraja, R. & Weisberg, R. A. (1990). Specificity determinants in the attachment sites of bacteriophages HK022 and lambda. *J. Bacteriol.* **172**, 6540–6550.
20. Cheng, Q., Swalla, B. M., Beck, M., Alcaraz, R., Jr, Gumport, R. I. & Gardner, J. F. (2000). Specificity determinants for bacteriophage Hong Kong 022 integrase: analysis of mutants with relaxed core-binding specificities. *Mol. Microbiol.* **36**, 424–436.
21. Gottfried, P., Yagil, E. & Kolot, M. (2000). Core-binding specificity of bacteriophage integrases. *Mol. Gen. Genet.* **263**, 619–624.
22. Dorgai, L., Yagil, E. & Weisberg, R. A. (1995). Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases. *J. Mol. Biol.* **252**, 178–188.
23. Yagil, E., Dorgai, L. & Weisberg, R. A. (1995). Identifying determinants of recombination specificity: construction and characterization of chimeric bacteriophage integrases. *J. Mol. Biol.* **252**, 163–177.
24. Shimada, K., Weisberg, R. A. & Gottesman, M. E. (1972). Prophage lambda at unusual chromosomal locations. I. Location of the secondary attachment sites and the properties of the lysogens. *J. Mol. Biol.* **63**, 483–503.
25. Weisberg, R. A. & Landy, A. (1983). Site-specific recombination in phage lambda. In *In Lambda II* (Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A., eds), pp. 211–250, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
26. Kolot, M. & Yagil, E. (1994). Position and direction of strand exchange in bacteriophage HK022 integration. *Mol. Gen. Genet.* **245**, 623–627.
27. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
28. Ross, W., Shulman, M. J. & Landy, A. (1982). Biochemical analysis of att-defective mutants of the phage lambda site-specific recombination system. *J. Mol. Biol.* **156**, 505–522.
29. Shulman, M. J., Mizuuchi, K. & Gottesman, M. M. (1976). New att mutants of phage lambda. *Virology*, **72**, 13–22.
30. Schneider, T. D. (1997). Sequence walkers: a graphical method to display how binding proteins interact with DNA or RNA sequences. *Nucl. Acids Res.* **25**, 4408–4415.
31. Mizuuchi, K., Weisberg, R. A., Enquist, L. W., Mizuuchi, M., Buraczynska, M., Foeller, C. *et al.* (1981). Structure and function of the phage lambda att site: size, int-binding sites, location of the crossover point. *Cold Spring Harbor Symp. Quant. Biol.* **45**, 429–437.
32. Tirumalai, R. S., Kwon, H. J., Cardente, E. H., Ellenberger, T. & Landy, A. (1998). Recognition of core-type DNA sites by lambda integrase. *J. Mol. Biol.* **279**, 513–527.
33. Dorgai, L., Sloan, S. & Weisberg, R. A. (1998). Recognition of core binding sites by bacteriophage integrases. *J. Mol. Biol.* **277**, 1059–1070.
34. Nunes-Düby, S. E., Azaro, M. A. & Landy, A. (1995). Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr. Biol.* **5**, 139–148.
35. Kitts, P. A. & Nash, H. A. (1988). Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. *J. Mol. Biol.* **204**, 95–107.
36. Nunes-Düby, S. E., Matsumoto, L. & Landy, A. (1987). Site-specific recombination intermediates trapped with suicide substrates. *Cell*, **50**, 779–788.
37. Sharples, G. J. (2001). The X philes: structure-specific endonucleases that resolve Holliday junctions. *Mol. Microbiol.* **39**, 823–834.
38. West, S. C. (1997). Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet.* **31**, 213–244.
39. Recchia, G. D. & Sherratt, D. J. (2002). Gene acquisition in bacteria by integron-mediated site-specific recombination. In *Mobile DNA II* (Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M., eds), pp. 162–176, ASM Press, Washington, DC.
40. Campbell, A., Schneider, S. J. & Song, B. (1992). Lambdoid phages as elements of bacterial genomes. *Genetica*, **86**, 259–267.
41. Goodman, S. D. & Nash, H. A. (1989). Functional replacement of a protein-induced bend in a DNA recombination site. *Nature*, **341**, 251–254.
42. Goodman, S. D., Nicholson, S. C. & Nash, H. A. (1992). Deformation of DNA during site-specific recombination of bacteriophage lambda: replacement of IHF protein by HU protein or sequence-directed bends. *Proc. Natl Acad. Sci. USA*, **89**, 11910–11914.
43. Hakimi, J. M. & Scocca, J. J. (1994). Binding sites for

bacteriophage HP1 integrase on its DNA substrates. *J. Biol. Chem.* **269**, 21340–21345.
44. Hauser, M. A. & Scocca, J. J. (1992). Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1. Identification of the points of recombinational strand exchange and the limits of the host attachment site. *J. Biol. Chem.* **267**, 6859–6864.
45. Peña, C. E., Lee, M. H., Pedulla, M. L. & Hatfull, G. F. (1997). Characterization of the mycobacteriophage L5 attachment site, attP. *J. Mol. Biol.* **266**, 76–92.
46. Campbell, A., Campillo-Campbell, A. & Ginsberg, M. L. (2002). Specificity in DNA recognition by phage integrases. *Gene*, **300**, 13–18.
47. Campbell, A. M. (2002). Preferential orientation of natural lambdoid prophages and bacterial chromosome organization. *Theor. Popul. Biol.* **61**, 503–507.
48. Campbell A. M. (2003). *DNA Research*, in the press.
49. Winans, S. C., Elledge, S. J., Krueger, J. H. & Walker, G. C. (1985). Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. *J. Bacteriol.* **161**, 1219–1221.
50. Oberto, J., Weisberg, R. A. & Gottesman, M. E. (1989). Structure and function of the nun gene and the immunity region of the lambdoid phage HK022. *J. Mol. Biol.* **207**, 675–693.
51. Silhavy, T. J., Berman, M. L. & Enquist, L. W. (1984). *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
52. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
53. Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517.
54. Enquist, L. W. & Weisberg, R. A. (1977). A genetic analysis of the att-int-xis region of coliphage lambda. *J. Mol. Biol.* **111**, 97–120.
55. Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G. & Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* **15**, 1637–1651.
56. Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G., Margalit, H. & Altuvia, S. (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.* **11**, 941–950.

Edited by M. Gottesman

(Received 15 November 2002; received in revised form 21 March 2003; accepted 21 March 2003)