Secondary Attachment Site for Bacteriophage Lambda in the *guaB* Gene of *Escherichia coli*

MARK S. THOMAS† AND WILLIAM T. DRABBLE*

Department of Biochemistry, University of Southampton, Southampton S09 3TU, United Kingdom

Received 26 March 1986/Accepted 26 July 1986

λ *gua* transducing bacteriophages were used to identify and sequence the secondary attachment site for λ in the *guaB* gene of *Escherichia coli*. The sequence matched the primary core sequence at nine positions, and a putative integrase binding-site overlapped the left core-arm junction. Recombinational crossover occurred between nucleotides −3 and +2 of the core region.

Bacteriophage λ encodes a site-specific recombination system that is used to integrate the phage chromosome into the chromosome of its host *Escherichia coli* (for a review, see reference 22). Integration involves recombination between the phage attachment site *attP* (POP') and the bacterial attachment site *attB* (BOB'). This process requires both the product integrase (Int) of the phage *int* gene and integration host factor, a host-encoded protein (5). A single reciprocal crossover between *attP* and *attB* occurs within a short region of sequence common to the two sites, the 15-base-pair (bp) core designated 'O' (9). The inserted prophage is thus flanked by the hybrid attachment sites *attL* (BOP') to the left and *attR* (POB') to the right. Apart from the core region, *attP* and *attB* are very different in both size and protein-binding characteristics (7). The *attP* site is quite large, encompassing approximately 240 bp (7, 11), and contains three binding sites for integration host factor (3), two classes of Int-binding site (arm type and junction type) (14), and two sites for binding Xis, a phage-encoded protein that is involved specifically in excision (23). The *attB* site is much smaller, spanning −11 to +11 (0 is the center of the common core sequence), and contains only two junction-type Int-binding sites (12).

The frequency of λ insertion into an *E. coli* mutant deleted for *attB* is less than 1% of the frequency of insertion into the chromosome of the parent strain (18). Insertion, which still requires Int and *attP*, occurs at a number of secondary attachment sites within the host chromosome (referred to as ∆OAs), and its efficiency varies over a wide range (10-fold). Insertion is analogous to the normal integration process and generates ∆OP' on the left and POA' on the right of the prophage (20).

A secondary attachment site occurs within *guaB*, the structural gene for IMP dehydrogenase, and accounts for 0.025% of the lysogens isolated from an *attB*-deleted strain (17). This site was located by restriction mapping approximately 570 bp from an *EcoRI* site in *guaB* and was shown subsequently to lie within the promoter proximal region of the gene (19) (Fig. 1). DNA sequence analysis of the promoter region of *guaB* (20) revealed a potential secondary attachment site for λ. The site is on the sense (transcribed) strand a few base pairs downstream of the translation initiation codon for IMP dehydrogenase and shows conservation of nine core nucleotides. To confirm this as the site of prophage insertion into *guaB*, we sequenced across the left (∆OP) and right (∆OA') prophage *att* sites and compared these sequences with that of the presumptive secondary attachment site (∆OA') in *guaB*.

DNA fragments that would include prophage ∆OP' and ∆POA' were identified by using the published sequences for DNA flanking *attP* (16) and the presumptive ∆OA' in *guaB* (20). Appropriate restriction fragments were then isolated from two transducing phages, λ *pguaA504* and λ *pguaOS04* (17, 19), and cloned into M13mp8 and M13mp9 (Fig. 1). Both strands of the λ *pguaA PvuII-BamHI* and the λ *pguoO HindIII-TaqI* DNA fragments were sequenced by using the dideoxynucleoside triphosphate chain termination procedure (15). Inspection of these sequences (Fig. 2) reveals that a crossover between bacterial and phage DNAs occurs within the predicted secondary attachment site for λ in *guaB*. The remaining nucleotide sequences of both phage and bacterial portions of the restriction fragments (data not shown) agree with the published sequences of λ (16) and the *gua* promoter region (20). It is evident that the point of crossover on insertion of λ into *guaB* is between nucleotides −3 and +2 of the att△OA' core sequence.

The nucleotide sequences of several bacterial secondary attachment sites are shown in Fig. 3. Each sequence displays homology to the common core region, and insertional crossover usually occurs in a block of conserved nucleotides that contains three T residues at positions −2, −1, and 0 (exceptionally, as in *galT*, crossover occurs between positions +4 and +5). The region of central core homology is present in *guaB* and is located where recombinational crossover occurs.

During site-specific recombination between *attP* and *attB*, the DNA strands are cut at positions −3/−2 in the top strand and +4/+5 in the bottom strand to generate 7-bp overlap regions (2) in the recombinant *att* sites. Both breaking and rejoining reactions are mediated by Int acting as a topoisomerase. Recombination is postulated to proceed through the formation of a Holliday structure intermediate by reciprocal single-stranded exchange, which is resolved, after branch migration and isomerization, to the prophage *att* sites (6). This mechanism, involving staggered cuts in DNA, requires a region of uninterrupted homology between the two *att* sites. Indeed, homology over the 7-bp overlap region (or between *attP* and at least one strand of *attB*) is necessary for efficient recombination (1). This homology is clearly not present with bacterial secondary attachment sites (Fig. 3), nor do the sequences of the corresponding prophage *att* sites indicate staggered cuts, but rather can be interpreted as
FIG. 1. Schematic representation of DNA regions flanking the attachment sites of λ pgua transducing phages (17, 19). Restriction fragments are shown (sizes in base pairs) carrying hybrid (prophage) att sites. Single lines, Phage DNA; double lines, bacterial DNA. P1 is the promoter for the gua operon, and the arrow indicates the direction of transcription (20). gua promoter and promoter-proximal portion of the guaB structural gene. The numbers above the BamHI and HindIII restriction sites refer to the positions of the sites in the complete nucleotide sequence of the λ genome (16). λ pguaA DNA was digested simultaneously with BamHI and EcoRI, and the 0.8-kilobase-pair BamHI-EcoRI fragment was separated by electrophoresis in low-melting-temperature agarose. This fragment was then treated with PvuII and ligated without further purification into M13mp8 and M13mp9 which had been cut previously with BamHI and SmaI. λ pguaO DNA was digested simultaneously with HindIII and SstII, and the 307-bp HindIII-SstII fragment was separated by electrophoresis in a 7.5% polyacrylamide gel. This fragment was treated with TaqI and ligated without further purification into HindIII-Accl-digested M13mp8 and M13mp9. The methods used have been described previously (19, 20).

resulting from simple reciprocal crossovers. An alternative mechanism for prophage insertion at secondary sites is that flush cuts are made at either the −3/−2 or the +4/+5 internucleotide bonds of both recombining partners. This need not be a concerted reaction but could proceed via a Holliday intermediate in which branch migration is impeded by heterology between the overlap regions. Exchange of the second strand would then occur at or near the position of −7 0 +7

FIG. 2. Comparison of DNA sequences within and adjacent to the hybrid attachment site core regions with those of attP (16) and the secondary attachment site within guaB (20). The 15-bp core sequences are numbered −7 to +7. Nucleotides in boldface type are from attΔ, those in lowercase type are of uncertain origin, and the remainder are from attP. Numbers below the sequence for the recombinant phage λ pguaOBA (17, 19, 20) refer to residues in the amino acid sequence of IMP dehydrogenase and are placed below the middle nucleotide of their corresponding codons.

FIG. 3. Comparison of secondary attachment site sequences. The secondary att sites are arranged in order of decreasing homology with the 15-bp core sequence of attP and attB. Insertional crossover occurs between adjacent nucleotides within the sequences underlined. Nucleotides identical to those in the consensus sequence for the Int junction-type binding site are shown in boldface type. Secondary site sequence data (from guaB) are from a previous study (22).

first-strand exchange, being promoted by an enzyme analogous to either T4 endonuclease VII (8) or T7 endonuclease I (4). A mechanism involving DNA repair synthesis within the heteroduplex overlap region is also conceivable (10), although Int-dependent recombination between wild-type attachment sites does not depend on degradation or synthesis of DNA nor does it require any high-energy cofactors.

Pinkham et al. (13) noted that the degree of core homology of secondary attachment sites does not correlate with the ability of the sites to promote phage insertion. Correctly spaced junction-type recognition sequences for Int are also necessary for phage insertion (14). These sequences (consensus CAACCTNNT) occur as imperfect inverted repeats across the core-arm junctions of both attP and attB. Homologous regions are found in most sequenced secondary att sites (22), including guaB, in which the putative left junction-type sequence has a five-of-seven match to the consensus (Fig. 3). The cut site for Int at the left junction-type sequence falls within the region of recombinational crossover for the guaB secondary att site.

Tiedeman and Smith (21) have suggested an alternative location for the guaB secondary att site, running from nucleotide 278 of their published sequence. This region, however, conserves only six core nucleotides and has poor homology to both left and right junction-type Int recognition sites. Moreover, the proposed att sequence is on the antisense (nontranscribed) strand, so that phage integration would have to occur with opposite orientation to that determined experimentally by prophage deletion mapping as guaA-guaBΔΔP′-cl-POA′-guaB′-guaO′ (17).

The location of the secondary att site at the extreme promoter-proximal end of guaB (Fig. 1 and 2) explains why λ pguaA is unable to transduce to prototrophy a set of guaB strains carrying point mutations and why the frequency of transduction of these mutants by λ pguaA depends on the map position of the guaB mutation (17, 19).

We thank K. Shimada for providing strains of phage λ.

We are grateful to the Science and Engineering Research Council for project grants GR/B8582 and GR/C64240.