Sequence of a secondary phage λ attachment site located between the pentitol operons of *Klebsiella aerogenes*

Thérèse LOVINY, Michael S. NEUBERGER* and Brian S. HARTLEY

Department of Biochemistry, Imperial College of Science and Technology, London SW7 2AZ, U.K.

(Received 16 September 1980/Accepted 2 October 1980)

We have determined the nucleotide sequence of a secondary phage λ attachment site (*att*) located between the structural genes of the ribitol and D-arabitol catabolic operons of *Klebsiella aerogenes*. The core region of this secondary attachment site (sequence: GGT TTT TTT CGATTAT) shows considerable homology with the 15-base-pair core region common to both the phage *att* and the primary bacterial *att* of *Escherichia coli* K12 (sequence: GCT TTT TTT ATA CTAA); however, there is no such clear homology between the sequences flanking the cores of the primary *att* and this secondary *att*. Integration of phage λ into the *K. aerogenes* secondary *att* occurred by recombination between the core region of the phage *att* and an oligo(T·A) stretch located within the *K. aerogenes* secondary *att*.

The integration of bacteriophage λ into the chromosome of its normal host, *Escherichia coli*, is one of the best studied systems of site-specific recombination. Linear insertion of the phage DNA is achieved by a single reciprocal recombination between a site on the λ genome and a site on the *E. coli* chromosome (Campbell, 1962). The regions on the phage and bacterial genomes in which the recombination occurs are termed *att* sites. The phage attachment site, *att*POP', consists of a 15-base-pair core region, O, and two flanking regions, P and P'. The bacterial site, *att*BOB', contains the same core region, but flanked by different sequences, B and B'. The four flanking sequences P, P', B and B' are all structurally non-homologous and functionally distinct (Guerrini, 1969). Phage integration is catalysed by a product of the λ *int* gene, although at least two host-coded proteins are also required (Miller & Friedman, 1977; Williams *et al.*, 1977).

Phage λ will integrate into the chromosome of strains in which the primary bacterial attachment site (*att*BOB') has been deleted, albeit at a much decreased frequency (Shimada *et al.*, 1972). This low-frequency integration into *att*BOB'-deleted

strains is *int*-dependent and occurs by recombination between the phage *att*POP' site and one of several sites on the *E. coli* chromosome. These secondary *atts* are designated *att*ΔO*Δ' by analogy with the primary bacterial *att*. It has previously been shown (Shimada *et al.*, 1973) that, when λ integrates into a secondary site located within a gene, the gene function is destroyed. Excision of the secondary site prophage requires expression of the phage *int* and *xis* genes and this excision is precise, as witnessed by the restoration of gene function.

The DNA sequences of *att*POP' and *att*BOB' have already been determined (Landy & Ross, 1977). A knowledge of the sequences of several secondary attachment sites will yield information about the specificity of the site-specific recombination. Here we present the sequence of a secondary *att* located between the structural genes of the pentitol operons of *Klebsiella aerogenes*.

Materials and methods

Nomenclature

Restriction enzymes are described by the abbreviations suggested by Smith & Nathans (1973) with the prefix *endo*R omitted. Sequences are written in the sense 5' to 3', left to right. For reasons of clarity, hyphens between deoxyribonucleotides have been omitted as well as their prefix d.

DNA

The isolation and characterization of specialized λ transducing phages λ prbt (full genotype:

Abbreviations used: *att*, attachment site of bacteriophage λ ; POP' and BOB', structural components of λ attachment sites on phage and bacterium, respectively; ΔO*Δ', secondary λ attachment site on the bacterial chromosome; O and O*, core regions of the primary and secondary attachment sites, respectively.

* Present address: M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

$\lambda c1857$ S7 $nin5$ $srI\lambda 4^\circ$ $srI\lambda 5^\circ$ $rbt-101$ $rbtD^+$ $rbtK^+$ and $\lambda prbt\ dal$ (full genotype: $\lambda c1857$ S7 $nin5$ $srI\lambda 4^\circ$ $srI\lambda 5^\circ$ $rbt-101$ $rbtD^+$ $rbtK^+$ dal^+) have been described previously (Neuberger & Hartley, 1979). Phage DNA was extracted with sodium dodecyl sulphate from phage purified by centrifugation in CsCl (Miller, 1972). The DNA was then extracted twice with phenol, precipitated with ethanol and stored in portions at -20°C as dry pellets. Plasmid DNA was purified by density-gradient centrifugation in CsCl/ethidium bromide (Clewell, 1972).

Enzymes

Restriction endonuclease *Hinf*I was purified by the method described by Roberts *et al.* (1975) for *Hae*III. *Bst*I was a gift from Dr. C. M. Clarke (Imperial College, London, U.K.) and *Hind*III was obtained from the Centre for Applied Microbiology and Research, Porton Down, Wilts., U.K. Purified alkaline phosphatase from *E. coli* was a gift from Dr. D. M. Glover (Imperial College, London, U.K.); phage T4 polynucleotide kinase was prepared essentially as described by Richardson (1971) except that the active fractions from the last chromatography step were concentrated by pressure dialysis. The Klenow fragment of *E. coli* DNA polymerase I was obtained from Boehringer, Mannheim, Germany.

Isolation of ^{32}P -end-labelled restriction fragments

Plasmid pRD101 was used as the source of DNA for the determination of the $\lambda prbt\ att\Delta(O^*/O)P'$ sequence. This plasmid contains two *Bst*I fragments of $\lambda prbt$ (*Bst*I $\lambda 1$ -*Bst*I λ and *Bst*I λ -*Bst*I $\lambda 3$; see Fig. 2) cloned into the *Bst*I site of plasmid pBR313 [C. L. Koh, unpublished work; Bolivar *et al.* (1977) for pBR313]. ($\lambda prbt$ DNA would, however, have served equally well as a DNA source, but pRD101 DNA was readily available.) pRD101 DNA (60 μg) was digested to completion with *Bst*I, the 5'-ends were labelled with [γ - ^{32}P]ATP (sp. radioactivity >2000 Ci/mmol; The Radiochemical Centre, Amersham, Bucks, U.K.) by using alkaline phosphatase and polynucleotide kinase as described by McReynolds *et al.* (1978). The 5'- ^{32}P -labelled DNA was then digested with *Hinf*I and the products were separated by electrophoresis in a 4% (w/v) non-denaturing polyacrylamide gel by using the Tris/borate buffer system described by Peacock & Dingman (1968). The size of the fragment containing the $att\Delta(O^*/O)P'$ sequence was known by a previous restriction mapping study of the whole plasmid DNA. This fragment was located on the gel by autoradiography and was eluted electrophoretically in 40 mM-Tris/acetate (pH 8.2)/0.1 mM-EDTA, as described by Galibert *et al.* (1974).

For the determination of the $att\Delta O^*\Delta'$ sequence, 180 μg of $\lambda prbt\ dal$ DNA was digested with *Bst*I and the fragments were separated by electrophoresis in a 0.8% (w/v) agarose/ethidium bromide gel in Tris/borate buffer. The section of the gel containing the

Fig. 1. Model for the formation of specialized transducing phages $\lambda prbt$ and $\lambda prbt\ dal$.

Phage $\lambda 627$ (a deletion derivative of $\lambda c1857$ S7) was infected into the *E. coli* K12/K. *aerogenes* hybrid strain NC596 [HfrH(*gal-att ψ -bio*)*thi rbt-101 rbtD⁺ rbtK⁺ dal⁺*] and a population of secondary site lysogens induced for phage production; $\lambda prbt$ was isolated from this phage lysate. It is envisaged that, in the secondary site lysogen which gave rise to $\lambda prbt$, phage $\lambda 627$ had integrated into the NC506 chromosome by *int*-mediated site-specific recombination at a secondary *att* located between *rbtD* and *dalD*. Phage $\lambda prbt\ dal$ was isolated from a phage lysate induced from a population of NC596 ($\lambda prbt$) lysogens and it is envisaged that, in the lysogen which gave rise to $\lambda prbt\ dal$, phage $\lambda prbt$ had integrated into the NC596 chromosome in the region of the pentitol operons by *rec*-mediated homologous recombination. The evidence for this model is presented in Neuberger & Hartley (1979).

*Bst*IA-*Bst*IB fragment was visualized with u.v. illumination, and the DNA was eluted from the gel electrophoretically. The isolated DNA fragment was then labelled either (a) at the 5'-ends by using [γ - 32 P]ATP as described above or (b) at the 3'-ends with [α - 32 P]dATP (sp. radioactivity 2000–3000 Ci/mmol; The Radiochemical Centre) and the Klenow fragment of *E. coli* DNA polymerase I as described by Schwartz *et al.* (1978). The labelled DNA from both procedures was digested with *Hind*III, fractionated by electrophoresis in a 4% (w/v) non-denaturing polyacrylamide gel, and the relevant fragment of known size was located by autoradiography and eluted electrophoretically as described above.

DNA sequence determination

Sequences were elucidated by the chemical degradation method of Maxam & Gilbert (1977) with the alternative strong adenine/weak cytosine cleavage, and fractionation in 8 or 12% (w/v) polyacrylamide thin (0.3 mm $\times$ 20 cm $\times$ 40 cm) gels containing 7M-urea (Sanger & Coulson, 1978). The sequences at the labelled ends of 5'- or 3'-labelled DNA fragments could be obtained up to the penultimate base by running the four chemical degradation products on a thicker (2 mm $\times$ 20 cm $\times$ 40 cm) 25% (w/v) polyacrylamide gel containing 7M-urea for no more than 2 h at 30 mA (Dr. G. Brownlee, personal communication).

Results

Strategy

We have previously described the transfer of the

K. aerogenes pentitol operons into *E. coli* K12 by transduction with phage P1 (Rigby *et al.*, 1976) and the subsequent isolation of specialised transducing phages carrying the *K. aerogenes* ribitol (*rbt*) and D-arabitol (*dal*) catabolic operons (Neuberger & Hartley, 1979). The primary transducing phage, λ *prbt*, was isolated by using the secondary-site-attachment technique described by Schrenk & Weisberg (1975) and selection was performed for lysogenation of phage into a site near the *rbt* operon. λ *prbt dal* is a secondary transducing phage which was constructed by *rec*-mediated recombination between λ *prbt* and the chromosomal pentitol operon genes.

The structure of phage λ *prbt* indicates that it is a 'gal-type' transducing phage formed by abnormal excision of a prophage which had integrated into a secondary *att* located between the *rbt* and *dal* structural genes (Neuberger & Hartley, 1979). This follows the classical model for the generation of specialised transducing phages (Campbell, 1962) and is illustrated in Fig. 1. It will be seen that the λ *prbt att* site is predicted to be of the form $\Delta(O^*/O)P'$, indicating that it is expected to contain a core region which is a hybrid between the wild type (O) and that of the secondary *att*(O*) with Δ and P' flanking regions. Thus by sequencing only the λ *prbt att* site it is not possible to deduce the full sequence of the secondary attachment site at which λ integrated prior to λ *prbt* formation. However, this information can be gleaned by DNA sequence analysis of λ *prbt dal*. As we have shown previously (Neuberger & Hartley, 1979), the structure of phage λ *prbt dal* suggests that it was formed by abnormal excision of a λ *prbt* prophage that had integrated into

Fig. 2. Restriction maps of λ *prbt* and λ *prbt dal* DNAs for *Bst*I.

The left- and right-hand ends of the DNAs are designated *m* and *m'* respectively and the positions of some λ genes are indicated. The *Bst*I sites on the λ DNA parts of the transducing phages are numbered according to the corresponding sites on wild-type λ DNA (Haggerty & Schleif, 1976). The restriction fragments containing the *att* sequences are shown on an enlarged scale and the distances between restriction sites are indicated in base pairs. The thick arrows under *Bst*A show the start points and the direction of the DNA sequence determinations.

Fig. 3. Sequencing gels of the *λprbt* and *λprbt dal* attachment sites

DNA sequencing was performed by the chemical degradation method of Maxam & Gilbert (1977). The lanes on the gels are marked T + C, C, A > C and G, to indicate the nature of the base-specific strand scission in the respective samples. (a) Autoradiograph of a sequencing gel showing part of the sequence obtained for the attachment site *λprbt attΔ(O*/O)P'*: 60 μg of plasmid pRD101 DNA was digested, labelled at the 5'-end and the labelled fragment was submitted to the chemical degradation as described in the Materials and methods section. Each specific degradation mixture (2000 c.p.m.) was loaded onto a 12% (w/v) 7M-urea thin polyacrylamide gel and subjected to electrophoresis at 30 mA; after 2 h, a second loading of the same samples was applied to the right part of the gel and the run continued at 30 mA for a further 2 h. The gel was then taken off the glass plates, covered with Saran wrap and radioautographed on a Kodak XH-1 film for 6 days, with the enhanced detection technique of Laskey & Mills (1977). The core sequence (O*/O) only is presented along the radioautogram. (b) Autoradiograph of a sequencing gel showing part of the sequence obtained for the secondary attachment site *λprbt dal attΔO*Δ'*: 180 μg of *λprbt dal* DNA was digested, labelled at the 3'-end and the labelled fragment was submitted to the chemical degradation as described in the Materials and methods section, thus showing the complementary sequence to the *λprbt attO** sequence (a). The same gel concentration and conditions of electrophoresis were used as in (a), except the running time was 5 h. The pseudo-core sequence O* only is presented.

Fig. 4. Extent of the sequences determined of (a) λ prbt and (b) λ prbdal attachment regions. The boxed sections include the core regions. Following the numbering system of Landy & Ross (1977), the central base of the core region was chosen as the number 0, with negative numbering on its left and positive on its right. In the λ prbdalatt sequence (b), palindromes are indicated by (▶) and (◀) under the corresponding sequences and two sets of different direct repeats are marked under the corresponding sequences by (—) and (----) respectively.

the bacterial chromosome by *rec*-mediated homologous recombination. Phage λ *prbt dal* thus contains the entire secondary attachment site region, *att* Δ O* Δ' (Fig. 1). Thus by sequencing relevant parts of λ *prbt* and λ *prbt dal* it should be possible to determine both the sequence of the secondary attachment site and the position at which cross-over occurred when λ integrated into this site.

Identification of the *att*-containing restriction fragments

The cleavage maps of phages λ *prbt* and λ *prbt dal* for restriction endonuclease *Bst*I are presented in Fig. 2. The *rbt* substitution on λ *prbt* contains a single *Bst*I site (*Bst*IA), the right-hand junction between the bacterial substitution and λ DNA being contained within a 325-base-pair *Bst*I fragment, *Bst*IA–*Bst*II3. As the core region of the wild type λ *att*POP' site is known to be located at about 260 base pairs to the left of *Bst*II3 (Davies *et al.*, 1977) the λ *prbt att* sequence can be easily determined by sequencing rightwards from *Bst*IA on λ *prbt*. Similarly, the *att* Δ O* Δ' sequence can be determined by sequencing rightwards from *Bst*IA on λ *prbt dal*.

DNA sequence of *att* Δ (O*/O)P' and *att* Δ O* Δ'

The sequences of the *att* sites *att* Δ (O*/O)P' and *att* Δ O* Δ' were determined as described in the Materials and methods section. Two of the gels from which some of the data were obtained are shown in Fig. 3 and the full extent of the sequence determined is presented in Fig. 4. The sequence that we have determined for *att*P' agrees entirely with that previously published by Landy & Ross (1977). The

30-nucleotide-long regions on each side of the pseudocore O* are 70% G + C rich, the core itself being A + T rich. Comparison of the flanking regions of the *rbt dal* secondary *att* with the flanking regions of the primary and secondary bacterial *atts* already sequenced (Landy & Ross, 1977; Christie & Platt, 1979; Csordás-Tóth *et al.*, 1979; Bidwell & Landy, 1979) reveals that only the *galT att* resembles the *rbt dal att* in this high G + C content, although to a lesser extent. The sequencing gel of *att* Δ (O*/O)P' (Fig. 3a) presents an anomaly at nucleotide –37 in the Δ arm, since the position on the 'ladder' is occupied by T and G simultaneously; the base was nevertheless confirmed to be a G by several independent determinations of the same Δ sequence and also by determining the sequence of the complementary strand. Similarly, several independent determinations of the Δ' sequence of *att* Δ O* Δ' have shown the +42 base to be a T, although the sequencing gel of Fig. 3(b) presents the four bases simultaneously at this position.

Discussion

Although there is a clear homology between the core region of the secondary *att*(O*) (Fig. 4) and that of the primary attachment site, no such homology is apparent between Δ and B or Δ' and B' [see Landy & Ross (1977) for the sequences of O, B and B']. Similar results have been obtained for the secondary sites located within *trpC* (Christie & Platt, 1979), *bfe* (Csordás-Tóth *et al.*, 1979) and *galT* (Bidwell & Landy, 1979). It has been previously calculated (Shimada *et al.*, 1975) that, from

Fig. 5. Comparison of the core regions of various attachment sites

The wild type common core sequence is taken from Landy & Ross (1977) and the *trpC*, *bfe* and *galT* core sequences are from Christie & Platt (1979), Csordás-Tóth *et al.* (1979) and Bidwell & Landy (1979) respectively. The arrows indicate the positions at which cross-over must have taken place on integrative recombination of phage λ . Bases identical with the corresponding ones in the primary attachment core are underlined. The sequence of only one of the DNA strands is presented for each attachment site.

the observed number of secondary attachment sites on the *E. coli* chromosome, the recognition elements of the secondary atts consist, on average, of five or six base pairs. Taken in conjunction with the observation that all the four secondary attachment sites which have been sequenced so far show homology only to the core region of the primary attachment site, these findings suggest that the pseudo-core regions of the secondary attachment sites provide the sole recognition element for int-dependent integration. Similarly, it would be inappropriate to attribute any evolutionary significance to the presence of a secondary attachment site in *K. aerogenes*, as the limited stringency of the base sequence requirement for a secondary attachment site suggests that such sequences would occur several times on all bacterial genomes.

Comparison of the core regions of the primary attachment site and the *galT*, *trpC*, *bfe* and *rbt-dal* secondary sites (Fig. 5) shows that they all share the trinucleotide sequence (T)₃ at positions -2 to 0 and the A + T-rich nature of these sequences suggests that they would be relatively easily denatured.

As the core regions of the phage and primary bacterial attachment sites are identical, it is not possible to determine whether, on prophage integration, the recombinational cross-over always takes place at a specific site, or randomly within the core. However, as the core region of the *rbt-dal* secondary att differs from that of the phage, it is possible to state from examination of the λ prbt core sequence that λ recombined into the *rbt-dal* secondary att by a cross-over somewhere within the oligo(T · A) region located at positions -5 to 0. Similarly, integration into *trpC* must have occurred between positions -2 and +3 (Christie & Platt, 1979) and into *bfe* at between positions -3 and +1 (Csordás-Tóth *et al.*, 1979), whereas integration into *galT* occurred at between +4 and +5 (Bidwell & Landy, 1979). It would certainly be of interest to determine whether, in independent lysogens with the prophage integrated at the same secondary att, λ had integrated by cross-over at an identical position within the core. It would also be of interest to correlate the sequences of secondary attachment sites with the efficacy with which λ integrates into them. However, at present, whereas λ integration into secondary atts may not truly reflect normal integration, there is certainly no reason to assume that there is a highly preferred region within the the core for cross-over on integrative recombination.

We thank the Science Research Council for support. One of us (T. L.) has pleasure in thanking Dr. George Brownlee (Laboratory of Molecular Biology, Cambridge, U.K.) for demonstrating methods of DNA sequencing and we are very grateful to Mr. Chong-Lek Koh for plasmid pRD101.

References

- Bidwell, K. & Landy, A. (1979) *Cell* **16**, 397–406.
- Bolivar, Fr., Rodriguez, R. L., Betlach, M. C. & Boyer, H. W. (1977) *Gene* **2**, 75–93.
- Campbell, A. M. (1962) *Adv. Genet.* **11**, 101–145.
- Christie, G. E. & Platt, T. (1979) *Cell* **16**, 407–413.
- Clewell, D. B. (1972) *J. Bacteriol.* **110**, 667–676.
- Csordás-Tóth, E., Boros, I. & Venetianer, P. (1979) *Nucleic Acids Res.* **7**, 1335–1341.
- Davies, R. W., Schreier, P. H. & Büchel, D. E. (1977) *Nature (London)* **270**, 757–760.
- Galibert, F., Sedat, J. & Ziff, E. (1974) *J. Mol. Biol.* **87**, 377–407.
- Guerrini, F. (1969) *J. Mol. Biol.* **46**, 523–542.
- Haggerty, D. M. & Schleif, R. F. (1976) *J. Virol.* **18**, 659–663.
- Landy, A. & Ross, W. (1977) *Science* **197**, 1147–1160.
- Laskey, R. A. & Mills, A. D. (1977) *FEBS Lett.* **82**, 314–316.
- Maxam, A. M. & Gilbert, G. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560–564.
- McReynolds, L., O'Malley, B. W., Nisbet, A. D., Fothergill, J. E., Givol, D., Fields, S., Robertson, M. & Brownlee, G. G. (1978) *Nature (London)* **273**, 723–728.
- Miller, H. I. & Friedman, D. I. (1977) in *DNA Insertion Elements, Plasmids and Episomes* (Bukhari, A. I., Shapiro, J. A. & Adhya, S. L., eds.), pp. 349–356, Cold Spring Harbor Laboratory, New York.
- Miller, J. H. (1972) in *Experiments in Molecular Genetics*, pp. 331–334. Cold Spring Harbor Laboratory, New York.
- Neuberger, M. S. & Hartley, B. S. (1979) *J. Mol. Biol.* **132**, 435–470.
- Peacock, A. C. & Dingman, C. W. (1968) *Biochemistry* **7**, 668–674.
- Richardson, C. C. (1971) in *Procedures in Nucleic Acid Research* (Cantoni, G. L. & Davies, D. R., eds.) vol. 2, pp. 815–828, Harper and Row, New York.
- Rigby, P. W. J., Gething, M. J. & Hartley, B. S. (1976) *J. Bacteriol.* **125**, 728–738.
- Roberts, R. J., Breitmeyer, J. B., Tabachnik, N. F. & Myers, P. A. (1975) *J. Mol. Biol.* **91**, 121–123.
- Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
- Schrenk, W. J. & Weisberg, R. A. (1975) *Mol. Gen. Genet.* **137**, 101–107.
- Schwartz, E., Scherer, G., Hobom, G. & Kössel, H. (1978) *Nature (London)* **272**, 410–414.
- Shimada, K., Weisberg, R. A. & Gottesman, M. E. (1972) *J. Mol. Biol.* **63**, 483–503.
- Shimada, K., Weisberg, R. A. & Gottesman, M. E. (1973) *J. Mol. Biol.* **80**, 297–314.
- Shimada, K., Weisberg, R. A. & Gottesman, M. E. (1975) *J. Mol. Biol.* **93**, 415–429.
- Smith, H. O. & Nathans, D. (1973) *J. Mol. Biol.* **81**, 419–423.
- Williams, J. G. K., Wulff, D. L. & Nash, H. A. (1977) in *DNA Insertion Elements, Plasmids and Episomes* (Bukhari, A. I., Shapiro, J. A. & Adhya, S. L., eds.), pp. 357–361, Cold Spring Harbor Laboratory, New York.