JOURNAL OF BACTERIOLOGY, Aug. 1989, p. 4232–4240
0021-9193/89/084232-09$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 171, No. 8

Nucleotide Sequence and Expression of the Gene for the Site-Specific Integration Protein from Bacteriophage HP1 of *Haemophilus influenzae*

STEVEN D. GOODMAN† AND JOHN J. SCOCCA*

Department of Biochemistry, The Johns Hopkins University School of Hygiene and Public Health, 615 North Wolfe Street, Baltimore, Maryland 21205

Received 27 February 1989/Accepted 7 May 1989

The nucleotide sequence of the leftmost 2,363 base pairs of the HP1 genome, which includes the attachment site (*attP*) and the integration region, was determined. This sequence contained an open reading frame encoding a 337-residue polypeptide, which is a member of the integrase family of site-specific recombination proteins as judged by sequence comparison. The open reading frame was located immediately adjacent to the *att* site and was oriented so that initiation of translation would begin distal to the *att* site and end in its immediate vicinity. Expression of this DNA segment in *Escherichia coli* provided extracts which promoted site-specific recombination between plasmids containing cloned HP1 *attP* and *Haemophilus influenzae attB* sites. This recombination was directional, since no reaction was observed between plasmids containing *attR* and *attL* sites. The reaction was stimulated by the accessory protein integration host factor of *E. coli*. Evidence was also obtained that the integration host factor influenced the levels of HP1 integrase expression. The deduced amino acid sequence of HP1 integrase has remarkable similarity to that deduced for the integrase of coliphage 186.

Bacteriophage HP1 is a temperate phage which infects and lysogenizes *Haemophilus influenzae* Rd (8). In lysogenic strains, the prophage genome is inserted into the *H. influenzae* chromosome at a single locus (40) as a consequence of a conservative recombination reaction between specific phage (*attP*) and host (*attB*) chromosomal sites (41). The HP1 insertion system is distinguished by the large size of the core, which is common to the *attB* and *attP* regions. This core contains 182 residues and consists of two blocks of 93 and 62 residues identical in both phage and host sites which are separated by a 27-base-pair (bp) block containing 6 mismatched residues; recombination occurs in the 62-bp block (41). HP1 belongs to a diverse group of lysogenic phages and integrating plasmids which use tRNA genes as the targets for insertion into the host genome (30, 32). The common core of the HP1 *att* sites contains a complete lysine tRNA gene sequence in the 93-bp segment and a sequence corresponding to the 3′ end of a leucine tRNA gene in the 62-bp crossover block (32).

The genome structure of HP1 lysogens resembles that of λ and other temperate coliphages, making it likely that insertion of the HP1 genome into the cognate host site results from a similar recombination process. The site-specific integration of the λ genome into the *Escherichia coli* chromosome has been analyzed in detail (for reviews, see references 2, 26, 36, and 42). Two proteins are required for λ integration, the phage-encoded recombinase Int (13) and the host DNA-binding protein integration host factor (IHF) (13, 21). Int is a member of the integrase family of site-specific recombinases, which has been defined on the basis of sequence similarities and functional relationships (1).

Studies on plasmids containing the HP1 *attP* and *attB* sites showed that specific intermolecular recombination between these sites occurred in *E. coli*; recombination depended on the presence of a 1.5-kilobase (kb) phage-derived DNA segment located immediately adjacent to the right end of the *attP* site. Efficient recombination required the presence of *E. coli* IHF (2). We hypothesized that the 1.5-kb HP1 segment contained the structural gene for a protein which activated site-specific recombination between the *att* sites. The location of the structural gene for the recombination protein immediately adjacent to the *attP* site is a common feature of all similar systems examined to date (15). To validate this hypothesis, we determined the nucleotide sequence of the segment. The sequence contained a long open reading frame (ORF) encoding a polypeptide product of 38.6 kilodaltons; the deduced sequence was clearly related to the sequences of other phage integrases (1, 28). (The sequence reported here has been deposited in the GenBank data base [accession number M22941].) To establish the role of this presumptive HP1 integrase in integrative recombination, we inserted a restriction fragment containing the complete ORF into an expression vector system. Extracts from induced cultures containing this plasmid supported site-specific recombination of HP1 *attP* and *attB* sites in vitro; this recombination was stimulated by the presence of IHF. Further, IHF appeared to participate in regulating the levels of active HP1 integrase in *E. coli*. The deduced amino acid sequence of HP1 integrase is remarkably similar to that of the coliphage 186 integrase (12).

MATERIALS AND METHODS

Materials and analytical methods. NACS Prepac cartridges and low-gelation-temperature agarose were from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.). *Mae*I was obtained from Boehringer Mannheim Biochemicals (Indianapolis, Ind.); other restriction enzymes and DNA-modifying enzymes were purchased from either New England BioLabs, Inc. (Beverly, Mass.) or Bethesda Research Laboratories and were used under the conditions recommended by the supplier. IHF (fraction V) was purified from *E. coli* HB101 essentially as described by Nash and Robertson (27)

* Corresponding author.
† Present address: Laboratory of Molecular Biology, National Institute of Mental Health, Bethesda, MD 20892.

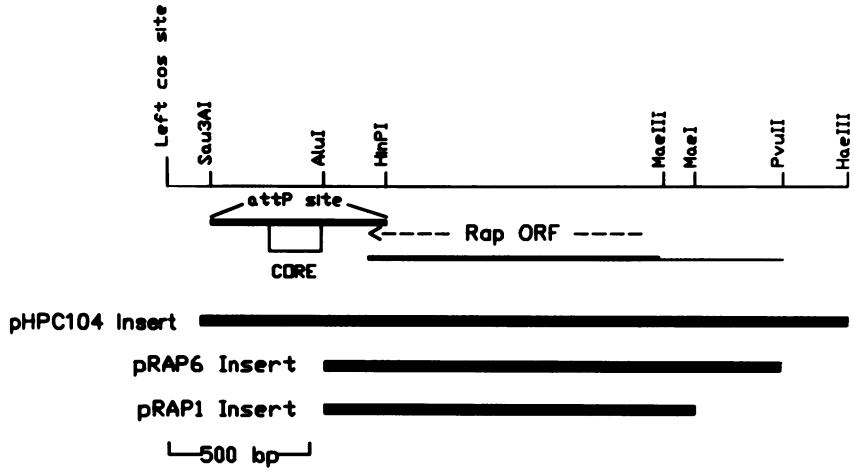

FIG. 1. Integration region of phage HP1. The 2,363-bp segment beginning with the left *cos* site is diagrammed; selected restriction sites are indicated. The approximate extent of the functional *attP* site and the location of the 182-bp core region are designated, as is the extent of the integrase ORF. The bars indicate the segments derived from this region which are present in the plasmids used in these studies; construction of these plasmids is described in the text. Rap, Recombination protein.

and was the generous gift of E. S. Hwang of this laboratory. DEAE-cellulose (Whatman DE52) and phosphocellulose (Whatman P11) were prepared according to the directions of the manufacturer. DNA (33) and protein (16) concentrations were determined colorimetrically.

Bacterial strains and plasmids. *E. coli* HB101 and the IHF$^-$ strain RW1815::Tn*10* (*recA hip*Δ*1*; from R. Weisberg, National Institutes of Health) were grown in solid or liquid LB medium supplemented with antibiotics as appropriate (19). Cultures were made competent and transformed by the method of Mandel and Higa (18). The plasmids pHPC104 containing the cloned HP1 *attP* site, pHPC102 containing the *H. influenzae attB* site, pHPC101 containing the *attR* site, and pHPC103 containing the *attL* site have been described previously (41). The expression vector pHE6, provided by G. Milman (The Johns Hopkins University), contains the p_L promoter, the temperature-sensitive *c*I857 repressor gene of λ, and a polylinker region situated to allow the expression of cloned DNA segments under the control of these elements (22).

Preparation and purification of DNA. Plasmids were isolated by published methods (19, 25) and purified by equilibrium centrifugation on CsCl-ethidium bromide density gradients. Ethidium bromide was removed by exhaustive extraction with saturated aqueous isobutanol. Linear recombination substrates were prepared by digestion with *Eco*RI (pHPC102 through 104) or with *Bam*HI (pHPC101) and labeled by the exchange reaction between 5′ DNA ends and [^{35}P]ATP catalyzed by T4 polynucleotide kinase (9). Specific restriction fragments were separated by electrophoresis on low-melting-point agarose gels (ultrapure grade; Bethesda Research Laboratories). Gel slices containing segments of interest were cut out, the agarose was melted by heating, and the DNA was purified with NACS cartridges.

Nucleotide sequences. The sequence of the HP1-derived segment of pHPC104 (41) (Fig. 1) was determined by chemical cleavage methods (20). Five reactions were used routinely: A+G; G; T+C; C; and A+C. Fragments were labeled at the 5′ termini by reaction with T4 polynucleotide kinase and [γ-^{32}P]ATP. Restriction sites for commercially available enzymes predicted by sequence results were verified by digestion. Both strands of duplex DNA were sequenced independently, and the sequences were determined at least twice on independent DNA preparations.

Construction of pRAP plasmids. A DNA segment containing the entire ORF of the HP1 insert was prepared by digestion of pHPC104 with *Alu*I. The products were separated by electrophoresis, and the 1,590-bp fragment was recovered and purified. It was then ligated with pHE6 which had been digested with *Pvu*II and treated with *E. coli* alkaline phosphatase. The ligation mixture was used to transform competent *E. coli* HB101; ampicillin-resistant colonies were isolated. Plasmids were purified, and the presence and orientation of the 1,590-bp insert were determined by restriction mapping. In pRAP6, the ORF was in parallel with the p_L promoter in pHE6, and in pPAR1, the insert was in the opposite orientation.

pRAP1 was constructed by digestion of pHPC104 with *Mae*I and *Alu*I. Unpaired ends were filled in by treatment with the large fragment of DNA polymerase I and deoxynucleoside triphosphates (19). The 1.3-kb *Mae*I-*Alu*I fragment containing the ORF, indicated in Fig. 1, was purified by electrophoresis, and the purified fragment was ligated with pHE6 which had been digested with *Pvu*II and dephosphorylated with *E. coli* alkaline phosphatase. The presence and orientation of the insert were established by restriction mapping.

Heat induction and preparation of extracts. *E. coli* HB101 containing an expression plasmid was grown in LB medium plus ampicillin at 30°C with shaking until the A_{600} was between 1.5 and 3.0. Sufficient prewarmed LB medium was added to give a final absorbance of 0.75 and a final temperature of 42°C, and incubation was continued at 42°C for the times indicated. Cells were collected by centrifugation, suspended in 1/15th culture volume of 10 mM Tris chloride–10 mM EDTA (pH 7.5) (TE buffer), and recentrifuged. The washed cells were resuspended in 1/30th culture volume of TE buffer at 4°C, frozen, and stored at −80°C. Frozen cells retained their original recombination activity for at least 6 months.

To prepare extracts, cell suspensions were thawed, and freshly prepared lysozyme solution (10 mg/ml in TE buffer) was added to a final concentration of 0.5 mg/ml. The suspension was incubated at 0°C for 20 to 30 min and

centrifuged for 10 min at 10,000 × *g* at 4°C, and the pellets were discarded. The supernatant solution (fraction 1) was adjusted to 50 mM Tris chloride buffer, (pH 7.5), 1 mg of bovine serum albumin per ml, 15 mM EDTA, 10% glycerol, and 2 mg of extract protein per ml. Fraction 1 could be frozen at −80°C and thawed once with no loss of activity.

Fractionation of extracts. Fraction 1 (0.6 ml) was prepared as above, except that 1 M KCl was added before centrifugation to dissociate proteins from nucleic acids. The supernatant solution was then diluted by the addition of 2.4 ml of TEG buffer (50 mM Tris chloride [pH 7.5], 15 mM EDTA, 10% [vol/vol] glycerol) and applied to a column (6-ml bed volume) of DEAE-cellulose equilibrated with TEG and eluted with 9 ml of TEG at a flow rate of 0.4 ml/min. Fractions (0.6 ml) were collected, and portions were dialyzed against TEG buffer and assayed. Recombination activity emerged in the flowthrough, while nucleic acids were retained by the column. Active fractions were pooled and applied to a phosphocellulose column (0.5-ml bed volume) which had been equilibrated with 0.2 M KCl in TEG. The column was eluted stepwise, first with 0.5 M KCl in TEG (5 ml) and then with 1.0 M KCl in TEG; fractions (1 ml) were collected. Portions of these fractions were dialyzed against TEG and assayed alone or in the presence of purified IHF (300 ng/ml).

Assay for in vitro recombination. The assay for in vitro recombination was based on an assay developed by Mizuuchi and Mizuuchi (23), in which supercoiled DNA containing an *attP* site recombines with a linear DNA molecule containing *attB* to yield a linear product equal in size to the sum of the two substrate DNA molecules. Recombination was quantitated with an end-labeled linear substrate.

Reaction mixtures (0.02 ml) contained 100 fmol each of undigested pHPC104 (donor) and linearized pHPC102 (acceptor) in 50 mM Tris chloride (pH 7.5)–15 mM EDTA–6 mM spermidine–1 mg of bovine serum albumin per ml–10% glycerol; the final concentration of NaCl (introduced with DNA solutions) was adjusted to 0.05 to 0.07 M. Fractions to be assayed were added, and incubation was at 30°C for 30 to 60 min. Reactions were stopped by the addition of proteinase K to 50 μg/ml followed by heating for 15 min at 65°C. Products and substrates were separated by electrophoresis and visualized by staining with ethidium bromide. Bands corresponding to the labeled product (11.9 kb) and labeled substrate (7 kb) were excised from the gel, and the radioactivity in each was determined. The percent recombination was calculated from the ratio of radioactivity in the product band to the total recovered radioactivity. One unit of activity is the amount of extract giving 40% recombination in 1 h at 30°C.

RESULTS

Earlier results showed that plasmids containing the HP1 *attP* and *attB* sites underwent site-specific dimerization when propagated in *E. coli* HB101. This dimerization was eliminated when the 1.5-kb HP1 *Bgl*I-*Hae*III segment immediately to the right of the *attP* site was deleted (2), suggesting that the deleted segment contained the structural gene for the HP1 integration protein. To confirm this, we determined the nucleotide sequence of this region.

Nucleotide sequence of left end of HP1 genome. The plasmid pHPC104, which contains the 2.2-kb *Bgl*II-*Hae*III fragment originating from the left end of the phage HP1 (41), was the source of DNA for this work. The sequence of the 0.8-kb *Bgl*II-*Bgl*I segment containing the *attP* site has been reported (41). The nucleotide sequence of the rest of the HP1 insert in pHPC104 was determined, and the results were combined with earlier data (6, 41) to provide the complete nucleotide sequence of the left end of the HP1 genome from the left *cos* site to the leftmost *Hae*III site. The organization of this region is presented in Fig. 1, and the DNA sequence is shown in Fig. 2.

The capacity of this sequence to code for one or more polypeptides in any of the six possible reading frames was examined. A long ORF was present in this sequence, with the initiation codon beginning at residue 1709 and terminating at residue 698. The ORF corresponds to a polypeptide of 337 amino acids, with a molecular size of 38.6 kilodaltons. The deduced amino acid sequence is shown in Fig. 3. The carboxy-terminal region of the predicted amino acid sequence contains the conserved -His-X_2-Arg- residues and an appropriately spaced tyrosyl residue which have been proposed as the catalytic segment of the integrase family of site-specific recombination proteins (1, 28). These sequence relationships show that the product of this ORF is most probably the HP1 site-specific integration protein or integrase.

Cloning and expression of the ORF. To establish the role of the predicted translation product in site-specific recombination, the DNA containing the ORF was inserted into the expression vector pHE6, and extracts of induced cultures containing these constructions were assayed for recombination in vitro. The complete ORF lies within a single 1,590-bp fragment bounded by an *Alu*I site at residue 543 and a *Pvu*II (*Alu*I) site at residue 2132; this fragment lacks the *attP* core sequence. DNA from pHPC104 was digested with *Alu*I, and the 1.6-kb fragment was isolated and inserted into the *Pvu*II site in the polylinker region of pHE6. The plasmids pRAP6 and pPAR1 were recovered from transformant clones. These plasmids differ in the orientation of the 1.6-kb segment; pRAP6 contains the coding insert in direct orientation relative to the p_L promoter, and pPAR1 contains the insert in opposed orientation.

Cultures of *E. coli* HB101(pRAP6) and HB101(pPAR1) were grown at 30°C and shifted to 42°C to inactivate the *c*I857 repressor. After various times of induction, extracts were prepared and assayed for the capacity to promote site-specific recombination as described in Materials and Methods. The time course of appearance of specific recombination activity after induction of *E. coli* HB101(pRAP6) is shown in Fig. 4. An 11.9-kb recombinant product was detectable by autoradiography with extracts prepared after 1 h of induction, and the level of product formation reached a steady-state level after 5 h. No recombinant product was detected with extracts from uninduced cultures or from induced cultures of *E. coli* HB101(pPAR1) (see Fig. 5, lane 5). The orientation dependence as well as the absence of detectable activity in extracts from uninduced cultures indicates that measurable levels of HP1 integrase accumulated only when expression was controlled by the derepressed λ p_L promoter.

Characterization of recombinant product. To confirm the site specificity of the reaction, the 11.9-kb band was isolated from several large-scale reactions and the pooled product was purified. The product was digested singly with *Pst*I, *Hin*dIII, *Pvu*II, or *Ava*I; the fragments produced in these digests had the sizes expected for site-specific recombination between the two *att* sites (data not shown). Autoradiographic analysis of reactions in which the linear substrate was labeled with ^{32}P showed that only the substrate and product bands contained measurable radioactivity and that

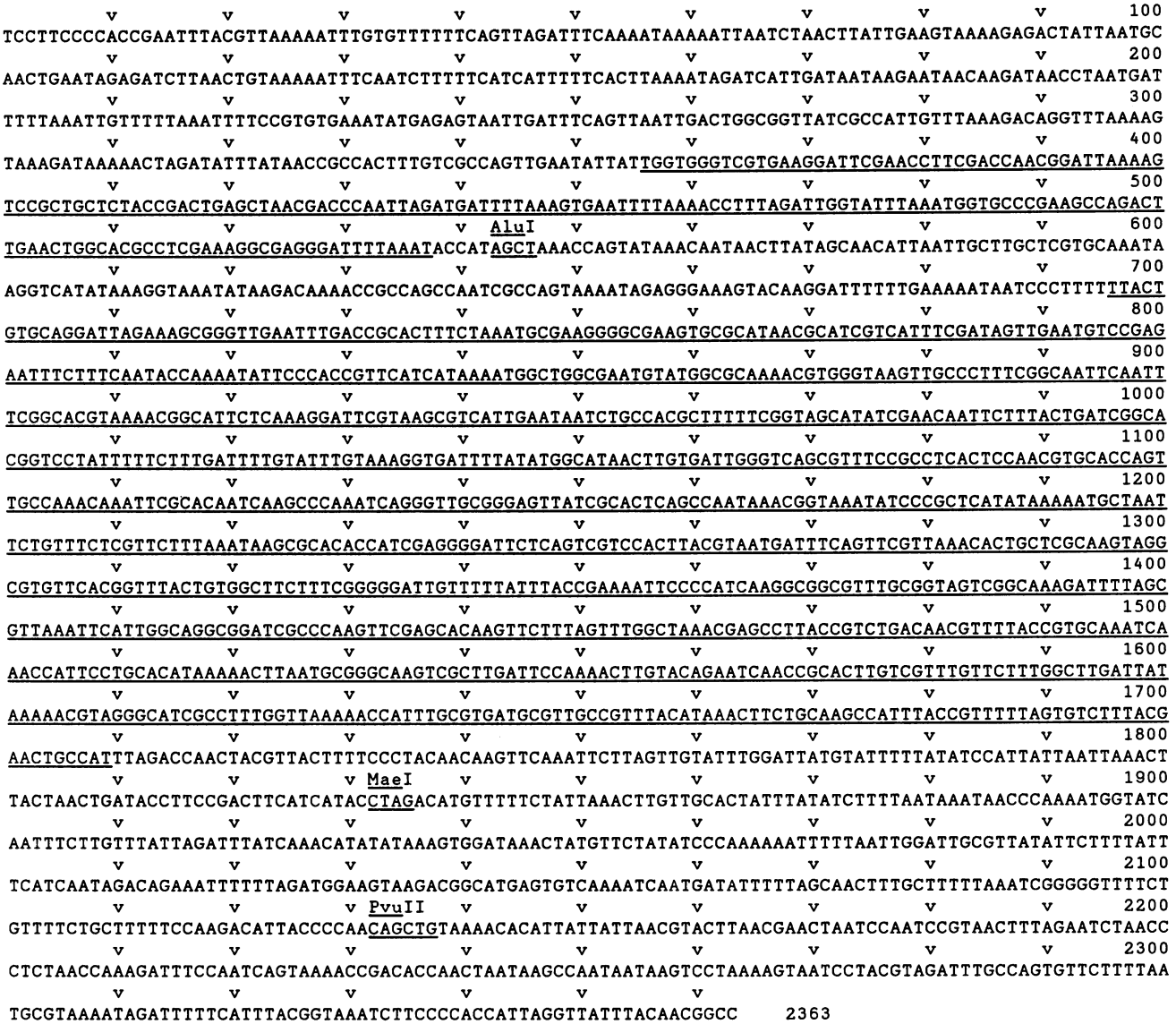

FIG. 2. Nucleotide sequence of the left end of the HP1 chromosome. Numbering begins with the first (unpaired) residue of the *cos* site and extends to the leftmost *Hae*III site. The locations of restriction sites used in the construction of pRAP6 and pRAP1 are indicated above the sequence. The *attP* core region (residues 356 to 537) and the sequence complementary to the integrase ORF are underlined.

the input radioactivity was accounted for by the sum of the radioactivity in these two bands.

Several additional properties of the recombination reaction promoted by crude extract were examined, and the results are shown in Fig. 5. Activity was heat labile; no recombination was observed with extracts heated at 60°C for 2 min. Preincubation of extracts at 45°C for 5 min also completely inactivated them (data not shown). The presence of spermidine was required for detectable recombination, and the reaction proceeded in the presence of 15 mM EDTA, indicating that divalent cations were not required. Recombination was inhibited by concentrations of NaCl or KCl above 50 mM. ATP did not stimulate the reaction (data not shown).

Substrate specificity. Studies on the in vitro integrative recombination promoted by λ Int have established that the two substrates are functionally distinct (11). The *attP* substrate contains multiple binding sites for Int (24) and for IHF (5) and is the site for the organization of the intasome complex (3). Under most conditions, this substrate must be supercoiled (23). The *attB* substrate, which may be linear or supercoiled, is captured by the *attP* intasome to produce the complete recombinational intermediate (34). Because of this functional asymmetry, the *attP* substrate is designated as the donor and the *attB* substrate as the acceptor in the integrative recombination reaction (11). Preliminary results in the HP1 system showed that when supercoiled pHPC104 was present in limited amounts, recombination ceased when the supercoiled plasmid was exhausted. Substituting nicked circular or linear forms of pHPC104 for the supercoiled plasmid yielded undetectable levels of recombination (data not shown). Therefore, under the conditions of reaction, one substrate must be supercoiled for recombination to occur.

The ability of the HP1 *attP*, *attB*, *attR*, and *attL* sites to serve as substrates for specific recombination in vitro was examined by measuring recombination of substrate pairs in

```
MetAlaValArgLysAspThrLysAsnGlyLysTrpLeuAlaGluValTyrValAsnGly      20
AsnAlaSerArgLysTrpPheLeuThrLysGlyAspAlaLeuArgPheTyrAsnGlnAla      40
LysGluGlnThrThrSerAlaValAspSerValGlnValLeuGluSerSerAspLeuPro      60
AlaLeuSerPheTyrValGlnGluTrpPheAspLeuHisGlyLysThrLeuSerAspGly      80
LysAlaArgLeuAlaLysLeuLysAsnLeuCysSerAsnLeuGlyAspProProAlaAsn     100
GluPheAsnAlaLysIlePheAlaAspTyrArgLysArgArgLeuAspGlyGluPheSer     120
ValAsnLysAsnAsnProProLysGluAlaThrValAsnArgGluHisAlaTyrLeuArg     140
AlaValPheAsnGluLeuLysSerLeuArgLysTrpThrThrGluAsnProLeuAspGly     160
ValArgLeuPheLysGluArgGluThrGluLeuAlaPheLeuTyrGluArgAspIleTyr     180
ArgLeuLeuAlaGluCysAspAsnSerArgAsnProAspLeuGlyLeuIleValArgIle     200
CysLeuAlaThrGlyAlaArgTrpSerGluAlaGluThrLeuThrGlnSerGlnValMet     220
ProTyrLysIleThrPheThrAsnThrLysSerLysLysAsnArgThrValProIleSer     240
LysGluLeuPheAspMetLeuProLysLysArgGlyArgLeuPheAsnAspAlaTyrGlu     260
SerPheGluAsnAlaValLeuArgAlaGluIleGluLeuProLysGlyGlnLeuThrHIS     280
ValLeuARGHisThrPheAlaSerHisPheMetMetAsnGLYGlyAsnIleLeuValLeu     300
LysGluIleLeuGlyHisSerThrIleGluMetThrMetArgTYRAlaHisPheAlaPro     320
SerHisLeuGluSerAlaValLysPheAsnProLeuSerAsnProAlaGln>>>           337
```

FIG. 3. Predicted amino acid sequence of the HP1 integrase.

which each of the four *att* sites was provided either as a supercoiled plasmid or as a linear molecule. The results are summarized in Table 1. The only supercoiled plasmid which gave detectable levels of recombination was pHPC104, which contains the HP1 *attP* site. This substrate is the donor, by analogy with the λ system. Linear plasmids which recombined with pHPC104 under these conditions were pHPC102, containing the *attB* site from *H. influenzae*, and, to a lesser extent, pHPC101, containing *attR*. Extracts containing HP1 integrase promoted recombination only in the direction corresponding to phage integration; no recombinant products were obtained when plasmids containing the *attL* and *attR* sites were the substrate pair. The absence of detectable recombination with acceptor substrates containing *attP* or *attL* is similar to the specificity pattern of λ integration (11) and suggests that HP1 sequences lying to the right of the core interfere with the acceptor function of DNA molecules containing them.

Since linear molecules containing *attR* are active acceptors and are produced by recombination between pHPC104 and pHPC102, detectable recombination might occur between pHPC104 and the principal recombination product (11.9 kb) to give a secondary product 16.8 kb in size. A product migrating at this position has been observed in reaction mixtures after extensive recombination (e.g., Fig. 5, lane 4).

Role of IHF in activity and expression of HP1 integrase. Our sequence data showed that the region neighboring the HP1 *attP* site contains a number of sequences similar to the consensus binding site ($YAAN_4TTGAT^T/_A$) for the specific DNA-binding protein IHF (4). Further, IHF appeared to be required for the integrative site-specific recombination which occurred between HP1 and *H. influenzae att* sites when plasmids containing these sites were propagated in *E. coli* (2). To characterize the role of IHF in this system further, it was necessary to resolve integrase activity from endogenous IHF. Two approaches were used: fractionation of extracts, and production of HP1 integrase activity in an *E. coli* strain deficient in IHF.

Crude extracts of induced *E. coli* HB101(pRAP6) were fractionated as described in Materials and Methods. Nucleic acids were removed by chromatography on DEAE-cellulose, and active fractions were further purified by absorption to and stepwise elution from phosphocellulose. The fraction eluting at 0.5 M KCl promoted recombination (Fig. 6); but the detection of this activity required the addition of purified IHF to the reaction mixtures. IHF eluted from phosphocellulose at salt concentrations above 0.7 M (27; E. S. Hwang, unpublished data). The results showed that IHF stimulates the recombination reaction promoted by HP1 integrase. IHF serves the same accessory function in site-specific recombination of λ *att* sites (13, 21).

We expected that induction of *E. coli* RW1815(pRAP6) would provide crude extracts in which site-specific recombination depended on the addition of purified IHF, since this host strain contains a deletion in the *hip* gene specifying one of the subunits of the protein. Surprisingly, this was not the case; no activity was detected in crude extracts from induced cultures of this strain, despite supplementation with purified IHF. This result raises the possibility that IHF acts as a regulator of HP1 integrase expression; IHF has been implicated in regulating expression of a diverse group of genes in *E. coli* (7). The nucleotide sequence in the segment upstream from the integrase initiation codon contains sev-

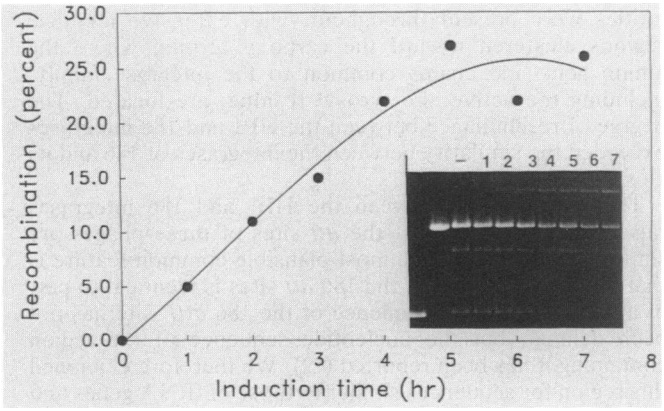

FIG. 4. Time course for induction of recombination activity. *E. coli* HB101(pRAP6) was grown and induced as described in Materials and Methods. At the indicated times, cells were harvested, washed, lysed and extracts (2 μg) were assayed for recombination activity. The inset photograph shows the pattern of ethidium bromide staining of the reaction mixtures assayed in this experiment. The unmarked lanes are the *attP* substrate (4.9-kb supercoil) and the *attB* substrate (7-kb linear); the lane marked – is an assay with an extract from uninduced cells, and the numbers above the other lanes are the times (hours) after heat induction when the extracts were prepared.

eral candidate IHF-binding sites between residues 1900 and 1940; specific interaction of IHF with this region of the sequence has been demonstrated (E. S. Hwang and J. J. Scocca, manuscript in preparation). To test the role of this upstream IHF-binding region in regulating integrase expression, we constructed a derivative of pRAP6 in which the 300-bp segment between residues 1831 and 2132 was deleted; this eliminated the upstream IHF-binding segment. The construction of this plasmid, designated pRAP1, is described in Materials and Methods, and the HP1 segment present in this plasmid is shown in Fig. 1.

TABLE 1. Substrate specificity of recombination[a]

Supercoiled substrate	Linear substrate	Relative activity (%)[b]
pHPC104	pHPC101	50
	pHPC102	100
	pHPC103	<2
	pHPC104	<2
pHPC103	pHPC101	<2
	pHPC102	<2
	pHPC103	<2
	pHPC104	<2
pHPC102	pHPC101	<2
	pHPC102	<2
	pHPC103	<2
	pHPC104	<2
pHPC101	pHPC101	<2
	pHPC102	<2
	pHPC103	<2
	pHPC104	<2

[a] Reaction mixtures contained 100 fmol each of the indicated plasmids; the linear substrate in each experiment was end labeled. Crude extract (1 U) was the source of integrase activity. The plasmids were pHPC101, containing *attR*; pHPC102, containing *attB*; pHPC103, containing *attL*; and pHPC104, containing *attP*.

[b] Activities are relative to the level measured with supercoiled pHPC104 and linear pHPC102.

Comparisons of the specific activity of extracts prepared from *E. coli* HB101(pRAP6) and HB101(pRAP1) and from *E. coli* RW1815(pRAP6) and RW1815(pRAP1) are summarized in Table 2. Extracts prepared from RW1815(pRAP6) had undetectable levels of activity, while induced cultures of RW1815(pRAP1) supported substantial levels of recombination when purified IHF was added to reaction mixtures. Deletion of the 300-bp upstream IHF-binding segment yielded plasmids which directed the expression of significant

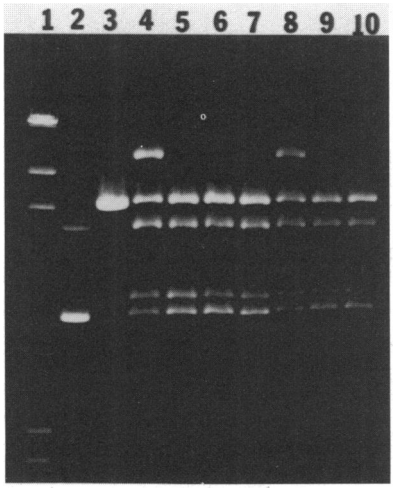

FIG. 5. Properties of the recombination reaction promoted by HP1 integrase. Lanes: 1, size standards; 2, pHPC104; 3, linear pHPC102; 4, standard reaction mixture with 2 μg of extract from *E. coli* HB101(pRAP6); 5, extract from *E. coli* HB101(pPAR1); 6, as in lane 4, except that the extract was heated to 60°C for 2 min; 7, as in lane 4 with the omission of spermidine; 8 to 10, effect of added KCl on the reaction; conditions were as in lane 4, with the addition of 50 mM (lane 8), 100 mM (lane 9), or 150 mM (lane 10) KCl.

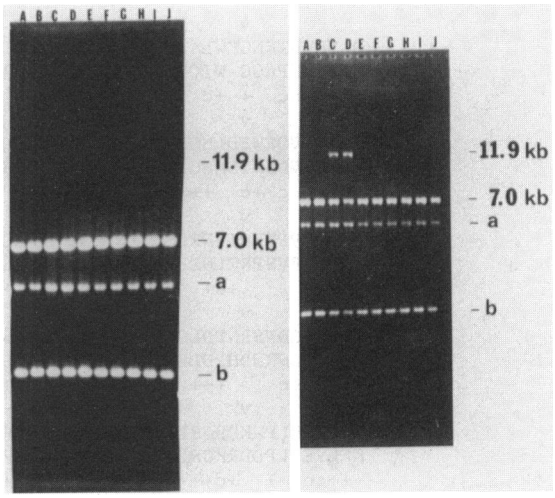

FIG. 6. Stimulation of recombination by IHF. Extracts were fractionated as described in Materials and Methods, and the fractions eluting from phosphocellulose were assayed. Left panel, Fractions assayed without added IHF; right panel, fractions assayed in the presence of 300 ng of IHF. Lanes A to E, Fractions eluted with 0.5 M KCl; lanes F to J, fractions eluted with 1.0 M KCl. The positions of relaxed (a) and supercoiled (b) pHPC104 are indicated.

TABLE 2. Effect of IHF status and upstream sequences on the expression of HP1 integrase activity in *E. coli*[a]

Plasmid	Host strain	Added IHF	Sp act[b]
pRAP6	HB101	+	700
pRAP6	HB101	−	500
pRAP1	HB101	+	1,600
pRAP1	HB101	−	550
pRAP6	RW1815	+	0
pRAP6	RW1815	−	0
pRAP1	RW1815	+	950
pRAP1	RW1815	−	50

[a] Strains containing the indicated plasmids were grown and induced and extracts were prepared and assayed as described in Materials and Methods. Where indicated, purified IHF (0.3 μg/ml) was added to the assay mixture.
[b] Units per milligram of protein.

levels of integrase activity in the absence of cellular IHF. In addition, HB101(pRAP1) yielded substantially higher levels of activity than did HB101(pRAP6), suggesting that the 300-bp upstream DNA segment contains one or more elements which reduced the expression of the integrase from the p_L promoter even in the presence of normal levels of IHF. The low but reproducible levels of recombination promoted by RW1815(pRAP1) extracts without added IHF most likely reflect the ability of high concentrations of HP1 integrase to effect significant levels of recombination in the absence of an accessory protein (J. H. Astumian, work in progress).

Relationship of HP1 integrase to other site-specific integrases. As noted earlier, the deduced amino acid sequence of the HP1 integrase contains the conserved residues which characterize the integrase family of site-specific recombination proteins (1). Pairwise comparisons of the HP1 sequence with other integrase sequences revealed a remarkable similarity with the sequence of the integrase of phage 186 (12); the comparison is shown in Fig. 7. The HP1 and 186 sequences could be aligned with few gaps. Sequence similarities were present throughout, with extensive identical regions clustered toward the carboxy termini where the amino acid side chains common to the integrase family, including the active site tyrosyl residue, are located. The degree of resemblance between the HP1 and 186 integrases exceeded the similarity between the integrases of 186 and its close relative P2 (1).

The resemblance between the HP1 and 186 integrases raises the possibility that the *att* sites of these phages are similar in some way. The most plausible common feature is that the common core of the 186 *att* sites is identical to part of a tRNA gene. The sequence of the 186 *attP* site has not been defined, but the nucleotide sequence of the region containing it has been reported (12). We therefore examined this region for sequences similar to those of tRNA genes and found that a 30-bp segment of 186 DNA in the *att* region (residues 671 to 700 [18]) is identical to the 3′ end of *E. coli* isoleucyl-tRNA (37). We speculate that this sequence constitutes the common core of the *att* site for 186.

DISCUSSION

The results presented here demonstrate that integrative recombination of HP1 *attP* and *attB* sites requires a gene product encoded by the DNA segment which lies immediately to the right of the *attP* region (2). The ORF in this region specifies a 38.6-kilodalton polypeptide which functions as the HP1 site-specific integrase, as shown by its activity in recombining the HP1 *attP* and the *H. influenzae attB* sites in vitro. The organization of the integration region of HP1 is similar to that of other lysogenic phages, with the structural gene for the recombinase located immediately adjacent to the *attP* site. In HP1, the integrase gene is oriented so that transcription must initiate at a site distal to *attP* and proceed toward it. This arrangement is shared by λ, P22, and 186, while in φ80 and in P4, the recombinase gene is oriented so that transcription must begin proximal to or within *attP* (15).

The amino acid sequence of HP1 integrase is clearly related to the integrase family of site-specific recombination

```
         v         v         v         v         v         v
MAVRKDTKNGKWLAEVYVNGNASRKWFLTKGDALRFYNQAKEQTTSAVDSVQVLESSDLP    60
MTVRKNPAGG-WICELYPNG-A-KGKRIRKKFATKGEALAFEQYTVQNPWQEEKEDRRT-    56
+c+++cc  + +c +c+ ++ + c   c +  + c    + ++ +      c  +
         v         v         v         v         v         v
ALSFYVQEWFDLHGKTLSDGKARLAKLKNLCSNLGDPPANEFNAKIFADYRKRRLDGEFS-   120
-LKELVDSWYSAHGITLKDGLKRQLAMHHAFECMGEPLARDFDAQMFSRYREKRLKGEYAR   116
 +   +c +c  ++ ++ ++  +   cc     c+c+ + c+c+ c+c ++ c++ ++cc
         v         v         v         v         v         v
VNKNNPPKEATVNREHAYLRAVFNELKSLRKWTTENPLDGVRLFKERETELAFLYERDIY    180
SNRVKEVSPRTLNLELAYFRAVFNELNRLGEWKGENPLKNMRPFRTEEMEMTWLTHDQIS    176
 +c       +c+ + ++ +++++++  +  + c+++++ c+ +c  + +ccc+   c+
         v         v         v         v         v         v
RLLAECDNSRNPDLGLIVRICLATGARWSEAETLTQSQVMPYKITFTNTKSKKNRTVPIS    240
QLLGECNRHDHPDLETVVRICLATGARWSEAESLRKSQLAKYKITYTNTKGRKNRTVPIS    236
 ++c++c    +++  c+++++++++++++++c+  ++c  ++++c++++cc++++++++
         v         v         v         v         v         v
KELFDMLP-KKRGRLFNDAYESFENAVLRAEIELPKGQLTHVLRHTFASHFMMNGGNILVL   300
KELYESLPDDKKGRLFSDCYGAFRSALERTGIELPAGQLTHVLRHTFASHFMMNGGNILVL   297
+++cc ++  +c++++ + + c+  +c +c ++++ ++++++++++++++++++++++++++
         v         v         v
KEILGHSTIEMTMRYAHFAPSHLESAVKFNPLSNPAQ   337
QRVLGHTDIKMTMRYAHFAPDHLEDAVKLNPLVHITNSK   336
  c+++c + ++++++++++ +++ +++ +++   cc
```

FIG. 7. Comparison of the amino acid sequences of HP1 and 186 integrases. HP1 integrase is the upper sequence. The sequence of the 186 integrase is from reference 12. Sequences were aligned manually; insertions were placed to maintain the alignment of regions of local identity. Residues identical in the two sequences are indicated by +, and conservative substitutions are indicated by c.

proteins. Integrases have diverse amino-terminal segments which are probably involved in specific recognition, but they share three common residues at relatively constant locations near the carboxyl termini, suggesting that the catalytic site is in this region (1, 28, 31, 43). Mutational analysis of P1 Cre (43), yeast FLP (31), and λ Int (28) have shown that alterations in conserved residues abolish recombination function but not sequence recognition. Covalent linkage of the substrate DNA to the conserved tyrosyl residue of Int through a 3′-phosphodiester bond has been demonstrated (28). Since HP1 integrase is a member of this family, it seems probable that it functions similarly in the breakage and rejoining steps of recombination.

The HP1 integrase sequence showed a striking similarity to the deduced sequence for the integrase of coliphage 186. The extent of identity for the HP1 and 186 sequences was greater than that between 186 and its close relative P2 (1) and is remarkable given the differences in host bacteria. The *attP* region of 186 contains a 30-bp sequence identical to the 3′ end of isoleucyl-tRNA, suggesting that the 186 prophage chromosome is inserted into this tRNA gene. Other members of the P2 phage family may also integrate within host tRNA genes. The *att* site core of the satellite phage P4, which consists of 20 residues at the 3′ end of the *leuX* allele of leucyl-tRNA (30), has considerable similarity to the corresponding portion of the HP1 *att* site, but the sequence of the P4 integrase shows only modest relationships of the HP1 or 186 integrase. When the precise sequence requirements for *att* site function in these phages are determined, it may be possible to relate similarities in protein sequences to any common elements in their binding sites.

Efficient recombination in vitro by purified HP1 integrase was stimulated by the accessory DNA-binding protein IHF. IHF is required for integration of λ and φ80 (21) and binds to the *att* regions of other lambdoid phages (14). IHF stimulates recombination by λ Int by binding to specific sites in *attP* and inducing bends in the DNA (35, 38). These bends facilitate the interactions between Int and the DNA substrate to form the recombinationally active nucleoprotein complex. The stimulation of HP1 recombination by IHF in vitro suggests that an intasomelike complex is formed during site-specific recombination in this system as well.

The levels of HP1 integrase expressed in *E. coli* showed an unexpected dependence on IHF. Deletion of the 300-bp segment located approximately 100 bp upstream from the initiation codon eliminated this IHF dependence; this deletion also appeared to increase integrase expression in cells with normal IHF levels. IHF binds to this segment, as judged by DNase I footprinting (Hwang and Scocca, in preparation). The 300-bp IHF-binding segment appears to contain a regulatory element which makes downstream genes responsive to cellular IHF levels. IHF has been implicated in regulating the expression of an increasing number of *E. coli* genes (reviewed in reference 7). It is intriguing that the expression of λ Int is indirectly regulated by IHF through its effect on the expression of the *c*II gene (10, 17, 29). It appears that in HP1 the expression of integrase activity may respond directly to the analogous host factor. Taken together, the effects of IHF in this system suggest that a protein functionally analogous to *E. coli* IHF is present in *H. influenzae*. Preliminary data (E. S. Hwang, unpublished data) indicate that extracts of *H. influenzae* contain an IHF-like activity.

HP1 integrase promoted recombination only with a supercoiled *attP* substrate. Analogy with the λ system suggests that supercoiled *attP* molecules provide the sites at which the donor complex or intasome (3) is organized. The intasome then captures the acceptor substrate to form the complete recombination complex (34).

Recombination promoted by HP1 integrase was directional, since the substrates for phage integration (*attP* and *attB*) were efficiently recombined under conditions in which the substrates for excision (*attL* and *attR*) did not react. Recognition of the six clustered mismatches which distinguish the *attP* and *attB* core segments cannot explain the observations. Molecules containing *attR* (with a core identical to *attP*) acted as acceptors but not as donors. The presence of the 182-bp *attP* core did not eliminate acceptor activity or confer donor activity. Conversely, the failure of *attL* segments with cores identical to *attB* to act as acceptors means that this core sequence is not sufficient for acceptor activity; the right arm of *attP* must be absent as well. Directional reaction and inhibition of *attB* function by arm sequences suggest that critical integrase-DNA interactions must occur at sites (arm sites) which lie outside the 182-bp common core. The existence of arm sites combined with the size of this core in the HP1 system make it likely that the functional HP1 *attP* site spans a DNA segment more extensive than the 240-bp λ *attP* site (11, 24). Resection studies on the HP1 *attP* site (M. A. Hauser, unpublished data) and DNA footprinting studies (J. Hakimi Astumian, unpublished data) suggest that the HP1 *attP* site is 410 bp long, 170 bp longer than λ *attP*. The presence of right arm (P′) sites reduces or eliminates acceptor (*attB*) function in λ integrative recombination (11), suggesting that directional reaction is governed by similar factors in the two systems. Excisive recombination between the *attR* and *attL* sites of HP1 is likely to require additional protein factors to activate the reaction. Efficient excisive recombination of λ requires a specific phage-encoded protein, Xis, and another host factor, FIS (39), in addition to Int and IHF (4, 42). The requirements for excision of HP1 from the host are as yet uncharacterized; preliminary data indicate that insertional inactivation of the integrase gene segment abolished detectable phage production by *H. influenzae* lysogenic for HP1 (T. Nguyen, unpublished data).

The reaction requirements, the stimulation by IHF, and the inhibitory action of arm sites on *attB* function constitute significant common elements in HP1 and λ integrative recombination in vitro. These similarities suggest that the recombining complexes in the two systems share a fundamentally similar architecture despite the differences in the host organisms and in the extents of the *att* sites. Studies on the interaction of purified HP1 integrase with its substrates should provide direct evidence concerning the organization of the recombination complex; these investigations are in progress. Studies on the integration of HP1 should also assist in understanding the mechanisms and consequences of insertions targeted to host tRNA genes.

ACKNOWLEDGMENTS

We thank H. A. Nash for valuable advice and discussions, R. Weisberg for providing bacterial strains, E. S. Hwang for the gift of purified IHF, and W.-D. Reiter for communicating results prior to publication. We also thank J. Hakimi Astumian, M. A. Hauser, E. S. Hwang, and W.-D. Reiter for comments on the manuscript.

This work was supported by grant MV-385 from the American Cancer Society.

LITERATURE CITED

1. **Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljundquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. L.**

Narayana, L. S. Pierson III, N. Sternberg, and J. M. Leong. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. **5**:433–440.
2. **Astumian, J. H., A. S. Waldman, and J. J. Scocca.** 1989. Site-specific recombination between cloned *attP* and *attB* sites from the *Haemophilus influenzae* bacteriophage HP1 propagated in recombination-deficient *Escherichia coli*. J. Bacteriol. **171**:1747–1750.
3. **Better, M., C. Lu, R. C. Williams, and H. Echols.** 1982. Site-specific DNA condensation and pairing mediated by the Int protein of bacteriophage λ. Proc. Natl. Acad. Sci. USA **79**: 5837–5841.
4. **Craig, N. L.** 1988. The mechanism of conservative site-specific recombination. Annu. Rev. Genet. **22**:77–105.
5. **Craig, N. L., and H. A. Nash.** 1984. *E. coli* integration host factor binds to specific sites in DNA. Cell **39**:707–716.
6. **Fitzmaurice, W. P., A. S. Waldman, R. C. Benjamin, P. C. Huang, and J. J. Scocca.** 1984. Nucleotide sequence and properties of the cohesive DNA termini from bacteriophage HP1*c*1 of *Haemophilus influenzae* Rd. Gene **31**:197–203.
7. **Friedman, D. I.** 1988. Integration host factor: a protein for all reasons. Cell **55**:545–554.
8. **Harm, W., and C. S. Rupert.** 1963. Infection of transformable cells of *Haemophilus influenzae* by bacteriophage and bacteriophage DNA. Z. Vererbungsl. **94**:336–348.
9. **Harrison, B., and S. B. Zimmerman.** 1986. T4 polynucleotide kinase: macromolecular crowding increases the efficiency of reaction at DNA termini. Anal. Biochem. **158**:307–315.
10. **Hoyt, A., D. Knight, A. Das, H. I. Miller, and H. Echols.** 1982. Control of phage λ development by stability and synthesis of cII protein: role of the viral *cIII* and host *hflA*, *himA*, and *himD* genes. Cell **31**:565–573.
11. **Hsu, P. L., W. Ross, and A. Landy.** 1980. The phage λ *att* site: functional limits and interaction with Int protein. Nature (London) **285**:85–91.
12. **Kalionis, B., I. B. Dodd, and J. B. Egan.** 1986. Control of gene expression in the P2-related template coliphages. III. DNA sequence of the major control region of phage 186. J. Mol. Biol. **191**:199–209.
13. **Kikuchi, Y., and H. A. Nash.** 1978. The bacteriophage λ *int* gene product. A filter assay for genetic recombination, purification of Int, and specific binding to DNA. J. Biol. Chem. **253**:7149–7157.
14. **Leong, J. M., S. Nunes-Duby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy.** 1985. The φ80 and P22 attachment sites. Primary structure and interaction with *Escherichia coli* integration host factor. J. Biol. Chem. **260**:4468–4477.
15. **Leong, J. M., S. E. Nunes-Duby, A. B. Oser, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy.** 1986. Structural and regulatory divergence among site-specific recombination genes of lambdoid phage. J. Mol. Biol. **189**:603–616.
16. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurement with the Folin phenol reagent. J. Biol. Chem. **193**:265–275.
17. **Mahajna, J., A. B. Oppenheim, A. Rattray, and M. Gottesman.** 1986. Translation initiation of bacteriophage λ gene *c*II requires integration host factor. J. Bacteriol. **165**:167–174.
18. **Mandel, M., and A. Higa.** 1970. Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. **53**:159–162.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
21. **Miller, H. I., and D. I. Friedman.** 1980. An *E. coli* gene product required for λ site-specific recombination. Cell **20**:711–719.
22. **Milman, G.** 1987. Expression plasmid containing the λ P_L promoter and cI857 repressor. Methods Enzymol. **153**:482–491.
23. **Mizuuchi, K., and M. Mizuuchi.** 1979. Integrative recombination of bacteriophage λ: in vitro study of the intermolecular reaction. Cold Spring Harbor Symp. Quant. Biol. **43**:1111–1114.
24. **Mizuuchi, K., R. Weisberg, L. Enquist, M. Mizuuchi, M. Buraczynska, C. Foeller, P. L. Hsu, W. Ross, and A. Landy.** 1981. Structure and function of the phage λ *att* site: size, Int-binding sites, and location of the crossover point. Cold Spring Harbor Symp. Quant. Biol. **45**:429–437.
25. **Mukhopadhyay, M., and N. C. Mandal.** 1983. A simple procedure for large-scale preparation of pure plasmid DNA free from chromosomal DNA from bacteria. Anal. Biochem. **133**:265–270.
26. **Nash, H. A.** 1981. Integration and excision of bacteriophage λ: the mechanism of conservative site specific recombination. Annu. Rev. Genet. **15**:143–167.
27. **Nash, H., and C. Robertson.** 1981. Purification and properties of the *E. coli* protein factor required for λ integrative recombination. J. Biol. Chem. **256**:9246–9253.
28. **Pargellis, C. A., S. E. Nunes-Duby, L. Moitoso de Vargas, and A. Landy.** 1988. Suicide recombination substrates yield covalent λ integrase-DNA complexes and lead to identification of the active site tyrosine. J. Biol. Chem. **263**:7678–7685.
29. **Peacock, S., H. Weissbach, and H. A. Nash.** 1984. In vitro regulation of phage λ *cII* gene expression by *Escherichia coli* integration host factor. Proc. Natl. Acad. Sci. USA **81**:6009–6013.
30. **Pierson, L. S., III, and M. L. Kahn.** 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*. DNA sequences of the phage and host regions involved in recombination. J. Mol. Biol. **196**:487–496.
31. **Prasad, P. V., L. J. Young, and M. Jayaram.** 1987. Mutations in the 2-micrometer circle site specific recombinase that abolish recombination without affecting substrate recognition. Proc. Natl. Acad. Sci. USA **84**:2189–2193.
32. **Reiter, W.-D., P. Palm, and S. Yeats.** 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17**:1907–1914.
33. **Richards, G. M.** 1974. Modifications of the diphenylamine reaction giving increased sensitivity and simplicity in the estimation of DNA. Anal. Biochem. **57**:369–376.
34. **Richet, E., P. Abcarian, and H. A. Nash.** 1988. Synapsis of attachment sites during λ integrative recombination involves capture of a naked DNA by a protein-DNA complex. Cell **52**:9–17.
35. **Robertson, C. A., and H. A. Nash.** 1988. Bending of the bacteriophage λ attachment site by *Escherichia coli* integration host factor. J. Biol. Chem. **263**:3554–3557.
36. **Sadowski, P.** 1986. Site-specific recombinases: changing partners and doing the twist. J. Bacteriol. **165**:341–347.
37. **Sprinzl, M., T. Vorderwuelbecke, and T. Hartmann.** 1985. Compilation of sequences of tRNA genes. Nucleic Acids Res. **13**:r51–r103.
38. **Thompson, J. F., and A. Landy.** 1988. Empirical estimation of protein-induced DNA bending angles: application to λ site-specific recombination. Nucleic Acids Res. **16**:9687–9705.
39. **Thompson, J. F., L. Moitoso de Vargas, C. Koch, R. Kahmann, and A. Landy.** 1987. Cellular factors couple recombination with growth phase: characterization of a new component in the λ site-specific recombination pathway. Cell **50**:901–908.
40. **Waldman, A. S., W. P. Fitzmaurice, and J. J. Scocca.** 1986. Integration of the bacteriophage HP1*c*1 genome into the *Haemophilus influenzae* Rd chromosome in the lysogenic state. J. Bacteriol. **165**:297–300.
41. **Waldman, A. S., S. D. Goodman, and J. J. Scocca.** 1987. Nucleotide sequence and properties of the site involved in lysogenic insertion of the bacteriophage HP1*c*1 genome into the *Haemophilus influenzae* chromosome. J. Bacteriol. **169**:238–246.
42. **Weisberg, R. A., and A. Landy.** 1983. Site-specific recombination in phage λ, p. 211–250. *In* R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg (ed.), Lambda II. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
43. **Wierzbicki, A., M. Kendall, K. Abremski, and R. Hoess.** 1987. A mutational analysis of the bacteriophage P1 recombinase Cre. J. Mol. Biol. **195**:785–794.