JOURNAL OF BACTERIOLOGY, Sept. 1990, p. 4852–4860
0021-9193/90/094852-09$02.00/0
Copyright © 1990, American Society for Microbiology
Vol. 172, No. 9

Interaction of Integration Host Factor from *Escherichia coli* with the Integration Region of the *Haemophilus influenzae* Bacteriophage HP1

EUN SEONG HWANG 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 20 April 1990/Accepted 16 June 1990

The specific DNA-binding protein integration host factor (IHF) of *Escherichia coli* stimulates the site-specific recombination reaction between the *attP* site of bacteriophage HP1 and the *attB* site of its host, *Haemophilus influenzae*, in vitro and also appears to regulate the expression of HP1 integrase. IHF interacts specifically with DNA segments containing the *att* sites and the integrase regulatory region, as judged by IHF-dependent retardation of relevant DNA fragments during gel electrophoresis. The locations of the protein-binding sites were identified by DNase I protection experiments. Three sites in the HP1 *attP* region bound IHF, two binding sites were present in the vicinity of the *attB* region, and one region containing three partially overlapping sites was present in the HP1 integrase regulatory segment. The binding sites defined in these experiments all contained sequences which matched the consensus IHF binding sequences first identified in the λ *attP* region. An activity which stimulated the HP1 site-specific integration reaction was found in extracts of *H. influenzae*, suggesting that an IHF-like protein is present in this organism.

The temperate bacteriophage HP1 can lysogenize its host, *Haemophilus influenzae* Rd (13). In the course of this process the circularized 32.4-kilobase (kb) HP1 genome is inserted into the host chromosome by recombination between specific phage (*attP*) and host (*attB*) sites (37). The HP1 *attP* site and the corresponding *H. influenzae attB* site are distinctive in that they share a common core which extends over a distance of 182 base pairs (bp). This common core consists of two completely identical segments of 93 and 62 bp, which are separated by an internal 27-bp segment which contains six mismatched residues. Recombinational crossover occurs within the 62-bp segment (38). The HP1 *att* cores include sequences which are virtually identical to *Escherichia coli* tRNA genes (30); the precise insertion point is located within one of these tRNA gene sequences (M. A. Hauser, unpublished data). The HP1 insertion system belongs to an otherwise disparate family of site-specific recombination systems which are targeted to host tRNA gene sequences (28, 30).

Recombination between HP1 *attP* and *attB* sites cloned in *E. coli* or in vitro is similar in a number of respects to the site-specific recombination reactions of other temperate phages, of which λ is the most extensively studied (reviewed in references 3, 18, 32, and 39). Integrative recombination of HP1 *attP* and *attB* sites is promoted by a phage-specified protein belonging to the integrase family of site-specific recombinases (1, 11). Recombination of HP1 *attP* and *attB* sites in *E. coli* also requires the specific DNA-binding protein integration host factor (IHF) (12), and the reaction in vitro is stimulated three- to fivefold by *E. coli* IHF (9; J. Hakimi, unpublished data). These protein requirements for HP1 recombination are similar to the requirements for integrative recombination of λ *att* sites by the λ integrase (Int) (26). Int-mediated recombination is almost completely dependent on IHF (16). The λ Int protein binds to several sites on the *attP* and *attB* substrates (15, 24), brings the recombining sites into apposition, and promotes the specific breakage and rejoining reactions (4, 18). IHF recognizes and binds to specific sequences in *attP* (5); detailed studies have indicated that IHF interacts with the minor groove of DNA and that the binding species is a single heterodimer of the two subunits of IHF (40). This binding introduces sharp DNA bending (31, 35); the resulting bends facilitate the simultaneous binding of Int to separated binding sites (6) and assist in forming the intasome, a condensed nucleoprotein structure in which recombination occurs (2, 18).

IHF has been implicated in a wide spectrum of regulatory and DNA recognition processes in *E. coli* and associated replicons (reviewed in reference 7). In every case IHF binds specifically to the relevant DNA at sequences that are identical or nearly identical to the consensus binding sequence that has been identified in the *attP* region of λ and other lambdoid phages (5, 20). How IHF binding and the consequent bending of DNA produce these diverse effects is presently unclear.

Because *E. coli* IHF increases the efficiency of specific recombination between HP1 *attP* and *H. influenzae attB* sites in *E. coli* and in vitro and also increases the expression of HP1 integrase in *E. coli*, specific interactions between *E. coli* IHF and sites in the integration region of HP1 are expected. To identify these interactions, we examined the IHF-dependent shifts in electrophoretic mobility of specific DNA fragments and the patterns of protection from DNase I cleavage produced by IHF. These experiments showed that *E. coli* IHF bound specifically and with high affinity to three sites in the HP1 *attP* region and to two sites in the *H. influenzae attB* region. IHF also interacted with the upstream DNA segment which was implicated in the IHF-dependent regulatory effects on HP1 integrase expression in previous studies (11). These results clearly require that a protein which is similar to *E. coli* IHF be present in *H. influenzae*, the natural host of HP1. We also found evidence showing that suitable extracts of *H. influenzae* are able to

* Corresponding author.

replace *E. coli* IHF in stimulating site-specific recombination of HP1 *att* sites in vitro.

MATERIALS AND METHODS

Bacterial strains. *E. coli* HB101 grown in LB medium supplemented with appropriate antibiotics (22) was used for maintenance and propagation of plasmids. IHF was prepared from *E. coli* HB101 or from *E. coli* K5746 (27) (obtained from H. A. Nash, National Institute of Mental Health, Bethesda, Md.). *H. influenzae* Rd001 was grown in heart infusion broth (Difco Laboratories) supplemented with hemin and NAD (33).

Plasmids. pWR1 (15) contains the 493-bp *Hin*dIII-*Bam*HI fragment of λ (spanning the *attP* site) in place of the *Hin*dIII-*Bam*HI segment of vector pBR322. pHPC101 contains the HP1 *attR* region on a 7.3-kb fragment; pHPC102 contains the HP1 *attB* region on a 2.6-kb segment from *H. influenzae*; and pHPC104 contains the 2.25-kb *Bgl*II-*Hae*III segment of HP1 which includes the *attP* and integrase gene regions (15). These plasmids were derivatives of pBR322 and were constructed as described previously (15, 38). pHPC201 was constructed by inserting the 471-bp *Sau*3AI-*Hin*PI fragment from pHPC104 into the *Sma*I site of pUC19; this fragment contained a fully functional *attP* site (Hauser, unpublished data).

Expression plasmids have been described previously (11); pRAP6 contains a 1,590-bp *Alu*I fragment from HP1 which includes the structural gene for HP1 integrase and 423 bp of HP1 DNA that is 5′ of the initiation codon. pRAP1 contains the open reading frame and 125 bp of HP1 upstream sequence. pRAP6Z is pRAP6 with the 545-bp sequence encoding the α peptide of β-galactosidase inserted downstream from the open reading frame (S. D. Goodman, Ph.D. thesis, Johns Hopkins University, Baltimore, Md., 1987).

Plasmids were isolated from chloramphenicol-treated cultures by using the boiling method (14), which was modified for large culture volumes as described previously (38). Supercoiled plasmid DNA was purified by equilibrium centrifugation on cesium chloride-ethidium bromide density gradients.

Labeling of DNA fragments. Plasmid DNA was first digested with one or more appropriate restriction enzymes, dephosphorylated with *E. coli* alkaline phosphatase, and then labeled at the 5′ ends by reaction with [γ-^{32}P]ATP (Amersham Corp.) and T4 polynucleotide kinase (New England BioLabs, Inc.). Singly end-labeled fragments were generated by a second restriction digestion and then isolated by electrophoresis on 1.3% low-gelation-temperature agarose gels. Segments containing the desired fragment were excised and melted, and the DNA was recovered and purified by using NACS Prepac columns (Bethesda Research Laboratories, Inc.).

Preparation of IHF. Extracts of *E. coli* K5746, which overproduces IHF, were used to prepare and purify IHF by previously described methods (27); the purification was monitored by specific retardation of the HP1 *attP* fragment, as described below.

Assay for site-specific recombination. Recombination was measured as described previously (11); HP1 integrase purified to apparent homogeneity (J. Hakimi and J. J. Scocca, unpublished data) was a generous gift from J. Hakimi. Reaction mixtures (20 μl) contained 100 fmol of supercoiled pHPC202 (*attP* donor), 100 fmol of linear pHPC301 (*attB* acceptor), and 640 fmol of integrase in 50 mM Tris phosphate buffer (pH 7.5) supplemented with 5 mM EDTA, 6 mM spermidine, and 20% (vol/vol) glycerol. IHF or extracts from *H. influenzae* cultures were added as indicated below. After incubation for 4 h at 25°C, reaction mixtures were heated at 65°C with 50 μg of proteinase K (Boehringer Mannheim Biochemicals). The reaction products were separated by electrophoresis on 0.7% agarose gels.

Fragment retardation assays. Fragments for analysis were produced by digesting pWR1 (λ *attP*) with *Taq*I, pHPC104 (HP1 *attP*) with *Rsa*I, and pHPC102 (*H. influenzae attB*) with *Pst*I and *Hin*dIII. DNA digests (0.1 pmol) were incubated with purified IHF in 15 μl of the binding buffer of Craig and Nash (5) at 37°C for 30 to 60 min. IHF-dependent retardation of specific fragments was analyzed by electrophoresis (2.7 V/cm) on 1.7% agarose gels that were formed and run in 25 mM Tris–190 mM glycine (pH 8.5)–1 mM EDTA and stained with ethidium bromide.

DNase I footprinting. The method of Galas and Schmitz (8) as modified for IHF by Craig and Nash (5) was used for DNase I footprinting. End-labeled DNA (5 to 15 fmol) was incubated with different concentrations of IHF in binding buffer (100 μl) (5) containing 7 mM $MgCl_2$ and 3 mM $CaCl_2$ for 30 to 60 min at 37°C. Pancreatic DNase I (Worthington Diagnostics) (4 to 6 μl of a 0.5-μg/ml solution in 50 mM Tris chloride [pH 7.4]–1.0 mM EDTA–1.0 mM 2-mercaptoethanol) was added, and incubation was continued for 4 to 5 min. Samples were treated with 50 μg of proteinase K and extracted with phenol-chloroform-isoamyl alcohol. DNA samples were precipitated, washed, and separated by electrophoresis on standard sequencing gels (1 m); standards were prepared by base-specific chemical cleavage (23). Separation was visualized by autoradiography.

Preparation of extracts from *H. influenzae*. Cultures (50 ml) of *H. influenzae* were grown to a density of 1×10^9 to 2×10^9 CFU/ml, harvested, washed, and osmotically shocked as described previously (34). The shocked cells were suspended in 3.8 ml of 20 mM potassium phosphate buffer (pH 7.5) containing 0.15% (wt/vol) Brij 35 and 20 μg of phenylmethylsulfonyl fluoride per ml, incubated for 25 min at 0°C, and centrifuged. The supernatant fraction was used.

RESULTS

Interaction of IHF with DNA fragments containing HP1 *att* sites. The binding of IHF in the vicinity of the HP1 *attP* and *attB* sites was examined by measuring the protein-dependent retardation of specific restriction fragments in gel electrophoresis. Purified IHF was incubated with restriction fragment mixtures from pHPC104 (containing HP1 *attP*), pHPC102 (containing HP1 *attB*), and pWR1 (containing λ *attP*). To facilitate comparison, digestion preparations were chosen so that the three *att* regions were located on fragments of approximately the same size (λ *attP* on a 774-bp *Taq*I fragment, HP1 *attP* on a 763-bp *Rsa*I fragment, and *H. influenzae attB* on a 774-bp *Hin*dIII-*Pst*I fragment) (Fig. 1B). As shown in Fig. 1, the fragment containing the HP1 *attP* region and the fragment containing the *H. influenzae attB* region were both specifically retarded by IHF in the concentration range which retarded the migration of the fragment containing the λ *attP* site. IHF bound specifically to the HP1 *attP* region and to the cognate *attB* site with affinities similar to the affinity in its interaction with λ *attP*. The HP1 *attP* fragment was completely retarded at a molar ratio of IHF to DNA of 18:1; the λ *attP* fragment was 75% retarded under these conditions. The binding of IHF to the *H. influenzae attB* region was weaker since at a 90:1 ratio of protein to DNA an appreciable fraction of the fragment remained

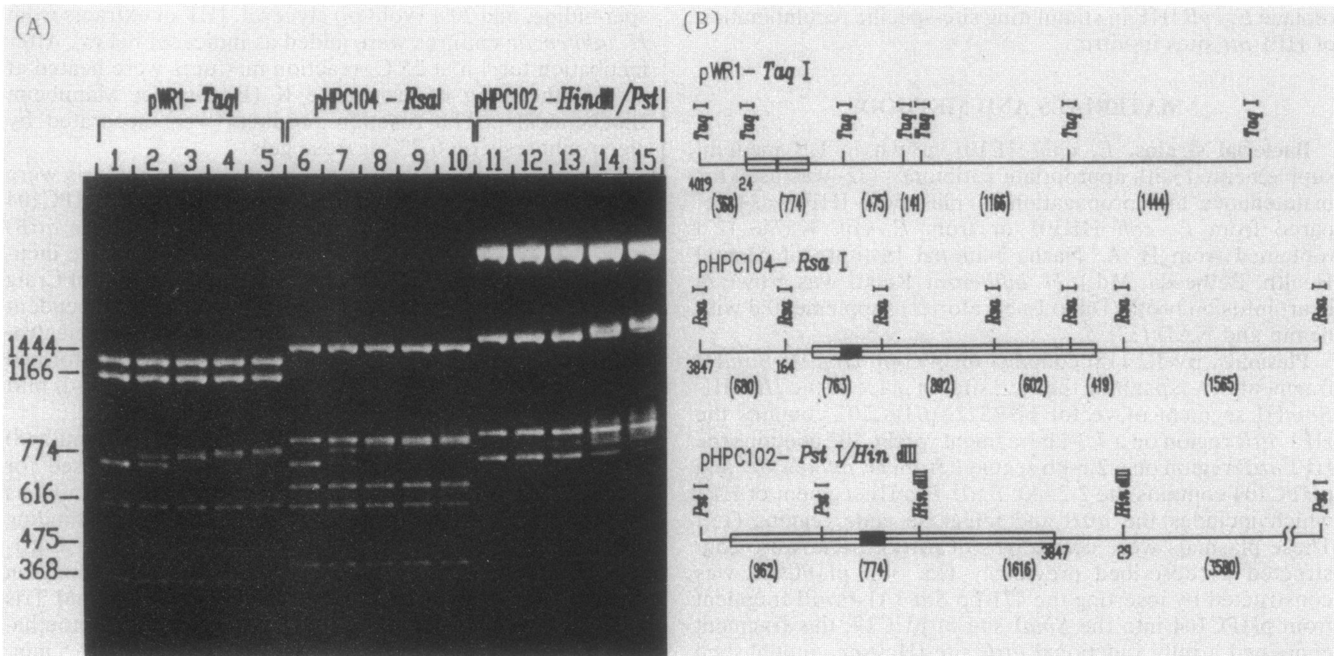

FIG. 1. IHF-induced retardation of DNA fragments containing integration regions. (A) Restriction digests (0.1 pmol) were incubated with various concentrations of IHF as described in the text and separated by electrophoresis. Lanes 1 through 5, pWR1 (λ *attP*) digested with *Taq*I; lanes 6 through 10, pHPC104 (HP1 *attP*) digested with *Rsa*I; lanes 11 through 15, pHPC102 (*H. influenzae attB*) digested with *Hin*dIII and *Pst*I. Lanes 1, 6, and 11 contained no IHF; lanes 2, 7, and 12 contained 0.9 pmol of IHF; lanes 3, 8, and 13 contained 1.8 pmol of IHF; lanes 4, 9, and 14 contained 3.6 pmol of IHF; and lanes 5, 10, and 15 contained 9.0 pmol of IHF. The fragment sizes of the pWR1 digest are indicated. The 774-bp fragment containing λ *attP* was the third largest band in lane 1, the 763-bp fragment containing HP1 *attP* was the third largest band in lane 6, and the 774-bp fragment containing *H. influenzae attB* was the fourth largest band in lane 11. (B) Maps of the fragments used. The plasmids and digests are the plasmids and digests used for the experiment shown in panel A. The lines indicate pBR322 DNA, the open boxes indicate the inserted DNA, and the solid boxes indicate the locations of the *att* sites. The numbers in parentheses are the fragment sizes (in base pairs).

unretarded. At IHF concentrations lower than those needed for complete retardation, each fragment containing an *att* site had at least one discrete intermediate species which migrated between the unretarded and fully retarded species. This suggests that each fragment contained more than one IHF binding site and that these sites differed significantly in affinity. No evidence for cooperative binding of IHF to the HP1 *att* sites was observed (data not shown).

The shifts in mobility produced by IHF binding varied for each fragment. These differences may have reflected conformational differences in the individual complexes which might have been due to different spacings of the individual binding sites and to varying distances between the binding sites and the fragment ends (31, 35).

The mobilities of several other fragments were retarded by IHF. One group of these fragments included fragments containing residues 4025 to 4325 of pBR322, where the three IHF-binding sites associated with a hot spot for IS*1* insertion are located (21). Two of these fragments, the 680-bp *Rsa*I fragment from pHPC104 and the 1,645-bp *Hin*dIII-*Pst*I fragment from pHPC102, were retarded under the conditions used in the experiments shown Fig. 2, and the 368-bp *Taq*I fragment from pWR1 was retarded at higher IHF concentrations (data not shown). The 602-bp *RsaI* fragment from pHPC104 was also retarded specifically; this fragment contained the HP1-derived segment upstream from the integrase coding region. The presence of this region was associated with the IHF requirement for efficient expression of HP1 integrase activity in *E. coli*, and specific interaction of IHF with this region was expected and was analyzed further (see below). Retardation of the 965-bp *Hin*dIII-*Pst*I fragment from pHPC102 suggested that IHF bound weakly to sites that were within approximately 400 bp of the *attB* core.

Identification of IHF binding sites in the HP1 *attP* region. The IHF-binding sites in the HP1 *attP* and *H. influenzae attB* regions were located by using DNase I footprinting analysis. Binding sites in the HP1 *attP* region were identified by locating the positions of IHF-dependent protection on the 454-bp *Ava*I-*Bgl*II fragment from pHPC101 labeled at the *Ava*I end and a 471-bp *Bam*HI-*Eco*RI fragment from pHPC201 labeled either at the *Bam*HI end or at the *Eco*RI end. These fragments enabled us to scan the 711-bp segment (residues 111 to 821) (11) surrounding the HP1 *attP* site. Three regions of this segment were specifically protected from DNase I digestion by IHF (Fig. 2). Two of these sites, designated sites P1 and P2, were located to the left of the crossover point, and one site, designated site P′, was located to the right of the recombination point. Sites P1 and P2 were separated by approximately 240 bp, and sites P2 and P′ were approximately 115 bp apart. The extent of IHF protection ranged from 30 to 45 bp; the exact boundaries of protection at sites P1 and P2 were difficult to locate because of the low levels of DNase I digestion in these regions in the absence of IHF. The sequences surrounding these three sites are shown in Fig. 3, and a map of their arrangement in HP1 *attP* is shown in Fig. 4.

IHF binding sites in the *H. influenzae attB* region. IHF binding to a 770-bp segment containing the *H. influenzae attB* region was examined. The sites of IHF-induced protection from DNase I were determined on the 511-bp *Ava*I-*Pst*I

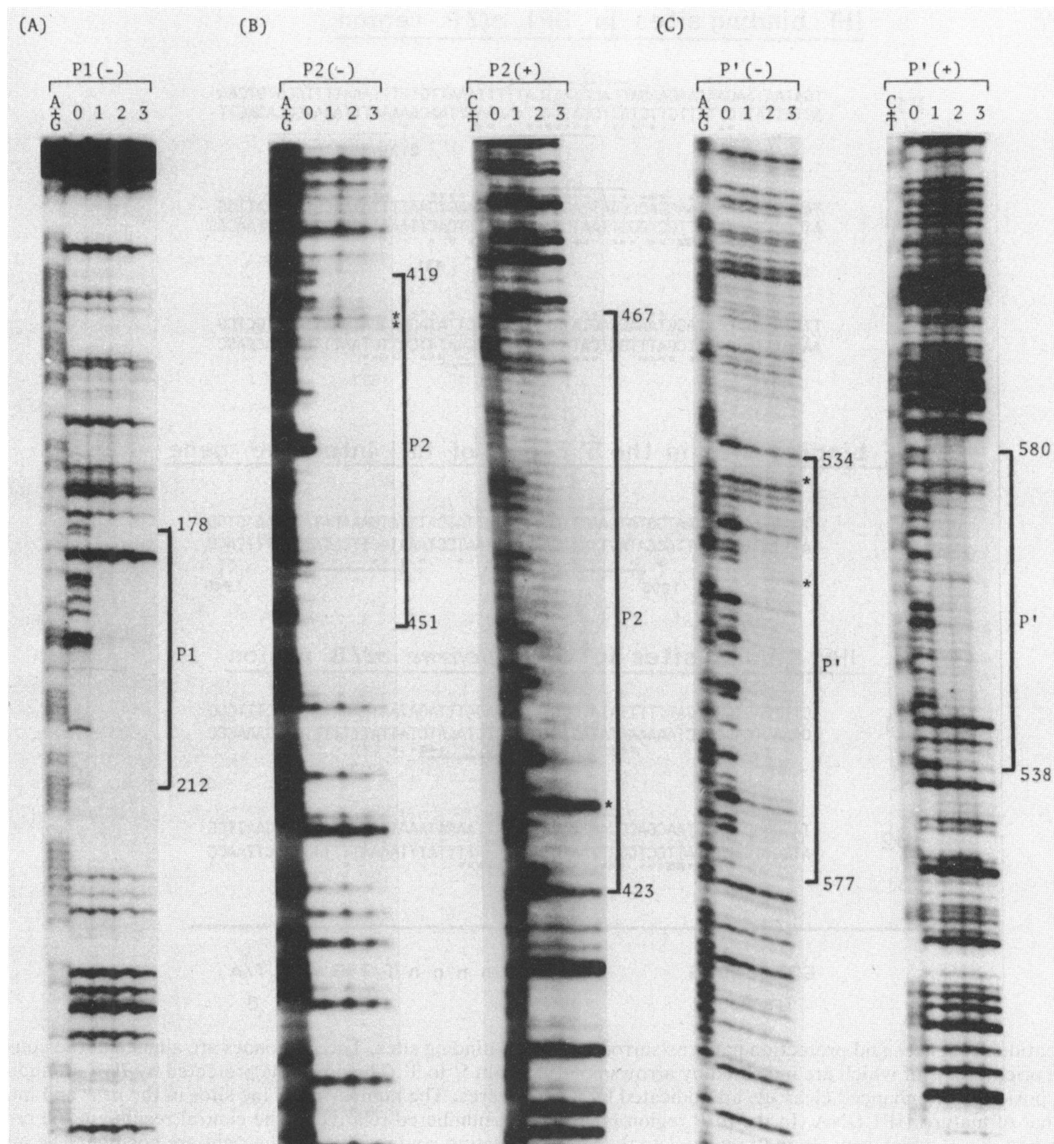

FIG. 2. IHF protection of sites in the HP1 *attP* region. DNase I digestions were performed as described in the text. Except where indicated, the footprints on both the top (plus) and bottom (minus) strands were determined. The extents of protection are indicated by brackets, and the positions of enhanced cleavage are indicated by asterisks. The A+G- or C+T-specific cleavage reactions (23) were used as markers. (A) Footprinting of the bottom strand of site P1. The DNA (0.1 nM) was the 454-bp *Ava*I-*Bgl*II fragment from pHPC101 labeled at the *Ava*I site. Lane 0, No IHF; lane 1, 17 nM IHF; lane 2, 34 nM IHF; lane 3, 68 nM IHF. (B) Site P2. For the bottom (minus) strand the 454-bp *Ava*I-*Bgl*II fragment from pHPC101 (0.12 nM) was labeled at the *Ava*I site (lane 0, no IHF; lane 1, 17 nM IHF; lane 2, 25 nM IHF; lane 3, 75 nM IHF). For the top (plus) strand the 471-bp *Bam*HI-*Eco*RI fragment from pHPC201 (0.15 nM) was labeled at the *Eco*RI site (lane 0, no IHF; lane 1, 25 nM IHF; lane 2, 38 nM IHF; lane 3, 113 nM IHF). (C) Site P′. For the bottom (minus) strand the 471-bp *Bam*HI-*Eco*RI fragment from pHPC201 (0.1 nM) was labeled at the *Bam*HI site (lane 0, no IHF; lane 1, 40 nM IHF; lane 2, 80 nM IHF; lane 3, 133 nM IHF). For the top (plus) strand the 471-bp *Bam*HI-*Eco*RI fragment from pHPC201 (0.15 nM) was labeled at the *Eco*RI site (lane 0, no IHF; lane 1, 25 nM IHF; lane 2, 38 nM IHF; lane 3, 113 nM IHF).

fragment from pHPC102 labeled at the *Ava*I end and the 264-bp *Ava*I-*Hin*dIII fragment from pHPC101 labeled at the *Ava*I end. As shown in Fig. 5, IHF protected two sites in this region, which were designated sites B1 and B2. The sequences and protection patterns for these sites are shown in Fig. 3. Site B2 was located within the 182-bp core region at the same relative position as site P2 in HP1 *attP*. The protection pattern on the bottom strand of site B2 was almost identical to that of site P2 (there were slight differences at positions where the two core regions differed in sequence). The other site, site B1, was approximately 140 bp 5′ to site B2 and was weakly protected by IHF compared with site B2. Both site B1 and site B2 were within regions that were relatively rich in A·T residues; the sequences surrounding these sites are shown in Fig. 3.

IHF binding to sequences 5′ to the HP1 integrase gene. The results of our previous studies indicated that efficient expression of HP1 integrase activity controlled by the λ p_L promoter in *E. coli* depended almost completely on IHF. The IHF requirement was eliminated by deleting the 300-bp

IHF binding sites in HP1 *att*P region

P1
TGATAATAAGAATAACAAGATAACCTAATGATTTTTAAATTGTTTTTAAATTTTCCGTGTGAA
ACTATTATTCTTATTGTTCTATTGGATTACTAAAAATTTAACAAAAATTTAAAAGGCACACTT
178 212

P2
423 467
TACCGACTGAGCTAACGACCCAATTAGATGATTTTAAAGTGAATTTTAAAACCTTTAGATTGG
ATGGCTGACTCGATTGCTGGGTTAATCTACTAAAATTTCACTTAAAATTTTGGAAATCTAACC
419 451

P'
538 580
TTTTAAATACCATAGCTAAACCAGTATAAACAATAACTTATAGCAACATTAATTGCTTGCTCG
AAAATTTATGGTATCGATTTGGTCATATTTGTTATTGAATATCGTTGTAATTAACGAACGAGC
534 577

IHF binding site in the 5' region of HP1 integrase gene

TAAATAACCCAAAATGGTATCAATTTCTTGTTTATTAGGATTTATCAAACATATATAAAGTGG
ATTTATTGGGTTTTACCATAGTTAAAGAACAAATAATCCTAAATAGTTTGTATATATTTCACC
1895 1940

IHF binding sites in *H. influenzae att*B region

B1
CTTTGAGTCGGGGATTTTTTATATGAGTTTCAGAGATTTAGATAATAAAAAAACACCTTTCGG
GAAACTCAGCCCCTAAAAAATATACTCAAAGTCTCTAAATCTATTATTTTTTTGTGGAAAGCC
-257 -213

B2
TACCGACTGAGCTAACGACCCAATTAGATGATTTTAAAATAAATTTTAAAATCTTTGAATTGG
ATGGCTGACTCGATTGCTGGGTTAATCTACTAAAATTTTATTTAAAATTTTAGAAACTTAACC
-114 -74

	*							*				
CONSENSUS	Y	A	A	n	n	n	n	T	T	G	A	T	T/A
frequency	8	8	7					4	8	8	6	7	8

FIG. 3. Nucleotide sequences and protection patterns surrounding IHF-binding sites. The sequences are aligned at the consensus specific binding sites, the orientations of which are indicated by arrows pointing from 5′ to 3′. The positions protected by IHF are indicated by small squares, and the positions of enhanced cleavage are indicated by large squares. The numbering of the sites in the *attP* and integrase regions corresponds to that of mature HP1 DNA. In the *attB* region positions are numbered relative to the central residue in the crossover region (Hauser, unpublished data), which is position 0; residues to the left are negative, and residues to the right are positive. The number of times that each of the eight sites agrees with the consensus sequence is also indicated; the asterisks indicate residues which are identical in all of the IHF sites described so far.

segment beginning 425 bp and ending 123 bp upstream from the initiation codon (11). These findings suggested that IHF bound specifically within this regulatory segment. The existence of one or more IHF binding sites in this segment was apparent from the results of gel retardation experiments (Fig. 1). The locations of these IHF-binding sites as determined by DNase I footprinting are shown in Fig. 6. IHF protected one segment in the 577-bp *Hin*dIII-*Rsa*I fragment of pRAP6Z, which contained the regulatory segment, at a site approximately 210 bp upstream from the initiation codon. The protected region contained two plausible IHF-binding sequences, one of which had the consensus sequence on both strands of the duplex (Fig. 3). The symmetric IHF site was closer to the integrase initiation codon, and the other site began 10 bp further upstream (Fig. 4). Despite the presence of these multiple candidate binding sites, the extent of protection was no greater than the protection associated with the more widely separated asymmetric sites in the *att* regions.

Nucleotide sequences and relative affinities of the IHF-binding sites. The nucleotide sequences of the six IHF-binding sites which we identified are compared with the consensus IHF-binding site (5, 7, 20) in Fig. 3. All of the sites contained a sequence which was essentially identical to the consensus sequence that was deduced from binding sites on DNA segments derived from *att* sites of lambdoid phages. Each site was bordered by sequences that were high in A·T base pairs.

The relative affinities of the IHF-binding sites which we identified were estimated and compared with the affinities of the sites in λ *attP* by comparing footprint intensities at various protein concentrations (data not shown). The approximate order was as follows: P1 > P2 ≈ H′ ≈ H2 ≈ P′ > H1 ≈ B2 > B1. IHF bound to the sites in the HP1 *attP* region with affinities which equaled or exceeded those for the sites in λ *attP*, while the binding to the two sites in the *H. influenzae attB* region was markedly weaker. Interaction with the upstream regulatory region was similar to interac-

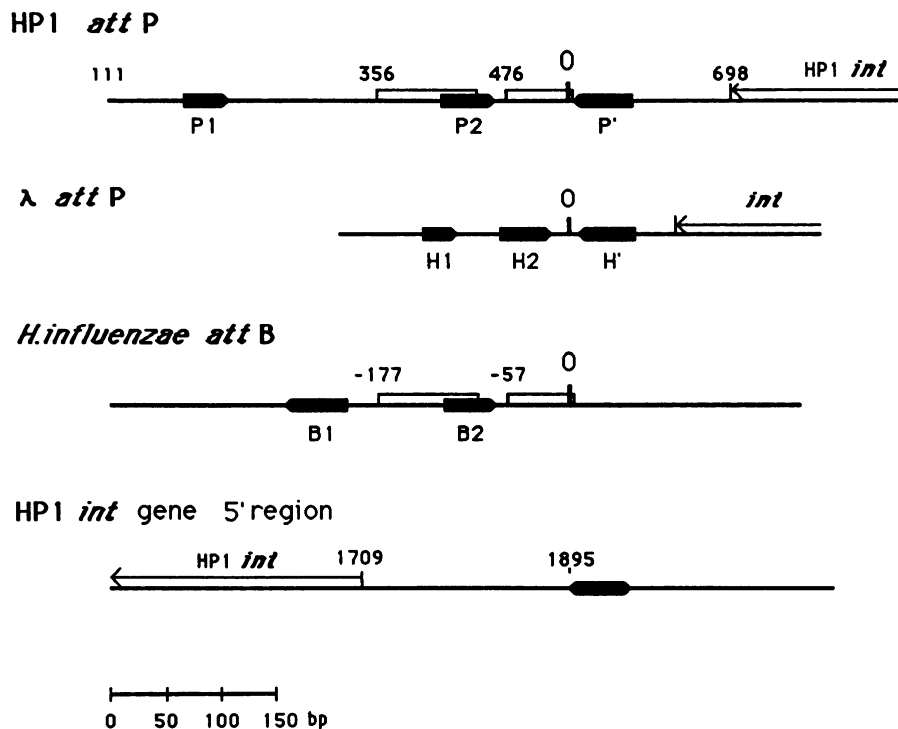

FIG. 4. Organization of HP1 and *H. influenzae* IHF-binding sites. The solid boxes indicate the locations and orientations (5' to 3') of the IHF-binding sites. The 92- and 62-bp blocks of identical sequences common to the HP1 *attP* and *H. influenzae attB* regions are indicated by open boxes. The locations and orientations of the IHF-binding sites in the λ *attP* site are provided for comparison. The three sites are aligned at the respective crossover points; the central residues of these recombination sites are indicated by zeros (19).

tion with sites P2 and P'. These qualitative estimates were in accord with the mobility retardation results and with estimates of the relative affinities of the λ sites for IHF (9).

Stimulation of recombination by extracts of *H. influenzae*. *H. influenzae* cells treated with Brij 35 as described above yielded extracts which increased the rate of integrative recombination promoted by HP1 integrase. *H. influenzae* extracts were similar to purified *E. coli* IHF in stimulating recombination; the results of a representative experiment are shown in Fig. 7. Inclusion of both *H. influenzae* extract and *E. coli* IHF produced no more stimulation than did optimal concentrations of either factor added singly (data not shown). These results indicate that the HP1 host, *H. influenzae*, contains an activity which is functionally similar to IHF in stimulating site-specific recombination. The presence of nonspecific DNA-binding proteins in these preparations precluded the use of mobility retardation assays for characterization.

DISCUSSION

The results of the binding studies described here confirmed our previous findings that *E. coli* IHF interacts with the integration region of *H. influenzae* bacteriophage HP1. Three specific high-affinity binding sites were identified in the vicinity of the HP1 *attP* site, two sites of moderate affinity were identified in the *H. influenzae attB* region, and one high-affinity site was identified in the regulatory region located upstream from the initiation codon for the HP1 integrase gene. More than 35 IHF-binding sites having varying affinities have been identified to date (reviewed in reference 7), including those reported here. The affinity of IHF for a given site appears to depend on how well its sequence agrees with the consensus sequence $YAAN_4TT$ GATW (5, 7, 20); the affinity may be enhanced when the site is in an A·T-rich environment, especially toward the 3' end (29). The 3' ends of the HP1 binding sequences are bordered by A·T-rich stretches which are likely sites for local intrinsic bending (17). The binding of IHF introduces sharp bends in DNA (31, 35); intrinsic bends might facilitate IHF interactions by reducing the barriers to distortion induced by IHF.

The function of IHF in λ site-specific recombination is precisely to introduce bends at the correct locations, since replacing an IHF site with another site for either protein-induced or intrinsic bending preserves the function of the λ *attP* segment (10). These bends facilitate the simultaneous binding of λ Int to separated binding sites on the substrate DNA (6) to form the condensed nucleoprotein structure or intasome needed for the breakage and joining reactions of recombination (2, 18). Since the recombination reaction promoted by HP1 integrase is stimulated by *E. coli* IHF and by the analogous activity in extracts of the authentic host *H. influenzae*, protein-induced bends are likely to be important for assembling the HP1 intasome structure also. The sequences of the binding sites for these proteins appear to be well conserved, suggesting that the proteins are similar as well.

IHF-binding sites have been mapped in the *att* regions of λ (5), φ80, and *Salmonella typhimurium* phage P22 (20). The λ *attP* site contains three IHF-binding sites; two of these (sites H1 and H2) lie to the left and one (site H') lies to the right of the strand exchange region (5). The consensus sequences in these sites have their 3' ends directed toward

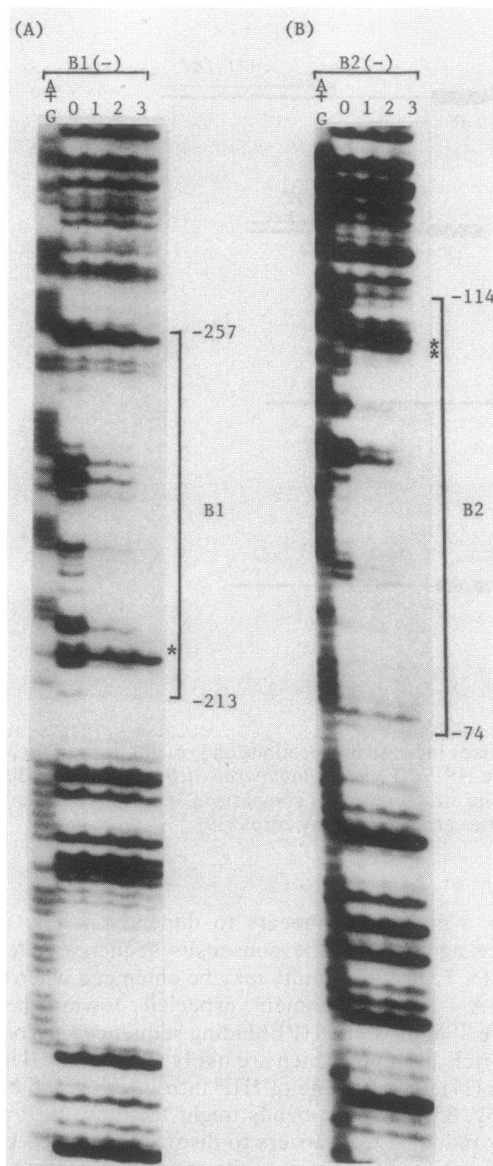

FIG. 5. IHF protection of sites in the *H. influenzae attB* region. The footprints of sites B1 and B2 were obtained by using a single fragment, the 511-bp *Ava*I-*Hin*dIII fragment from pHPC102 labeled at the *Ava*I site. DNA (0.1 nM) was incubated with IHF and DNase I as described in the text. Lanes 0, No IHF; lanes 1, 17 nM IHF; lanes 2, 25 nM IHF; lanes 3, 75 nM IHF. The footprints of the bottom strand are shown. (A) Site B1. (B) Site B2.

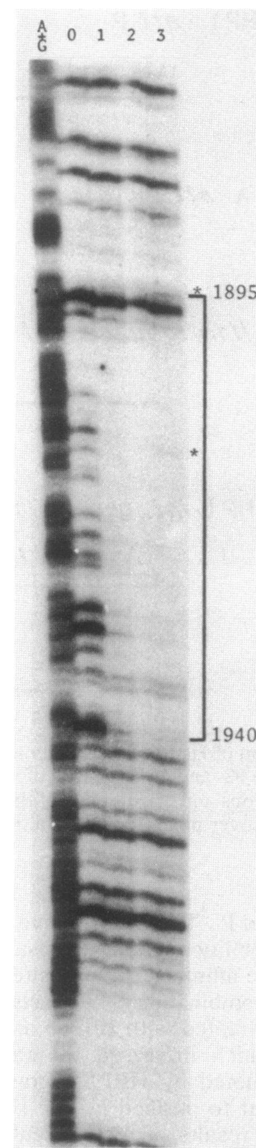

FIG. 6. IHF protection in the regulatory segment that is 5' of the HP1 integrase coding segment. The 577-bp *Hin*dIII-*Rsa*I fragment from pRAP6Z (0.1 nM) labeled at the *Hin*dIII site was mixed with 0, 40, 80, or 133 nM IHF (lanes 0, 1, 2, and 3, respectively). Only the footprint on the strand which is anticoding for the integrase gene is shown.

the crossover point. The ϕ80 *attP* region contains a single IHF site to the left of the presumptive crossover point, while the P22 *attP* region contains two IHF sites which flank the proposed recombination region (20). The 3' ends of the IHF consensus sequences in these *attP* sites are directed toward the crossover points.

A comparison of the organization of the IHF-binding sites in the HP1 *attP* region with the organization of the sites in λ *attP* is shown in Fig. 4. When the two *attP* sites are aligned at the positions of strand exchange (24; M. A. Hauser and J. J. Scocca, unpublished data), it is apparent that the λ site H' and the HP1 site P' are at nearly identical positions relative to the crossover points. HP1 binding sites P1 and P2, which are to the left of the crossover point, are more widely separated than sites H1 and H2 of λ, and the distance between sites P2 and P' exceeds the distance which separates sites H2 and H' of λ. Resection studies on the HP1 region have shown that the leftmost site (site P1) can be removed without affecting *attP* function (Hauser, unpublished data). Therefore, this strong IHF-binding site does not participate in the integration reaction. While it is possible that this site is involved in excisive recombination only, its position relative to the sites of integrase binding (Hakimi, unpublished data) make this unlikely. Site P1 is within 200 bp of the left *cos* site of HP1 and may function in producing cohesive termini or in DNA packaging. The presence of at most two IHF sites in the functional HP1 *attP* region makes this region similar to the P22 *attP* site in this respect. These

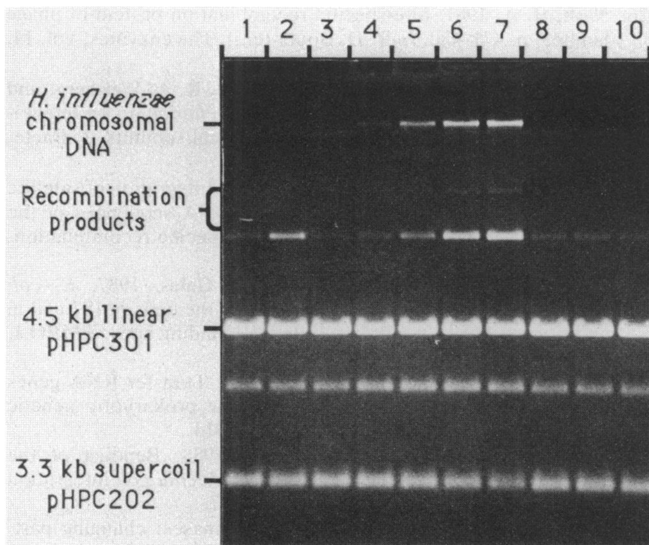

FIG. 7. Stimulation of site-specific recombination by *H. influenzae* extracts. Reaction mixtures contained 100 fmol of supercoiled HP1 *attP* donor plasmid pHPC202 (3.28 kb), 100 fmol of linear *attB* acceptor plasmid pHPC301 (4.5 kb), and 640 fmol of HP1 integrase in 50 mM Tris phosphate buffer (pH 7.5) supplemented with 5 mM EDTA, 6 mM spermidine, and 20% glycerol. Purified *E. coli* IHF or detergent extracts from *H. influenzae* were added, and mixtures were incubated at 25°C for 4 h. Lanes 1 and 3, No addition; lane 2, 43 ng of *E. coli* IHF; lane 4, 0.5 μl of *H. influenzae* extract; lane 5, 2 μl of extract; lane 6, 4.5 μl of extract; lane 7, 6.5 μl of extract. Lanes 8 through 10 contained 0.7, 4.5, and 6.5 μl, respectively, of the Brij 35 solution used to prepare the extracts.

different *attP* segments are most similar in the regions which separate the respective crossover points and integrase genes (20, 21).

Studies on the expression of HP1 integrase in *E. coli* cultures have indicated that efficient expression of this activity from the λ p_L promoter depends on active IHF; this effect of IHF can be eliminated by deleting the 300-bp HP1 DNA segment beginning 425 bp 5′ of the integrase initiation codon and ending 125 bp 5′ of this point (11). A high-affinity binding site for IHF is clearly present in this putative regulatory segment, as shown by the DNase I protection results described above. The consensus sequence of this binding site is located 200 bp 5′ of the integrase initiation codon, and unlike the oriented sites in the *attP* region, this regulatory IHF site has overlapping consensus sequences on both strands and is therefore symmetric. The consensus IHF-binding site possesses a modest degree of interrupted symmetry in the AAN_4TT sequence. It remains to be determined whether more extensive symmetry influences the binding affinity or the degree of bending by IHF. IHF regulates the expression of a number of genes in *E. coli* (7). Since IHF binding induces an acute bend in the DNA (31, 35), these regulatory effects may require DNA bending as well. Specific DNA bends could bring separated protein-binding sites into apposition, or a specific DNA conformation generated by bending might be required for efficient expression of a given gene.

The *attB* sites of HP1 and P22 contain IHF-binding sites, while the *attB* sites of λ and φ80 do not (20); introduction of an IHF site has no effect on *attB* function in λ recombination (25). Resection studies have shown that both IHF-binding sites may be deleted from the *H. influenzae attB* region without affecting the ability of the site to act as a recombination acceptor in vitro (Hauser, unpublished data). While HP1 site B1 is well outside the sequence common to *attP* and *attB*, site B2 is within the common core at the same relative location as site P2 within *attP*. One common property of the *att* sites of HP1 and P22 is that their extended common cores constitute parts of tRNA genes. P22 inserts within a threonyl tRNA gene (28), and HP1 inserts within a leucyl tRNA segment which is part of an operon of three *H. influenzae* tRNA genes (Hauser, unpublished data). It is possible that these IHF sites are present because these *attB* sites are tRNA genes and have no integration-related function. Transcripts of tRNA sequences contain substantial secondary structure which could lead to premature termination, so it is possible that IHF binding in these sequences prevents incomplete transcription. It would be interesting to determine whether tRNA genes which are not *att* sites also bind IHF.

The regulatory and stimulating effects of *E. coli* IHF on the HP1 integration system provided circumstantial evidence for the existence of an analogous or possibly homologous protein in the HP1 host, *H. influenzae*. Direct evidence for such a protein has now been obtained, as described above. In addition to *E. coli*, from which IHF-like activity was first isolated and extensively studied, *S. typhimurium* has been reported to have such an activity (20), and regulatory sequences which interact with IHF have been isolated from *Chlamydomonas* chloroplast DNA (36). Taken together, these findings support the hypothesis that IHF-like activity is widely distributed and serves important functions in a broad spectrum of organisms. Purification and characterization of the IHF-like activity from *H. influenzae* should assist in identifying conserved structural and functional features of these proteins. These studies are in progress.

ACKNOWLEDGMENTS

We thank J. Hakimi for generous gifts of purified HP1 integrase, H. Nash for providing strain K5746, and S. Goodman, J. Hakimi, and M. Hauser for assistance and advice.

This work was supported by grant MV385 from the American Cancer Society.

LITERATURE CITED

1. **Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, B. Kalionis, S. V. L. Narayana, L. S. Pierson, N. Sternberg, and J. N. Leong.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. **5**:433–440.
2. **Better, M., C. Lu, R. 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.
3. **Craig, N. L.** 1988. The mechanism of conservative site-specific recombination. Annu. Rev. Genet. **22**:77–105.
4. **Craig, N. L., and H. A. Nash.** 1983. The mechanism of phage λ site-specific recombination: site-specific breakage of DNA by Int topoisomerase. Cell **35**:795–803.
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. **de Vargas, L. M., C. A. Pargellis, N. M. Hasan, E. W. Bushman, and A. Landy.** 1988. Autonomous DNA binding domains of λ integrase recognize two different sequence families. Cell **54**: 923–929.
7. **Friedman, D. I.** 1988. Integration host factor: a protein for all reasons. Cell **55**:545–554.
8. **Galas, D., and A. Schmitz.** 1978. DNase footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5**:3157–3170.
9. **Gardner, J. F., and H. A. Nash.** 1986. Role of *Escherichia coli* IHF protein in lambda site-specific recombination. A mutational analysis of binding sites. J. Mol. Biol. **191**:181–189.

10. **Goodman, S. D., and H. Nash.** 1989. Functional replacement of a protein-induced bend in a DNA recombination site. Nature (London) **341**:251–254.
11. **Goodman, S. D., and J. J. Scocca.** 1989. Nucleotide sequence and expression of the gene for the site-specific integration protein from bacteriophage HP1 of *Haemophilus influenzae*. J. Bacteriol. **171**:4232–4240.
12. **Hakimi Astumian, J. H., A. S. Waldman, and J. J. Scocca.** 1988. Site-specific recombination between cloned *attP* and *attB* sites from the *Haemophilus influenzae* phage HP1 propagated in recombination-deficient *Escherichia coli*. J. Bacteriol. **171**: 1747–1750.
13. **Harm, W., and C. S. Rupert.** 1963. Infection of transformable cells of *Haemophilus influenzae* by bacteriophage and bacteriophage DNA. Z. Vererbungsl. **94**:336–348.
14. **Holmes, P. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
15. **Hsu, P. L., W. Ross, and A. Landy.** 1980. The λ phage *att* site: functional limits and interaction with Int protein. Nature (London) **285**:85–91.
16. **Kikuchi, Y., and H. A. Nash.** 1978. The bacteriophage λ *int* gene product. J. Biol. Chem. **253**:7149–7157.
17. **Koo, H. S., H. M. Wu, and D. M. Crothers.** 1986. DNA bending at adenine · thymine tracts. Nature (London) **320**:501–506.
18. **Landy, A.** 1989. Dynamic, structural, and regulatory aspects of λ site-specific recombination. Annu. Rev. Biochem. **58**:913–950.
19. **Landy, A., and W. Ross.** 1977. Viral integration and excision: structure of the λ *att* sites. Science **197**:1147–1158.
20. **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 integration host factor. J. Biol. Chem. **260**:4468–4477.
21. **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.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**: 499–560.
24. **Mizuuchi, K., R. Weisberg, L. Enquist, M. Mizuuchi, M. Buraczynska, C. Foeller, P. Hsu, W. Ross, and A. Landy.** 1981. Structure and function of the phage lambda *att* site: size, Int-binding sites, and location of the crossover point. Cold Spring Harbor Symp. Quant. Biol. **45**:429–437.
25. **Mizuuchi, M., and K. Mizuuchi.** 1980. Integrative recombination of bacteriophage λ: extent of the DNA sequence involved in attachment site function. Proc. Natl. Acad. Sci. USA **77**: 3220–3224.
26. **Nash, H. A.** 1981. Site-specific recombination protein of phage lambda, p. 471–480. *In* P. D. Boyer (ed.), The enzymes, vol. 14, Academic Press, Inc., New York.
27. **Nash, H. A., C. A. Robertson, E. Flamm, R. A. Weisberg, and H. I. Miller.** 1987. Overproduction of *Escherichia coli* integration host factor, a protein with nonidentical subunits. J. Bacteriol. **169**:4124–4127.
28. **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 site-specific recombination. J. Mol. Biol. **196**:487–496.
29. **Prentki, P., M. Chandler, and D. J. Galas.** 1987. *E. coli* integration host factor bends the DNA of the ends of IS*1* and in an insertion hot spot with multiple IHF binding sites. EMBO J. **6**:2479–2487.
30. **Reiter, W., P. Palm, and S. Yeats.** 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17**:1907–1914.
31. **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.
32. **Sadowski, P.** 1986. Site-specific recombinases: changing partners and doing the twist. J. Bacteriol. **165**:341–347.
33. **Scocca, J. J., R. L. Poland, and K. C. Zoon.** 1974. Specificity in deoxyribonucleic acid uptake by transformable *Haemophilus influenzae*. J. Bacteriol. **118**:369–373.
34. **Sutrina, S. L., and J. J. Scocca.** 1983. DNA-binding proteins of *Haemophilus influenzae*: purification and characterization of a major intracellular binding protein. J. Bacteriol. **155**:246–253.
35. **Thompson, J. F., and A. Landy.** 1988. Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. Nucleic Acids Res. **16**: 9687–9705.
36. **Thompson, R. J., and G. Mosig.** 1988. Integration host factor (IHF) represses a *Chlamydomonas* chloroplast promoter in *E. coli*. Nucleic Acids Res. **16**:3313–3326.
37. **Waldman, A. S., W. P. Fitzmaurice, and J. J. Scocca.** 1986. Integration of the bacteriophage HP1c1 genome into the *Haemophilus influenzae* Rd chromosome in the lysogenic state. J. Bacteriol. **165**:297–300.
38. **Waldman, A. S., S. D. Goodman, and J. J. Scocca.** 1987. Nucleotide sequence and properties of the sites involved in lysogenic insertion of the bacteriophage HP1c1 genome into the *Haemophilus influenzae* chromosome. J. Bacteriol. **169**:238–246.
39. **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.
40. **Yang, C. C., and H. A. Nash.** 1989. The interaction of *E. coli* IHF protein with its specific binding sites. Cell **57**:869–880.