Cell, Vol. 39, 707–716, December 1984 (Part 2), Copyright © 1984 by MIT 0092-8674/84/130707-10 $02.00/0

E. coli Integration Host Factor Binds to Specific Sites in DNA

Nancy L. Craig* and Howard A. Nash†
Laboratory of Neurochemistry
National Institute of Mental Health
Bethesda, Maryland 20205

Summary

E. coli integration host factor (IHF) both participates directly in phage lambda site-specific recombination and regulates the expression of phage and bacterial genes. Using protection from nuclease and chemical attack as an assay, we examined the interaction of IHF with DNA. We found that IHF is a specific DNA binding protein that interacts with three distinct segments of *att*P, the recombination site carried by phage lambda. We also found that specific IHF binding sites are located in non-*att* DNA. Several non-*att* IHF binding sites that we have identified are adjacent to genes whose expression is altered in IHF mutants. From comparison of the sequences protected by IHF, we suggest that the critical determinant in specific IHF–DNA interaction is contained in the sequence $\begin{matrix}\text{T}\cdot\text{PyAA}\cdots\text{PuTTGaT}\cdot\\ \text{A}\cdot\text{PuTT}\cdots\text{PyAACtA}\cdot\end{matrix}$

Introduction

The integration of bacteriophage lambda into the E. coli chromosome occurs by reciprocal recombination between specific sites in the phage and bacterial chromosomes, *attP* (POP′) and *attB* (BOB′) (for review see Nash, 1981; Weisberg and Landy, 1983). Each *att* site consists of a core (0) flanked by two arms. The core is a 15 bp segment common to *att* sites in which the exchange of DNA strands occurs during recombination. The flanking arms of the *att* sites (P, P′, B, and B′) are unique. *attP* is approximately 240 bp in length, extending from about −150 to +90 (measured with respect to the center of the core); *attB* is approximately 25 bp in length, extending from about −11 to +11. The products of integrative recombination are the hybrid sites *attL* (BOP′) and *attR* (POB′) which flank the integrated prophage. Prophage excision occurs by reciprocal recombination between *attL* and *attR*. Integrative recombination in vitro requires only two proteins: Int and integration host factor (IHF). Int is a single polypeptide of molecular weight about 40,000 daltons; it is the product of the phage *int* gene. IHF contains two polypeptides, α and β, which are products of bacterial genes. The *himA* gene encodes α, a polypeptide of molecular weight about 11,000 daltons; β, a polypeptide of molecular weight about 9,500 daltons, is the product of the *hip* (also called *himD*) gene (see also Kikuchi et al., 1984; Flamm and Weisberg, 1984). The active component of IHF is likely a heterodimer of α and β.

How do Int and IHF promote recombination between *attP* and *attB*? Int is a specific DNA binding protein that can recognize two distinct nucleotide sequences: an "arm-type" recognition sequence (Ross and Landy, 1982) and a "junction-type" recognition sequence (Ross and Landy, 1983). The interaction of Int with the multiple copies of these sequences that exist in *att* sites mediates at least some of Int's activities in recombination. The interaction of Int with "junction-type" sequences that bridge the arms and cores of *att* sites prompts the action of Int topoisomerase within the core (Craig and Nash, 1983). This reaction underlies the breakage and rejoining reactions that accomplish strand exchange. The role(s) of Int's interaction with the "arm-type" recognition sequences remains undefined. Although efficient in vitro integrative recombination requires both Int and IHF, Int alone is capable of promoting integrative recombination, albeit at a much reduced efficiency (Lange-Gustafson and Nash, 1984). This indicates that IHF functions as an accessory protein in recombination and enhances Int's intrinsic capacity to promote recombination. Previous work using a DNA filter-binding assay demonstrated that purified IHF binds to DNA (Nash and Robertson, 1981). However, no specificity for IHF binding to *att* sites was observed. In this work, we examine the interaction of purified IHF with *att* sites using the protection of *att* DNA against attack by both nuclease and chemical agents as an assay. We report that IHF is a specific DNA binding protein that binds to three distinct sites in *attP*. The IHF binding sites are interspersed among and closely adjacent to the Int binding sites in *attP*.

In addition to its direct role in lambda site-specific recombination, IHF also participates in several other cellular processes. *himA* and *hip* mutants are defective in the expression of some phage (Miller and Friedman, 1980; Miller, 1981) and bacterial genes (Miller et al., 1981; Szekely et al., 1984; Friden et al., 1984; Friedman et al., 1984), indicating that IHF has a regulatory role in which it mediates gene expression. We also examine the interaction of purified IHF with DNA containing the regulatory regions of several genes whose expression is altered in *himA* and *hip* mutants. We report that IHF binds specifically to these non-*att* DNAs and that this binding is promoted by recognition of the same nucleotide sequence that directs the binding of IHF to *att* sites.

Results

The Interaction of IHF with *attP* as Revealed by Protection Against Nuclease Attack

To probe for a specific interaction of IHF and DNA, we used the "footprinting" method of Galas and Schmitz (1978). The experiment shown in Figure 1 displays digestion of an *attP*-containing fragment by neocarzinostatin (NCS) in the presence and absence of IHF. Comparison of lanes 3 and 4 demonstrates that IHF protects three regions in *attP* against NCS digestion; we designate these

* Present address: Department of Microbiology and Immunology and George W. Hooper Foundation, University of California, San Francisco, California 94143.
† Present address: Laboratory of Molecular Biology, National Institute of Mental Health, Bethesda, Maryland 20205.

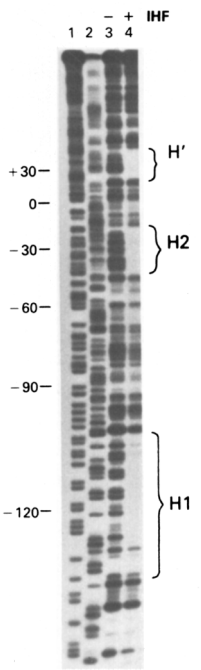

Figure 1. Protection Against Neocarzinostatin Attack of an *attP*-Containing DNA Fragment by IHF

The *attP*-containing restriction fragment Eco RI to Bam HI from pPH54 was 5′ end-labeled at the Eco RI site with ^{32}P. This fragment contains the Eco RI to Hind III segment of pBR322 and an *attP*-containing segment extending from −160 (at the Hind III site) to +242 (at the Bam HI site). The nucleotide sequence of this fragment is presented in Landy and Ross, 1977 and Hsu et al., 1980. The sequence is numbered with 0 as the central base of the common core: negative numbers extend leftward into the P arm and positive numbers extend rightward into the P′ arm. Lane 1 contains the products of an A+G sequencing reaction of the fragment. Lane 2 contains the products of a T+C sequencing reaction of the fragment. The interaction of IHF with the fragment was examined by NCS attack as described in Experimental Procedures. Lane 3 = 0 μg/ml IHF. Lane 4 = 0.8 μg/ml IHF.

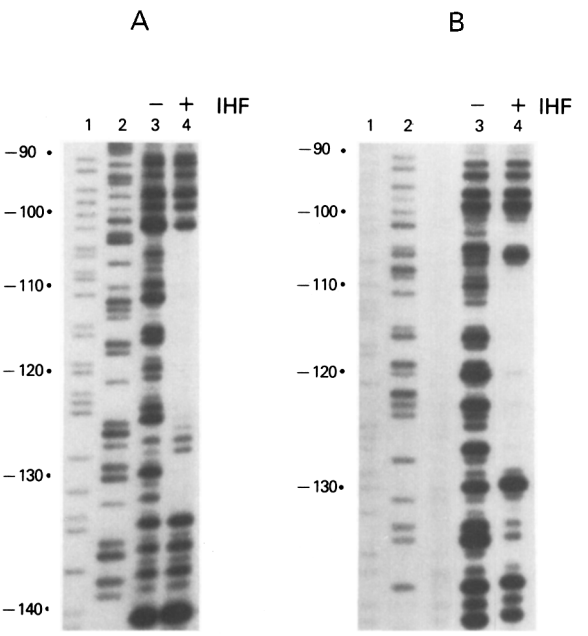

Figure 2. Protection Against DNAase Attack of DNA Fragments Containing the H1 Region of *attP* by IHF

In each panel, lane 1 contains products of an A+G sequencing reaction of the substrate fragment and lane 2 contains the products of a T+C sequencing reaction of the fragment. The sequence is numbered as described in Figure 1. The nucleotide sequence of this region is also presented in Figure 3. The interaction of IHF with the substrate fragment was examined by DNAase attack as described in Experimental Procedures. Lane 3 = 0 μg/ml IHF. Lane 4 = 0.8 μg/ml IHF.

(A) The top strand of H1. The same end-labeled DNA fragment described in Figure 1 was used.

(B) The bottom strand of H1. The *attP*-containing restriction fragment Hind III (−160) to Bam HI (+242) from pPH54 was 3′ end-labeled with ^{32}P at the Hind III site.

IHF binding sites H1, H2, and H′. Because our preparations of IHF are highly purified and because no other components of the lambda site-specific recombination machinery such as Int or a partner *att* site have been included in this experiment, we conclude that IHF is itself a specific DNA binding protein.

We also examined the ability of IHF to protect *attP* against digestion with DNAase. We find that IHF also protects H1, H2, and H′ from digestion by this attack reagent. The experiments displayed in Figure 2 present a detailed view of the protection against DNAase attack provided by IHF to the top (A) and bottom (B) strands of DNA from the H1 region; a summary of these experiments is shown in Figure 3B. The outermost positions that become inaccessible to nuclease in the presence of IHF define the minimal extent of the binding site. As indicated by the filled bar, the minimal extent of the IHF binding site in the top strand of the H1 region is from at least the junction between −133 and −132 to the junction between −107 and −106. The junctions between −135 and −134 and between −103 and −102 (indicated by the open circles) remain accessible to nuclease attack in the presence of IHF and thus lie outside the IHF binding site. We are unable to evaluate protection of the DNA that lies between the position of the open circles and the filled bar because DNAase does not efficiently attack these bonds even in the absence of IHF. The experiment shown in Figure 2 also reveals that some internucleotide junctions within the IHF binding site in the H1 region are accessible to DNAase attack even in the presence of IHF. The positions of these interior yet accessible junctions are indicated by the arrows in Figure 3B. As shown in Figure 2A, some of these bonds are similarly susceptible in the presence and absence of IHF (for example, between −128 and −127) whereas attack at other positions is enhanced by the presence of IHF (for example, between −129 and −128 and between −127 and −126).

Using DNAase as an attack reagent with a variety of *attP*-containing fragments, we similarly analyzed the interaction of IHF with the H2 and H′ regions (data not shown). The results of these experiments are presented in Figure 3B. The sizes and locations of H1, H2, and H′ are summarized in Figure 3A. The IHF binding sites in *attP* are not identical. They range in size from about 32 bp (H1) to 48 bp (H′). Moreover, the sites also differ in their apparent

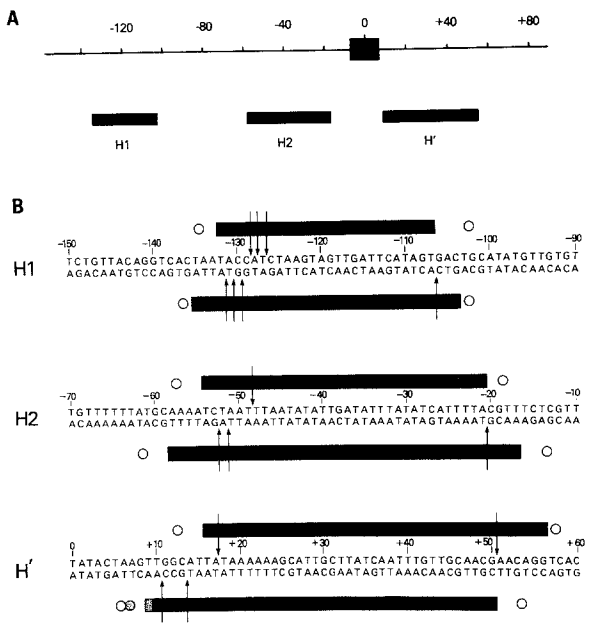

Figure 3. IHF Binding Sites in *attP*

(A) The positions of H1, H2, and H′. The linear map depicts the extent of a functional *attP* site (Hsu et al., 1980; Mizuuchi and Mizuuchi, 1980) and the relative size and position of the 16 bp core (Landy and Ross, 1977). The bars depict the sizes and positions of H1, H2, and H′ as defined by protection against DNAase attack by the presence of IHF.

(B) Sequences in H1, H2, and H′ regions of *attP* protected against DNAase attack. The bars indicate the sequences protected by IHF against DNAase attack; the procedures for aligning DNAase cleavages with the DNA sequence is as described in Ross et al. (1979). Bars drawn above the sequence indicate the extent of protection of the top strand and bars drawn below the sequence indicate the extent of protection of the bottom strand. The solid bars extend to the outermost nucleotide junctions that become inaccessible to DNAase in the presence of IHF. The stippled bar indicates a region of ambiguity for the position of the outermost nucleotide junction. The open circles indicate the nucleotide junctions closest to the protected region that are accessible in the presence of IHF. The stippled circles indicate a region of ambiguity for the location of the unprotected region. The arrows indicate positions within the IHF protected region that are accessible to DNAase even in the presence of IHF.

affinity for IHF. For example, we found that only the H2 site is protected at a low concentration of IHF and that only the H2 and H′ sites are protected in the presence of high concentrations of non-*att* DNA (data not shown). We have also observed that, at relatively high concentrations of IHF, an apparently nonspecific interaction of IHF with *attP* DNA occurs such that the susceptibility of the entire fragment to nuclease attack is reduced (data not shown). We note that because all of our protection experiments analyzing the interaction of IHF with *attP* utilized fragments that contain all three IHF binding sites, we have no information about the intrinsic affinity of each site for IHF or about the effect of one IHF binding site on another. However, as described below, we have identified single IHF binding sites on non-*att* DNA fragments, indicating that the presence of multiple IHF binding sites is not essential for the specific binding of IHF to DNA.

Using protection from nuclease attack as an assay we also examined the interaction of IHF with a fragment containing *attB*. We have not detected any specific binding of IHF to the region that constitutes a functional *attB* site although there is an IHF binding site, which we have not characterized in detail, about 50 bp to the right of the core (data not shown).

The Interaction of IHF with *attP* as Revealed by Protection from Chemical Attack

Another method by which the interaction of a protein with DNA can be evaluated is by examining the ability of the protein to alter the susceptibility of purines to methylation by dimethylsulfate (DMS) (Gilbert et al., 1976). DMS methylates the N7 position of guanine (in the major groove) and the N3 position of adenine (in the minor groove). An alteration in the susceptibility of a purine to methylation in the presence of a protein suggests a close proximity of the protein and the affected base. The experiments shown in Figure 4 display a detailed view of the methylation pattern of the top strand of *attP* in the H2 region (A) and a survey view of the methylation pattern of the bottom strand of *attP* (B). We observe that IHF protects some purines from methylation (at the positions marked by "–" in Figure 4) and enhances the methylation of others (at the positions marked by "+" in Figure 4).

The results of the experiments presented in Figure 4 and of other methylation experiments utilizing a variety of *attP*-containing fragments (data not shown) are summarized in Figure 5. In presenting the results of our experiments, we assigned a numerical value to each alteration (either protection or enhancement) in the methylation pattern provided by IHF which reflects both the degree and reproducibility of the alteration. The IHF-dependent alterations in purine methylation are clustered in specific regions of *attP*. Indeed, they lie in the H1, H2, and H′ regions of *attP* that were identified as sites of specific IHF–DNA interaction by our nuclease protection experiments. We note that the presence of IHF provides the most dramatic protection to adenine rather than guanine residues in *attP*. The positions marked by the arrows in Figure 4B identify pairs of adjacent adenine residues in the bottom strand of the H1 (at positions −118 and −117) and H2 (at positions −40 and −39) regions that are strongly protected against methylation by IHF.

The Interaction of IHF with non-*att* DNA

To characterize more fully the DNA binding properties of IHF, we also analyzed the interaction of IHF with non-*att* DNA fragments. We targeted our studies to two genes that appear to be regulated by IHF, the E. coli *hag* and lambda *cII* genes. The *hag* gene encodes flagellin, a major structural component of the flagellar filaments (Makela, 1964; Szekely and Simon, 1983). Several pieces of evidence suggest that IHF participates in regulation of the expression of this gene (Szekely et al., 1984) although the nature of the regulation (i.e., transcriptional or posttranscriptional) is not known. We find that IHF binds specifically

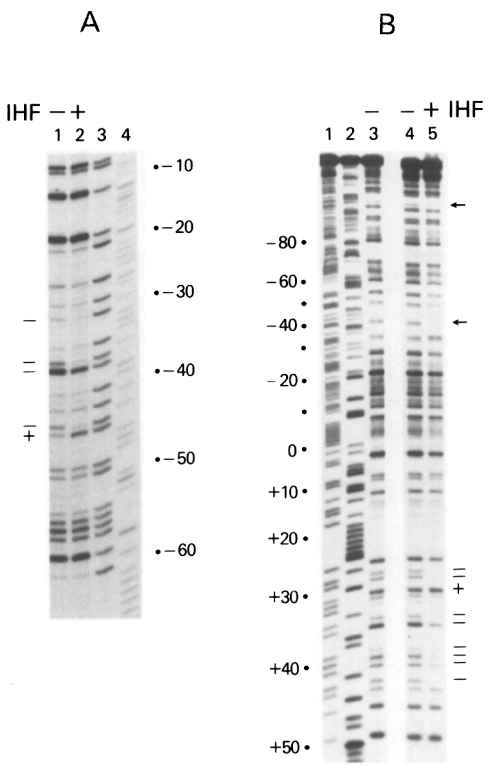

Figure 4. Protection Against Methylation by IHF

In each panel, the sequence is presented as described in Figure 1. "−" marks purines that are protected against methylation by the presence of IHF. "+" marks purines whose methylation is enhanced in the presence of IHF.

(A) The top strand of the H2 region. The methylation pattern of the uniquely end-labeled restriction fragment described in Figure 1 was determined in presence or absence of IHF as described in Experimental Procedures. Lane 1 = 0 μg/ml IHF. Lane 2 = 4.0 μg/ml IHF. Lane 3 contains the products of an A+G sequencing reaction of the fragment. Lane 4 contains the products of a T+C sequencing reaction of the fragment.

(B) The bottom strand of *attP*. The *attP*-containing fragment Hind III (−160) to Bam HI (+82) from pPH507 was 5′ end-labeled at the Bam HI site. The arrows mark the position of strongly protected pairs of A residues in H1 and H2 regions (see text). Lane 1 contains the products of an A+G sequencing reaction of the fragment. Lane 2 contains the products of a T+C sequencing reaction of the fragment. The methylation pattern of the fragment was determined in the presence and absence of IHF as described in Experimental Procedures. Lanes 3 and 4 = 0 μg/ml IHF. Lane 5 = 4.0 μg/ml IHF.

to a region centered about 20 bp upstream of the translational initiation codon of the *hag* gene (Figure 6A). The size of this IHF binding site is similar to the size of H1 in *attP*, about 33 bp and 32 bp, respectively. We also found that IHF protects this region from attack by NCS (data not shown). As summarized in Figure 6B, IHF also alters the methylation pattern of some purines in the region upstream of the *hag* gene. As we observed in *attP*, the most dramatic protection against methylation is provided to adenine residues. We note that specific binding of IHF to the *hag* site occurs under conditions (i.e., IHF and DNA concentrations) similar to those we have used to observe the specific binding of IHF and *attP*.

Lambda *cII* protein is a positive activator of functions essential to lysogenic development (reviewed in Wulff and Rosenberg, 1983). *himA* and *hip* mutants are defective in expression of the *cII* gene (Miller, 1981; Hoyt et al., 1982; Oppenheim et al., 1982). Using DNAase and NCS as attack reagents, we found that IHF binds specifically to a site centered about 25 bp upstream of the translational initiation codon of the *cII* gene (data not shown). A portion of the sequence protected by IHF is presented in Figure 9. The region of DNA protected from DNAase attack by IHF is about 33 bp in length, a size similar to the site of the H1 site in *attP* and the *hag* site. We found that IHF also alters the methylation pattern of several purines in this region upstream of the *cII* gene; the two most prominent positions of protection are indicated in Figure 9. The binding of IHF to the *cro-cII* site occurs under conditions similar to those we used to detect the binding of IHF to *attP* and to the *hag* site.

Taken together, these experiments demonstrate that IHF has the capacity to bind specifically to non-*att* DNA. They also demonstrate that the specific binding of IHF to DNA does not require the presence of multiple, distinct binding sites nor does the binding of IHF to a single site result in multiple regions of protection.

The Interaction of Mixtures of IHF and Int with *attP*

The experiment shown in Figure 7A displays the protection against DNAase attack provided to a segment of the P arm by a mixture of IHF and Int. We find that this mixture protects a region of DNA extending from about −145 to −95. The results of this experiment and other similar experiments (data not shown) are summarized in Figure 8A. The region of the distal segment of the P arm that is protected from DNAase attack by IHF + Int is, to a first approximation, the simple sum of the regions protected by each protein alone. This segment of the P arm contains the IHF H1 binding site and "arm-type" Int binding sites P1 and P2. Although these sites (as defined by protection experiments using the individual proteins) overlap slightly, we observe simultaneous occupation of all the sites in the presence of IHF + Int.

The experiment shown in Figure 7B displays the protection against DNAase attack provided to the top strand of the core and P′ regions of *attP* by a mixture of IHF and Int (lane 1), IHF (lane 3), and Int (lane 5). The results of this experiment are summarized in Figure 8B, together with a similar analysis of protection of the bottom strand. We observe that IHF + Int protects a large region extending from about −55 to about +90. The region protected by IHF + Int is, to a first approximation, the simple sum of the regions protected by the proteins individually. We observe that the IHF H2 and H′ sites as well as Int binding sites in the core and P′ arm are all protected by IHF + Int. It has previously been observed that Int alone binds to core and to P′ (Ross et al., 1979; Ross and Landy, 1983). However, as typified by the experiment shown in Figure 7B, lane 5, we consistently observe strong protection of only P′. At low Int concentrations, we usually observe only very modest protection of the core (not clearly evident in Figure 7);

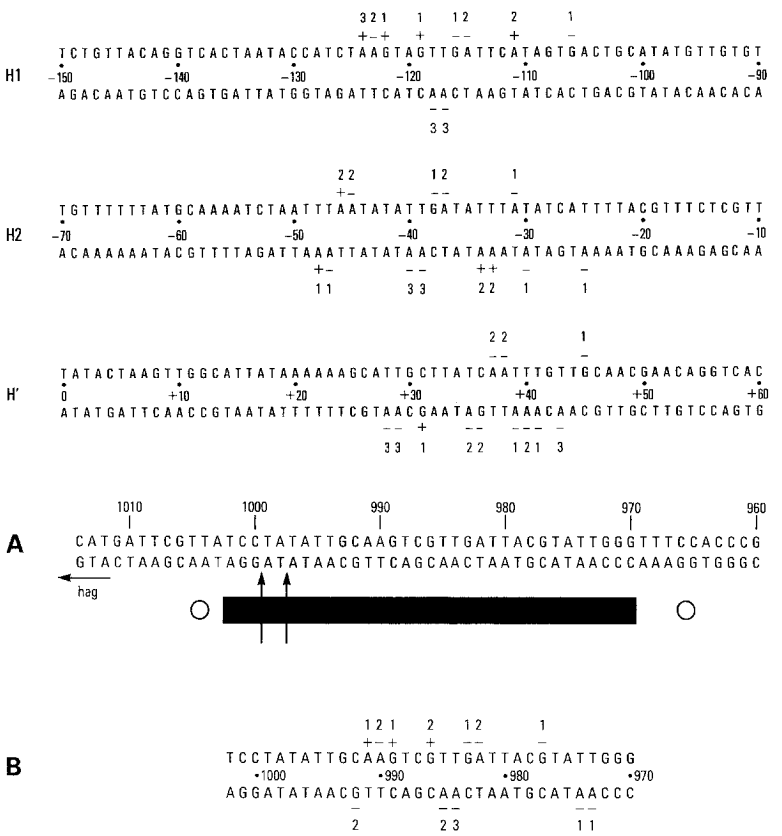

Figure 5. Purines in *attP* whose Methylation Is Altered by the Presence of IHF

Portions of sequence from *attP* are presented as described in Figure 1. "+" indicates a purine whose methylation by dimethylsulfate is enhanced by the presence of IHF. "−" indicates a purine that is protected from methylation by the presence of IHF. The numbers reflect the extent and consistency of the enhancement or protection observed: 3 = an alteration dramatic in extent and consistently observed; 2 = an alteration modest in degree and consistently observed; and 1 = an alteration small in degree and usually observed.

Figure 6. Interaction of IHF with a Region Adjacent to the *hag* Gene

The nucleotide sequence of the 5′ region of the *hag* gene is presented (Szekely and Simon, 1983). The horizontal arrow marks the ATG translational initiation codon of the *hag* gene.

(A) A region adjacent to the *hag* gene protected against DNAase attack by IHF. An approximately 145 bp restriction fragment extending from a Taq I site (at about 1085) to a Hinc II site (942) was 3′ end-labeled with ^{32}P at the Taq I site. The interaction of IHF with this fragment was analyzed by DNAase attack as described in Experimental Procedures. The results of these experiments are presented as described in Figure 3.

(B) Purines adjacent to the *hag* gene whose methylation is altered by IHF. Both the restriction fragment described in (A) and a similar fragment that was 5′ end-labeled with ^{32}P at the Taq I site were used. The interaction of IHF with these fragments was examined by methylation analysis as described in Experimental Procedures. The alterations in purine methylation observed are presented as described in Figure. 5.

at higher Int concentrations, the entire fragment is nonspecifically protected (data not shown). However, we find that addition of IHF to concentrations of Int that do not provide obvious core protection (lane 5) results in clear protection of the core region (lane 1). This result reveals that although Int has the capacity to recognize and bind to the core, this activity can be enhanced by the presence of IHF.

Discussion

Determinants of Specific IHF–DNA Interaction

Through characterization of five sites in *att* and non-*att* DNA whose susceptibility to nuclease or chemical attack is altered by the presence of IHF, we have determined that IHF is a specific DNA binding protein. How does IHF recognize these specific sites in DNA? Inspection of the nucleotide sequences and methylation patterns of the IHF binding sites reveals several common features. Each site contains at least one pair of adjacent adenine residues that are strongly protected from methylation by the presence of IHF. Indeed, the protection of these residues is the most dramatic and striking feature of the methylation pattern at each IHF binding site. In Figure 9, the sequences of the IHF binding sites are presented with their strongly protected adenines (enclosed by circles) aligned. This alignment reveals several other features common to the binding sites we have examined. Each IHF binding site contains a sequence that matches the sequence $\begin{matrix}\text{T}\cdot\text{PyAA}\cdots\text{PuTTG}\cdot\text{T}\\ \text{A}\cdot\text{PuTT}\cdots\text{PyAAC}\cdot\text{A}\end{matrix}$ exactly. Moreover, this conserved sequence lies at a similar position in each IHF binding site: the −7 position of this sequence (see Figure 9) is located 3–5 nucleotides from an outermost edge of the region protected from DNAase attack. The conservation of this sequence and its position suggests that it is an important component in specific interaction between IHF and DNA. Further evidence is provided by the fact that methylation of several purines within the conserved sequence is similarly altered by the presence of IHF at each binding site. First methylation of the A at position −4 of the conserved sequence is often (four of five sites) enhanced. Second, the A at position −3 is always protected from methylation.

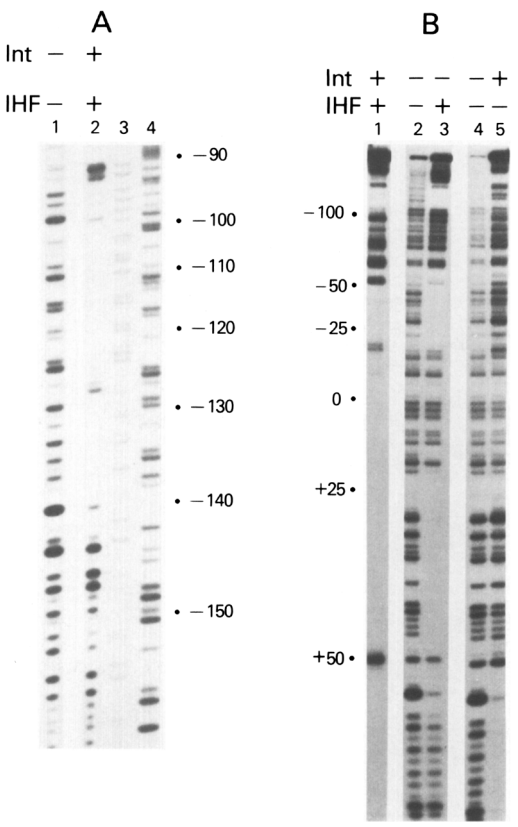

Figure 7. Protection against DNAase Attack by a Mixture of IHF and Int

In each panel, the sequence is numbered as described in Figure 1.

(A) The distal region of the P arm of *attP*. The same restriction fragment described in Figure 1 was used. Its susceptibility to attack by DNAase in the presence or absence of a mixture of IHF and Int was examined as described in Experimental Procedures. Lane 1 = 0 μg/ml IHF and 0 μg/ml Int. Lane 2 = 0.8 μg/ml IHF + 3.2 μg/ml Int. Lane 3 contains the products of an A+G sequencing reaction of the fragment. Lane 4 contains the products of a T+C sequencing reaction of the fragment.

(B) The core and P′ arm regions of *attP*. The same restriction fragment described in Figure 2B 3′ end-labeled at the Bam HI site was used. The susceptibility of this fragment to DNAase attack was examined as described in Experimental Procedures. Lane 1 = 0.8 μg/ml IHF + 1.6 μg/ml Int. Lane 2 = 0 μg/ml IHF + 0 μg/ml Int. Lane 3 = 0.8 μg/ml IHF + 0 μg/ml Int. Lane 4 = 0 μg/ml IHF + 0 μg/ml Int. Lane 5 = 0 μg/ml IHF + 1.6 μg/ml Int.

Third, methylation of the purine at position +1 is often altered. In the two sites where this position is G, its methylation is enhanced and in two of the three sites where this position is A, it is protected from methylation. Fourth, the As at position +2 and +3 are protected from methylation. These are the strongly protected residues used to align the sites. Fifth, the G at position +4 is protected from methylation.

In addition to the conserved sequence $\begin{matrix}\text{T·PyAA···PuTTG·T}\\ \text{A·PuTT···PyAAC·A}\end{matrix}$ just considered, the alignment of the IHF binding sites sequence presented in Figure 9 reveals several other conserved features. In four of the five binding sites, position +5 is an A/T base pair and in the remaining site, the *cro-cII* site, this position is a T/A base pair. Furthermore, we find that the A at position +5 is always protected from methylation by IHF, suggesting that position +5 may also be important to specific IHF–DNA interaction. Thus we propose that the sequence $\begin{matrix}\text{T·PyAA···PuTTGaT}\\ \text{A·PuTT···PyAACtA}\end{matrix}$, where the letters in lower case indicate an important but non-unique base, contains the critical determinant for IHF binding. In addition, several other positions near the conserved sequence share similar features: an A/T or T/A base pair is present in all five sites at positions −12, −8, +7, and +11; in four of five sites a Py/Pu pair is at positions −10 and +10 and a Pu/Py pair is at position +12. We do not know if these additional features are important to the specific binding of IHF to DNA. Such insight must await analysis of variants of the IHF binding sites we have identified and characterization of other IHF binding sites. Leong and Landy (personal communication) have identified specific IHF binding sites in the DNA segments that participate in site-specific recombination promoted by the phages ϕ80 and P22 and their results are in general agreement with ours. However, we note that several of the IHF sites they have characterized differ at positions −7 and +1 from the conserved sequence we have identified.

The strongest and most consistent modifications in the methylation pattern promoted by IHF that we have observed occur at adenine residues (for example, at positions −3, +2, and +3 of the IHF consensus sequence). The changes in G methylation are modest (for example, at position 4). This suggests the possibility that, in contrast to the majority of other specific DNA binding proteins that have been characterized (Pabo and Sauer, 1984), critical contacts between IHF and DNA may occur in the minor groove of the helix. Such alteration in the methylation pattern could also reflect a change in DNA conformation which renders this site inaccessible to methylation (Frederick et al., 1984).

We also note that the IHF consensus sequence contains some dyad symmetry. Specifically, the center of the consensus sequence (from position −5 to position +4) is an imperfect palindrome containing 5′-PyAA...PuTTG-3′ on the top strand and 5′-CAAPy...TTPu-3′ on the bottom strand. Although each protomer of IHF is itself asymmetric, containing one *himA* polypeptide and one *hip* polypeptide (Nash and Robertson, 1981), these two subunits are clearly related (Flamm and Weisberg, 1984; Miller, 1984). Perhaps the central palindrome of the consensus sequence is recognized by the homologous segments of the IHF subunits while the nonsymmetrical features of the sequence are recognized by the unique segments of each subunit.

Given the small size of IHF (M_r ~20,000) we are struck by the large size of the IHF binding sites; four of the five regions of protection are 30–40 bp long and the H′ region of *attP* is even larger (about 50 bp). We wonder if more than one IHF protomer is involved in each binding region, or if wrapping of DNA around the rather elongated protomer of IHF (Nash and Robertson, 1981) can account for the size of the footprint. The H′ site may be larger than the others because it contains two IHF recognition se-

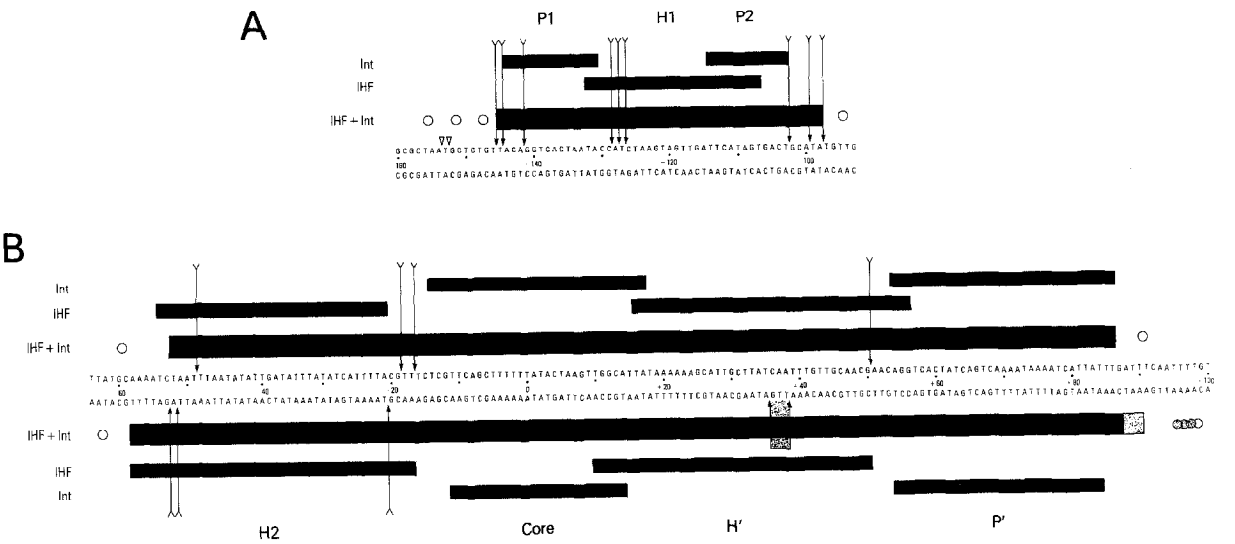

Figure 8. The Interaction of IHF + Int with *attP*

In both panels, the sequence is presented as described in Figure 1. The bars labeled "Int" indicate the sequences protected against NCS digestion by Int and are taken from the literature (Ross et al., 1979; Hsu et al., 1980). The bars labeled "IHF" indicate the sequences protected against DNAase attack by IHF (Figure 3). The bars labeled "IHF + Int" indicate the sequences protected against DNAase attack by a mixture of IHF and Int. The symbols marking the protected regions, its boundaries and accessible sites are as in Figure 3. The stippled segment bounded by the vertical arrows indicates a DNAase accessible position that could not be more precisely mapped.

(A) The region of the P arm distal from the core.

(B) The core and P′ arm regions of *attP*.

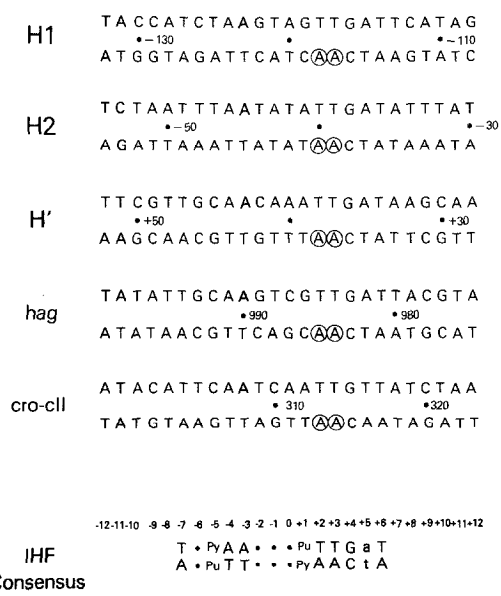

Figure 9. A Consensus Sequence Shared by IHF Binding Sites

Portions of the sequences to which IHF binds specifically are shown. In addition to the H1, H2, and H′ sequences of *attP* (Figure 3) and the *hag* sequence (Figure 6), a portion of the sequence of the *cro-cII* region of phage lambda (Rosenberg et al., 1978) protected by IHF is presented (see text). The circles indicate adjacent A residues which are strongly protected against methylation by the presence of IHF. A consensus sequence for specific IHF–DNA interaction is presented. The base pairs in upper case letters are those conserved in all five IHF binding sites. The base pair in lower case letters may also be important even though it is conserved in only four of the five sites (see text).

quences. As presented in Figure 9, the H′ region contains a sequence extending from +47 to +34 that exactly matches the IHF consensus sequence. The methylation pattern of this sequence is altered by IHF in a manner similar to alterations observed at the other IHF binding sites. However, the H′ region also contains another sequence extending from +19 to +32 that matches the IHF consensus sequence at seven of the ten specified positions. Moreover, the adjacent adenines at +28 and +29 (at positions +2 and +3 in this second putative IHF recognition sequence) are strongly protected from methylation by IHF, a feature we have argued is characteristic of specific IHF–DNA interaction. Furthermore, position −7 of this sequence lies three nucleotides from an outermost edge of the protected region. Thus the larger size of the H′ protected region could reflect the binding of an additional unit of IHF. However, the binding of this additional unit may be unusual since methylation at the −4, −3, +1, and +4 positions in this second putative IHF recognition sequence is not altered by IHF.

The Role of IHF in Recombination

Critical steps in recombination are the recognition of *att* sites by the recombination proteins; the synapsis of *att* sites to align them; and the breakage of the parental DNA strands and their rejoining in novel combinations to produce recombinant molecules. Int has the capacity to execute each of these steps (Lange-Gustafson and Nash, 1984) so that the role of IHF is that of an accessory protein

that enhances Int's ability to carry out recombination. Our analysis has revealed that IHF, a host-encoded protein, interacts with multiple, specific sites in *attP,* the phage partner in recombination, rather than with *attB,* the bacterial partner. Our analysis has also revealed the highly complex character of *attP*. This approximately 240 bp DNA segment contains seven distinct sites of specific protein–DNA interaction: *attP* includes three IHF binding sites and four Int binding sites (see Figure 8). Interactions between proteins bound to these sites may be important to recombination. For example, we have found that IHF can promote the specific binding of Int to the core region. This extends the earlier observation using DNA filter binding as an assay that IHF promotes the interaction of Int with *att* sites (Nash and Robertson, 1981). These findings suggest that IHF may promote recombination by stimulating Int to recognize the core. This activity could also underlie a role of IHF in excisive recombination (Abremski and Gottesman, 1982). However, several observations suggest that this is not the only major role of IHF in recombination. First, Int can bind to the core in the absence of IHF (Ross et al., 1979; Ross and Landy, 1983). Second, efficient recombination requires IHF even at Int concentrations that might be expected to compensate for the modest affinity of Int for the core (Kikuchi and Nash, 1978; Nash and Robertson, 1981).

Several observations suggest that higher order structures involving *att* sites and recombination proteins are critical components of recombination (Better et al., 1982; Pollock and Nash, 1983) and we are attracted to the hypothesis that a major role for IHF in recombination is to assist Int in the formation of such structures. Condensed structures containing Int and *attP* (intasomes) have been observed by electron microscopy, although those formed in the absence of IHF show no capacity to interact with *attB* (Better et al., 1982). Perhaps IHF is needed to make a functional intasome. IHF could facilitate the formation of functional higher order structures by critical protein–protein contacts or by introducing bends or kinks into DNA (Wu and Crothers, 1984; Frederick et al., 1984).

The Role of IHF in Gene Expression and Its Relation to DNA Binding Protein II

Our finding that IHF binds to DNA near genes that show altered regulation in *himA* and *hip* mutants is provocative. Indeed, examination of the sequences adjoining other genes whose expression is altered in *himA* and *hip* mutants reveals the presence of sequences related to the IHF consensus sequence (Goosen and van de Putte, 1984; Friden et al., 1984; Szekely et al., 1984; Flamm and Weisberg, 1984). However, it is important to realize that there is no evidence (other than their tantalizing location) that these putative IHF binding sites are regulatory elements. Moreover, in vivo studies of the regulation of *cII* by IHF suggest that IHF acts at a posttranscriptional step (Hoyt et al., 1982; A. Oppenheim, personal communication), suggesting interaction with RNA rather than DNA. We also note that we failed to detect the binding of IHF (as assayed by protection from nuclease or chemical attack) to DNA segments containing the regulatory regions of the Salmonella typhimurium HU gene (Craig, Nash, and Simon, unpublished observations) or the E. coli *ilvB* gene (Craig, Nash, and Freundlich, unpublished observations), two loci whose expression is altered in *himA* and *hip* mutants (Szekely et al., 1984; Friden et al., 1984) although these regions do contain sequences related to the IHF consensus sequence. Thus our identification of IHF binding sites does not suggest a simple model for the mechanism by which IHF may participate in gene regulation. However, it is tempting to speculate that some of the actions that we suggest may underlie the role(s) of IHF in site-specific recombination (such as promoting a transition in the structure of DNA or direct interaction with another DNA binding protein) may also underlie the role of IHF in gene expression.

The predicted amino acid sequence for the genes that encode the subunit of IHF suggest that IHF is related to the widely distributed family of DNA binding protein that include the E. coli protein HU (Flamm and Weisberg, 1984; Miller, 1984). Moreover, the crystal structure of the B. stearothermophilus member of this family has been determined (Tanaka et al., 1984) and many residues that are involved in critical structures in this protein are conserved in the products of he *himA* and *hip* genes (R. Weisberg, personal communication). Although HU and its relatives are thought to interact with DNA in a nonspecific manner, our observations on IHF suggest that these proteins might also display specific interactions with DNA. We make this suggestion with reservations since the relationship between IHF and HU may not be too intimate. HU protein will not replace IHF in recombination reactions (Nash and Robertson, 1981) and footprinting experiments with the purified protein (a gift from A. Kornberg) show no specific protection of *attP* (J. Gardner, C. Robertson, and H. Nash, unpublished observations).

Experimental Procedures

Preparation of Plasmid DNA

Plasmid-containing derivatives of E. coli strains 204 (Kikuchi and Nash, 1978) were grown and supertwisted DNA was prepared from them as previously described (Nash and Robertson, 1981). The plasmids used (Hsu et al., 1980) were pPH54, in which an *attP*-containing fragment (−160 to +242) derivatized with a Hind III linker at −160 replaces the Hind III–Bam HI fragment of pBR322, and pPH507, in which an *attP*-containing fragment (−160 and +82) derivatized with a Hind III linker at −160 and a Bam HI linker at +82 replaces the Hind III–Bam HI fragment of pBR322.

Preparation of End-Labeled DNA Fragments

^{32}P end-labeled restriction fragments were prepared from supertwisted plasmid DNA according to the procedures of Maxam and Gilbert (1980) with modifications as indicated. Supertwisted plasmid DNA was first digested with a restriction enzyme, thus generating the end to be labeled. Restriction enzymes were obtained from Bethesda Research Laboratories or New England Biolabs. After removal of endogenous terminal phosphates with calf intestine alkaline phosphatase (Boehringer Mannheim), the digest was 5′ end-labeled with ^{32}P-γ-ATP (New England Nuclear) using polynucleotide kinase. Alternatively, the digest was 3′ end-labeled with ^{32}P-α-dCTP or ^{32}P-α-dATP (New England Nuclear) using the Klenow fragment of E. coli DNA polymerase (Boehringer Mannheim). A second restriction digestion was performed and the fragments separated on 6% or 7.5% polyacrylamide

gels. Fragments were eluted from the gel (without the addition of tRNA to Maxam-Gilbert extraction buffer) and concentrated by ethanol precipitation.

Preparation of IHF and Int

Purified IHF was prepared through step V as previously described (Nash and Robertson, 1981). Immediately before use, dilutions of IHF were prepared in 50 mM Tris-HCl (pH 7.4), 10% (v/v) glycerol, 800 mM KCl, and 2 mg/ml bovine serum albumin (Pentex). Purified Int was prepared as previously described (Kikuchi and Nash, 1978; Nash, 1983). Immediately before use, dilutions of Int were prepared in 50 mM Tris-HCl (pH 7.4), 10% (v/v) glycerol, 600 mM KCl, 1 mM EDTA, and 2 mg/ml bovine serum albumin.

Attack with Neocarzinostatin

Approximately 0.01 to 0.10 pmole of uniquely end-labeled restriction fragment was incubated with IHF in a 100 μl reaction mixture containing 52 mM Tris-HCl (pH 7.4), 70 mM KCl, 10% (v/v) glycerol, 1.1 mM EDTA, 10 mM β-mercaptoethanol, and 200 μg/ml bovine serum albumin. The reaction mixture was incubated for 20 min at 25°C; 5 μl of 40 mM Tris-HCl (pH 7.4), 8 mM β-mercaptoethanol, and 2.8 mM NaAc (pH 5.0) containing 0.9 U of neocarzinostatin (Bristol Laboratories) was added and incubation continued for 2 min. One hundred microliters of 0.6 M NH_4Ac, 0.1 M EDTA, and 20 μg/ml sonicated calf thymus DNA was added and the samples put on ice. The samples were then extracted with phenol/chloroform/isoamyl alcohol (1:1:0.04) that had been saturated just before use with 50 mM Tris-HCl (pH 8.0)–1.0 mM EDTA; subsequent extractions were performed with chloroform and ether. DNA was precipitated from the sample by the addition of 2.5 volumes of ethanol. The resulting pellet was resuspended with 0.3 M NaAc and reprecipitated by the addition of 2.5 volumes of ethanol. The resulting pellet was washed with ethanol and then dried under vacuum.

Attack with DNAase

Approximately 0.01 to 0.10 pmole of uniquely end-labeled restriction fragment was incubated with IHF and/or Int in a 100 μl reaction mixture containing 52 mM Tris-HCl (pH 7.4), 70 mM KCl, 10% (v/v) glycerol, 1.1 mM EDTA, 1.0 mM β-mercaptoethanol, 7.0 mM $MgCl_2$, 3.0 mM $CaCl_2$, and 200 μg/ml bovine serum albumin. After 20 min at 25°C, 5.0 (or sometimes 2.5) μl of 0.5 μg/ml pancreatic DNAase I (Worthington) in 50 mM Tris-HCl (pH 7.4), 1.0 mM EDTA, and 1.0 mM β-mercaptoethanol was added and incubation continued for 2 min. The reactions were then stopped and processed as described above.

Attack with Dimethylsulfate

Procedures were adopted from chemical sequence methods (Maxam and Gilbert, 1977). Approximately 0.1 to 1.0 pmole/ml of uniquely end-labeled restriction fragment was incubated with IHF in a 100 μl (or sometimes 200 μl) reaction mixture containing 5 mM Tris-HCl (pH 7.4), 2.0 mM Tris-HCl (pH 8.0), 1% (v/v) glycerol, 0.1 mM EDTA, 70 mM KCl, 200 μg/ml bovine serum albumin, 50 mM Na-cacodylate (pH 8.0), and 10 mM $MgCl_2$ for 20 min at 25°C. A 1/100 volume of 2.7 M dimethylsulfate (Aldrich Chemical Co.) in 55.6 mM Na-cacodylate (pH 8.0), 2.2 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, and 11.1 mM $MgCl_2$ was added to the reaction mixture and incubation continued for 1 min. A 1/4 volume of a 4°C solution containing 1.0 M Tris-Ac (pH 7.5), 1.5 M Na-Ac, 50 mM Mg-Ac, 1.0 mM EDTA, 1.0 M β-mercaptoethanol, and 200 μg/ml tRNA was added. DNA was precipitated from the sample by the addition of 3 volumes of −20°C ethanol. The resulting pellet was resuspended in 0.3 M Na-Ac and reprecipitated by the addition of 2.5 volumes of ethanol. The resulting pellet was washed with ethanol and then dried under vacuum. The dried pellet was resuspended in 200 μl 50 mM Tris-HCl (pH 8.0) and 1 mM EDTA, extracted with phenol as described above, and concentrated by ethanol precipitation. The dried pellet was resuspended with 20 μl of 20 mM KPO_4 (pH 7.0) and 1 mM EDTA, incubated for 10 min at 90°C and placed on ice. Two microliters of 1.0 M NaOH was added, the sample incubated for 5 min at 90°C and placed on ice. One hundred and eighty microliters of 10 mM Tris-HCl (pH 8.0) was added and the DNA concentrated by ethanol precipitation.

Preparation of DNA Sequence Markers and Electrophoresis Conditions

A+G and T+C DNA sequencing reactions were carried out as described by Maxam and Gilbert (1980). Hydrazine was used to generate fragments broken at T or C residues and pyridinium formate was used to generate fragments broken at A or G residues.

Electrophoretic analysis was carried out by resuspending the products of DNA sequencing reactions or the dried pellets from protection reactions with 80% (v/v) formamide, 50 mM Tris-HCl, 50 mM borate, 1.4 mM Na_2 EDTA, 0.1% (w/v) xylene cyanol, and 0.1% (w/v) bromphenol blue. After denaturation by incubation at 90°C followed by chilling in ice water, the samples were electrophoresed through 8% (w/v) acrylamide (1:20 cross-linked)–8.3 M urea gels. The electrophoresis buffer was 90 mM Tris-HCl, 89 mM borate, and 2.5 mM Na_2 EDTA. After electrophoresis, the gels were exposed to Kodak XAR-2 film at −70°C with intensifying screens.

Acknowledgments

We thank M. Feiss, M. Freundlich, H. Miller, M. Rosenberg, A. Landy, M. Simon, and their coworkers for helpful discussions and for communication of results prior to publication. We are especially grateful to M. Freundlich, M. Rosenberg, and M. Simon for supplying us with *ilv, cro-cII, hag,* and *H2*-containing restriction fragments. We thank K. McKenney for help with the computer analysis of recognition sequences and R. Weisberg for his comments on this manuscript. Carol Robertson provided excellent technical assistance. N. L. C. was supported in part by National Research Service Award GM07498.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 30, 1984; revised October 5, 1984

References

Abremski, K., and Gottesman, S. (1982). Purification of bacteriophage λ *xis* gene product required for λ excisive recombination. J. Biol. Chem. *257,* 9658–9662.

Better, M., Lu, C., Williams, R., and Echols, H. (1982). Site-specific DNA condensation and pairing mediated by the Int protein of bacteriophage λ. Proc. Nat. Acad. Sci. USA *79,* 5837–5841.

Craig, N., and Nash, H. (1983). The mechanism of phage λ site-specific recombination: site-specific breakage of DNA by Int topoisomerase. Cell *35,* 795–803.

Flamm, E., and Weisberg, R. (1984). Primary structure of the *hip* gene of *E. coli* and of its product, the β subunit of integration host factor. J. Mol. Biol., in press.

Frederick, C., Grable, J., Melia, M., Samundi, C., Jen-Jacobson, L., Wang, B-W., Greene, P., Boyer, H., and Rosenberg, J. (1984). Kinked DNA in crystalline complex with *Eco*RI endonuclease. Nature *309,* 327–331.

Friden, P., Voelkel, K., Sternglanz, R., and Freundlich, M. (1984). Reduced expression of the isoleucine and valine enzymes in integration host factor mutants of *Escherichia coli.* J. Mol. Biol. *172,* 484–489.

Friedman, D., Olson, E., Carver, D., and Gellert, M. (1984). Regulation of mu transposition. I. Localization of the presumed recognition sites for HimD and Ner functions controlling bacteriophage mu transcription. J. Bacteriol. *157,* 484–489.

Galas, D., and Schmitz, A. (1978). DNase footprinting: a simple method for the detection of protein–DNA binding specificity. Nucl. Acids Res. *5,* 3157–3170.

Gilbert, W., Maxam, A., and Mirzabekov, A. (1976). Contacts between the *lac* repressor and DNA revealed by methylation. In Control of Ribosome Synthesis, The Alfred Benson Symposium IX, N. Kjeldgaard and O. Maale, eds. (Copenhagen:,Munksgaard), pp. 139–148.

Goosen, N., and van de Putte, P. (1984). Synergistic effect of *him*A and *gyr*B mutations: evidence that Him functions control expression of *ilv* and *xyl* genes. Gene, in press.

Hoyt, M., Knight, D., Das, A., Miller, H., and Echols, H. (1982). Control of phage lambda development by stability and synthesis of cII protein: role of the viral cII and host *hfl*A, *him*A and *him*D genes. Cell *31,* 565–573.

Hsu, P-L., Ross, W., and Landy, A. (1980). The phage λ *att* site: functional limits and interaction with Int protein. Nature *285,* 85–91.

Kikuchi, Y., and Nash, H. (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.

Kikuchi, A., Flamm, E., and Weisberg, R. (1984). An *E. coli* mutant unable to support site-specific recombination of bacteriophage λ. J. Mol. Biol., in press.

Landy, A., and Ross, W. (1977). Viral integration and excision: structure of the lambda *att* sites. Science *197,* 1147–1160.

Lange-Gustafson, B., and Nash, H. (1984). Purification and properties of Int-h, a variant protein involved in site-specific recombination of bacteriophage lambda. J. Biol. Chem. *259,* 12724–12732.

Makela, P. (1964). Genetic homologies between flagellar antigens of *Escherichia coli* and *Salmonella abony.* J. Gen. Microbiol. *35,* 503–510.

Maxam, A., and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA *74,* 560–564.

Maxam, A., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. *65,* 499–560.

Miller, H. I. (1984). Primary structure of the *himA* gene of *E. coli:* homology with DNA binding protein HU and association with the phenylalanyl tRNA synthetase operon. Cold Spring Harbor Symp. Quant. Biol. *49,* in press.

Miller, H. (1981). Multi-level regulation of bacteriophage λ lysogeny by the E. coli *himA* gene. Cell *25,* 269–276.

Miller, H., and Friedman, D. (1980). An E. coli gene product required for λ site-specific recombination. Cell *20,* 711–719.

Miller, H., Kirk, M., and Echols, H. (1981). SOS induction and autoregulation of the *him*A gene for site-specific recombination in *Escherichia coli.* Proc. Nat. Acad. Sci. USA *78,* 6754–6760.

Mizuuchi, M., and Mizuuchi, K. (1980). Integrative recombination of bacteriophage λ: extent of the DNA sequence involved in attachment site function. Proc. Nat. Acad. Sci. USA *77,* 3220–3224.

Nash, H. (1981). Integration and excision of bacteriophage λ: the mechanism of conservative site-specific recombination. Ann. Rev. Genet. *15,* 143–167.

Nash, H. (1983). Purification and properties of the bacteriophage lambda Int protein. Meth. Enzymol. *100,* 210–216.

Nash, H., and Robertson, C. (1981). Purification and properties of the *E. coli* protein factor required for λ integrative recombination. J. Biol. Chem. *256,* 9246–9253.

Oppenheim, A., Gottesman, S., and Gottesman, M. (1982). Regulation of bacteriophage lambda *int* gene expression. J. Mol. Biol. *158,* 327–346.

Pabo, C., and Sauer, R. (1984). Protein–DNA recognition. Ann. Rev. Biochem. *53,* 293–321.

Pollock, T., and Nash, H. (1983). Knotting of DNA caused by a genetic rearrangement. Evidence for a nucleosome-like structure in site-specific recombination of bacteriophage lambda. J. Mol. Biol. *170,* 1–18.

Rosenberg, M., Court, D., Shimakake, H., Brady, C., and Wolff, D. (1978). The relationship between function and DNA sequence in an intercistronic regulatory region in phage λ. Nature *272,* 414–423.

Ross, W., and Landy, A. (1982). Bacteriophage λ Int protein recognizes two classes of sequence in the phage *att* site: characterization of arm-type sites. Proc. Nat. Sci. USA *79,* 7724–7728.

Ross, W., and Landy, A. (1983). Patterns of λ Int recognition in the regions of strand exchange. Cell *33,* 261–272.

Ross, W., Landy, A., Kikuchi, Y., and Nash, H. (1979). Interaction of Int protein with specific sites on λ *att* DNA. Cell *18,* 297–307.

Szekely, E., and Simon, M. (1983). DNA sequence adjacent to flagellar genes and evolution of flagellar-phage variation. J. Bacteriol. *155,* 74–81,

Szekely, E., Miller, H., and Simon, M. (1984). Regulation of flagellin gene expression by integration host factor. J. Bacteriol., in press.

Tanaka, I., Appelt, K., Dijk, J., White, S. W., and Wilson, K. S. (1984). 3-A resolution structure of a protein with histone-like properties in prokaryotes. Nature *310,* 376–381.

Weisberg, R., and Landy, A. (1983). Site-specific recombination in phage lambda. In Lambda II, R. Hendrix, J. Roberts, F. Stahl, and R. Weisberg, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 221–250.

Wu, H-M., and Crothers, P. (1984). The locus of sequence-directed and protein-induced DNA bending. Nature *308,* 509–513.

Wulff, D., and Rosenberg, M. (1983). Establishment of repressor synthesis. In Lambda II, R. Hendrix, J. Roberts, F. Stahl, and P. Weisberg, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Labortory), pp. 53–74.