The EMBO Journal vol.11 no.7 pp.2695-2705, 1992

Sequence-specific interaction of the *Salmonella* Hin recombinase in both major and minor grooves of DNA

Kelly T.Hughes, Peter C.W.Gaines[1], Joyce E.Karlinsey, Ravi Vinayak[2] and Melvin I.Simon[3]

Department of Microbiology SC-42, University of Washington, Seattle, WA 98195, [1]Department of Biology, KBT 1126, Yale University, New Haven, CT 06510, [2]Applied Biosystems, Inc., 850 Lincoln Center Drive, Foster City, CA 94404 and [3]Division of Biology 147-75, California Institute of Technology, Pasadena, CA 91125, USA

Communicated by H.Eisen

The Hin recombinase of *Salmonella* catalyzes a site-specific recombination event which leads to flagellar phase variation. Starting with a fully symmetrical recombination site, *hixC*, a set of 40 recombination sites which vary by pairs of single base substitutions was constructed. This set was incorporated into the *Salmonella*-specific bacteriophage P22 based challenge phage selection and used to define the DNA sequence determinants for the binding of Hin to DNA *in vivo*. The critical sequence-specific contacts between a Hin monomer and a 13 bp *hix* half-site are at two T:A base pairs in the major groove of the DNA which are separated by one base pair, and two consecutive A:T contacts in the minor groove. The base substitutions in the major groove recognition portion which were defective in binding Hin still retained residual binding capability *in vivo*, while the base pair substitutions affecting the minor groove recognition region lost all *in vivo* binding. Using *in vitro* binding assays, Hin was found to bind to *hix* symmetrical sites with A:T base pairs or I:C base pairs in the minor groove recognition sequences, but not to G:C base pairs. In separate *in vitro* binding assays, Hin was equally defective in binding to either a G:C or a I:C contact in a major groove recognition sequence. Results from *in vitro* binding assays to *hix* sites in which 3-deazaadenine was substituted for adenine are consistent with Hin making a specific contact to either the N3 of adenine or O2 of thymine in the minor groove within the *hix* recombination site on each symmetric half-site. These results taken with the results of previous studies on the DNA binding domain of Hin suggest a sequence-specific minor groove DNA binding motif.

Key words: DNA base analogs/DNA-protein interactions/Hin recombinase/P22 challenge phage

Introduction

Flagellar phase variation refers to the alternate expression of two antigenically distinct flagellar filament proteins (Stocker, 1949). This alternate expression results from a reversible inversion of the *Salmonella* chromosome which is catalyzed by a site-specific DNA recombination mechanism (Glasgow *et al.*, 1989b) (Figure 1). The Hin recombinase recognizes and binds to two 26 bp homologous sequences, *hixL* and *hixR*, that flank the invertible segment (Johnson and Simon, 1985). In the presence of a third DNA site, the recombinational enhancer sequence (Johnson and Simon, 1985), and the Fis protein which interacts directly with the recombinational enhancer element (Bruist *et al.*, 1987a; Johnson *et al.*, 1986, 1987), a specific protein-DNA complex is formed (Johnson and Bruist, 1989; Heichman and Johnson, 1990). This complex is thought to form through protein-protein interactions between Hin molecules bound to the *hix* sites and Fis molecules bound to the enhancer element (Heichman and Johnson, 1990). This specific protein-DNA complex traps negative modes in the DNA so that subsequent DNA cleavage and strand rotation results in loss of supercoils in the recombinant product (Boocock *et al.*, 1986, 1987). This was first demonstrated in the Tn*3* resolvase site-specific recombinase system (Wasserman *et al.*, 1985).

Binding of the Hin recombinase to the *hix* recombination sites has been extensively characterized both *in vivo* and *in vitro*. The DNA binding domain of Hin comprises the C-terminal 52 amino acids of the 190 amino acid protein (Bruist *et al.*, 1987b). Methylation inference assays suggest that Hin interacts with the *hix* sites in both the major and minor grooves of the DNA (Glasgow *et al.*, 1989a). Covalent attachment of Fe-EDTA cleaving moieties to different positions within the binding domain followed by mapping of the cleavage patterns with high resolution gel electrophoresis was used to determine the organization of the folded 52 amino acid protein domain with respect to the DNA (Sluka *et al.*, 1987, 1990; Mack *et al.*, 1990). Results of these studies support a model of a helix-turn-helix motif near the C-terminus of the DNA binding domain contacting *hix* sequences in the major groove while an Arg-Pro-Arg sequence recognizes DNA sequences in the minor groove of the *hix* site. The binding of Hin to the various *hix* sites was analyzed *in vivo* using the *Salmonella* bacteriophage P22 based challenge phage selection devised to characterize the binding of sequence-specific DNA binding proteins (Hughes *et al.*, 1988). These studies imply that there are critical DNA base contacts between Hin and the DNA, both in the presumed major groove recognition region of the *hix* half-site and in the presumed minor groove recognition region. In this manuscript we report *in vivo* and *in vitro* results which characterize the sequence-specific recognition of DNA by Hin. These results support the notion that specific interactions between Hin recombinase and the minor groove of the *hix* DNA sites play a substantial role in overall binding in addition to contacts in the major groove.

Because sequence-specific recognition of DNA by proteins is a fundamental aspect of many biological processes, it has been studied in great detail using genetic, biochemical, spectral and crystallographic methods. In prokaryotes, early crystal structures of coliphage λ repressor (Lewis *et al.*, 1982; Pabo and Lewis, 1982). λ Cro protein (Anderson

et al., 1981), *Escherichia coli* CAP protein (McKay and Steitz, 1981; KcKay *et al.*, 1982; Steitz *et al.*, 1982) and Trp repressor (Schevitz *et al.*, 1985) revealed the well known helix−turn−helix major groove DNA binding motif (Sauer *et al.*, 1982). A very different prokaryotic DNA binding motif was uncovered in the crystal structure of *Eco*R1 restriction endonuclease (McClarin *et al.*, 1986). As three-dimensional structures of sequence-specific DNA binding proteins from eukaryotic systems have been reported it has become clear that nature has developed a variety of mechanisms through which proteins can recognize specific DNA sequences (Mitchell and Tijan, 1989). The common feature of all these systems is that sequence-specific recognition of bases occurs through contacts in the major groove

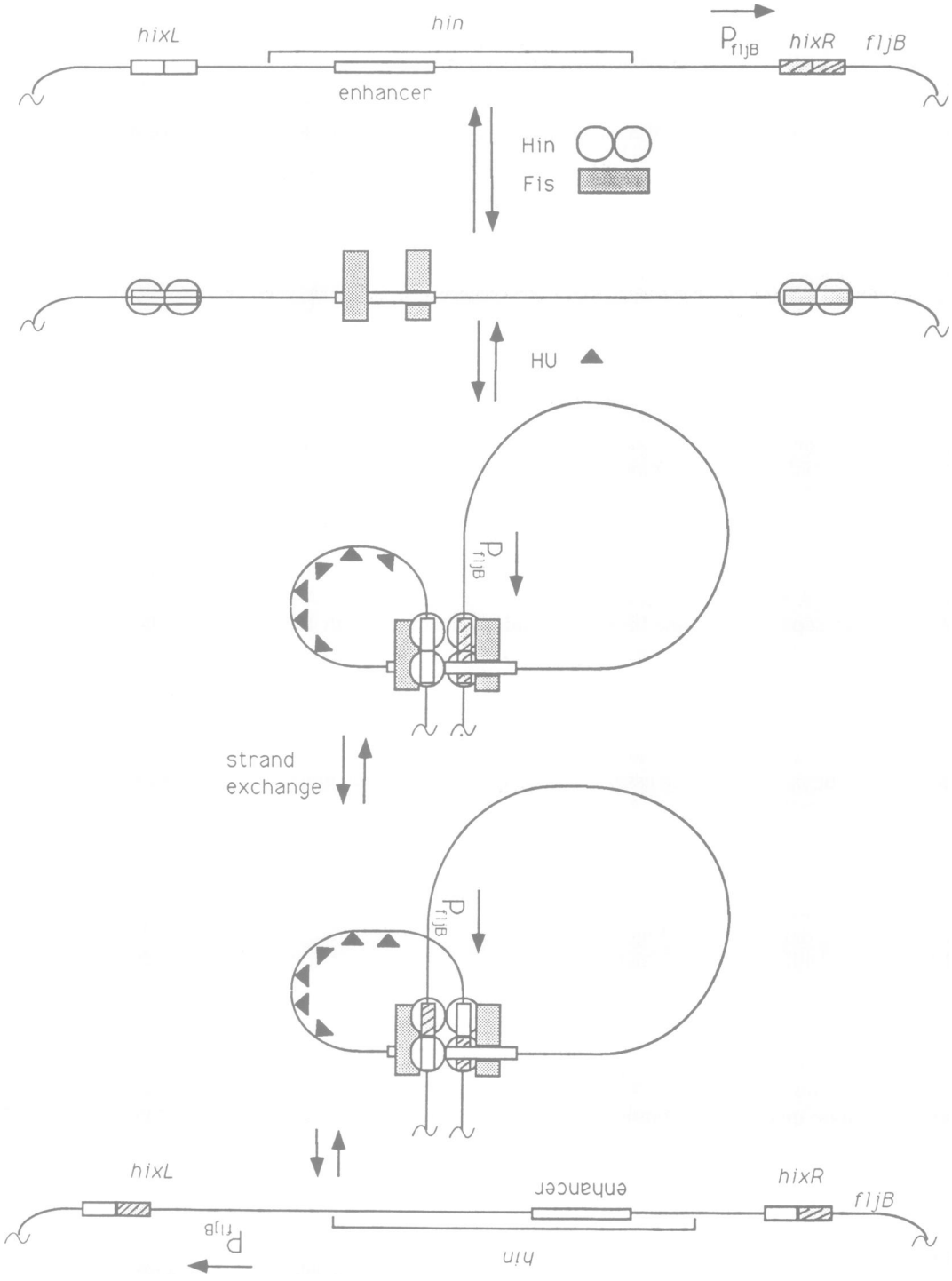

Fig. 1. Model for Hin-mediated Hin inversion. Hin inverts a 1 kb segment of the *Salmonella* chromosome which includes the *hin* structural gene, the promoter for the *fljBA* (flagellin) operon and a recombinational enhancer sequence. Hin recognizes and binds two homologous 26 bp recombination sites, *hixL* and *hixR*. The Fis protein interacts with the recombinational enhancer element. Hin−*hix* complexes interact with Fis−enhancer complexes to direct the formation of a specific protein−DNA complex which is necessary for strand exchange at the *hix* sites so that after recombination, DNA supercoils are lost, providing the energy for the reaction. HU protein binds non-specifically to DNA and bends it. It is only required when the distance between the recombinational enhancer and the nearby *hix* site is short (Johnson *et al.*, 1986).

of the DNA. More recently, some proteins have been found which make sequence-specific contacts to the minor groove of the DNA. In bacterial systems, these include the IHF protein of *E.coli* which is required for bacteriophage λ integration and excision (Yang and Nash, 1989) and the Hin recombinase of *Salmonella* (Glasgow *et al.*, 1989a; Sluka *et al.*, 1990). The minor groove sequence-specific interaction in the Hin system appears to be between an arginine residue in the Hin protein and a pair of adjacent A:T base pairs in the *hix* site. The significance of minor groove specificity to overall DNA recognition in the Hin system is becoming clearer as more examples of regulatory proteins that have specific minor groove binding domains are found. This group of proteins includes the *Drosophila* Antennapedia (Otting *et al.*, 1990) and engrailed homeodomains (Kissinger *et al.*, 1990) and the eukaryotic transcription factor TFIID (Starr and Hawley, 1991). The extensive family of homeodomain proteins may all share the minor groove binding motif. The Hin recombinase system provides a model in which sophisticated techniques of bacterial genetics can be used to explore the details of these interactions.

Results

In vivo DNA binding assay

The P22 based challenge phage assay is designed to characterize the binding of sequence-specific DNA binding proteins to their target DNA sites *in vivo* (Benson *et al.*, 1986). Using the P22-based challenge phage selection, a fully symmetrical site, *hixC*, was tested and found to bind Hin as well as the wild type *hixL* and *hixR* recombination sites (Hughes *et al.*, 1988). The *hixC* sequence shown in Figure 2 is derived from the half-site that is common to both *hixL* and *hixR* and is exchanged during the recombination reaction. The P22 challenge phage assay takes advantage of the regulatory properties of the immunity I region of the *Salmonella*-specific bacteriophage P22. Bacteriophage P22 is similar to the *E.coli* bacteriophage λ in the immunity C region of P22 (Susskind and Botstein, 1978). P22 carries a lysogenic repressor, C2 (equivalent to the CI repressor of λ), which represses other phage genes during lysogeny. Unlike bacteriophage λ, P22 has a second immunity region, the immunity I region, which is responsible for the regulation of expression of the *ant* gene product. Antirepressor (Susskind and Youderian, 1983). Antirepressor inhibits the C2 lysogenic repressor of P22 thereby inducing lytic phage growth. During lysogeny, the Mnt repressor binds at an operator site for the *ant* gene promoter, P_{ant}, to repress transcription of the *ant* gene. In the challenge phage, the *mnt* gene is replaced by a kanamycin resistance gene and the normal operator sequence for P_{ant} is replaced by a DNA site whose *in vivo* binding properties to a particular sequence-specific DNA binding protein is to be assayed (Figure 3). When the DNA recognition protein binds to its specific DNA sequence placed at the P_{ant} operator site, it represses transcription of the *ant* gene from P_{ant} permitting lysogeny. Lysogens are selected directly by selecting for kanamycin resistance encoded within the challenge phage. The greater the affinity a protein has to an artificial P_{ant} operator site, the greater the frequency of kanamycin resistance due to lysogenization of the phage.

```
                 -13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1  +1 +2 +3 +4 +5 +6 +7 +8 +9 +10 +11 +12 +13
hixL:  5'-TTCTTGAAAACCAAGGTTTTTGATAA-3'
       3'-AAGAACTTTTGGTTCCAAAAACTATT-5'

hixR:  5'-TTATCAAAAACCTTCCAAAAGGAAAA-3'
       3'-AATAGTTTTTGGAAGGTTTTCCTTTT-5'

hixC:  5'-TTATCAAAAACCATGGTTTTTGATAA-3'
       3'-AATAGTTTTTGGTACCAAAAACTATT-5'
```

Fig. 2. The sequence of the *hixL* and *hixR* recombination sites as they are read anticlockwise on the standard *Salmonella* linkage map. The *hixC* sequence is fully symmetrical and is derived from the *hix* half-site sequence that is common to both *hixL* and *hixR*.

DNA recognition by Hin is by both major and minor groove interactions

In a previous study, important contacts between the Hin recombinase and the *hix* recombination sites were defined by mutant analysis (Hughes *et al.*, 1988). A strain lysogenic for the P22 c2 repressor and carrying a Hin overproducing plasmid was constructed. This strain is immune to infection by challenge phages with either the *hixL*, *hixR* or *hixC* sequences presented at the operator site for the *ant* promoter. Mutant *hix* challenge phages were isolated which would plaque on the immune host. The DNA sequence analysis of the *hix* regions of numerous such phages resulted in the identification of single base substitution mutations of the *hix* sequences which were defective in binding Hin *in vivo*. These single base changes (see Figure 2) include a change of T:A to A:T or G:C at position −10, a change of A:T to C:G or G:C at either position −5 or −6 and a change of T:A to G:C or C:G at either position +5 or +6. Recent experiments support the model that mutants at position −10 are defective in the ability of Hin to interact with the major groove of the *hix* sequences while mutants defective in positions −5, −6, +5 or +6 are defective in minor groove interactions. These experiments include methylation

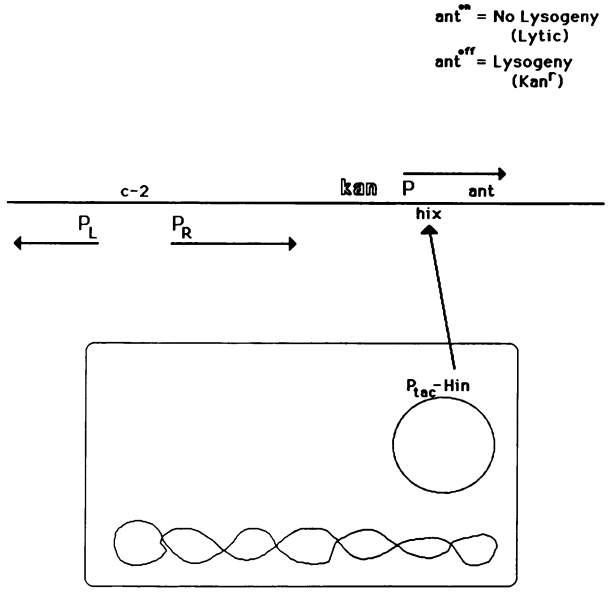

Fig. 3. The P22 challenge phage assay with a *hix* challenge phage. A *hix* site is placed at the normal operator site for the *ant* gene promoter in phage P22. When this *hix* challenge phage is used to infect *Salmonella* cells producing Hin, Hin binds at the *hix* site in the phage and represses transcription from the *ant* gene promoter. Repression of the *ant* promoter is necessary for the establishment of lysogeny. Lysogens are selected directly by selecting for kanamycin resistance encoded within the challenge phage.

Table I. Effects of symmetric pairs of single base substitutions on the binding of *Salmonella* Hin recombinase to the *hix* DNA sites

OPERATOR	SEQUENCE	Frequency of Lysogenization [IPTG] 0μM	10μM	100μM	1,000μM	P_{ant} transcription β-gal activity
hixC	TTATCAAAAACCATGGTTTTTGATAA	10^{-5}	10^{-5}	0.4	0.4	2200
1C	CG	0.1	0.1	0.4	0.4	3400
1G	GC	10^{-5}	10^{-5}	0.4	0.3	3700
1T	TA	10^{-5}	10^{-5}	0.4	0.4	3000
2A	A T	10^{-5}	10^{-5}	0.4	0.3	2600
2G	G C	10^{-5}	10^{-5}	0.4	0.4	4400
2T	T A	10^{-5}	10^{-5}	0.4	0.3	3600
3A	A T	10^{-7}	10^{-6}	10^{-3}	0.4	2700
3G	G C	10^{-5}	10^{-5}	0.2	0.3	4200
3T	T A	10^{-5}	10^{-5}	0.4	0.3	2500
4C	C G	10^{-5}	10^{-5}	0.2	0.4	4100
4G	G C	10^{-8}	10^{-5}	10^{-4}	0.4	2600
4T	T A	10^{-5}	10^{-5}	0.3	0.4	3400
5C	C G	$<10^{-8}$	$<10^{-8}$	$< 10^{-8}$	$<10^{-8}$	6000
5G	G C	$<10^{-8}$	$<10^{-8}$	$< 10^{-8}$	$<10^{-8}$	4400
5T	T A	0.1	0.1	0.4	0.3	1100
6C	C G	$<10^{-8}$	$<10^{-8}$	$< 10^{-8}$	$<10^{-8}$	6100
6G	G C	$<10^{-8}$	$<10^{-8}$	$< 10^{-8}$	$<10^{-8}$	4200
6T	T A	10^{-5}	10^{-5}	0.3	0.3	3600
7C	C G	10^{-7}	10^{-6}	10^{-2}	0.3	3300
7G	G C	10^{-6}	10^{-6}	0.3	0.3	4200
7T	T A	10^{-5}	10^{-5}	0.4	0.4	3300
8C	C G	10^{-6}	10^{-6}	0.4	0.3	3100
8G	G C	10^{-6}	10^{-5}	0.4	0.4	2600
8T	T A	10^{-5}	10^{-5}	0.4	0.4	2100
9A	A T	0.1	0.2	0.4	0.4	700
9G	G C	10^{-5}	10^{-5}	10^{-3}	0.3	960
9T	T A	10^{-3}	10^{-3}	0.4	0.3	1200
10A	A T	10^{-6}	10^{-6}	10^{-6}	10^{-6}	2300
10C	C G	10^{-2}	0.1	0.4	0.4	1000
10G	G C	10^{-7}	10^{-7}	10^{-6}	10^{-5}	2200
11C	C G	10^{-5}	10^{-4}	0.4	0.4	2000
11G	G C	10^{-6}	10^{-6}	10^{-4}	0.2	2000
11T	T A	10^{-6}	10^{-5}	0.4	0.3	2000
12A	A T	10^{-6}	10^{-6}	0.2	0.3	3000
12C	C G	10^{-5}	10^{-5}	0.4	0.3	1500
12G	G C	10^{-7}	10^{-7}	10^{-8}	10^{-7}	2700
13A	A T	10^{-4}	10^{-4}	0.3	0.4	1900
13C	C G	10^{-4}	10^{-4}	0.3	0.4	1200
13G	G C	10^{-5}	10^{-4}	0.3	0.3	1400

The Hin recombinase is expressed in the recipient cells from the *tac* promoter present in plasmid pKH66 (Hughes *et al.*, 1988). Challenge phage assays were performed with challenge phages containing each of the symmetrically altered *hix* sites (see Materials and methods). Increasing levels of IPTG inducer resulted in increasing levels of Hin produced *in vivo*. Hin binds the *hix* site in the P22 challenge phage which represses the *ant* promoter allowing lysogenization (see Results). The more tightly that Hin binds to the *hix* site, the more that lytic functions are repressed and the greater the frequency of lysogenization. Lysogenization was measured as the number of kanamycin resistant lysogens isolated after infection divided by the total number of cells infected. In separate experiments, the *lac* operon was fused downstream of the *ant* promoter in each of the challenge phages with the 40 different *hix* symmetrical sites and integrated into the *Salmonella* chromosome in single copy into strain MS1389. The expression of the *lac* operon was measured in the absence of Hin to assay the effect of each symmetrical *hix* sequence on promoter expression.

interference assays between Hin and DMS-methylated *hixL* and *hixR* sequences (Glasgow *et al.*, 1989a), DNA cleavage assays with Fe–EDTA covalently linked to particular amino acid residues of the Hin binding domain (Sluka *et al.*, 1990) and mutant analysis (Hughes *et al.*, 1988). In order to characterize the critical base interactions between Hin and the *hixC* sequence completely, and to quantify the relative contribution of any base pairs to overall binding, a set of 40 symmetrically altered pairs of single base substitution mutations in the *hixC* sequence was constructed and inserted into the challenge phage system. These *hixC* symmetrical double mutants were assayed for the effect of each alteration on the abiltiy of Hin to bind to these sites *in vivo*. The results presented in Table I and summarized in Figure 4 agree, with one exception, with the mutant results and the *in vitro* results obtained in earlier studies. Symmetrical changes of −10T,+10A to either −10A,+10T or −10G,+10C severely diminished the ability of Hin to bind to these *hixC* mutant sites. Symmetrical changes of −6A,+6T to either −6G,+6C or −6C, +6G and symmetrical changes of −5A,+5T to either −5G,+5C or −5C,+5G completely eliminated any detectable binding by Hin to these mutant *hixC* sequences *in vivo*. One important contact not observed in the original mutant studies was also found using the set of symmetrical double mutant *hixC* sites. A change of −12T,+12A to −12G,+12C is defective in binding Hin even at the highest induced levels of Hin in the cell. These results define positions 10 and 12 as the primary contact points for Hin in the major groove region of the *hix* sequences and positions 5 and 6 as the primary contact points for Hin in the minor groove region of the *hix* sequences. As discussed below, however, the change of −12T,+12A to −12G,+12C probably affects multiple contacts in the major groove due to the creation of a Dam-methylation site.

The P22-based challenge phage assay does not result in a direct measurement of binding or dissociation constants for interactions of DNA-specific binding proteins to their respective DNA binding sites. However, the challenge phage assay does allow for the direct, *in vivo*, comparision of relative binding measurements of Hin for the different symmetrically altered *hixC* sequences providing a means to order the importance of each symmetric double base pair in *hixC* for binding to Hin. The data in Table I show that positions 5 and 6 in the *hixC* sequence are the most critical single contacts between Hin and the DNA. Symmetrical changes to G:C or C:G base pairs at positions 5 and 6 eliminate the ability of Hin to bind to the *hixC* sequence as measured in the challenge phase assay. The symmetrical changes of −10T,+10A to either −10A,+10T or −10G,+10C as well as the symmetrical changes of −12T,+12A to −12G,+12C result in a substantial reduction in the ability of Hin to bind these sequences, but residual binding to these changes remains. Considering the fact that Hin recognizes positions 5 and 6 in the minor groove and positions 10 and 12 in the major groove, these data support the model that critical contacts between Hin and the *hix* recombination sites lie within the minor groove portion of the binding site as well as the major groove region.

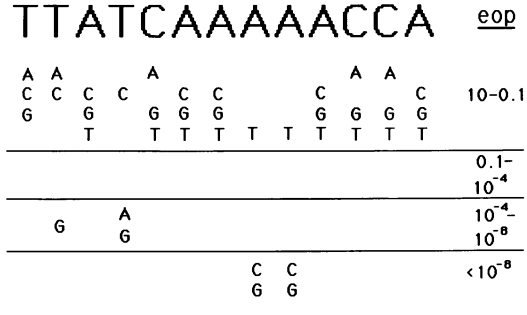

Fig. 4. Efficiency of *in vivo* binding of Hin recombinase to symmetrically altered *hix* DNA sites. This figure represents a summary of the data presented in Table I. Only one half of the symmetrical *hixC* sequence is presented here. The efficiency of plating (eop) represent the frequency of lysogeny in the standard challenge phage assay (see Materials and methods) as compared with the 'wild-type' *hixC* sequence.

To assay the effect of the symmetric changes of the *hixC* operator sequence on the intrinsic levels of transcription from the *ant* promoter, P_{ant}, the region of the P22 genome downstream of P_{ant} was replaced with the *lac* operon. This region of P22 was then recombined into the *Salmonella* chromosome in order to assay P_{ant} expression in single copy. Levels of β-galactosidase for construction with all 40 symmetrical *hixC* operator sites were measured and are included in Table I. These results show that the β-galactosidase activity varies within a 9-fold range. These assays were done in strains without the Hin-producing plasmid pKH66. The β-galactosidase activities represent constitutive expression of P_{ant} with different *hix* operator sites downstream. There is no significant correlation between β-galactosidase activity and the frequency of lysogeny, suggesting that expression from the *ant* promoter is not significantly affected by the different operator mutations in the absence of Hin. The 9-fold range of the effect of different *hix* operator sequences on promoter expression is somewhat different than the effects of symmetrical mutant operators in other systems on promoter strength which can vary >1000-fold (Staacke *et al.*, 1990). It may be that by recombining our fusions into the chromosome and thus measuring transcription from a single copy, problems associated with plasmid copy number and instability are avoided resulting in a much narrower range of promoter transcription *in vivo*.

Effect of Dam-methylation on Hin binding to symmetric hix sites

Because position 12 was not identified as an important base despite a thorough search among hundreds of *hix* mutant sites defective in binding Hin (Hughes *et al.*, 1988), it was surprising to find that the symmetrical *12G* site is very defective in binding to Hin in the challenge phage assay (Table I). It appears that a change of T to G at position 12 results in the creation of a GATC *Dam*-methylation site. Had this change occurred during mutant selection, any GATC sequence that might have arisen would have been unmethylated due to the nature of the selection (Hughes *et al.*, 1988). In the original selection, *hix* challenge phages were mutagenized by UV irradiation prior to infection of a host immune to *hix* challenge phage infection. The immune host also carried a plasmid expressing the *mucAB* operon of pKM101. The *mucAB* genes of pKM101 are analogous to the *umuCD* genes of *E.coli* and result in a high frequency of base pair substitution mutagenesis following repair of UV damage (Perry and Walker, 1982). Thus, had the change of −12T to −12G or +12A to +12C occurred during MucAB repair of the UV-mutagenized *hix* challenge phages, the resulting *dam* site would not have been methylated. However, when the symmetrical *12G hixC* site was cloned

into the challenge phage, it would have been fully methylated. Thus, the *12G* symmetrical mutant constructed *in vitro* would be multiply modified having not only the change of T:A to G:C at position 12, but also having the adenine residues at positions 10 and 11 methylated in the N6 position in the major groove by Dam-methylase. Since position 10 is a critical contact, we could not immediately rule out the possibility that methylation of the adenine at position 10 may contribute more to the defect in Hin binding to the *12G* sequence rather than the change of T:A to G:C at position 12.

To test the effect of Dam-methylation on the inability of Hin to recognize the *12G* symmetrical site, challenge phage assays were performed with the *12G* challenge phage in strains that were defective in Dam-methylation so that binding in the presence and absence of Dam-methylation could be compared. The results presented in Table II show that Dam-methylation had no effect on Hin binding the *hixC* sequence which lacks a Dam-methylation site, but made a substantial contribution on the inability of Hin to bind the *12G* site. This result supports the idea that position 12 is not as critical a contact in the major groove as is position 10 and that the defect in binding to *12G* site results from both changing the normal T:A base pair to G:C and from methylation of the adenine residues at positions 10 and 11. Still, the frequency of lysogeny of the *12G* challenge phage in the *dam* mutant recipient is 1000-fold lower than the wild-type *hixC* phage, suggesting that interaction at this position is important to binding.

The above results were confirmed *in vitro* using filter binding assays to the following four *hix* sequences: the *12G hix* sequence, the wild type *hixC* sequence and the same two sequences with N6-methyladenine incorporated in place of adenine at positions 10 and 11. The results presented in Figure 5A demonstrate that methylation of the adenine at positions 10 and 11 as occurs by Dam methylase *in vivo* results in a significant reduction in binding by Hin *in vitro* as well. As expected, the efficiency of binding Hin was greater for the *hixC* sequence with N6-methyladenine at positions 10 and 11 than for the *12G* sequence with N6-methyladenine at the same positions. Also predicted from the *in vivo* results above, the affinity of Hin for the unmethylated *12G* site is less than that for the *hixC* site and greater than that for the *hixC* sequence with N6-methyladenine at positions 10 and 11. Thus, the change of T:A to G:C at position 12 reduces Hin binding for both the wild type *hixC* sequence and the methylated *hixC* sequence consistent with the results found *in vivo*.

Table II. Effect of *dam*-methylation on the binding of Hin to the *−12G + 12C* site.

Challenge phage donor	Recipient	Frequency of lysogeny (per 10^8 cells)
hixC	*dam*$^+$	40%
hixC	*dam*$^-$	40%
−12G + 12C	*dam*$^+$	10^{-8}
−12G + 12C	*dam*$^-$	10^{-3}

Donor phages grown on *dam* mutant cells were used to infect MS1868 cells containing the Hin producing plasmid pKH66 under conditions of full Hin induction (1 mM IPTG). Lysogenization was measured as the number of kanamycin resistant lysogens isolated after infection divided by the total number of cells infected.

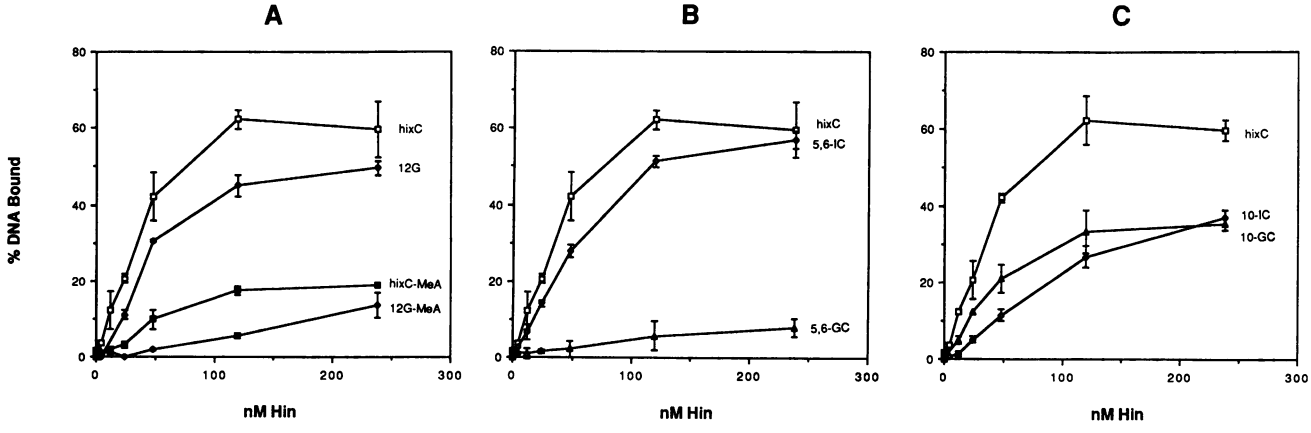

Fig. 5. Filter binding assays of purified Hin to various *hix* sequences. The following oligonucleotide was used in all the filter binding assays presented in this figure: 5′-GCATGGAGGATTGCGT-*hix*-AGCAATCCTCCATGC-3′. Binding conditions were done as previously described for Hin (Glasgow *et al.*, 1989a), and the binding assays were performed as described by Halvorsen *et al.* (1991). **A.** Methylation effects on the binding of Hin to the *hixC* and *12G* sequences. The following *hix* sequences were used in these assays: *hixC*: 5′-TTATCAAAAACCATGGTTTTTGATAA-3′; *12G*: 5′-TGATCAAAAACCATGGTTTTTGATCA-3′; *hixC-MeA*: 5′-TT(MeA)TCAAAAACCATGGTTTTTG(MeA)TAA-3′; and *12G-MeA*: 5′-TG(MeA)TCAAAAACCATGGTTTTTG(MeA)TCA-3′ where (MeA) is N6-methyladenine. **B.** Binding of Hin to *hix* sites with I:C base pairs in the minor groove recognition region. The following *hix* sequences were used in these assays: *hixC*: 5′-TTATCAAAAACCATGGTTTTTGATAA-3′; *5,6-IC*: 5′-TTATCAAIIACCATGGTCCTTGATAA-3′; and *5,6-GC*: 5′-TTATCAAGGACCATGGTCCTTGATAA-3′ where I is inosine. **C.** Binding of Hin to *hix* sites with I:C base pairs in the major groove recognition region. The following *hix* sequences were used in these assays: *hixC*: 5′-TTATCAAAAACCATGGTTTTTGATAA-3′; *10-IC*: 5′-TTAICAAAAACCATGGTTTTTGCTAA-3′; and *10-GC*: 5′-TTAGCAAGGACCATGGTTTTTGCTAA-3′ where I is inosine.

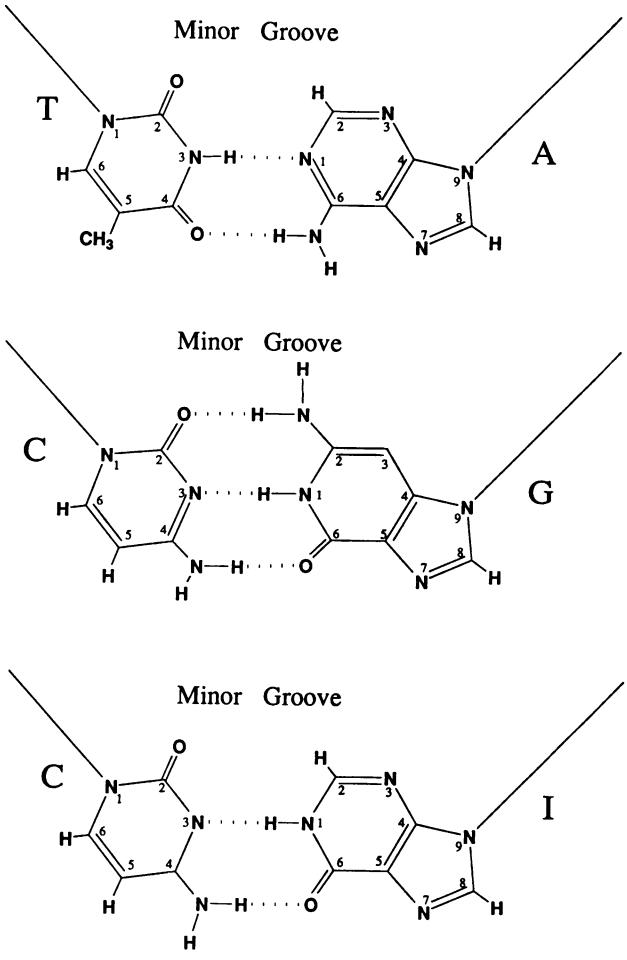

Fig. 6. Diagram indicating the atomical structure of an adenine:thymine base pair, a guanine:cytosine base pair and an inosine:cytosine base pair. The minor groove region of the I:C base pair is identical to that of an A:T base pair while the major groove region of an I:C base pair is identical to that of a G:C base pair.

Hin binds A:T and I:C but not G:C base pairs in the minor groove recognition region of the hix sequence

The results of the challenge phase assays with the symmetrical *hix* sites suggest that the more critical contacts between Hin and the *hix* sequences are through minor groove contacts at positions 5 and 6. To test further the model that Hin recognizes the A:T base pairs in the minor groove at positions 5 and 6, *in vitro* binding studies were performed on *hix* symmetrical sequences with A:T, G:C or I:C base pairs at positions 5 and 6. *In vivo* data already presented support the result that Hin does not bind to *hix* sequences with G:C base pairs at positions 5 and 6. If the inability of Hin to bind *hix* sites with G:C or C:G base pairs in positions 5 and 6 is due to a loss of interaction in the minor groove of the DNA, then Hin should bind a *hix* sequence with I:C base pairs at positions 5 and 6. The I:C base pair is chemically identical to an A:T base pair from the perspective of the minor groove and to a G:C base pair in the major groove (Figure 6). Symmetrical *hix* sites were synthesized with A:T, G:C or I:C base pairs at positions 5 and 6 and binding of Hin to these sites was determined using both gel retardation assays and filter binding assays. The assays were performed using purified Hin and the results are shown in Figures 5B and 7. Hin binds very well to *hix* sites with I:C base pairs in these positions. These results support the model that contacts between Hin and positions 5 and 6 of the *hix* sites are in the minor groove of the DNA.

Hin does not bind I:C pairs in the major groove recognition region of the hix sequence

The change of T:A to G:C at position 10 was also shown to be defective in binding Hin (Table I). Since this is thought to be in the major groove recognition region of the *hix* sequence, one would predict that substitution of I:C for G:C at this position would still be defective in binding Hin. Filter binding assays were performed using purified Hin and the wild-type *hixC* sequence, the *10G* sequence (−10G, +10C) and the same sequence with inosine in place of guanine

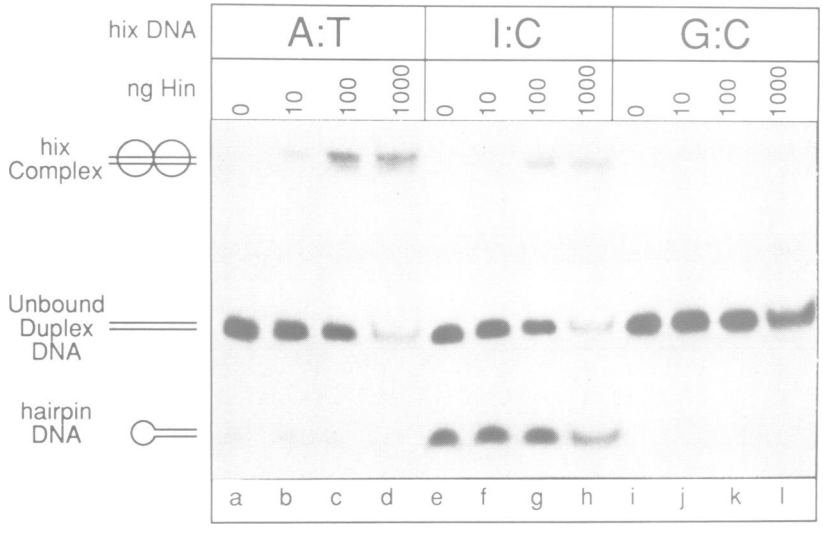

Fig. 7. *In vitro* binding assay. This is a gel retardation assay demonstrating that Hin binds to either A:T (lanes a–d) or I:C (lanes e–h) base pairs at positions 5 and 6 in the *hix* sites, but not to G:C (lanes i–l) base pairs at those positions. The exact sequence of the 56 bp oligonucleotides used is given in Materials and methods. It appears that the *hix* sequences containing I:C base pairs are not as stable as the otherwise identical DNA with A:T or G:C base pairs in the same positions. The exact sequences of the *hix* sites are given in Materials and methods.

(−10I, +10C). The results shown in Figure 5C demonstrate that Hin binding is reduced with either the G:C or I:C substitutions at position 10. These results support the model that Hin binds the *hix* sequence through major groove contacts at position 10.

Effect of deazaadenine on sequence-specific binding to A:T base pairs in the minor groove

Minor groove recognition by Hin may play a larger role in overall binding than major groove binding. Mutants in the minor groove recognition region were no longer recognized by Hin while the mutants most defective in major groove recognition retained some binding activity *in vivo*. Two DNA recognition models were envisioned to explain sequence-specific binding in the minor groove of the *hix* sites. One model is that Hin makes sequence-specific hydrogen-bond contacts with either the N3 position of adenine or the O2 position of thymine with A:T base pairs in the minor groove. This would explain the high degree of specificity for A:T base pairs, since the NH_2 group of guanine would block hydrogen bond formation, and the fact that orientation of the A:T base pairs does not affect binding, since N2 and O2 can be considered equivalent H-bond acceptors. A second model is that interaction between Hin and the minor groove is structural in nature, binding through the charge contacts with the phosphate backbone and through van der Waals contacts within the groove; the difference in surface contacts between Hin and A:T or G:C base pairs in the minor groove explains the difference in binding.

The ability of Hin to make direct hydrogen bond contacts to A:T base pairs in the minor groove was tested by synthesizing *hixC* sites with 3-deazaadenine in place of adenine at positions 5 and 6 in the *hixC* site. Deazaadenine is identical to adenine except that the N3 group is replaced by a CH group (Figure 8). Sequence-specific minor groove interactions with Hin would occur at the A:T base pairs in positions 5 and 6. Gel retardation assays were performed on *hix* sites with either adenine or 3-deazaadenine residues incorporated at positions 5 and 6 and the apparent K_D for each sequence was determined. When 3-deazaadenine was incorporated into the same *hixC* sites in place of adenine at positions 5 and 6, loss of the hydrogen bond acceptors sites in the minor groove affected Hin binding when the 3-deazaadenine residues were on one side of the minor groove and not on the other (Table III). The apparent K_D increased by 3-fold when the N3 positions are replaced by a CH group using the 3-deazaadenine substitutions. This result suggests that Hin does make a specific interaction in the minor groove with the wild-type *hixC* site.

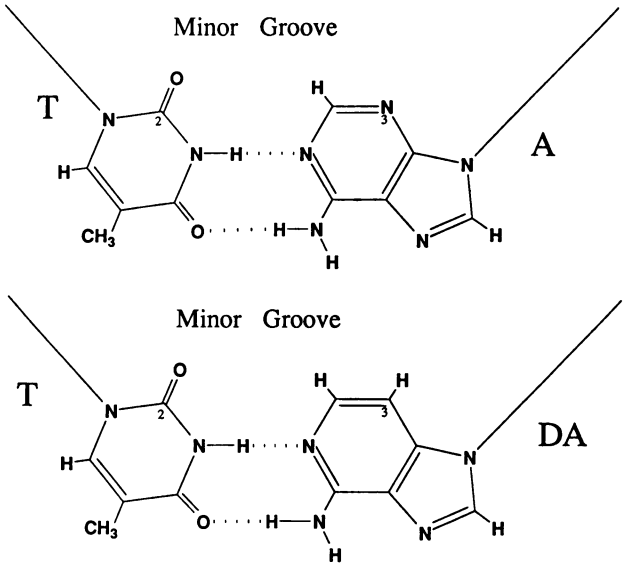

Fig. 8. Diagram indicating the atomical structure of an adenine:thymine base pair as compared with a deazaadenine:thymine base pair. The deazaadenine residue would be capable of making duplex DNA just like adenine, but is lacking the N3 hydrogen bond acceptor site in the minor groove.

Discussion

The bacteriophage P22-based challenge phage selection was used to characterize the binding of the Hin recombinase to the *hix* recombination sites and thereby separate the binding reaction from the other aspects of the site-specific recombination process. The importance of each base pair in DNA recognition by Hin was assayed by screening the *in vitro* binding of Hin to a set of 40 recombination sites which vary by symmetrical pairs of single base substitutions. Only a single base pair at position 10 shows strong interaction in sequence-specific DNA recognition in the presumed major groove contact region, although position 12 does make an important contribution. Positions 5 and 6 appear to be critical contacts between Hin and the DNA. These contacts are

Table III. *In vitro* binding of Hin to *hix* sites with deazaadenine substituted for adenine at positions 5 and 6

hix site	Apparent K_D ([Hin] at 50% bound)
T T A T C A A **A** **A** A C C A T G G T **T** **T** T T G A T A A A A T A G T T **T** **T** T G G T A C C A **A** **A** A A C T A T T	33 ± 3 nM
T T A T C A A **T** **T** A C C A T G G T **A** **A** T T G A T A A A A T A G T T **A** **A** T G G T A C C A **T** **T** A A C T A T T	24 ± 3 nM
T T A T C A A **D** **D** A C C A T G G T **T** **T** T T G A T A A A A T A G T T **T** **T** T G G T A C C A **D** **D** A A C T A T T	39 ± 2 nM
T T A T C A A **T** **T** A C C A T G G T **D** **D** T T G A T A A A A T A G T T **D** **D** T G G T A C C A **T** **T** A A C T A T T	61 ± 3 nM

In vitro binding of Hin to the various *hix* sequences depicted was measured using gel retardation assays with ^{32}P-labeled *hix* DNA and purified Hin protein. The apparent K_D is the concentration of Hin at which 50% of the DNA is retarded (bound).

presumed to be with A:T base pairs in the minor groove based on several criteria. (i) Previous *in vitro* results with Fe–EDTA cleaving moieties attached to the Hin binding domain followed by mapping of the cleavage sites is consistent with the Arg 140–Pro–Arg residues in the N-terminal portion of the 52 amino acid Hin binding domain recognizing the A:T base pairs at positions 5 and 6 in the minor groove (Sluka *et al.*, 1987, 1990). (ii) DMS-methylation of the N3 adenine in the minor groove of *hixL* at either positions −5, 6, +5, or +6 interferes with Hin binding *in vitro* (Glasgow *et al.*, 1989a). (iii) Substitution of T:A base pairs for A:T base pairs and vice versa at positions 5 and 6 of the *hixC* sequence has no effect on Hin binding (Table I). In their classic paper on sequence-specific DNA recognition by proteins, Seeman *et al.* (1976) had predicted that recognition of A:T base pairs in the minor groove would be relatively insensitive to A:T base pair reversals (Seeman *et al.*, 1976). Finally, Hin will bind *hixC* with I:C base pairs substituted for A:T base pairs at positions 5 and 6 but not to G:C base pairs (Figure 5B and 7). Since an I:C base pair is identical to an A:T base pair in the minor groove and a G:C base pair in the major groove, this result is consistent with a minor groove interaction at positions 5 and 6. Also, Hin was found to be equally defective in binding to either a G:C or an I:C base pair at position 10. This result is consistent with a major groove interaction between Hin and the *hix* sequence at this position. Furthermore, the results of the challenge phage assays with the symmetrically altered *hixC* sequences suggest that sequence-specific interactions in the minor groove play a substantial role in overall binding and may be more critical than major groove contacts. An advantage of the challenge phage system over other systems is the ability to compare numerous binding sites directly for their ability to bind Hin *in vivo* using a simple genetic assay.

Three of the *hix* symmetrical sites tested in the challenge phage assay showed high frequencies of lysogenization (0.1) even in the absence of IPTG. These three include the symmetrical sites *1C*, *5T* and *9A*. These sequences changes may have created a more tightly binding *hix* site and even the basal level of Hin from the *tac* promoter is enough to give repression. This is unlikely because when challenge

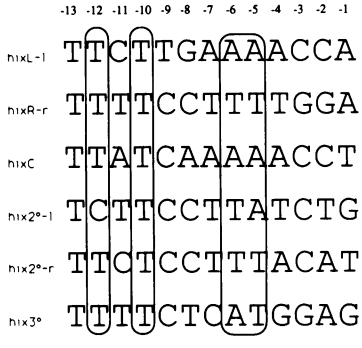

Fig. 9. Comparison of DNA sequences recognized by the Hin recombinase. The *hixC* sequence represents the portions of the *hixL* and *hixR* recombination sites which are identical to each other and are exchanged during the DNA inversion reaction (*hixL*-r and *hixR*-l). The *hixL*-l and *hixR*-r sequences represent the portions of the *hixL* and *hixR* recombination sites bound by Hin but not exchanged during the DNA inversion reaction. The left and right portions of the *hix2°* site (*hix2°*-l and *hix2°*-r, respectively), are located upstream of the Hin initiation codon and may be a site for autogenous regulation by Hin. The *hix3°* sequence maps upstream of the *hix2°* site, but is only bound by a Hin binding domain lacking the Arg140 residue (Sluka *et al.*, 1990).

phage assays were performed with these three phages in MS1868 without the pKH66 plasmid, the same frequency of lysogeny was obtained (data not shown). It is unlikely that the high frequency of lysogeny in the absence of Hin for the *1C*, *5T* and *9C* sites is due to a reduction in the intrinsic levels of transcription from the P_{ant} promoter since sites with similar effects on transcription from P_{ant} give low lysogenization frequencies in the challenge phage assay. Another possibility is that these sequences are bound by some other protein which represses P_{ant} in the challenge phage assay. Such alternative repressors were observed when a set of symmetrical changes were constructed in the *trp* operator sequence and tested in the challenge phage system (Bass *et al.*, 1987; D.N.Arvidson and P.Youderian, personal communication).

The identification of the critical contacts between Hin and the DNA allows the prediction of a minimal consensus sequence for a Hin recognition sequence. The Hin binding domain recognizes the individual *hix* 13bp half-sites. Six such sequences have been identified in *in vitro* DNA binding studies and are diagrammed in Figure 9. These include the *hixL* and *hixR* unique half-sites (Johnson and Simon, 1985), the *hixC* half-site which is present in both *hixL* and *hixR*, the left and right half-sites of the *hix* secondary site (*hix2°*) which is present upstream of the Hin initiation codon (Bruist *et al.*, 1987b), and the *hix* tertiary half-site (*hix3°* (Sluka *et al.*, 1990). All of these sequences have the minimal sequence requirements defined in these studies. These include a T:A base pair at position-10 (−10T, +10A), either A:T or T:A base pairs at positions 5 and 6, and the lack of a G residue at position −12 or a C residue at position +12.

The possibility that Hin makes direct hydrogen bond contacts to the A:T base pairs in the minor groove was tested. Binding of Hin to *hix* sites in which adenine at positions 5 and 6 was replaced by the adenine analog, 3-deazaadenine, was measured. A wide variety of DNA base analogs can be used in studying specific protein–DNA interactions (Aiken and Gumport, 1991). Incorporation of 3-deazaadenine in place of adenine does not destabilize duplex DNA and has essentially no effect on overall DNA conformation (Newman *et al.*, 1991). 3-Deazaadenine is capable of making duplex DNA through hydrogen bond contacts to thymine, but lacks the N3 hydrogen bond acceptor site in the minor groove. The results of these studies are consistent with the idea that Hin makes a specific contact to the thymine side of the minor groove recognition region (positions 5 and 6) in the wild-type *hixC* sequence. The reduction in binding with the 3-deazaadenine substitutions is consistent with the loss of a hydrogen bond contact. In the worst case, the complete loss of a 2 Kcal hydrogen bond contact might lead to a 10-fold increase in the apparent K_D. However, in the absence of the correct hydrogen bond contact, the Hin protein is likely to make the next best contact available either through an alternative hydrogen bond contact or through the van der Waals contacts. This would lead to a small increase in the apparent K_D than if the energy from a hydrogen bond contact was completely eliminated. *In vitro* results are consistent with the Arg 140–Pro–Arg residues of Hin making the minor groove contacts at positions 5 and 6 in *hixC* (Sluka *et al.*, 1990). This Arg140–Pro–Arg peptide may be a member of a new family of minor groove recognition motifs. Recently, NMR structure of the Antennapedia homeodomain demonstrates that the Arg5 residue contacts the DNA in the

minor groove (Otting *et al.*, 1990). Also, the crystal structure of an engrailed homeodomain–DNA complex has been reported which also contains an Arg–Pro–Arg peptide that interacts with the minor groove of its DNA recognition sequence (Kissinger *et al.*, 1990). In fact, it was proposed, based on the crystal structure, that Arg3 of the homeodomain, which would be equivalent to Arg140 of Hin for minor groove interaction, makes a specific hydrogen bond contact to O2 thymine in what would be the equivalent placement on the DNA to O2 thymine at position 5 of the *hixC* sequence. The similarity between Hin and the homeodomain proteins has been reviewed recently (Affolter *et al.*, 1991). Our current results emphasize how these evolutionary distinct systems employ almost identical DNA recognition elements.

The Hin family of site-specific recombinase includes the Gin protein of phage Mu, the Cin protein of phage P1 and the Pin protein of phage e14. These proteins will all complement a Hin mutant for site-specific recombination at the *hix* sites (Kutsukake and Iino, 1980; Plasterk *et al.*, 1983). These proteins share 55% amino acid identity in their peptide sequence. In comparing these proteins in the minor groove recognition regions it appears that only Arg140, which is thought to make the specific contact in the minor groove, is conserved. The minor groove recognition peptides are Gly138-Gly-Arg-Pro-Arg-Ala143 in Hin, Gly137-Gly-Arg-Pro-Pro-Lys142 in Gin, Gly137-Gly-Arg-Arg-Pro-Lys142 in Cin and Gly137-Gly-Arg-Arg-Pro-Lys142 in Pin. The differences in amino acids beyond the Gly–Gly–Arg residues suggests that these residues do not make sequence-specific contacts in the minor groove. It seems more likely that they are involved in contacts to the phosphate backbone in the minor groove through either a lysine or arginine residue as the Hin protein sequence crosses the backbone of the minor groove to extend to the major groove recognition region (Mack *et al.*, 1990).

Still, it is possible that the most important aspect of the minor groove recognition may not be due to a specific contact at all. Rather, the lack of a negative contact may be the most significant contribution of the A:T base pairs in allowing Hin to make tight interactions in the minor groove. For example, either a change of A:T to G:C or C:G, or methylation of the N3 adenine position by dimethylsulfate, introduces a bulky amino or methyl group into the minor groove; both strongly inhibit Hin binding to the *hix* sites. The decrease in binding with the 3-deazaadenine substitutions may simply be the effect of placing a hydrogen atom into the minor groove at those positions. So while the deazaadenine studies suggest a specific contact is made in the minor groove and the placement of Arg140 at this location would support the idea that the specific contact is through a hydrogen bond, other modes of interaction are still possible.

Materials and methods

Bacteria and bacteriophage strains

All *S.typhimurium* strains are derived from strain LT2. Isogenic strain MS1883 (*leuA414 hsdSB supE40* Fels$^-$) and MS1868 (*leuA414 hsdSB Fels*$^-$) were used for phage growth and challenge phage assays, respectively (Graña *et al.*, 1985). TT11691 (*dam-101*::Tn*10*dTc), kindly provided by Chris Conner and John Roth was used to introduce a *dam*-insertion mutant allele into other strains. P22 *Kn9 arcH1065* (Am), the parent of the *hix* challenge phages, has been described by Youderian *et al.* (1983). The *hix* challenge phages were constructed as described previously (Hughes *et al.*, 1988) by crosses between P22 *Kn9 arcH1605* and pPY190-derived plasmids in which various *hix* sequences had been cloned in place of the *ant* operator. Strain MS1389 is a defective P22 lysogen deleted from the bacterial *proA* gene through the P22 immunity C region (Δc2). This strain still has the immunity I region and was used to recombine the *lac* operon fused to the *ant* promoter with all 40 symmetrical *hix* mutant sites in P22 into the chromosome in single copy.

Plasmids

The construction of plasmids carrying *hix* DNA sequences at the P22 *ant* operator locus was done as previously described (Benson *et al.*, 1986). Symmetrical *hix* 26 bp oligonucleotides were synthesized on an Applied Biosystems automated DNA synthesizer, purified on a 20% polyacrylamide gel and resuspended in 10 mM Tris–HCl, 0.1 mM EDTA and 50 mM NaCl. They were self-hybridized by placing them in a boiling water bath and allowing them to cool slowly to room temperature. The duplex oligos were cloned into the *Sma*I site of pPY190 and recombined into the P22 challenge phage as described. The plasmid pKH66 contains the *hin* gene under the control of the *tac* promoter and genes conferring resistance to the antibiotics streptomycin and spectinomycin. Its construction has been previously described (Hughes *et al.*, 1988).

Media

The E medium of Vogel and Bonner supplemented with 0.2% glucose was used as minimal medium (Vogel and Bonner, 1956). Alternative carbon sources were supplemented to 0.2% in E medium lacking citrate. Luria–Bertani (LB) medium (Difco Bacto tryptone, 10 g/l; Difco yeast extract, 5g/l, NaCl 5 g/l) was used as rich medium. Auxotrophic supplements were included in media at final concentrations described previously (Davis *et al.*, 1980). Antibiotics were included in media as needed (final concentrations given): ampicillin (100 μg/ml for plasmid selection, 30 μg/ml for single copy selection in rich media, 15 μg/ml for single copy selection in minimal media), chloramphenicol (12.5 μg/ml), kanamycin sulfate (50 μg/ml in rich media, 125 μg/ml in minimal media), spectinomycin (100 μg/ml), streptomycin (50 μg/ml for plasmid selection, 1 mg/ml for chromosomal *rpsL* mutant selection), tetracycline hydrochloride (15 μg/ml in rich media, 10 μg/ml in minimal media). Bacterial cell growth was monitored at 540 nm in a Klett–Summerson colorimeter.

Transduction

For all transductional crosses, the high-frequency generalized transducing mutant of bacteriophage P22 (*HT105/1 int-201*) was used (Sanderson and Roth, 1988). Selective plates were spread with $2 \times 10^8 - 2 \times 10^9$ phage. Transductants were purified by two successive single colony isolations on non-selective medium. For antibiotic selections other than ampicillin and tetracycline, phage and cells were mixed and left for at least one hour prior to plating on selective media to allow for expression of the antibiotic-resistance phenotype.

Challenge phage assay

An overnight culture of strain MS1868 carrying the Hin-producing plasmid pKH66 was diluted 100-fold into LB plus streptomycin and spectinomycin to maintain plasmid selection and was grown to a density of 100 Klett units ($\sim 6 \times 10^8$ cells/ml). Cells were diluted 4-fold into the same medium plus varying amounts of IPTG and grown for 1 h to permit the induction of Hin expression from plasmid pKH66. P22 *hix* challenge phage were added to a m.o.i. of 20, and infected cells were incubated for 1 h at room temperature to allow expression of the Kan^r phenotype. After 1 h, dilutions of the adsorption mixture were plated on LB-Kan plates containing the same concentration of IPTG used for induction and incubated overnight at 37°C. β-Galactosidase assays were performed according to Miller (1972).

In vitro binding of Hin to synthetic hix sites

The following oligonucleotide background was used for the DNA binding assays using deazazdenine: 5′-GGATTGCTTTATCAANNACCATGG-TNNTTGATAAAGCAATCC-3′. The nucleotides at positions 5 and 6 in the *hix* sequence (N) were varied as depicted in Table III. The following oligonucleotide was used for the DNA binding assay using inosine: 5′-GCATGGAGGATTGCTTTATCAANNACCATGGTNNTTGATA-AAGCAATCCTCCATGC-3′. The nucleotides at positions 5 and 6 in the *hix* sequence (N) were either A,A at −5 and −6 and T,T at +5 and +6, or G,G at −5 and −6 and C,C at +5 and +6, or I,I at −5 and −6 and C,C at +5 and +6. The DNA was purified and self-hybridized as described above. The oligonucleotides were labeled with [λ^{32}P]dATP and T4 polynucleotide kinase and the duplex DNA was separated from non-duplex DNA by PAGE. The gel retardation assays with purified Hin were performed as described by Glasgow *et al.*, (1989a). Filter binding assays were done according to Halvorsen *et al.* (1991) using the binding conditions identical to the gel retardation assays (Glasgow *et al.*, 1989).

Synthesis of 5′-O-dimethoxytrityl-2′-deoxy-3-deazaadenosine ('Compound 1')

A sample (66 mg, 0.265 mmol) of 2′-deoxy-3-deazaadenosine, generously given to us by Dr Thomas Krenitsky, dried three times by coevaporation with 1 ml each of pyridine, was dissolved in 0.5 ml of anhydrous pyridine. To this solution was added 107 mg (0.315 mmol) of dimethoxytritylchloride and the reaction mixture was stirred under argon for 4 h. It was then evaporated to dryness and the residue taken up in 5 ml of CH_2Cl_2. The organic layer was washed successively with two 1.5 ml portions of saturated aqueous NaCl, dried over Na_2SO_4 and evaporated. The residue was chromatographed on a short silica gel column (2 cm × 10 cm) using chloroform:methanol (9:1) as eluant. Evaporation of appropriate fractions yielded 98 mg (67%) of Compound 1 as a white foam and was used without further purification in the next reaction.

Synthesis of 3′-O-β-cyanoethyl-N,N-diisopropylaminophosphoramidite of 2′-deoxy-5′-O-dimethoxytrityl-3-deazaadenosine ('Compound 2')

Compound 1 (98 mg, 0.177 mmol) obtained from the previous reaction was placed in a 10 ml round bottom flask with a stir bar, capped with a rubber septum. A small needle was inserted through the septum and the entire assembly was dried overnight over KOH pellets. The flask was flushed with argon and dry CH_2Cl_2 (1.5 ml, filtered over basic Al_2O_3) was added through a glass syringe. Diisopropylethylamine (0.123 ml, 0.706 mmol) was added followed by the addition of 2-cyanoethyl-*N,N*-diisopropylchlorophosphoramidite (0.079 ml, 0.354 mmol). The reaction mixture was stirred at room temperature for 0.5 h. TLC (EtOAc:Et_3N = 9:1) found no trace of the starting material. Methanol (0.01 ml) was added to destroy the excess phosphine. The entire contents of the flask were then transferred to a 25 ml separatory funnel with ~7 ml of EtOAc. The organic layer was washed with two 5 ml portions of 10% aqueous Na_2CO_3, one 5 ml portion of saturated aqueous NaCl and dried over Na_2SO_4. Evaporation of this solution yielded an oil which was chromatographed on a silica gel column (2 cm × 10 cm) using EtOAc:Et_3N (9:1) as eluant. Evaporation of desired fractions afforded 65 mg (45%) of the Compound 2 as a foam which was used without further purification in the DNA synthesis of oligonucleotides containing deazaadenine residues.

Acknowledgements

We are very grateful to Dr Thomas Krenitsky and Burroughs Wellcome Co. for generously giving 3-deaza-2′-deoxyadenosine. We thank Anna Glasgow, Heon Lim and Reid Johnson for providing some of the purified Hin protein used in this study, Stanley Brown for pointing out to us that the 12T:A to 12G:C symmetrical change in the *hix* sequence resulted in a *dam*-methylation site, Phoebe Rice for suggesting that we look at binding to *hix* sites with I:C base pairs, and Howard Nash and Kevin Plaxco for critically reviewing the manuscript. This work was supported by Public Health Service grants GM43149 and RR-07096 from the National Institutes of Health (awarded to K.T.H.). K.T.H. is a recipient of a Junior Faculty Research Award from the American Cancer Society.

References

Affolter,M., Percival-Smith,A., Müller,M., Billeter,M., Qian,Y.Q., Otting,G., Wüthrich,K. and Gehring,W.J. (1991) *Cell*, **64**, 879–880.
Aitken,C.R. and Gumport,R.I. (1991) *Methods Enzymol.*, **208**, 433–457.
Anderson,W.F., Ohlendorf,D.H., Takeda,T. and Matthews,B.W. (1981) *Nature*, **290**, 754–758.
Bass,S., Sugiono,P., Arvidson,D.N., Gunsalas,R.P. and Youderian,P. (1987) *Genes Dev.*, **1**, 565–572.
Benson,N., Sugiono,P., Bass,S., Mendelman,L.V. and Youderian,P. (1986) *Genetics*, **118**, 21–29.
Boocock,M.R., Brown,J.L. and Sherratt,D.J. (1986) *Res. Biochem. Soc. Trans.*, **14**, 214–216.
Boocock,M.R., Brown,J.L. and Sherratt,D.J. (1987) In Kelly,T.J. and MacMacken,R. (eds), *DNA Replication and Recombination*. Alan R.Liss, New York, pp. 703–718.
Bruist,M.F., Glasgow,A.C., Johnson,R.C. and Simon,M.I. (1987a) *Genes Dev.*, **1**, 762–772.
Bruist,M.F., Horvath,S.J., Hood,L.E., Steitz,T.A. and Simon,M.I. (1987b) *Science*, **235**, 777–780.
Davis,R.W., Botstein,D. and Roth,J.R. (1980) *Advanced Bacterial Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Glasgow,A.C., Bruist,M.F. and Simon,M.I. (1989a) *J. Biol. Chem.*, **264**, 10072–10082.
Glasgow,A.C., Hughes,K.T. and Simon,M.I. (1989b) In Berg,D. and Howe,M. (eds), *Mobile DNA*. American Society for Microbiology, Washington, DC, pp. 637–659.
Graña,D., Youderian,P. and Susskind,M.M. (1985) *Genetics*, **110**, 1–16.
Halvorsen,Y.C., Nandabalan,K. and Dickson,R.C. (1991) *Mol. Cell. Biol.*, **11**, 1777–1784.
Heichman,K.A. and Johnson,R.C. (1990) *Science*, **249**, 511–517.
Hughes,K.T., Youderian,P. and Simon,M.I. (1988) *Genes Dev.*, **2**, 937–948.
Johnson,R.C. and Bruist,M.F. (1989) *EMBO J.*, **8**, 1581–1590.
Johnson,R.C. and Simon,M.I. (1985) *Cell*, **41**, 781–791.
Johnson,R.C., Bruist,M.F. and Simon,M.I. (1986) *Cell*, **46**, 531–539.
Johnson,R.C., Glasgow,A.C. and Simon,M.I. (1987) *Nature*, **329**, 462–465.
Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.
Kutsukake,K. and Iino,T. (1980) *Nature*, **284**, 479–481.
Lewis,M., Jeffrey,A., Wang,J., Ladner,R., Ptashne,M. and Pabo,C.O. (1982) *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 435–440.
Mack,D.P., Sluka,J.P., Shin,J.A., Griffin,J.H., Simon,M.I. and Dervan,P.B. (1990) *Biochemistry*, **29**, 6561–6567.
McClarin,J.A., Frederick,C.A., Wang,B.-C., Greene,P., Boyer,H.W., Grable,J. and Rosenberg,J.M. (1986) *Science*, **234**, 1526–1541.
McKay,D.B. and Steiz,T.A. (1981) *Nature*, **290**, 744–749.
McKay,D.B., Weber,I.T. and Steitz,T.A. (1982) *J. Biol. Chem.*, **27**, 9518–9524.
Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mitchell,P.J. and Tjian,R. (1989) *Science*, **245**, 371–378.
Newman,P.C., Nwosu,V.U., Williams,D.M., Cosstick,R., Seela,F. and Connolly,B.A. (1991) *Biochemistry*, **29**, 9891–9901.
Otting,G., Qian,Y.Q., Billeter,M., Müller, Affolter,M., Gehring,W.J. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.
Pabo,C.O. and Lewis,M. (1982) *Nature*, **298**, 443–447.
Perry,K.L. and Walker,G.C. (1982) *Nature*, **300**, 278–281.
Plasterk,R.H.A., Brinkman,A. and van de Putte,P. (1983) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5476.
Sanderson,K.E. and Roth,J.R. (1983) *Microbiol. Rev.*, **47**, 410–453.
Sauer,R.T., Yocum,R.R., Doolittle,R.F., Lewis,M. and Pabo,C.O. (1982) *Nature*, **298**, 447–451.
Schevitz,R.W., Otwinowski,O., Joachimiak,A., Lawson,C.L. and Sigler,P.B. (1985) *Nature*, **317**, 782–786.
Seeman,N.C., Rosenberg,J.M. and Rich,A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 804–808.
Sluka,J.P., Horvath,S.J., Bruist,M.F., Simon,M.I. and Dervan,P.B. (1987) *Science*, **238**, 1129–1132.
Sluka,J.P., Horvath,S.J., Glasgow,A.C., Simon,M.I. and Dervan,P.B. (1990) *Biochemistry*, **29**, 6551–6561.
Staacke,D., Walter,B., Kisters-Woike,B., Wilcken-Bergmann,B.V. and Müller-Hill,B. (1990) *EMBO J.*, **9**, 1963–1967.
Starr,D.B. and Hawley,D.K. (1991) *Cell*, **67**, 1231–1240.
Steitz,T.A., Weber,I.T. and Matthews,B.M. (1982) *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 419–426.
Stocker,B.A.D. (1949) *J. Hygiene.*, **47**, 398–413.
Susskind,M.M. and Botstein,D. (1978) *Microbiol. Rev.*, **42**, 385–413.
Susskind,M.M. and Youderian,P. (1983) In Hendrix,R.W., Roberts,J.W., Stahl,F.W. and Weisberg,R.A. (eds), *Lambda II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 347–363.
Vogel,H.J. and Bonner,D.M. (1956) *J. Biol. Chem.*, **218**, 97–106.
Wasserman,S.A., Dungan,J.M. and Cozzarelli,N.R. (1985) *Science*, **229**, 171–174.
Yang,C.C. and Nash,H.A. (1989) *Cell*, **57**, 869–880.
Youderian,P., Vershon,A., Bouvier,S., Sauer,R.T. and Susskind,M.M. (1983) *Cell*, **35**, 777–783.

Received on August 2, 1991; revised on March 27, 1992.