Nucleic Acids Research, 2001, Vol. 29, No. 19 **3955–3964**

Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation

Dominic Esposito*, Julia S. Thrower and John J. Scocca

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

Received June 26, 2001; Revised and Accepted August 17, 2001

ABSTRACT

A fundamental step in site-specific recombination reactions involves the formation of properly arranged protein–DNA structures termed intasomes. The contributions of various proteins and DNA binding sites in the intasome determine not only whether recombination can occur, but also in which direction the reaction is likely to proceed and how fast the reaction will go. By mutating individual DNA binding sites and observing the effects of various mixtures of recombination proteins on the mutated substrates, we have begun to categorize the requirements for intasome formation in the site-specific recombination system of bacteriophage HP1. These experiments define the binding site occupancies in both integrative and excisive recombination for the three recombination proteins: HP1 integrase, HP1 Cox and IHF. This data has allowed us to create a model which explains many of the biochemical features of HP1 recombination, demonstrates the importance of intasome components on the directionality of the reaction and predicts further ways in which the role of the intasome can be explored.

INTRODUCTION

The tyrosine recombinase family of proteins mediates the joining and separation of multiple fragments of DNA by means of a precisely controlled set of site-specific recombination reactions (1–3). The numerous family members are classified by their cellular function, the complexity of their DNA substrates and the proteins required to carry out the reaction. Simple systems, including the Cre recombinase of phage P1 (4), the Flp recombinase of the yeast 2μ plasmid (5) and the Xer proteins found in numerous bacteria (6), mediate recombination at short sequence-specific pieces of DNA and regulate plasmid copy number, decatenate daughter molecules, or resolve dimeric species in order to properly segregate the DNA during cell division. Considerably more complex substrates are used in many bacteriophage recombination systems, such as that of phage λ, which are required for the insertion and excision of phage DNA from the host bacterial chromosome (7–10). In a number of these systems additional host- or phage-encoded proteins are required and more complicated arrangements of protein-binding sites on the DNA are observed. The complexity observed in these systems may be a sign of the requirement for directionality; protein requirements needed for integration of the phage into the host are different from those needed for excision. In fact, in several systems conditions favoring integration inhibit excision, and vice versa.

We are interested in the mechanism and regulation of a complex site-specific recombination system from the *Haemophilus influenzae* temperate phage HP1. HP1, a member of the P2 family of bacteriophages, integrates and excises from the *H.influenzae* genome in a reaction superficially analogous to the reaction first studied in bacteriophage λ (9). The HP1 genome is inserted into a unique site in the host chromosome by a site-specific recombination reaction carried out by the HP1 integrase protein, stimulated by the action of the host-encoded integration host factor (IHF) (11). The integration substrates are a 500 bp phage attachment site (*attP*), containing binding sites for integrase and IHF, and an 18 bp host site (*attB*), which lies at the anticodon stem–loop of a host leucyl-tRNA gene (12). HP1 is a member of a diverse group of bacteriophages that insert into tRNA genes (13). The phage carries with it a large region of DNA identical to the host which encodes for the rest of the downstream tRNA operon. Recombination restores the complete tRNA genes, ensuring that they can still be produced after integration. Once recombination takes place, the integrated prophage is flanked by recombinant *attL* and *attR* sites, which serve as the substrates for excisive recombination (14). This reaction again requires HP1 integrase and IHF, but also requires the action of an additional phage-encoded protein, Cox (15).

The actual chemical steps of recombination take place in synaptic complexes, called intasomes, in which the two substrate DNA molecules are held together by integrase and accessory proteins. Little is known about the specific nature of these complexes. In order to understand the mechanism of recombination, it is essential that we identify the structure of

*To whom correspondence should be addressed at present address: Invitrogen Corporation, 9800 Medical Center Drive, Rockville, MD 20850, USA. Tel: +1 301 610 8206; Fax: +1 301 610 8251; Email: dominic.esposito@invitrogen.com
Present address:
Julia S. Thrower, Department of Biochemistry, University of California at Berkeley, Berkeley, CA, USA

these intasomes and determine the differences between the intasome structures utilized for integration and excision. For instance, it is known that HP1 Cox functions by binding to the *attL* DNA and altering the binding properties of a neighboring integrase-binding site (15). Presumably, this interaction alters the structure and composition of the intasome to create a complex which favors excision, but inhibits integration. The binding sites for the various recombination proteins involved in HP1 recombination (see Fig. 6) have previously been identified (11,16). The *attP* substrate contains six sets of integrase-binding sites (IBS) arranged in pairs of direct or inverted repeats. HP1 integrase has two DNA binding specificities: it binds tightly to type II sites (IBS1, IBS2 and IBS5) and more weakly to type I sites (IBS3, IBS4 and IBS6) (16). IBS4 is the site at which recombination actually occurs and recombination proceeds between this site and the single type I site, IBSB, in the *H.influenzae* chromosomal *attB* site. Two adjacent Cox-binding sites are present in *attP*, positioned between the IBS4 and IBS5 sites on the *attL* arm of the substrate (15), and two IHF-binding sites were identified, one on either side of the site of recombination (11). Here we describe experiments in which these binding sites were individually mutated and the effects on recombination with the mutated substrates were studied. By observing the effects of mutations at various sites and their differential effect on integration and excision, we propose models for the structure of the HP1 integrative and excisive intasomes. Such models serve as useful templates for future experiments to examine the detailed mechanism of the recombination reaction and may prove valuable in understanding recombination reactions catalyzed by other members of the tyrosine recombinase family.

MATERIALS AND METHODS

Proteins and DNA substrates

Highly purified HP1 integrase was obtained from S. Waninger and was purified as previously described (17). *Escherichia coli* IHF protein was purified from *E.coli* K5746 as described (18). HP1 Cox was purified as previously described (14); Cox protein lacking contaminating IHF was produced by expressing the pCOX1 plasmid in *E.coli* RW1815, which carries a deletion in the gene encoding one of the IHF subunits, and was purified as described (15). Recombination substrates have been previously detailed; pHPC120 carries the *attP* segment, pHPC121 carries the *attB* segment, pHPC122 carries the *attL* segment and pHPC123 carries the *attR* segment (15,16).

Oligonucleotide primers

Synthetic oligonucleotides used in this study were obtained from Life Technologies. The primers used in the construction of plasmids were: IBS1M, 5′-GATTTCAGTTAATTGACTTAAGGTTATATTCATTGTTTAAAGACAGG; IBS2M, 5′-CTAGATATTTATAACATACACTTTGTGATCAGTTGAATATTATTG; IBS3M, 5′-CGAACCTTCGACCATTAATTAAAAGTAAACTGCTCTACCGACTG; IBS4M, 5′-CGCCTCGAAAGGCGATTTATTTTAAATAAAATAGCTAAACCAG; IBS5M, 5′-CCCTCTATTTTACTTTATATTGGCTTTATGTTTTGTCTTA; IBS6M, 5′-GGGGAAGCTTGTACTTTTCCCTCTATTTTACTGGC; PPLI5, 5′-CCCTGTATTTTACTGAATATTGGCTGTATGTTTTGTCTTATATTTAC; H1M, 5′-AAAATTCACTTTAAAATATTCTACCGGGGTCGTTAGCTCAGTC; H2M, 5′-GCAATTAATGTTGCTATCGGTAATATTTTATACTGGTTTAGC.

Recombination assays

Single time point reactions were carried out in 20 µl volumes containing 50 mM Tris–phosphate, pH 7.5, 15 mM EDTA, 20% glycerol, 6 mM spermidine, 90 mM KCl and 100 fmol (5 nM) each substrate. Linear substrates were cut with restriction endonucleases and radiolabeled using the Klenow fragment of *E.coli* DNA polymerase I and [^{32}P]dATP as previously described (14). The specific activity of the radioactive substrates was 500 d.p.m./fmol. A standard integration assay contained 75 ng integrase (60 nM) and 80 ng IHF (150 nM). A standard excision assay contained 180 ng Cox (1 µM), 80 ng IHF (150 nM) and 600 ng integrase (770 nM). After incubation at 37°C for 30 min, 10× stop mix (10% SDS, 0.1% bromophenol blue and 0.1% xylene cyanol) and 1 µl of proteinase K (4 mg/ml) were added, the samples were heated for 15 min at 65°C and the DNA was separated on 1% agarose gels. The gels were stained with ethidium bromide, visualized with an Ambis CCD camera and then dried down onto DE81 paper and exposed to a Fuji BAS1000 imaging plate for 1–2 h. MacBAS software (Fuji) was used to quantitate the substrate and product bands. Reaction rates were measured with time courses of varying lengths. Master recombination reactions were set up, 20 µl aliquots were removed at the required times and the reactions were stopped by addition of 10× stop mix. Rates in fmol product formed per min were determined by calculating the slope of the linear portion of a plot of percent recombination versus time. In general, time points within the first 15 min of the start of the reaction were used to determine reaction rates; in reactions with significantly reduced activity time points of up to 1 h were used.

Construction of attachment sites with mutated binding sites

Mutations introduced into the Cox-binding sites were previously described (15). Mutations were introduced into the integrase- and IHF-binding sites using one of two methods. The first method, the Transformer Site-Directed Mutagenesis kit (Clontech), involves the use of two primers; one to eliminate a restriction site and the other to introduce the desired mutation. Additionally, some mutants were constructed using MR–PCR (19), in which a mutated restriction site is used to select against molecules which fail to receive the desired mutation. The mutations introduced are summarized in Tables 1 (for the IBS sites) and 2 (for the IHF sites). Mutations in IBS1, IBS2 and IBS3 were constructed in *attP* and in *attR* using the IBS1M, IBS2M and IBS3M primers, respectively. The IBS4 mutation in *attP* was introduced with primer IBS4M. A plasmid containing a deletion of the right half (IBS6b) of IBS6 (pPL675) has been described previously (12). A separate plasmid containing a deletion of the whole IBS6 site (pPL665) was constructed by PCR amplification of the pPL675 template using primer IBS6M. Vector sequences which replaced the deleted regions of the IBS6 sites are indicated in Table 1. An IBS6b mutant in *attL* was constructed by recombining pPL675 with wild-type *attB*, isolating the recombinant product, digesting the DNA with *Eco*RI and recircularizing the products

Table 1. Mutations introduced into the integrase-binding sites (IBS)

		Wild-type site	Mutant site
Type II integrase-binding sites	IBS1a	ACTGGCGGTT	ACTtaaGGTT
	IBS1b	AATGGCGATA	AATGaatATA
	IBS2a	ACTGGCGACA	ACTGatcACA
	IBS2b	ACTGGCGGTT	ACTGtatGTT
	IBS5a	ACTGGCGATT	ACTttatATT
	IBS5b	GCTGGCGGTT	GCTttatGTT
	Consensus	ACTGGCGRTW	
Type I integrase-binding sites	IBSBa	AGGGATTT	AtttATTT
	IBSBb	AGGGATTT	AtttATTT
	IBS4a	AGGGATTT	AtttATTT
	IBS4b	TGGGATTT	TtttATTT
	IBS3a	ACGGATTA	AttaATTA
	IBS3b	GCGGACTT	GtttACTT
	IBS6a	AAGGATTT	Agcttgcc
	IBS6b	AGGGATTA	ggtcataa
	Consensus	AGGGATTW	

The consensus sequence of the type I and type II integrase-binding sites, the sequences of the 12 individual half-sites and the mutations introduced into the IBS mutants are shown. All of the IBS mutants described in the text, with the exception of IBS6, were made simultaneously in both half-sites (a and b) of the site. The IBS6 mutants were constructed by deleting DNA from the end of the *attP* fragment by PCR; the sequences indicated are vector sequences which substitute for the IBS6 binding sites.

Table 2. Mutations introduced into the IHF-binding sites

		Score	Relative affinity (%)
Wild-type H1	AATCATCTAATTG	48.3	100
Wild-type H2	AAACAATAACTTA	50.2	100
Mutant H1	AATatTCTAccgG	12.4	2.3
Mutant H2	AAAatATtACcgA	16.8	3.0
Consensus	WWWCAAnnAnTTR		

The published consensus sequence of IHF-binding sites, the sequences of the two HP1 *attP* IHF sites and the mutations introduced into the IHF mutants are shown. The predicted site score as determined by a computer analysis modified from Goodrich *et al*. (20) and the relative IHF binding affinity predicted for the mutant sites as compared to their wild-type counterparts are also indicated.

with T4 DNA ligase. The *attL* IBS6b mutant was identified by the pattern of bands produced by *Dra*I restriction digestion. An *attL* plasmid containing a deletion of the entire IBS6 site was constructed by PCR-mediated deletion of the wild-type *attL* plasmid using primer IBS6M. Double mutations in IBS5 and IBS6, as well as IBS5 and IBS6b, were constructed using the *attL* IBS6 mutants as the template for MR–PCR mutagenesis using the PPLI5 primer. Mutations were also introduced into the IHF sites of *attP*, *attL* and *attR*. The H1M primer was used to make mutations in the H1 site of *attP* and *attR*, while the H2M primer was used to make mutations in the H2 site of *attP* and *attL*. The H2 mutant in *attL* was also combined with the IBS5 mutation through another round of MR–PCR using primer IBS5M.

Retardation on agarose gels

Plasmids containing the *attP* or *attR* sites were digested with *Hin*dIII and *Eco*RI and the fragments purified using Magic columns (Promega). DNA fragments (2 µg) were incubated with IHF in 0.1 M Tris–phosphate, pH 7.5, 5 mM EDTA, 0.07 M KCl, 5% glycerol and marker dyes (bromophenol blue/xylene cyanol). After 15 min at 25°C, the samples were separated by electrophoresis on 2% agarose gels in 0.5× TBE buffer (25 mM Tris, 50 mM boric acid, 2.5 mM EDTA) at 4 V/cm for 2 h. The gel was stained with ethidium bromide, destained and digitized with an Ambis CCD system.

RESULTS

Introduction of mutations into the recombinant protein-binding sites

Tables 1 and 2 summarize the mutations made in the binding sites as described in Materials and Methods. IBS site mutations (Table 1) were designed to eliminate the strictly conserved GGCG sequence of the type II sites or the SGG sequence of the type I sites. These mutant binding sites were unable to be shifted by integrase in gel retardation experiments and showed no detectable DNase I footprinting, demonstrating that these mutations effectively eliminated integrase binding (data not shown). Because the spacing between the two halves of the IBS6 site differs from that seen in the other sites (9 bp between 8 bp repeats, as opposed to 2 bp in IBS3 and 1 bp in IBS4), we decided to construct mutants with both the entire IBS6 site removed or with just the right half-site removed (referred to as IBS6b) to see if the two half-sites had different effects on recombination. Cox-binding site mutants and gel retardation experiments showing their elimination of Cox binding have been previously described (15). IHF has a defined binding sequence; computer analysis using an IHF site searching algorithm can predict the relative binding affinity of various IHF-binding sites and mutants (20,21). As shown in Table 2, the mutated H1 and H2 sites are predicted to have <3% of the binding affinity of the wild-type binding sites. Gel retardation analysis of the IHF site mutants demonstrates that the sites failed to bind IHF, as shown in Figure 1. The wild-type *attP* substrate shows two IHF-induced shifts (due to binding to a single site or to both the H1 and H2 sites), while the *attP* H2 mutant exhibits only a single shift (due to the presence of only the wild-type H1 site). The *attR* H1 mutant is not shifted by IHF, as no functional IHF sites are present on the plasmid.

The role of the integrase-binding sites in recombination

The results of integrative recombination assays on the *attP* IBS mutants are shown in Figure 2A. Mutations in IBS2, IBS4 and IBS5 completely abolish integration. In assays using the mutated IBS2 or IBS5 sites, no detectable recombination was

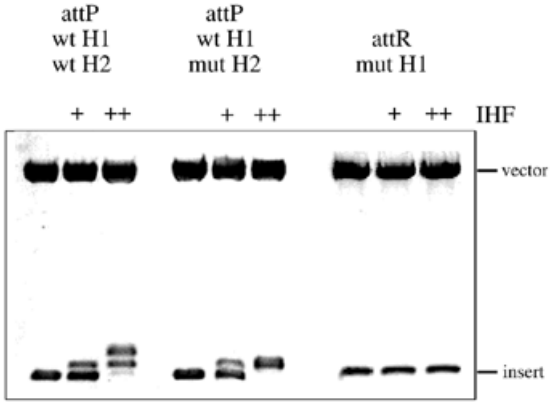

Figure 1. Gel retardation of mutated IHF-binding sites. Restriction fragments (1.5 µg) of wild-type *attP*, *attP* H2 mutant and *attR* H1 mutant were incubated with varying amounts of IHF (–, 0 ng; +, 20 ng; ++, 50 ng) for 15 min as described in Materials and Methods. Samples were loaded on a 2% agarose gel and electrophoresed for 2 h at 60 V. The gel was stained with ethidium bromide and digitized with an Ambis CCD camera. The top band is the 2700 bp vector backbone, from which the 400 bp insert containing the IHF sites has been excised.

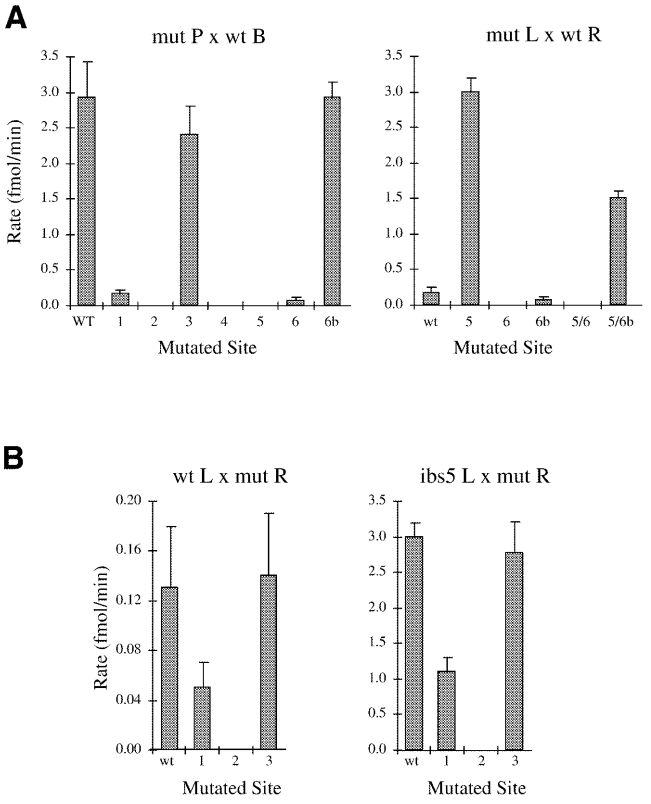

Figure 2. Effect of the IBS mutants on integrative and excisive recombination. (**A**) (Left) standard integration time courses with 5 nM linearized *attB* were performed with 5 nM supercoiled *attP* plasmids containing mutations in the indicated IBS sites and 100 nM integrase. Time points were taken at 0, 5, 10, 20 and 30 min (0, 60, 120 and 240 for IBS6a). Rates of integration were determined by plotting the product formed over time and are the averages of three separate determinations for each mutant. Error bars correspond to the standard deviation of the results. (Right) Standard excision time courses were carried out in a similar manner with combinations of supercoiled *attL* and linear *attR* plasmids with mutations in the indicated IBS sites. Wild-type *attR* substrates were recombined with mutant *attL* substrates. (**B**) (Left) Reactions involving wild-type *attL* substrates and mutant *attR* substrates. (Right) The same reactions but using an *attL* substrate with the IBS5 site mutated.

observed even after overnight incubation. The lower limit of detection of this assay is estimated to be <0.2% recombination. Substrates containing the IBS1 mutation produced a very small but detectable amount of activity, ~6% that of wild-type, while substrates containing the IBS3 mutation maintained 82% of wild-type activity. The loss of IBS6b had no effect on the reaction, while loss of the whole of the IBS6 site eliminated all but 2% of wild-type activity.

The effects of the IBS mutations on the excision reaction were investigated next. The right side of Figure 2A shows the activity of IBS mutants in *attL* in reactions with wild-type *attR*. As we have previously shown, mutation of IBS5 dramatically enhances the amount of excision; this mutation mimics the removal of integrase from IBS5, which is normally carried out by the Cox protein (15). Deletion of both repeats of the IBS6 site eliminates all excision activity, whether IBS5 is wild-type or mutated. However, removal of the IBS6b half-site only reduces the level of excision 2-fold, again independent of the mutation status of IBS5. Figure 2B shows the effects on excision of mutants in *attR*, using wild-type *attL* or IBS5 mutant *attL* as substrate. In both sets of data it is clear that the loss of IBS2 completely eliminated excision. Also, as in integration, the loss of IBS3 had minimal effect; in excisive recombination, the mutated IBS3 substrate was recombined at a rate equivalent, within error, to the wild-type substrate. The effects of IBS1 mutations were less dramatic in excision than in integration. Loss of IBS1 reduced activity by slightly more than 2-fold. As expected, there were no qualitative differences between the results seen with IBS5 mutant *attL* and wild-type *attL*.

Requirement for IHF in HP1 recombination

Unlike integration carried out by phage λ Int, HP1 integrative recombination does not require the presence of IHF (16). Instead, IHF has a significant stimulatory effect on the reaction, as shown in Figure 3A. Over a broad range of IHF concentrations integrative recombination is stimulated up to 5-fold over reactions carried out in the absence of IHF. Initial results suggested that IHF was also not required for excisive recombination. As shown by the first column of Table 3, our original data seemed to suggest that addition of IHF had no effect on the levels of excisive recombination. However, quite different results were seen when samples of ^{3}H-labeled Cox were used in the assay. The purification of this protein included a step in which guanidine hydrochloride was used to denature the protein and thus it was possible that our original preparations of Cox contained a contaminant which was no longer present in these differentially purified samples. When overloaded on an SDS–polyacrylamide gel, our original Cox protein clearly showed a contaminating doublet of 12 kDa proteins representing <2% (w/w) of the total protein. N-terminal sequencing showed that these contaminants were in fact the two subunits of *E.coli* IHF. Using a strain deleted for one of the IHF subunits (*E.coli* RW1815), Cox extracts were produced which were free of contaminating IHF, and the

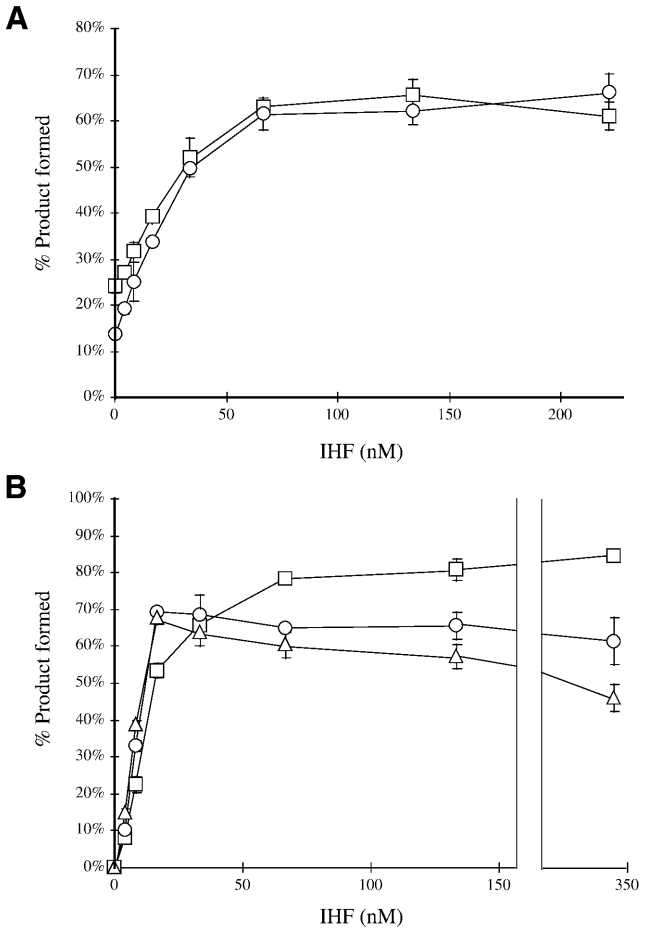

Figure 3. Stimulation of recombination by IHF. (**A**) Standard integration assays were carried out at varying concentrations of IHF, using linear *attB* and supercoiled *attP* templates, 40 nM (circles) or 100 nM (squares) integrase and a 30 min reaction time. (**B**) Standard excision assays were carried out at varying concentrations of IHF, using linear *attR*, supercoiled *attL* mutated at IBS5 and 1 μM Cox and a 20 min reaction time. The concentration of integrase was either 1.4 μM (squares), 710 nM (circles) or 355 nM (triangles). Each data point represents the percentage of labeled substrate converted to product and is the average of three separate experiments, with error bars corresponding to the standard deviation of the three values.

results of excision assays using this protein are shown in the right column of Table 3. It is clear from this data that IHF in fact has a dramatic effect on the excision reaction. Unlike integration, which proceeds to a limited extent in the absence of IHF, no detectable excision activity was observed in the absence of IHF. Stimulation by IHF, however, required only very low levels of the protein; saturation of the effect occurred around 20–30 nM IHF (Fig. 3B). Curiously, in the presence of sub-optimal levels of integrase, IHF produced a low, but reproducible, level of inhibition at concentrations >150 nM, while in the presence of optimal levels of integrase, IHF continued to slightly stimulate the reaction at high concentrations.

The role of the IHF-binding sites in recombination

The results of integration assays with substrates containing IHF site mutations are shown in Figure 4A. Mutation of the H2 site appears to have little effect on the integration reaction, but removal of the H1 site nearly eliminates IHF stimulation of integration. In the presence of a mutated H1 site, the reaction is stimulated less than 2-fold by IHF; levels of IHF up to 2 μM do not further stimulate the reaction (data not shown). The effects of the IHF site mutations on excision are shown in Figure 4B. The substrate containing a mutation in the H2 site was completely unable to support excision; even after 6 h, no recombination product was detected. Recombination with *attR* containing a mutated H1 site produced a 2- to 3-fold reduction in excision rate compared to wild-type *attR*. In addition, as shown in Figure 5, the substrate with a mutation in the H1 site showed a slight inhibition by high concentrations of IHF at sub-optimal integrase levels; this effect is identical to that seen with the wild-type substrates (see Fig. 3B), suggesting that this phenomenon is related to IHF binding at the H2 site and not to interactions at H1.

Table 3. The source of Cox protein affects the IHF requirements of excisive recombination

IHF concentration (nM)	Excisive recombination (% product formed)		
	Cox(DH5α)	[^{3}H]Cox	Cox(RW1815)
0	65 ± 8	0	0
8	69 ± 7	18 ± 4	23 ± 3
33	72 ± 4	58 ± 7	66 ± 5
67	70 ± 3	72 ± 5	78 ± 3
133	77 ± 6	84 ± 7	81 ± 4

Levels of excisive recombination at a range of IHF concentrations in the presence of Cox protein from three different sources are given. Cox(DH5α) was purified as previously described from *E.coli* DH5α containing the Cox expression vector pCOX1. [^{3}H]Cox was produced from pCOX1 in *E.coli* BL21(DE3) and was purified by a method involving denaturation with guanidine hydrochloride. Cox(RW1815) was produced from pCOX1 in *E.coli* RW1815, which contains a deletion of one subunit of IHF, and was purified by the native purification procedure. Values shown are the percent of substrate converted to product in a 30 min reaction and are the averages (and standard deviations) of three experiments.

DISCUSSION

Mutation of the binding sites for the three recombination proteins has allowed us to determine the regions necessary for integrative and excisive recombination in HP1. These data are graphically summarized in Figure 6, which separates the sites into four categories based on the observed effects of binding site mutations: required if mutation eliminates all recombination activity; stimulatory if some, but not all, recombination activity is lost; dispensable if >80% of the wild-type level of recombination remains; inhibitory if mutation of a site enhances recombination activity.

HP1 integrase is the protein directly involved in the chemistry of the recombination reaction and is the only one of the three proteins with any known catalytic activity. In addition, HP1 integrase is intimately involved in formation of the proper protein–DNA structures, known as intasomes, which are needed for recombination and which in some way

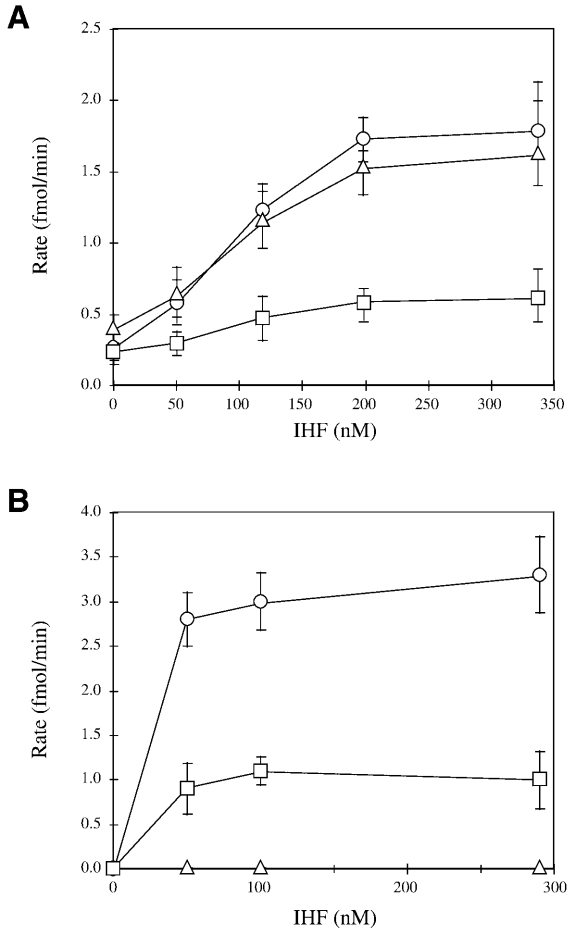

Figure 4. Effect of IHF-binding site mutations on integrative and excisive recombination. (**A**) Standard integration time courses using supercoiled *attP* and linearized *attB* were carried out at varying IHF concentrations using 100 nM integrase and taking time points at 0, 10, 20 and 30 min. The rate of integration was plotted against IHF concentration for the reaction with wild-type *attP* (circles), the reaction with *attP* mutated at H1 (squares) and the reaction with *attP* mutated at H2 (triangles). Each data point is the average of three separate determinations; error bars represent the standard deviation of the values. (**B**) Standard excision time courses using supercoiled *attL* and linearized *attR* were carried out at varying IHF concentrations using 800 nM integrase and 1 μM Cox and taking time points at 0, 5, 10, 15 and 30 min. The rate of integration was plotted against IHF concentration for the reaction with wild-type IHF sites (circles), the reaction with a mutant H1 site in *attR* (squares) and the reaction with a mutant H2 site in *attL* (triangles). In the case of the mutant H2 site, no detectable excisive recombination was observed, indicating a rate of recombination of <0.02 fmol/min.

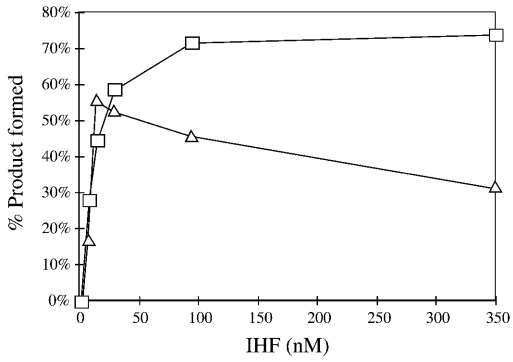

Figure 5. H1 IHF site mutants still inhibit at high IHF levels and low integrase concentrations. This set of excision reactions contains supercoiled *attL* with an IBS5 mutation and linearized *attR* containing an H1 mutation. The percent recombination at varying IHF concentrations is shown. Cox was present at 1 μM in these reactions and integrase was present at either 1410 nM (squares) or 300 nM (triangles).

must control the directionality of recombination. From the data in Figure 2 it is clear that other than IBS4, where the actual cleavage and strand exchange reactions occur, only IBS2 is completely essential for both integration and excision. Presumably, this site is vitally important for forming the basic structure of the reactive intasome. However, since its effect is seen equally in both integration and excision, IBS2 is unlikely to affect directionality. IBS5, on the other hand, is clearly a determinant of directionality. In the absence of IBS5, integration does not occur; in fact, inhibition of integration by Cox occurs through Cox-mediated abrogation of integrase binding at IBS5. However, IBS5 is not required for excision and, in fact, is inhibitory to excisive recombination, as shown by the dramatic increase in excision activity in the absence of an active IBS5 site (15). This implies a role for integrase binding at IBS5 in the formation of a protein–DNA structure which somehow favors integration but which is detrimental to excision. The roles of IBS1 and IBS6 are less clear; both show differential effects on integration and excision, but are at least partially dispensable for activity. These data also show that IBS3 is completely dispensable for both integrative and excisive recombination. The IBS3 site lies in a region of homology between *attP* and the host DNA which contains the operon of three tRNA genes into which HP1 inserts itself. HP1 integrates into the IBS4 site, which lies in the anticodon stem–loop of the $tRNA^{Leu}$ gene. The DNA making up the IBS3 site lies in the anticodon stem–loop of the neighboring $tRNA^{Lys}$ gene, implying that it may simply be a fortuitous binding sequence and is not actually used in the reaction. The sequence has most likely remained intact because mutations in this sequence could destroy the structure and activity of the essential $tRNA^{Lys}$ gene. We have never observed integration into the IBS3 site either *in vitro* or *in vivo*, which suggests that the subtle difference in DNA sequence of IBS3 or the slight alteration in half-site spacing (1 bp between 8 bp repeats in IBS4, 2 bp between repeats in IBS3) must provide a very high level of specificity against recombination at IBS3 (D.Esposito and J.J.Scocca, unpublished observations). It is also possible, as our model may suggest, that formation of an active intasome at IBS3 might be prohibited by the spacing and orientation of the other protein-binding sites.

Surprisingly, IHF appears to provide an additional source of directional control in the HP1 recombination reaction. The data in Figure 4 offer evidence that the presence or absence of IHF at the H2 site acts as another determinant of directionality. The H2 mutation in *attL* completely eliminates excision, therefore, IHF must be bound at H2 for excision to proceed. In the integrative reaction, loss of IHF binding at H2 has a negligible effect on the reaction. The unlikely possibility that Cox binding could stimulate binding of IHF to H2, in much the same way that λ Xis affects the binding of IHF to a neighboring site (22,23), was also examined. The H2 site lies adjacent to the Cox site, but unlike the situation observed with

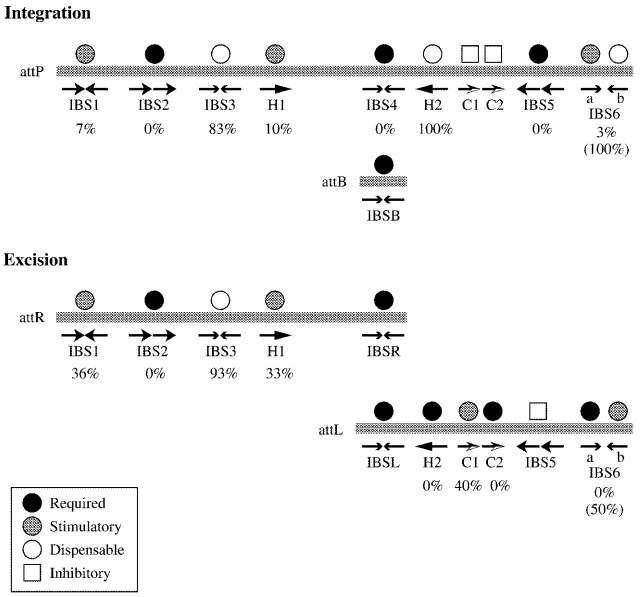

Figure 6. A graphical summary of the requirements for HP1 recombination. The recombination protein-binding sites in the HP1 attachment sites are shown, and the requirements for either integration (top) or excision (bottom) are shown. Percentages represent the amount of activity obtained in experiments using substrates mutated at the given site. The percentage indicated for IBS6 represents the complete deletion of both half-sites; the percentage of activity remaining when IBS6b alone was deleted is shown in parentheses. Circles schematically represent the strength of the requirement for bound protein at a given site. Black circles represent a completely required site while gray circles signify sites which stimulate recombination. Open circles represent sites which when mutated still produce levels of activity >80% of wild-type; these are identified as dispensable sites. Open squares represent sites which when bound by protein are inhibitory to recombination.

integrase at IBS5, there is no overlap of DNase I footprints between Cox and IHF bound at H2. Experiments using DNase I failed to detect any effect on the IHF occupancy of the H2 site in the presence of Cox (data not shown). The use of IHF to control directionality may represent a means of introducing cell cycle control on recombination. It has been shown that IHF levels in the cell vary depending on both growth conditions and growth stage (24,25). It is possible that the requirement for H2 binding for excision ensures that during exponential growth, when cellular levels of IHF are lowest, excision is inhibited, but during stationary phase, when IHF levels rise nearly 10-fold, excision can occur. Generally, it is believed that temperate phages prefer to stay integrated under conditions that favor cell growth; this enables the phage to be vertically transmitted to large numbers of progeny cells with no effort on its part. However, if growth conditions become poor, cells are unlikely to divide and may even begin to die. In this case, the phage may favor the excision pathway as a means of escaping death of the cell. IHF levels could be used to sense the state of cellular health and allow the phage to take this route.

Taken together, these data allow us to formulate a preliminary model of intasome formation in HP1 recombination. A number of facts are known about the binding of HP1 integrase to its binding sites. First, integrase in solution exists primarily in monomeric form under normal reaction conditions; however, gel retardation and equilibrium sedimentation experiments have shown that integrase is capable of forming dimers and tetramers through highly cooperative interactions (S.Waninger and J.J.Scocca, manuscript in preparation). The current data suggest that single integrase monomers bind sequentially and cooperatively to an IBS site, producing a complex with two protomers of integrase and one DNA dyad. In addition, HP1 integrase is capable of bridging two different types of binding sites. In phage λ recombination, the λ Int protein has been shown to bind core and arm sites (analogous to type II and type I sites in HP1) with two different binding regions on the same Int protein (26). Proteolysis and footprinting experiments with HP1 integrase also confirm that it can bind both type I and type II binding sites, bridging one type I and one type II IBS sites (16). Finally, we must note that the IBS2 and IBS5 sites are unique among the HP1 IBS sites in that they are direct repeats. It is unlikely that a dimer of integrase is capable of binding simultaneously to type I and type II sites with opposite orientations, but instead a dimer presumably must bridge a leftward-directed type I repeat and a leftward-directed type II repeat or two rightward-directed repeats. The symmetry observed in the crystallographic dimer of the HP1 integrase C-terminal region also predicts this mode of binding (27). Therefore, our model must take into account the interactions of integrase molecules bridging a direct and inverted repeat.

The basic model for the formation of the integrative intasome is shown in Figure 7. The model shows the synapsis of a complex of protein-bound *attP* with an unbound *attB* site. The *attP* DNA contains a DNA loop on the left (*attR*) arm mediated by IHF binding at the H1 site, which brings the IBS2 and IBS4 sites into proximity to each other (Fig. 7A). Since IBS2 is a direct repeat and IBS4 is an inverted repeat, only a half-site of each can be bridged by integrase. IBS5, the other vital site for integration, can then be aligned with IBS4 by forming a loop in the DNA on the other side of the attachment site (the *attL* arm). This will bring the IBS5 right half-site adjacent to the unoccupied half-site of IBS4, and the similar orientation of the repeats will allow another molecule of integrase to bridge these two sites. This leaves one half-site of IBS2 and one half-site of IBS5 free. Due to the high cooperativity of integrase binding, these sites would presumably be occupied by additional integrase from solution. These newly bound integrase protomers will have a free type I binding site available and can be used to recruit *attB* DNA (Fig. 7B). The core intasome thus constructed will be made up of an integrase tetramer and four IBS sites. Viewed from the top (Fig. 7C), this structure clearly allows synapsis of the two recombining strands (those containing IBS4 and IBSB). Such a model also explains the absolute requirement in integration for IBS2 and IBS5, which are both essential to forming the proper structure. The remainder of the *attP* intasome is held together by bridging of the inverted repeat sites of IBS1 and IBS6. Mutation of either of these sites is extremely detrimental to the reaction, which suggests that formation of this bridge is a key factor in stabilizing the structure of the integrative intasome.

Though Figure 7A and B seems to suggest that a significant amount of DNA looping must occur between IBS4 and IBS5 in order to form the final structure, this is more a result of viewing the complex in two dimensions from the side. The top view in Figure 7C demonstrates that if the type I and type II sites are

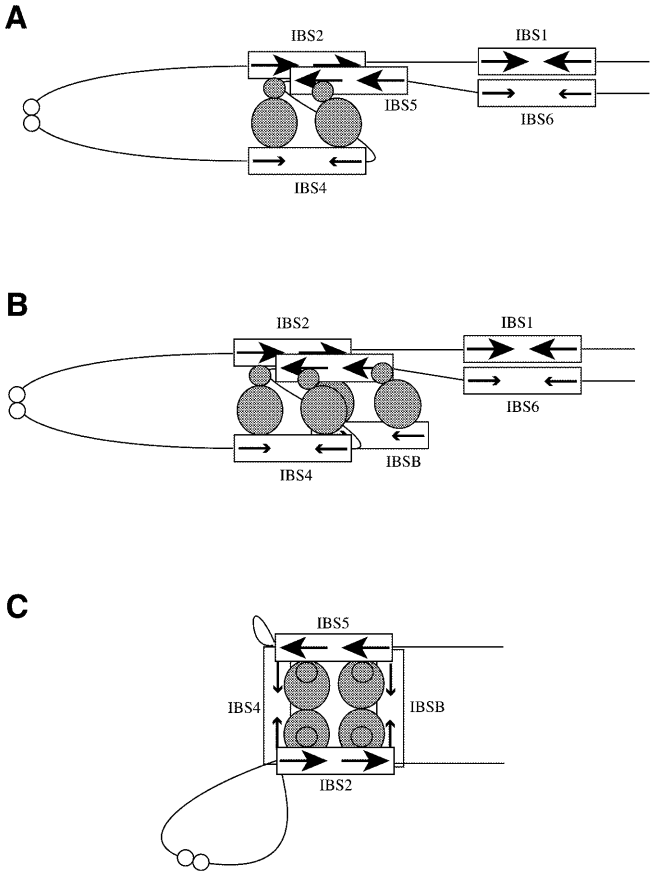

Figure 7. A model for integrative recombination. Rectangular boxes represent integrase-binding sites; type I sites are represented by thin arrowheads, while type II are represented by thick arrowheads. The arrows indicate the direction of the repeats in the binding site. IHF is represented by two small white circles, while integrase molecules are shown in gray. The larger gray oval represents the C-terminal domain of Int which binds the type I sites, while the smaller gray oval represents the N-terminal domain which binds the type II sites. (**A**) A model of the pre-synapsis *attP* intasome. Additional integrase monomers may be bound to the remaining portions of sites 2 and 5; these are omitted for clarity, as are the integrase molecules bridging sites 1 and 6. (**B**) The *attP* intasome synapsed to the *attB* DNA. (**C**) A top-down view of the core of the synapsis shown in (B).

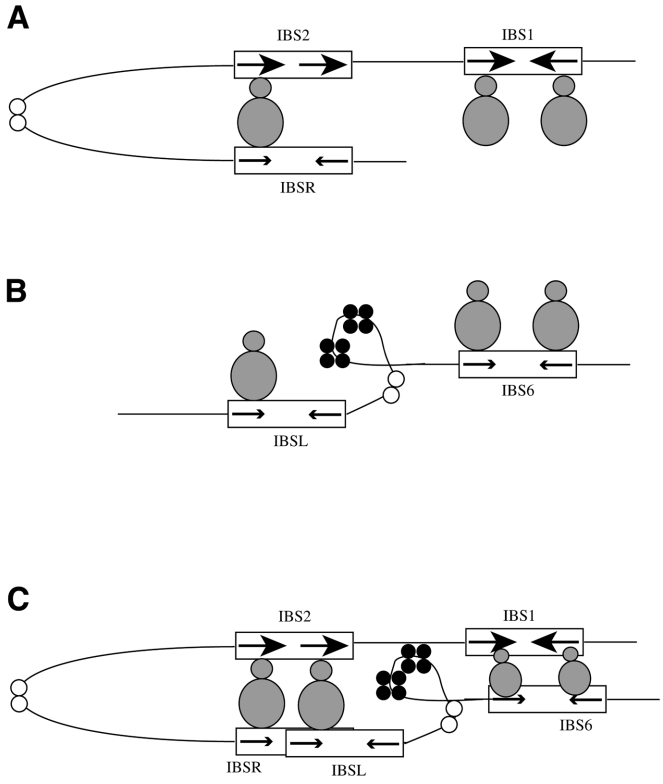

Figure 8. A model for excisive recombination. See Figure 7 for descriptions of the symbols used. Cox is represented by a tetramer of small black circles. (**A**) The proposed structure of the *attR* intasome. (**B**) The proposed structure of the *attL* intasome. (**C**) The synapsed complex of *attL* and *attR* intasomes. As in Figure 7, some Int monomers are omitted for clarity.

not colinear then the looping required is actually quite minimal. It is likely that integrase binding at type II IBS sites is capable of producing some level of DNA bending, as suggested by hypersensitivity to DNase I cleavage observed at these sites (16); it is possible that such bending could provide enough distortion to form this loop. Another explanation may be that the required supercoiling of *attP* permits the IBS5 site to 'slither' into the proper position for interaction with IBS4. The bridging of sites 1 and 6 could then produce enough force to trap the core intasome into the required structure.

A similar model can be proposed for the excisive recombination reaction, as shown in Figure 8. Excisive recombination requires the formation of two separate intasomes, one on *attL* and one on *attR*. Since the binding requirements for *attR* in excision are identical to the binding requirements for the left half of the *attP* site for integration, the *attR* structure in the excision model (Fig. 8A) is identical to that arm of *attP* in integration. Looping of the DNA by IHF bound at H1 allows bridging of the IBS2 and IBSR sites, which are held together by interactions at a single half-site. The *attL* structure, however, looks quite different from its integration counterpart (Fig. 8B). During formation of the *attL* intasome, Cox binding prohibits binding of integrase to IBS5 (15). With IBS5 unoccupied, and no other type II integrase-binding sites present on *attL*, there can be no bridging structure formed on this substrate. Instead, we believe that the binding of Cox and the required IHF bound at H2 must impart the necessary structure to this complex. This may also explain why *in vitro* a supercoiled *attL* substrate produces much higher activity than a supercoiled *attR* substrate; the supercoiling may assist in the formation of the *attL* intasome or may help to stabilize it after the proteins are bound. In addition to the ability of IHF to bend DNA, evidence suggests that Cox can also strongly distort the structure of DNA when it binds (15). Most known helix–turn–helix proteins create significant DNA bends (28,29), and the appearance of strong enhancements of DNase I cleavage at the two ends of the Cox-binding sites predicts that the DNA is dramatically distorted upon binding.

When the two excision substrates come together (Fig. 8C), the integrase bound at IBSL can bridge the appropriately oriented free half-site on IBS2, while the IBS1 and IBS6 sites can form a bridge similar to the one predicted for integration in Figure 7. Interestingly, IBS6 is completely essential for

excision, while in integration its absence still permits very low levels of recombination. This may suggest that the role of IBS6 in helping to synapse the *attR* and *attL* complexes in *trans* is essential, whereas synapsis in *cis* is needed for integration to occur, at low frequency, by either non-specific binding or random collisions. HP1 integrase is the only tyrosine recombinase studied so far which contains additional low affinity (type I or λ core-like) binding sites other than those present at the site of synapsis. The essential role of IBS6 in bridging the *trans* synapsis may be unique to this system because the HP1 *attL* site during excision, lacking integrase at IBS5, contains no bound high affinity integrase binding sites which could otherwise assist in bridging interactions.

After the initial stage of synapsis the remaining unbound half-sites on IBSL and IBSR would likely be quickly filled by free integrase protomers from solution to form the final excision-competent intasomes. This offers an explanation for the higher levels of integrase needed for excision as compared to integration. If the *attR* and *attL* substrates form as pictured in Figure 8, two independent complexes containing mostly 'unbridged' integrase are present, whereas in integration nearly all of the integrase-binding sites are involved in bridging interactions and less integrase protomers would be required. One important feature of the excision reaction is that excision *in vivo* is actually a *cis* event, though we model it *in vitro* as a *trans* synapsis. In the phage, the *attL* and *attR* sites are separated by 32 kb of DNA. It is easy to assume that this large stretch of DNA allows the two attachment sites to act independently as if they were in *trans*. However, it is possible that other factors come into play *in vivo* which might assist or affect synapsis of the *attL* and *attR* sites. Further investigation will be needed in order to clarify this issue.

The models presented here are completely compatible with information gleaned from detailed structural studies of the related tyrosine recombinase (Cre) from bacteriophage P1. The structure of Cre bound to Holliday junction DNA shows the presence of four monomers of Cre binding to a synapsed set of two loxP sites, which contain a total of four individual Cre-binding sites (30,31). This DNA is functionally equivalent to the four type I binding sites in a synapsed IBS4–IBSB complex, which we presume are bound by four monomers of HP1 integrase. In the case of Cre, the monomers are held together by protein–protein interactions which are established by contacts of the C-terminal helix of one monomer with a hydrophobic cleft on the adjacent monomer in a head-to-tail fashion. The tail of each monomer interacts with its neighboring monomer in a circular pattern, with the tail of monomer 1 binding to the cleft of monomer 2, the tail of monomer 2 with the cleft of monomer 3, and so on. In contrast, the crystal structure of the HP1 integrase dimer contains two integrase monomers bound together, with the tail of monomer 1 bound to monomer 2 and the tail of monomer 2 bound to monomer 1 (27). Whether the HP1 integrase is capable of reorganizing its multimerization contacts to form a Cre-like structure or whether two dimers of HP1 integrase have additional protein–protein contacts which allow them to synapse is unclear. A plausible hypothesis is that Cre requires complex circular multimerization because those are the only contacts holding the synapsed DNA together. In HP1, the other domain of the integrase is bound to type II sites and these tight interactions may be the driving force for synapsis. Our model, in which Int 'partners' bound to a single type I site are bound to different type II sites, predicts that this is the case. This kind of multisite interaction may also provide a higher level of control which will ensure that recombination occurs only at the correct time and in the correct direction.

ACKNOWLEDGEMENT

We thank Shani Waninger for supplying HP1 integrase protein, as well as for valuable discussions and advice.

REFERENCES

1. Esposito,D. and Scocca,J.J. (1997) The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.*, **25**, 3605–3614.
2. Nunes-Duby,S.E., Kwon,H.J., Tirumalai,R.S., Ellenberger,T. and Landy,A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.*, **26**, 391–406.
3. Argos,P., Landy,A., Abremski,K., Egan,J.B., Haggard-Ljungquist,E., Hoess,R.H., Kahn,M.L., Kalionis,B., Narayana,S.V., Pierson,L.S. *et al.* (1986) The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.*, **5**, 433–440.
4. Abremski,K. and Hoess,R. (1984) Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.*, **259**, 1509–1514.
5. Andrews,B.J., Proteau,G.A., Beatty,L.G. and Sadowski,P.D. (1985) The FLP recombinase of the 2 micron circle DNA of yeast: interaction with its target sequences. *Cell*, **40**, 795–803.
6. Hayes,F. and Sherratt,D.J. (1997) Recombinase binding specificity at the chromosome dimer resolution site dif of *Escherichia coli*. *J. Mol. Biol.*, **266**, 525–537.
7. Landy,A. (1989) Dynamic, structural and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.*, **58**, 913–949.
8. Yu,A., Bertani,L.E. and Haggard-Ljungquist,E. (1989) Control of prophage integration and excision in bacteriophage P2: nucleotide sequences of the int gene and att sites. *Gene*, **80**, 1–11.
9. Waldman,A.S., Fitzmaurice,W.P. and Scocca,J.J. (1986) Integration of the bacteriophage HP1c1 genome into the *Haemophilus influenzae* Rd chromosome in the lysogenic state. *J. Bacteriol.*, **165**, 297–300.
10. Lee,M.H. and Hatfull,G.F. (1993) Mycobacteriophage L5 integrase-mediated site-specific integration in vitro. *J. Bacteriol.*, **175**, 6836–6841.
11. Hwang,E.S. and Scocca,J.J. (1990) Interaction of integration host factor from *Escherichia coli* with the integration region of the *Haemophilus influenzae* bacteriophage HP1. *J. Bacteriol.*, **172**, 4852–4860.
12. Hauser,M.A. and Scocca,J.J. (1992) Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1. Identification of the points of recombinational strand exchange and the limits of the host attachment site. *J. Biol. Chem.*, **267**, 6859–6864.
13. Reiter,W.D., Palm,P. and Yeats,S. (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res.*, **17**, 1907–1914.
14. Esposito,D. and Scocca,J.J. (1994) Identification of an HP1 phage protein required for site-specific excision. *Mol. Microbiol.*, **13**, 685–695.
15. Esposito,D. and Scocca,J.J. (1997) Purification and characterization of HP1 Cox and definition of its role in controlling the direction of site-specific recombination. *J. Biol. Chem.*, **272**, 8660–8670.
16. Hakimi,J.M. and Scocca,J.J. (1994) Binding sites for bacteriophage HP1 integrase on its DNA substrates. *J. Biol. Chem.*, **269**, 21340–21345.
17. Hakimi,J.M. and Scocca,J.J. (1996) Purification and characterization of the integrase from the *Haemophilus influenzae* bacteriophage HP1; identification of a four-stranded intermediate and the order of strand exchange. *Mol. Microbiol.*, **21**, 147–158.
18. Nash,H.A. (1990) Bending and supercoiling of DNA at the attachment site of bacteriophage lambda. *Trends Biochem. Sci.*, **15**, 222–227.
19. Ito,W., Ishiguro,H. and Kurosawa,Y. (1991) A general method for introducing a series of mutations into cloned DNA using the polymerase chain reaction. *Gene*, **102**, 67–70.
20. Goodrich,J.A., Schwartz,M.L. and McClure,W.R. (1990) Searching for and predicting the activity of sites for DNA binding proteins: compilation

and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). *Nucleic Acids Res.*, **18**, 4993–5000.
21. Staden,R. (1989) Methods for discovering novel motifs in nucleic acid sequences. *Comput. Appl. Biosci.*, **5**, 293–298.
22. Bushman,W., Yin,S., Thio,L.L. and Landy,A. (1984) Determinants of directionality in lambda site-specific recombination. *Cell*, **39**, 699–706.
23. Bushman,W., Thompson,J.F., Vargas,L. and Landy,A. (1985) Control of directionality in lambda site specific recombination. *Science*, **230**, 906–911.
24. Friedman,D.I. (1988) Integration host factor: a protein for all reasons. *Cell*, **55**, 545–554.
25. Freundlich,M., Ramani,N., Mathew,E., Sirko,A. and Tsui,P. (1992) The role of integration host factor in gene expression in *Escherichia coli*. *Mol. Microbiol.*, **6**, 2557–2563.
26. Moitoso de Vargas,L., Pargellis,C.A., Hasan,N.M., Bushman,E.W. and Landy,A. (1988) Autonomous DNA binding domains of lambda integrase recognize two different sequence families. *Cell*, **54**, 923–929.
27. Hickman,A.B., Waninger,S., Scocca,J.J. and Dyda,F. (1997) Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. *Cell*, **89**, 227–237.
28. Dodd,I.B. and Egan,J.B. (1990) Improved detection of helix–turn–helix DNA-binding motifs in protein sequences. *Nucleic Acids Res.*, **18**, 5019–5026.
29. Brennan,R.G. and Matthews,B.W. (1989) The helix-turn-helix DNA binding motif. *J. Biol. Chem.*, **264**, 1903–1906.
30. Guo,F., Gopaul,D.N. and Van Duyne,G.D. (1999) Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Natl Acad. Sci. USA*, **96**, 7143–7148.
31. Guo,F., Gopaul,D.N. and van Duyne,G.D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.