THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 275, No. 29, Issue of July 21, pp. 22255–22267, 2000

Printed in U.S.A.

DNA Recognition, Strand Selectivity, and Cleavage Mode during Integrase Family Site-specific Recombination*

Received for publication, October 8, 1999, and in revised form, February 7, 2000
Published, JBC Papers in Press, March 9, 2000, DOI 10.1074/jbc.M90826199

Gena Tribble‡§, Yong-Tae Ahn‡, Jehee Lee¶, Thomas Dandekar‖, and Makkuni Jayaram‡**

From the ‡Department of Microbiology, University of Texas, Austin, Texas 78712, ¶Faculty of Applied Marine Sciences, Cheju University, Cheju City 690756, South Korea, and ‖European Molecular Biology Laboratory, Postfach 102209, Heidelberg, Germany

We have probed the association of Flp recombinase with its DNA target using protein footprinting assays. The results are consistent with the domain organization of the Flp protein and with the general features of the protein-DNA interactions revealed by the crystal structures of the recombination intermediates formed by Cre, the Flp-related recombinase. The similarity in the organization of the Flp and Cre target sites and in their recognition by the respective recombinases implies that the overall DNA-protein geometry during strand cleavage in the two systems must also be similar. Within the functional recombinase dimer, it is the interaction between two recombinase monomers bound on either side of the strand exchange region (or spacer) that provides the allosteric activation of a single active site. Whereas Cre utilizes the cleavage nucleophile (the active site tyrosine) in *cis*, Flp utilizes it in *trans* (one monomer donating the tyrosine to its partner). By using synthetic Cre and Flp DNA substrates that are geometrically restricted in similar ways, we have mapped the positioning of the active and inactive tyrosine residues during *cis* and *trans* cleavage events. We find that, for a fixed substrate geometry, Flp and Cre cleave the labile phosphodiester bond at the same spacer end, not at opposite ends. Our results provide a model that accommodates local heterogeneities in peptide orientations in the two systems while preserving the global functional architecture of the reaction complex.

The Flp protein encoded by the 2-μm plasmid of *Saccharomyces cerevisiae* and the Cre protein encoded by the bacteriophage P1 are members of the Integrase family (also called the tyrosine family) of "conservative" site-specific recombinases (1, 2). Members of this family carry out a wide range of biological functions but utilize a common biochemical mechanism to accomplish them (3). They mediate the breakage and joining between two double helical DNA partners harboring their respective target sequences (4–6). Four recombinase monomers participate in the reaction to complete two sequential pairs of strand cleavage and joining steps. A Holliday junction is therefore a central intermediate in recombination.

The Integrase family members follow the type IB topoisomerase chemistry to break the DNA strands, forming a 3′-phosphotyrosine linkage and a 5′-hydroxyl group at the site of the strand nick. A number of biochemical observations and recent structural data strongly suggest that the active site tyrosine nucleophile effects strand cleavage with catalytic assistance from a conserved triad consisting of two arginines and one histidine (the RHR triad) (7–15). Two other conserved family residues also contribute to catalysis as follows: a lysine (or arginine) and a histidine (or tryptophan). As demonstrated for the Flp recombinase and as suggested by the Cre-DNA crystal structure (10, 16–18), the catalytically relevant interactions for one strand cleavage event are mediated by two monomers of the recombinase, bound across from each other on either side of the strand exchange region (or spacer). However, the utilization of the cleavage nucleophile is not uniform within the family (1, 2, 19–21). For example, the Flp dimer establishes a shared active site, the RHR triad from one monomer and the tyrosine from the second monomer contributing to its functionality (Fig. 1*A, left*). As a result, Flp cleaves DNA in the *trans* mode, the tyrosine from the monomer bound on the left arm reaching across the spacer to attack the scissile phosphodiester at the right or vice versa (7, 16). By contrast, within the Cre dimer, the triad and the tyrosine contributions to an active site are made by only one monomer (Fig. 1*A, right*; 10). Hence, Cre cleaves DNA in the *cis* mode, the tyrosine of the left or the right bound monomer attacking the adjacent labile phosphodiester at the left or the right spacer end, respectively. Regardless of the cleavage mode, however, neither the Flp nor the Cre dimer can assemble two active sites at the same time. This was revealed by solution experiments with Flp (22) and is suggested by the crystal structure of the cleaved Cre-DNA complex (10). This active site exclusion is consistent with the two-step strand exchange mechanism that both proteins follow.

In the present study, we first probed the changes in protease susceptibility of Flp upon binding to its DNA substrate. The derived Flp footprints are consistent with the features of DNA-protein interactions observed in the Cre-DNA crystal structures (9, 10). Given that the DNA-protein associations are similar for the *trans* cleaving Flp and the *cis* cleaving Cre systems, we wished to decipher which of the two monomers that constitute an active dimer serves as the source of the catalytic tyrosine during the two cleavage modes. The two possible alternatives are diagrammed in Fig. 1*B*. The *cis versus trans* distinction may be represented by I and II or by I and III. In the former case, equivalent recombinase monomers (shown in *green*) provide the tyrosine, yielding cleavage at opposite ends of the spacer. In the latter case, non-equivalent monomers (*green* during *cis* cleavage and *purple* during *trans* cleavage) provide the tyrosine, causing cleavage to occur at the same end of the spacer. In order to maintain the same geometry of the DNA-protein complex for the *cis* and the *trans* cleavage modes (as is the case in Fig. 1*B*), our assays utilized synthetic DNA

* This work was supported in part by grants from the National Institutes of Health and the Robert F. Welch foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Recipient of a National Institutes of Health postdoctoral fellowship.
** To whom correspondence should be addressed. Tel.: 512-471-0966; Fax: 512-471-5546; E-mail: jayaram@icmb.utexas.edu.

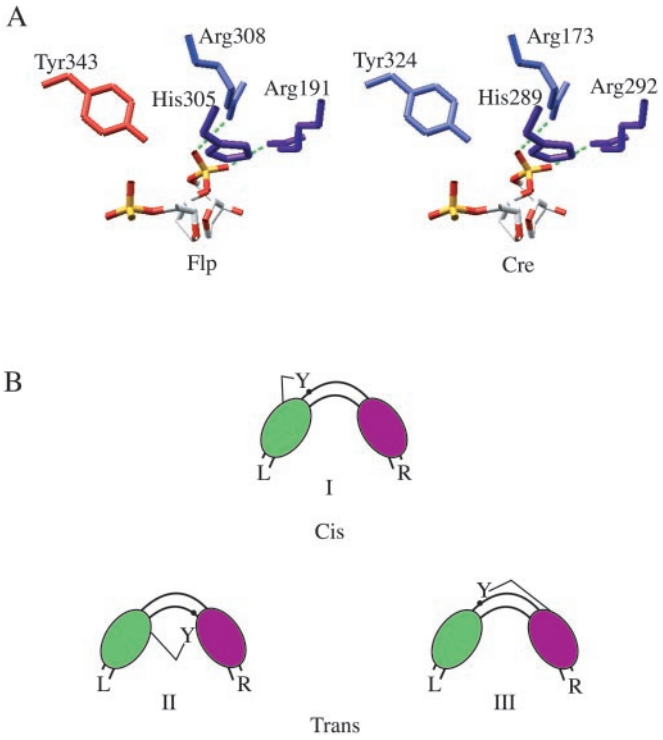

FIG. 1. ***cis* and *trans* cleavages within the Integrase family recombinases.** *A,* the positioning of the catalytic Arg-His-Arg (RHR) triad and the tyrosine nucleophile around the scissile phosphodiester bond is shown. In the case of Flp (*left*), which cleaves DNA in *trans*, the RHR triad (shown in *blue*) is contributed by one protein monomer, and the tyrosine (shown in *red*) is contributed by a second monomer. For Cre (*right*), the triad as well as the tyrosine are derived from a single Cre monomer (all residues shown in *blue*). The arrangement of the phosphate group and the amino acid side chains are redrawn from the Cre-DNA crystal structure (10). *B,* a recombinase dimer that cleaves in *cis* (*I*; Cre, for example) can, in principle, be related to one that cleaves in *trans* (Flp, for example) in one of two ways (*II* or *III*). The tyrosine (from the *green* monomer) that cleaves at the left end of the spacer can be pushed to the right end in its functional orientation as shown in *II*. Or, the tyrosine from the monomer at the right (*purple*) can be moved to the left end as shown in *III*. The cleavage susceptible phosphodiester positions are indicated by the *dots.* The representation of the cleavage active and inactive recombinase monomers in *green* and *purple*, respectively, in the *cis* cleaving dimer (*I*) follows the color codes used by Guo *et al.* (10).

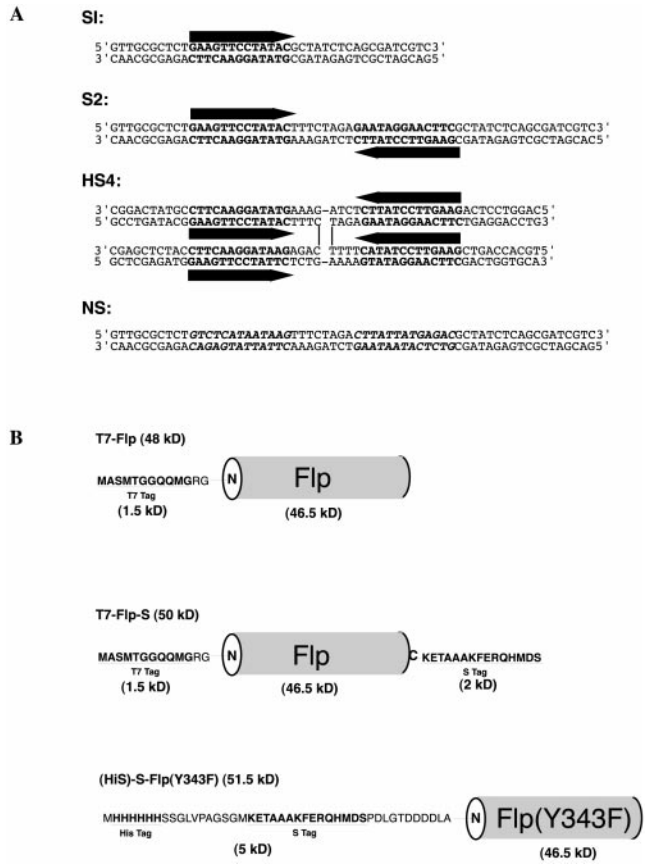

FIG. 2. **Synthetic DNA substrates and tagged variants of Flp used in the Flp footprinting assays.** *A,* the sequences of the individual substrates are shown with the Flp-binding elements written in *bold uppercase letters*, and the *arrows* indicate their relative orientations. The linear substrates *S1* and *S2* and the Holliday junction *HS4* contain one, two, and four Flp-binding elements, respectively. The Holliday junction is represented in an antiparallel configuration. Note that the two Flp-binding elements within a recombination substrate (say, S2) differ in one of the 13-bp positions; the nearest neighbors of the GC pair immediately flank the strand exchange region. It is a TA pair in the left binding element of S2 and an AT pair in the right binding element. Interaction between two Flp monomers, one bound at the left and one bound at the right of the 8-bp spacer, is essential for the assembly of a functional strand cleavage pocket. The non-binding substrate *NS* was obtained by altering 12 positions each within the left and right binding elements. The defunct binding elements are shown in *italics*. *B,* the wild type Flp protein was tagged at either the amino terminus with the T7 tag (*row 1*) or at both the amino and carboxyl termini with the T7 and S tag (*row 2*), respectively. Flp(Y343F) was tagged at the amino terminus with a combined His_6-S tag (*row 3*). The sequences of the tags and the additional mass contributed by them to the 46.5 kDa of native Flp are indicated.

substrates that were constrained in their configuration by strand-specific nucleotide bulges. Our results demonstrate that it is the I/III pair in Fig. 1*B* (and not the I/II pair) that fits the cleavage modes of Cre and Flp. The peptide orientations that would accommodate this mechanism support the intersubunit interactions proposed by Gopaul and Van Duyne (23) for a *trans* cleaving recombinase tetramer.

MATERIALS AND METHODS

Purification of Flp, Flp Derivatives, and Cre—Flp, Flp(Y343F), and their tagged derivatives used in these experiments were >90% pure and were obtained by previously published procedures (22). The Flp derivative containing the His_6 tag was run over a nickel column prior to affinity chromatography on a specific oligonucleotide matrix. The Cre protein (approximately 40% pure) was prepared using a T7 expression vector kindly provided by Ron Hoess (Du Pont Merck Pharmaceutical Co.). Identical results were obtained when the assays were repeated with a highly pure Cre preparation provided by Enoch Baldwin (University of California, Davis).

Synthetic DNA Substrates—All DNA substrates were assembled using synthetic oligonucleotides purchased from Integrated DNA Technologies, Coralville, IA. Hybridization of the appropriate oligonucleotides dissolved in 50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10% glycerol was carried out by first heating them to 65 °C for 5 min in a water bath, followed by slow overnight cooling to room temperature. The substrates present in the footprinting assays are displayed in Fig. 2*A*. The strand cleavage substrates without spacer bulges were as follows: 5′AGCGCTCT**GAAGTTCCTATAC***TTTCTAGA***GAATAGGAACTTC**GGAATAGGAACTTCA3′/3′TCGCGAGA**CTTCAAGGATATG***AAAGATCT***CTTATCCTTGAAG**CCTTATCCTTGAAGT5′ for Flp, and 5′CCCTTAAT**ATAACTTCGTATAA***TGT ATG***CTATACGAAGTTAT**TAGGTCTGAAGAGGGA3′/3′GGGAATTA**TATTGAAGCATATT***ACATAC***GATATGCTTCAATA**ATCCAGACTTCTCCT 5′ for Cre. The positions that contribute to the recombinase binding elements are shown in bold letters. The strand exchange region (spacer) is written in italics. The top strands are shown 5′ to 3′ and bottom strands 3′ to 5′. To create bulges, three additional adenine nucleotides were placed at the required position within the spacer.

Binding of Flp to DNA Substrates—The binding reactions were done in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 10% glycerol. Each incubation mixture contained approximately 5 μg (100 pmol) of Flp or Flp(Y343F) and a DNA substrate equivalent to 500 pmol of the Flp-binding element. At this ratio of roughly 5-fold molar excess of the binding site per Flp monomer, essentially all Flp was bound to the DNA.

Partial Proteolysis—Proteolytic digestions were performed using sequencing grade chymotrypsin (70 ng), trypsin (30 ng), or V8 protease (35 ng) and high purity proteinase K (10 ng) for every 5 μg of Flp. Typical reactions were done in 100-μl volume and utilized 25 μg of Flp per reaction. All enzymes were purchased from Roche Molecular Biochemicals. Reactions were stopped after 20 min incubation at 30 °C by adding 0.2 the volume of 6× SDS sample buffer (0.12 M Tris-HCl, pH 6.8, 100 mM dithiothreitol, 10% SDS, 10% glycerol, 0.001% bromphenol blue). Samples were heated to 90 °C for 5 min and electrophoretically fractionated in 12% SDS-polyacrylamide gels. Gels were either stained with Coomassie Brilliant Blue to visualize protein bands or were prepared for Western blotting.

Western Blotting of Gel-fractionated Proteins—Following electrophoresis, the polypeptides were electrophoretically transferred overnight to PVDF[1] membranes (Gelman Sciences) at 4 °C in transfer buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM glycine, and 20% methanol. The PVDF membranes were then blocked with 1% TBST buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5% Tween 20). After three washes in 1× TBST, the membranes were incubated with an appropriate dilution of the specific monoclonal antibody (1:10,000 for the T7 tag) or with horseradish peroxidase-conjugated S protein (1:5000; supplied by Novagen, Madison, WI) for 1 h at room temperature. Membranes containing the T7-tagged polypeptides were further incubated, following three washes with 1× TBST, with the horseradish peroxidase-conjugated secondary antibody (1:3000 dilution; obtained from Bio-Rad) Positive polypeptide bands were detected using the ECL detection system purchased from Amersham Pharmacia Biotech. Native Flp protein (without the tags) did not cross-react with the T7-specific monoclonal antibody. It also tested negative during probing with the peroxidase-conjugated S protein.

Determination of Molecular Masses—Molecular masses were estimated by comparison to Western blot standards provided by Novagen or rainbow markers supplied by Bio-Rad. The reference curves were prepared by plotting relative mobilities against $\log(M_r)$. Variations in the estimates from experiment to experiment were within 10%.

Peptide Sequencing—Protease-treated samples were fractionated in 15% SDS-polyacrylamide gels and electrophoretically transferred to PVDF membranes. Peptide sequencing was performed either in a gas-phase protein sequencer (LF3000; Beckman Instruments) or in a pulsed liquid protein sequencer (Applied Biosystems) using Edman chemistry (8).

Strand Cleavage Reactions with Flp and Cre—The conditions of the Flp and Cre cleavage reactions were the same and followed previously described protocols (22). Reactions were carried out in 50-μl volume. The samples were fractionated by electrophoresis in 10% denaturing polyacrylamide gels. The cleavage bands were visualized by autoradiography or by phosphorimaging. When reactions were directly analyzed in non-denaturing gels, cleavage bands could not be detected, indicating that strand cleavages within a substrate were exclusively single-stranded. Double-stranded cleavages, if they did occur, should have been apparent in the absence of denaturation.

Quantitation of Strand Cleavage Products—The left and right cleavage bands were quantitated using a Bio-Rad GS-525 Molecular Imaging System PhosphorImager and the image analysis software Molecular Analyst. The cleavage yield was normalized with respect to the substrate amount in the following way. First the intensities of the cleavage bands in a reaction set (labeled as *CL* or *CR* in the Figs. 5–8) were computed after phosphorimaging the fractionated samples for an appropriate length of time. Since the doublet of substrate bands (see Fig. 5, for example) were overexposed and poorly resolved under these conditions, their intensities were derived after reducing the imaging time appropriately. The ratio of the intensity of a cleavage band to that of the corresponding substrate band in a given reaction was designated as the cleavage yield. There was a significant bias in the native Cre substrate toward one cleavage event over the other (CL > CR). In the two experiments used to compose Fig. 8*B*, the ratios of CL to CR in the unbulged control substrate were 3.5 and 4.5, respectively. The "cleavage yields" of CL and CR from the bulged Cre substrates were first estimated as described here for the Flp substrates. The CL yields were then divided by the average bias factor of 4 before plotting the cleavage curve shown in Fig. 8*B*.

Structural Representations and Graphics—The molecular details of the Cre-DNA structures were adapted from Guo *et al.* (10) using molecular visualization softwares RasMol and the Swiss Protein Viewer.

[1] The abbreviations used are: PVDF, polyvinylidene difluoride; bp, base pair; NS, nonspecific; Int, Integrase.

RESULTS

General Strategies, Experimental Rationale, and Interpretations—This section can be divided into two parts. The first set of experiments (Figs. 3 and 4) was aimed at mapping the peptide regions of Flp that are protected from protease digestion when it is bound to its DNA substrate or to the Holliday intermediate of recombination. The protection profile could be overlaid on the Cre-DNA crystal structure to assess the overall degree of fit between the modes of substrate recognition in the two systems. The second set of experiments (Figs. 5–8) addresses how two alternative active site orientations, one for *cis* and the other for *trans* cleavage, can be derived from the same global architecture of DNA-protein interactions.

Protease Footprinting—The DNA substrate in the majority of protease footprinting experiments was the Flp full site (containing two Flp-binding elements, S2; see Fig. 2*A*). In one set of experiments, a half-site (containing one Flp-binding element, S1; Fig. 2*A*) and a Holliday junction substrate (containing four Flp-binding elements, HS4; Fig. 2*A*) were also employed. The nonspecific DNA used as a control in these reactions was the same size as S2 and was obtained by changing 12 positions within the left and right Flp-binding elements (Fig. 2*A*) to the "least favored" bp, as deduced from the analysis of Senecoff *et al*. (24).

The Flp protein derivatives used for the footprinting assays were tagged using the T7 epitope tag or the S peptide tag at either the amino terminus or at the carboxyl terminus. They are schematically represented in Fig. 2*B*. T7-Flp (*row 1*, Fig. 2*B*) and His_6-S-Flp(Y343F) (*row 3,* Fig. 2*B*) harbored the T7 and S tags, respectively, at their amino termini; T7-Flp-S contained both tags, T7 at the amino terminus and S at the carboxyl terminus (*row 2,* Fig. 2*B*).

Wild type Flp or Flp(Y343F) was subjected to partial proteolysis in the presence or absence of DNA under a set of standardized conditions (Figs. 3 and 4). The protease-digested samples were gel-fractionated, transferred to PVDF membranes, and probed with the T7 antibody or the S protein probe. The ladder of bands observed, in the order of increasing molecular mass, provides a measure of the relative distances of the protease-sensitive sites from the position of the tag. In principle, for every band derived from the protein tagged at the amino terminus upon digestion with a given protease, there should be a corresponding band from the protein tagged at the carboxyl terminus such that the two together add up to the full-length protein. The footprinting profiles obtained with wild type Flp or Flp(Y343F) were virtually the same, except that the protein-DNA adduct resulting from Flp cleavage was absent in the Flp(Y343F) reactions.

For each of the proteases used, the digestion patterns of Flp in the absence and presence of the nonspecific DNA fragment (NS) were not identical. This is easily seen by comparing *lanes 4* to *lanes 2* in Fig. 3*C*, for example. The difference could be either due to an effect of DNA on the activity of the proteases or due to some nonspecific Flp-DNA interaction. In deriving the footprint profiles, only those proteolytic bands that could be distinguished between the Flp/S2 and the Flp/NS samples were considered as significant, except in one case that is explained below. Prior to assigning the composite protected peptide regions (see Fig. 4*C*), the molecular masses of proteolytic products were corrected for the size differences in the respective tags they harbored (1.5 kDa for the amino-terminal T7 tag; 2 kDa for the carboxyl-terminal S tag; 5 kDa for the amino-terminal His_6-S tag).

Chymotryptic Footprinting of Flp Bound to DNA Substrates Containing One, Two, and Four Flp-binding DNA Arms—A normal Flp recombination event is initiated by the binding,

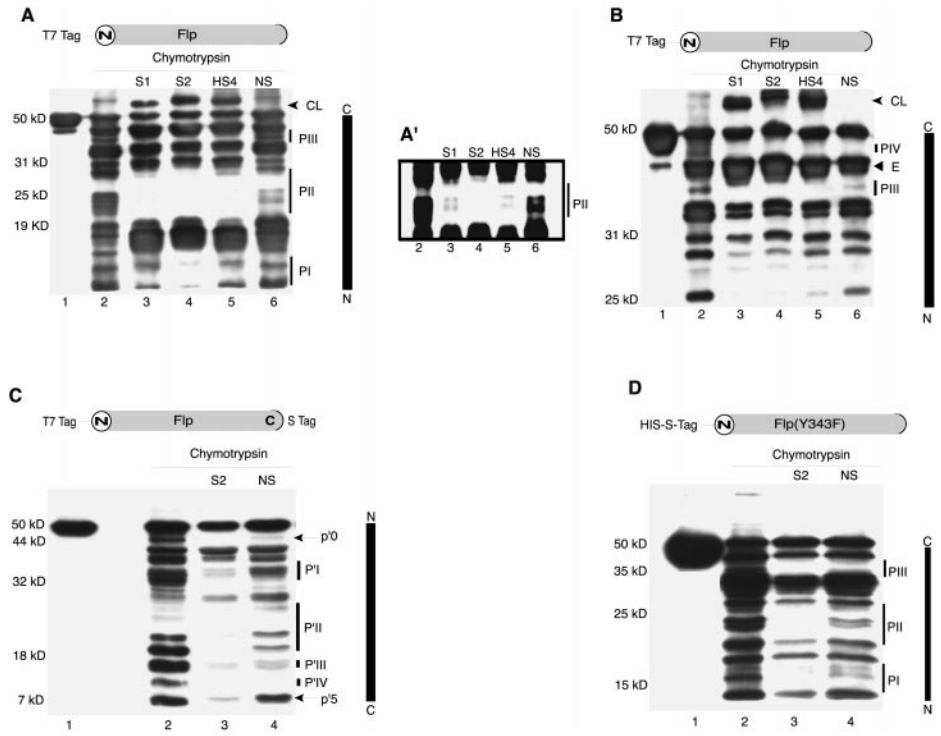

FIG. 3. **Chymotryptic footprint patterns of N-tagged Flp or Flp(Y343F) and C-tagged Flp.** The reactions in *A–D* were carried out in identical manner ("Materials and Methods"). The samples in *A* and *B* were electrophoretically fractionated for different lengths of time to highlight individual features within the footprints. The Western blots were probed with a monoclonal antibody against the T7 peptide in *A* and *B* and with the S probe in *C* and *D*. The N-C ruler placed at the *right* of each panel measures the distance of a protected region from the tagged end. *Lanes 1* and *2* are untreated Flp and Flp treated with chymotrypsin in the absence of DNA. *S1, S2, HS4,* and *NS* refer to the presence of substrates with one, two, four, and no Flp-binding elements (see Fig. 2*A* for details), respectively, in the protease reactions. Regions of protection for the N-tagged proteins are marked *P* with an appropriate suffix. The P′ patches refer to the footprints from the C-tagged protein. The suffixes I–IV match them to the P footprints from the N-tagged protein. The *arrows* with the p′ label (p′0 and p′5 in *C*) indicate protected bands whose counterparts from the N-tagged profile were not apparent. The band migrating above the Flp band (*CL* in *A* and *B*) is due to the covalent attachment of Flp to DNA as a result of strand cleavage. *E* refers to enhanced chymotryptic cleavage in *lanes 3–5. A′* represents a longer exposure of the *luminogram* of *A* to accentuate features of the PII footprint.

separately, of two Flp monomers from solution to each of the two binding elements of the minimal recombination target site. Dimerization of Flp occurs only between two DNA-bound Flp monomers. Furthermore, the strand exchange mechanism gives rise to a Holliday junction as an obligatory intermediate in the reaction. We wished to probe the footprinting patterns of Flp bound to DNA harboring a single Flp-binding element (*S1*, Fig. 2*A*), or to a functional substrate containing two binding elements (*S2*, Fig. 2*A*), or to a Holliday junction containing four Flp binding arms (*HS4*, Fig. 2*A*). The results of these assays using partial proteolysis by chymotrypsin are presented in Fig. 3. The two profiles shown in *A* and *B* correspond to identical sets of chymotryptic reactions fractionated differentially to highlight the region corresponding to 30–14 kDa in *A* and to expand that spanning 50–25 kDa in *B*.

The overall protection patterns were quite similar with the three types of substrates (*lanes 3–5* of Fig. 3, *A* and *B*). A conspicuous region of protection spanning the 22–30-kDa range was seen in all three cases (*PII* in Fig. 3*A*). Within PII, a doublet at approximately 24–25 kDa showed a higher degree of protection in Flp bound to S2 than in Flp bound to S1 or HS4. Although not readily apparent in Fig. 3*A*, this difference can be seen by comparing *lanes 3* and *5* to *4* in Fig. 3*A′* (*A′* shows the relevant region of *A* for a longer exposure of the Western blot luminogram). Similarly, the two bands at 14–16 kDa (*PI*) showed significant protection in the Flp-S2 complex but were essentially unprotected in the Flp-S1 and Flp-HS4 complexes (compare *lane 4* to *lanes 3* and *5* in Fig. 3*A*). Finally, a 34–35-kDa proteolytic fragment was absent from the S1, S2, and HS4 lanes, defining a third chymotryptic footprint, PIII (Fig. 3, *A* and *B*). The possible significance of the band marked *PIV* in Fig. 3*B* is discussed in reference to the results shown in Fig. 4*A* (see below).

An interesting feature of the footprints was the enhancement of a doublet band just above PIII (labeled *E*; approximately 38–39 kDa) in the presence of the Flp substrates (*lanes 3–5;* Fig. 3*B*) but not in the presence of nonspecific DNA (compare *lane 6* to *lane 2* in Fig. 3*B*). The cleavage enhancement was stronger for S1 and S2 compared with HS4. This hypersensitivity to chymotrypsin is suggestive of a potential conformational transition, commensurate with DNA binding, near the carboxyl-terminal portion of Flp adjacent to PIII. This region maps to the amino-terminal side of the active site tyrosine (Tyr-343) of Flp.

The band migrating slower than 50 kDa in *lanes 3–5* of Fig. 3, *A* and *B* (*CL*), was the result of DNA cleavage by Flp and the formation of the covalent Flp-DNA adduct. It should be pointed out that, *a priori*, the presence of CL in the Flp/S1 case (*lanes 3*) was not expected, since it takes two Flp monomers to assemble a single active site for strand cleavage (7, 16). However, once a Flp monomer has bound its cognate sequence in S1 and oriented the scissile phosphodiester, fortuitous transient binding of a second Flp monomer on the same DNA or encounter with a second DNA-bound Flp monomer may result in strand cleavage.

Substrate-mediated Protection against Chymotrypsin in Flp Tagged at the Carboxyl Terminus or Flp(Y343F) Tagged at the Amino Terminus—To verify the authenticity of the footprints obtained with Flp tagged at the amino terminus, a similar chymotryptic digestion was carried out using Flp tagged at the

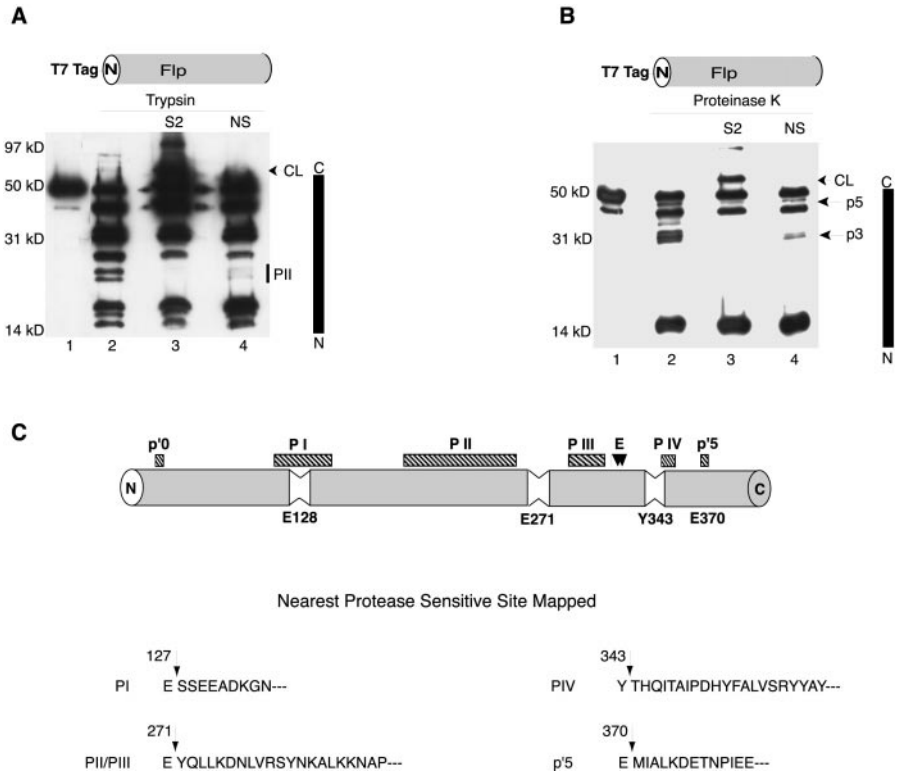

FIG. 4. **Footprinting profiles of N-tagged Flp with trypsin and proteinase K, distribution of protected regions on the Flp protein.** *A* and *B,* the footprints of N-tagged Flp using trypsin and proteinase K are shown. Lane designations and other details are as in Fig. 3. The probe was a monoclonal antibody directed against the T7 peptide. The PII footprint observed with chymotrypsin (see Fig. 3) was also revealed by trypsin (*A*). Two protected bands p3 and p5 were displayed by proteinase K (*B*). The Flp-DNA covalent adduct is labeled *CL*. *C,* the Flp protein is schematically represented as a *cylinder*. The cleavage protection (denoted by *P, p′*) and enhancement (denoted by *E*) data from Fig. 3 are mapped on Flp after compensating for the sizes of the tags used for end labeling (see Fig. 2; also explanation in text). The approximate ranges of PI, PII, and PIII, based on molecular mass estimates, are 114–150, 186–259, and 295–314, respectively. Potential chymotryptic sensitive sites near the PI borders are Tyr-107, Tyr-108, and Trp-155. The amino-terminal PII border matches Phe-182, Phe-186, or Phe-192. Phe-248 or Tyr-271 form the likely carboxyl-terminal border of PII. Tyr-293, Phe-296, and Phe-314 are plausible residues that delimit PIII. The protease-sensitive sites in Flp or Flp(Y343F) adjacent to or overlapping a subset of the protected regions were determined by amino-terminal microsequencing (this study; see Ref. 8).

carboxyl terminus in the presence of S2 or NS. The results are shown in Fig. 3*C*. The sum of the molecular masses of a protected fragment from amino-terminally tagged Flp and its complement from carboxyl-terminally tagged Flp, corrected for the differences in tag sizes, should be approximately 46.5 kDa, the mass of native Flp. Since the amino-terminal T7 tag and the carboxyl-terminal S tag are approximately 1.5 and 2.0 kDa, respectively, the observed sum should be 46.5 + 3.5 = 50 kDa. The regions of protection marked *P′I* (32–36 kDa), *P′II* (20–28 kDa), and *P′III* (14–16 kDa) match the PI (14–16 kDa), PII (22–30 kDa), and PIII (34–35 kDa) protections, respectively, deduced from the data in Fig. 3, *A* and *B*.

The protected band at 7–8 kDa (*p′5*, Fig. 3*C*) is likely the result of chymotryptic cleavage at one or more tyrosine residues at positions 361, 362, and 364. This assignment, although not confirmed by sequencing, is consistent with a previously mapped V8 cleavage site at Glu-370 (8, 28). These are the only aromatic residues within a 20 amino acid segment spanning Glu-370 (from 361 to 380). The approximately 10-kDa band P′IV (*lane 2*, Fig. 3*C*) protected in the S2 and NS lanes (*lanes 3* and *4*, Fig. 3*C*) signifies cleavage at Tyr-343 (or Phe-343 for Flp(Y343F)), as determined by amino-terminal sequencing (see Fig. 4*C*). The high protease sensitivity of the peptide segment housing the Flp catalytic tyrosine was noted in a previous study as well (8). Protease susceptibility in the vicinity of the catalytic tyrosine has been also observed for the lambda Int protein (29). We interpret the P′IV protection to be functionally relevant even though it was observed in both S2- and NS-containing reactions. Since Tyr-343 forms a covalent adduct with the DNA backbone upon strand cleavage, it has to be in close contact with DNA. The band corresponding to P′IV expected from Flp tagged at the amino terminus is approximately 40 kDa (the combined masses of Flp together with the carboxyl- and amino-terminal tags should be 50 kDa). Such a band was observed in the partial chymotryptic digest of N-tagged Flp (*PIV, lane 2*, Fig. 3*B*) and was not detected in the presence of the Flp substrates (*lanes 3–5*, Fig. 3*B*). As with P′IV, PIV was also absent in the reaction containing the nonspecific DNA fragment (*lane 6*, Fig. 3*B*).

The pattern with the C-tagged Flp also included a 46-kDa protected polypeptide (*p′0,* Fig. 3*C*). The equivalent 4-kDa protection from N-tagged Flp would not have been detected under the electrophoretic conditions employed in Fig. 3, *A, B,* and *D*.

Note that the Flp cleavage product (>50 kDa; *CL*) was not seen with the C-tagged Flp (absence of CL in *lane 3* of Fig. 3*C*). This result is consistent with previous observations that even minor alterations of the carboxyl terminus of Flp (in the present instance, the addition of the S tag) result in loss of catalytic activity (30). The crystal structure of the Cre-DNA complex also indicates that the carboxyl-terminal region is responsible for the allosteric activation of one Cre monomer within a dimer to cleavage competence (10).

The chymotryptic footprints obtained with N-tagged Flp(Y343F) in the presence of the S2 DNA (Fig. 3*D*) were qualitatively identical to those yielded by wild type Flp and S2 (compare *lane 3* of Fig. 3*D* with *lanes 4* in Fig. 3, *A* and *B*). The three protected areas (*PI, PII,* and *PIII*) were clearly discernible, and specific bands within a region of protection (for exam-

ple, the doublets harbored by PI and PII) were slightly shifted up in agreement with the longer amino-terminal tag in Flp(Y343F). As expected from the inability of Flp(Y343F) to cleave DNA, no band migrating above 50 kDa was observed in its footprint.

Footprints of Flp Derived by Tryptic and Proteinase K Digestion—The general inferences drawn from the chymotrypsin assays were further tested by probing N-tagged Flp in the presence of S2 with two other proteases (Fig. 4, *A* and *B*). Individually, the information contained within each of the two footprint patterns was sparse relative to the chymotryptic data. Collectively, however, they support and augment the conclusions derived from the latter. For instance, the PII footprint (the protected doublet at approximately 24 kDa) was detected in the trypsin digestion (*lane 3*, Fig. 4*A*). Similarly, the approximately 33-kDa band protected in the proteinase K assays (*p3*, *lane 3*, Fig. 4*B*) maps adjacent to the amino-terminal border of the PIII patch. The proteinase K profile also contained a 47–48-kDa protected band (*p5*, *lane 3*, Fig. 4*B*), placing the corresponding proteolytic site near the extreme carboxyl-terminal region of Flp.

Summary of the Footprints of the Flp-DNA Complexes—The outcomes of the footprint analyses are schematically diagrammed in Fig. 4*C*. The *diagram below* relates the protected regions to the location of the principal protease-sensitive sites mapped within Flp when it was not bound to DNA. Because the cleanest internally cleaved and electrophoretically fractionated polypeptide fragments (suitable for amino-terminal microsequencing) were obtained with V8 protease, the primary landmarks in Fig. 4*C* are provided by Glu residues (Glu-128, Glu-271, and Glu-370). In addition, the protected P′IV cleavage fragment from the chymotrypsin digest of C-tagged Flp was characterized as the product of cleavage at Tyr-343. Hypercleavage of unbound Flp at Asp-134 and Asp-332 that we observed during digestion with endoproteinase Asp-N would be consistent with V8 and chymotryptic cleavages at Glu-128 and Tyr-343, respectively (data not shown). Earlier experiments had revealed that one of the preferred cleavages by trypsin also occurs proximal to Tyr-343, at Arg-340 (8).

The three areas of protection (PII, PIII, and PIV/p′5) spanning the major portion of the carboxyl-terminal domain of Flp would be consistent with the extensive contacts between the corresponding region of the Cre protein and its binding site revealed by the crystal structure (10; see Fig. 9 also). The two protected segments within the Flp amino-terminal domain (p′0 and P1) would also agree with the crystal structure in which the amino- and carboxyl-terminal domains of Cre form a crescent over the DNA target. However, the predicted secondary structure of Flp, as well as the footprinting data, suggests that the amino-terminal domains of Flp and Cre may be less related to each other. The significance of the footprints in DNA recognition by Flp is further addressed under "Discussion."

Overall, the footprinting analysis on Flp conforms to the reasonable expectation that the mode of substrate binding is conserved among members of the Int family. This provides the justification for the functional assays for Flp and Cre described below.

Cleavage Assays in Bulge-containing DNA Substrates—The logic of forcing Flp and Cre substrates into similar configurations for mapping strand cleavage was based on our previous work (22) with Flp substrates containing nucleotide bulges in specific strands. NMR and FRET studies have shown that the bulge acts as a one-way hinge in DNA (25–27). Depending on the strand that it resides in and its position within that strand, a bulge promotes the bending of the flanking DNA arms in one direction (away from the bulge) but not the other. Thus, by placing nucleotide bulges at equivalent positions within DNA sequences of similar size and organization (as is the case for the Flp and Cre target sites), they can be forced to assume similar geometries. The structural features of the different bulged substrates are illustrated in the appropriate figures and described in the corresponding text.

Preferential Cleavage of Bulge-containing Strands by Cre and Flp—The experiments displayed in Fig. 5 test whether *trans* cleavage by Flp and *cis* cleavage by Cre can be related by I and II or by I and III in Fig. 1*B*.

For the Flp system, strand-specific nucleotide bulges between the central two nucleotides of the spacer can constrain the substrate geometry to two nearly exclusive states that yield left cleavage in one case and right cleavage in the other (22). It is always the strand harboring the bulge that becomes cleavage-susceptible. Furthermore, this selective cleavage occurs by the *trans* mechanism, as would be expected from the known behavior of the Flp protein (7, 16, 31). We have now constructed bulged synthetic Cre substrates (analogous to the Flp substrates, containing nucleotide bulges in the middle of the spacer), and we assayed their cleavage preferences in the presence of wild type Cre (Fig. 5*A*). The corresponding results with bulged Flp substrates and Flp are shown for comparison (see Ref. 22).

In the control substrate with no bulges, there was an intrinsic 3–4-fold bias that favors left end cleavage by Cre over the right end (Fig. 5*A, lane 2*). When the bulge was located in the top strand (the strand containing the labile phosphate at the left end), cleavage was directed almost exclusively to the left end, magnifying the natural cleavage bias (Fig. 5*A, lane 4*). When the bulge was present on the bottom strand, cleavage became biased in the opposite direction, toward the right end by a factor of 3 (Fig. 5*A*, *lane 6*; CR:CL ~3:1). By taking into account the inherent cleavage preference of the control substrate, the bottom strand bulge enhances the odds of right cleavage by a factor of 9–12. This strong proclivity for cleavage of the bulge-containing strand was identical to that observed with Flp (Fig. 5*A, lanes 7–12*). The Flp substrates in the reactions represented by *lanes 10* and *12* contained the bulge on the top and bottom strands, respectively.

Thus, for a defined geometry of their substrates (imposed experimentally by the strand bulge), the *cis* cleaving Cre and the *trans* cleaving Flp orient the reactive tyrosine toward the same end of the spacer (as in I and III of Fig. 1*B*). It should be noted that the bulge geometry affects cleavage directly and not indirectly by slowing down the reverse reaction that reseals a broken strand. The cleavage bias observed in substrates containing spacers with centrally placed bulges was identical for Flp as well as the joining incompetent variant Flp(H305L) (22). We assume that this condition holds for the Cre reactions as well.

There is one report in literature (32) that, in contradiction to the crystal structure (10), claims *trans* cleavage by Cre. Our repeated efforts have failed to provide any evidence of Cre cleaving substrates with or without bulges in the *trans* mode.[2] The criterion set by us for *trans* cleavage was a cleavage followed by joining reaction when a recombinase mutant lacking the active site tyrosine was paired with an RHR triad mutant. This is because a cleavage produced adjacent to the tyrosine mutant (as is the case for *trans*) can readily proceed through the joining step. The active site tyrosine is not required for strand joining. In every situation that Flp answered the test (7, 31), Cre failed. Thus, in our assays, Cre and Flp are truly *cis* and *trans* cleaving recombinases.

[2] J. Lee and M. Jayaram, unpublished data.

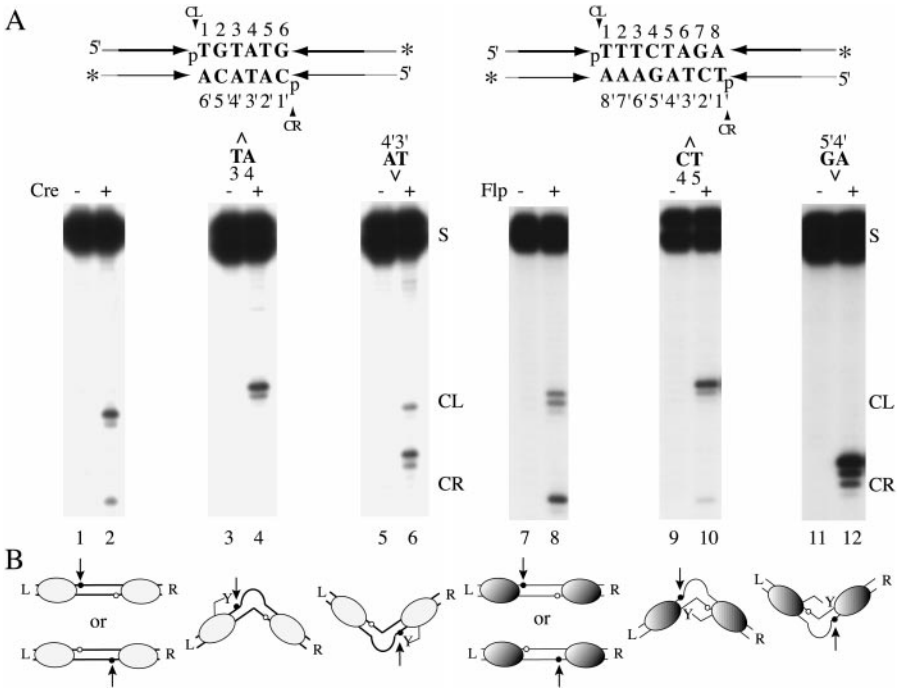

FIG. 5. **Selective strand cleavage by Cre and Flp in substrates containing strand-specific bulges.** *A,* the Cre and Flp substrates (at *left* and *right*, respectively) are schematically shown with their spacer sequences written out, and the scissile phosphodiesters indicated as *p*. The recombinase binding elements are indicated by the *parallel arrows* that also represent their head to head orientation. The *top* and *bottom* strands (arbitrary designations) are distinguished by *thick* and *thin lines,* respectively. The spacer positions are numbered from *left* to *right* on the top strand and from *right* to *left* on the bottom strand (in the 5′ to 3′ direction). The A_3 bulges were introduced between positions 3 and 4 or 3′ and 4′ in the Cre substrate and between 4 and 5 or 4′ and 5′ in the Flp substrate. The reactions were carried out on substrates labeled with ^{32}P at the 3′ ends on both strands (indicated by the *asterisks*). The bands *CL* and *CR* represent cleavage fragments derived from the *left* and *right* ends of the spacer, respectively. The cleavage bands from the bulge containing strands would be 3 nucleotides longer than those from the control substrate. The unreacted substrate bands are denoted by *S*. The presence of more than one cleavage band at a given spacer end is due partly to substrate heterogeneity and partly to the occurrence of an occasional aberrant cleavage in linear substrates. *B,* the experimental results from *A* are schematically diagrammed at the *bottom.* The Cre and Flp substrates are drawn in *thick* and *thin lines,* respectively; the corresponding recombinase proteins are represented by the *unshaded* and *shaded ovals*, respectively. The phosphodiester that is reactive during a cleavage event is shown by the *filled circle.* The unreactive one is shown by the *open circle.*

Scanning the Flp and Cre Spacer with Bulges for Position-dependent Cleavage Bias—If the bulge induces similar DNA geometry in the Flp and Cre substrates (as we have supposed), the position effects on strand cleavage displayed by the spacer bulges should be similar for both the Flp and Cre systems. The experiments that address this issue are summarized in Figs. 6–8.

To simplify interpretations, the following points may be emphasized. Strand cleavage assays for the wild type recombinases normally measure the balance between the cutting (forward) and joining (reverse) reactions. In the case of Flp, strand joining can be suppressed by introducing mismatches in the spacer immediately adjacent to the cleavage site or by using the mutant protein Flp(H305L). The assays shown in Fig. 6*A* utilize a combination of both these conditions to circumvent possible differential effects on the joining reaction of bulges neighboring a scissile phosphodiester *versus* bulges internal to the spacer. Currently, we do not have a mutant of Cre that catalytically mimics Flp(H305L), being normal in cleavage but defective in strand joining. In the Cre assays, done with the wild type protein (Fig. 8), bulges immediately adjacent to the cleavage positions were omitted to avoid "mismatch" effects on the joining reaction. Comparison of the cleavage results with wild type Cre and Flp(H305L) is valid qualitatively. Quantitatively, the Flp (H305L) reactions yielded higher levels of cleavage as expected (compare Fig. 6*A* to Fig. 8*A*).

Bulges within the Flp Spacer—In Fig. 6*A*, the outcomes from a cleavage reaction of Flp(H305L) with substrates containing spacer bulges at various positions are displayed. The lack of cleavage by Flp(H305L) at a given spacer end in the presence of a neighboring bulge (absence of CL in *lane 6* and of CR in *lane 15* of Fig. 6*A*) was expected. The mutant is known to cut poorly at mismatched spacer ends (a bulge would effectively mimic base noncomplementarity; see also Fig. 6*B*). By contrast, as expected from the blockage in strand sealing, wild type Flp yielded enhanced cleavage adjacent to the bulge (CL $\gg$ CR in *lane 5* and CR $\gg$ CL in *lane 14* of Fig. 6*A*). In plotting the cleavage bias in Fig. 7, the values of CL and CR for bulges located at the left and right spacer ends, respectively, were derived from the wild type Flp reaction (*CL* from *lane 5* and *CR* from *lane 14* of Fig. 6*A*). All other values were from the Flp(H305L) reactions. Thus, Fig. 7 represents the cleavage profile, with minimal strand religation, as a function of the bulge location within the spacer. Even without these adjustments to the cleavage outputs by Flp(H305L), the patterns in Fig. 7 (see below) would not be significantly altered qualitatively.

The results displayed in Fig. 6*B* document the differential cleavage competence of Flp(H305L) in response to mismatches at the ends of the spacer or internal to it. As shown in *lanes* 2 and *3* Fig. 6*B*, end mismatches severely compromised the activity of Flp(H305L) but not of wild type Flp. By contrast, base non-complementarity at the spacer center did not affect cleavage by Flp(H305L) (*lane 6* of Fig. 6*B*). Note that the differences in cleavage by wild type Flp in *lanes 2* and *5* of Fig. 6*B* reflect the role of "end homology" in cleavage reversal, namely the slow down in joining due to mismatched bases adjoining the nick (33, 34).

The cleavages at the left and right spacer ends for the different bulge positions (Fig. 6*A*) were quantified as a weighted

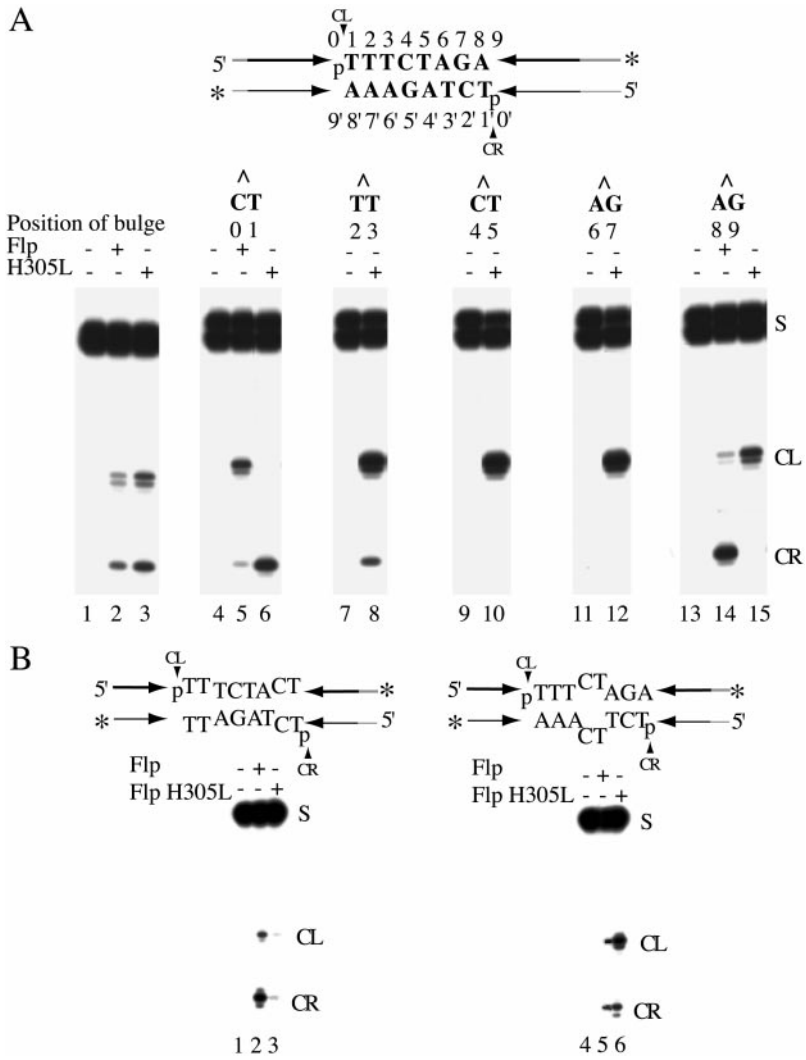

FIG. 6. **Strand cleavage by Flp(H305L) and Flp in substrates with spacer bulges and spacer mismatches.** *A*, in the normal Flp substrate shown at the *top*, the numbers 0 and 0′ (*C*) and 9 and 9′ (*G*) refer to the nucleotide positions immediately flanking the spacer. The bulge positions on the top strand are denoted *above* the respective reactions. The reactions in *lanes 2* and *3* were done with the control substrate harboring no spacer bulges. *B,* in the schematic representation of Flp substrates, the mismatched nucleotides are indicated. The designations of substrate and cleavage bands in *A* and *B* follow the symbols used in Fig. 5.

fraction of the substrate strand converted to CL or CR (see "Materials and Methods") and are graphically represented in Fig. 7. Large cleavage discriminations were observed when the bulge was located at or near the center of the spacer. A strong diminution in the bias, or even its reversal, was observed at the spacer extremities. This pattern holds true for bulges on the other strand as well. The *dashed line* in Fig. 7 represents the cleavage profile obtained for substrates containing bulges on the bottom strand (raw data not shown). The two cleavage curves, drawn as the algebraic sum of the left and right cleavages, demonstrate the inverse relationship in the choice of the phosphodiester target imposed by a matched pair of bulges located at equivalent positions on opposite strands.

The above results suggest that the DNA arms for the centrally placed bulges assume a geometry that most closely resembles their relative disposition within the cleavage-competent complex formed between a Flp dimer and a normal DNA substrate. However, large cleavage preferences were observed even with bulges placed eccentrically within the spacer (between 2 and 3 or 6 and 7; *lanes 8* and *12* of Fig. 6*A*). Presumably, the Flp dimer can manipulate these substrates sufficiently to establish a correctly oriented active site. The cleavage bias for wild type Flp for substrates with bulges at the internal spacer positions (between 2′ and 3′, 4′ and 5′, and 6′ and 7′) was qualitatively the same as that for Flp(H305L) (data not shown), further indicating that cleavage bias was a function of substrate geometry and did not result from indirect effects of bulges on strand ligation.

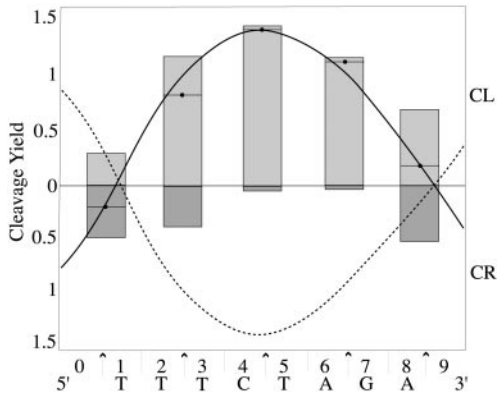

FIG. 7. **The cleavage yields by Flp(H305L) from Fig. 6*A*, *CL* and *CR*, respectively, were calculated (as described under "Materials and Methods") and graphed as a function of the bulge position (indicated by the *vertical arrows* on the abscissa).** The cleavage curve was drawn (*solid line*) by connecting points representing the algebraic sum of CL and CR for each bulge position. The values of CL for the bulge placed between 0 and 1 and that of CR for the bulge placed between 8 and 9 were obtained from the Flp reactions (*lanes 5* and *14* of Fig. 6*A*). For a detailed explanation, see text. This manipulation was necessitated by the severely diminished cleavage capacity of Flp(H305L) at a scissile phosphodiester position when it was flanked immediately by a bulge on the same strand or the opposite strand (*lanes 6* and *15* of Fig. 6*A*; see also text for details). The *dashed line* is the cleavage curve computed for bottom strand bulges from an experiment analogous to that shown in Fig. 6*A*.

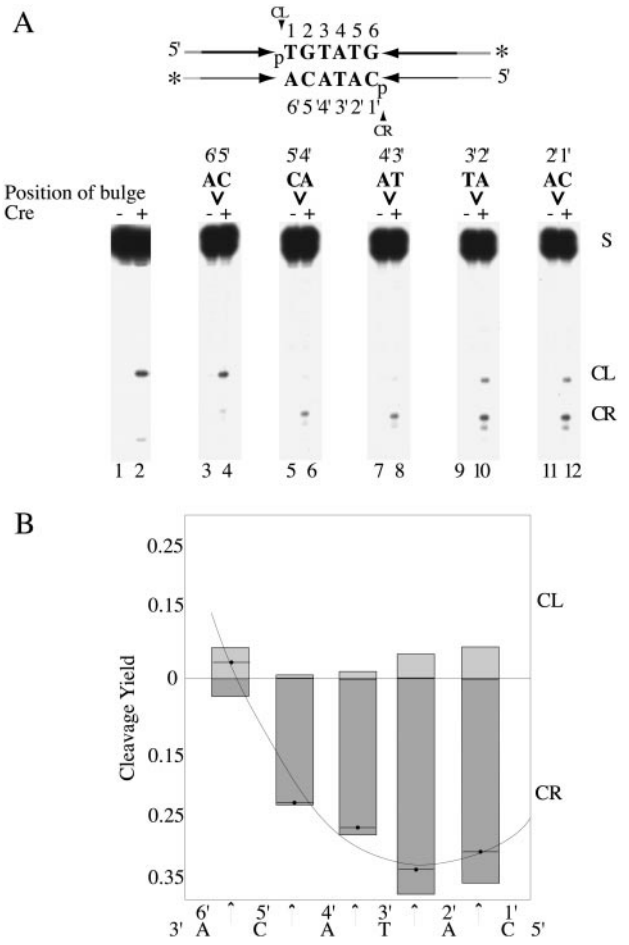

FIG. 8. **Strand cleavage bias as a function of the bulge position within the Cre spacer.** *A*, the bulge locations are indicated *above* the respective reactions. A control assay with the normal substrate (schematically depicted at the *top*) is shown in *lane 2*. *B,* the amounts of CL and CR products from reactions with the bulge-containing substrates were normalized against the corresponding amounts from the control substrate (*lane 2*) as described under "Materials and Methods." The *graph* represents the averaged values from two reaction sets, one of which is shown in *A*.

Bulges within the Cre Spacer—The cleavage patterns obtained with wild type Cre for substrates containing bottom strand spacer bulges are arranged in Fig. 8, *A* and *B*. In the plot showing the cleavage yields as a function of the bulge location (Fig. 8*B*), the CL yields were divided by a correction factor of 4 to compensate for the cleavage bias in the native substrate (*lane 2* of Fig. 8*A*; see also "Materials and Methods"). The dependence and the magnitude of the bias on the bulge position qualitatively paralleled those observed for the Flp substrate (shown in Figs. 6*A* and 7). The bias was strongest at an internal position within the Cre spacer (bulge between 2′ and 3′; Fig. 8, *A* and *B*) and either faded or reversed itself as the bulge was moved to either side (between 1′ and 2′ to the *right*; and between 3′ and 4′, 4′ and 5′, and 5′ and 6′ to the *left*; Fig. 8, *A* and *B*). As already noted, the intrinsic cleavage bias of the control Cre substrate significantly favors the top strand (CL > CR; *lane 2*, Fig. 5*A*). Because of the low cleavage susceptibility of the native bottom strand, spacer bulges introduced on this strand provide a strikingly clear cut demonstration of their role in determining the choice of strands as targets for cleavage.

The similarity in the correlation between cleavage bias and bulge location for the Flp and Cre substrates supports the functional similarity of the bulge-containing substrates in the cleavage reactions by the two recombinases. The *cleavage curves* in Figs. 7 and 8*B* are suggestive of a sinusoidal trend, as would be expected for the phasing of the DNA kinks induced by bulges located along the helical path of a DNA strand (reviewed in Ref. 35). Since the functional spacer length for recombination (the distance between the scissile phosphodiester bonds) is less than 10 bp (6 bp for Cre and 8 bp for Flp), it is not possible to test directly this prediction. In addition, if the primary determinant of the geometry of the protein-DNA complex is the interactions between the two bound recombinase monomers, the "cleavage periodicity" may differ significantly from the helical periodicity of DNA. The effects of the nucleotide bulges are most easily explained by their role in offering steric inducements or posing steric barriers, in a position dependent manner, to the establishment of a particular cleavage geometry as follows: one that targets the left spacer end, or one that targets the right spacer end.

DISCUSSION

Footprinting methods (36, 37), despite their relatively low resolution compared with x-ray crystallography or NMR analysis, provide useful information on structure-function relationships in proteins. Footprint analysis of the DNA-bound Flp has revealed several protected peptide regions and one hypersensitive region. The deduced footprints shed light on peptide segments that contact DNA (or are in proximity to DNA) or undergo conformational changes as a result of DNA binding. Some contributions to the footprints are likely from interactions between Flp monomers that are bound to DNA. Our interpretations assume that the absence or reduction of a proteolytic band is the result of protection from protease and not due to rapid degradation of the primary cleavage product. The latter possibility is extremely unlikely under the limiting protease amounts used in our assays. Furthermore, in a time course assay from 5 to 30 min, no evidence for such degradation was seen.[3] The potential significance of the observed footprints (Fig. 9) may be assessed in the context of the known DNA-binding properties of Flp (28, 38, 39) as follows: a previously proposed model structure for Flp (30), the recently solved structures of the catalytic domains of lambda Int and lambda HP Int proteins (11, 12), the structure of the *Escherichia coli* XerD protein (15), and the co-crystal structures of Cre-DNA complexes (9, 10, 40).

Carboxyl-domain Footprints of Flp—The PII and PIII footprints (Fig. 9*A*) observed in the carboxyl-terminal domain of Flp show overlap with two highly conserved amino acid regions within the Int type yeast recombinases, referred to as Box I and Box II, respectively (41). The hallmark catalytic triad of the Int family (Arg-191, His-305, and Arg-308 in Flp; Refs. 42 and 43) are located within the following boxes: Arg-191 in Box I (spanning positions 185–203; Fig. 9*A*) and His-305 and Arg-308 in Box II (spanning positions 295–313; Fig. 9*A*). The Box I (also called Box A) and Box II homologies extend to the prokaryotic members of the Int family as well (1, 2). However, Box II has been expanded in these protein alignments and includes the invariant active site tyrosine as well. Note that Box II is a composite of the B and C boxes in the classification of Esposito and Scocca (1). The PIV Flp footprint includes Tyr-343, the active site nucleophile for the strand cleavage reaction.

The rather extended series of footprints covering the carboxyl-terminal domain of Flp is consistent with the observation that, in the crystal structure (10) of Cre, DNA contacts essentially span the entire carboxyl-terminal portion of the protein. Furthermore, the predicted secondary structural features of the Flp carboxyl-terminal domain align reasonably well with

[3] G. Tribble, Y.-T. Ahn, J. Lee, T. Dandekar, and M. Jayaram, unpublished data.

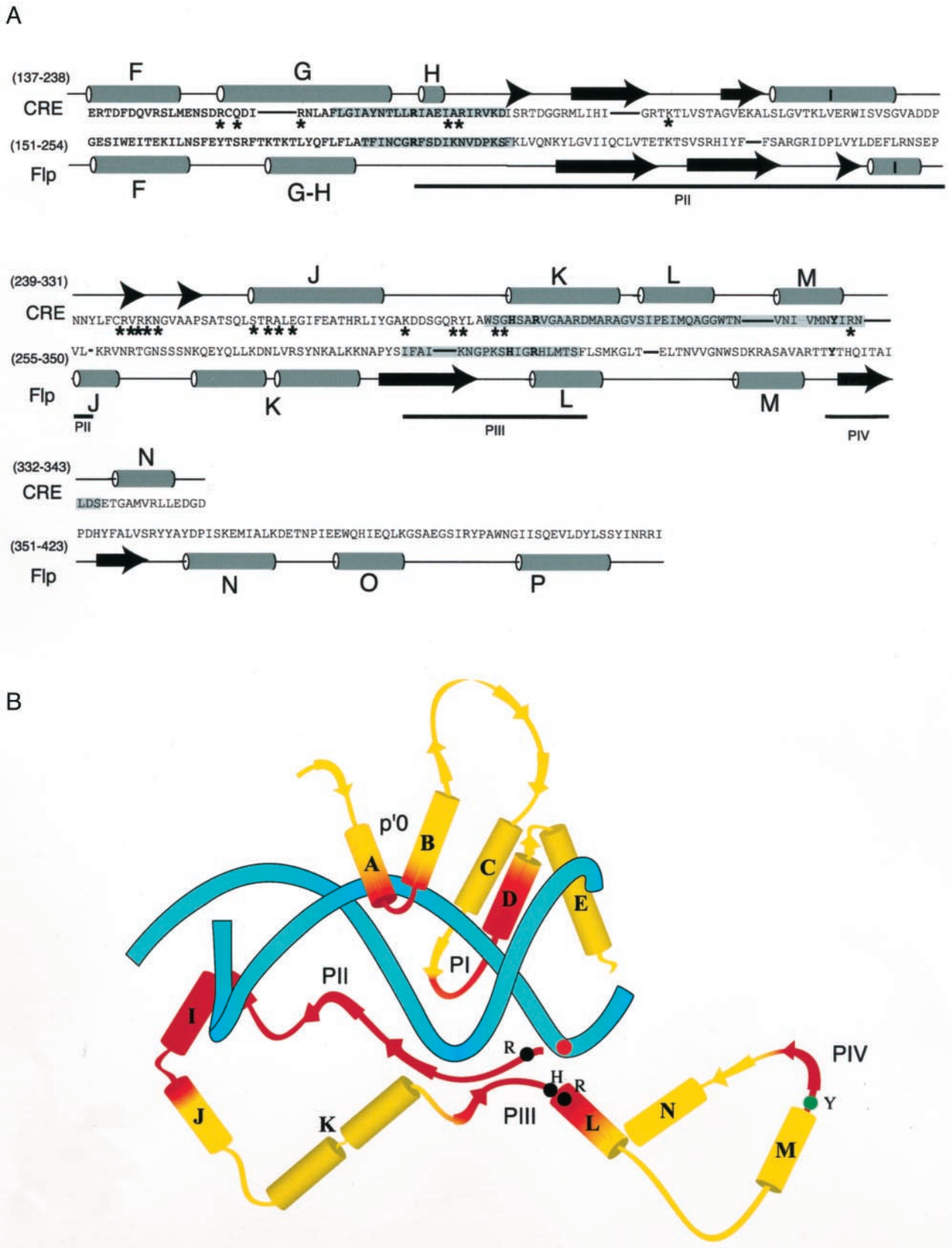

FIG. 9. **Comparison of Flp footprints to DNA contacts observed in the crystal structure of Cre bound to DNA; a model for DNA recognition by Flp.** *A*, the secondary structure representation of the carboxyl-terminal domain of Cre is taken from Guo *et al*. (10) and that of Flp is based on the Protein Secondary Structure Prediction Program provided by the Baylor College of Medicine, Houston. The predicted Flp helices (*bottom panel*) are named arbitrarily to obtain reasonable matches with the Cre nomenclature. The *asterisks* in Cre indicate amino acids that are in close proximity to DNA. The *shaded sequences* in Cre represent two conserved regions in prokaryotic Int type recombinases as follows: Box I (or Box A) and Box II (or Box B plus Box C) (1, 2). In Flp, the regions of protease protection are *underlined* and labeled as in Fig. 4*C*. The conserved

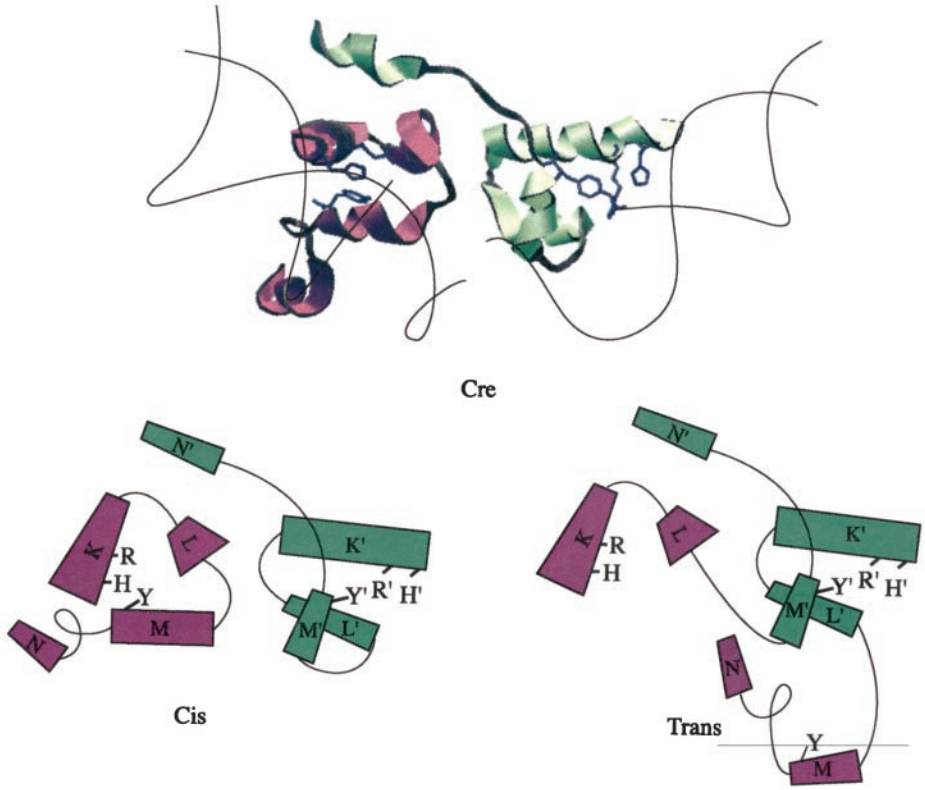

FIG. 10. **Peptide connectivities that determine *cis* or *trans* cleavage by a recombinase dimer.** *Top,* perspective of the cleaved Cre-DNA complex (10) as viewed from a vantage point on the carboxyl-terminal side of the protein is shown. The DNA bend corresponds to a bottom strand bulge near the center of the spacer. Cleavage has occurred at the *right-hand side*, executed in the *cis* mode by the Cre monomer in *green. Bottom,* the same perspective as in the *top panel* is redrawn at the *left* to represent the helices K–N and their linkages. At the *right*, the change in connectivity between the *purple* and *green* monomers to mediate *trans* cleavage is shown. Effectively, the *purple* L helix is linked to the *green* M′–N′ helices. Conversely, the *green* L′ helix is linked to the purple M–N helices. The new location of the M–N helices approximates the situation for the recombinase tetramer seen in DNA-protein cocrystal structures (23).

the corresponding domain of Cre (Fig. 9*A*). The similarity breaks down toward the extreme carboxyl terminus, where Flp has a relatively long peptide tail compared with Cre.

The PI and PII protections and the cleavage enhancement adjacent to PIII (see Fig. 4*C*) showed subtle differences depending on the type of DNA substrate to which Flp was bound. The protections were more prominent in the S2 substrate (full-site) compared with S1 (half-site) or HS4 (Holliday junction). The cleavage enhancement, on the other hand, was less conspicuous in HS4 compared with S1 or S2. These variations in the footprint profiles likely reflect differences in DNA-protein or protein-protein interactions and conformational dynamics within the recombination complexes at individual stages of the reaction.

Amino-terminal Footprints of Flp—The amino-terminal domain of Flp contains two footprints, p′0 and PI, located near the opposite ends of this domain (Fig. 4*C* and Fig. 9*B*). The PI footprint would correspond to the DNA contact points observed in the D and E helices and the interhelical segment between E and F helices of the Cre protein (10). In the Cre structure, the amino-terminal helices B and D are seen to contact the DNA major groove. The Flp results do not reveal an obvious footprint that can be interpreted as equivalent to the B helix-DNA interaction of Cre. The p′0 protection in Flp (proximal to the carboxyl-terminal side of helix A; Fig. 9*B*) may represent the corresponding, if somewhat altered, DNA contact by Flp. The secondary structural correspondence between the amino-terminal domains of Flp and Cre are quite poor and are therefore not aligned in Fig. 9. We suspect that DNA recognition by the amino-terminal domains of Flp and Cre may be less well conserved than that of their carboxyl-terminal domains.

Substrate Recognition by the Int Family Recombinases, DNA Binding by C-shaped Protein Clamps—The protease footprinting results obtained for Flp, together with the available structural and biochemical information for the Int family members, suggest a reasonable global conformation for the Flp-DNA complex. In Fig. 9*B*, the Flp protein is shown to form a C-clamp around the DNA in a manner analogous to Cre (10). A similar mode of DNA binding is accommodated by the Flp model structure of Saxena *et al*. (30), provided the amino-terminal domain is swung into position to contact DNA. The picture of DNA being cradled within a protein crescent agrees well with the bidomain architecture of Flp deduced from partial proteolysis of the free protein or the protein bound to DNA (8, 28, 38). This mode of DNA interaction is also supported by UV cross-linking studies demonstrating that the amino-terminal domain of Flp

boxes of the yeast family recombinases, Box I and Box II (41), are shown by the *shaded areas.* Note that the demarcation of the yeast family Box II places the active site tyrosine (shown in *bold letter*) outside its boundary, whereas the prokaryotic Box II includes this residue. The RHR catalytic triads (within Box I and Box II) are also shown in *bold letters.* The observed protections in Flp are in fair agreement with the expectations based on the protein-DNA and protein-protein contacts revealed by the Cre structure (10). The Flp to Cre alignment is similar to those published previously (2, 12). *B,* the global three-dimensional topology of the Flp protein is obtained by fitting the secondary structure of Flp (predicted by Protein Secondary Structure Prediction Program) to a proposed model structure for Flp (30). To do this, the amino-terminal domain of the protein is moved with respect to the carboxyl-terminal domain using the presumably flexible interdomainal peptide loop so as to bring the former into close proximity with DNA. In the resulting protein-DNA model, the DNA is held between the two domains in a manner that is analogous to the mode of substrate association by the Cre protein (10). The peptide regions protected in the protease footprinting assays (*p′0, PI–IV*) are indicated in *red.* Note that the A/B helices together with the D helix could form a scissors-like grip on the DNA primarily through the major groove. The clustering of the RHR triad residues around the scissile phosphodiester bond (shown in *red*) and the positioning of the active site tyrosine (shown in *green*) away from it are consistent with the mechanism of phosphate orientation and the *trans* mode of nucleophile donation during strand cleavage by Flp. The protection of this tyrosine and the adjacent peptide segment must result from proximity to DNA at the other end of the spacer.

contacts the spacer-proximal 4 bp of the binding element, whereas the carboxyl-terminal domain contacts the spacer-distal 9 bp (39). The general view of the Flp-DNA complex presented here will be refined by crystal structures of Flp-DNA complexes that are currently being solved.[4] However, the inferred similarity between Flp and Cre in their DNA association validates the rationale for inquiring how two similar protein-DNA assemblies yield different modes of strand cleavage.

Directed DNA Bends as the Agent of Strand Cleavage Selectivity—The results from the present study, together with those of Lee *et al.* (22), have shown that nucleotide bulges located centrally within the spacer induce the same directionality of cleavage bias in synthetic Cre and Flp substrates. It is the strand that contains the bulge that is the preferred target for cleavage in both cases. Based on the geometric effects of nucleotide bulges in double-stranded nucleic acids (25–27, 44, 45), the biased cleavages can be accounted for by two distinct DNA bends, equivalent in magnitude but opposite in orientation. In the absence of the bulge, a substrate bound by a Flp or Cre dimer has a finite probability of establishing either of these two DNA bends. The bulge, depending on the strand that contains it and its relative location within that strand, can enrich one type of DNA bend by selectively blocking the other. The bulge-mediated cleavage exclusion on one of the two strands in Flp substrates is consistent with the overall features of the Flp-induced DNA bend described extensively by Sadowski and colleagues (46–48). Furthermore, the most recently solved structure of the pre-cleavage synaptic complex formed by Cre (40) also suggests that it is the DNA bend direction that specifies the strands to be cleaved and exchanged in the initial step of recombination.

The positional effects of the bulge on the magnitude and direction of the bias indicate similar trends for both Flp and Cre. The implication is that both recombinases induce identical or similar DNA configurations when bound to the two substrate arms flanking the spacer. By extrapolation, a similar geometry also exists in the recombination complexes, in which all four arms of the DNA partners are protein-bound.

Global Change in Strand Cleavage Mode by Local Change in Peptide Geometry—The Cre crystal structure reveals that it is the extreme carboxyl-terminal region of Cre consisting of the K, L, M, and N helices and the tail end of the protein that partake in the interactions responsible for the allosteric activation of one monomer within a dimer to cleavage competence (Fig. 10, *top*; see Refs. 9 and 10). The perspective in Fig. 10 brings into view the relevant portions of the carboxyl-terminal domain of Cre. The outline of the DNA backbone is included to indicate the direction of the bend (corresponding to bottom strand bulge and cleavage at the right spacer end by the *green* monomer). A two-dimensional projection of the K-N helices and their connectivities is shown in the *bottom left panel* of Fig. 10. Gopaul and Van Duyne (23) have pointed out that, for a given recombinase subunit, an extensible tether between the L and M helices could, in principle, confer the *trans* configuration on M (which harbors the catalytic tyrosine), whereas the rest of the carboxyl-terminal peptide folds back to retain the *cis* configuration of N. The strand cleavage patterns imposed by the spacer bulges in Cre and Flp reactions are consistent with this model. In the representation of the Flp dimer in Fig. 10 (*bottom right*), the *purple* L helix is connected to M′–N′, defining the peptide connectivity within a monomer of the *trans* cleaving recombinase. The M–N helices (now linked to the *green* L′ helix) are swung away to designate their plausible location within a recombination complex containing a recombinase tetramer (23). Although the cyclic contact between adjacent recombinase monomers is counterclockwise for the Cre tetramer (Fig. 10, *bottom left*), it becomes clockwise for Flp (Fig. 10, *bottom right*) in order to retain identity in strand cleavage. It seems plausible that a *trans* cleaving recombinase could have evolved differently, by maintaining the counterclockwise interaction seen in Cre, but adopting strand selectivity in cleavage opposite to that of Cre.

Evolution of Divergent Protein Folds for a Conserved Function—The evolution of two distinct DNA cleavage modes within the Int family recombinases by localized amino acid changes among its members does not contradict the rules of protein structure and folding. A relevant example is the conversion of staphylococcal nuclease from the native monomer form into a stable dimer by a short peptide deletion in a surface loop (49). The resultant stripping of the carboxyl-terminal α helix from its normal position (*cis*) and its incorporation into an equivalent position on the adjoining monomer (*trans*) is analogous to the switching of peptide connectivities diagrammed in Fig. 10. More recently, it has been shown that interchanging two amino acid neighbors in the Arc repressor transforms the local peptide fold from a β strand into a right-handed helix (50). This structural alteration has little effect on the global protein stability or folding cooperativity.

Acknowledgments—We acknowledge the gift of the Cre expression vector by Ron Hoess and a sample of pure Cre preparation by Enoch Baldwin. We thank Klaus Linse and associates at the Core Facility of the Institute for Cell and Molecular Biology (University of Texas, Austin) for performing protein sequencing. We are grateful to Shailja Pathania for helping with the preparation of figures and to all our colleagues for comments and criticism. Cecil Harkey provided excellent technical assistance.

[4] P. Rice and M. M. Cox, personal communication.

REFERENCES

1. Esposito, D., and Scocca, J. J. (1997) *Nucleic Acids Res.* **25,** 3605–3614
2. Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., and Landy, A. (1998) *Nucleic Acids Res.* **26,** 391–406
3. Nash, H. A. (1996) in *Escherichia coli and Salmonella, Cellular and Molecular Biology* (Neidhart, F. C., ed) Vol. 2, pp. 2363–2376, American Society for Microbiology, Washington, D. C.
4. Sadowski, P. D. (1993) *FASEB J.* **7,** 760–767
5. Landy, A. (1993) *Curr. Opin. Genet. & Dev.* **3,** 699–707
6. Lilley, D. M. (1997) *Chem. Biol.* **4,** 717–720
7. Chen, J.-W., Lee, J., and Jayaram, M. (1992) *Cell* **69,** 647–658
8. Evans, B. R., Chen, J. W., Parsons, R. L., Bauer, T. K., Teplow, D. B., and Jayaram, M. (1990) *J. Biol. Chem.* **265,** 18504–18510
9. Gopaul, D. N., Guo, F., and Van Duyne, G. D. (1998) *EMBO J.* **17,** 4175–4187
10. Guo, F., Gopaul, D. N., and Van Duyne, G. D. (1997) *Nature* **389,** 40–46
11. Hickman, A. B., Waninger, S., Scocca, J. J., and Dyda, F. (1997) *Cell* **89,** 227–237
12. Kwon, H. J., Tirumalai, R., Landy, A., and Ellenberger, T. (1997) *Science* **276,** 126–131
13. Pargellis, C. A., Nunes-Duby, S. E., de Vargas, L. M., and Landy, A. (1988) *J. Biol. Chem.* **263,** 7678–7685
14. Parsons, R. L., Prasad, P. V., Harshey, R. M., and Jayaram, M. (1988) *Mol. Cell. Biol.* **8,** 3303–3310
15. Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J., and Wigley, D. B. (1997) *EMBO J.* **16,** 5178–5187
16. Lee, J., and Jayaram, M. (1995) *J. Biol. Chem.* **270,** 23203–23211
17. Lee, J., Jayaram, M., and Grainge, I. (1999) *EMBO J.* **18,** 784–791
18. Voziyanov, Y., Lee, J., Whang, I., Lee, J., and Jayaram, M. (1996) *J. Mol. Biol.* **256,** 720–735
19. Jayaram, M. (1997) *Science* **276,** 49–51
20. Jayaram, M., and Lee, J. (1995) *Trends Genet.* **11,** 432–433
21. Stark, W. M., and Boocock, M. R. (1995) *Trends Genet.* **11,** 121–123
22. Lee, J., Tonozuka, T., and Jayaram, M. (1997) *Genes Dev.* **11,** 3061–3071
23. Gopaul, D. N., and Van Duyne, G. D. (1999) *Curr. Opin. Struct. Biol.* **9,** 14–20
24. Senecoff, J. F., Rossmeissl, P. J., and Cox, M. M. (1988) *J. Mol. Biol.* **201,** 405–421
25. Gohlke, C., Murchie, A. I., Lilley, D. M., and Clegg, R. M. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11660–11664
26. Rosen, M. A., Shapiro, L., and Patel, D. J. (1992) *Biochemistry* **31,** 4015–4026
27. Rosen, M. A., Live, D., and Patel, D. J. (1992) *Biochemistry* **31,** 4004–4014
28. Chen, J. W., Evans, B. R., Yang, S. H., Teplow, D. B., and Jayaram, M. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 5944–5948
29. Tirumalai, R. S., Healey, E., and Landy, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 6104–6109
30. Saxena, P., Whang, I., Voziyanov, Y., Harkey, C., Argos, P., Jayaram, M., and Dandekar, T. (1997) *Biochim. Biophys. Acta* **1340,** 187–204
31. Lee, J., Whang, I., Lee, J., and Jayaram, M. (1994) *EMBO J.* **13,** 5346–5354
32. Shaikh, A. C., and Sadowski, P. D. (1997) *J. Biol. Chem.* **272,** 5695–5702

33. Lee, J., and Jayaram, M. (1995) *J. Biol. Chem.* **270,** 4042–4052
34. Zhu, X. D., Pan, G., Luetke, K., and Sadowski, P. D. (1995) *J. Biol. Chem.* **270,** 11646–11653
35. Lilley, D. M. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 7140–7142
36. Baichoo, N., and Heyduk, T. (1997) *Biochemistry* **36,** 10830–10836
37. Heyduk, T., Heyduk, E., Severinov, K., Tang, H., and Ebright, R. H. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10162–10166
38. Pan, H., Clary, D., and Sadowski, P. D. (1991) *J. Biol. Chem.* **266,** 11347–11354
39. Panigrahi, G. B., and Sadowski, P. D. (1994) *J. Biol. Chem.* **269,** 10940–10945
40. Guo, F., Gopaul, D. N., and Van Duyne, G. D. (1999) *Proc. Natl. Acad Sci. U. S. A.* **96,** 7143–7148
41. Utatsu, I., Sakamoto, S., Imura, T., and Toh-e, A. (1987) *J. Bacteriol.* **169,** 5537–5545
42. Abremski, K. E., and Hoess, R. H. (1992) *Protein Eng.* **5,** 87–91
43. Argos, W., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H., Kahn, M. L., Kalionis, W., Narayama, S. V. L., Pierson, L. S., III, Sternberg, N., and Leong, J. M. (1986) *EMBO J.* **5,** 433–440
44. Bhattacharyya, A., and Lilley, D. M. (1989) *Nucleic Acids Res.* **17,** 6821–6840
45. Bhattacharyya, A., Murchie, A. I., and Lilley, D. M. (1990) *Nature* **343,** 484–487
46. Luetke, K. H., and Sadowski, P. D. (1995) *J. Mol. Biol.* **251,** 493–506
47. Luetke, K. H., and Sadowski, P. D. (1998) *Mol. Microbiol.* **29,** 199–208
48. Luetke, K. H., and Sadowski, P. D. (1998) *Nucleic Acids Res.* **26,** 1401–1407
49. Green, S. M., Gittis, A. G., Meeker, A. K., and Lattman, E. E. (1995) *Nat. Struct. Biol.* **2,** 746–751
50. Cordes, M. H. J., Walh, N. P., McKnight, C. J., and Sauer, R. T. (1999) *Science* **284,** 325–327