A STRUCTURAL VIEW OF Cre-loxP SITE-SPECIFIC RECOMBINATION

Gregory D. Van Duyne
Department of Biochemistry and Biophysics, Howard Hughes Medical Institute,
University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104;
e-mail: vanduyne@mail.med.upenn.edu

Key Words  recombinase, integrase, synapse, Holliday junction, strand exchange

Abstract  Structural models of site-specific recombinases from the lambda integrase family of enzymes have in the last four years provided an important new perspective on the three-dimensional nature of the recombination pathway. Members of this family, which include the bacteriophage P1 Cre recombinase, bacteriophage lambda integrase, the yeast Flp recombinase, and the bacterial XerCD recombinases, exchange strands between DNA substrates in a stepwise process. One pair of strands is exchanged to form a Holliday junction intermediate, and the second pair of strands is exchanged during resolution of the junction to products. Crystal structures of reaction intermediates in the Cre-loxP site-specific recombination system, together with recent biochemical studies in the field, support a "strand swapping" model for recombination that does not require branch migration of the Holliday junction intermediate in order to test homology between recombining sites.

INTRODUCTION

Site-specific recombinases from the lambda integrase family of enzymes catalyze DNA rearrangements that are critical for a variety of important biological functions. Perhaps the best known example is the integration and excision of the bacteriophage λ genome into and out of the Escherichia coli host chromosome (9). Other functions include the resolution of multimeric plasmids and chromosomes to monomers in order to ensure faithful segregation upon cell division, the amplification of yeast 2µ circle copy number, and the regulation of gene expression (reviewed in 24, 44, 57, 59). The λ-integrase family, also referred to as the "tyrosine recombinases" (58), includes over 100 members identified on the basis of sequence similarity (47). The most well-studied examples include, in addition to the integrase protein from bacteriophage λ (34), the bacterial XerC and XerD recombinases (57), Cre recombinase from bacteriophage P1 (1), and the Flp recombinase from Saccharomyces cerevisiae (53).
The tyrosine recombinases carry out the site-specific recombination reaction in a stepwise manner. One pair of DNA strands is first exchanged to form a Holliday junction intermediate, and then the Holliday junction is converted to recombinant products by exchange of the second pair of DNA strands. In contrast to the many years of biochemical and molecular genetic studies on the various integrase family members by several laboratories, structural models for the tyrosine recombinases have only recently become available, beginning with the report of the λ-integrase catalytic domain crystal structure in early 1997 (33). Since then, structures of the HP1 integrase catalytic domain (27), the XerD protein (62), and three Cre recombinase/DNA complexes (17, 22, 23) have been described. Most recently, the structure of an Flp recombinase/DNA complex has been determined (11a). The resulting structural models for the tyrosine recombinase family have provided both a view of the enzyme active sites and a three-dimensional perspective on the reaction pathway (for minireviews of the structural work, see references 20, 36, 40, 58, 68). This review focuses on the three-dimensional nature of the Cre-loxP site-specific recombination pathway, where models of the recombinase-DNA complexes for three intermediates in the reaction are available.

THE Cre-loxP SYSTEM

Cre recombinase is a 38-kDa protein encoded by bacteriophage P1. Its roles in the P1 life cycle are thought to include cyclization of the linear genome and resolution of dimeric chromosomes formed following DNA replication as a result of homologous recombination (61). The DNA sequences where recombinase binding and strand exchange take place are named loxP in the Cre recombinase system. These sites are composed of two recombinase binding elements (RBEs) arranged as nearly perfect inverted repeats surrounding a central strand exchange or crossover region (Figure 1). Two recombinase subunits bind to each core site (one to each

![Figure 1](The 34-bp loxP site is composed of two 14 bp recombinase binding elements (RBEs) arranged as inverted repeats around a central 6-bp crossover region (28). Cleavage of the sites occurs at the borders between the crossover region (in boldface) and the RBEs. Two Cre recombinase subunits bind cooperatively to the loxP site (one to each RBE) with no direct contact between the recombinases and bases in the crossover region.)
Cre-loxP RECOMBINATION

RBE) with a high level of cooperativity (7, 51). The phosphoryl transfer strand exchange chemistry between substrates occurs within the 6-bp crossover region. The crossover sequence is asymmetric and therefore provides directionality to the site. For the simplest systems, exemplified by Cre recombinase, this core recombination site is sufficient for the complete recombination reaction (3, 53). For more complex systems, such as λ-integrase and the XerC/D recombinases acting on plasmid substrates, the sites required for recombination contain additional sequences recognized by accessory proteins and auxiliary DNA-binding domains. The simplicity of the Cre recombinase system has led to its widespread use as a tool in the manipulation of DNA molecules both in vivo and in vitro (54).

The tyrosine recombinases use a topoisomerase I-like mechanism to cleave and religate DNA strands during recombination. An historical model for the recombination pathway is shown in Figure 2, which illustrates the stepwise nature of the reaction. This view of the reaction involves only the core recombination sites and the core-binding components of the recombinase enzymes. For simple systems (such as Cre) recombining linear substrates, this model represents all of the interacting components. For more complex systems, there are additional factors that play architectural and/or energetic roles that could influence the various steps in the recombination pathway. For reviews of the λ-integrase family recombination mechanism as discussed prior to 1995, see references (13, 34, 35, 44, 59).

In the mechanism shown in Figure 2, two Cre-bound loxP sites associate to form a recombination synapse. Two of the four recombinase subunits in the synapse cleave the DNA substrates using conserved tyrosine side chains as nucleophiles, forming covalent 3'-phosphotyrosine linkages to DNA and releasing free 5'-hydroxyl groups. Either the 5'-hydroxyl groups can reseal the nicks to restore the original substrates and complete one round of topoisomerase-I-like cleavage and religation, or the 5'-hydroxyl groups can attack the 3'-phosphotyrosine linkages of the partner substrates, resulting in the exchange of one pair of strands and formation of a Holliday junction (HJ) intermediate. After branch migration through the crossover region, the second pair of strands in the HJ intermediate are then cleaved and exchanged by the second pair of recombinase subunits to form recombinant products.

The tyrosine recombinases display a strong requirement for sequence identity in the crossover region between recombining sites. Early models have explained how each site senses homology with its partner by requiring the HJ intermediate to branch migrate through this region (66). Mismatches in sequence would provide energetic barriers to this process and ensure that only identical crossover regions could be efficiently recombined. The branch migration model is satisfying in terms of explaining the requirements for homology and the observed reversibility of the recombination pathway, but it is difficult to reconcile with the motions required of the recombinase subunits and DNA arms of the HJ in three dimensions (60). In order to branch migrate 6-8 bp, the recombinase-bound arms of the HJ intermediate would need to undergo extensive rotations, requiring that protein-protein interfaces formed in the synapse be disrupted. In the mid 1990s, Landy and coworkers...
Figure 2 The branch migration model for integrase family site-specific recombination (for a review of this mechanism, see reference 13). Two recombinase-bound sites associate to form a recombination synapse (top left). Two subunits cleave the DNA substrates with conserved tyrosine side chains to form 3\'-phosphotyrosine linkages and release free 5\'-hydroxyl groups (middle left). The 5\'-hydroxyl groups undergo intermolecular attack of the partner phosphotyrosine to complete the exchange of one pair of DNA strands between the two substrates and form a Holliday junction intermediate (lower left). The branch point of the junction starts at the site of initial strand exchange and then migrates through the crossover region to the second set of cleavage sites. The second pair of subunits is then activated and/or positioned for cleavage of the bottom substrate strands, which are exchanged to form recombinant products. Heterology between crossover sequences would block efficient branch migration and prevent the reaction from proceeding to the second strand exchange. For simplicity, the DNA sites are shown associating in a parallel orientation in this figure. The same mechanism can also be described with an antiparallel alignment of sites (see reference 4).

proposed an alternative “strand swapping-isomerization” model for integrase family site-specific recombination (45) that was supported by experimental results from a number of laboratories reported at nearly the same time (4, 8, 37, 69). In this mechanism, strand exchange occurs following cleavage of the site by melting 2–3 bases from their complementary strand and annealing to the corresponding complementary strand in the recombining partner. Those bases are then effectively tested for homology with the partner by Watson-Crick base-pairing and subsequent ligation. As discussed in this review, the Cre-DNA structural models strongly
support the strand swapping, but not the branch migration model for integrase family site-specific recombination.

**THE STRUCTURE OF Cre RECOMBINASE**

Before three-dimensional structural data became available, most of what was known about protein structures in the tyrosine recombinase family came from sequence comparisons (2, 6) and limited proteolysis experiments (29, 48). The tyrosine recombinases share limited sequence similarity overall, although it is now clear that family members have a common catalytic domain fold (15, 47). Early sequence analyses revealed four strictly conserved amino acids required for catalysis: the Arg-His-Arg catalytic triad and the tyrosine nucleophile (2, 6).

All four residues were found to be in the C-terminal halves of the recombinase sequences, which show higher sequence similarity than do the more divergent N-terminal halves. As illustrated in Figure 3, Cre is a two-domain protein with a helical N-terminal domain and a larger, primarily helical C-terminal domain.

The complexes of Cre recombinase with DNA substrates revealed an extensive protein-DNA interface between each protein subunit and its contacted RBE (17, 22, 23). The two domains of Cre form a C-shaped clamp that grasps the DNA from opposite sides, where the N-terminal domain interacts primarily with the major groove proximal to the crossover region on one face, and the C-terminal domain interacts with successive minor, major, and minor grooves on the opposite face (Figure 3a). A similar organization has been observed in the Flp-DNA complex (11a) despite clear differences in the structures of the proteins N-terminal to the conserved core catalytic domains. Biochemical evidence in the Xer and λ-int systems, together with the available structural models, indicates that this recombinase-DNA organization is likely to be a common feature of the tyrosine recombinases (26, 63).

In addition to the extensive protein-DNA interface formed between Cre and its recombination site \textit{loxP}, a substantial protein-protein interface is formed between the two subunits bound to RBES on the same site. On one side of the DNA substrate, the two N-terminal domains interact with one another primarily through helix-helix contacts, but they do not interact substantially with the catalytic domains (Figure 4a). On the opposite side, the two catalytic domains interact via the exchange of helices located at the carboxyl termini (Figure 4b). The C-terminal helix in one Cre subunit (helix-N) buries its hydrophobic surface in an acceptor pocket on the adjacent subunit. The network of contacts between N- and C-terminal domains in DNA-bound Cre subunits may in part explain the large cooperativity of binding to the \textit{loxP} site (1, 51).

The C-terminal helix exchange is not reciprocal between Cre subunits bound to the same \textit{loxP} site in the Cre-DNA crystal structures because a second Cre-\textit{lox} complex interacts to form a synaptic assembly and the helix exchange is cyclic among the four subunits (Figure 4b). In the HP1-integrase catalytic domain struc-
ture, however, a C-terminal helix swap occurs in a reciprocal fashion, leading to an HP1-integrase dimer (27). An HP1-like reciprocal exchange of C-terminal helices between DNA-bound Cre subunits appears to be precluded for geometric reasons if the loxP site is bent even slightly. Modeling exercises indicate that even in the case of an unbent loxP site, mutual exchange of C-terminal helices would be difficult and may require partial unfolding or repositioning of the penultimate helix (M) in order to form the correct helix-acceptor interactions (21).

The helix-swapping arrangement observed in the Cre-DNA structures not only provides a model for understanding cooperative binding, bending, and synapsis of loxP sites, but it also provides a provocative mechanistic argument for control of the cleavage reaction. The conserved tyrosine nucleophile (Tyr324) is located close to the peptide linker (Arg326-Gly333) that connects helix-N to the rest of the domain. Changes in quaternary structure of the recombination complex could therefore be sensed by the tyrosine nucleophile through this peptide linker, providing a stereochemical coupling between the overall structure of the recombination assembly and the positioning of a critical catalytic amino acid.

SYNAPSIS AND STRAND EXCHANGE

In order for recombination to occur, two DNA sites must associate to form a synaptic complex, within which the cleavage and strand exchange reactions take place. In the structures of two Cre mutants bound to symmetrized loxP sites (22), the Cre-bound DNA duplexes are bent sharply within the crossover regions and are brought together with all four arms (defined as the DNA segments on either sides of the bends) lying nearly in the same plane (Figure 4). The four recombinase monomers create a pseudo fourfold-symmetric network of protein-protein interactions responsible for holding the synapse together, with a set of interactions between the N-terminal domains on one side of the DNA plane and an independent set of interactions between C-terminal domains on the opposite side of the DNA substrate plane. The complex is strictly twofold symmetric overall, with the dyad of symmetry passing through the center of the complex in a direction perpendicular to the plane of the DNA substrates. A similar twofold-symmetric arrangement of recombinase and substrate is present in the structure of the cleaved covalent Cre-DNA intermediate (23) and in the structures of the Cre- and Flp-bound Holliday junction intermediates (11a, 17). Although the DNA sites in these cases are symmetric and therefore have no sequence-derived directionality that would specify parallel versus antiparallel alignment, the twofold-symmetric structures strongly imply an antiparallel alignment of wild-type recombination sites. The three-dimensional mechanism discussed for the Cre-loxP pathway (18, 22) in fact requires that the sites be aligned in an antiparallel fashion in order for productive strand exchange to occur in the absence of large structural rearrangements.

The nature of the protein-protein contacts in the Cre-DNA synaptic complexes, and more recently in the Flp-HJ complex, have provided a great deal of insight into
many aspects of the recombination process. A striking feature of these structures is the types and the extents of the protein-protein interfaces. In the Cre tetramer, the self-association of N-terminal domains alone buries \( \sim 4000 \ \AA^2 \) of solvent accessible surface area, with the interactions between C-terminal domains burying a further \( \sim 7000 \ \AA^2 \). The N-terminal domains primarily interact with one another (and with the DNA) to form a nearly fourfold-symmetric tetramer via helix-helix and loop-helix interactions. As a consequence of this symmetry, the interactions between two N-terminal domains bound to the same loxP site are nearly identical to the synaptic interactions between N-terminal domains on different sites. This arrangement of N-terminal domains is capable of accommodating the synapsed substrates, the cleaved substrates, and the Holliday junction intermediate in both halves of the pathway with virtually no changes in quaternary structure. The N-terminal domains therefore appear to form a relatively rigid structure that persists throughout the reaction.

Some of the most informative insights from structural models in the tyrosine recombinase system concern the geometry of the cleavage and strand exchange steps in the reaction. Our first snapshot of this process in action came from the crystal structure of Cre recombinase covalently bound to a suicide DNA substrate (23). This intermediate was formed using symmetrized loxP sites that contained nicks adjacent to the scissile phosphates. Cleavage by the recombinase proteins resulted in formation of 3'-phosphotyrosine linkages to the DNA and release of free cytidine molecules, which diffuse away from the active sites. The resulting complex was trapped at the strand exchange step because the departing cytidine takes with it the 5'-hydroxyl required to either continue or reverse the strand exchange reaction, and the remaining strand is too short on the 5'-end to reach either of the activated phosphotyrosines.

The covalent Cre-DNA complex revealed a synaptic assembly with architecture nearly identical to that later observed in the precleavage synaptic complex (Figure 5). The recombinase proteins form a pseudo-fourfold-symmetric tetramer with the DNA duplex substrates in a distorted square planar arrangement. Two of the four protein subunits diagonally related in the complex have cleaved the DNA to form covalent bonds to the scissile phosphate, but the other two subunits have not cleaved the DNA. Remarkably, the 5'-ends of the DNA strands that would be transferred to the partner substrates in the normal reaction (if the 5'-cytidines were present) have partially melted away from their complementary strands, with their 5'-thymidine ends converging in the middle of the synaptic complex (see Figure 5b).

It is tempting to think of this cleaved intermediate structure as a snapshot of the strand exchange process in action. In this context, the complex structure could equally well represent a trapped strand transfer reaction in the HJ-forming step or in the HJ-resolving step. One of the interesting questions concerning this view of the structural model is whether the observed protein-DNA architecture needs to undergo a significant change in quaternary structure in order to accommodate the transfer and ligation of DNA strands between opposing halves of the recombination
Figure 6  Isomerization model for the HJ intermediate in Cre-loxP site-specific recombination. The conformer on the left differs from that on the right by an exchange in the interarm angles and an exchange in the positions of the branch-point phosphates. The stereochemical identities of the dark (crossing) strands on the left are identical to those of the light (crossing) strands on the right. Likewise, the light strands (continuous) on the left are equivalent to the dark (continuous) strands on the right. The strands labeled “crossing” are activated for cleavage and exchange in the two conformers.

assembly. Modeling exercises suggest that the intact strands (with the 5'–cytidine restored) would be able to reach the complementary strands of the partner substrate and form at least two Watson-Crick base pairs immediately adjacent to the scissile phosphate. This annealing process could occur by moving only the three nucleotides at the 5'-ends of the DNA strands and leaving the remainder of the protein-DNA assembly unchanged. With more extensive movement of the DNA strands, a third base pair (and thus, a fully base-paired HJ intermediate) could be formed. The structure of the Cre-HJ intermediate (discussed below) indicates that three bases are exchanged between duplex substrates, placing the branch point of the HJ formed at the center of the loxP site. This implies that exactly half of the 6-bp crossover site between scissile phosphates are exchanged in forming the HJ (Figure 6) and the remaining half are exchanged in the resolution of the HJ intermediate to products.

THE ORDER OF STRAND EXCHANGE

In principle, either of the DNA strands in the two duplex substrates could be cleaved and exchanged first in the reaction pathway. From a structural point of view, it is useful to consider whether there are differences between the DNA strands and/or the active sites in the precleavage synaptic complex that would favor the initial
Cre-loxP RECOMBINATION

cleavage and exchange of one strand over the other. In the Cre-DNA synapse formed with inactive Cre mutants and symmetrized loxP sites (22), there is a clear distinction between the two strands. The strand that corresponds to the one caught in the act of swapping has a helicoidal trajectory that takes it through the middle of the open strand exchange region between duplex substrates. This strand is also relatively free from recombinase-backbone contacts that might limit its mobility. In contrast, the complementary strand is embedded in a more extensive protein interface, which is primarily composed of sugar and phosphate contacts located on the side of the DNA duplex pointing away from the strand exchange region. If this strand were to be exchanged first, a more complicated unraveling from the duplex and dissociation from recombinase interactions would be required.

These observations from the precleavage Cre-DNA complex are consistent with the DNA strand that was caught in the act of exchange in the covalent intermediate structure. Together they suggest that the order of strand exchanges in the recombination pathway is determined once the synaptic complex is formed. However, two distinct antiparallel (and two distinct parallel) synapses are possible, which differ by the direction of bending of the sites. Recombination would therefore be expected to proceed with an opposite order of strand exchanges in these two cases (22). For Cre, a preference for cleaving at the “G/C” end of the loxP crossover (left side in Figure 1) has been reported (30), which indicates that formation of one of the two antiparallel synapses is slightly favored over the other. In principle, a bending preference arising from the crossover sequence itself could be responsible for the observed bias, although there are no published experimental data to support this idea for the Cre-loxP system. Independent biochemical evidence based on studies of bulged DNA substrates in the Cre and Flp systems also supports this relationship between bending of the core recombination site and strand cleavage preference in the synaptic complex (38).

The Flp recombinase, which has a larger crossover region compared to Cre, shows little or no preference for the order of strand exchange during recombination (38,39). Although bending of the frt site by Flp has been well studied (42,43), there is no strong evidence that the bend occurs in a preferred direction, which is consistent with a lack of initial cleavage preference. The λ-Int and Xer systems, on the other hand, show a clear preference for cleaving one strand first during recombination (5,31,46).

THE HOLLIDAY INTERMEDIATE

At present, there are two crystal structures available representing the Cre-HJ intermediate (17) and, more recently, a crystal structure of Flp recombinase bound to an HJ substrate with a 7-bp crossover site (11a). In one of the Cre-bound junctions (HJ2 in reference 16), eight overlapping DNA strands are assembled to form an HJ with four nicks (missing phosphates) that is fourfold symmetric with respect to sequence symmetry and in principle is free to branch migrate within the crossover
region. The branch point is located at the center of the crossover region in this structure. The other Cre-bound HJ (HJ1 in reference 17) is an immobile junction with the branch point fixed by design at the center of the crossover region. The structures of the two junctions are nearly superimposable, indicating that the presence of the nicks and the asymmetries imposed to form an immobile junction most likely do not distort the observed structures significantly from that of the true HJ intermediate in Cre-loxP site-specific recombination.

The Cre-bound HJ arms, defined here as the duplex DNA segments extending away from the branch point, lie nearly in the same plane with a small amount of curvature within each arm that creates a slightly concave surface on the side of the junction plane where the recombinase catalytic domains bind. The arms adopt a distorted square planar arrangement that is strictly only twofold symmetric, with interarm angles of ∼75° and ∼105° (Figure 7). At the center of the junction all bases are Watson-Crick paired, and all eight bases are unstacked and exposed to solvent. The nearly square planar arms and the unstacking at the branch point in the Cre-HJ complexes most closely resemble the square planar form of the HJ characterized in solution by Lilley and coworkers (14) and observed in the crystal structures with bound RuvA protein (25, 52). This free junction conformer exists in the absence of divalent ions or high concentrations of monovalent ions, where phosphate-phosphate repulsions presumably lead to a fully extended fourfold symmetric structure (41).

In addition to the unequal interarm angles, the geometry of the deoxyribose-phosphate backbone is quite different for the two pairs of DNA strands in the Cre-bound HJ. The strands that span the obtuse interarm angle form a more or less uninterrupted helicoidal trajectory through the branch point of the junction. The DNA strands that span the acute interarm angles have a rather sharp discontinuity at the branch point, where the branch point phosphate rotates toward the center of the junction, away from the path traced by the flanking phosphates (Figure 7b). Thus, in the Cre-bound HJ, like in the stacked-X form of the junction in solution (14), there are a pair of continuous DNA strands and a pair of crossing DNA strands that are stereochemically distinct. It follows, therefore, that the recombinase active sites that surround the scissile phosphates on the continuous versus crossing strands in the junction are not stereochemically equivalent. Based on a comparison with the covalent intermediate complex (Figure 5), the crossing strand active sites are activated for cleavage and the continuous strand active sites are prevented from cleaving the DNA during recombination.

It has been understood for many years that a change in the structure of the HJ intermediate in tyrosine recombinase recombination must serve the pivotal role of deciding whether the reaction should proceed forward to recombinant products or backward to restore the original substrates (see discussion in references 13, 17, 45). Early models of this isomerization step involved a migration of the junction branch point between the sites of cleavage within the crossover region (Figure 2). The strands cleaved in resolution of the HJ would be those whose
scissile phosphates were nearest the junction branch point in a given branch-point isomer. In addition to providing a mechanism for reading homology between sites, this model provided a rationale for how pairs of active sites could be turned on and off in response to the quaternary structure of the recombination complex. The Cre-DNA complex structures, and in particular the structures of the Cre-HJ intermediates, have provided a different view of this central isomerization step in the recombination pathway, which along with biochemical studies from several laboratories has led to an alternative model (17, 18).

In the specific case of the Cre-bound HJ complex, there are two different ways one could draw the structure shown in Figure 7b. The conformers differ by which arms form acute and which arms form obtuse angles, and by which DNA strands adopt the crossing configuration with inverted phosphates at the branch point and which strands adopt the more continuous configuration. The Cre-loxP model for HJ isomerization shown in Figure 6 involves a planar scissoring motion of the HJ arms that swap the pairs of interarm angles along with adjustments to the deoxyribose-phosphate backbone torsion angles that lead to a swap of branch point phosphate configurations. The analogy with crossover isomerization of the Holliday intermediate in homologous recombination is evident. In both cases, the generation of strand equivalence by quaternary changes in the junction structure is accomplished.

The effect of HJ isomerization on the catalytic domains of the recombinase subunits mirrors that of the DNA strands. Those pairs of adjacent subunits that are separated by a larger distance in one conformer are closer together in the alternative conformer. Accordingly, the nature of the interactions between the catalytic domains in these two distinct interfaces is exchanged between recombinase pairs. These differences in domain-domain interfaces may directly relate to the positioning of active site residues and therefore in determining which subunits are activated for cleavage. The situation with the N-terminal domains of the Cre subunits is somewhat different. These domains maintain a nearly fourfold-symmetric arrangement in both isomers, although the interactions formed with the crossing and continuous strands of the junction are slightly different.

In terms of active versus inactive subunits in the Cre-HJ model, the subunits whose active sites are positioned on the crossing strands are predicted to be active for strand exchange, and the pair of subunits on the continuous strands are predicted to be rendered inactive. These predictions follow from a comparison with the cleaved Cre-DNA intermediate (23), where the active site that has catalyzed formation of a covalent Cre-DNA linkage corresponds to one containing the crossing strand phosphates in the HJ complex. A similarity also exists at the quaternary structural level between the synapsed substrates and the HJ intermediate. Aside from differences in DNA structure at the center of the synaptic complex due to the strand exchange event, the structures of the synapsed sites, the covalent Cre-DNA intermediate, and the Cre-HJ intermediate are nearly superimposable, allowing for easy prediction of the active subunits in the two complexes where direct evidence of cleavage is not available.
It is clear from Figure 6 that the proposed isomerization of the Cre-HJ intermediate does not involve migration of the branch point from the center of the crossover region. In fact, given the extent of the protein-protein interfaces in the Cre tetramer (average of \( \sim 3400 \) Å\(^2\) surface buried in each pair of adjacent subunits), it is difficult to imagine how this HJ intermediate could undergo any branch migration. A coordinated rotation of the protein-bound DNA arms to move the branch point would require that the observed protein-protein interface be disrupted and a new interface be formed. Even if the old and new interfaces were isoenergetic, considerable activation energy would no doubt be required to accomplish this type of rearrangement.

RECOMBINASE ACTIVE SITES AND PHOSPHORYL TRANSFER

The active site model that has emerged from the integrase family protein and protein domain structures (27, 33, 62), the closely related eukaryotic topoisomerase Ib structures (12, 49), and an analysis of three intermediates in the Cre-DNA pathway is illustrated in Figure 8. The previously identified Arg-His-Arg side chains coordinate the scissile phosphate during recombination, while the conserved tyrosine side chain is positioned near the scissile phosphate, poised for nucleophilic attack. In addition to these four residues, two new participants in the enzyme active

![Active site cleavage model for Cre recombinase.](attachment:image.png)

**Figure 8** Active site cleavage model for Cre recombinase. The general base responsible for accepting a proton from the tyrosine nucleophile during cleavage has not been clearly identified biochemically but could be His-289. The general acid responsible for protonating the 5'-OH leaving group has been identified as the equivalent of Lys-201 in the vaccinia virus topoisomerase and is therefore likely to play the same role in the tyrosine recombinases, as shown here (32). The ligation reaction is simply the reverse of that shown above, beginning with nucleophilic attack of the 3'-phosphotyrosine intermediate by a 5'-hydroxyl group.
site have been identified. A tryptophan forms a hydrogen bond to the scissile phosphate via the indole nitrogen in Cre, whereas a histidine side chain occupies this position in the sequences of λ-Int, HP1-Int, XerD, and the vaccinia and human topoisomerases. A comparison of tyrosine recombinase sequences indicates that this His/Trp hydrogen bond is likely to be conserved throughout the protein family.

The second previously unidentified active site component is a lysine residue located in the loop between β-strands 2 and 3 (Lys201 in Cre). Structural and biochemical studies in the vaccinia virus topoisomerase system were key to realizing the importance of this residue in the site-specific recombinases (12, 67). In the vaccinia virus enzyme, mutation of the corresponding residue (Lys165) results in a loss of $10^4$ in catalytic efficiency (67). The alanine mutants of this lysine in XerD (10) and Cre (F Guo & G Van Duyne, unpublished observations) are each defective in recombination. In the covalent Cre-DNA and human topoisomerase-Ib DNA structures, this lysine contacts a base adjacent to the cleaved phosphate in the minor groove (23, 49). In the synaptic Cre/DNA complexes and in the Cre/HJ complex structures, the loop containing Lys-201 is not well ordered, most likely indicating a high degree of mobility.

A general model for acid-base catalysis of the initial phosphoryl transfer step in the tyrosine recombinase cleavage reaction is shown in Figure 8 (see also references 19 and 58). The Arg-His-Arg triad of conserved side chains, along with the Trp side chain observed in Cre, provides an array of hydrogen bond donors surrounding the scissile phosphate that fulfills the requirement of stabilizing a pentacoordinate transition state of the phosphate and probably contributes an electrophilic catalysis component to the reaction. The equivalent of Lysine-201 has been shown to be the catalytic acid responsible for protonating the 5'-OH leaving group in the vaccinia virus topoisomerase system (32), and it seems quite likely that the same is true in the tyrosine recombinases. The identity of a general base that accepts the Tyr-324 hydroxyl proton is still uncertain, although His-289 is a prime candidate.

**SUMMARY**

The convergence of biochemical, genetic, and structural studies in an area of biological research generally results in a strengthened view of the underlying molecular processes. This is clearly the case for the tyrosine recombinase family of site-specific recombination enzymes. Structural data for five of the integrase family recombinase systems and two eukaryotic topoisomerase Ib systems have provided a three-dimensional framework for understanding years of results obtained by a number of different laboratories. Biochemical studies in the vaccinia virus topoisomerase Ib enzyme have also contributed greatly to our understanding of catalysis in this superfamily.

Our current understanding of the tyrosine recombinase site-specific recombination pathway in the three-dimensional sense has been guided primarily by structures of reaction intermediates in the Cre-loxP system. The most surprising
feature of what we have learned from this system is that the reaction appears to proceed with a single protein-DNA architecture that requires no large changes in quaternary structure in converting substrates to products. A subtle isomerization of the nearly coplanar Holliday junction intermediate between dyad-symmetric forms may be sufficient to act as the active site switch that triggers the exchange of only one of the two pairs of DNA strands to generate either substrates or products. The current model for the Cre-\textit{loxP} recombination pathway is outlined in Figure 9. This mechanism requires antiparallel synapsis of the sites in order to lead to productive strand exchange. The strand exchange process occurs by a strand-swapping mechanism based on that proposed in 1995 by Nunes-Düby et al (45). Three bases are tested for homology with the recombining site and exchanged to form the Holliday intermediate, and the remaining three bases in the crossover region are tested during the second strand exchange step.

Perhaps the most compelling questions remaining at the structural and mechanistic level for the tyrosine recombinases involve the detailed stereochemical interfaces between subunits that regulate the recombination process. While Cre and XerC/D have similar N-terminal domain structures and might be expected to form similar protein-protein contacts in the recombination assembly that could mediate an allosteric regulatory process, it is clear that these contacts have not been conserved in Flp recombinase (11a). A discussion of the stereochemical components of this regulatory system, given the available structural and biochemical data, is presented elsewhere (64).

A closely related question for the future regards the isomerization of the HJ-intermediate in tyrosine recombinase recombination. While the isomerization appears to be quite subtle in the Cre system, requiring only small changes in quaternary structure and no branch migration of the junction, the same need not be true of Flp, Int, and XerCD, where the crossover sequence lengths are in most cases different. In the Flp recombinase 8-bp crossover system, both a limited branch migration model and a simple 4-bp swap model would seem possible (65). In the XerCD 8 bp crossover and \textit{λ}-Int 7 bp crossover systems, the formation of specific nucleoprotein architectures adjacent to the recombination sites leads to interesting questions regarding the role of branch migration in the recombination pathway. The extent to which these nucleoprotein assemblies might inhibit, promote, or simply tolerate a limited branch migration process during recombination (16) will be an important issue in understanding the more complex tyrosine recombination systems in the future.

ACKNOWLEDGMENTS

I gratefully acknowledge the valuable comments, discussions, and support received from members of the site-specific recombination community during my laboratory’s work in the Cre-\textit{loxP} system. I also thank Drs. Feng Guo and Michael Gopaul, who carried out the crystallographic studies in the Cre-\textit{loxP} system described here, and the National Institutes of Health for financial support.
LITERATURE CITED

42. Luetke KH, Sadowski PD. 1995. The role of DNA bending in Flp-mediated...
Cre-loxP RECOMBINATION

53. Sadowski PD. 1995. The FLP recombinase of the 2-microns plasmid of Saccha-


55. Schwartz CJ, Sadowski PD. 1989. FLP recombinase of the 2 microns circle plas-


57. Sherratt DJ, Arciszewska LK, Blakely G, Colloms S, Grant K, et al. 1995. Site-


nases. Trends Genet. 8:432–39


61. Sternberg N, Hamilton D, Austin S, Yarmolinsky M, Hoess RH. 1981. Site-

nase, XerD. EMBO J. 16:5178–87

63. Tirumalai RS, Kwon HJ, Cardente EH, Ellenberger T, Landy A. 1998. Recognition of core-type DNA sites by lambda inte-
grase. J. Mol. Biol. 279:513–27

64. Van Duyne GD. 2001. A structural view of


Figure 3  Domain structure and DNA-binding of Cre recombinase. (a) Ribbon/cylinder drawing of one Cre subunit bound to a \( \text{loxP} \) half-site (representing one half of Figure 1). Helices are labeled as in (b) and active site residues are drawn as red sticks. The scissile phosphate is drawn as a red sphere. (b) Folding pattern of Cre recombinase. Helices A through E comprise the N-terminal domain, and the remainder forms the catalytic domain. A similar fold has been observed for the catalytic domains of \( \lambda \)-integrase (33), HP1-integrase (27), and XerD recombinase (62). This fold forms a subset of the eukaryotic topoisomerase IB (12) and the yeast (Flp) recombinase (50) catalytic domains. The locations of conserved catalytic residues (Cre numbering) are indicated. Ribbon/cylinder illustrations in Figures 3–7 were produced with the program RIBBONS (11).
Figure 4. Structure of the synapsed substrates (or synapsed products) Cre-DNA complex (22). (a) View of the synapse from the amino-terminal domain side. Helices in the N-terminal domain of one subunit are labeled. Subunits colored green are active for substrate cleavage. (b) View of the synapse from the catalytic domain side, rotated 180 degrees about a horizontal axis with respect to (a). Helices in one of the active (green) subunits are labeled and the C-terminal helices-N that interact with adjacent subunits are circled. The helix swap is cyclic among the four subunits.
Figure 5  Structure of the Cre-DNA covalent intermediate (23). 5’-cytidine residues in nicked suicide substrates were cleaved and allowed to diffuse away in order to trap this intermediate (46). (a) View from the N-terminal domain side, as in Figure 4a. (b) The DNA alone in the cleaved intermediate structure, viewed from the catalytic domain side of the complex, as in Figure 4b. Missing 5’-cytidine residues and the direction of strand movement and nucleophilic attack during strand exchange are indicated.
Figure 7  Structure of the Cre-Holliday junction intermediate (17). (a) View of the complex from the N-terminal domain side, as in Figures 4a and 5a. (b) The DNA in the Cre-HJ intermediate structure viewed from the catalytic domain side as in Figures 4b and 5b. The formation of three base-pairs with the partner substrates as a result of strand exchange is indicated by arrows (compare with Figure 5b). Only the phosphodiester backbone of the exchanged bases needs to move in order to form the HJ intermediate.
Figure 9 Mechanism of Cre-\textit{loxP} site-specific recombination, based on the strand-swapping model of Nunes-Düby et al (45) and on structural models of Cre recombinase/DNA complexes. Green subunits are active for cleavage in the top half of the pathway, and purple subunits are active for cleavage in the bottom half of the pathway. The DNA substrates lie nearly in the same plane and undergo only a subtle scissoring motion at the HJ isomerization step of the reaction that serves to switch the roles of the protein subunits and switch strands that are activated for exchange. The mechanism does not require branch migration of the HJ intermediate.
CONTENTS

HYDROGEN BONDING, BASE STACKING, AND STERIC EFFECTS IN DNA REPLICATION, Eric T. Kool 1

STRUCTURES AND PROTON-PUMPING STRATEGIES OF MITOCHONDRIAL RESPIRATORY ENZYMES, Brian E. Schultz, Sunney I. Chan 23

MASS SPECTROMETRY AS A TOOL FOR PROTEIN CRYSTALLOGRAPHY, Steven L. Cohen, Brian T. Chait 67

A STRUCTURAL VIEW OF Cre-loxP SITE-SPECIFIC RECOMBINATION, Gregory D. Van Duyne 87

PROBING THE RELATION BETWEEN FORCE--LIFETIME--AND CHEMISTRY IN SINGLE MOLECULAR BONDS, Evan Evans 105

NMR PROBES OF MOLECULAR DYNAMICS: Overview and Comparison with Other Techniques, Arthur G. Palmer III 129

STRUCTURE OF PROTEINS INVOLVED IN SYNAPTIC VESICLE FUSION IN NEURONS, Axel T. Brunger 157

AB INITIO PROTEIN STRUCTURE PREDICTION: Progress and Prospects, Richard Bonneau, David Baker 173

STRUCTURAL RELATIONSHIPS AMONG REGULATED AND UNREGULATED PHOSPHORYLASES, Jenny L. Buchbinder, Virginia L. Rath, Robert J. Fletterick 191

BIOMOLECULAR SIMULATIONS: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions, Wei Wang, Oreola Donini, Carolina M. Reyes, Peter A. Kollman 211

CHAPERONIN-MEDIATED PROTEIN FOLDING, D. Thirumalai, George H. Lorimer 245

INTERPRETING THE EFFECTS OF SMALL UNCHARGED SOLUTES ON PROTEIN-FOLDING EQUILIBRIA, Paula R. Davis-Searles, Aleister J. Saunders, Dorothy A. Erie, Donald J. Winzor, Gary J. Pielak 271

PHOTOSYSTEM II: The Solid Structural Era, Kyong-Hi Rhee 307

BINDING OF LIGANDS AND ACTIVATION OF TRANSCRIPTION BY NUCLEAR RECEPTORS, Anke C. U. Steinmetz, Jean-Paul Renaud, Dino Moras 329

PROTEIN FOLDING THEORY: From Lattice to All-Atom Models, Leonid Mirny, Eugene Shakhnovich 361

STRUCTURAL INSIGHTS INTO MICROTUBULE FUNCTION, Eva Nogales 397

PROPERTIES AND BIOLOGICAL ACTIVITIES OF THIOREDOXINS, Garth Powis, William R Montfort 421

RIBOZYME STRUCTURES AND MECHANISMS, Elizabeth A. Doherty, Jennifer A. Doudna 457