Title

The Order of Strand Exchanges in Cre-LoxP Recombination and its Basis Suggested by the Crystal Structure of a Cre-LoxP Holliday Junction Complex

Journal

Journal of Molecular Biology, 319(1)

ISSN

0022-2836

Authors

Martin, Shelley S
Pulido, Erik
Chu, Victor C
et al.

Publication Date

2002-05-01

DOI

10.1016/s0022-2836(02)00246-2

Peer reviewed

doi:10.1016/S0022-2836(02)00246-2 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 319, 107–127

JMB

The Order of Strand Exchanges in Cre-LoxP Recombination and its Basis Suggested by the Crystal Structure of a Cre-LoxP Holliday Junction Complex

Shelley S. Martin[1], Erik Pulido[2], Victor C. Chu[1], Tyson S. Lechner[1] and Enoch P. Baldwin[1,3]*

[1]*Section of Molecular and Cellular Biology University of California Davis, 1 Shields Avenue Davis, CA 95616, USA*

[2]*Department of Chemistry San Jose State University 1 Washington Square San Jose, CA 95192-099, USA*

[3]*Department of Chemistry University of California Davis, 1 Shields Avenue Davis, CA 95616, USA*

**Corresponding author*

Cre recombinase uses two pairs of sequential cleavage and religation reactions to exchange homologous DNA strands between 34 base-pair (bp) LoxP recognition sequences. In the oligomeric recombination complex, a switch between "cleaving" and "non-cleaving" subunit conformations regulates the number, order, and regio-specificity of the strand exchanges. However, the particular sequence of events has been in question. From analysis of strand composition of the Holliday junction (HJ) intermediate, we determined that Cre initiates recombination of LoxP by cleaving the upper strand on the left arm. Cre preferred to react with the left arm of a LoxP suicide substrate, but at a similar rate to the right arm, indicating that the first strand to be exchanged is selected prior to cleavage. We propose that during complex assembly the cleaving subunit preferentially associates with the LoxP left arm, directing the first strand exchange to that side. In addition, this biased assembly would enforce productive orientation of LoxP sites in the recombination synapses. A novel Cre-HJ complex structure in which LoxP was oriented with the left arm bound by the cleaving Cre subunit suggested a physical basis for the strand exchange order. Lys86 and Lys201 interact with the left arm scissile adenine base differently than in structures that have a scissile guanine. These interactions are associated with positioning the 198–208 loop, a structural component of the conformational switch, in a configuration that is specific to the cleaving conformation. Our results suggest that strand exchange order and site alignment are regulated by an "induced fit" mechanism in which the cleaving conformation is selectively stabilized through protein–DNA interactions with the scissile base on the strand that is cleaved first.

Keywords: protein–DNA complex co-crystal; integrase family; tyrosine site-specific recombinase; sequence-specific DNA recognition; allosteric regulation of DNA specificity

Introduction

Site-specific recombinases promote precise recombination events by catalyzing cleavage and religation at particular DNA sequences.[1] General features of the reaction include protein recognition of the target DNA, synapsis to bring the recombining strands into proximity, and formation of covalent protein–DNA

Abbreviations used: Amino acids, and nucleotides are abbreviated using the three-letter code; 13 bp repeat(s), 13 nucleotide Cre-binding sequences found in inverted orientation in LoxP; 8 bp spacer, asymmetric eight nucleotide region between the inverted repeats in LoxP; Cre*, His-tagged Cre recombinase, wild-type Cre with six histidine residues inserted between Met1 and Ser2; HJ, Holliday junction; MIR, multiple isomorphous replacement; MPD, 2-methyl-2,4-pentanediol; rmsd, root-mean-squared deviation in atomic positions; TBE, Tris–borate–EDTA; CRX, series of structures of Cre-Lox reaction intermediates determined by the van Duyne Laboratory using symmetrized Lox DNA; 1CRX, covalent intermediate structure; 2CRX and 3CRX, Holliday intermediate structures; 4CRX and 5CRX, precleavage intermediate structures.

E-mail address of the corresponding author: epbaldwin@ucdavis.edu

0022-2836/02/$ - see front matter

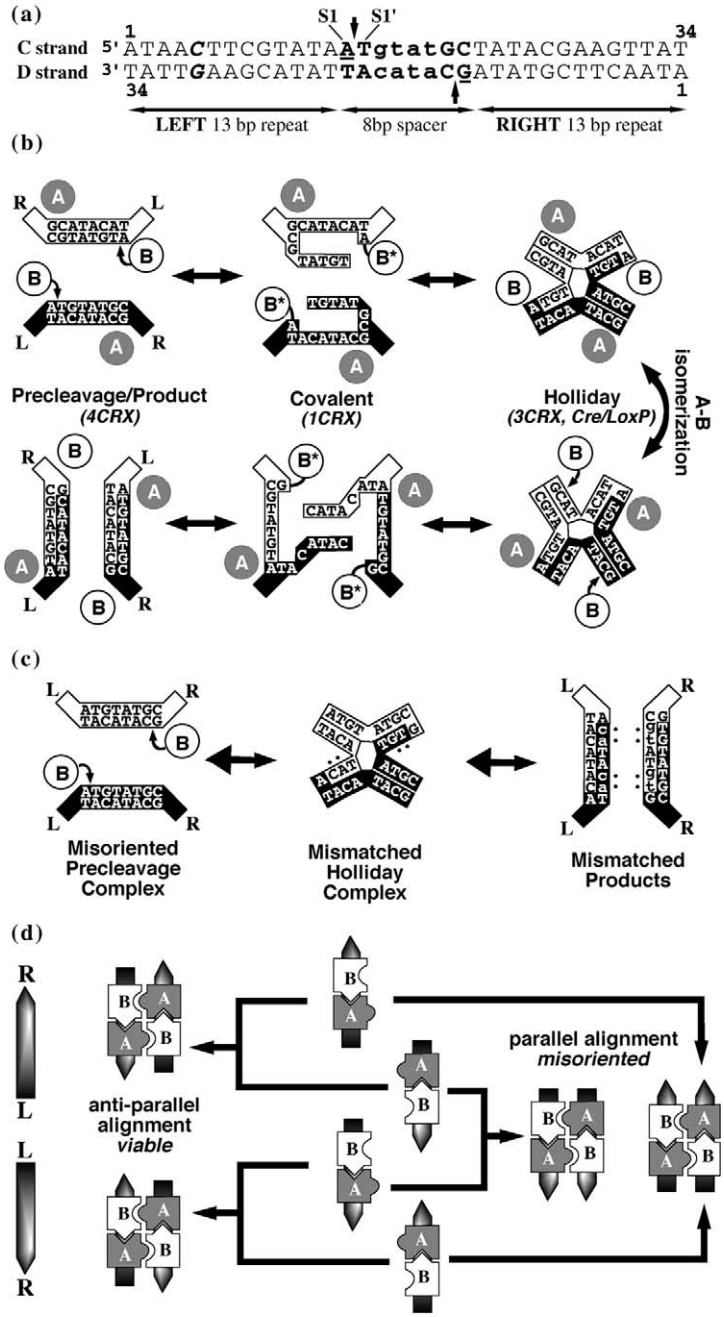

Figure 1. Lox DNA sequences, the mechanism of site-specific recombination, and the importance of site alignment. (a) Diagram of the LoxP recombination site, highlighting the left and right 13 bp repeats (normal type), the 8 bp spacer (bold), and the scissile sites S1 (underlined) and S1′ (upper case bold) and the cleavage site (arrowheads). The S1 and S1′ bases of the left arm are further identified explicitly. The strand numbering starts at the 5′ ends. The LoxP numbering in the text is taken as that for the upper C strand. Position 5, the site of the left arm C5 → G substitution is indicated (bold italic). (b) Proposed mechanism of site-specific recombination showing the correspondence between putative reaction intermediates and X-ray structures 1CRX,[26] 3CRX,[27] and 4CRX[25] (adapted from Guo *et al*.[26]). The 8 bp spacer sequences are arranged in the productive anti-parallel orientation. Analogous intermediates in the upper and lower pathways are aligned vertically, and differ in the relative orientation of the Lox DNA with respect to the non-cleaving A and cleaving B subunits, shown as gray and white circles, respectively. The A–B isomerization step, or "conformational switch", interconverts cleaving and non-cleaving conformations. The covalently attached Cre monomers are marked with an asterisk (∗). (c) Misorientation of LoxP sites in a parallel alignment leads to mismatches in the HJ and product spacers. The mismatched base-pairs are indicated by the asterisks. The higher energy of the products would be expected to skew the reaction equilibrium towards reactants. (d) The alternating subunit arrangement in the synapse dictates anti-parallel orientation of the two A–B Cre dimers in the tetramer. Therefore, the substrate orientations are established by the positioning of the two subunits on the LoxP site. Random association of Cre and LoxP would lead to an equal mixture of viable (left) and inviable misaligned (right) complexes.

intermediates upon cleavage to preserve the phosphodiester bond energy for religation.[2] Because of the precise positioning of strand exchanges, and the fact that no free broken DNA ends are created, site-specific recombinases have proven to be extremely useful for programmed DNA rearrangements in living cells.[3,4]

Site-specific recombination systems originate in bacteria, fungi and archae.[5] The biological roles for recombinase-induced rearrangements are diverse, including control of gene expression,[6] prophage integration,[7] control of plasmid copy number,[8,9] and segregation of circular chromosomes after conservative replication.[10] The Int family of tyrosine recombinases and type 1B topoisomerases is characterized by a conserved tyrosine nucleophile, active-site structure, and protein fold.[11–15] During the reaction, the DNA strands are cleaved, exchanged and religated in two stages, one homologous pair at a time. The two cleavages occur 6–8 bp apart yielding an overhanging 5′-hydroxyl end and a 3′-phosphotyrosyl DNA–protein

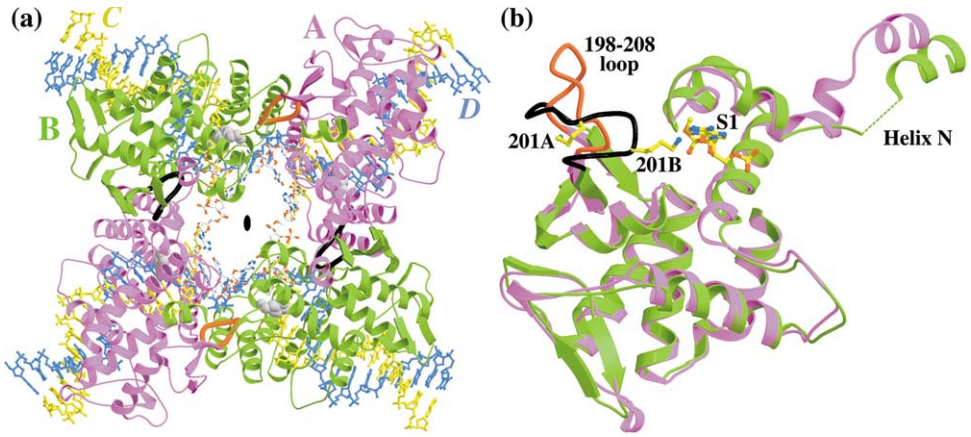

Figure 2. Cre-Lox complex structure, represented by Cre*/LoxP-G5. (a) Overall architecture of the synapse. The non-cleaving A (purple) and cleaving B (green) Cre subunits are shown in ribbon representation. The active-site tyrosine residues for both subunits are shown for the upper dimer in gray CPK spheres. The crossover C (yellow) and non-crossover D (blue) DNA strands are shown as ball-and-sticks. The 3CRX-like alternative conformation for the HJ crossover strand is also shown (white). The tetramer is generated by crystallographic 2-fold symmetry from the unique Cre_2-Lox dimer. The position of the 2-fold axis is indicated by the black oval at the center of the complex. (b) Main-chain structural differences between residues 138–341 of the cleaving (green) and non-cleaving subunits (purple) after superposition. The broken green line depicts the connection made by residues 329–332, which are disordered in the cleaving subunit of Cre*/LoxP-G5. The largest difference between subunits is in the 198-208 loop, shown in thick cable representation. In the cleaving subunit loop (black), Lys201 contacts the scissile base (201B, S1) while in the non-cleaving subunit (red), the loop positions this residue (201A) far from the scissile base.

intermediate. Crossing-over is effected by reciprocal exchange of the 5′ ends, followed by religation through in-line displacement of the tyrosine residue. The first cleavage and religation steps generate a Holliday junction (HJ) and the second leads to the recombined products.[16]

Several Int enzymes function autonomously and require only 30–40 bp DNA recognition sequences to effect recombination. Phage P1 Cre recombinase (38 kDa) promotes strand exchange within the 34 bp LoxP site, as well as other Lox sequences (Figure 1(a)).[17] Cre-Lox recombination has been utilized extensively for generating chromosomal deletions, insertions, and translocations in a wide variety of cell types and live organisms.[18] LoxP contains two 13 bp Cre binding sequences in an inverted orientation (13 bp repeats).[19] The 13 bp repeats are separated by an asymmetric 8 bp intervening sequence (8 bp spacer) in which strand cleavage, transfer, and religation occur. The 8 bp spacer asymmetry divides LoxP into two 17 bp domains, the left arm and the right arm half-sites. The DNA cleavages occur at the scissile sites between the first and second bases of each strand of the spacer,[20] which we denote as S1 and S1′. The left arm scissile site contains A and T on the upper strand S1 and S1′ positions, respectively, while the right arm has G and C on the lower strand. The interior six nucleotides of each strand of the spacer are swapped during the reaction. Homology is required for efficient recombination of some but not all substrates.[21,22]

Cre is the most structurally well-characterized Int family member.[23,24] In conjunction with a number of biochemical studies, crystal structures of Cre-Lox complexes reported by Guo, Gopaul, and van Duyne defined the main mechanistic features of Int site-specific recombination. These structures depict three key reaction intermediates, represented by the PDB entries 4CRX and 5CRX,[25] 1CRX,[26] and 2CRX and 3CRX.[27] They are quite similar and contain a pseudo 4-fold Cre tetramer bound to two bent Lox duplexes arranged in a nearly planar X-shaped structure. The Cre/LoxP HJ complex described here is representative of this architecture, as depicted in Figure 2(a). Two conformationally and functionally distinct Cre monomers, represented by the A and B chains, are bound to each Lox site. The two strands of the Lox duplex are represented by the C and D chains. In the crystals, the tetrameric synapse is generated by 2-fold crystallographic symmetry from the unique dimer–DNA half-complex. The DNA arms are unequally splayed with 102–106° angles between arms of the same Lox site, and 65–72° angles between arms of partner Lox sites. Positional shifts in the active site, the 8 bp spacer, protein–protein interfaces, and the relative orientations of the subunits and DNA arms account for the structural variation between intermediate structures. The A and B subunit conformations differ primarily in the disposition of the hinge between the C-terminal domain and the N helix, active-site structure, and of interest here, the position of the 198–208 loop, which contains the catalytic residue Lys201 (Figure 2(b)).

The recombination mechanism, as defined by the existing crystal structures, is depicted in Figure 1(b).[26] The strategies used to crystallize these putative intermediates are summarized in

Table 1. Summary of DNA and protein components in Cre–DNA crystal structures

Structure[a] (resolution (Å))	Mechanistic intermediate[b]	Lox DNA/Cre protein[c]	8 bp spacer DNA configuration[d]	Reference
1KBU (2.2)	HJ	LoxP with C5 → G/WT[e]	5'-pApTpGpTpApTpGpCp-3' 3'-pTpApCpApTpApCpGp-5	This work
1CRX (2.4)	Covalent	Symmetric half-site suicide substrate/WT	5'-pGpC*ApTpApTpGpCp-3' 3'-pCpGpTpApTpA*CpGp-5'	26
2CRX (2.7)	HJ	Symmetric half-site with gap at cleavage site/WT	5'-pG*CpApTpApTpGpCp-3' 3'-pCpGpTpApTpApC*Gp-5'	27
3CRX (2.5)	HJ	Pre-formed HJ from four oligonucleotides/non-cleaving R173K[f]	3'-pCpGpTpA-pApCpApTp-5' 5'-pGpCpApT pTpGpTpAp-3' | | 3'-pTpApCpAp ApTpGpCp-5' 5'-pApTpGpTp-TpApCpGp-3'	27
4CRX (2.2)	Pre-cleavage	Symmetric full site/non-cleaving R173K[f]	5' pGpCpApTpApTpGpCp-3' 3' pCpGpTpApTpApCpGp-5'	25
5CRX (2.7)	Pre-cleavage	Symmetric full site/non-cleaving Y324F[f]	5'-pGpCpApTpApTpGpCp-3' 3'-pCpGpTpApTpApCpGp-5'	25
1F44 (2.05)	Trimer	1CRX half-site/non-cleaving Y324F[e,f]	5'-pGpC*ApTp-ApTpGpCp-3' 3'-pCpGpTpAp TpApCpGp-5' pT-pA pA-pT pC-pG 5'-pG-pC-3'	37

[a] PDB accession number of crystal structures. With the exception of 1F44, all complexes were tetrameric and isomorphous, space group $C222_1$, cell dimensions $a = 108$ Å, $b = 122$ Å, $c = 180$ Å, and contain one Cre dimer and a full Lox site in the asymmetric unit. The trimer 1F44 was solved in space group $I23$, $a = 160.5$ Å. MIR methods were used to solve 1CRX and 1F44, while Fourier difference methods were used to solve the others.

[b] Intermediates according to the proposed reaction mechanism from Guo *et al.*[26] as depicted in Figure 1(b). The trimer 1F44 is not an intermediate in this scheme but represents Cre bound to an alternative substrate, a Y-DNA junction.

[c] Crystallization substrates for 1CRX, 2CRX, 4CRX, 5CRX, and 1F44, contained the LoxP 13 bp repeat, ATAACTTCGATAT. The 1KBU and 3CRX substrates were substituted at the fifth position of the 13 bp repeat (underlined, see also Figure 1(a)). The 1KBU substrate contained a minimal 34 bp duplex. All other substrates were appended by an additional Thy residue external to the Lox sequence, which constituted a single-base 5′ overhang. Half-site duplexes formed complete lox sites when annealed *via* symmetric 4–6 bp 5′ overhangs (1CRX, 2CRX, 1F44). Full sites were formed by annealing two continuous strands (4CRX, 5CRX, and 1KBU). For 3CRX, formation of preformed HJ was promoted by using four 35 nt duplexes containing distinct 8 bp spacer sequences and substitutions at positions 5 and 30 (see Figure 1(a) for numbering).

[d] Nucleotides in the 8 bp spacer of each substrate. The 1KBU substrate contained the native LoxP spacer. Substrates for 1CRX, 2CRX, 4CRX and 5CRX had symmetric spacers. In 3CRX, the spacer sequences for each strand were distinct. The 1KBU and 2CRX HJs and the 1F44 three-way junction formed spontaneously. Half-site substrates contained a gap between nucleotides, indicated by an asterisk (*), that lacked a bridging phosphate group. The base that is 5′ to the scissile phosphate group, the "scissile base", that was observed in the cleaving subunit active site is underlined. For 3CRX, the scissile bases are those given in the published model, which were interpreted to be averaged about the central axis of the tetramer and allowed for a Gua or a Cyt scissile base. However, the data did not allow for their unambiguous identification (see Gopaul *et al.*[27] for details).

[e] These Cre proteins contain an addition six histidine residues fused to the N terminus to facilitate purification, Cre* in this work (see Woods *et al.*[37]).

[f] These Cre proteins were inactivated by substitution of essential catalytic residues.

Table 1. The reaction is initiated when inactive Cre monomers[28] assemble cooperatively with LoxP sites[29] into the synaptic complex. In the pre-cleavage complex, represented by 4CRX and 5CRX, the 8 bp spacers are brought into proximity in the central "strand transfer channel" (Figure 2(a)).[26] After synapsis, the DNA strands are exchanged in two sequential transesterification reactions. In the covalent complex (1CRX), each Lox site is nicked on the "crossover" strand and becomes attached covalently to Tyr324.[2,20] After cleavage, the free 5′-ends diffuse across the channel and transesterify to the opposite strands, creating a Holliday structure (2CRX and 3CRX). To complete the reaction, second strand cleavage and religation proceed *via* an analogous set of steps (Figure 1(b), lower path). Biochemical and crystallographic analysis of the related FLP-FRT system support

the generality of the Cre-Lox mechanism for Int family recombination, but with significant differences in the details.[2,30–33]

A key aspect of the Cre mechanism is the regiotemporal regulation of the four required strand cleavages.[23,24] If cleavages occurred randomly, only a small percentage of complexes would be competent to exchange homologous strands. Instead, they are coordinated by a switch between "non-cleaving" and "cleaving" conformations, represented by the A and B subunits, respectively. At any particular step, only the cleaving B subunit is involved in strand exchange with the crossover C strand located in its active site, while the non-cleaving A subunit performs an architectural role and has the non-crossover D strand in its active site. After creation of the HJ intermediate, the complex undergoes a conformational switch that exchanges the cleaving and non-cleaving subunit conformations (A–B isomerization, Figure 1(b)). In conjunction with isomerization, the DNA arms "scissor",[26,27] converting the non-crossover strands into crossover strands for cleavage and religation. The presence of only two active conformers guarantees that only two strands will be cleaved at any one time and explains the "half-of-the-sites" reactivity of Int recombinases.[34] Cleavages are restricted to the diagonally related strands, since the cleaving subunits are oriented across the complex (Figure 2(a)). The conformational switch ensures that the two sets of cleavages will occur in an ordered fashion.

Given the near-symmetry of the LoxP substrate, the recombination mechanism appears symmetrical, centered around the A–B isomerization step. However, asymmetry of LoxP within the 8 bp spacer has several mechanistic consequences. An asymmetric substrate generates six distinct reaction intermediates. Analogous pairs, depicted in the upper and lower paths in Figure 1(b), are differentiated by the positions of the cleaving and non-cleaving subunits relative to the Lox site orientation. This arrangement introduces directionality into the reaction. Is there a preferred order to strand exchanges? Geometric considerations suggest that cleavages would likely be ordered. Anti-parallel assembly of LoxP sites leads to the expected recombination products (Figure 1(b)), whereas assembly in parallel orientation would lead to products with four mismatches in the spacer (Figure 1(c)). Since the reaction is under equilibrium control, conversion to the higher-energy products would be inefficient. Although the 8 bp spacer determines the site alignment,[22] these nucleotides do not interact in the initial synapsis step. Instead, their orientations would be determined by the initial positioning of the cleaving and non-cleaving subunits as illustrated in Figure 1(d). If this positioning were random, then one-half of the precleavage complexes would be expected to have parallel Lox orientations, as in Figure 1(c), and would need to dissociate, isomerize, and reassemble in order to reorient them. Of the correctly oriented anti-parallel complexes, half would have the cleaving subunit bound to the left arm, the other half bound to the right arm, leading to an equal mixture of upper and lower strand exchanges in the first step. On the other hand, if the cleaving subunit consistently assembled on one arm, a single precleavage complex would be produced, and strand exchange would always be initiated from that arm. Consequently, maximizing assembly of properly oriented complexes would seem to impose a step order.

While the CRX structures were key to defining the mechanism, they are not as informative with regards to Lox orientation and the potential role of the 8 bp spacer in specifying it. Although the Cre dimers are asymmetric, the Lox DNA used for crystallization was not (Table 1). In 1CRX, 2CRX, 4CRX, and 5CRX, crystallization substrates were composed of 2-fold related LoxP right arm half-sites yielding a symmetric **GC**ATAT**GC** spacer, while in 3CRX, the pre-formed HJ DNA was rotationally averaged. The occurrence of asymmetric Cre dimers in the context of the symmetric substrates indicates that subunit conformations are determined by the Lox DNA geometry and the pattern of protein–protein interactions (Figure 1(d)). However, if cleavages are ordered, the relative bias in subunit positioning would necessarily be determined by the Cre–DNA interaction.

The available biochemical data suggest that cleavages are ordered,[16,20] but the precise order has not been addressed straightforwardly. Analysis of unusual *in vivo* recombination products suggested that cleavage was initiated on the lower strand,[20] a view that has been restated[21,35] but not verified. In apparent accord with these data, the symmetric CRX substrates were based in the right arm sequence (Table 1, Figure 1(a)).[25–27] On the other hand, upper strand cleavage was favored fourfold in covalent complex formation with LoxP[36] and in LoxP HJ resolution.[35] While these studies indicate thermodynamic preference for covalent interaction with the left arm, they do not directly resolve the issue of which strand is exchanged first in Cre-LoxP recombination.

To study further the recombinase–DNA interaction, we have begun crystallizing Cre-LoxP complexes. While the CRX crystals were obtained using variant DNA substrates and mutant proteins, we wondered which intermediate structure would crystallize spontaneously with active enzyme and native substrate. From wild-type Cre with a minimal LoxP DNA site, we obtained a distinct Cre-HJ complex, in which the cleaving subunit associated with one arm of LoxP. However, we could not directly assign this structure to the "first" or "second" HJ-bound species (Figure 1(b)) without knowing the order of strand exchanges. By determining the strands transferred to form the HJ intermediate, we found that Cre cleaves this same arm first. Inspection of the crystal structure suggested that a sequence-specific interaction at the S1 scissile site contributes to the conformational

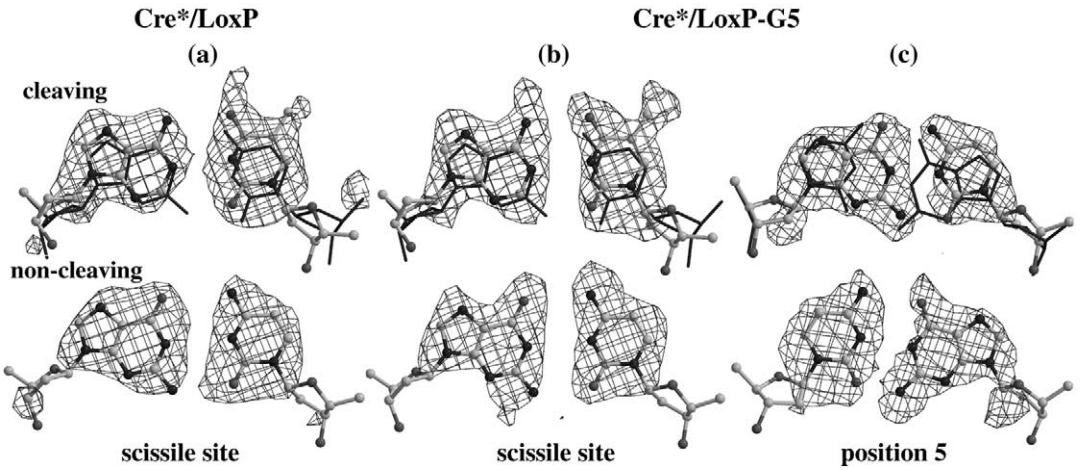

Figure 3. Evidence for the orientation of LoxP in the Cre-HJ complexes. Omit-refine $F_o - F_c$ difference maps were calculated using phases from a partially (Cre*/LoxP) or fully refined (Cre*/LoxP-G5) model in which atoms for the indicated bases were removed, and the models further refined as described in Materials and Methods. The Cre/LoxP atoms are shown in grayscale ball-and-sticks while the 4CRX atoms[25] are shown as black sticks. (a) 2.7 Å $F_o - F_c$ density for S1 bases in the cleaving (upper, +2.3σ) and non-cleaving (lower, +3σ) active sites in the Cre*/LoxP complex. Note the clearly delineated densities for the left arm A–T scissile base-pair in cleaving subunit active sites. (b) The 2.2 Å omit-refine $F_o - F_c$ density for S1 bases in the cleaving (+2.3σ) and non-cleaving (+3σ) subunit active sites in the Cre*/LoxP-G5 complex. Relative to Cre*/LoxP-G5, the R_{work} and R_{free} values after 50 refinement cycles changed $< \pm 0.1\%$ and +0.7%. (c) Omit-refine $F_o - F_c$ Difference map indicating the presence of the left arm substituted guanine base at position 5 associated with the cleaving subunit (upper, +3.5σ) and not the non-cleaving subunit (lower, +2.9σ). Relative to Cre*/LoxP-G5, the R_{work} and R_{free} values after 50 refinement cycles changed $< \pm 0.1\%$ and +0.3%.

Table 2. Data collection and refinement statistics for the Cre*/LoxP-G5 complex

Data set[a]	
Resolution (Å) (last shell)[b]	2.2 (2.32–2.20)
No. measurements	161,433
No. unique reflections	44,903
Coverage (%)	89
R_{sym} (%) (last shell)[b]	3.3 (28.7)
Wavelength (Å)	1.0
Wilson B (Å^2)	52
Refinement	
Resolution (Å)	5–2.2
Space group	$C222_1$
Unit cell (Å)	$a = 107.17$, $b = 121.60$, $c = 179.31$
No. reflections (completeness (%))	
Working set	45,726 (84)
Free set	2441 (4)
No. atoms	6673 (6691)[c]
No. protein atoms	5089
No. DNA atoms	1347 (1511)[c]
No. solvent atoms	237
Average B (Å^2)	57.9
Protein	58.5
DNA	55.6
Solvent	55.0
Rms bond lengths[d] (Å)	0.006
Rms bond angles[d] (deg.)	1.61
Rms B (overall)[d] (Å^2)	2.60
R_{factor}[e]	0.231
R_{free}[e]	0.279

[a] Data were collected at SSRL, beamline 9-2 using an ADSC CCD detector.
[b] $\Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$ for all of the data.[62,63]
[c] The numbers in parentheses consider the DNA two conformations as separate atoms.
[d] Calculated by TNT[66] using Engh & Huber parameters.[67]
[e] $\Sigma|F_o - F_c|/\Sigma|F_o|$ calculated by TNT.[66]

preference of the complex, which ultimately determines strand exchange order.

Results

Structure solution of Cre-LoxP complexes

Complex crystals containing histidine-tagged wild-type Cre[37] (Cre*) and LoxP DNA (Figure 1(a)) were obtained using slight variation of previously published conditions.[26] The crystals were isomorphous to the CRX crystals and diffracted to 2.7 Å using Cu Kα radiation at 100 K. Our initial observations were made from 2.7 Å resolution density maps calculated using home source data phased with a partially refined Cre*/LoxP structure (see Materials and Methods). Omit-refine difference maps exhibited electron density consistent with an adenine base at the S1 position in the cleaving subunit active site and a guanine base in the non-cleaving subunit active site, indicating that LoxP was oriented preferentially (Figure 3(a)). To further investigate this biased LoxP orientation, we subsequently crystallized and determined the structure of a complex containing a LoxP derivative in which the left arm was tagged specifically with a cytidine-to-guanine substitution at position 5, LoxP-G5 (Figure 1(a)). This substitution had no detectable effect on reaction yield (data not shown) and Cre has no specific interaction with these bases.[27] A 2.2 Å synchrotron data set from a Cre*/LoxP-G5 complex crystal was obtained at SSRL beamline 9-2. A new hybrid model was

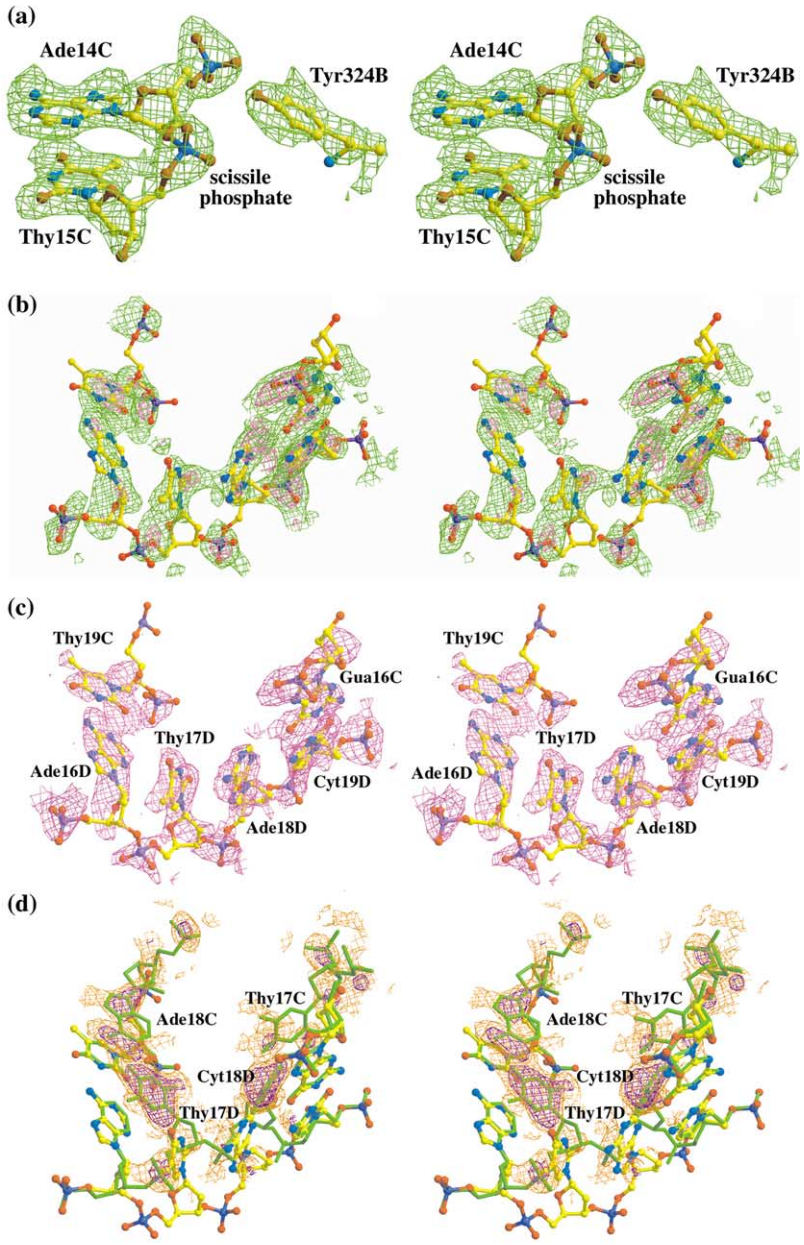

Figure 4. (a) Stereo view and omit-refine $F_o - F_c$ difference map (green, $+2.5\sigma$) of the cleaving subunit active site showing that the scissile bond is intact in Cre*/LoxP-G5. Relative to Cre*/LoxP-G5, the R_{work} and R_{free} values after 50 refinement cycles changed $< \pm 0.1\%$ and $+0.9\%$, repectively. (b) Stereo view of the $F_o - F_c$ difference density (magenta, $+2.3\sigma$, and green, $+1.9\sigma$) in the junction region after initial refinement. The 4CRX model with the four central base-pairs of LoxS omitted was subjected to rigid body, positional and *B*-factor refinement against the Cre*/LoxP-G5 data set. The R_{work} and R_{free} values were 27.2% and 31.0%, respectively. (c) and (d) Stereo views of the two conformations of the DNA junction in Cre*/LoxP-G5, along with omit-refine $F_o - F_c$ maps (see Materials and Methods). For each conformation, the positions and *B*-factors of the alternate conformation were fixed, while parameters for the rest of the structure were refined, to prevent distortion of the omitted density from positional or *B*-factor shifts in the overlapping alternate conformation. (c) Omit-refine $F_o - F_c$ difference map (magenta, $+2.0\sigma$) of the dominant HJ DNA conformation, superimposed on the structure (atom-colored ball-and-stick). Relative to Cre*/LoxP-G5, the R_{work} and R_{free} values after 50 refinement cycles changed -0.4% and $+0.2\%$, respectively (d) Omit $F_o - F_c$ difference map of the minor 3CRX-like conformation (magenta, $+2.3\sigma$, and orange, $+1.3\sigma$) superimposed on the structure (atom-color ball-and-stick) with the dominant conformation indicated for comparison (green sticks). Relative to Cre*/LoxP-G5, the R_{work} and R_{free} values after 50 refinement cycles changed -0.5% and $< \pm 0.1\%$, respectively.

constructed from 1CRX and 4CRX models refined against the 2.2 Å data set (see Materials and Methods). The structure was adjusted for the Lox substitution and further improved by several rounds of model building and refinement against these data. The final model had an R_{factor} of 23.1% and a free R_{factor} value (R_{free}) of 27.9%. The data collection and refinement statistics are shown in Table 2. This model was deposited in the Protein Data Bank (see Materials and Methods). The structure descriptions below refer to this complex, but the electron density and model are comparable to the Cre*/LoxP complex except near the substituted bases.

The majority of the structure was well defined except for the region adjacent to the active site and the DNA junction. Particularly, the protein electron density near the hinge of the cleaving subunit was discontinuous (see Materials and Methods). The DNA was better defined overall with a lower overall *B*-factor (Table 2), except in the central region of the junction, where the electron density was blurred and two conformations were apparent. However, its appearance was essentially identical in four independently solved Cre*/LoxP derivative structures (S.S.M., T.S.L. & E.P.B., unpublished results), suggesting that the apparent disorder

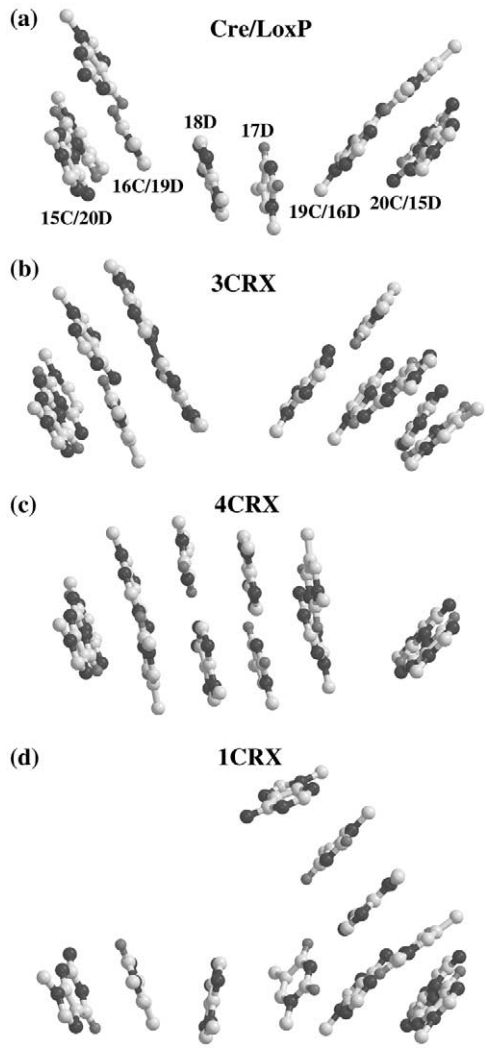

Figure 5. Base-stacking differences in the central six base-pairs of Lox sites in different complex structures. The numbering of the bases and the chain designators are shown for: (a) the dominant conformation of Cre*/LoxP-G5, but apply also to the others; (b) 3CRX HJ complex; (c) 4CRX precleavage complex; and (d) 1CRX covalent complex. Note that the Cre*/LoxP-G5 complex has a unique pattern of unstacked bases and is lacking base-pair partners for 17D and 18D.

is intrinsic rather than being due to phase errors from the model.

Overall structure, intermediate state and LoxP orientation

Although similar in organization, this complex is distinct from the CRX structures. The overall positioning of the Cre subunits and DNA is most similar to 4CRX, rmsd = 0.8 Å for protein backbone atoms of the AB dimer (main-chain atoms of residues common to the compared structures, see Materials and Methods for details of structural comparisons), compared to 1.2–1.3 Å for others. Further, the center-of-mass distance between subunits bound to the same LoxP site is 0.8 Å closer than in 3CRX, while that between subunits across the junction is 1.3 Å closer. These differences in quaternary structure from that of 3CRX are comparable to those from 1CRX (0.8 Å and 1.6 Å), and in contrast the closer correspondence to 4CRX (0.1 Å and 0.7 Å). However, on the basis of the electron density in the active site and strand transfer channel, we suggest that Cre*/LoxP-G5 contains an HJ. As indicated from omit-refine $F_o - F_c$ maps (see Materials and Methods for details of map calculations), the scissile phosphodiester bond is intact and Tyr324 is positioned inappropriately to participate in a covalent bond, eliminating the possibility of an intact or hydrolyzed covalent intermediate (Figure 4(a)). Further, the distance between the Gua16C 3′-hydroxyl group and the Thy19C phosphate group within the same dimer is 23 Å, too long to be bridged by the missing nucleotides, confirming that Cre*/LoxP-G5 is not a precleavage analog. In agreement with this idea, there was no electron density corresponding to the bridging bases in 4CRX. On the other hand, the equivalent connection can made with the symmetry-related dimer, 11.4 Å, consistent with an inter-duplex bridge of a four-way junction, 10.9 Å in 3CRX.

Having identified the Cre-HJ complex structure, we wanted to know the orientation of the LoxP DNA. Two configurations were possible, one with the Lox left arm bound by the cleaving subunit (Figure 1(b), upper HJ intermediate) or one with the right arm bound (Figure 1(b), lower HJ intermediate). We distinguished them from the identity of the bases at the S1 positions in each subunit. As was observed for the Cre*/LoxP data, omit-refine $F_o - F_c$ maps of these regions showed an A–T base-pair in the B subunit active site and a G–C base-pair in the non-cleaving subunit active site (Figure 3(b)). The orientation was corroborated by clearly defined difference density for the Gua5 substitution in the half-site bound by the cleaving subunit, and the wild-type C–G base-pair at the corresponding position in the non-cleaving subunit-bound half-site (Figure 3(c)).

The LoxP Holliday junction

The junction exhibited two distinct conformations. Predominantly, the electron density indicated a novel positioning for the non-crossover D strand that was observable in unbiased and omit-refine $F_o - F_c$ maps (Figure 4(b) and (c)). After accounting for this novel configuration, residual density in the junction region overlapped the HJ DNA in 3CRX. While Thy17D was readily identifiable (Figure 4(d)), further corresponding difference density was observed only at very low contour levels ($+1.0 - +1.3\sigma$, not shown). In the final model, both the novel and 3CRX-like conformations for residues C16 to C19 and D16 to D19 are incorporated. These were weighted at 0.73 and 0.27, respectively, as determined by refinement of

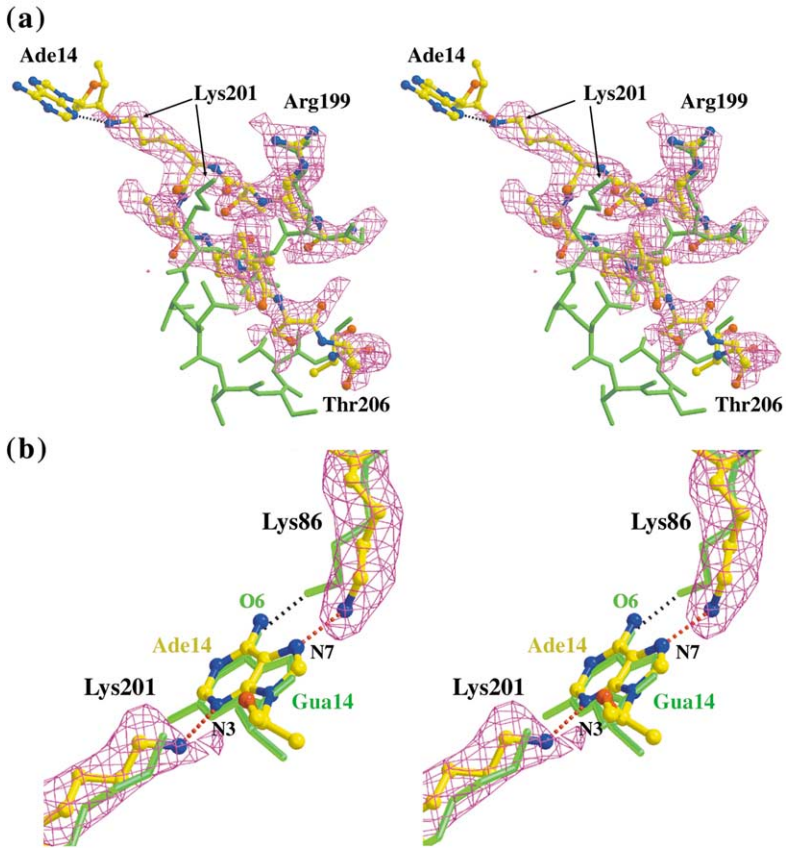

Figure 6. Stereo views of positional differences in the 198–208 loop and scissile base regions of Cre*/LoxP-G5 and other structures. (a) The 198–208 loop in Cre*/LoxP-G5 (atom colored ball-and-stick) is displaced 3.4 Å (main-chain rmsd for residues 198–208) compared to 3CRX, after superposition of residues 20–197 and 209–341 (green sticks). Note the large shift in the Lys201 position (arrow). The density in the omit-refine $F_o - F_c$ map (magenta, $+2.5\sigma$) corroborates the placement of this segment. Relative to Cre*/LoxP-G5, the R_{work} and R_{free} values after 50 refinement cycles changed $< \pm 0.1\%$ and $+0.2\%$. (b) Comparison of interactions between Lys86 and Lys201 in Cre*/LoxP-G5 and 1CRX, superimposed and colored as in (a). In 1CRX, Lys86 NZ makes a hydrogen bond with the scissile guanine O6 atom (black dotted line), while in Cre*/LoxP-G5 it forms hydrogen bonds with N7 (red dotted line). Lys201 NZ atom forms a hydrogen bond with the Ade14 N3 in Cre*/LoxP-G5 (red dotted line), while in 1CRX atoms CE and NZ make only van der Waals contacts with the Gua14. The omit-refine $F_o - F_c$ omit map density (magenta, $+2.3\sigma$) corroborates the Cre*/LoxP-G5 atom placements. Relative to Cre*/LoxP-G5, the R_{work} and R_{free} values after 50 refinement cycles changed -0.3% and $+0.1\%$.

group occupancies. In the 3CRX-like conformation, base stacking and hydrogen bonding are continuous up to the junction, where unstacked 17C-18D and 18C-17D base-pairs form a diamond-shaped channel (Figures 2(a) and 5(b)). In the dominant novel conformation, base-pairing and stacking patterns are altered, due to the shift of the non-crossover strand. The appearance is that of the 3CRX junction in which the two central nucleotides, 17D and 18D, are displaced away from the center and unstack with their neighbors 16D and 19D through shear movements along the base planes (compare Figure 5(a) and (b)). The arrangement accommodated the compressed HJ resulting from the reduced intersubunit distances mentioned above, and the distances between scissile base C1′ atoms are 0.9 Å closer than those in 3CRX. The 4.7 Å difference between the DNA sugar-phosphate backbone positions of residues D16 to D19 was accompanied by multiple rotations of phosphodiester linkages. This outward shift blocked the formation of the 17C-18D and 18C-17D base-pairs observed in the 2CRX and 3CRX junctions (Figure 5(b)), and the disorder of bridging nucleotides is likely a consequence of this disruption. This junction configuration is in contrast to the 4CRX arrangement, in which bases 19C and 20C are unstacked (Figure 5(c)). However, similar to 4CRX and 1CRX (Figure 5(d)), the backbone of the non-crossover D strand at the junction is underwound relative to that of *B*-DNA.

Protein structural differences: interaction with the scissile base

While the polypeptide conformation in Cre*/LoxP-G5 would be expected to most resemble 3CRX, it matched 1CRX and 4CRX more closely. For the non-cleaving subunit, 1CRX was most similar, (main-chain rmsd, 0.7 Å). A better model for the cleaving subunit was a hybrid composed of 1CRX (residues 20–300, main-chain rmsd, 0.34 Å from the final Cre*/LoxP-G5 model) and 4CRX (residues 301–325 and 333–341, main-chain rmsd, 0.8 Å). The most striking departure from 3CRX was the positioning of the 198–208 loop of the cleaving subunit. As shown in Figure 6(a), the loop is well defined but traces a completely different path from that in 3CRX (main-chain rmsd, 3.4 Å for residues 198–208 after superposition of the B subunits), where it folded away from the active site. In 1CRX and 4CRX, this loop extended to the scissile site where Lys201 contacted the S1 base in the minor groove. The Cre*/LoxP-G5 198–208 loop is most similar to 1CRX (main-chain rmsd, 0.54 Å, compared to 4CRX, rmsd 1.9 Å). In 1CRX and 4CRX, atoms CE and NZ of Lys201 make van der Waals contacts with

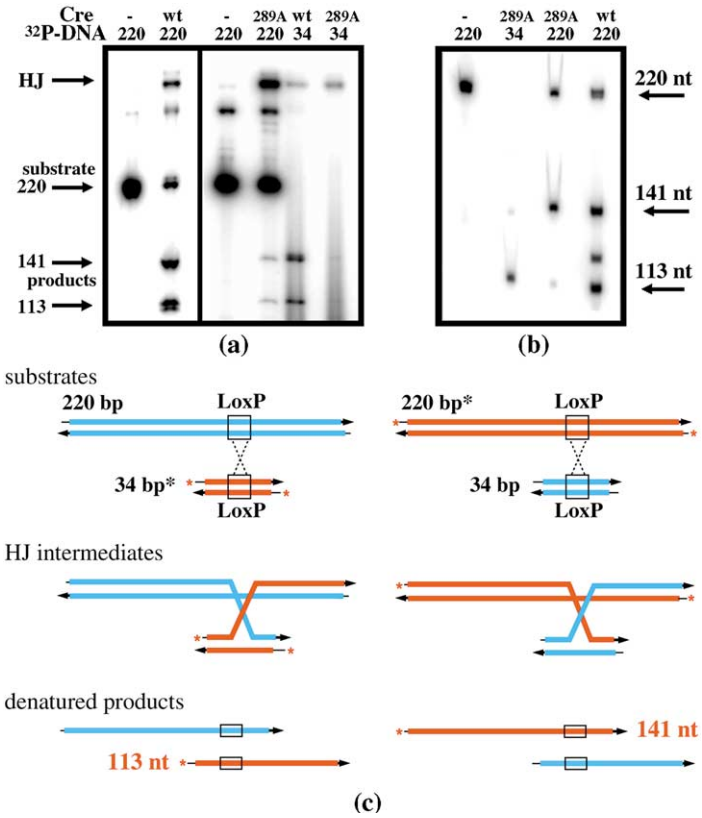

Figure 7. Generation and characterization of differentially labeled HJ intermediates to determine the identity of the exchanged strand. (a) Cre* recombination reactions with LoxP, performed and analyzed by SDS-PAGE as described in Materials and Methods. When 5′-labeled 220 bp LoxP-containing substrate (lane 1) was reacted with wild-type Cre*, two 113 bp and 141 bp product duplexes were generated, but HJ intermediates (~6%) were observed as a slower-migrating species (lane 2). Using Cre*(H289A), a larger proportion of HJ intermediates accumulated using labeled 220 bp fragment (20–40%, lane 4) or labeled 34 bp LoxP site (lane 6), compared to the reaction with Cre* (lane 5). (b) Analysis of the size of transferred strands in HJ intermediates with denaturing 8 M urea-containing gels. Gel-purified HJs from reactions of Cre with labeled 34 bp LoxP site (lane 2) or labeled 220 bp fragment (l ane 3) were denatured with formamide prior to loading. The HJ generated from the labeled 220 bp fragment contained 220 and 141 nt labeled strands (lane 3), while HJ generated from the labeled 34 bp fragment contained 113 nt and 34 nt strands. The 34 nt strand appeared on the lower part of the gel (not shown). The sizes of the denatured reactant strands (lane 1) and the product strands from the Cre* reactions (lane 4) are shown as size markers. The intermediate band in l ane 4 (~121 nt) corresponds to left arm covalent intermediate hydrolysis product that accumulates in the 16 hour Cre* reactions. The 93 nt right arm hydrolysis product was observed also (not shown) but at a very low level, suggesting that the left arm covalent intermediate selectively hydrolyzes. (c) Relationship of strand lengths and label position for upper strand transfer from left arm cleavage in HJ formation. The labeled strands are shown in red, and the unlabeled ones in blue. Note the correspondence between these product sizes and those actually found in (b).

the S1 guanine N2, N3 and C2 atoms, but NZ is not positioned to form an effective hydrogen bond with the base (Figure 6(b)). In Cre*/LoxP-G5, omit maps revealed that Lys201 reoriented, placing NZ in the base plane and in position to hydrogen bond with Ade14C N3 (Figure 6(b)). Lys86 was altered in its interaction with the scissile base, forming a hydrogen bond with the Ade14C N7 atom, rather than with O6 of the S1 guanine base, as in 1CRX and 4CRX. In 3CRX, Lys86 is rotated away and makes no contact with the scissile base.

Determination of the order of strand exchanges

The orientation of the LoxP site raised the question of whether the complex represents the HJ intermediate formed immediately after the first strand exchange, or whether A–B isomerization had occurred. The strand exchange order is intrinsic to differentiating these possibilities, as it is related to the positioning of the cleaving subunit in the first step. Initially, we investigated the order of appearance of left and right-arm covalent intermediates in LoxP recombination reactions (data not shown). In parallel experiments, Cre* was reacted with a synthetic LoxP site that was 5′-labeled with ^{32}P on either strand. Covalent intermediates, visualized as shifted DNA bands by SDS-PAGE and autoradiography, formed within seconds after mixing and accumulated up to 6% of the total radioactivity over 30 minutes. Although three to five times more left arm covalent intermediate was formed, the rates of accumulation on both strands, normalized to the maximal amount, were indistinguishable. Even at five to ten second intervals, our shortest reliable sampling rate, we could not reproducibly discern differences in time-dependent accumulation between the two sides. To reduce the speed of the reaction, we utilized a catalytically impaired mutant, Cre*(H289A), which is 60-fold slower in the overall recombination rate (data not shown) and accumulates HJ structures (Figure 7(a)). Even at the slower rate, the time-courses of the two intermediates were similar. The differences in the covalent product yields are in good agreement with similar previous work by Tribble *et al.*[36] One possible explanation for the similarities in rates is that the

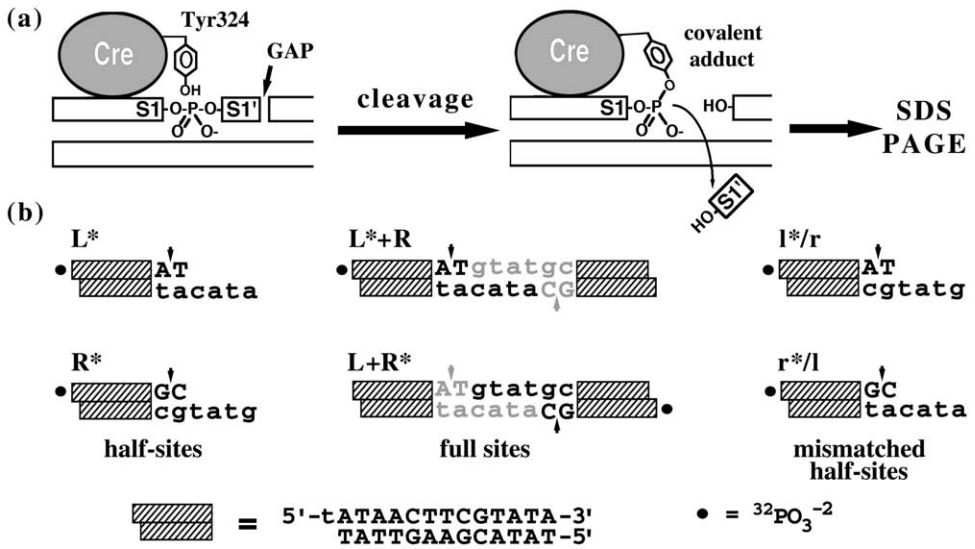

Figure 8. Scheme for suicide substrate reaction and LoxPsui substrates. (a) Suicide substrates contain a gap on the 3′ side of S1′ of the upper "suicide" strand. Cleavage releases the single S1′ nucleotide, which diffuses from the active site, preventing resolution of the covalent intermediate. The covalent intermediates were separated from unreacted material by SDS-PAGE and quantified by autoradiography. (b) LoxPsui substrates consisted of left (L) and right arm (R) LoxP half-sites. These were combined to yield the full site L + R, which was labeled either on the left arm (L* + R) or right arm (L + R*). Reactions with only labeled L (L*) or R (R*) were performed. In addition, upper and lower strands of L and R were swapped to generate the mismatched half-sites with either the labeled left arm suicide strand paired with the right arm lower strand (l*/r), or the labeled right arm suicide strand paired with the left arm lower strand (r*/l). The cleavage sites are indicated by the arrows. The labeled strands are indicated by a dot (·).

second cleavage was much faster than the first, and the complexes equilibrated within the reaction time-scale. In this case, differences in accumulation would reflect the relative stabilities of the two covalent intermediates. Alternatively, the first cleavage occurred at similar rates on both arms, but was biased to the left arm.

To determine directly which LoxP strand is exchanged first, we examined the strand lengths in HJ structures generated by recombination between differentially labeled LoxP-containing DNA fragments. In this reaction, a 220 bp LoxP-bearing DNA fragment was reacted with an unlabeled minimal 34 bp LoxP site, and reaction progress was scored by the amount of 113 bp and 141 bp fragments formed (Figure 7(a)). A slower-migrating product was observed, corresponding to the HJ intermediate, since it persisted after treatment with proteinase K. Since only one strand transfer has occurred in this species, the relative amounts of the two possible product strands would indicate the bias for the first strand exchange. By radiolabeling only one of the substrate duplexes, recombination between heterologous DNAs would yield unique lengths of 5′-labeled single strands, depending on which side of LoxP recombination was initiated from. If the 34 bp LoxP site is 5′-labeled, labeled strands of 113 nt or 141 nt are formed from left or right-arm cleavage and religation, respectively (Figure 7(c)). If the 220 bp fragment is labeled, the size order of these labeled fragments would be reversed. We purified HJ intermediates to prevent contamination from recombination products and substrate, and analyzed the constituent labeled strand lengths (Figure 7(b)). To maximize the amount of HJ produced, we utilized the Cre*(H289A) mutant, which accumulated HJ intermediates at levels up to 40%. In two separate reactions, either the 34 bp or 220 bp fragments were labeled and reacted with the unlabeled counterpart for 16 hours. The concentrated, isolated, gel-purified HJ was then electrophoresed on a denaturing 8 M urea-containing gel along with the total products of Cre* reactions as markers. In both reactions, greater than 95% of the labeled transfer product from the isolated HJ was observed in a single band, an indication that cleavage occurred predominantly on one strand (Figure 7(b)). But which arm is represented? The 141 nt exchange product was detected in HJs from the reaction containing labeled 220 bp substrate, while the 113 nt product was detected in the reaction with labeled 34 bp substrate. In both cases, the size corresponded to the expected left arm strand exchange product (Figure 7(c)). Thus, Cre initiates strand exchange almost uniformly on the left arm of LoxP by cleaving the upper strand first.

Kinetic mechanism of strand discrimination

How does Cre differentiate the two arms? One hypothesis is that the left arm reacts more rapidly, partitioning the reaction at the first cleavage. To investigate reactivity differences between the left and right arms of LoxP, we utilized half-site

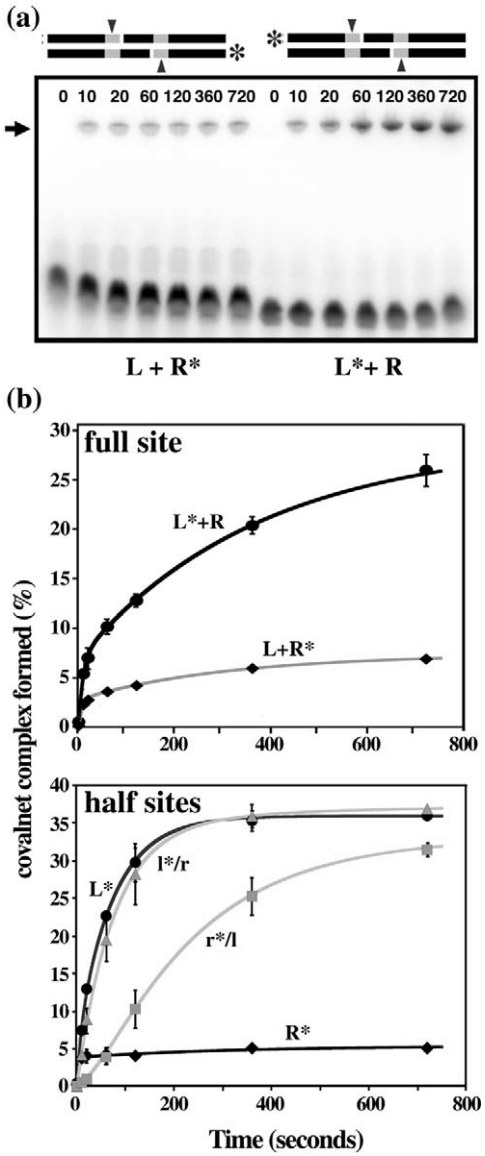

Figure 9. Time-course assays of covalent modification by half-site suicide substrates. (a) An example of SDS-PAGE analysis of time-course reactions with L + R* (lanes 2–7) or L* + R (lanes 9–14). The covalent product bands are indicated by the arrow. (b) Plotted results from LoxPsui time-courses. The individual time-points were normalized to the mean extents from Table 3 and then replotted with the theoretical curves generated using the averaged parameters from Table 3. The mean renormalized values for three independent determinations are indicated for L* + R (black filled circles) and L + R* (black filled diamonds). Error bars correspond to ±SD. The curves were generated from simulation of the sequential model using the Table 3 parameters. (c) Plotted results for half-sites L* (black filled circles), R* (black filled diamonds), l*/r (gray filled triangles), and r*/l (gray filled rectangles). These data were processed as in (b).

suicide substrates. Half-site suicide substrates irreversibly trap covalent complexes of λ integrase,[38] FLP,[39,40] XerCD[41] and Cre[26,42] after the first cleavage event. The essential characteristic of these substrates is a gap in the cleavage strand one to three nucleotides 3′ to the scissile bond. When the 3′-phosphotyrosyl bond is formed, the remaining nucleotides 5′ to the gap diffuse from the active site, preventing resolution or equilibration of the covalent intermediate in either direction (Figure 8(a)). The trapped covalent complex is visualized by denaturing SDS-PAGE as a slower-moving labeled band (Figure 9(a)). As a result, the first cleavage can be separated from the overall reaction. A larger signal is provided through accumulation of the irreversible reaction products, in contrast to the low abundance of covalent intermediates in intact LoxP reactions. Potential perturbations from these gaps are a complicating factor and preclude direct comparisons with LoxP. However, since such influences might be expected to be similar for both sides, the reactivity differences due to different scissile nucleotides can still be inferred.

We utilized LoxPsui, a LoxP version of the symmetric LoxA suicide substrate.[26] LoxPsui is composed of two duplexes, L and R, for the left and right arm half-sites, respectively, which reconstitute a complete LoxP sequence upon annealing. The suicide gap is 3′ to the S1′ nucleotide (Figure 8(a)). The suicide strand of each arm was labeled in parallel experiments giving either L* + R or L + R* (Figure 8(b)) and reacted with Cre. Reactions were performed with 1:1 Cre to Lox half-site at concentrations that were 200–400-fold higher than that required to give half-maximal recombination. Reaction rates have no concentration-dependence from 12–50 μM, suggesting that the rate-limiting steps do not occur during the aggregation steps of complex assembly (S.S.M. & E.P.B., unpublished results). This procedure allowed direct measurement of arm bias in the first cleavage from the ratios of the fraction of covalent product formed (extents) in each reaction, while relative rates were ascertained from kinetic modeling of the time-courses.

While Cre preferred to react with the left arm of LoxPsui, the cleavage rates were similar for both sides (Table 3). The extent of the L* + R reaction, obtained by curve fitting (see below), was fourfold more than L + R*, 28% *versus* 7%, indicating a significant but not complete preference. However, the unimolecular rates of cleavage, in which the observed rates were normalized to the extents, were similar for both substrates (Table 3). To quantify the reaction rates, we fit the time-course data to an arbitrary kinetic model. Since assembly was not rate-limiting and DNA was saturated with protein, the reaction time-courses were expected to reflect single turnover rates for the complex. However, the time-evolution of covalent product was fit less well by a two-parameter, single-exponential model, $R^2 = 0.92 - 0.94$, than by a four-parameter, two-exponential model, $R^2 > 0.99$ (see the footnotes to Table 3A for details of the model). In Figure 9(b), the time-course data are plotted with theoretical curves generated from

Table 3. Kinetic parameters for Cre covalent modification by LoxP-derived suicide substrates

Substrate	f	A	k_1 (10^3 s^{-1})	B	k_2 (10^3 s^{-1})
A. *Two-exponential model*[a]					
L* + R	28 ± 5	0.78 ± 0.07	3.0 ± 1.0	0.22 ± 0.07	120
L + R*	7 ± 1	0.61 ± 0.10	3.3 ± 0.6	0.39 ± 0.10	120
L*	36 ± 7	0.81 ± 0.08	12 ± 4	0.19 ± 0.08	120
R*	5.2 ± 0.4	0.24 ± 0.21	3.0 ± 0.0	0.76 ± 0.21	180 ± 72
l*/r	37 ± 7	1.0	12.5		
		(0.96 ± 0.04)	(12 ± 4)	(0.045 ± 0.04)	(120)
r*/l	NA				
B. *Sequential mechanism model*[b]					
L* + R	29 ± 7	0.74 ± 0.07	2.7 ± 0.2	0.26 ± 0.07	100 ± 30
R* + L	7 ± 1	0.61 ± 0.04	3.2 ± 0.9	0.39 ± 0.05	123 ± 7
L*	36 ± 7	0.74 ± 0.03	13.5 ± 2.3	0.26 ± 0.03	117 ± 12
R*	5.4 ± 0.4	0.29 ± 0.16	3.1 ± 0.6	0.71 ± 0.16	280 ± 40
l*/r	37 ± 7	1.0	12.5		
		(0.06 ± 0.11)	(3.0 ± 0.0)	(0.94 ± 0.11)	(13 ± 2)
r*/l	33 ± 5	0.78 ± 0.09	19 ± 5	0.22 ± 0.09	4.7 ± 0.7

The parameters were fairly robust, in that variation of rate constants by more than 20% or f, A and B by more than 10% could not be fully compensated by refitting other parameters.

[a] Data from each of three independent experiments, described in Figure 8 were fit to the bi-exponential model, $P = f(Ae^{-k_1 t} + Be^{-k_2 t})$. There were four fit parameters: f, the overall extent of the reaction at $t = \infty$ as a percentage of substrate conversion; first-order rate constants for each exponential process k_1 and k_2; and A, the fractional amplitude of process 1. The amplitude of process 2 was obtained from $B = 1 - A$. Manual optimization was performed to maximize the R^2 value (>0.99) for A, k_1 and k_2, and to minimize the residual $\Sigma|\text{exp} - \text{calc}|/\Sigma\text{exp}$ for f, at each experimental time-point.The parameters are expressed as the average value ± the standard deviation of those obtained from the individual experiments.

[b] Data from three independent experiments were fit by comparison of the experimental data to a graphic simulation for the minimal sequential mechanism: Model : $a \xrightarrow{k_1} b \xrightarrow{k_2} c$. The concentration of species a, b, and c were calculated at each step adding the change in concentration over the interval Δt to the concentration from the previous step. For example $[\text{b}]_{t+\Delta t} = [\text{b}]_t + \Delta t(k_1[\text{a}]_t - k_2[\text{b}]_t)$, where $\Delta t = 1$ second. There were four fit parameters: f, the percentage of substrate conversion, i.e. the amount of c at $t = \infty$; first-order rate constants for each step k_1 and k_2; and A, the fractional amount of a at $t = 0$. The fractional amount of b at $t = 0$ was obtained from $B = 1 - A$. Manual optimization was performed to maximize the R^2 value (>0.99) for A, k_1 and k_2, and minimize the residual $\sigma|\text{exp} - \text{calc}|/\sigma\text{exp}$ for f at each experimental time-point.The parameters are expressed as the average value ± the standard deviation of those obtained from the individual experiments. Simulation time-steps of less than one second had no discernible effect on the values of the optimized parameters.

the averaged parameters. The rate parameters in Table 3 are given for comparison, since their correspondence to actual mechanistic steps is unknown. Nonetheless, it is clear that, although Cre distinguishes the arms of LoxPsui, it is not through differences in cleavage rates (upper panel, Figure 9(b), Table 3). In spite of the bias in extents, L* + R and L + R* had similar rate parameters, indicating that the choice for strand cleavage is not made at the rate-limiting step for cleavage.

We were concerned that the less than quantitative strand preference and low extents were due to substrate heterogeneity. The four constituent oligonucleotides of LoxPsui can, due to homology, assemble into four half-site duplexes, L, R, and the mis-paired sites r/l, and l/r (Figure 8(b)): 64 possible synapses of uncertain reactivity result, half of which could yield a signal. To assess the possible contributions of different complexes, isolated half-sites R* or L*, or the heterologous half-sites l*/r or r*/l (Figure 8(b)), were reacted separately with stoichiometric Cre. In spite of the lack of homology in the 4 nt internal overhangs, covalent products accumulated to high levels, suggesting that assembly of an active complex does not require prior annealing of half-sites (lower panel, Figure 9(b)). This ability to cleave non-complementary half-site substrates has been documented for Cre,[42] and underscores the primary role of protein- protein interactions in complex stabilization. Each of the half-sites had a unique reactivity profile that differed from that of LoxPsui (Table 3), suggesting that the dominant contribution in the LoxPsui reactions comes from the "full site" complex. The extents are actually twice as large as those in the LoxPsui experiments, since half of the substrates would be bound by the non-cleaving subunit upon trapping. Even more marked than the left arm preference in LoxPsui, L* reacted to a sevenfold greater extent than R*, 36% *versus* 5% (Table 3). This left arm bias in the absence of a choice for which sequence to cleave, along with the distinctive reaction kinetics for each half-site, suggests an additional specificity mechanism that does not rely on Cre asymmetry, but instead on the ability to achieve an active conformation.

While the L* and R* half-site time-courses were well described by the two-exponential model, the mis-paired substrates l*/r and r*/l behaved differently (Table 3). The l*/r time-course was well fit by a single-exponential equation. More interestingly, although r*/l contained the right arm sequence on the suicide strand, it reacted to a similar high extent, 33%, as the l*/r substrate, 37%. Furthermore, r*/l had an unusual time-course and could not be fit to either model (lower panel, Figure 9(b)). At first, product formed slowly,

and after a 10–20 second lag, a second phase was observed. The minimal scheme that described this behavior adequately incorporated two sequential first-order irreversible reactions, $a \rightarrow b \rightarrow c$ (see Table 3(b) footnotes for details of this model). The lagging behavior occurs when the second step is similar in time-scale to the first, and initially b is less populated. When the first step is much slower than the second, this mechanism is compatible also with the two-exponential model described above. All of the data were refit to this sequential model and yielded very similar rate constants (Table 3B).

Discussion

In Int family mediated recombination, there is an intimate association between subunit arrangement, site alignment, and strand exchange order. The cleaving and non-cleaving subunits in the Cre dimers present distinct inter-subunit interfaces that orient them anti-parallel in the synapse (Figure 1(d)). Consequently, anti-parallel orientation of DNA substrates, common in Int recombination,[43–46] is required to place homologous strands in the cleaving subunits' active sites (Figures 1(d) and 2(a)). While this does not enforce a particular strand exchange order, the minimal hypothesis is that the arm of DNA substrate bound by a cleaving subunit in the precleavage complex would be the location of the first strand exchange. In Cre-LoxP recombination, we propose that this anti-parallel alignment is achieved by preferential association of the cleaving subunit with the LoxP left arm upon dimer formation. While Cre can misalign LoxP sites *in vivo* under certain conditions,[47] experiments *in vitro* done by us (Figure 7(b)) and by others[47] indicate that this occurs rarely. In the data presented above, we demonstrated that Cre initiates recombination predominantly on the left arm of LoxP. This strand selectivity is achieved not by cleaving the left arm faster but through biased partitioning of the complexes towards the left arm-cleaving form. DNA–protein interactions with the 8 bp spacer, the only asymmetric element of LoxP, must play a role in discriminating between the two Lox orientations. The Cre-LoxP HJ structure revealed novel interactions between the cleaving subunit and the adenine residue that is 5′ to the cleavage site on the left side of the 8 bp spacer, that are absent or altered when this nucleotide is guanine. We suggest that these interactions stabilize the left arm-bound arrangement. The left arm biases towards forming covalent intermediates[36] and resolving HJs[35] also point to a thermodynamic preference for interaction with the left arm, presumably due to the spacer composition. Spacer sequences modulate site alignment[48] and HJ resolution directionality[49] in other, but not all,[50] Int reactions. The elegance of this mechanism is in contrast to XerCD/dif-mediated chromosome segregation, in which strand exchange order and site alignment are assured by two distinct recombinase proteins that interact specifically with their unique half-sites.[51]

Cre initiates recombination on the left arm of LoxP

From analysis of single-strand transfer products in the first HJ intermediate, we determined that the upper strand had been cleaved and transferred greater than 95% of the time. This conclusion is in direct contrast to the generally stated view that exchange is initiated from the right arm.[20,21,35] However, this outcome was corroborated by previous experiments in which the strand composition of HJs generated from intramolecular recombination was characterized.[16] While it was concluded that one set of strands had been exchanged, the particular strand was not identified explicitly. Our analysis of these data, based on the substrate sequence,[52] indicated that these HJs were also formed from left arm cleavage.

In LoxP reactions, left arm covalent intermediates were formed preferentially over right arm intermediates,[36] but their order of appearance was indistinguishable because of low levels and rapid rates. The simultaneous accumulation of both intermediates can be accounted for in two ways: (1) the first cleavages occur on either arm but at the same rates; or (2) the rate-limiting step for recombination occurs after second strand cleavage. The latter view is supported by our observations that both covalent intermediates accumulate more rapidly than the products (data not shown).

Suicide substrate reactivities support the idea that Cre does not discriminate by cleaving the left arm faster. Although Cre cleaved the left arm to a fourfold greater extent, the similar kinetic parameters for the $L^* + R$ and $L + R^*$ LoxPsui reactions suggest the same rate-limiting steps (Table 3). The comparable cleavage rates but differing extents are explained most easily by preferentially partitioning complexes towards the left arm-cleaving form prior to the rate-limiting step for covalent bond formation, consistent with the idea of arm-biased complex assembly. However, there appear to be different limitations on homogeneous half-site reactivities. While differential half-site reactivity was noted previously,[42] our quantitative results showed an even more marked product bias for L^*, six- to sevenfold over R^*. Why is the R^* extent still low, even when the cleaving subunit must react with that arm? While partitioning explains the right arm suppression in LoxPsui reactions, it does not explain the low reactivity of R^*. This silencing of the R^*-containing complex suggests another specificity mechanism, perhaps through adopting an alternative, inactive conformation(s). One explanation for rapidity but small extent of R^* reactions might be that, while the majority of complexes were inactive, there was a

small amount of a "fast-reacting" conformer with lower specificity. This sort of behavior, partitioning between faster less-specific and slower more-specific complexes, affected in part by substrate interactions, is remarkably reminiscent of the behavior of RNA polymerase towards nucleotide misincorporation.[53,54] What element of the right arm sequence suppresses activity? Clearly, the presence of a non-adenosine nucleotide at S1 is insufficient. While l*/r reacted similarly to L*, r*/l reacted to the same extent, sevenfold greater than R*. Energetics of deforming the scissile base-pairs may influence the distribution of alternative states. This idea is supported by the enhancement of reactivity accompanying mispairing of S1 and S1′ in r*/l, hinting that an inactive state is not formed, or is reactivated if the scissile base-pair can be disrupted. Although inspection of existing structures does not indicate such deformations, they may occur transiently.

The LoxPsui reactions required two exponential terms to describe the time course data adequately. The lag in r*/l product evolution was indicative of a sequential reaction mechanism but, for the other substrates, it cannot be discerned whether the two processes are sequential or parallel. The complex kinetic behavior of this multi-step reaction is not surprising but it defines the minimum analysis required for characterization. Time-courses are the most informative, whereas endpoints, initial rates, or half-times may be misleading in describing substrate reactivities. For example, the least reactive right arm substrates had faster initial rates. On the other hand, r*/l, had the slowest initial rate but reacted to levels similar to those of the left arm substrates.

Determinants of strand exchange order in LoxP

Although the 8 bp spacer sequence is not required absolutely for function, the scissile bases play a key role in governing strand reactivity in both recombination[21] and HJ resolution.[35] Systematic mutagenesis identified the S1 base as the key determinant of strand selection[35] in HJ resolution. Indeed, swapping S1 adenine and guanine nearly completely reversed the left arm bias. For recombination overall, specific 8 bp spacer substitutions, particularly at S1 and S1′, had detrimental effects on reaction yield and progress.[21] The five- to tenfold reduction in products, and accumulation of HJ intermediates, were attributed to blockage of first strand cleavage, and inhibition of HJ resolution, respectively. However, this behavior is consistent also with the idea that the scissile nucleotide identity orients the LoxP site in the synapse. Less preferential interaction with the left arm would lead to a preponderance of misaligned complexes that either resolve to the reactants, or accumulate as HJs. HJ accumulation was observed in reactions between Lox sites with heterologous spacer sequences,[21] which is analogous to the misaligned case. On the other hand, disrupting the cleaving subunit interaction with the right arm would be expected to have little influence on the initial partitioning of the complex and alignment. Indeed, right arm spacer substitutions had more variable and, generally, less detrimental, effects.[21]

The persistence of a fourfold left arm preference in suicide substrate reactivity, covalent intermediate formation,[36] and HJ cleavage,[35] indicate similar contributing factors, and suggest that the overall energy difference between left arm and right arm-cleaving complexes is ~800 cal/mol (1 cal = 4.184 J). The difference between Lys201 interactions with a scissile adenine or guanine residue could account for this value reasonably well. It does not account completely for the over 20-fold left arm preference in the first LoxP strand exchange, indicating additional contributions from the 8 bp spacer. The greater geometric constraints from the uninterrupted strands in LoxP could amplify the bias. Indeed, the maximal amount of covalent intermediates was achieved within 60 seconds with LoxP (data not shown) compared to greater than 15 minutes for LoxPsui. Additionally, the energetics of structural distortions in the 8 bp spacer, which is asymmetrically bent in the pre-cleavage complex, may favor one isomer.[36] In LoxP, this would involve unstacking a CpA/TpG base-step at positions 19 and 20 (Figure 5(c)), which has one of the lowest stacking energies and is deformed most easily.[55,56] In the suicide substrates, the suicide gap occurs at this same base-step, which may release some of the constraints and lead to the reduced preference. These constraints would be absent from the HJ complexes also, since the DNA is already bent. Such "indirect" readout of 8 bp orientation is attractive, given the paucity of Cre-Lox contacts in the spacer region. Indeed, bends introduced by 2 nt insertions into different positions of the spacer modulated strand selectivity, perhaps by assembling one complex orientation preferentially.[36]

Implications of the Cre-LoxP HJ structure

We applied the CRX crystallization conditions to obtain Cre/LoxP complex crystals, to address the role of the 8 bp spacer in directing strand exchange order and site alignment. Crystals we have grown in other conditions have thus far diffracted poorly, or contained an unusual trimeric synapse structure.[37] For Cre*/LoxP, we expected one of several configurations: (1) a homogeneous complex consisting of a single intermediate; (2) crystals containing a mixture of intermediates; and (3) homogeneous crystals of different intermediates from the same drop or under different conditions. Nine crystals grown from Cre* and LoxP or variant LoxP sites analyzed thus far contained similar Cre-HJ complexes. In these complexes, the HJ DNA was characterized by two conformations. In the dominant conformation, disrupted base-pairing at the junction and disorder of the bridging nucleotides suggests that these interactions may

contribute little to HJ complex energetics, and thus, spacer homology may be sensed less effectively in this intermediate. This might explain why recombination between mismatched spacers occurs or leads to the accumulation of HJ structures[21,22] rather than precleavage complexes.

In addition to providing the first view of a Cre-LoxP HJ complex, we have developed a facile method for visualization of variant complexes without potential interference by unrelated Cre or Lox substitutions. Why did this form crystallize? In hindsight, the manipulations utilized for 2CRX and 3CRX to obtain the HJ intermediate were not required. The spontaneous crystallization of both 2CRX and Cre-LoxP HJ complexes suggest that this intermediate is the lowest-energy form. Our kinetic studies indicate that the ten minutes or longer pre-incubation prior to mixing crystallization drops was sufficient for a sizable fraction of complexes to react, whereas crystallization conditions would preclude the reaction (data not shown). Cre binding to "pre-formed" LoxP HJs seems unlikely, given the expected junction heterogeneity due to branch migration. However, the possibility of Cre-induced assembly of these species cannot be ruled out at this time.

The relevance of LoxP alignment in Cre*/LoxP is underscored by the left arm bias in HJ resolution.[35] Given that the various reaction intermediates are accommodated in this crystal form, we feel that the dominance of the left arm-bound HJ complex reflects energetic differences between isomers rather than selective crystallization. Although the electron density suggests an obvious preference (Figure 3), presently we are experimentally limited in quantifying the fraction of each orientation in the crystals. In the future, we will determine this through the use of asymmetric 5-bromouridine substitutions.

S1 base-dependent positioning of the 198–208 loop is a good candidate for directing the cleaving subunit orientation in the precleavage and HJ complexes, since it corresponds to a major structural difference between the A and B conformations (Figure 2(b)). The contacts between Lys86 and Lys201 and the S1 adenine base, which are completely absent from the non-cleaving subunit, and different in the CRX structures, apparently play a key role. The Lys201–Ade14 interaction appears sufficient to reposition this loop in Cre*/LoxP complexes relative to the 2CRX and 3CRX HJ complexes (Figure 6(a)). In the 2CRX and 3CRX HJ complexes,[27] which lack the S1 adenine residue, neither lysine residue contacts the scissile base. In 1CRX and 4CRX, these residues contact the S1 guanine base but only *via* a single hydrogen bond with Lys86. Differential recognition of adenine over guanine by these residues was corroborated by the effects of the Lys86-Ala mutation, which reduced the HJ resolution bias by 60%.[35] However, since Lys86 forms hydrogen bonds with both adenine and guanine, its role may be to position Ade14 for sequence-specific recognition by Lys201.

The reorientations of the 198–208 loop and lysine side-chains indicate strongly that Cre can discriminate between left and right S1 bases. From inspection of the electronic structure and properties of the bases,[57] such interactions do not appear to allow differentiation between adenine and guanine *per se*, but the rearrangements imply that in the context of the Cre active site, they can distinguish the two. To investigate whether this is an unusual mode of discrimination, we analyzed interactions with guanine and adenine N3 and N7 atoms in a limited database of 96 protein–DNA structures derived from the NDB.[58] Phenomenologically, the N3 interactions appear to offer greater discrimination. Adenine was recognized by polar contacts to N3 23% of the time, 18% of these by lysine NZ atoms. However, guanine was recognized rarely at N3 (3%) and never by a lysine residue. With regards to N7, both adenine and guanine are commonly recognized at this position (44–53%) but, while guanine recognition was often mediated by lysine NZ (20%), no similar lysine–adenine hydrogen bond has been observed.

Unfortunately, this role of Lys201 cannot be probed directly by mutagenesis, as it is essential for recombination activity.[25] In Int enzymes, it is believed to act as a general acid that protonates the leaving 5′-hydroxyl group.[59] We propose that Lys201 also specifically recognizes the scissile nucleotide, which is not incompatible with a catalytic role, as the side-chain is positioned appropriately for both tasks. Lys201 in Cre is apparently critical, participating in catalysis, proper alignment of Lox sites, and setting the position of the conformational switch. Whether this role is generalizable remains to be determined. In the FLP-FRT structure,[30] Lys223 also contacts the scissile base in the minor groove, but is positioned similarly in all four subunits, implying that any role in strand selectivity occurs *via* a mechanism different from that proposed here.

New views of the Cre-LoxP recombination mechanism

The observations made here suggest some amendments and pose additional questions regarding the Cre-Lox recombination mechanism. First, strand exchanges proceed in a specified order, starting on the left arm. Second, the lowest-energy state of the complex appears to be the first HJ intermediate. Rather than a different directional bias for the LoxP duplexes and HJs,[35] which might be expected to promote the reaction, the equilibrium in Cre-LoxP interaction favors intermediates in the first half of the pathway. The preference for the left arm HJ isomer is hard to reconcile mechanistically because product complex dissociation would be required to drive the reaction to completion. Flux through the pathway, albeit less than optimal, would be promoted by differential left arm-cleavage bias for the two substrates, with >95% preference for LoxP but only

80% for the HJ, allowing 20% to react further. This property suggests that Cre's velocity could be improved severalfold if the conformational equilibrium was skewed towards the second HJ structure without affecting the precleavage partitioning. Third, Cre*/LoxP and Cre*/LoxP-G5 likely reflect the actual HJ intermediate structure, in that the positioning of the 198–208 loop is maintained during each set of exchanges. Given the S1 guanine base, 1CRX and 4CRX may be more representative of the intermediates in the second strand exchange cycle (Figure 1(b)). Finally, the complex kinetic behaviors of half-site substrates indicate that Cre may possess yet-unidentified off-pathway conformations that modulate specificity. The structural characterization of these conformations awaits crystallographic analysis of the inactive complexes.

Materials and Methods

Plasmids, proteins, oligonucleotides, and ^{32}P-labeling

His-tagged Cre (Cre*) was purified to homogeneity from induced lysates of *Escherichia coli* BL21(∷DE3, pET28b–His$_6$Cre) by Ni-chelate and cation-exchange chromatographies, and treated as described.[37] All preparations were free of measurable endo- and exonuclease activities under assay conditions. Mutant Cre*(H289A) was created by the method of Kunkel,[60] and purified similarly from lysates of RR1(∷DE3, F′tet) infected with f1-packaged pET28b–His$_6$Cre(H298A). The properties of this variant will be described elsewhere.

The LoxP-containing substrate plasmid, pLox38(+), was created by ligation of a synthetic LoxP duplex into *Sna*BI/*Eco*RV-digested pLITMUS38 (NEB). The 220 bp fragment for assays was prepared by digestion of 100 μg of Quiagen-purified plasmid with *Ban*I and *Stu*I (NEB) followed by purification on an agarose gel (Nusieve, 3%) with Glassmilk (BIO101).

Synthetic Lox oligonucleotides for crystallization and activity assays were obtained from DNA Express (1 μmol scale, RPC-purified), resuspended in 100 μl of TE (10 mM Tris–HCl (pH 8.0), 1 mM EDTA), at concentrations of 0.5 to 2 mM. Concentrations were determined by measuring UV absorbance.[61]

DNA fragments (0.1–10 pmol) were 5′-end labeled with 2 μM [γ-^{32}P]ATP (3000 Ci/mmol; Amersham) using 100 units/ml of bacteriophage T4 polynucleotide kinase (NEB) in the supplied buffer followed by purification using a Quiagen Nucleotide Purification Kit.

Crystallization and data collection

LoxP and LoxP-G5 (Figure 1(a)) duplexes were generated by annealing corresponding pairs of 34 nt oligonucleotides at 500 μM in TE with 100 mM NaCl; heating to 95 °C and cooling to 35 °C over 30 minutes using a thermocycler (MJ Research Minicycler). The duplexes were mixed with Cre (35–45 mg/ml), left for ten minutes, and diluted to 40 μM duplex and 50 μM protein. The same crystals were obtained when protein and DNA were mixed at concentrations as low as 10 μM and 12.5 μM, respectively. Protein–DNA solution (3 μλ) was mixed in 1:1 ratio with well solution and inverted over a 500 μl reservoir containing 17–30% MPD, 12.5–100 mM sodium acetate (pH 5.0–5.5) or Na–Mes (pH 5.75), and 40–60 mM NaCl, at 21 °C. Crystals appeared within 3–21 days, reaching maximum size (100 μm × 300 μm × 400 μm) with one month. Crystals grown in greater than 22% MPD were flashed-cooled in liquid nitrogen prior to data collection. Those grown in 17–22% MPD were equilibrated to 24% MPD by vapor-diffusion for 1 week. Crystals were isomorphous to those described by van Duyne and co-workers,[27] space group $C222_1$, with unit cell dimensions $a \sim 108$ Å, $b \sim 122$ Å, and $c \sim 180$ Å, and diffracted to 2.7–3.0 Å using Cu Kα radiation and 2.2–2.5 Å at SSRL. The Cre*/LoxP data were obtained on the home source using a RAXIS-IV image plate (Molecular Structures Corporation). The Cre*/LoxP-G5 data were collected at SSRL beamline 9-2 using an CCD camera (ADSC, special thanks to James Endrizzi, James Holton and Tom Alber). Diffraction data were indexed and processed with MOSFLM[62,63] and scaled using SCALEA[64] assisted by the Wedger program from the Elves package[65] (special thanks to James Holton and Tom Alber, U.C. Berkeley). Data collection statistics for Cre*/LoxP-G5 are given in Table 2.

X-ray structure solution and refinement

Refinement was carried out using TNT.[66] The geometry restraints of Engh & Huber were imposed.[67] The atomic *B*-factors were restrained using a BCORRELS library[68] modified by us to contain information from high-resolution DNA structures. Of the available data, 5% were set aside prior to refinement for the free *R* calculation. The free R_{factor}[69] was used to optimize the number of refinement cycles and other variables throughout the procedure.

First, starting models consisting of either 1CRX, 2CRX, 3CRX or 4CRX in which residues B324–333 of the *B* subunit and the 8 bp spacer nucleotides had been omitted, were refined in parallel in the following manner.[70] Initial rigid-body refinements were performed using first the entire asymmetric unit at 4 Å, then individual subunits at 4 Å, and finally elements of secondary structure at 3.3 Å. All data were used and observed structure factors were scaled to the model using the parameters *B* 0.0, KSOL 0.9, BSOL 150.0. All *B*-factors were set initially to the Wilson *B* and in between positional refinements, an overall *B*-factor for each rigid body was refined. Subsequently, positional refinement, and then positional with restrained *B*-factor refinements were performed at 2.7 Å resolution. The resulting structures had R_{free} values ranging from 33–37% but yielded nearly identical $2F_o - F_c$ and $F_o - F_c$ maps. The best fit was obtained using 1CRX but no single model accounted for the electron density satisfactorily. The best overall agreement was achieved by combining the initially refined models of 1CRX (A20-A341, B20-B298) and 4CRX (B299-326, B333-B341, DNA). The hybrid model was refined at 2.7 Å resolution to yield the "partially refined Cre*/LoxP model", $R_{\text{work}} = 27.6\%$, $R_{\text{free}} = 32.8\%$.

An independent but identical approach was taken for the 2.2 Å data set, except that additional positional and *B*-refinements were carried out at 2.2 Å. A similar hybrid model was constructed with sequence adjustments incorporated, and further refined. Difference density indicated a unique position for the junction region of the DNA, which was incorporated at this stage ($R = 31\%$). To guide model building, $F_o - F_c$ difference maps were

generated from models in which 30 residue segments were omitted prior to refinement. Water molecules were added by inspection of difference maps. In the later rounds of model building and refinement, only the 5–2.2 Å data and no solvent model were used because of poor low-resolution scaling. Later in the process, residual density corresponding to the 3CRX junction DNA was observed and an alternate model for only residues C16–19 and D16–19 was incorporated. The occupancies of the two conformations were set to 0.5 and the *B*-factors for each conformer were combined. Refinement of group occupancy and *B*-factors limited to these atoms gave a ratio of 0.73 and 0.27 for the novel and 3CRX-like conformations, respectively. The occupancies were fixed to these values and model building proceeded as before.

The final model contains an asymmetric Cre dimer, a complete LoxP DNA site, and 237 solvent atoms. The final refinement statistics are given in Table 2. Of the protein and DNA atoms, 16% and 4% have *B*-factors of 80–99 $Å^2$ and 100 $Å^2$, respectively, but are included in the model because there is corresponding electron density visible in omit maps. Density for the hinge residues B329–332 was completely absent and these atoms were not included in the final model. This region is absent from the HJ-containing 2CRX model and the precleavage analog 5CRX. There was no apparent electron density for the entirety of residues A191, A215, A312 and A316, and the side-chains of A182, A221, A222, A234, A302, A307, A311, A317, A323, B22, B28, and B317. The occupancies of these atoms are set to zero and their inclusion in the model is based primarily on geometric constraints. In addition, residues A184–206, A275–278, A303–319, and B300–334 have interrupted or blurry electron density. The combination of density and geometric constraints is adequate to approximate their positions but the model has less precision in these regions.

Structural comparisons and difference map calculations

Superpositions, structural comparisons, rmsd calculations, and other parameters for model analysis were performed with the aid of EDPDB.[71] Unless specified otherwise, structural shifts were measured as the rmsd of common sets of atoms between structures. For protein comparisons, the following sets of main-chain atoms were used to account for the different omitted residues in each structure: 1CRX, A20–341, B20–328, and B333–341; 2CRX, main A19–198, A210–275, A283–341, B19–198, B207–326, and B333–341; 3CRX, A19–341, B20–328, and B333–341; and 4CRX, A20–341, B20–328, and B333–341. For DNA comparisons, only the atoms of the sugar–phosphate backbone were used.

Except where noted, refined-omit difference maps were calculated by first removing atoms from the final structure and then refining with 50 cycles of conjugate directions minimization using TNT to reduce the contribution of model bias. The refinement weights were adjusted to maintain geometry and *B*-correlation deviations. Figures were created with MOLSCRIPT[72] followed by rendering with Raster 3-D.[73]

DNA cleavage assays with half-site suicide substrates

The duplex half-site substrates consisted of 16 nt strands, 5′-tATAACTTCGTATAGC (16A) or 5′-tATAACTTCGTATAAT (16B), annealed with 19 nt strands, 5′-ATACATATATCGAAGTTAT (19A) or 5′-GTATGCATATCGAAGTTAT (19B), at room temperature after heating to 70 °C for 30 seconds (the lower case *t* indicates a non-LoxP 5′-Thy extension). The composition of each half site was L = 16B + 19A; R = 16A + 19B; l/r = 16B + 19B; and r/l = 16A + 19A. In all experiments, one 16 nt oligonucleotide was end-labeled and purified as described above. The LoxPsui full site substrates consisted of equimolar L and R (L + R), annealed with trace labeled 16B (L^* + R) or 16A (L + R^*). The half-site reactions contained twice the concentration of the unlabeled half-site annealed with trace 16B (L^* or l^*/r) or 16A (R^* or r^*/l). Reactions contained 25 μM LoxPsui, or 50 μM homogeneous half-site substrate, 50 μM Cre monomer, in optimized Cre reaction buffer.[37] The labeled cleavage strand oligonucleotide, 20 nM, was annealed with 100 μM each unlabeled substrate strand in TE, by heating briefly at 70 °C and allowing to cool in air to room temperature. Reactions were initiated by combining 30 μl of the labeled DNA mixture with 30 μl of 100 μM Cre in a twofold concentrated reaction buffer solution at 21 °C. Time-point samples of 8 μl were removed from the reaction at 10, 20, 60, 120, 360 and 720 seconds, and quenched in 8 μl of gel loading buffer (TE containing 2% (w/v) SDS, 30% (v/v) glycerol, and 0.01 % (w/v) bromophenol blue). Aliquots (7 μl) were loaded onto 10% (w/v) polyacrylamide gels, (20:1 (w/w) bis acrylamide to acrylamide) containing Lammelli SDS-PAGE buffers[74] and were electrophoresed at 200 V for 2–2.5 hours. Vacuum-dried gels (80 °C for one hour) were used to expose an image plate (Fuji BAS-MS 2040), which was then scanned with an image plate reader (Molecular Dynamics Storm 860). The data were visualized and analyzed with ImageQuant (Molecular Dynamics). Kinetic constants were determined as described in the legend to Table 3.

Determination of strand exchange order

Synthetic LoxP duplex was obtained by annealing the component strands at 12 μM as described for crystallization experiments. Recombination reactions were carried out for 16 hours at 21 °C in optimized reaction buffer, 4.8 μM Cre^* or Cre^*(H289A), 1.2 μM 34 bp LoxP fragment and approximately 5–10 nM 220 bp *Ban*I/*Stu*I fragment. Either substrate was ^{32}P-labeled on the 5′-ends. Reactions were quenched with 1% SDS, digested with proteinase K (0.5 mg/ml, 30 minutes at 37 °C) and analyzed on SDS/8% polyacrylamide gels.[74] For HJ preparative reactions, 400 μl of a reaction of Cre^*(H289A) with labeled LoxP, or 50 μl of reaction of Cre^*(H289A) with labeled 220 bp fragment were generated, quenched, proteinase K-digested, and loaded onto 1–3 lanes of a 1.5 mm preparative SDS/polyacrylamide gel. HJ-containing bands were located by autoradiography and cut out of the gel. The gel slices were soaked for 12–48 hours at room temperature in 1–2 ml of 100 mM ammonium acetate, 0.5 mM Na–EDTA (pH 7.0), supplemented with 1 μg of carrier oligonucleotide DNA. The extracts were concentrated *in vacuo* and the labeled HJ was precipitated with two volumes of ethanol. The dried pellets were resuspended in 90% (v/v) formamide in TE buffer, heat-denatured at 95 °C, cooled on ice immediately, and electrophoresed on a 10% polyacrylamide gel with 8 M urea and Tris-borate-EDTA (TBE) buffer. The gel was dried and processed as above.

Protein Data Bank accession numbers

The structure factors and coordinates for the Cre*/LoxP-G5 structural model have been deposited in the RCSB Protein Data Bank with accession number 1KBU. The abbreviations used for the reaction intermediates are their PDB accession numbers, e.g. the precleavage intermediate has the PDB accession number 4CRX.

Acknowledgments

This work was funded, in part, by an American Cancer Society Institutional Research Grant (#36769) through U.C. Davis Cancer Center and U.C. Davis Molecular and Cellular Biology startup funds. E.P. was supported by a U.C. Davis SURPRISE summer fellowship. T.L. was supported by the U.C. Davis SHARP program funded by the Howard Hughes Medical Institute. Many thanks to James Endrizzi, James Holton, and Tom Alber (U.C. Berkeley) for synchrotron access and data collection assistance. Protein purifications, home-source data collections, and all computations were carried out in the W.M. Keck Protein Expression and X-ray Crystallographic Facilities at U.C. Davis. Synchrotron data were obtained at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences."

References

1. Sadowski, P. (1986). Site-specific recombinases: changing partners and doing the twist. *J. Bacteriol.* **165**, 341–347.
2. Stark, W. M., Boocock, M. R. & Sherratt, D. J. (1992). Catalysis by site-specific recombinases. *Trends Genet.* **8**, 432–439.
3. Kilby, N. J., Snaith, M. R. & Murray, J. A. (1993). Site-specific recombinases: tools for genome engineering. *Trends Genet.* **9**, 413–421.
4. Metzger, D. & Feil, R. (1999). Engineering the mouse genome by site-specific recombination. *Curr. Opin. Biotechnol.* **10**, 470–476.
5. Sadowski, P. D. (1993). Site-specific genetic recombination: hops, flips, and flops. *FASEB J.* **7**, 760–767.
6. Van de Putte, P. & Goosen, N. (1992). DNA inversions in phages and bacteria. *Trends Genet.* **8**, 457–462.
7. Nash, H. A. (1981). Integration and excision of bacteriophage lambda: the mechanism of conservation site specific recombination. *Annu. Rev. Genet.* **15**, 143–167.
8. Volkert, F. C. & Broach, J. R. (1986). Site-specific recombination promotes plasmid amplification in yeast. *Cell*, **46**, 541–550.
9. Summers, D. (1998). Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. *Mol. Microbiol.* **29**, 1137–1145.
10. Sherratt, D. J., Arciszewska, L. K., Blakely, G., Colloms, S., Grant, K., Leslie, N. & McCulloch, R. (1995). Site-specific recombination and circular chromosome segregation. *Philos. Trans. R. Soc. ser. B*, **347**, 37–42.
11. Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Ljungquist, E., Hoess, R. H. *et al.* (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *Eur. Mol. Biol. Org. J.* **5**, 433–440.
12. Esposito, D. & Scocca, J. J. (1997). The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucl. Acids Res.* **25**, 3605–3614.
13. Cheng, C., Kussie, P., Pavletich, N. & Shuman, S. (1998). Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell*, **92**, 841–850.
14. Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucl. Acids Res.* **26**, 391–406.
15. Sherratt, D. J. & Wigley, D. B. (1998). Conserved themes but novel activities in recombinases and topoisomerases. *Cell*, **93**, 149–152.
16. Hoess, R., Wierzbicki, A. & Abremski, K. (1987). Isolation and characterization of intermediates in site-specific recombination. *Proc. Natl Acad. Sci. USA*, **84**, 6840–6844.
17. Sternberg, N. (1981). Bacteriophage P1 site-specific recombination. III. Strand exchange during recombination at lox sites. *J. Mol. Biol.* **150**, 603–608.
18. Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. *Genesis*, **26**, 99–109.
19. Hoess, R. H., Ziese, M. & Sternberg, N. (1982). P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc. Natl Acad. Sci. USA*, **79**, 3398–3402.
20. Hoess, R. H. & Abremski, K. (1985). Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *J. Mol. Biol.* **181**, 351–362.
21. Lee, G. & Saito, I. (1998). Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene*, **216**, 55–65.
22. Hoess, R. H., Wierzbicki, A. & Abremski, K. (1986). The role of the loxP spacer region in P1 site-specific recombination. *Nucl. Acids Res.* **14**, 2287–2300.
23. Van Duyne, G. D. (2001). A structural view of cre-loxp site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 87–104.
24. Gopaul, D. N. & Duyne, G. D. (1999). Structure and mechanism in site-specific recombination. *Curr. Opin. Struct. Biol.* **9**, 14–20.
25. Guo, F., Gopaul, D. N. & Van Duyne, G. D. (1999). Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Natl Acad. Sci. USA*, **96**, 7143–7148.
26. Guo, F., Gopaul, D. N. & van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.
27. Gopaul, D. N., Guo, F. & Van Duyne, G. D. (1998). Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *Eur. Mol. Biol. Org. J.* **17**, 4175–4187.
28. Abremski, K. & Hoess, R. (1984). Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.* **259**, 1509–1514.

29. Ringrose, L., Lounnas, V., Ehrlich, L., Buchholz, F., Wade, R. & Stewart, A. F. (1998). Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. *J. Mol. Biol.* **284**, 363–384.
30. Chen, Y., Narendra, U., Iype, E. L., Cox, M. M. & Rice, A. P. (2000). Crystal structure of a Flp recombinase–Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell*, **6**, 885–897.
31. Yang, W. & Mizuuchi, K. (1997). Site-specific recombination in plane view. *Structure*, **5**, 1401–1406.
32. Jayaram, M. (1997). The *cis–trans* paradox of integrase. *Science*, **276**, 49–51.
33. Lee, J., Jayaram, M. & Grainge, I. (1999). Wild-type Flp recombinase cleaves DNA in *trans*. *Eur. Mol. Biol. Org. J.* **18**, 784–791.
34. Lee, J., Tonozuka, T. & Jayaram, M. (1997). Mechanism of active site exclusion in a site-specific recombinase: role of the DNA substrate in conferring half-of-the-sites activity. *Genes Dev.* **11**, 3061–3071.
35. Lee, L. & Sadowski, P. D. (2001). Directional resolution of synthetic holliday structures by the Cre recombinase. *J. Biol. Chem.* **276**, 31092–31098.
36. Tribble, G., Ahn, Y. T., Lee, J., Dandekar, T. & Jayaram, M. (2000). DNA recognition, strand selectivity, and cleavage mode during integrase family site-specific recombination. *J. Biol. Chem.* **275**, 22255–22267.
37. Woods, K. C., Martin, S. S., Chu, V. C. & Baldwin, E. P. (2001). Quasi-equivalence in site-specific recombinase structure and function: crystal structure and activity of trimeric cre recombinase bound to a three-way lox dna junction. *J. Mol. Biol.* **313**, 49–69.
38. Nunes-Duby, S. E., Matsumoto, L. & Landy, A. (1987). Site-specific recombination intermediates trapped with suicide substrates. *Cell*, **50**, 779–788.
39. Chen, J. W., Lee, J. & Jayaram, M. (1992). DNA cleavage in *trans* by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell*, **69**, 647–658.
40. Whang, I., Lee, J. & Jayaram, M. (1994). Active-site assembly and mode of DNA cleavage by Flp recombinase during full-site recombination. *Mol. Cell. Biol.* **14**, 7492–7498.
41. Blakely, G. W., Davidson, A. O. & Sherratt, D. J. (1997). Binding and cleavage of nicked substrates by site-specific recombinases XerC and XerD. *J. Mol. Biol.* **265**, 30–39.
42. Shaikh, A. C. & Sadowski, P. D. (1997). The Cre recombinase cleaves the lox site in *trans*. *J. Biol. Chem.* **272**, 5695–5702.
43. Lee, J., Tribble, G. & Jayaram, M. (2000). Resolution of tethered antiparallel and parallel holliday junctions by the Flp site-specific recombinase. *J. Mol. Biol.* **296**, 403–419.
44. Grainge, I., Buck, D. & Jayaram, M. (2000). Geometry of site alignment during int family recombination: antiparallel synapsis by the Flp recombinase. *J. Mol. Biol.* **298**, 749–764.
45. Arciszewska, L. K., Grainge, I. & Sherratt, D. J. (1997). Action of site-specific recombinases XerC and XerD on tethered Holliday junctions. *Eur. Mol. Biol. Org. J.* **16**, 3731–3743.
46. Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W. & Cozzarelli, N. R. (1999). The topological mechanism of phage lambda integrase. *J. Mol. Biol.* **289**, 747–775.
47. Aranda, M., Kanellopoulou, C., Christ, N., Peitz, M., Rajewsky, K. & Droge, P. (2001). Altered directionality in the Cre-LoxP site-specific recombination pathway. *J. Mol. Biol.* **311**, 453–459.
48. Senecoff, J. F. & Cox, M. M. (1986). Directionality in FLP protein-promoted site-specific recombination is mediated by DNA–DNA pairing. *J. Biol. Chem.* **261**, 7380–7386.
49. Azaro, M. A. & Landy, A. (1997). The isomeric preference of Holliday junctions influences resolution bias by lambda integrase. *Eur. Mol. Biol. Org. J.* **16**, 3744–3755.
50. Dixon, J. E. & Sadowski, P. D. (1993). Resolution of synthetic chi structures by the FLP site-specific recombinase. *J. Mol. Biol.* **234**, 522–533.
51. Colloms, S. D., McCulloch, R., Grant, K., Neilson, L. & Sherratt, D. J. (1996). Xer-mediated site-specific recombination *in vitro*. *Eur. Mol. Biol. Org. J.* **15**, 1172–1181.
52. Abremski, K., Hoess, R. & Sternberg, N. (1983). Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell*, **32**, 1301–1311.
53. Erie, D. A., Hajiseyedjavadi, O., Young, M. C. & von Hippel, P. H. (1993). Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science*, **262**, 867–873.
54. Foster, J. E., Holmes, S. F. & Erie, D. A. (2001). Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell*, **106**, 243–252.
55. Packer, M. J., Dauncey, M. P. & Hunter, C. A. (2000). Sequence-dependent DNA structure: dinucleotide conformational maps. *J. Mol. Biol.* **295**, 71–83.
56. el Hassan, M. A. & Calladine, C. R. (1995). The assessment of the geometry of dinucleotide steps in double-helical DNA; a new local calculation scheme. *J. Mol. Biol.* **251**, 648–664.
57. Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
58. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T. *et al.* (1992). The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759.
59. Krogh, B. O. & Shuman, S. (2000). Catalytic mechanism of DNA topoisomerase IB. *Mol. Cell*, **5**, 1035–1041.
60. Kunkel, T. A., Bebenek, K. & McClary, J. (1991). Efficient site-directed mutagenesis using uracil-containing DNA. *Methods Enzymol.* **204**, 125–139.
61. Cantor, C. R., Warshaw, M. M. & Shapiro, H. (1970). Oligonucleotide interactions. 3. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
62. Powell, H. R. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallog. sect. D*, **55**, 1690–1695.
63. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image data. *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography*, vol. 26, Daresbury Laboratory, Warrington, UK.
64. Collaborative computational project, No. 4 (1994). The CCP4 suite programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
65. Holton, J. H. (2001). *The ELVES Manual. Glossary of X-ray Terms*, University of California Berkeley, Berkeley, CA.

66. Tronrud, D. (1997). TNT refinement package. *Methods Enzymol.* **277**, 306–319.
67. Engh, R. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A*, **47**, 392–400.
68. Tronrud, D. (1996). Knowledge-based *B*-factor restraints for the refinement of proteins. *J. Appl. Crystallog.* **29**, 100–104.
69. Brunger, A. (1993). Assessment of phase accuracy by cross validation: the free *R* value. Methods and applications. *Acta Crystallog. sect. D*, **49**, 24–36.
70. Baldwin, E., Xu, J., Hajiseyedjavadi, O., Baase, W. A. & Matthews, B. W. (1996). Thermodynamic and structural compensation in size-switch core repacking variants of bacteriophage T4 lysozyme. *J. Mol. Biol.* **259**, 542–559.
71. Zhang, X.-J. & Matthews, B. W. (1995). EDPDB: a multi-functional tool for protein structure analysis. *J. Appl. Crystallog.* **28**, 624–630.
72. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
73. Merritt, E. A. & Murphy, M. E. P. (1994). Raster3d version 2.0: a program for photorealistic molecular graphics. *Acta Crystallog. sect. D*, **50**, 869–873.
74. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.

Edited by K. Morikawa

(Received 27 November 2001; received in revised form 11 March 2002; accepted 18 March 2002)