SYNAPSIS OF loxP SITES BY CRE RECOMBINASE*
Kaushik Ghosh1, Feng Guo2, and Gregory D. Van Duyne1
1Department of Biochemistry & Biophysics and Howard Hughes Medical Institute, University of Pennsylvania, Philadelphia PA 19104
2Present Address: Department of Biological Chemistry, UCLA School of Medicine Los Angeles, CA 90095
Running head: Synapsis of loxP sites
Address correspondence to: Gregory D. Van Duyne, 242 Anatomy-Chemistry Building, University of Pennsylvania, Philadelphia, PA 19104-6059. Phone: 215-898-3058; Fax: 215-573-4764; Email: vanduyne@mail.med.upenn.edu

Cre recombinase catalyzes site-specific recombination between 34-bp loxP sites in a variety of topological and cellular contexts. An obligatory step in the recombination reaction is the association, or synapsis, of Cre-bound loxP sites to form a tetrameric protein assembly that is competent for strand exchange. Using analytical ultracentrifugation and electrophoresis approaches, we have studied the energetics of Cre-mediated synapsis of loxP sites. We found that synapsis occurs with a high affinity (Kd = 10 nM) and is pH-dependent, but does not require divalent cations. Surprisingly, the catalytically inactive Cre K201A mutant is fully competent for synapsis of loxP sites, yet the inactive Y324F and R173K mutants are defective for synapsis. These findings have allowed us to determine the first crystal structures of a pre-cleavage Cre-loxP synaptic complex in a configuration representing the starting point in the recombination pathway. When combined with a quantitative analysis of synapsis using loxP mutants, the structures explain how the central eight base-pairs of the loxP site are able to dictate the order of strand exchange in the Cre system.

Site-specific recombinases from the tyrosine recombinase family are used by bacteria and yeast to mediate the integration, excision, resolution, and inversion of DNA segments in order to carry out a wide variety of biological functions (1,2). In the simplest systems, typified by the bacteriophage P1 Cre recombinase (3) and the Saccharomyces cerevisiae Flp recombinase (4), only the recombinase enzymes and DNA substrates containing 34-bp recombination sequences are required for efficient recombination. In others, such as the bacteriophage λ integrase (λ-int)(5) and the E. coli XerC and XerD recombinases (6), efficient recombination requires more complex recombination sites and the activities of auxiliary proteins that tightly regulate the forward and reverse reactions. A remarkable property of both the simple and complex systems is their ability to efficiently synapse (associate) recombining sites that can be located far from one another on the same chromosome or, for bacteriophage integration, are in separate DNA molecules.

Synapsis of recombining sites has been visualized in tyrosine recombinase systems using a variety of complementary experimental approaches. For example, synaptic λ-int att and Cre-loxP complexes have been observed by electron microscopy (7,8) and as slower-migrating bands on non-denaturing polyacrylamide gels (9,10); atomic force microscopy has been used to probe synaptic complexes in the Int, Flp, and Cre systems (11,12); and single molecule light microscopy has recently been used to observe formation of synaptic complexes during λ-int recombination (13). In addition, crystal structures have been reported for synaptic Cre-DNA complexes (14,15) and for a truncated λ-int/DNA complex (16). Despite the fundamental importance of this initial step in the recombination reaction pathway, however, there have been no quantitative studies of the energetics of synapsis in these systems.

Cre recombinase has proven to be a useful model system for biochemical and structural investigations (17). During synapsis, Cre sharply bends the loxP sites to form an assembly that is stabilized by protein-protein interactions between Cre subunits bound to the associating sites (14,15). Once this complex is formed, the recombinase subunits are committed to cleavage and exchange.
of one of the two pairs of DNA strands. As shown in Fig. 1, however, this initial step in the recombination pathway is actually quite complex. There are four distinct synaptic complexes that can form between two Cre-bound \( \text{loxP} \) sites if the sites are arranged in an approximately co-planar configuration; two of the complexes are anti-parallel (I and II in Fig. 1c) and two are parallel (III and IV in Fig. 1c).

Structural, biochemical, and topological data strongly support a model in which only an anti-parallel alignment of sites will lead to efficient strand exchange (2,17-19). One might predict that Cre-\( \text{loxP} \) site-specific recombination should be able to initiate recombination equally well from either of the anti-parallel synaptic configurations. However, several independent experimental approaches have demonstrated that Cre preferentially exchanges the \( \text{loxP} \) bottom strands (the black strands in Fig. 1) first to generate the Holliday junction (HJ) intermediate, and then exchanges the top (red) strands during HJ resolution to form recombinant products (19-23). Indeed, recent data further indicate that preferential formation of complex II in Fig. 1c is responsible for this bias in the order of strand exchanges (23).

The structural and biochemical bases for this asymmetry in the Cre recombination pathway have only been partially addressed, leaving a number of unanswered questions. For example, why should one of the anti-parallel synapses be preferred over the other? What is the structure of this synaptic complex? What is the binding affinity associated with Cre-\( \text{loxP} \) synopsis? Is synopsis best viewed as an interaction between the protein surfaces from two rigid Cre-bound \( \text{loxP} \) sites or do the dynamics of \( \text{loxP} \) bending also play a role? One of the primary gaps in our mechanistic understanding of the Cre-\( \text{loxP} \) recombination pathway has been the lack of an experimental system to quantitatively study the requirements for efficient synopsis. A second issue has been the lack of a structural model that represents the actual starting point in the recombination pathway. Crystal structures of synaptic complexes containing Cre bound to modified \( \text{loxP} \) sites (14,15) have provided a great deal of general insight, but do not provide satisfactory explanations to these questions. Indeed, one of the reported Cre-DNA synaptic complex structures (15) argues against the currently accepted model for initiation of recombination.

Here, we report advances in understanding both the biochemistry and structural biology of synopsis in the Cre-\( \text{loxP} \) system. Using analytical ultracentrifugation and electrophoresis techniques, we have developed assays to measure equilibrium binding constants for Cre-mediated synopsis of \( \text{loxP} \) sites. Using these methods, we have examined the effects of pH, divalent ions, Cre mutants, and \( \text{loxP} \) modifications on synopsis. By using a cleavage-deficient Cre mutant that is able to synapse \( \text{loxP} \) sites as well as wild-type Cre, we have also determined two independent crystal structures of pre-cleavage Cre-\( \text{loxP} \) synaptic complexes that reveal the actual starting point for the recombination pathway. Together, the biochemical and structural data explain why Cre preferentially forms a bottom-strand cleavage synaptic complex to initiate recombination and provide new insights into this important first step in the recombination pathway.

**Experimental Procedures**

**Protein and DNA purification-** Wild-type Cre and Cre mutants were overexpressed and purified as described (24). The extinction coefficient for Cre at 280 nm was determined to be 49.1 mM\(^{-1}\) cm\(^{-1}\) by comparing protein absorbance before and after denaturation in 3 M guanidine hydrochloride (25). Oligonucleotides for ultracentrifugation experiments (\( \text{loxP} \)cent in Supplemental Fig. 1) were synthesized by the Keck Facility at Yale University with the 5’-dimethoxytrityl (DMT) group attached and purified by reverse-phase HPLC as described (24). Following detritylation and concentration, oligonucleotides were annealed in 10 mM TrisHCl, pH 8, 0.5 mM EDTA, 100 mM NaCl (annealing buffer) and further purified on a 5 ml type I hydroxyapatite column (BioRad) with sodium/potassium phosphate elution to remove small amounts of excess ssDNA. The extinction coefficient at 260 nm was determined experimentally as 661 mM\(^{-1}\) cm\(^{-1}\) for duplex 44-mer \( \text{loxP} \) sites by comparing the absorbance before and after total DNase digestion, and using literature values for the extinction coefficients of individual nucleotides. For crystallization, oligonucleotides were purified by reverse-phase HPLC as described above, but the hydroxyapatite
column step was omitted. For gel-based synopsis assays, oligonucleotides were gel purified.

**Analytical ultracentrifugation** - Sedimentation equilibrium ultracentrifugation experiments were performed at 20 °C using a Beckman Optima XL-A analytical ultracentrifuge fitted with an AN-60 Ti rotor and with six-sector centerpieces. To determine synopsis affinities, Cre-*loxP* complexes were prepared at 0.6, 0.3, and 0.15 μM 44-mer *loxP* site DNA and 1.8, 0.9, 0.45 μM Cre, respectively, in standard recombination buffer (20 mM sodium HEPES, pH 7.5, 150 mM NaCl, 10 mM MgCl2, 1 mM DTT; measured density of 1.004 g/cm^3). These ratios provide an excess of 0.5 Cre subunits per *loxP* half-site, which we found to be important in ensuring saturation of *loxP* sites throughout the experiment. Reference chambers were filled with either buffer alone or buffer containing 0.6, 0.3, and 0.15 μM Cre, respectively. Similar binding constants resulted from both approaches, but the fits were slightly better when the concentration of excess Cre was matched in the reference chambers due to the small absorbance contribution of unbound Cre that affected the absorbance baselines. The data fit in Fig. 2 and Supplemental Fig. 2 were measured using the latter approach. Samples were centrifuged at 6000, 9000, 12000 and 15000 rpm and radial absorbance scans were measured in step mode at 260 nm after 12 and 14 hours, at which time equilibrium had been achieved.

Radial absorbance data were fit to a monomer-dimer equilibrium model using the program SEDPHAT (26), with a fixed ‘monomer’ molecular weight of 104,000 Da, based on two 38.5 kDa Cre subunits and one 27 kDa *loxP*-containing 44mer duplex. The partial specific volume (\(\bar{\nu}\)) for Cre (0.73 cm^3/g) was determined from the protein amino acid composition. The \(\bar{\nu}\) value for *loxP* DNA (0.59 cm^3/g) was determined from sedimentation equilibrium analysis of the unbound DNA duplex, which behaved as an ideal monomer. The \(\bar{\nu}\) value of the Cre-*loxP* monomer (0.69 cm^3/g) was estimated as the mass-weighted average of the partial specific volumes for Cre and the *loxP* site. An estimated error for the equilibrium constant was determined from a 1000-iteration Monte Carlo simulation, as implemented in SEDPHAT.

To examine the effects of buffer composition on synopsis, Cre-*loxP* complexes were prepared in the appropriate buffers at 400 nM *loxP* and 1.2 μM Cre. Centrifugation and absorbance scans were performed as described above and relative synopsis association constants were estimated from global fits of the radial scans. These affinity constants are less well determined than the global fits involving three different Cre-*loxP* concentrations, but this approach allowed the simultaneous comparison of a large number of buffer conditions in the multi-chamber centerpieces.

Sedimentation equilibrium analysis of Cre alone was performed using 6 μM Cre in 20 mM sodium HEPES, pH 7.5, 275 mM NaCl, 1 mM DTT, with or without 5 mM MgCl2. Samples were centrifuged at 12, 15, 18, and 21k rpm and radial absorbance scans were measured at 280 nm after equilibrium had been reached at 14 hours. Absorbance data fit well to a single species model corresponding to a Cre monomer.

**Synapsis gel electrophoresis assay** - Ten pmol of 54-mer top strand was 5'-end labeled with ^32P and annealed to the complementary bottom strand by slow cooling (1 °C/min) in a thermal cycler. Synapsis reactions were carried out in 20 mM sodium HEPES buffer, pH 7.5, 5 mM MgCl2, 150 mM NaCl, 2 mM DTT and 50 μg/ml sheared salmon sperm DNA (1X NCB buffer). Each reaction in a given titration contained 200 pM radiolabeled 54-mer *loxP* DNA, 1 μM Cre or Cre mutant, and varying concentrations of unlabeled 54-mer DNA. Reactions were incubated for 20 minutes at 20° C, mixed with loading dye (0.001% bromophenol blue, 0.001% xylene cyanol and 3% Ficoll) and loaded onto a 6% non-denaturing polyacrylamide gel (29:1 acrylamide:bis-acrylamide) that had been pre-run at 10 V/cm for 45 minutes at 20° C, with constant temperature maintained by a circulating water bath. The gel polymerization and running buffers were 50 mM Hepes acid, 25 mM Tris base, at pH 7.0. Similar results were obtained when 1 mM EDTA was present in the buffers. After electrophoresis at 13 V/cm for 2 hours at 20 °C, gels were dried and quantitated by phosphorimager analysis (Molecular Dynamics).

Standard electrophoretic mobility shift assays with wild-type Cre and 54-mer *loxP* DNA performed under identical conditions were used to verify the assignments of free DNA, DNA bound by one or two Cre subunits, and synaptic complex
(as annotated in Fig. 3). The fraction synapsed (f) was calculated as the counts in the synaptic complex counts plus the unsynapsed Cre-\textit{loxP} counts in the corresponding lane and then plotted against the total DNA concentration. A monomer-dimer equilibrium model was utilized to fit the data using the equation

$$f = f_{\text{max}} \left[ 1 - \frac{K}{4c} \left( \sqrt{1 + \frac{8c}{K}} - 1 \right) \right]$$

where $c$ is the total concentration of 54-mer \textit{loxP} site, and $K$ is the monomer-dimer dissociation constant. For $D = [\text{dimer}] = [(\text{Cre-2loxP})_2]$ and $M = [\text{monomer}] = [\text{Cre-loxP}]$, then $K = M^2/D$, $c = 2D + M$, and $f = 2D/(2D+M)$. The equation used for fitting was derived by expressing $M$ as a function of $c$ and $K$ and substituting this expression into $f = (c-M)/c$. The $f_{\text{max}}$ term was introduced as an adjustable parameter and was fit to nearly 1.0 in most cases. The binding data cannot be properly fit to a simple binding isotherm, since the approximation that the total monomer concentration is equal to the free monomer concentration is invalid. Data were fit using Origin software (OriginLab Corp.).

**Structure determination and refinement** - Cre K201A/\textit{loxP} crystals were formed by vapor diffusion at 18°C from solutions containing 13 \textgreek{M} protein and 19 \textgreek{M} 35-mer \textit{loxP} DNA (the 34-base pair \textit{loxP} site containing 5'-T residues on each strand; Supplemental Fig. 1) in 20 mM sodium acetate, pH 5, 30% 2-methyl-2,4-pentanediol, and 20 mM CaCl$_2$. Both orthorhombic and trigonal crystal forms grew from the same conditions. Diffraction data were measured at the Advanced Light Source beamline 8.2.1 using a Quantum4 CCD detector at 100K and processed using the HKL suite (27). The trigonal crystal form was sensitive to radiation damage, resulting in a lower completeness and higher overall $R_{\text{sym}}$ value compared to the orthorhombic form (Table III).

Both orthorhombic and trigonal crystal forms grew from the same conditions. Diffraction data were measured at the Advanced Light Source beamline 8.2.1 using a Quantum4 CCD detector at 100K and processed using the HKL suite (27). The trigonal crystal form was sensitive to radiation damage, resulting in a lower completeness and higher overall $R_{\text{sym}}$ value compared to the orthorhombic form (Table III).

Both synaptic complex crystal forms were initially phased using molecular replacement with AMORE (28) and the recombination-DNA half-sites from the covalent intermediate structure 1CRX (29) as a search model. The resulting starting models were then optimized using rigid-body refinement of the individual protein domains and the 13-bp inverted repeat DNA arms at 3 \textAA resolution, resulting in correlation coefficients of 0.88/0.81 and $R$-factors of 0.35/0.38 for the orthorhombic and trigonal crystal forms, respectively. The 8-bp spacer region encompassing positions 4' through 4 was omitted from the model until these DNA residues could be unambiguously fit into unbiased electron density maps. Iterative cycles of positional refinement with REFMAC (30) and model building in O (31) were performed as the resolution was gradually increased to the limits of the individual data sets.

$\sigma_A$-weighted $2F_o-F_c$ and $F_o-F_c$ electron density maps showed clear density for the bases and for the backbone of the omitted central 8-base pairs of the \textit{loxP} site. This region was initially modeled as (dAdU)$_4$ until the final stages of refinement in order to avoid bias in the choice of sequence directionality as described (32). The sequence assignment was then made based on the presence or absence of difference map densities for Gua-N2 and Thy-C5, which uniquely identifies the direction of the pseudo-palindromic spacer. Electron density maps following refinement of the (dAdU)$_4$ model were consistent with only one direction for the \textit{loxP} spacer, corresponding to a bottom strand cleavage configuration. Refinement results are summarized in Table III and coordinates for the orthorhombic and trigonal forms have been deposited in the Protein Data Bank, with accession codes 2HOF and 2HOI, respectively.

**RESULTS**

**Cre-\textit{loxP} synapsis is pH-dependent, but does not require divalent cations** - Our initial attempts to monitor Cre-\textit{loxP} synapsis as a function of \textit{loxP} concentration were based on native PAGE analysis. Slower migrating bands representing higher order complexes have been reported in the Cre system (9,33) and these bands were shown to contain primarily synaptic complexes formed by association of two Cre-bound \textit{loxP} sites (9). When we attempted to extract binding constants using this approach, the resulting affinities were weaker than anticipated and the results were difficult to reproduce in separate experiments. After considering a number of plausible explanations (e.g. buffer requirements and stability of the synaptic complex during electrophoresis), we decided to first establish the basic requirements...
and affinities involved in Cre-loxP synapsis using sedimentation equilibrium analytical ultracentrifugation (hereafter referred to as AU).

The primary advantage of AU is the ability to monitor complex formation under equilibrium conditions using well-defined experimental parameters, such as buffer components, pH, and temperature (34,35). The equilibria involved in synapsis of loxP sites are complex if cooperative binding of Cre subunits to the loxP site is also considered. To simplify the analysis, we performed AU experiments with 44-bp loxP sites and saturating concentrations of Cre, taking advantage of observations that Cre binds to the loxP site with a sub-nanomolar K_d (36). Under these conditions, the loxP sites are fully occupied with Cre subunits and synapsis can be modeled as a simple monomer-dimer equilibrium. In this model, two associating 104 kDa Cre-loxP complexes (monomers) associate to form a 208 kDa synaptic complex (dimer) (Fig. 1b). The 44-bp loxP site used (Supplemental Fig. 1) dominates absorbance at 260 nm, with Cre contributing only a few percent of the signal at this wavelength. The AU experiment therefore monitors the oligomeric state of the loxP site, to a good approximation.

The results of an AU experiment for wild-type Cre-loxP synapsis are summarized in Fig. 2 and in Supplemental Fig. 2. Three different loxP concentrations (each with saturating amounts of Cre) were monitored at four different rotor speeds and the equilibrium binding constant was determined by a global fit to the resulting twelve radial distributions. The excellent fits and randomly distributed residuals indicate that the monomer-dimer model is a good description of loxP site association under these conditions.

The experiment described in Fig. 2 was performed using our standard recombination assay buffer, which is at near neutral pH and contains both monovalent (sodium) and divalent (magnesium) cations. To test the effects of different buffer conditions on the ability of Cre to synapse loxP sites, we performed AU experiments using the buffer conditions shown in Table I. We were particularly interested to learn if divalent ions are required for efficient synapsis. It was originally reported that divalent ions and small polyamines increase the efficiency of the overall Cre recombination reaction (3) and a role for divalent ions in stabilizing the synaptic complex has been suggested (14). We were therefore surprised to find that divalent ions have only a small effect on synapsis of loxP sites and that a physiological concentration of monovalent cations appears to be sufficient for this stage of the recombination pathway. A role for divalent ions or polyamines in facilitating cooperative Cre-loxP binding has also been suggested (37), which would not be manifested in these experiments since we are forcing full occupancy of the loxP sites by providing Cre in excess.

We also wanted to determine if Cre-loxP synapsis is pH dependent. As shown in Table I, synapsis is efficient from pH 5.5 to 7.5, which is consistent with our previous observations that the in vitro Cre recombination reaction is robust over a wide pH range, including that of most restriction enzyme buffers, which are typically buffered at pH 7.5-8.0. At pH 8.5 (Table I) and above (data not shown), Cre-loxP synaptic complexes become much less stable. We do not yet understand the cause of the breakdown in synapsis that occurs at pH 8.5, but ionization of two Cys residues located in the helical core-binding domains of Cre could alter the ability of these domains to interact in the synaptic complex.

The negative effect of alkaline pH on synapsis provides an explanation for the difficulties we encountered in performing synapsis assays using native PAGE, since these types of gels are commonly run with buffers with pH 8-9. Indeed, we had been using Tris-borate-EDTA (TBE) buffer systems at pH 8.5-9. As expected, AU synapsis experiments performed in 0.5X or 1X TBE result in only small amounts of synaptic complex formation even at high loxP concentration (data not shown).

In addition to studying synapsis of loxP sites under a variety of conditions, we also used AU to determine the oligomeric state of Cre in solution. An early report based on gel filtration and glycerol gradient ultracentrifugation indicated that Cre was monomeric, but that magnesium ions cause the sedimentation coefficient to increase (3). We conducted sedimentation equilibrium experiments for 6 µM wild-type Cre using buffers composed of 20 mM sodium Hepes, pH 7.5, 275 mM NaCl, 1 mM DTT, both with and without 5 mM MgCl₂. In both cases, we found that Cre is monomeric, with no evidence of dimer or higher order multimer formation (Supplemental Fig. 3). Similar results
were obtained with buffers containing 125 mM NaCl. The increase in sedimentation coefficient previously observed in the presence of magnesium ions is therefore most likely due to a change in shape of the free protein in solution.

Based on the findings from the AU experiments described above, we modified the gel-based synapsis assay by using Tris-Hepes or Tris-Hepes-EDTA at pH 7-7.5 as a buffer system. Although a number of neutral pH buffers have been described for native PAGE (38), we favored a system where two buffering components bracket the desired pH. An example of this improved assay for wild-type Cre-loxP synapsis is shown in Fig. 3. The effective Kd of 14 nM obtained from a fit of the binding data agrees reasonably well with the value of 10 nM obtained using AU. This indicates that the synaptic complexes formed are sufficiently stable under electrophoresis conditions to be able to extract useful quantitative binding data.

Cleavage is not required for efficient synapsis

One possible concern in interpreting the synapsis experiments shown in Figs 2 & 3 for wild-type Cre is that cleavage and strand exchange should be occurring at some rate during the course of the experiment. In particular, some fraction of the ‘dimer’ complex representing synapsed loxP sites may also include a contribution from HJ intermediates that have formed between sites undergoing recombination. To address this issue, we measured the ability of cleavage-deficient active site mutants to synapse loxP sites using the gel-based assay. As shown in Table II and Supplemental Fig. 4, the Cre K201A mutant is able to synapse loxP sites as well as wild-type Cre, despite being catalytically inactive for cleavage or recombination (23).

An interesting difference observed in the synapsis of loxP sites by wild-type Cre vs. the K201A mutant is a lower binding plateau (fmax ~ 0.75) at high concentrations of loxP site for the mutant. To determine if this is due to the inability of Cre K201A to cleave the loxP site, as opposed to an inherent property of this particular mutant, we tested the ability of this protein to synapse loxP sites containing a 5'-bridging phosphorothioate at the bottom strand scissile phosphate. Cre K201A can cleave this modified loxP site, due to the lowered pKₐ of the 5'-thiol leaving group vs. the normal 5'-hydroxyl group (23). We found that Cre K201A synapsis of the phosphorothioate-containing site is nearly identical to that observed for wild-type Cre (Supplemental Fig. 4 & Fig. 3). Thus, while cleavage is not a prerequisite for efficient synapsis of loxP sites, there is a subtle difference observed when a cleavage-competent recombinase enzyme is used.

Although this result indicates that cleavage of loxP sites is not required for synapsis, we were still interested to know if HJ intermediates were present at significant steady state levels under our experimental conditions in the case of wild-type Cre. To test this, synapsis binding assays were rapidly quenched with SDS/proteinase K at various time points and the extracted DNA was analyzed by native PAGE. Using our standard recombination buffer conditions (Table I), we were not able to detect HJ intermediates > 1% of the total loxP DNA, whereas accumulation of HJ intermediates by the HJ-generating Cre A312T mutant (20) was clear (data not shown). This result is consistent with early reports that HJ intermediates are present at very low steady-state levels in wild-type Cre-loxP recombination, under standard buffer conditions (20).

In contrast to the K201A mutant, two other inactive Cre mutants, Y324F and R173K, are both severely defective in synapsis (Table II). This was not anticipated in our earlier work, where we used these mutants bound to symmetric loxP variants to determine crystal structures of synaptic complexes (14). Presumably, the high concentrations used for crystallization were able to overcome the decreased affinity between mutant Cre-bound sites.

Neither Tyr324 or Arg173 is directly involved in the protein-protein interface formed between two associated loxP sites, therefore they must each affect synapsis indirectly. Arg173 forms an intimate, double hydrogen bonding interaction with the scissile phosphate, which is thought to stabilize the transition states of the catalytic phosphoryl transfer steps during recombination (29). Mutation of the corresponding residue in Flp recombinase (Arg191) to lysine causes altered DNA-bending of the frt site (39). Given that the loxP sites are sharply bent in synaptic Cre-loxP crystal structures, a deficiency in DNA-bending could be partly responsible for the synapsis defect of the R173K mutant.
The observation that Cre Y324F is defective in synapsis is more difficult to rationalize based on current structural data. Tyr324 is the catalytic nucleophile that forms transient 3’-phosphotyrosine linkages to the loxP sites during strand exchange. This tyrosine residue is observed ‘Docked’ to the upstream phosphate via a hydrogen bond from the phenol hydroxyl group prior to cleavage (14). One possible explanation for the synopsis defect is that the Tyr324-phosphate interaction may be important in stabilizing the helix L-M-N region of the recombinase catalytic domain (29). This region plays key roles in both synopsis and catalysis, since the C-terminal helix-N forms domain-swapping interactions across the synaptic interface and helices L and M contribute the conserved active site residues Trp315 and Tyr324, respectively. Indeed, the Cre ΔC331 mutant (deletion from Glu331 to the C-terminus) is unable to synapse loxP sites even at relatively high loxP concentrations (Table II).

We also tested the Cre A36V substitution, which was the first synopsis-deficient Cre mutant reported (9). Ala36 forms part of the protein-protein interface formed between core-binding domains in the synaptic complex and substitution of valine would be expected to disrupt this region of the structure, based on existing models. This mutant has wild-type loxP-binding activity (40) and is able to cleave suicide substrates (23). As expected, we found that Cre A36V is defective in synopsis, with Kd > 500 nM.

**Structural basis for selective DNA-bending in the Cre-loxP synaptic complex**- In order to determine the structure of a Cre-loxP synaptic complex representing the start of the recombination pathway, we made use of our observations that Cre K201A does not cleave loxP, but forms synaptic complexes as efficiently as the wild-type protein. Using this Cre mutant and unmodified loxP site DNA, we were able to grow diffraction quality crystals of the Cre K201A/loxP complex in orthorhombic and trigonal crystal forms, with resolution limits of 2.4 Å and 2.6 Å, respectively (Table III). The orthorhombic form is isomorphous with the structure of the covalent Cre/DNA intermediate (29) and contains one Cre/loxP complex in the crystallographic asymmetric unit. In this case, a crystallographic twofold symmetry axis relates the two Cre-bound loxP sites in an anti-parallel aligned synaptic complex. The trigonal crystal form contains the entire [Cre/loxP]2 synaptic complex in the asymmetric unit. As expected, this complex contains two anti-parallel loxP sites related by local (non-crystallographic) twofold symmetry.

Both structures were determined by straightforward molecular replacement and rigid body refinement of the protein domains and the 13-bp inverted repeats from the Cre/loxA structure (PDB code 1CRX; (29)) where the central 8 bp of the loxP sites were removed to eliminate model bias in this region. The central loxP sequence was fit into electron density as dA-dU base pairs, which allowed for refinement of the structure without committing to a particular direction for the loxP site (32). This simple approach is possible because the central 8 bp of loxP are pseudo-palindromic with respect to purine and pyrimidine nucleotides. The final assignment of sequence was then made from inspection of electron density maps late in the refinement process, where we observed a unique pattern of Thy-C5 and Gua-N2 densities that allowed unambiguous identification of the loxP direction. Unbiased electron densities used in the sequence assignment are shown in Supplemental Fig. 5. In both structures, the central region of loxP is very well ordered, with strong electron density for the sugar-phosphate backbone and for the bases.

The synaptic structures in the two crystal forms both reveal a bottom strand cleavage configuration, where the loxP site is sharply bent in the left half-site (defined in Fig. 1a) and the recombinase subunit poised to catalyze strand exchange is bound to the right half-site (Fig. 4 and complex II in Fig. 1c). The identification of which strand is committed for cleavage and exchange in the synaptic complex is based primarily on geometric and stereochemical arguments (14). In the structures described here, the loxP bottom strand trajectory is ideally suited for exchange between the two halves of the synaptic complex to form the HJ intermediate, with no protein-DNA contacts that might inhibit this process. In contrast, the top strand is intimately engaged in a network of interactions with the recombinase subunit and would be unable to migrate to the opposing loxP site even if the corresponding scissile phosphate were cleaved in the synaptic complex. As described for the earlier structures,
the ‘active’ and ‘inactive’ active sites are quite similar, with no large differences involving coordination of the scissile phosphate. Positioning of the Tyr324 nucleophile does differ between the two active sites, a likely consequence of allosteric regulation of catalysis via the domain-swapped carboxyl-terminal helices (41).

The structures in the two crystal forms are nearly identical, with a total of three independent views of the synapsed loxP site (one from the orthorhombic form and two from the trigonal form). When the central eight base-pairs of the loxP site are compared, the all-atom r.m.s. deviation between independent sites in the trigonal form is 0.4 Å and the r.m.s. deviation for comparison of each of these with the orthorhombic form is 0.8 Å. Similarly, comparison of the three independent ‘active’ and the three independent ‘inactive’ recombinase subunits shows r.m.s. deviations of 0.4-0.6 Å for well-ordered Cα atoms in residues 40-320. Most aspects of the recombinase subunits, the active sites, and the Cre-DNA interface are similar to that already described for previous Cre-DNA crystal structures. Here, we limit our discussion to the central region of the loxP site, which is most relevant to understanding why the bottom strand synaptic complex is favored to initiate recombination.

As shown in Fig. 4b, the loxP sites are bent sharply one base-pair step away from the top strand scissile phosphate. There is a large negative roll associated with the bend, which is a consequence of the major groove opening towards the synaptic interface (Fig. 4a). There is a corresponding compression of the minor groove, which is centered very close to the T3’ and T4’ backbones (Fig. 4c). This bend occurs in the left half-site, which is proximal to the ‘inactive’ recombinase subunit. The right half-site adopts a relatively undistorted B-DNA conformation, which includes the scissile bottom-strand phosphate that would normally be cleaved to initiate recombination in this step of the reaction pathway. The geometry of the DNA bends observed in both the orthorhombic and trigonal crystal forms described here are similar to those previously observed in the structure of Cre bound to a symmetric lox site (14) but involve entirely different DNA sequences (see Discussion).

The location and stereochemistry of the sharp loxP bend provide a compelling explanation for the observed preferential formation of a bottom-strand cleavage synapse. Compression of the minor groove would be expected to be more readily tolerated in the A/T-rich left half-site of loxP, relative to the right half-site. This argument is particularly striking when we consider that the center of the minor groove compression is very close to the 4’-scissile base (Fig. 4c), which implies that the minor groove functional groups of the 4 vs. 4’ bases (and to some extent the 3 vs. 3’ bases) should largely dictate whether Cre bends the left vs. right half-site of loxP. Since right vs. left half-site bending implies a difference in bend direction (i.e., complex I vs. complex II in Fig. 1c), the resulting synaptic complexes will be poised for top strand or bottom strand cleavage, respectively. Previous observations that addition of a 2-amino group to Ade-4’ to create a 2,6-diaminopurine-containing loxP eliminates the bottom strand cleavage preference (23) and that swapping of the 4’-scissile bases in the loxP substrate also reverses the preference (21) strongly support this simple model.

Asymmetric Cre-loxP contacts at the scissile bases- Only two recombinase residues interact directly with bases in the asymmetric central region of loxP (Fig. 5). Lys201 normally interacts with the N3 group of the scissile 4 or 4’-purine base via the minor groove and activates the scissile O5’ leaving group during catalysis (23). This residue has been mutated to alanine in the structures described here. The second interacting residue is Lys86, which directly contacts the scissile bases via the major groove and is well-defined in both structures. In the right half-site of loxP, Lys86 hydrogen bonds directly to O6 of Gua4, whereas in the left half-site, Lys86 hydrogen bonds directly to N7 of Ade4’. This asymmetric interaction at the ends of the 8-bp central region of loxP is similar to what was observed in the structure of an inactive Cre mutant bound to a loxP-derived HJ substrate (32). In that case, the resolution was higher (2 Å) and an additional water-mediated interaction was observed between the side chain of Lys86 and the exocyclic N6-amino group of Ade4’. We also observe a network of water-mediated interactions involving Lys86 in the somewhat lower resolution structures described here, but the details of these networks differ slightly in the three independent representations available from the two crystal forms.
forms. The well-defined set of direct interactions between Lys86 and the scissile bases in the current structures support the sequence assignment in the central region, since the direct hydrogen bond observed to O6 of Gua4 (~3 Å between Lys-Nε donor and Gua-O6 acceptor atoms) would not make sense if that base were adenine.

Despite the distinct hydrogen bonding interactions observed between Lys86 and the scissile bases in the 4 and 4′ positions (Fig. 5), this residue does not appear to play an important role in preferential formation of a synaptic complex poised for bottom-strand cleavage (21,23). This result can be rationalized in terms of the synaptic complex structures described here. The two alternative anti-parallel synapses differ by the bend directions of the sites and the DNA sequences distorted to achieve the bends (Fig. 1). However, the alternative complexes (I vs II in Fig. 1c) would be expected to have the same set of asymmetric interactions involving Lys86. There is no obvious structural reason why the specific interaction of Lys86 with Ade4′ would somehow facilitate bending in the left loxP half-site, or that the unique interaction with Gua4 would somehow disfavor bending in the right half-site.

In principle, Lys201 could contribute to selecting the observed synaptic configuration (i.e. complex II rather than complex I in Fig. 1c). This residue interacts directly with the 4′-purine base in the minor groove where cleavage occurs, but is excluded from making a similar interaction in the opposite half-site of the synaptic complex, covalent intermediate, and HJ intermediate structures (Fig. 5; 17). Since Lys201 is substituted by alanine in the structures described here, we cannot establish the nature of its interactions with the loxP half-sites. It is clear, however, that Lys201 is not required for preferential formation of this synaptic complex. The Cre K201A mutant displays the same bottom-strand cleavage preference with 5′-bridging phosphorothioate substrates and fluorescence resonance energy transfer experiments demonstrating a synaptic preference in solution made use of the K201A mutant to prevent cleavage of loxP sites (23).

Insights from synapsis of symmetric loxP variants- A complication in interpreting an equilibrium binding constant for synapsis of loxP sites in terms of the specific structural model shown in Fig. 4 is illustrated in Fig. 1. In principle, four distinct synaptic complexes can form between two loxP sites. Two have a relative anti-parallel orientation (I & II in Fig. 1c) and differ by both the bend direction of the sites and by which strand is committed for cleavage and exchange in the synaptic complex. The other two are parallel complexes (III & IV in Fig. 1c), containing loxP sites with opposite bend directions. Although the bottom strand cleavage complex II has been shown to be favored over the top-strand complex I in solution (23), there is currently no experimental evidence to indicate that parallel complexes do not readily form in solution as well. Indeed, parallel synaptic complexes for both the Cre and Flp systems have been observed by electron microscopy (7,42). Thus, the synapsis affinity constant measured in an experiment with loxP sites represents a macroscopic $K_d$ which includes contributions from individual equilibria involving the species shown in Fig. 1c.

To further simplify the interpretation of synapsis binding data and to test mechanistic hypotheses inferred from the Cre-loxP crystal structures, we compared the ability of Cre, Cre K201A, and Cre K86A to synapse the loxP site and the symmetric loxSL and loxSR sites (Fig. 6). The symmetric sites have two useful properties. First, they can form only one type of synaptic complex, since parallel and anti-parallel alignments are all the same. This allows for more direct interpretation of the interaction energies in terms of the DNA sequences involved. Second, since the DNA distortions involved in the synaptic loxP bend are localized to one of the two half-sites (as opposed to involving both half-sites), the two halves of the loxP spacer region can be separated and assayed independently by using symmetric sites. The results of these experiments are summarized in Fig. 7a.

Relative to loxP, synapsis of loxSR sites is less efficient by a factor of ~10-15 for wild-type Cre, Cre K201A, and Cre K86A. This is the trend that one would expect, given the structural model for DNA-bending shown in Fig. 4. The G/C-rich half-sites of loxSR are less able to accommodate the DNA-bend observed in the A/T-rich left half site, since the 2-amino groups of Gua-3 and Gua-4 would be inserted into the tightly compressed minor groove at the bend center. The previously reported crystal structures of the Cre R173K/loxSR and Cre Y324F/loxSR complexes...
indicate that this is exactly what happens when a synaptic complex is formed between loxSR sites. Presumably, there are no stereochemical alternatives for loxSR bending that would allow synapse formation with a lower energy.

In contrast to the results with loxSR, the synthesis of loxSL sites by Cre, Cre K201A, and Cre K86A is more efficient than synthesis of loxP by a factor of 3-4. In this case, both halves of the loxSL spacer region contain the sequence that readily accommodates a DNA bend in the synaptic complex. Since there are four identical ways of forming the synaptic complex from loxSL sites, the macroscopic $K_d$ value measured for loxSL synapsis should be roughly one fourth the value of the microscopic $K_s$ for synthesis of loxP sites to form antiparallel complex II in Fig. 1. This estimate is based on the assumption that the right half-site sequence of loxP (which is not significantly distorted from B-form DNA when the left half-site is bent) does not contribute substantially to the energetics of synapsis in complex II. If wild-type Cre-loxP synapsis is dominated by formation of complex II, then one might also expect to see a fourfold difference in macroscopic $K_d$ values between loxP and loxSL, based on a simple statistical factor. The experimental ratios $K_s$-loxP/$K_s$-loxSL for the Cre K201A and Cre K86A mutants are both 3.7, suggesting that this is a reasonable model. The same ratio for wild-type Cre is 2.9.

A comparison of loxSL vs. loxSR synapsis provides an estimate of the energy difference between bending the alternative loxP half-sites. Using the Cre K86A mutant synapsis data, the ratio of $K_s$ values is 41.8, corresponding to a free energy difference of $\sim 2.2$ kcal/mol at $25^\circ$ C for bending at the A/T-rich left vs. the G/C-rich right half-site of loxP. Since the Cre K86A mutant cannot make potentially discriminating interactions with the scissile bases in loxSL and loxSR, the assumption that the difference in synapsis affinity is primarily derived from DNA-bending energetics seems reasonable. The calculated free energy differences for wild-type Cre and Cre K201A are 2.2 and 2.1 kcal/mol, respectively, indicating that Lys86 interactions with the scissile bases do not have a large affect. Although this analysis is admittedly a simplification, it does show that a quantitative understanding of individual steps in the Cre-loxP site-specific recombination can yield useful insights into fundamental mechanistic issues.

**Synopsis of loxP mutants** The loxSL and loxSR sites simplify the interpretation of synapsis affinities, but each of these sites involves three base-pair changes relative to the loxP sequence (Fig. 6). To probe the synapsis of loxP sites with smaller changes, we tested the ability of Cre, Cre K201A, and Cre K86A to synapse a series of loxP sites modified at the scissile bases (Fig. 7b). The loxP-DAP site, for example, differs from loxP by only a single functional group. In this site, Ade-4' is replaced by 2,6-diaminopurine (DAP) and the complementary thymidine residue is left unchanged. The DAP/T base pair is therefore identical to an A/T base-pair, except for the presence of an adenine 2-amino group and an additional DAP-T hydrogen bond (Fig. 5). If facile minor groove compression in the left loxP half-site is responsible for selecting which synaptic configuration is favored, then the loxP-DAP site should de-stabilize complex II (Fig. 1c) due to the insertion of the exocyclic N2-amino group into the narrowed minor groove. Indeed, loxP-DAP synapsis is less efficient than loxP for all three Cre variants (Fig. 7b). Interestingly, the decrease in synapsis affinity is most extreme for the Cre K86A mutant, despite the identical major groove composition in loxP and loxP-DAP.

We made a similar substitution in loxP to give the loxP-GG site, where the A/T base pair at the 4’ position is replaced by G/C. Like loxP-DAP, the loxP-GG site would insert a 2-amino group into the minor groove near the bend center of the left half-site. However, loxP-GG also presents a different functional group for interaction with Lys86 in the major groove. We observed a similar decrease in affinity for wild-type Cre and Cre K201A synthesis of loxP-GG (relative to the loxP site) as was observed for loxP-DAP. However, the decrease in synthesis affinity for Cre K86A-loxP-GG is only about a factor of three relative to loxP. Rather than the factor of eight observed for the loxP-DAP site.

Both the loxP-DAP and loxP-GG sites were predicted to destabilize complex II by perturbing the left half-site sequence that most readily accommodates the bend shown in Fig. 4. We also tested loxP variants that would be predicted to improve the ability of the right half-site to accommodate the bend, leading to a more highly represented top-strand cleavage synaptic complex
The loxAA site contains an A/T base-pair in the 4-position of \textit{loxP}, so that both the 4- and 4'-scissile bases are now adenine (Fig. 6). The lox4 site is a \textit{loxP} variant where the bases in the 4 and 4'-position have been swapped (43). In both cases, wild-type Cre synapses these sites more efficiently than \textit{loxP}, by about a factor of 2. Surprisingly, however, the Cre K201A mutant synopsis affinity is slightly decreased, and the Cre K86A mutant synopsis affinity is decreased by a factor of 2-3 for these modified sites. For both loxAA and lox4, we predicted that complexes I, III, & IV (Fig. 1c) would be more likely to contribute to synopsis observed in solution.

Indeed, the lox4 site shows a complete reversal of preferential strand exchange order, cleaving and exchanging the top strands first during recombination (21). This suggests that synaptic species I is preferred over species II for lox4, although this has not yet been demonstrated experimentally.

The modest decrease in synopsis affinity for the Cre K86A mutant with loxAA and lox4 sites is difficult to explain based on existing structures of the Cre-\textit{loxP} synaptic complex. The Lys86 side-chain interacts differently with scissile-A vs. scissile-G, but it does not appear that the nature of these interactions differs depending on which base is present in the bent vs. unbent half-site. However, it is important to note that a change in measured synopsis affinity could be due to either an altered stability of the synaptic complex, or to a change in stability of the unsynapsed sites, or both. It may be unreasonable to assume that changes in the \textit{loxP} site will only affect the synaptic complex and not the unsynapsed sites.

Unfortunately, there is currently no structural information available for the unsynapsed Cre\textsubscript{2}\textit{loxP} site, or for any corresponding full site among the tyrosine recombinases. However, there is evidence that the synapsed and unsynapsed Cre\textsubscript{2}\textit{loxP} complexes must differ in structure. For example, DNA-bending analysis of unsynapsed \textit{loxP} sites indicates that the bend location and magnitude is different than that observed in structures of the synaptic complex (44). It has also been shown that the top strands are preferentially cleaved in isolated \textit{loxP} sites that have not synapsed, an opposite bias to that observed upon synopsis at the start of the recombination pathway (23). These results imply that the \textit{loxP} site is bent differently and may even be bent in the opposite direction prior to association with a partner site.

**DISCUSSION**

The biochemical studies described here were initially motivated by the need to have a more quantitative understanding of the energetics of \textit{loxP} site synopsis. Although higher order complexes representing synapsed Cre-\textit{loxP} assemblies have been visualized in gel-based assays (9,33), an assay capable of providing binding affinities has not been described. We found analytical ultracentrifugation to be a powerful tool for establishing the basic properties of synopsis in the Cre system, which we were then able to use to develop and validate a more practical gel-based approach. Although widely used to study protein-DNA interactions, the electrophoretic mobility shift assay is not routinely used to study interactions dominated by protein-protein contacts and it was essential to establish whether the gel-based assay accurately reports the equilibrium distributions of species in solution.

The 10 nM \(K_d\) that we measured for synopsis of \textit{loxP} sites explains a number of observations, including those from applications of Cre in DNA manipulations of transgenic systems. For example, Cre will excise genetic loci flanked by \textit{loxP} sites when the sites are separated by megabases of sequence on a eukaryotic chromosome or on different chromosomes in site-directed translocation experiments in (45). The dissociation constant for wild-type Cre also agrees remarkably well with a previously estimated product dissociation \(K_d\) based on kinetic analysis of an intramolecular excision reaction (36).

Our finding that divalent ions are not required for efficient Cre-\textit{loxP} synopsis was somewhat surprising. Although there is a small, positive effect on synopsis for Mg\(^{2+}\), Ca\(^{2+}\), and Mn\(^{2+}\), synopsis occurs efficiently in buffers containing 150 mM NaCl without divalent ions. Initial crystal structures of Cre-\textit{loxP} synaptic complexes revealed that the phosphate backbones of the synapsed \textit{loxP} sites approach each other quite closely, with a minimum distance between non-bridging phosphate oxygen atoms of \(\sim 4\) Å (14). Indeed, the new structures reported here confirm this close contact. Since divalent ions and polyamines were known to stimulate Cre-\textit{loxP}
recombination (3), it seemed reasonable to assume that a cationic species might bridge the close contact between phosphate backbones, thereby explaining this stimulatory effect. Interestingly, replacement of sodium by potassium as the monovalent cation also resulted in modest enhancement of synapsis that was similar in magnitude to the effects of divalent cations.

Our second goal in this work was to establish a structural model for the synaptic complex that forms at the start of Cre-loxP recombination. Since our initial structural models of this reaction intermediate made use of mutants that we later learned were defective in synapsis, we questioned whether these results were biased in some way. The initial studies also made use of a symmetric loxP variant that contained two identical G/C-rich right half-sites (i.e., loxSR), which left open the possibility that the synaptic complex with wild-type loxP could be different. Indeed, a third synaptic complex structure was subsequently described that contained a loxP sequence with non-bridging phosphorothioates at the scissile phosphates (15). This structure revealed a loxP bend that is quite different than that originally observed in the symmetric loxSR-containing complexes. However, the anti-parallel alignment of sites observed in this unique synaptic complex structure corresponds to a top strand cleavage configuration, which existing biochemical data strongly argue is not the preferred starting configuration for Cre-loxP recombination. Thus, despite the reports of three Cre-DNA crystal structures with uncleaved, synapsed sites, a structural model that could clearly explain the biochemical properties of this system was still lacking.

The new Cre-loxP synaptic complex crystal structures described here were formed with the cleavage-deficient, but synapsis-competent Cre K201A mutant. Indeed, the analysis of synapsis efficiencies of Cre mutants was a crucial step in being able to crystallize this intermediate. Three independent representations of the bent loxP site are virtually identical in these structures, revealing a bottom-strand cleavage synaptic complex that explains the existing biochemical data regarding reaction directionality in the Cre system. In particular, the tightly compressed minor groove adjacent to the ‘inactive’ scissile base explains how the identity of the bases in the 4 and 4’ positions can have such a dramatic influence on the directionality of the recombination pathway (21,23). Interestingly, our earlier attempts to determine the structure of a synaptic Cre-loxP complex included experiments using the inactive R173K mutant with wild type loxP sites. However, this combination of Cre mutant and DNA site reproducibly resulted in crystallization of a Cre-Holliday junction complex, where junction DNA formed spontaneously from melting and annealing of loxP ends was trapped by the recombinase (32). HJ formation from loxSR sites was less efficient with the Cre R173K mutant, leading instead to formation of a synaptic complex in spite of the unfavorable combination of a synapse-defective mutant and a poorly synapsing lox site (14).

It is interesting to note that each of the five synaptic Cre-DNA complex crystal structures (three previously described plus two described in this work) reveals a nearly co-planar arrangement of loxP sites. The four duplex segments of the complex that will ultimately become the ‘arms’ of the HJ intermediate are arranged such that their ends trace out a parallelogram. A topological and mathematical analysis of site-specific recombination on circular DNA substrates has indicated that λ-int, Cre, and Flp share an intrinsic chirality of recombination (46). This chirality was interpreted as a right-handed crossing of sites in the synaptic complex that is maintained through the reaction to provide recombinant products with the same handedness. More recently, a topological analysis based on AFM images following in vitro recombination by the Cre and Flp recombinases has supported the chiral recombination mechanism (11). In the latter work, a model was proposed in which a chiral HJ intermediate is responsible for the biased topological outcomes of recombination. In both cases, formation of a right-handed crossing of loxP sites would require that the loxP arms adopt a non-coplanar arrangement. This could be envisioned as a complex where the arms pass through the vertices of a disorted tetrahedron, which need only differ slightly from the planar arrangement observed in crystal structures.

The structural basis for the chirality of Cre-loxP recombination is not apparent from existing structural models of synaptic complexes and HJ intermediates, all of which show a planar
arrangement of DNA arms (17). In principle, the lack of structural insights into this phenomenon could be due to crystal lattice effects that force the complexes into adopting misleading quaternary structures. Since crystals of protein-DNA complexes are often formed from repeating units of continuous DNA duplexes (47), this could be a general source of concern. However, the Cre-\textit{loxP} synaptic complex crystal structures described here, as well as the Cre-HJ intermediate crystal structures that have been described (32,48,49), do not contain continuous DNA duplexes. In fact, there are no DNA-DNA contacts at all in either the orthorhombic or trigonal forms shown in Table III. It therefore seems unlikely that a chiral crossing of \textit{loxP} sites in the Cre-\textit{loxP} synaptic complex has been masked by crystal packing effects in several independent experiments. The source of the observed chirality in integrase family recombination is likely to lie in an intermediate stage of the pathway for which we do not yet have an adequate structural or biochemical understanding.

The work described here advances our understanding of the initial stages of Cre-\textit{loxP} recombination, but these insights also impact related areas of investigation. One such area is the ongoing effort of a number of laboratories to modify Cre and/or the \textit{loxP} site for more optimal use in genetic engineering and DNA manipulation applications. For example, since the initial report that mutant \textit{loxP} sequences exist that will recombine with one another, but not with wild-type \textit{loxP} (50), applications have been designed in a number of recombinase systems that make use of mutually exclusive pairs of sites. The recombinase-mediated cassette exchange reaction is perhaps the most widely used example (51,52). In this context, it is interesting to note that altered \textit{loxP} sites generally recombine less efficiently than wild-type sites (53). Of the \textit{loxP} variants described here, the \textit{loxGG} and \textit{loxAA} sites have been shown to be extremely poor at recombination in vitro, and the \textit{lox4} site has been shown to be moderately defective relative to \textit{loxP} (23).

A simplistic view of the recombination process might be that synapsis efficiency should be a primary limitation of the ability of Cre to recombine \textit{loxP} sites with altered spacer sequences. If synapsis can be achieved, then one might argue that spacer sequences with similar G/C-content should then be able to undergo strand exchange to generate recombinant sites with efficiencies similar to \textit{loxP}. This is clearly not the case. The \textit{loxAA} and \textit{lox4} sites are synapsed more efficiently than \textit{loxP}, but both are poor substrates for recombination. The formation of unproductive dead-end complexes (such as parallel synapses) and the inability to efficiently pass through formation and resolution of the HJ intermediate (19) are likely to be the primary reasons for the poor recombination behavior of \textit{loxAA} and \textit{lox4}, although this remains to be established. In order to understand and anticipate these effects in the design of alternative recombination sequences, it will be important to be able to dissect and understand in energetic terms the individual biochemical steps of the recombination pathway.

REFERENCES

1. Nash, H. A. (1996) in \textit{Escherichia coli and Salmonella: Cellular and Molecular Biology}. (Neidhardt, F. C. e. a., ed), pp. 2363-2376, ASM Press, Washington, D.C.
2. Grindley, N. D., Whiteson, K. L., and Rice, P. A. (2006) \textit{Annu Rev Biochem} 75, 567-605
3. Abremski, K., and Hoess, R. (1984) \textit{Journal of Biological Chemistry} 259, 1509-1514
4. Jayaram, M., Grainge, I., and Tribble, G. (2002) in \textit{Mobile DNA II} (Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A. M., eds), pp. 192-218, ASM Press, Washington DC
5. Azaro, M. A., and Landy, A. (2002) in \textit{Mobile DNA II} (Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A. M., eds), pp. 118-148, ASM Press, Washington D.C.
6. Sherratt, D. J., Soballe, B., Barre, F. X., Filipe, S., Lau, I., Massey, T., and Yates, J. (2004) \textit{Philos Trans R Soc Lond B Biol Sci} 359, 61-69
7. Hamilton, D. L., and Abremski, K. (1984) \textit{Journal of Molecular Biology} 178, 481-486
8. Better, M., Lu, C., Williams, R. C., and Echols, H. (1982) *Proc Natl Acad Sci U S A* **79**, 5837-5841
9. Hoess, R., Wierzbicki, A., and Abremski, K. (1990) in *Structure & Methods* (Sarma, R. H., and Sarma, M. H., eds) Vol. 1, pp. 203-213, Adenine Press, New York
10. Segall, A. M., and Nash, H. A. (1993) *Embo Journal* **12**, 4567-4576
11. Vetcher, A. A., Lushnikov, A. Y., Navarra-Madsen, J., Scharein, R. G., Lyubchenko, Y. L., Darcy, I. K., and Levene, S. D. (2006) *J Mol Biol* **357**, 1089-1104
12. Cassell, G., Moision, R., Rabani, E., and Segall, A. (1999) *Nucleic Acids Research* **27**, 1145-1151
13. Mumm, J. P., Landy, A., and Gelles, J. (2006) *Embo J* **25**, 4586-4595
14. Guo, F., Gopaul, D. N., and Van Duyne, G. D. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 7143-7148
15. Ennifar, E., Meyer, J. E., Buchholz, F., Stewart, A. F., and Suck, D. (2003) *Nucleic Acids Res* **31**, 5449-5460
16. Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005) *Nature* **435**, 1059-1066
17. Van Duyne, G. D. (2001) *Annu Rev Biophys Biomol Struct* **30**, 87-104
18. Grainge, I., Buck, D., and Jayaram, M. (2000) *J Mol Biol* **298**, 749-764
19. Lee, L., and Sadowski, P. D. (2005) *Prog Nucleic Acid Res Mol Biol* **80**, 1-42
20. Hoess, R., Wierzbicki, A., and Abremski, K. (1987) *Proceedings of the National Academy of Sciences of the United States of America* **84**, 6840-6844
21. Lee, L., and Sadowski, P. D. (2003) *J Mol Biol* **326**, 397-412
22. Gelato, K. A., Martin, S. S., and Baldwin, E. P. (2005) *J Mol Biol* **354**, 233-245
23. Ghosh, K., Lau, C. K., Gupta, K., and Van Duyne, G. D. (2005) *Nat Chem Biol* **1**, 275-282
24. Ghosh, K., and Van Duyne, G. D. (2002) *Methods* **28**, 374-383
25. Gill, S. C., and von Hippel, P. H. (1989) *Anal Biochem* **182**, 319-326
26. Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A., and Schuck, P. (2004) *Anal Biochem* **326**, 234-256
27. Otwinowski, Z., and Minor, V. (1997) *Methods in Enzymology* **276**, 307-326
28. Navaza, J. (1994) *Acta Cryst.* **A50**, 157-163
29. Guo, F., Gopaul, D. N., and Van Duyne, G. D. (1997) *Nature* **389**, 40-46
30. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr D Biol Crystallogr* **53**, 240-255
31. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Cryst.* **A47**, 110-119
32. Ghosh, K., Lau, C. K., Guo, F., Segall, A. M., and Van Duyne, G. D. (2005) *J Biol Chem* **280**, 8290-8299
33. Lee, L., and Sadowski, P. D. (2003) *J Biol Chem* **278**, 36905-36915
34. Hensley, P. (1996) *Structure* **4**, 367-373
35. Lebowitz, J., Lewis, M. S., and Schuck, P. (2002) *Protein Sci* **11**, 2067-2079
36. Ringrose, L., Lounnas, V., Ehrlich, L., Buchholz, F., Wade, R., and Stewart, A. F. (1998) *J Mol Biol* **284**, 363-384
37. Rufer, A., Neuenschwander, P. F., and Sauer, B. (2002) *Anal Biochem* **308**, 90-99
38. Sealy, P., and Southern, E. (1990) in *Gel Electrophoresis of Nucleic Acids* (Rickwood, D., and Hames, B., eds), pp. 51-99, IRL Press, Oxford
39. Friesen, H., and Sadowski, P. D. (1992) *Journal of Molecular Biology* **225**, 313-326
40. Wierzbicki, A., Kendall, M., Abremski, K., and Hoess, R. (1987) *Journal of Molecular Biology* **195**, 785-794
41. Van Duyne, G. D. (2002) in *Mobile DNA II* (Craig, N. L., Craige, R., Gellert, M., and Lambowitz, A. M., eds), pp. 93-117, ASM Press, Washington DC
42. Huffman, K. E., and Levene, S. D. (1999) *J Mol Biol* **286**, 1-13
43. Lee, L., and Sadowski, P. D. (2001) *J Biol Chem* **276**, 31092-31098
44. Lee, L., Chu, L. C., and Sadowski, P. D. (2003) *J Biol Chem* **278**, 23118-23129
45. Sauer, B. (2002) in *Mobile DNA II* (Craig, N. L., Craigie, R., Gellert, M., and Lambowitz, A. M., eds), pp. 38-58, ASM Press, Washington D.C.
46. Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W., and Cozzarelli, N. R. (1999) *J Mol Biol* **289**, 747-775
47. Tan, S., Hunziker, Y., Pellegrini, L., and Richmond, T. J. (2000) *J Mol Biol* **297**, 947-959.
48. Gopaul, D. N., Guo, F., and Van Duyne, G. D. (1998) *EMBO J.* **17**, 4175-4187
49. Martin, S. S., Pulido, E., Chu, V. C., Lechner, T. S., and Baldwin, E. P. (2002) *J Mol Biol* **319**, 107-127
50. Hoess, R. H., Wierzbicki, A., and Abremski, K. (1986) *Nucleic Acids Research* **14**, 2287-2300
51. Bouhassira, E. E., Westerman, K., and Leboulch, P. (1997) *Blood* **90**, 3332-3344
52. Seibler, J., and Bode, J. (1997) *Biochemistry* **36**, 1740-1747
53. Lee, G., and Saito, I. (1998) *Gene* **216**, 55-65
54. Delano, W. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos

**FOOTNOTES**

* We are grateful to Ron Hoess, Scott Pesiridis, Philip Dreyfuss, and G.V. lab members for insightful discussions and comments regarding this work. Supported by NIH grant GM55041 to G.V. G.V. is an Investigator of the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

**FIGURE LEGENDS**

Figure 1. Synapsis of *loxP* sites by Cre recombinase. (a) Sequence and organization of the *loxP* site. The 8-bp spacer between inverted repeats is in upper case and is numbered. Left and right half-sites and the location of cleavage sites are indicated. (b) Synapsis of *loxP* sites. Two Cre-bound *loxP* sites (monomers) associate to form a synaptic complex (dimer) to begin recombination. (c) Four ways to form a synaptic complex between Cre-bound *loxP* sites. Complexes I and II have the sites arranged in an anti-parallel alignment and complexes III and IV are parallel alignments. If complex I forms to initiate recombination, the top (red) strands are cleaved and exchanged to form the HJ intermediate. If complex II forms to initiate recombination, the bottom (black) strands are exchanged first. Arrows indicate the direction of the site, as defined in (a). Red and black lines represent the top and bottom strands of *loxP*, respectively, as defined in (a).

Figure 2. Sedimentation equilibrium analytical ultracentrifugation analysis of wild type Cre-*loxP* synapsis. Three concentrations of Cre-*loxP* complex (150, 300, 600 nM) were centrifuged at the four rotor speeds shown until an equilibrium radial distribution of unsynapsed and synapsed complex was achieved. The synapsis equilibrium constant was determined from a global fit of the 12 resulting radial absorbance scans at 260 nm. (a) Radial absorbance at equilibrium for 600 nM *loxP*. Solid lines are from the global fit to a monomer-dimer equilibrium model. (b) Residuals for the fits shown in (a). Fits and residuals for all three concentrations are provided in Supplemental Fig. 2.

Figure 3. Electrophoretic mobility shift assay to monitor wild type Cre-*loxP* synapsis. (a) Example of a synapsis titration with increasing concentration of unlabeled *loxP* site. Cre is present at saturating concentrations in order to ensure full occupancy of *loxP* sites at all *loxP* concentrations. (b) Fit of a monomer-dimer equilibrium model based on quantitation of the bands shown in (a). The fraction synapsed and estimated errors were calculated from
at least three independent titrations for wild type Cre-loxP synapsis and for the results summarized in Table II and Fig. 7.

Figure 4. Structure of the pre-cleavage Cre-loxP synaptic complex. (a) Overall structure, as viewed from the face of the complex where the N-terminal recombinase domains interact. This structure corresponds to complex II in Fig. 1c. Arrows indicate the loxP direction, as defined in Fig. 1. Asterisks indicate the recombinase subunits that are poised to catalyze strand cleavage and exchange. These recombinase subunits are bound to the loxP half-sites that do not contain the sharp DNA bend associated with synapsis. (b) Closeup of the loxP spacer corresponding to the boxed region in (a). The top and bottom strands (as defined in Fig. 1) are labeled with red and black fonts, respectively. The major groove opening and minor groove compression at the kink in the loxP site are indicated. (c) Closeup of the DNA segment shown in (b), as viewed along the compressed minor groove. The 2-positions of Ade-3' and Ade-4' are indicated with arrows. This figure was prepared with PyMol (54).

Figure 5. Asymmetric interactions with the scissile base-pairs of loxP. The unique interactions between Lys86, Lys201 and the base-pairs in the 4'- and 4-positions of loxP are illustrated schematically. The contacts shown for Lys86 are observed in the structures described here. The contacts shown for Lys201 are based on structures of Cre-DNA reaction intermediates in which this residue has not been mutated (17) and are expected to be present in the wild-type Cre-loxP synaptic complex that initiates recombination.

Figure 6. Core sequences of modified loxP sites used in synapsis assays. The loxSR site contains two right half-sites (defined in Fig. 1a) arranged as inverted repeats. The loxSL site contains two left half-sites arranged as inverted repeats. The loxP-DAP site differs from loxP only by the addition of a 2-amino group to Ade-4'. LoxGG and loxA4 are single base-pair changes in the top and bottom strand scissile bases of loxP, respectively. Lox4 is a mutant loxP in which the 4 and 4'-scissile base pairs are swapped. Coloring and upper/lower case follows that defined in Fig. 1a. Bases that are changed relative to the wild-type loxP site are boxed. The full sequences of DNA duplexes used in the experiments are given in Supplemental Fig. 1.

Figure 7. Synapsis of modified loxP sites by wt-Cre, Cre K201A, and Cre K86A. (a) Synapsis of symmetric lox sites. (b) Synapsis of singly substituted loxP sites. Core sequences of the modified sites are given in Fig. 6. Dissociation constants were measured using the EMSA method as shown in Fig. 3.
Table I. Effects of pH and divalent ions on Cre-\textit{loxP} synapsis.

| Buffer            | Monovalent | Divalent    | Relative $K_d$ |
|-------------------|------------|-------------|----------------|
| 20 mM Hepes, pH 7.5 | 150 mM NaCl | 5 mM MgCl$_2$ | 1.0            |
| 20 mM Na Hepes, pH 7.5 | 150 mM KCl | 5 mM MgCl$_2$ | 0.4            |
| 20 mM Hepes, pH 7.5 | 150 mM NaCl | none        | 1.7            |
| 20 mM Acetate, pH 5.5 | 150 mM NaCl | 5 mM MgCl$_2$ | 0.6            |
| 20 mM Hepes, pH 6.5 | 150 mM NaCl | 5 mM MgCl$_2$ | 0.9            |
| 20 mM Tris, pH 8.5 | 150 mM NaCl | 5 mM MgCl$_2$ | 8              |
| 20 mM Hepes, pH 7.5 | 150 mM NaCl | 5 mM CaCl$_2$ | 1.3            |
| 20 mM Hepes, pH 7.5 | 150 mM NaCl | 5 mM MnCl$_2$ | 0.8            |

$K_d$ values were determined by sedimentation equilibrium ultracentrifugation of 400 nM Cre$_2\textit{loxP}$ at four rotor speeds as described in Experimental Methods. $K_d$ values are relative to the standard buffer condition, which was also used for the experiment in Fig. 2. All buffers also contained 1 mM dithiothreitol. Sodium was used as counterion for Hepes and Acetate buffers and chloride was used for Tris.
Table II. Synapsis of \textit{loxP} sites by Cre mutants

| Mutant   | Description                                                                 | Synapsis \(K_d\) (nM) |
|----------|-----------------------------------------------------------------------------|------------------------|
| Wild-type|                                                                              | 14 ± 1                 |
| K201A    | Active site mutant; no cleavage/recombination                                | 8.9 ± 0.6              |
| K86A     | Contacts scissile base; wild-type activity                                   | 8.5 ± 1.4              |
| H289A    | Active site mutant; slower recombination                                     | 15 ± 1                 |
| R173K    | Active site mutant; no cleavage/recombination                                | 130 ± 30               |
| Y324F    | Active site mutant; no cleavage/recombination                                | > 200 nM               |
| A36V     | Protein-protein interface formed during synapsis; slower recombination;      | > 500 nM               |
| ΔC331    | C-terminal helix deletion; no cleavage/recombination; deficient in synapsis   | No synapsis observed   |

*\(K_d\) values were determined by the gel-based assay, as described in Methods.
Table III. Crystallographic data and results for Cre-loxP Synaptic complexes.

|                      | C222₁       | P3₂21       |
|----------------------|-------------|-------------|
| Space group          |             |             |
| Unit cell (a, b, c, Å)| 107.6, 122.1, 178.7 | 136.8, 136.8, 218.3 |
| X-ray source         | ALS 8.2.1   | ALS 8.2.1   |
| Resolution (Å)       | 30.5 – 2.4  | 19.9 – 2.6  |
| Completeness (%)a    | 98.7 (94.3) | 81.4 (81.6) |
| Rsym b               | 5.5 (33.2)  | 9.0 (19.8)  |
| Rwork c              | 18.7        | 18.0        |
| Rfree d              | 25.7        | 24.3        |
| R.m.s.d. bond length (Å)| 0.025   | 0.007       |
| R.m.s.d. bond angles (deg)| 2.674 | 1.195       |

*a Values in parentheses refers to the highest resolution shell.

$b R_{sym} = \frac{\sum_h |I_h - \langle I_h \rangle|}{\sum I_h}$, where $\langle I_h \rangle$ is the average intensity over the symmetry equivalents.

$c R_{work}$ includes 95% of the reflection data used in refinement.

$d R_{free}$ includes 5% of the reflection data excluded from refinement.
Figure 1, Ghosh et al.

(a) Schematic representation of the TS and BS sequences with their respective cleavage sites.

(b) Diagram showing the molecular weights of the cleavage products.

(c) Comparison of TS-, BS-, and TS/BS-cleavage with anti-parallel and parallel orientations.
Figure 2, Ghosh et al.

Absorbance (260 nm) vs. radius (cm)

K_d = 10.4 ± 0.3 nM

Legend:
- 6k
- 9k
- 12k
- 15k
Figure 3, Ghosh et al.

(a) 

(b) 

K_d = 14 ± 1 nM
Figure 4, Ghosh et al.
Figure 5, Ghosh et al.

Ade4'  

DAP4'  

Gua4

Lys201

NH₃⁺  

Lys₈₆

NH₃⁺  

Lys₂₀₁

NH₃⁺  

Lys₈₆

NH₃⁺
| Site   | 5'  | 3'  |
|--------|-----|-----|
| loxP   | ...tataATGTATG | ...atatTACATACG |
| loxSR  | ...tataSCATATG | ...atatCGTATACG |
| loxSL  | ...tataATGTACAT | ...atatTACATGA |
| loxP-DAP | ...tataATGTATG | ...atatTACATACG |
| loxGG  | ...tataATGTATG | ...atatACATACG |
| loxAA  | ...tataATGTATG | ...ATATTACATAC |
| lox4   | ...tataATGTATG | ...atatACATACG |
Figure 7, Ghosh et al.
