Published online 12 May 2010

Nucleic Acids Research, 2010, Vol. 38, No. 17 ***5817–5832***
doi:10.1093/nar/gkq384

Requirements for catalysis in the Cre recombinase active site

Bryan Gibb, Kushol Gupta, Kaushik Ghosh, Robert Sharp, James Chen and Gregory D. Van Duyne*

Department of Biochemistry and Biophysics and Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

Received April 7, 2010; Revised April 26, 2010; Accepted April 27, 2010

ABSTRACT

Members of the tyrosine recombinase (YR) family of site-specific recombinases catalyze DNA rearrangements using phosphoryl transfer chemistry that is identical to that used by the type IB topoisomerases (TopIBs). To better understand the requirements for YR catalysis and the relationship between the YRs and the TopIBs, we have analyzed the *in vivo* and *in vitro* recombination activities of all substitutions of the seven active site residues in Cre recombinase. We have also determined the structure of a vanadate transition state mimic for the Cre–loxP reaction that facilitates interpretation of mutant activities and allows for a comparison with similar structures from the related topoisomerases. We find that active site residues shared by the TopIBs are most sensitive to substitution. Only two, the tyrosine nucleophile and a conserved lysine residue that activates the 5′-hydroxyl leaving group, are strictly required to achieve >5% of wild-type activity. The two conserved arginine residues each tolerate one substitution that results in modest recombination activity and the remaining three active site positions can be substituted with several alternative amino acids while retaining a significant amount of activity. The results are discussed in the context of YR and TopIB structural models and data from related YR systems.

INTRODUCTION

Members of the tyrosine recombinase (YR) family of site-specific recombinases catalyze a variety of sequence-specific DNA rearrangements in biological systems, including the integration and excision of phage genomes into and out of their bacterial hosts (1). The phage λ integrase (2), the yeast Flp recombinase (3), the phage P1 Cre recombinase (4) and the *Escherichia coli* XerC/XerD recombinases (5) are examples of well-studied YR systems. The YR enzymes bind to two distinct, specific sequences (e.g. *attP* and *attB* for λ integrase; two *loxP* sites for Cre) and catalyze a stepwise exchange of strands between the two sites to generate recombinant products (6).

The site-specific recombination reaction begins with cleavage of the recombining sites by conserved tyrosine residues to form covalent 3′-phosphotyrosine linkages to the DNA substrates and release free 5′-hydroxyl groups (Figure 1A). Reversal of this cleavage reaction restores the starting configuration and completes a cycle of phosphoryl transfer chemistry that is identical to that performed by type Ib topoisomerases (TopIBs; 7). Alternatively, the strands bearing free 5′-hydroxyl groups can be exchanged between recombining sites and subsequent ligation results in formation of a four-way Holliday junction (HJ) intermediate. The HJ intermediate can then serve as the substrate for a second round of cleavage, strand exchange and ligation steps. If the same strands are processed to resolve the junction, the starting sites are re-generated. If the HJ intermediate isomerizes to allow processing of the opposite pair of DNA strands, recombinant products are formed.

The YR active site is composed of seven conserved residues (Table 1). In addition to the tyrosine nucleophile, one lysine and two arginine residues are thought to be essential for efficient recombination. Two histidine residues are also found in most YR active sites; one of these (His/Trp III) is tryptophan in the Cre and Flp recombinases. The seventh active site residue is either Glu or Asp in the YR enzymes. A comparison of ∼100 tyrosine recombinases in 1998 revealed aspects of the protein sequences that were strongly conserved, that were similar and that differed among family members

*To whom correspondence should be addressed. Tel: +1 215 898 3058; Fax: +1 215 573 4764; Email: vanduyne@mail.med.upenn.edu
Present address:
Kaushik Ghosh, Syngene International Ltd, Jigani Link Rd, Bangalore, India.

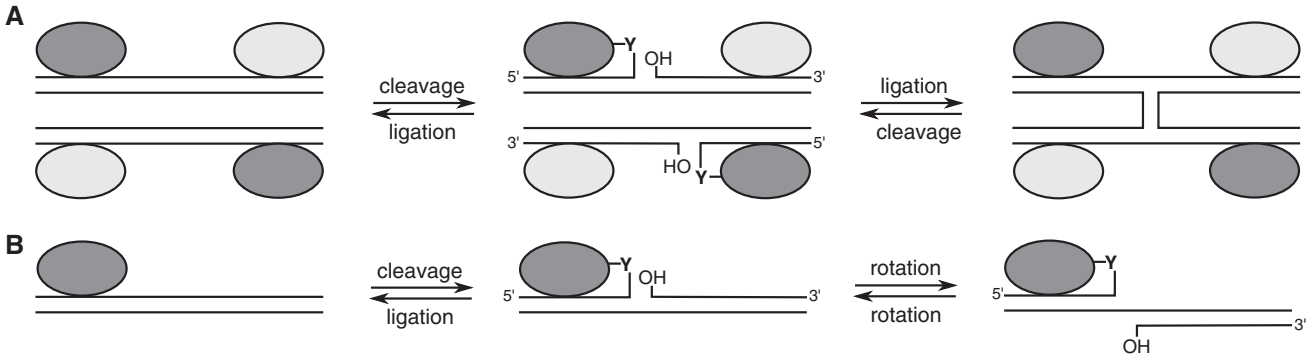

Figure 1. Tyrosine recombinases and type IB topoisomerases. (**A**) Tyrosine recombinase strand-exchange mechanism to generate a HJ intermediate. The HJ intermediate can be converted to starting substrates by exchange of the same strands or to recombinant products by exchange of the second pair of strands by the remaining two protein subunits. (**B**) The cleave–rotate–ligate mechanism of supercoil relaxation used by type IB topoisomerases. The two enzyme families use nearly identical mechanisms to catalyze the phosphoryl transfer reactions.

Table 1. Tyrosine recombinase and topoisomerase IB active site residues

	Arg I	Glu/Asp I	Lys	His II	Arg II	His/Trp III	Tyr
P1 Cre	R173	E176	K201	H289	R292	W315	Y324
λ Int	R212	D215	K235	H308	R311	H333	Y342
S. cerevisiae Flp	R191	D194	K223	H305	R308	W330	Y343
E. coli Xer C/D	R148	D151	K172	H240/244	R243/247	H243/247	Y275/279
Poxvirus TopIB[a]	R130	None	K167	None[b]	R223	H265	Y274
Conservation[c]	91.2% R 1.1% K	79% E 16.4% D	94.7% K 3.8% R	87.3% H 6.8% Y	99.1% R 0.5% K	92% H 4% W	99.6% Y
Functions[d]	TS	S	GAB	TS, GAB	TS	S, TS	N

[a]Residue numbers refer to the vaccinia/variola virus topoisomerases.
[b]A lysine residue is located in this position of the sequence in the TopIB enzymes, but is directed away from the active site.
[c]The values given are based on analysis of PSI-BLAST hits from the NCBI RefSeq database using the Cre catalytic domain as query; the percentages for the catalytic lysine are estimates, as described in 'Materials and Methods' section.
[d]Roles of active site residues in catalysis, based on published biochemical and structural data and on the work described here.
TS, transition state stabilization; S, structural; GAB, general acid/base catalysis; N, nucleophile.

(8). To obtain an updated estimate of the degree of conservation of the active site residues that includes data from microbial genomes sequenced over the past decade, we analyzed over 12 000 PsiBlast (9) hits from the National Center for Biotechnology Information (NCBI) RefSeq protein database (10) using the Cre recombinase catalytic domain as a query sequence. As shown in Table 1, conservation ranges from the more variable box II histidine (87%) to the highly conserved catalytic tyrosine (99.6%) and box II arginine (99.1%). Two residues (His/Trp III and Glu/Asp I) are present as one of two alternative amino acid types in the YR enzymes.

The properties of a number of YR active site mutants have been described in the literature, including their ability to catalyze recombination *in vivo* or *in vitro* and to carry out specific steps of the recombination pathway such as cleavage, ligation, and synapsis of the recombining sites. We have compiled a summary of these results in Table S1 that augments a similar compilation reported by Nunes-Duby *et al*. (8). In addition to the biochemical studies that have been reported for YR systems, a number of high-resolution structural models are available that provide snapshots of the enzyme active sites at specific steps of recombination (11).

As shown in Figure 1B, TopIBs use a similar strategy of cleavage with an active site tyrosine to form a 3′-phosphotyrosine linkage that allows strand rotation and relaxation of supercoils. This similarity also extends to the level of structure and catalytic mechanism, since the TopIB enzymes have similar catalytic domain folds and share a subset of the YR active site residues (12). Indeed, four of the seven active site residues listed in Table 1 are strictly conserved among the TopIBs. The topoisomerases encoded by poxviruses are type IB enzymes and the vaccinia virus TopIB has been extensively studied (7). Investigations into the function of poxvirus TopIB have not only been instrumental in developing a mechanistic understanding for this family of topoisomerases but also had an important impact on understanding the YR enzymes. As is the case for the YRs, there are now several high-resolution crystal structures available for TopIBs bound to DNA substrates, providing a corresponding set of snapshots of the enzyme active sites at different steps in the phosphoryl transfer reaction. One such model is a vanadate transition state mimic for both trypanosomal and variola virus TopIB (13,14) which has provided additional insight into the nature of TopIB catalysis. A corresponding transition state mimic structure for a YR system has not yet been reported.

To extend our structural views of the YR active site, we have determined the structure of a Cre–DNA–vanadate transition state mimic, where the active site phosphate has been replaced by pentavalent vanadium. The transition state mimic structure allows a more informative interpretation of active site mutant activities and also allows for a detailed comparison with similar structures from the TopIB family. To improve our understanding of the requirements for catalysis in the tyrosine recombinases and the differences with the closely related TopIB enzymes, we have created a library of active site mutants for Cre recombinase where each of the seven residues listed in Table 1 has been replaced by all 19 alternatives. Here, we report the *in vivo* and *in vitro* recombination activities of these mutants. We find that despite the high degree of conservation, many active site substitutions retain a significant amount of recombination activity, including one substitution of an active site arginine previously thought to be essential for efficient catalysis.

MATERIALS AND METHODS

Reagents and strains

Reagents were obtained from Sigma, unless otherwise noted. Strains CSH100 (*F' lacproA+,B+(lacI^q lacPL8)/ara Δ(gpt-lac)5*) and CSH142 (*F-/ara Δ(gpt-lac)5*) were obtained from the *E. coli* Genetic Stock Center (CGSC8105 and CGSC8083) at Yale University. Plasmids pFW11 and pTSA29 were gifts from Dr. Fred Whipple and Dr. Gregory Phillips, respectively.

Crystallization and diffraction data

Cre was over-expressed and purified as previously described (15). Oligonucleotides were synthesized by the Keck Biotechnology Facility at Yale University and purified by ion-exchange chromatography. Sodium orthovanadate (Sigma) was activated as described (16) and added to a premixed solution containing 40 μM Cre, 75 μM half-site DNA, 20 mM $CaCl_2$, 20 mM sodium acetate, pH 5.6, 1 mM dithiothreitol and 25% 2-methyl-2,4-pentanediol (MPD) to a final concentration of 2 mM. Crystals were grown by hanging drop vapor diffusion against a well reservoir of 20 mM sodium acetate, pH 5.0, 20 mM $CaCl_2$ and 30–45% MPD. Crystals were flash-frozen in liquid nitrogen prior to data collection. Diffraction data were measured at the Advanced Light Source beamline 8.2.1 and processed using HKL2000 (17).

Structure solution and refinement

The structure was determined by molecular replacement with AMORE (18), using a search model based on the pre-cleavage synaptic complex (pdb entry 2HOI) where crossover region DNA and all active site residues had been removed. Initial rounds of refinement were performed with CNS (19) following model building and adjustment with COOT (20). Final refinements were performed with REFMAC (21). Active site residues, the vanadate linkage and omitted DNA residues were fit into clear, unbiased electron density. Vanadate geometry was restrained during refinement as described for the viral TopIB–DNA complex (14). Coordinates for the refined structure have been deposited with the Protein Data Bank with accession code 3MGV.

Generation of active site mutants

An XbaI–HindIII fragment containing the wild-type Cre coding sequence was cloned into the same sites of pBluescriptIIKS+ (Stratagene) and used as a template for polymerase chain reaction (PCR)-based mutagenesis using a modified version of the Phusion protocol (New England Biolabs; NEB). Most substitutions were generated using primers with degenerate codons and those remaining following initial sequencing were obtained using specific primers. In both cases, only the primer containing the altered active site codon was phosphorylated during oligonucleotide synthesis (Integrated DNA Technologies; IDT) and a 1-h DpnI digestion following ligation of the PCR products was used to remove the template. All mutant plasmids were sequenced twice to verify the correct substitution.

In vivo recombination reporter

A single copy reporter was generated by homologous recombination of an intermediate plasmid based on pFW11 into an F' containing an intact *lac* operon (22). A modified *sulA* promoter (23) was cloned into the EcoRI and BamHI sites of pFW11 using oligonucleotides P1 and P2 (Table S2) and a fragment containing the *aadA* coding sequence with a 5′-KpnI site was amplified using primers P3 and P4 and ligated into the BamHI and SalI sites of sulA-FW11 to give pGV1026. A terminator cassette that could be excised by Cre was generated by inserting the *E. coli rpoC* terminator (24) amplified using primers P5 and P6 into the BsrGI and HincII sites of pGV697, a pBluescript derivative containing directly repeating *loxP* sites flanking an 867-bp segment (GV, unpublished data), to generate pGV1025. The resulting construct contains *loxP* sites flanking a 920-bp segment containing *TrpoC*. The excisable element was then inserted into the BamHI–KpnI sites of pGV1026 to give pGV1035. pGV1035 was then transformed into CSH100 and double crossovers were selected following transfer of the recombinant episome into CSH142/pTSA29 by conjugation as described (22). pTSA29 provides ampicillin resistance and a temperature-sensitive pSC101 origin, allowing facile removal following incubation at 42°C (25). The relevant segment of the resulting F' (pGV1048) is shown schematically in Figure 4A.

In vivo recombination

Cre mutant plasmids were transformed into chemically competent CSH142/pGV1048 cells and after a one hour incubation at 37°C were plated on LB containing ampicillin (100 μg/ml), kanamycin (35 μg/ml) and X-gal (20 μg/ml). An equal volume was plated on LB containing the same components plus 20 μg/ml streptomycin. Activity was recorded as the ratio of the number of colonies after 16 h at 37°C on the streptomycin plate divided by

the number on the plate lacking streptomycin. For wild-type Cre, this ratio was reproducibly close to 1, with a colony count of ~1000. For the empty pBluescript and Cre Y324F negative controls, the background was zero; no streptomycin-resistant colonies were observed in multiple experiments and colonies on the non-selection plate were white. For weakly active mutants, where a small number of colonies were observed on the streptomycin plate, *in vivo* activity was corroborated by the appearance of blue colonies in the absence of streptomycin selection. *In vivo* assays were repeated three to four times and the results were tabulated as a percentage of wild-type activity with typical standard deviations of 9% for the most weakly active mutants.

In vitro recombination assay

The substrate for *in vitro* recombination was a linear 602-bp DNA fragment digested from pBG1701, dephosphorylated and ^{32}P end-labeled with polynucleotide kinase. pBG1701 was constructed by ligation of a DNA segment containing two *loxP* sites flanking a 334-bp spacer, shown schematically in Figure 5A. Cre-containing lysates were prepared from CSH142/pGV1048 transformants by sonication of cells from 2 ml culture grown at 37°C to an optical density of ~1.0 in 100 µl lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 2 mM dithiothreitol, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.5% Triton-X100 and protease inhibitors). Cleared lysates were diluted in lysis buffer and used directly for *in vitro* binding and recombination assays. Recombination reactions were carried out in 25 µl assay buffer (50 mM Tris pH 7.4, 150 mM NaCl, 2 mM dithiothreitol, 1 mM EDTA, 50 µg/ml sheared salmon DNA) containing 200 pM labeled substrate and a 50-fold dilution of lysate. After incubation at 37°C for 1 h, reactions were quenched with 0.5% sodium dodecyl sulfate (SDS), heated at 70°C for 1 min and digested with proteinase K for 1 h at 37°C. Reactions were analyzed on a 6% acrylamide gel in 1 × TBE for 2 h at 10 V/cm. Gels were dried and quantified on a Phosphorimager. Errors for the *in vitro* recombination assays were estimated at 5–10% of wild-type activity based on multiple repetitions of a subset of weak and moderately active mutants.

In vitro loxP binding

The probe for electrophoretic mobility shift assays (EMSAs) was a 64-mer duplex containing the *loxP* site flanked by 15-bp G·C-rich segments. Ten picomoles of polyacrylamide gel electrophoresis (PAGE)-purified top strand were ^{32}P end-labeled with polynucleotide kinase and annealed to a purified, unlabeled bottom strand. Cleared lysate was added (50:1 final dilution) to 200 pM labeled substrate in assay buffer and the reactions were incubated for 30 min at 20°C. Binding reactions were electrophoresed on a 6% polyacrylamide gel (29:1 acrylamide:bisacrylamide) in TBE that was pre-run for 10 V/cm for 45 min at 20°C prior to loading. The gel was run at 13 V/cm for 2 h at 20°C, dried, and analyzed with a phosphoimager.

Tyrosine recombinase sequence alignments

The Cre catalytic domain (residues 134–343) was used as a query sequence for three iterations of PSI-BLAST (9) searching of the NCBI RefSeq database (10) and 12 049 alignments (maximum *E*-value = 0.01) that included all seven active site residues were analyzed by quantifying the amino acid frequency at each of the seven positions. The catalytic lysine is not well aligned due to its position in a loop of variable size and sequence. For this residue, we counted lysine (or arginine) as present at this site if it was located within five residues of Cre Lys201 in the alignment. We obtained similar results when the λ integrase catalytic domain was used as a search query.

RESULTS

Structure of a Cre–*loxP* transition state mimic

To trap a vanadate transition state mimic for the Cre–*loxP* reaction, we used a symmetric version of the *loxP* site where the scissile phosphates are missing in both the top and bottom strand cleavage positions (Figure 2A). We chose to symmetrize the more G·C-rich half-site of *loxP* for these experiments [resulting in a site which we refer to as loxS1 (26)], which provides a transition state model depicting the DNA functional groups that would be present in the initial cleavage step of the reaction (27), as opposed to the HJ-resolution step later in the pathway. When activated potassium orthovanadate was added to the Cre–loxS1 complex, one VO_2^+ equivalent was incorporated into the active site and covalent linkages were formed with O3′, O5′ and Tyr324 as observed in the related TopIB transition state mimic structures (13,14).

Crystals of the Cre–loxS1–vanadate complex were grown from hanging drops using conditions similar to those used previously for several Cre–DNA structures (28–30) and diffracted to 2.3 Å resolution at synchrotron sources. The structure was determined by molecular replacement using the synaptic Cre–DNA complex (26) lacking active site residues as a search model. A summary of diffraction data and refinement results is given in Table 2 and electron density within the active site following refinement is shown in Figure 2B.

The overall structure is similar to that observed in the pre-cleavage and covalent intermediate Cre–DNA structures previously described (26,30), with a tetramer of interlocked Cre subunits bound to two bent DNA substrates (Figure 2C and D). With the exception of the active sites, the protein subunits in the current structure superimpose well with those in the synaptic complex and covalent intermediate structures, with no significant differences observed. Our discussion of the transition state mimic structure will therefore focus only on the active site of the enzyme.

In principle, vanadate could be incorporated into each of the four active sites present in the Cre–loxS1 complex shown in Figure 2C. In the Cre–DNA reaction intermediate structures that have been reported, two of the active sites in the Cre tetramer adopt 'active' configurations that

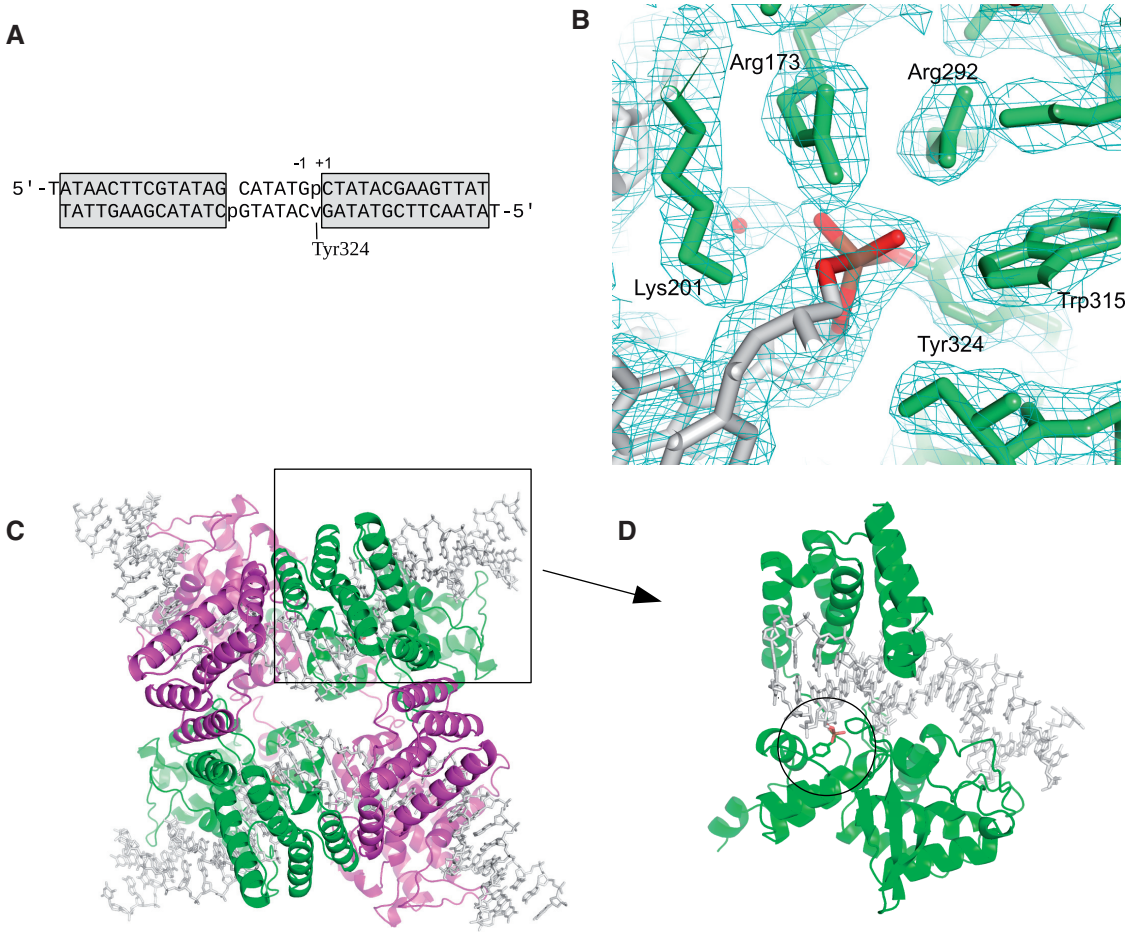

Figure 2. Structure of Cre–DNA–vanadate transition state mimic. (**A**) Sequence of the DNA substrate used, with vanadium incorporation indicated by 'v'. The 14-bp recombinase binding elements that surround the central 6-bp crossover region are shaded. The '+1' and '–1' base pair labels are used to facilitate comparison with related TopIB structures. (**B**) Weighted $2F_o$–F_c electron density following refinement at 2.3 Å resolution. The map is contoured at 1.2σ. Glu176 and His289 are not visible in this view. (**C**) Overall structure of the synaptic complex, viewed from the N-terminal domains of the recombinase subunits. The activated subunits containing vanadate linkages are colored green. (**D**) View of a single Cre subunit bound to a *loxP* half-site within the tetrameric complex, rotated from the boxed subunit in (C). The tyrosine linkage to vanadate is circled.

are poised for catalysis of strand exchange (referred to here as the 'activated sites') and the other two are in 'inactive' configurations ('inactive sites') that are not competent to catalyze cleavage of the DNA (31). In the current structure, vanadium is incorporated only into the activated sites, reflecting the inability of the remaining two subunits to stabilize a transition state configuration. With the exception of Tyr324 and Lys201, the protein and DNA residues involved in coordinating vanadium in the activated sites (Figure 3) adopt similar positions in the inactive sites of the Cre–loxS1 complex. Tyr324 is shifted by 3.5 Å in the inactive site relative to the activated site position and Lys201 is withdrawn from the minor groove. These alternative configurations for Tyr324 and Lys201 are typical of those observed in the inactive sites of other Cre reaction intermediates and we assume that these effects are also responsible for preventing stable incorporation of penta-coordinate vanadium in the inactive sites of the Cre–DNA complex described here.

In the activated sites, vanadium is coordinated by five oxygen atoms in a trigonal bipyramidal arrangement with Tyr324-Oη and O5′ in the axial positions and with O3′ and the two non-bridging oxygen atoms in equatorial positions (Figure 3). The two conserved arginine residues each make double hydrogen bonding interactions; Arg173 hydrogen bonds to O5′ and the non-bridging oxygen corresponding to O2P and Arg292 hydrogen bonds to the Tyr324 phenolic oxygen and to O1P. Lys201 is within close hydrogen bonding distance of four groups: O5′ of the −1 sugar, N3 of the +1 guanine base, O4′ of the −1 sugar and a tightly bound water molecule. Each of these features is also found in the trypanosomal and viral TopIB transition state structures (13,14), supporting the close mechanistic and evolutionary relationship between the two enzyme families. Several other aspects of the Cre active site differ from the TopIB enzymes and will be discussed below in the context of the specific residues.

Comprehensive analysis of Cre active site substitutions

A number of active site mutants have been previously described for YR and TopIB enzymes (summarized in

Table 2. Crystallographic data for the Cre–DNA–vanadate transition state mimic

Diffraction data	
Resolution (Å)	2. Å
Space group	$P3_221$
Unit cell	a = 136.0 Å c = 218.4Å
Mosaicity (°)	0.53
Completeness (%)	97.2 (95.4)
R_{sym}	0.103 (0.462)
Total reflections	775 986
Unique reflections	102 079
I/σ	16.17 (4.15)
Redundancy	7.6 (7.3)
Refinement	
R_{free}	0.222 (0.285)
R_{work}	0.193 (0.238)
Number of atoms	13 407
Protein	10 204
DNA	2859
Water	346
Average B factors (Å^2)	
Protein	27.8
DNA	36.8
Water	25.6
Rmsd	
Bond lengths (Å)	0.01
Bond angles (°)	1.62

Values in parentheses refers to the highest resolution shell (2.38–2.30 Å).
$R_{sym} = \sum_h | I_h - \langle I \rangle_h | / \sum I_h$, where $\langle I \rangle_h$ is the average intensity over the symmetry equivalents.
R_{work} includes 95% of the reflection data used in refinement.
R_{free} includes 5% of the reflection data excluded from refinement.

Table S1). Even when the data for all YR enzymes are combined, however, the substitutions that have been studied for a given active site residue number only 6–7 at most out of the 19 alternatives and, in some cases, relatively conservative substitutions have not been reported. To establish which Cre active site substitutions would support recombination, we generated all single site variants for the residues shown in Table 1 and measured their ability to catalyze recombination *in vivo* and *in vitro*.

To test for recombination activity in bacterial cells, we designed the single copy reporter shown schematically in Figure 4A. The *aadA* streptomycin resistance and *lacZYA* genes located on an F′ episome are transcribed from a strong promoter when an intervening transcriptional terminator has been excised by Cre–*loxP* recombination, but full-length transcripts are only produced at low levels when the terminator is present. Cells that are Lac$^-$ and harbor the reporter F′ will be sensitive to streptomycin and Lac$^-$ in the absence of active Cre but will become streptomycin resistant and Lac$^+$ when active Cre is present. For the experiments described here, we used streptomycin resistance and β-galactosidase activity to score for Cre-mediated site-specific recombination. Growth on minimal lactose plates can also be used and is more effective in identifying mutants with very low activity.

To assay for *in vivo* Cre activity, we transformed *E. coli* strain CSH142 containing the reporter F′ with a plasmid containing a Cre coding sequence (described below) and compared the number of colonies that formed on plates containing 20 μg/ml streptomycin to the number that formed on plates lacking streptomycin. Approximately 45 min is required for expression of the *aadA* gene (data not shown). A 60-min outgrowth prior to plating therefore provided a ~15-min time window for recombination to occur. To facilitate *in vitro* mutagenesis and limit *in vivo* expression, we cloned the wild-type Cre coding sequence into pBluescriptIIKS+ where it is oriented opposite that of the *lac* and *bla* promoters. Due to the high plasmid copy number, this promoter-less construct still provides moderate levels of constitutively expressed Cre. When we compared Cre expression from this plasmid to basal Cre expression from uninduced pET21a (T7/lac promoter) and to a P1 lysogen by western blotting, we were not able to detect Cre in lysates from the P1 lysogen but could detect small amounts from pET21a and larger amounts from pBluescriptIIKS+ (data not shown). EMSAs confirmed that active Cre is present in all three cases and allowed us to estimate the relative concentrations. Thus, the plasmid used for Cre expression provides protein levels that are higher than that found in a P1 lysogen, but much lower than would be produced by a promoter-driven expression system.

Using the pBluescriptII-Cre plasmid as template, we generated all 19 amino acid substitutions at each of the seven active site residues shown in Table 1. Four of these (Arg173, Lys201, Arg292 and Tyr324) are strongly conserved among both the YRs and the TopIB enzymes. Trp315 is most often a histidine in the YRs and is always histidine for the TopIBs. The final two residues are Glu176 and His289, both of which are unique to the YR enzymes. Glutamic acid is favored in the Glu176 position, although a significant number of family members use aspartic acid. His289 is moderately conserved among the YRs but is absent from the TopIB proteins. The results of *in vivo* recombination for the 134 Cre variants are summarized in Figure 4B. We found a surprising range of flexibility at several active site positions, including a number of mutants that we had not predicted would be active.

In vitro recombination of Cre mutants

To determine if any of the active site substitutions disrupt the ability of Cre to bind to the *loxP* site, we used a gel-shift assay to test cleared lysates from the transformed reporter strain (Supplementary Figure S1). The lysates in these experiments were diluted in binding buffer so that wild-type Cre lysate had similar binding activity as that of a 4 nM Cre control. This allowed a qualitative comparison of DNA-binding activities in the lysates relative to wild-type Cre. As shown in Supplementary Figure S1, the lysates for all substitutions at His289 contain DNA-binding activities similar to wild-type Cre. Out of all 134 Cre variants tested, only the Y324T mutant displayed a significant *loxP*-binding defect, where the apparent K_d is more than 10-fold greater than wild-type Cre. For the remaining mutants, these results indicate that the Cre variants are being produced in a soluble form that

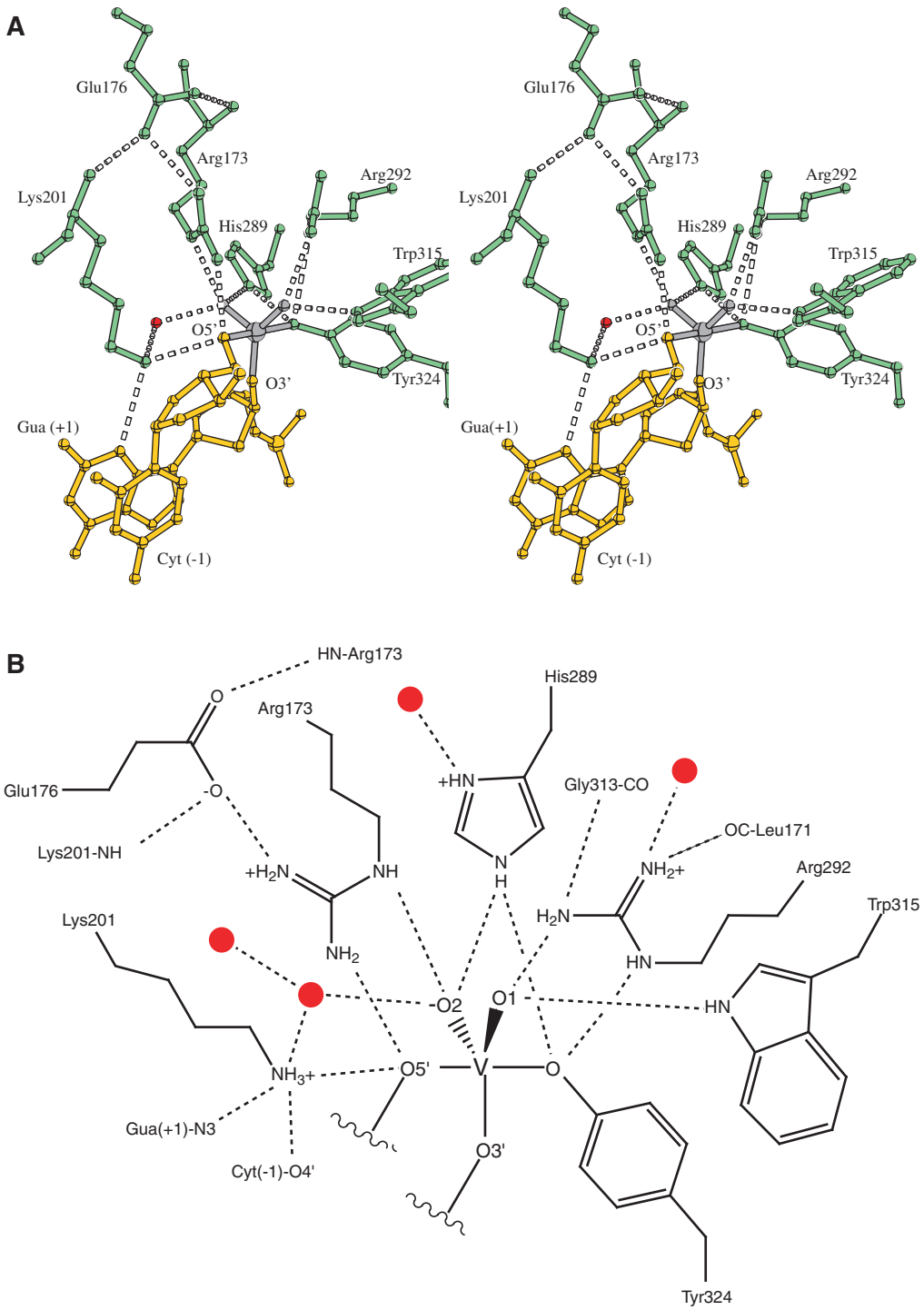

Figure 3. The Cre–DNA–vanadate transition state mimic active site. (**A**) Stereo view of the 'activated' site where vanadate has been incorporated. (**B**) Schematic of the activated site with additional residues and hydrogen bonding interactions indicated. Well-ordered water molecules are drawn as red spheres in (A) and (B).

bind to *loxP* site DNA, eliminating expression, folding, or gross DNA-binding defects as explanations for a lack of recombination activity. For many of the mutants, we noticed a small, but reproducible variation in mobility of the shifted EMSA bands that may be caused by subtle differences in DNA-bending.

To confirm and extend our findings from the *in vivo* recombination experiments, we tested each Cre mutant for *in vitro* recombination, again using lysates from the transfected reporter strain. In this assay, Cre excises a 334-bp segment flanked by *loxP* sites in a labeled, linear substrate, producing an unlabeled 368-bp circle and a

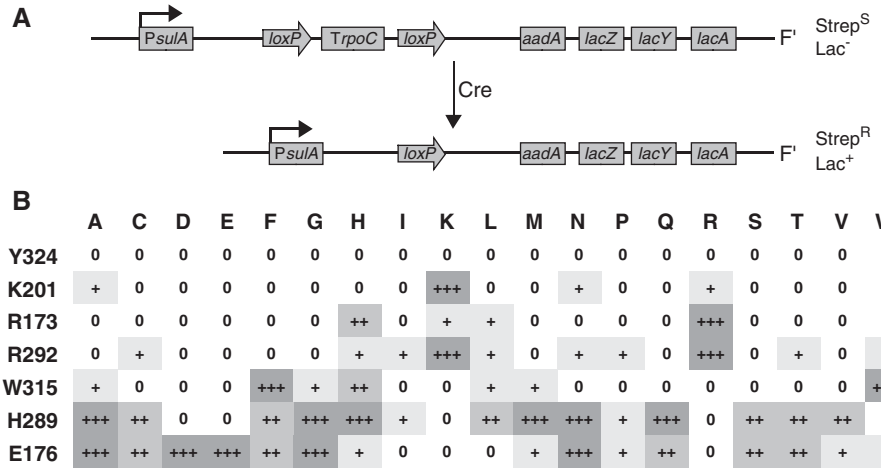

Figure 4. *In vivo* recombination activities of Cre active site mutants. (**A**) F' reporter used to score *in vivo* activity. If Cre excises a transcriptional terminator located downstream of the promoter, the streptomycin resistance and LacZYA gene products are expressed. (**B**) *in vivo* recombination results for Cre active site substitutions. Activity definitions: 0: no activity observed; +: <5% wild-type, ++: 5–50% wild-type, +++: >50% wild-type.

labeled 234-bp linear product (Figure 5A). The results of *in vitro* recombination for the H289 panel are shown in Figure 5B and a summary of results for all of the active site substitutions is shown in Figure 5C. In general, there is good agreement between *in vitro* and *in vivo* recombination, but for mutants with higher levels of activity the *in vitro* assay provides better discrimination. For example, the W315F and W315Y substitutions have near-wild-type activity in the *in vivo* assay, but are shown to have <50% wild-type activity in the *in vitro* excision assay. The *in vitro* recombination experiments also identified active site mutants that accumulate covalent and/or HJ intermediates in excess of wild-type Cre. These results are summarized in Table 3 and in Supplementary Figure S2.

Tyr324 substitutions

As anticipated, Tyr324 is strictly required for recombination. None of the 19 alternative residues showed any detectable recombination activity *in vivo* or *in vitro*. The requirement for tyrosine in this position of the active site is evident from Figure 3. Whereas some substitutions could, in principle, fulfill a role as nucleophile (e.g. serine, cysteine, lysine and histidine), the length and shape of the tyrosine side chain are unique. Indeed, a large superfamily of site-specific recombinases uses serine rather than tyrosine as a nucleophile to catalyze strand exchange by a related phosphoryl transfer mechanism (6). However, even if a suitable general base were present to activate a serine residue for nucleophilic attack in the Cre active site, it is clear that the side chain of a Y324S substitution is too short reach the scissile phosphate.

The requirement for tyrosine in the YR and TopIB active sites may also extend to the identity of the phenolic hydroxyl group. Semi-synthetic vaccinia virus TopIB mutants with the tyrosine oxygen replaced by nitrogen (tyramine) or sulfur (thio-tyrosine) are both inactive, suggesting a unique requirement for tyrosine in the cleavage and ligation chemistry catalyzed by this enzyme superfamily (32).

As noted above, Cre Y324T was the only mutant identified that is substantially defective in DNA-binding. Western blot analysis indicates that this mutant is expressed in a full-length, soluble form and at levels comparable to wild-type Cre (data not shown). It is possible that this mutant is misfolded or adopts an alternative structure that is not competent to bind to the *loxP* site.

Lys201 substitutions

Lys201 is located in the loop that connects conserved β-strands 2 and 3 in the YRs and in the TopIBs, where it is inserted into the minor groove near the cleavage site and hydrogen bonds to the +1 base (30,33). The important catalytic role of this conserved lysine was first demonstrated for poxvirus TopIB, where 5′-bridging phosphorothiolate DNA substrates were used to show that this residue is the general acid responsible for activation of the O5′ leaving group during the cleavage step of the reaction (34). The same conclusion was later reached for Cre, using a similar approach (35). Accordingly, the Cre K201A mutant and the corresponding mutants in Flp and Xer are inactive (36–38). Although Lys201 appears to be strongly conserved among the YRs, its position in a variable-length loop creates challenges for identifying the corresponding residue in large alignments of YR sequences (see 'Materials and Methods' section).

Consistent with previous studies, none of the Lys201 substitutions in Cre resulted in high recombination activity. However, several substitutions did have a low level of recombination activity *in vivo,* including the K201R mutant. Since the pKs of lysine and arginine differ by roughly two log orders, the arginine substitution would be expected to result in a 100-fold decrease in cleavage rate based only on its ability to activate the O5′-leaving group. This prediction holds true for poxvirus TopIB (34) and appears to be approximately true for Flp (37) and for Cre (this work). However, the

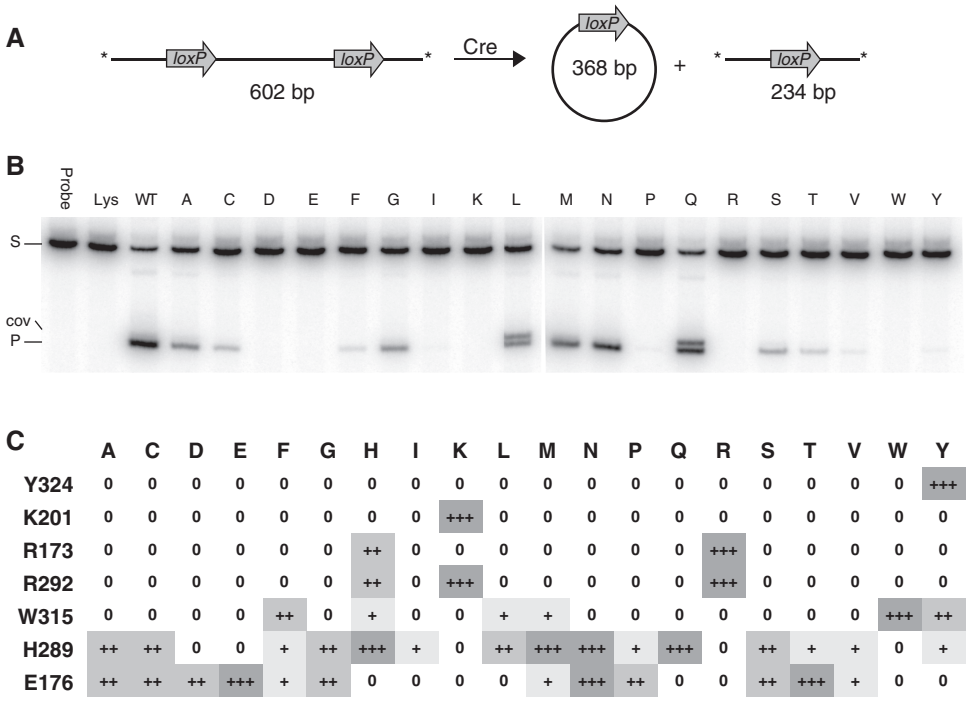

Figure 5. *In vitro* recombination activities of Cre active site mutants. (**A**) Schematic of the excision assay. (**B**) Native PAGE of *in vitro* recombination reactions following protein digestion for the H289 panel of substitutions. In addition to the 602 bp substrate (labeled 'S') and the 234 bp product (labeled 'P'), a band corresponding to a covalent Cre–DNA intermediate can be observed migrating slightly above the product band for some mutants. This is a nicked 234-bp covalent 3′-phosphotyrosine adduct. The corresponding covalent intermediate associated with the substrate is not well resolved. Identification of covalent and HJ intermediate bands are described in more detail in Supplementary Figure S2. (**C**) I*n vitro* recombination results. Activity definitions: 0: no activity observed; +: <5% wild-type, ++: 5-50% wild-type, +++: >50% wild-type.

Table 3. Accumulation of Holliday junction and covalent intermediates during *in vitro* recombination[a]

		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Y324	**HJ**																				+
	cov																				+
K201	**HJ**									+											
	cov									+											
R173	**HJ**							++								+					
	cov							++								+					
R292	**HJ**							++		+						+					
	cov									+++						+					
W315	**HJ**					+++														+	+++
	cov																			+	+
H289	**HJ**	++	+			+	++	+	+		+	++	+	+	+		+		+		
	cov							+			++	+	+		++						
E176	**HJ**	+++	+++	+++	+	+++	+++	+				++	++	++	+	+	+++	++	++		+
	cov	+			+								+					+			

Levels reported are after a 60-min reaction.
[a]Key: + indicates levels similar to wild-type Cre; ++ indicates >2× wild-type; +++ indicates >4× wild-type.

Ala and Asn substitutions at this position also have weak *in vivo* activity, despite their lack of ability to participate in general acid catalysis. None of the K201 substitutions showed measurable *in vitro* recombination activity.

The Cre transition state mimic structure strongly supports the biochemical data and related structural results from the TopIB systems regarding the active site lysine (13,14,34,39). Lys201 hydrogen bonds directly to the O5′ leaving group as well as to the +1 base, O4′ of the −1 sugar, and a well-ordered water molecule that is also found in the TopIB transition state mimic structures (Figure 3). The lysine side chain adopts a fully extended form, with all four side chain dihedral angles near 180°. This may explain in part why histidine is ineffective in this position; even if the bulkier imidazole ring could be accommodated in the minor groove, the ε-nitrogen would fall well short of forming effective hydrogen-bonding interactions with O5′. Although the weak activity of the arginine substitution can be rationalized, it is not clear why the Ala and Asn substitutions have a comparable level of activity *in vivo*.

Arg173 substitutions

Arg173 is strongly conserved among the tyrosine recombinase and TopIB enzymes and conservative lysine substitutions in this position show greatly diminished *in vitro* catalytic activity in the systems that have been tested (Table S1). Mutagenesis and biochemical experiments with poxvirus TopIB indicated a role in transition state stabilization for Arg130 and this idea was well supported by subsequent Cre–DNA and human TopIB–DNA crystal structures, where this arginine was observed making a double hydrogen-bonding contact to the scissile phosphate in the covalent intermediate of the reaction (30,33). More recent experiments using bridging phosphorothiolate substrates further indicated a role in leaving group activation, inasmuch as the poxvirus TopIB R130K mutant showed a substantial thio-effect (39). These observations have been largely reconciled by TopIB–DNA transition state mimic structures and now the Cre–DNA transition state mimic (Figure 3), all of which show this catalytic arginine making hydrogen bonding interactions with both a non-bridging phosphate (vanadate) oxygen and with the O5′ leaving group. Thus, Arg173 plays a dual role in the transition state of the cleavage and ligation reactions.

The results of our substitutions at Arg173 were somewhat surprising. We had purified the R173K mutant for a number of previous studies and have never observed *in vitro* cleavage or recombination activity using linear substrates. We therefore expected to find that no substitutions of Arg173 would result in recombination activity above the threshold of our *in vivo assay*. However, the R173K mutant is clearly active for recombination *in vivo*. We do not yet understand why Cre R173K is inactive *in vitro*, but we have previously shown that this mutant is severely defective in synapsis of *loxP* sites (26) and the corresponding R191K mutant in Flp shows altered patterns of DNA-bending (40). It is conceivable that substrate supercoiling or other differences in the *in vivo* reaction (41) are able to compensate to some extent for these deficiencies.

The second surprise was the relatively high activity observed for R173H, a mutant that to our knowledge has not been described for the tyrosine recombinases or the TopIB enzymes. Unlike R173K, which showed no *in vitro* recombination activity, the R173H mutant showed robust activity both *in vivo* and *in vitro*. These results suggest that a lack of interaction with the O5′ leaving group is unlikely to be responsible for the devastating loss of activity observed in other substitutions at this position (e.g. R173A). The hydrogen bond to O5′ made by R173 is via the η2-nitrogen, at the very end of the arginine side chain (Figure 3A). At best, the shorter histidine side chain in this position would be capable of a long-range interaction with O5′ via its ε-nitrogen. The more likely active site configuration for Cre R173H is one where His173 hydrogen bonds to the O2P non-bridging oxygen via Nε, just as Arg173 does in all Cre–DNA crystal structures that have been reported. Thus, the activity of R173H is most likely due to its ability to stabilize a buildup of negative charge on O2P rather than its ability to activate O5′.

The R173H mutant also accumulates more HJ and covalent intermediates of the reaction than are observed for wild-type Cre. The accumulation of covalent intermediate implies that $K_{eq} = k_{clv}/k_{lig}$ (where k_{clv} and k_{lig} are the first-order rate constants for cleavage and ligation, respectively) has increased for the covalent intermediates in the recombination pathway. Although Arg173 has been implicated in activation of the O5′ leaving group, the same interaction may also play a role in guiding O5′ into position for efficient ligation. Loss of this contribution could explain the mild ligation defect observed in the R173H mutant. Interestingly, the Flp R191K mutant also accumulates covalent intermediates *in vitro*, but in this case the ligation defect is more severe and no recombinant products are formed (40).

As shown in Table 3, most active site substitutions with modest *in vitro* recombination activities generate more HJ intermediate than wild-type Cre. Modeling of the recombination pathway suggests that this general phenomenon is most likely a consequence of a decrease in the rates of strand exchange, extending the time window for persistence of the HJ intermediate (data not shown). Analyses of reaction time courses would be required to determine if individual mutants are specifically defective in processing the HJ intermediate.

We did not anticipate that the more conservative R173K substitution would be less active than R173H. As a possible explanation for this result, we suggest that whereas histidine should be able to compensate for the loss of the R173–O2P interaction, lysine may not effectively replace either function of R173. Modeling of a Lys173 side chain in the transition state mimic active site confirms that the side chain is too short to interact effectively with O5′ and also indicates that it may be too long to form productive hydrogen-bonding interactions with O2P if the more fully extended rotamers are used. Alternative rotamers that would position the side chain for more optimal hydrogen-bonding geometry with O2P appear to introduce steric clashes with other active site residues, perhaps explaining why this residue is ineffective as an arginine substitute. An additional factor in the relative activities of Lys173 versus His173 may involve Glu176. As discussed below, Glu176 positions Arg173 in the active site by hydrogen bonding to its backbone amide and to its η1-nitrogen. It is possible that His173 is similarly positioned by Glu176 via its δ-nitrogen.

Substitutions at Arg292

Sequence alignments indicate that Arg292 is as conserved as Y324, with only a small number of lysine substitutions represented (Table 1). Structures of Cre–DNA (11), λ Int–DNA (42,43) and Flp–DNA (44) complexes all show this side chain hydrogen-bonding to the O1P non-bridging oxygen via Nη2. Since lysine in this position should also be capable of interacting with O1P (based on modeling), it is not surprising that the corresponding R223K mutant in vaccinia virus TopIB retains ~10% activity (45). We also found that the R292K mutant was moderately active.

In addition, we identified nine different substitutions for Arg292 that are weakly active *in vivo*. Of these, only the histidine mutant shows detectable recombination *in vitro*. The R292K substitution, on the other hand, is active both *in vivo* and *in vitro*. This mutant also accumulates the most covalent intermediate of any single active site substitution that we have studied (Table 3), indicating a substantial defect in substrate ligation. Similar observations have been reported for the corresponding mutants in vaccinia TopoIB (45), and XerCD (46).

The Cre–DNA–vanadate structure reveals a potentially important interaction involving Arg292 that was not evident in earlier recombinase–DNA crystal structures. In addition to forming a hydrogen bond with O1P, the arginine side chain hydrogen-bonds to the Tyr324 phenolic oxygen via Nε, forming a double hydrogen-bonding contact resembling that observed for Arg173 (but to different functional groups). A similar interaction was observed in the TopIB–vanadate structures (13,14), indicating that this is a conserved feature of the transition state in the YR and TopIB enzymes. As noted for the TopIB enzyme (47), the interaction of Arg292 with the Tyr324 oxygen suggests that Arg292 could play a functional role in stabilizing the buildup of negative charge of the phenolate form of the active site tyrosine during the transition state of the reaction. The accumulation of covalent intermediate by the R292K mutant would be consistent with a role for Arg292 in activation of the tyrosine leaving group during the ligation reaction, since lysine in this position would not be expected to form optimal interactions either with the Tyr324 oxygen or with O1P for the same reasons discussed above for the R173K mutant.

Substitutions at Trp315

Trp315 is an outlier among the YR enzymes in that >90% of family members have histidine in this position (Table 1). The Flp recombinase (and the small number of recombinases from saccharomyces-related yeast that harbor 2μ plasmids) also uses Trp in this position (8). In the covalent Cre–DNA complex crystal structure, Trp315 was observed hydrogen bonding to the O1P phosphate oxygen via its indole nitrogen, leading to the prediction that histidine most likely makes a similar interaction in the related enzymes (30). Although this turned out to be true for λ-integrase (42) and for the TopIB enzymes (33,48), subsequent work revealed that histidine is not effective when substituted for the corresponding tryptophan in Flp (37) or in Cre (49). In both cases, the Phe substitutions were reported to have higher activity than His *in vitro* and the Ala substitutions were inactive. These studies, combined with the structure of the Flp W330F–DNA complex (37), led to a model where the primary role of this residue involves buttressing the α-helix that contains the tyrosine nucleophile and not transition state stabilization via the indole-phosphate hydrogen bond.

Our analysis of Trp315 substitutions confirms the previous *in vitro* recombination data for Flp and Cre and extends the list of highly active substitutions to include tyrosine. In addition, we found that Ala, Gly, Leu and Met substitutions all have weak *in vivo* recombination activity. Of the weaker substitutions, His, Leu and Met have detectable recombination activity *in vitro* but the smaller Gly and Ala substitutions do not. The finding that W315M has recombination activity comparable to W315H further underscores the need for a large hydrophobic residue in this position. The *in vivo* activity of W315A and W315G would appear to contradict this model, but it is possible that these mutants adopt an alternative conformation in this region of the enzyme that is somehow able to partially compensate for the loss of W315.

The positions of Trp315 and Tyr324 in the transition state mimic structure (Figure 3) provide an opportunity to carefully examine the contacts that are present at this crucial stage of the reaction. Trp315 is located on a turn between the αL and αM helices, where it provides a substantial amount of hydrophobic surface onto which the αM helix docks via van der Waals interactions with Ile320 and Val321. Tyr324 is located at the carboxyl-terminus of αM, where the tyrosine ring 'sits' on top of Trp315. Each of the carbon atoms in the Tyr324 phenyl ring contacts Cζ of Trp315 with a distance of 3.5–4 Å. Although the Phe and Tyr substitutions are predicted to be less effective than Trp in supporting the Tyr324 ring system, it is clear from modeling that a similar set of substantial contacts could be made in the transition state in both cases. Indeed, the structure of the Flp W330F/DNA complex shows that this is the case when the catalytic tyrosine has not yet attacked the scissile phosphate (37). On the other hand, the smaller histidine in this position does not appear to be capable of forming similar contacts. We note that both the Phe and Tyr substitutions at Trp315 cause substantial accumulations of HJ intermediates (Table 3), suggesting that repacking this hydrophobic region of the enzyme may also lead to conformational changes that influence aspects of the reaction pathway other than catalysis.

If Trp315 plays a crucial role in positioning the active site nucleophile, then it is instructive to consider how the majority of tyrosine recombinases are able to function with histidine in this position. In both Cre and Flp (and most other YRs that use Trp in this position), glycine is located two residues before His/Trp III. A side chain in this position would occupy the same space as the tryptophan side chain in the absence of large structural changes in the αL–αM turn. This residue is normally Leu, Met or Val in most of the YR enzymes, each of which would likely be able to fulfill the role of the active site tryptophan in Cre and Flp. In λ integrase, for example, a leucine side chain in this position packs against the catalytic tyrosine (42).

In those cases where a smaller side chain is located two residues preceding His/Trp III, a large hydrophobic side chain is present four residues after Arg292 in the αK helix. This residue is normally an alanine in most YRs, but a larger side chain would also project into the space occupied by the bulky Trp side chain in Cre and Flp. Thus, there appear to be a number of solutions for stabilizing the local structure and defining the position of the active site nucleophile in the YR enzymes (49).

In vaccinia virus TopIB, the corresponding His265 can be substituted by Asn or Gln with only a modest loss of activity (50). In this case, the helix containing Tyr274 is anchored adjacent to the active site via direct contacts to the DNA backbone and to bases downstream of the cleavage site (14), so the role of this catalytic residue for the TopIBs appears to primarily involve transition state stabilization and not positioning of the active site tyrosine. In the Flp recombinase, the W330A mutant shows an unexpected thio effect on 5′-bridging phosphorothiolate substrates, suggesting a role in leaving group activation (49). This observation cannot be readily explained based on the Cre transition state mimic structure.

His289 is unique to the tyrosine recombinases

The TopIB enzymes have no equivalent to His289 in their active sites. Although lysine precedes Arg223 by three residues in vaccinia and variola virus TopIB, differences in local structure allow Lys220 to be directed away from the scissile phosphate, where it instead interacts with the upstream (+1) phosphate on the cleaved strand (48). Consequently, the space occupied by His289 in the YRs is filled with water molecules in the TopIB structures. His289 is less strongly conserved among the tyrosine recombinases compared to the active site arginines and the tyrosine (~87%), with ~7% of sequences showing Tyr in this position and ~1% with Gln. The original covalent Cre–DNA structure prompted the suggestion that His289 could be responsible for activation of Tyr324 during cleavage and ligation (30,51). Indeed, subsequent crystal structures of Cre–DNA, Flp–DNA and λ Int–DNA complexes supported this idea. Biochemical evidence for a role in general acid/base catalysis has come from the Flp system, where synthetic peptides with increased nucleophilicity of the catalytic tyrosine were shown to rescue cleavage activity of the His305Q mutant (52).

We found that all but five substitutions (Asp, Glu, Lys, Arg and Trp) at His289 in Cre are active for recombination *in vivo*. Of these, Ile and Pro show only weak activity but several are nearly as efficient as wild-type Cre. All of the mutants that were active *in vivo* also showed detectable *in vitro* activity (Figure 5B). Two of the active mutants, H289A and H289N, have been previously studied (53). In the transition state mimic structure, His289 adopts a curious hydrogen-bonding arrangement, with the Nε hydrogen directed midway between non-bridging O2P and the attacking/leaving oxygen of Tyr324. Thus, His289 is positioned in such a way that it could stabilize the buildup of negative charge on O2P and serve as a proton donor/acceptor for the Tyr324 phenol oxygen. A well-ordered water molecule (B-factor ~ 18 $Å^2$) hydrogen bonds to Nδ of His289, an anchoring interaction that was also observed in the covalent intermediate structure.

The variety of amino acid substitutions at His289, which are active for recombination, indicates that although moderately conserved, this active site residue is not strictly required for catalysis. This idea is consistent with the absence of an equivalent residue in the TopIB enzymes. Glutamine is probably the most conservative substitution for His289 in terms of size and ability to hydrogen bond, yet the Met substitution has similar activity *in vivo* and nearly the same activity *in vitro*. Even the Gly substitution is moderately active. As shown in Figure 5B and Table 3, both the Gln and Leu substitutions at His289 lead to an accumulation of covalent intermediate. This effect was also observed for several His305 mutants in Flp, including H305L (54). Although one might conclude from an analysis of these substitutions alone that His289 is required for efficient ligation, the data indicate that covalent intermediate accumulation is not a general property of His289 mutants. The Ala, Gly, Met and Asn substitutions are all moderately active, yet none of them appears to have shifted the cleavage–ligation balance in favor of cleavage. The H289M mutant does, however, accumulate a significant amount of HJ intermediate, with lesser amounts produced by the Ala and Gly substitutions. Interestingly, the Cre H289A mutant has also been shown to have an altered preference for the strand of *loxP* that is initially exchanged during recombination (53).

Of the inactive substitutions at His289, modeling indicates that Trp is too bulky to be accommodated in the active site without major changes in the positions of catalytic residues. Asp and Glu would, in principle, be capable of performing general acid/base catalysis, but the negative charges associated with the carboxylate side chains are clearly detrimental to recombination, presumably by destabilizing the transition state. The lack of detectable activity for the Lys and Arg substitutions was somewhat surprising. One possible explanation is that these side chains compete with Tyr324 for the same position adjacent to the scissile phosphate and sterically block the nucleophile. A large conformational change that occurs in helix-M and Tyr324 relative to the inactive active sites in the Cre tetramer suggests that this explanation is plausible. Alternatively, the side chains may simply be too long to be accommodated in the active site. It does not appear possible for Lys289 or Arg289 to adopt conformations that resemble Lys220 in the viral TopIB structures due to steric interference with the inter-domain linker that is quite different in Cre compared to the viral TopIB protein.

Glu176 organizes the active site

Glu176 does not interact directly with the scissile phosphate, but plays a role in organizing the Cre active site. Although Glu is favored in this position of the YR family (79%), Asp is also highly represented (16.4%), with over 95% using one or the other acidic side chain (Table 1). In the transition state mimic structure, Glu176 accepts three hydrogen bonds from catalytic residues. The side chain carboxylate forms a hydrogen-bonding bridge between the backbone amide and Nη1 of Arg173, appearing to lock this key residue into a conformation that is optimal for catalysis. The Glu176 carboxylate also hydrogen bonds to the backbone amide of Lys201, helping to position the β2–β3 loop in the minor groove adjacent to the active site.

Both λ integrase and Flp recombinase have aspartic acid in this position. In the λ Int–DNA complex, Asp215 anchors the β2–β3 loop by hydrogen bonding with the backbone amide of Ser234 (which precedes the conserved active site Lys235) and makes a long-range interaction with Arg212–Nη1 (42,43). In the Flp–DNA complex, Asp194 hydrogen bonds to the backbone amide of Arg191 and interacts with the amide of Lys223 indirectly via bridging water molecules (55). Thus, Asp and Glu in this position of the YR enzymes carry out similar structural roles, but the detailed interactions differ among the individual systems. The TopIB enzymes do not have a residue corresponding to Glu176 in Cre. Instead, tightly bound water molecules in the active site make a similar set of interactions with Arg130 and the Lys167 backbone (13,14).

We found that 15 substitutions for Glu176 in Cre were active for recombination *in vivo*, although the His, Met, Pro, Val, Trp and Tyr substitutions were very weak. *In vitro*, the His, Trp and Tyr mutants showed no detectable activity. Surprisingly, the strongest substitutions were Asn and Thr, each of which showed high *in vitro* recombination activity. The conservative E176D mutant was somewhat less active and the E176Q mutant showed no measurable activity at all in the *in vitro* assay. The behavior of E176Q was unexpected; presumably this residue stabilizes an alternative, but unproductive arrangement of active site residues. The modest activities resulting from Ala and Gly substitutions indicate that loss of the interactions involving Arg173 and Lys201 diminishes, but does not destroy catalysis in this system.

Several substitutions for the corresponding Asp194 in Flp recombinase have also been studied (Table S1). Of these, it is interesting to note that the D194N mutant is deficient for cleavage of DNA half-sites (56). As noted above, the corresponding conversion from acid to amide in Cre is also defective. Although none of the Cre Glu176 mutants accumulates significant amounts of covalent intermediate, many of them do accumulate Holliday intermediates.

DISCUSSION

The Cre active site residues can be grouped into three categories based on their tolerance to substitution. In the first category, Tyr324 and Lys201 are essential for catalysis; we found no active substitutions of Tyr324 and only very weak ones for Lys201. Thus, a nucleophile and a general acid catalyst to activate the 5′-hydroxyl leaving group play the most unique and essential roles in the reaction. In the second group, Arg173 and Arg292 each play dual roles in catalysis and are symmetrically organized in the transition state with respect to cleavage and ligation. Both residues stabilize the buildup of negative charge by hydrogen-bonding to a non-bridging phosphate oxygen. These residues also hydrogen bond to the nucleophile and leaving groups of the reaction: Arg173 to O5′ and Arg292 to the Tyr324 hydroxyl. Thus, the roles of the two residues become interchanged during cleavage versus ligation. The third category includes Trp315, His289 and Glu176, each of which is tolerant to substitutions that result in moderate-to-high levels of recombination activity, with additional substitutions that have lower levels of activity. No active site substitution resulted in an increase in *in vivo* or *in vitro* recombination activity relative to wild-type Cre.

For most active site residues, there is a high correlation between the degree of substitution allowed and the diversity of residues identified at the corresponding position in the protein database. The exception is Arg173, which is present as Arg in only 91% of the YR protein sequences. Interestingly, the lysine substitution is found in only ~1% of the sequences and the remainder are distributed among Ala, Ser and other residues. One explanation could be that there are alternative active site compositions and/or geometries that do not require the box I arginine for activity. Alternatively, the YRs lacking Arg I may represent a group that is catalytically inactive, but has nonetheless retained Arg II and the active site tyrosine.

The active site residues most sensitive to substitution are those shared by the TopIB enzymes, which catalyze the same basic phosphoryl transfer reactions very efficiently without His289 and Glu176 equivalents. Why, then, is His II moderately conserved in the YR family and why are Asp and Glu nearly always found in the position corresponding to Glu176? Part of the answer may be related to the differences between the reactions catalyzed by the YR and TopIB enzymes. The TopIBs cleave their targets to form covalent 3′-phosphotyrosine intermediates and allow rotation of DNA downstream of the cleavage site about the site of the nick. TopIB evolution has presumably tuned the relative rates of cleavage and ligation to strike a balance between allowing escape of the 5′-nucleoside (and subsequent relaxation of supercoils) and minimizing exposure of the nicked chromosome to potentially damaging processes. In contrast, the YR enzymes need to coordinate the cleavage, exchange and ligation of DNA strands between two active sites in the recombinase tetramer, while rendering the other two sites inactive. In the absence of such allosteric regulation, YR enzymes could generate double-strand breaks across the 6–8 bp overlap region.

The details of the inter-subunit interactions responsible for coordinating cleavage in the YR systems differ considerably among the Cre, λ integrase and Flp systems. However, one general feature in common to each of these systems involves positioning of the active site tyrosine, where availability of the nucleophile for catalysis is strongly influenced by interactions with the neighboring subunit (31,42,44,46). For Flp recombinase, the active site tyrosine is donated in *trans* from an adjacent subunit (57). An additional source of regulation may involve positioning of the active site lysine, which is also influenced by inter-subunit contacts. One consequence of evolving a regulated recombinase active site from an ancestral topoisomerase may therefore have been a loss of some catalytic power, as the dynamics of active site residues were modified in order to allow allosteric control of their positions. We suggest that the YR active site residues corresponding to His289 and Glu176 in Cre were added in order to compensate for this loss in activity. Indeed, loss of

either of these residues in Cre to generate a more 'Topo-like' active site results in a decrease in recombination activity and for those mutants in Cre and other systems where cleavage has been assayed specifically, a decrease in cleavage rates have been reported (Table S1).

Ironically, a comparison of the static transition state models between the TopIB and YR enzymes might lead to the conclusion that the YRs should be more efficient in catalyzing cleavage and ligation. In the Cre transition state mimic structure, His289 provides both additional transition state stabilization and a potential activating function for Tyr324. Glu176 plays a structural role in positioning Arg173 and Lys201. In the corresponding variola virus and trypanosomal TopIB transition state mimics, these functional roles appear to be carried out by ordered water molecules. However, estimates of the pseudo-first order cleavage rate constants for half-site and full-site DNA substrates for Cre range from 0.001 to $0.003\,s^{-1}$ (35,53,58), compared to 0.3 to $0.7\,s^{-1}$ for vaccinia virus TopIB (34,59). Flp recombinase, on the other hand, cleaves half-site and full-site DNA at rates that are closer to that of the TopIB enzymes (60). Thus, the additional active site residues found in the YR enzymes do not appear to provide a substantial increase in their ability to cleave and ligate DNA substrates relative to their TopIB cousins.

In addition to providing additional insight into the relationship between the YR and TopIB enzymes, the data presented here provide a basis for interpreting the results of experiments involving TopIB and YR cleavage of DNA substrates containing methylphosphonates at the scissile phosphate positions (58,60,61). Vaccinia virus TopIB preferentially cleaves DNA substrates with a scissle Sp-methylphosphonate (Sp-MeP) over the Rp-MeP diastereomer to form a covalent topo–DNA intermediate. This intermediate is hydrolyzed at a rate that is four orders of magnitude higher than observed with the normal phosphotyrosine linkage, leading to the conclusion that the negative charge on the phosphate (which is reduced in the case of the methylphosphonate) plays a role in protecting the covalent intermediate from hydrolysis by repelling potential water nucleophiles (61). When similar experiments were repeated with the Flp and Cre recombinases, however, the stereospecificity of methylphosphonate cleavage was found to be identical, but there was no rate enhancement of hydrolysis of the corresponding covalent intermediate (58,60).

A comparison of the Cre–DNA transition state mimic with similar TopIB structures provides some insight into these observations. In both Cre and TopIB, an S_p-MeP (which corresponds to O1 in Figure 3 being replaced by a methyl group) would generate a steric clash with Arg292 (Cre)/Arg223 (TopIB) in the transition state, requiring that this catalytic arginine be repositioned. In the TopIB active site, a number of alternative low-energy rotamers are available to Arg223. One of these side chain conformations results in Arg223 straddling the Tyr274 leaving group oxygen, where it can form two hydrogen bonds via Nε and Nη2. Thus, part of the hydrolysis rate enhancement observed for the TopIB enzyme could be due to an increase in leaving group activation due to an alternative conformation of Arg223. In the Cre and Flp active sites, this alternative side chain conformation for Arg292/Arg308 is not possible because His289/His305 (which is absent in the TopIBs) sterically limits the number of arginine rotamers than can be accommodated.

The active site substitutions that we have described, when combined with transition state mimic structures of Cre and TopIB, provide a useful foundation for comparing the two systems and for understanding the requirements for catalysis during Cre–*loxP* recombination and presumably in the other YR family members as well. In the YR systems, however, it is an oversimplification to think of the active site as being devoted entirely to catalysis of cleavage and ligation. The Cre R173K and H289A mutants are examples where this is clearly the case. Cre R173K is defective in synapsis of *loxP* sites, which may contribute to the lack of *in vitro* activity for this mutant. In the structure of the pre-cleavage synaptic Cre–*loxP* complex (using the Cre K201A mutant), Arg173 hydrogen bonds to the scissile phosphate in the 'inactive sites', stabilizing a phosphate backbone conformation adjacent to the asymmetric *loxP* bend (26). The H289A mutant has the unexpected property of promoting a top-strand-first order of *loxP* strand exchanges during recombination, the opposite of that preferred by wild-type Cre (53). His289 also plays an important role in the inactive site of the synaptic complex, nucleating a hydrogen-bonding network involving the Arg173 backbone, two water molecules and the active site phosphate (26). Related allosteric effects have been reported in the Xer system, where an R148K mutant in XerC is able to stimulate the catalytic activity of XerD on HJ substrates and vice versa (46). Thus, active site residues can also strongly influence the reaction via interactions involving the half-sites not undergoing catalysis.

Efforts thus far to correlate biochemical activities with structural models in the Cre, Flp and λ Int systems indicate that while the YRs have evolved strategies for inter-subunit communication that share some overall themes, they do differ considerably in their molecular details. This is especially apparent when comparing Flp recombinase to the bacterial YRs, where the inter-subunit contacts in the recombinase tetramers are quite different. Consequently, the roles of active site residues in mediating the allosteric regulation of cleavage and ligation in these enzymes are likely to differ as well. An example of this can be seen in a comparison of the Cre H289Q mutant with the corresponding H305Q mutant in Flp. The Cre mutant is nearly as active as wild-type Cre (this study), yet the Flp H305Q mutant is severely defective (54). Dissection of active site contributions to chemical catalysis versus mediating the conformational changes and dynamics of the recombinase tetramer through the recombination pathway may therefore prove to be challenging.

ACCESSION NUMBER

3MGV.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr. Kay Perry for insightful discussions.

FUNDING

The National Institutes of Health (GM55041 to G.V., T32GM008275 to B.G.); Howard Hughes Medical Institute. Funding for open access charge: Howard Hughes Medical Institute.

Conflict of interest statement. None declared.

REFERENCES

1. Nash,H. (1996) Site-specific recombination: integration, excision, resolution, and inversion of defined DNA segments. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*. ASM Press, Washington, D.C, pp. 2363–2376.
2. Azaro,M. and Landy,A. (2002) λ Integrase and the λ Int family. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington D.C, pp. 118–148.
3. Jayaram,M., Grainge,I. and Tribble,G. (2002) Site-specific recombination by the Flp protein of Saccharomyces cerevisiae. In Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (eds), *Mobile DNA II*. ASM Press, Washington DC, pp. 192–218.
4. Abremski,K. and Hoess,R. (1984) Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.*, **259**, 1509–1514.
5. Sherratt,D., Soballe,B., Barre,F., Filipe,S., Lau,I., Massey,T. and Yates,J. (2004) Recombination and chromosome segregation. *Philos. Trans. Roy. Soc. Lond. B Biol. Sci.*, **359**, 61–69.
6. Grindley,N., Whiteson,K. and Rice,P. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem.*, **75**, 567–605.
7. Shuman,S. (1998) Vaccinia virus DNA topoisomerase: a model eukaryotic type IB enzyme. *Biochim. Biophys. Acta*, **1400**, 321–337.
8. Nunes-Düby,S., Kwon,H., Tirumalai,R., Ellenberger,T. and Landy,A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.*, **26**, 391–406.
9. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
10. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
11. Van Duyne,G. (2008) Site-specific recombinases. In Rice,P. and Correll,C. (eds), *Protein–Nucleic Acid Interactions*. Royal Society of Chemistry, Cambridge, UK, pp. 303–332.
12. Cheng,C., Kussie,P., Pavletich,N. and Shuman,S. (1998) Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell*, **92**, 841–850.
13. Davies,D.R., Mushtaq,A., Interthal,H., Champoux,J.J. and Hol,W.G.J. (2006) The structure of the transition state of the heterodimeric topoisomerase I of Leishmania donovani as a vanadate complex with nicked DNA. *J. Mol. Biol.*, **357**, 1202–1210.
14. Perry,K., Hwang,Y., Bushman,F.D. and Van Duyne,G.D. (2010) Insights from the structure of a smallpox virus topoisomerase-DNA transition state mimic. *Structure*, **18**, 127–137.
15. Ghosh,K. and Van Duyne,G. (2002) Cre-loxP biochemistry. *Methods*, **28**, 374–383.
16. Gordon,J.A. (1991) Use of vanadate as protein-phosphotyrosine phosphatase inhibitor. *Methods Enzymol.*, **201**, 477–482.
17. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
18. Navaza,J. (1994) AMORE: an automated package for molecular replacement. *Acta Cryst.*, **A50**, 157–163.
19. Brünger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
20. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
21. Murshudov,G., Vagin,A. and Dodson,E. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.
22. Whipple,F. (1998) Genetic analysis of prokaryotic and eukaryotic DNA-binding proteins in Escherichia coli. *Nucleic Acids Res.*, **26**, 3700–3706.
23. Dmitrova,M., Younès-Cauet,G., Oertel-Buchheit,P., Porte,D., Schnarr,M. and Granger-Schnarr,M. (1998) A new LexA-based genetic system for monitoring and analyzing protein heterodimerization in Escherichia coli. *Mol. Gen. Genet.*, **257**, 205–212.
24. Brosius,J. (1984) Toxicity of an overproduced foreign gene product in Escherichia coli and its use in plasmid vectors for the selection of transcription terminators. *Gene*, **27**, 161–172.
25. Phillips,G. (1999) New cloning vectors with temperature-sensitive replication. *Plasmid*, **41**, 78–81.
26. Ghosh,K., Guo,F. and Van Duyne,G.D. (2007) Synapsis of loxP sites by Cre recombinase. *J. Biol. Chem.*, **282**, 24004–24016.
27. Hoess,R., Wierzbicki,A. and Abremski,K. (1987) Isolation and characterization of intermediates in site-specific recombination. *Proc. Natl Acad. Sci. USA*, **84**, 6840–6844.
28. Gopaul,D., Guo,F. and Van Duyne,G. (1998) Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *EMBO J.*, **17**, 4175–4187.
29. Guo,F., Gopaul,D. and Van Duyne,G. (1999) Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Natl Acad. Sci. USA*, **96**, 7143–7148.
30. Guo,F., Gopaul,D. and Van Duyne,G. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.
31. Van Duyne,G. (2001) A structural view of Cre-loxP site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 87–104.
32. Gao,R., Zhang,Y., Choudhury,A.K., Dedkova,L.M. and Hecht,S.M. (2005) Analogues of vaccinia virus DNA topoisomerase I modified at the active site tyrosine. *J. Am. Chem. Soc.*, **127**, 3321–3331.
33. Redinbo,M., Stewart,L., Kuhn,P., Champoux,J. and Hol,W. (1998) Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science*, **279**, 1504–1513.
34. Krogh,B. and Shuman,S. (2000) Catalytic mechanism of DNA topoisomerase IB. *Mol. Cell*, **5**, 1035–1041.
35. Ghosh,K., Lau,C., Gupta,K. and Van Duyne,G. (2005) Preferential synapsis of loxP sites drives ordered strand exchange in Cre-loxP site-specific recombination. *Nat. Chem. Biol.*, **1**, 275–282.
36. Lee,L. and Sadowski,P. (2003) Identification of Cre residues involved in synapsis, isomerization, and catalysis. *J. Biol. Chem.*, **278**, 36905–36915.
37. Chen,Y. and Rice,P.A. (2003) The role of the conserved Trp330 in Flp-mediated recombination Functional and structural analysis. *J. Biol. Chem.*, **278**, 24800–24807.
38. Cao,Y. and Hayes,F. (1999) A newly identified, essential catalytic residue in a critical secondary structure element in the integrase family of site-specific recombinases is conserved in a similar element in eucaryotic type IB topoisomerases. *J. Mol. Biol.*, **289**, 517–527.

39. Krogh,B. and Shuman,S. (2002) Proton relay mechanism of general acid catalysis by DNA topoisomerase IB. *J. Biol. Chem.*, **277**, 5711–5714.
40. Friesen,H. and Sadowski,P.D. (1992) Mutagenesis of a conserved region of the gene encoding the FLP recombinase of Saccharomyces cerevisiae. A role for arginine 191 in binding and ligation. *J. Mol. Biol.*, **225**, 313–326.
41. Adams,D., Bliska,J. and Cozzarelli,N. (1992) Cre-lox recombination in Escherichia coli cells. Mechanistic differences from the in vitro reaction. *J. Mol. Biol.*, **226**, 661–673.
42. Aihara,H., Kwon,H., Nunes-Duby,S., Landy,A. and Ellenberger,T. (2003) A conformational switch controls the DNA cleavage activity of lambda integrase. *Mol. Cell*, **12**, 187–198.
43. Biswas,T., Aihara,H., Radman-Livaja,M., Filman,D., Landy,A. and Ellenberger,T. (2005) A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature*, **435**, 1059–1066.
44. Chen,Y. and Rice,P.A. (2003) New insight into site-specific recombination from Flp recombinase-DNA structures. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 135–159.
45. Cheng,C., Wang,L.K., Sekiguchi,J. and Shuman,S. (1997) Mutational analysis of 39 residues of vaccinia DNA topoisomerase identifies Lys-220, Arg-223, and Asn-228 as important for covalent catalysis. *J. Biol. Chem.*, **272**, 8263–8269.
46. Arciszewska,L., Baker,R., Hallet,B. and Sherratt,D. (2000) Coordinated control of XerC and XerD catalytic activities during Holliday junction resolution. *J. Mol. Biol.*, **299**, 391–403.
47. Yakovleva,L., Chen,S., Hecht,S.M. and Shuman,S. (2008) Chemical and traditional mutagenesis of vaccinia DNA topoisomerase provides insights to cleavage site recognition and transesterification chemistry. *J. Biol. Chem.*, **283**, 16093–16103.
48. Perry,K., Hwang,Y., Bushman,F. and Van Duyne,G. (2006) Structural basis for specificity in the poxvirus topoisomerase. *Mol. Cell*, **23**, 343–354.
49. Ma,C., Kwiatek,A., Bolusani,S., Voziyanov,Y. and Jayaram,M. (2007) Unveiling hidden catalytic contributions of the conserved His/Trp-III in tyrosine recombinases: assembly of a novel active site in Flp recombinase harboring alanine at this position. *J. Mol. Biol.*, **368**, 183–196.
50. Petersen,B.O. and Shuman,S. (1997) Histidine 265 is important for covalent catalysis by vaccinia topoisomerase and is conserved in all eukaryotic type I enzymes. *J. Biol. Chem.*, **272**, 3891–3896.
51. Sherratt,D. and Wigley,D. (1998) Conserved themes but novel activities in recombinases and topoisomerases. *Cell*, **93**, 149–152.
52. Whiteson,K.L., Chen,Y., Chopra,N., Raymond,A.C. and Rice,P.A. (2007) Identification of a potential general acid/base in the reversible phosphoryl transfer reactions catalyzed by tyrosine recombinases: Flp H305. *Chem. Biol.*, **14**, 121–129.
53. Gelato,K.A., Martin,S.S. and Baldwin,E.P. (2005) Reversed DNA strand cleavage specificity in initiation of Cre-LoxP recombination induced by the His289Ala active-site substitution. *J. Mol. Biol.*, **354**, 233–245.
54. Parsons,R., Prasad,P., Harshey,R. and Jayaram,M. (1988) Step-arrest mutants of FLP recombinase: implications for the catalytic mechanism of DNA recombination. *Mol. Cell. Biol.*, **8**, 3303–3310.
55. Chen,Y., Narendra,U., Iype,E., Cox,M. and Rice,A. (2000) Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell*, **6**, 885–897.
56. Chen,J., Evans,B., Yang,S., Araki,H., Oshima,Y. and Jayaram,M. (1992) Functional analysis of box I mutations in yeast site-specific recombinases Flp and R: pairwise complementation with recombinase variants lacking the active-site tyrosine. *Mol. Cell. Biol.*, **12**, 3757–3765.
57. Chen,J., Lee,J. and 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.
58. Ma,C., Kachroo,A.H., Macieszak,A., Chen,T., Guga,P. and Jayaram,M. (2009) Reactions of Cre with methylphosphonate DNA: similarities and contrasts with Flp and vaccinia topoisomerase. *PLoS One*, **4**, e7248.
59. Stivers,J., Jagadeesh,G., Nawrot,B., Stec,W. and Shuman,S. (2000) Stereochemical outcome and kinetic effects of Rp- and Sp-phosphorothioate substitutions at the cleavage site of vaccinia type I DNA topoisomerase. *Biochemistry*, **39**, 5561–5572.
60. Ma,C., Rowley,P.A., Macieszak,A., Guga,P. and Jayaram,M. (2009) Active site electrostatics protect genome integrity by blocking abortive hydrolysis during DNA recombination. *EMBO J.*, **28**, 1745–1756.
61. Tian,L., Claeboe,C.D., Hecht,S.M. and Shuman,S. (2003) Guarding the genome: electrostatic repulsion of water by DNA suppresses a potent nuclease activity of topoisomerase IB. *Mol. Cell*, **12**, 199–208.