Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse

Feng Guo, Deshmukh N. Gopaul & Gregory D. Van Duyne

Johnson Research Foundation and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA

During site-specific DNA recombination, which brings about genetic rearrangement in processes such as viral integration and excision and chromosomal segregation, recombinase enzymes recognize specific DNA sequences and catalyse the reciprocal exchange of DNA strands between these sites. The bacteriophage recombinase Cre catalyses site-specific recombination between two 34-base-pair *loxP* sites. The crystal structure at 2.4 Å resolution of Cre bound to a *loxP* substrate reveals an intermediate in the recombination reaction, in which a Cre molecule has cleaved the substrate to form a covalent 3′-phosphotyrosine linkage with the DNA. Four recombinases and two *loxP* sites form a synapsed structure in which the DNA resembles models of four-way Holliday-junction intermediates. The Cre–*loxP* complex challenges models of site-specific recombination that require large changes in quaternary structure. Subtle allosteric changes at the carboxy termini of the Cre subunits may instead coordinate the cleavage and strand-exchange reactions.

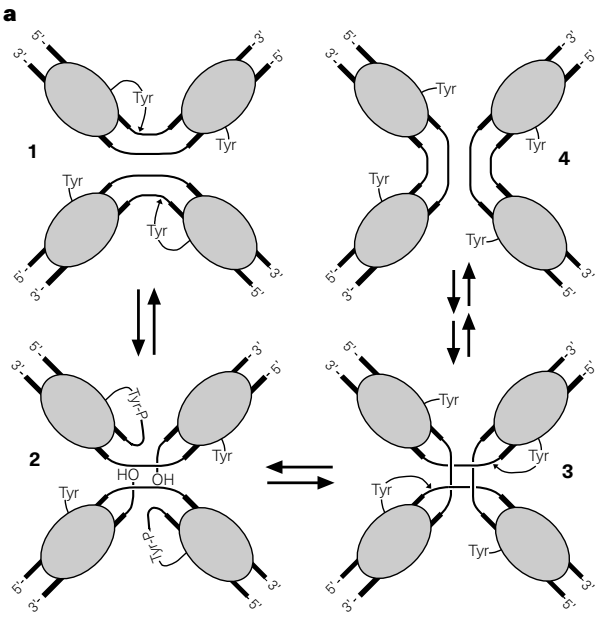
Based on sequence similarity, recombinases from bacteria and yeast have been classified into the integrase family and the resolvase–invertase family, which use two distinct mechanisms of recombination[1,2]. Integrase family members cleave their DNA substrates by a series of staggered cuts, during which the recombinase becomes covalently linked to the DNA through a catalytic tyrosine residue (Fig. 1). Recombination takes place by exchanging one set of strands to yield a Holliday structure intermediate, followed by the exchange of a second set of strands to resolve the intermediate into recombinant products.

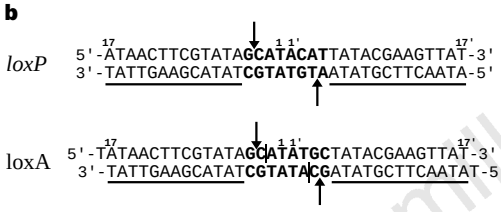
The >60 members of the integrase family[3] share poor sequence similarity overall, although four residues required for catalysis are strictly conserved[4,5]. The sequence dissimilarity among integrase family members may reflect the diversity of function and form in which these recombinases carry out their genetic rearrangements. Some members, such as the bacteriophage-λ integrase, require additional accessory proteins to facilitate the recombination process[6]. In contrast, the Cre recombinase (relative molecular mass, 38 K) from bacteriophage P1 mediates a site-specific recombination reaction between two 34-base-pair *loxP* sites (Fig. 1a, b) which does not require accessory factors and can be performed *in vitro* with a variety of DNA substrates[7,8]. The role of Cre recombinase in the bacteriophage P1 life cycle is to maintain the phage genome as a monomeric, unit-copy plasmid in the lysogenic state[7,9]. The simplicity of the Cre/*loxP* system has led to its use in a growing number of *in vitro* and *in vivo* applications[10,11], including the engineering of synthetic antibody libraries[12], site-directed chromosome translocations[13], targeted gene deletions[14], mosaic genomes[15] and gene-specific humanized animal models[16]. Advances in this technology have resulted in transgenic systems in which Cre-mediated genetic rearrangements can be controlled in site-specific, tissue-specific and time-specific manners[17,18].

The largest gap in our understanding of how Cre recombinase and other integrase family members catalyse the site-specific recombination reaction has been a result of a lack of three-dimensional structural models for the macromolecular complexes involved. However, the three-dimensional structures for the catalytic domains of the bacteriophage λ and HP1 integrases[3,19] and the structure of the *Escherichia coli* XerD recombinase[20] have recently been reported. Using the Cre/*loxP* system as a model, we are working to bridge this gap by studying the structures of intermediates in the site-specific recombination reaction. We have determined the three-dimensional structure at 2.4 Å resolution of a covalent intermediate in the site-specific recombination reaction between Cre recombinase and a symmetrized, suicide *loxP* substrate. This Cre–DNA complex is a structural model for a site-specific recombination synapse and provides several unexpected mechanistic insights into the recombination reaction.

Structure determination

Cre recombinase was expressed in soluble form in *E. coli* and purified to homogeneity. The purification and construction of DNA substrates for co-crystallization was simplified by replacing the asymmetric 8-base-pair spacer sequence between inverted repeats in *loxP* with the symmetric spacer shown in Fig. 1b. This modified substrate (*loxA*) was constructed from two modular half-sites that dimerize through 4-base-pair 5′ overhangs to yield a doubly nicked duplex[21]. The choice of positions for the missing phosphodiester linkages in *loxA* results in the formation of a trapped covalent intermediate following cleavage by Cre. Because the residue on the 3′ side of the cleaved phosphate is able to diffuse away, the 5′-hydroxyl group that would normally serve as a nucleophile in the subsequent strand-exchange step of the reaction is eliminated (Fig. 1c). A similar strategy was used to prepare covalent λ integrase–DNA intermediates[22]. SDS–PAGE analysis of Cre/*loxA* mixtures showed a band of reduced mobility corresponding to formation of a covalent Cre–DNA complex. Crystals of the Cre-*loxA* complex (see Methods), dissolved directly in SDS-containing buffer, showed the same slower-moving band. Crystallographic phases for the Cre-*loxA* structure were determined at 3 Å resolution using a combined multiple isomorphous replacement/anomalous scattering (MIRAS) and multiwavelength anomalous diffraction (MAD) approach using diffraction data measured at the NSLS X25 and X4A beamlines. This combined phasing approach resulted in an experimental electron density map that was readily interpretable, and refinement of the initial model with 2.4 Å data for the parent complex converged rapidly to $R = 0.201$ and $R_{free} = 0.264$ for 50–2.4 Å resolution data. Data collection statistics, phasing statistics and refinement results are summarized in Table 1.





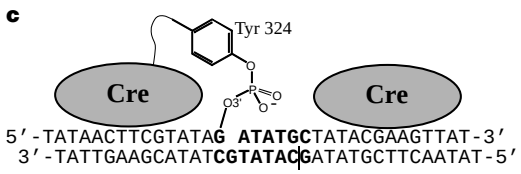


Figure 1 a, The Cre–*loxP* site-specific recombination reaction, based on studies in the lambda integrase family[1] and on the work described here. Two Cre molecules bind to *loxP* sites which then associate to form a recombination synapse (**1**). Conserved Tyr 324 cleaves each substrate to form covalent 3′-phosphotyrosine intermediates in an antiparallel arrangement (**2**). (Cleavage of the substrates could precede synapsis.) The 5′-hydroxyl groups released by cleavage of the phosphodiester linkages become nucleophiles in a strand-transfer reaction, in which they attack the phosphotyrosines of the partner substrates, yielding a Holliday-junction intermediate (**3**). A second round of cleavage and strand-exchange reactions gives recombinant products (**4**). The reaction intermediate shown in **2** represents the Cre–*loxA* synapse structure described here. **b**, Sequence of *loxP* and of the DNA substrate (which we have named *loxA*) used to form Cre/DNA cocrystals. The *loxA* construct was formed from 16-mer (5′-TATAACTTCGTATAGC-3′) and 19-mer (5′-ATATGCTATACGAAGTTAT-3′) oligonucleotides that were annealed together. The 13-bp inverted repeats in the *loxP* and *loxA* sites are underlined and the phosphate groups cleaved during the Cre–*lox* recombination reaction are indicated by arrows. The central 8 bp comprise the strand exchange, or crossover region. Vertical lines in the *loxA* sequence identify the location of missing phosphodiester linkages in the top and bottom strands. The number scheme follows refs 26, 30. **c**, The trapped 3′-phosphotyrosine intermediate formed when Cre cleaves the *loxA* substrate. The Cyt 3 residue on the 3′ side of the cleaved phosphate is able to diffuse away, eliminating the 5′-hydroxyl group that would normally serve as a nucleophile in the subsequent strand-exchange step of the reaction.

Table 1 Summary of crystallographic analysis

Data collection statistics

Derivative (no. of sites)		Source	Resolution limit (Å)	R_{sym}*	Redundancy†	Completeness (%)
Native	λ = 0.9790 Å	NSLS X25‡	2.4	0.069	5.0	93
Se-Met1§	λ = 0.9840 Å	NSLS X4A	3.0	0.067	2.7	84
Se-Met2 (22)	λ = 0.9793 Å	NSLS X4A	3.0	0.068	2.7	83
Se-Met3 (22)	λ = 0.9790 Å	NSLS X4A	3.0	0.068	2.7	83
Se-Met4 (22)	λ = 0.9790 Å	NSLS X25	2.7	0.041	2.0	97
Mersalyl (6)		Laboratory	3.0	0.051	2.6	88
Iodine-1‖ (4)	λ = 0.9790 Å	NSLS X25	3.0	0.076	4.6	87
Iodine-2 (4)		Laboratory	3.4	0.073	3.1	88
Iodine-3 (4)		Laboratory	4.0	0.087	3.7	96

Phasing statistics

Resolution limit (Å)	19.8	11.0	7.6	5.8	4.7	4.0	3.4	3.0	All
Phasing power/anomalous phasing power¶									
Se-Met2		1.5/0.93	1.6/0.84	1.6/0.83	1.4/0.72	1.1/0.59	0.83/0.52	0.66/0.44	0.98/0.58
Se-Met3		1.2/0.97	1.1/0.98	1.1/1.0	0.88/0.98	0.66/0.82	0.49/0.72	0.37/0.62	0.59/0.79
Se-Met4	0.43/0.41	0.72/0.77	0.77/0.74	0.93/0.85	0.76/0.85	0.52/0.73	0.45/0.63	0.46/0.54	0.56/0.68
Mersalyl	0.58	1.5	1.6	1.7	1.2	0.83	0.78	0.82	0.96
Iodine-1	0.88	0.76	0.79	0.88	0.62	0.41	0.36	0.37	0.47
Iodine-2	0.61	0.97	1.1	1.1	0.72	0.56	0.51		0.64
Iodine-3	0.66	1.3	1.0	0.96	0.70	0.52			0.70
Mean figure of merit	0.44	0.76	0.77	0.78	0.70	0.58	0.50	0.43	0.56

Refinement results

	Resolution (Å)	*R*-factor#	Free *R*-factor#	No. of reflections
Data with $F > 2\sigma$	50–2.4	0.201	0.264	41,368
All data	50–2.4	0.208	0.272	42,896

R.m.s. deviations: Bond lengths 0.005 Å; Bond angles 1.2°; *B* values (bonded) 3.7 Å

* $R_{sym} = \Sigma|I_h - \langle I_h\rangle|/\Sigma I_h$, where $\langle I_h\rangle$ is the average intensity over symmetry equivalent measurements.
† Redundancy is the number of measured data/number of unique data.
‡ NSLS X25, NSLS X4A: National Synchrotron Light Source beamlines X25 and X4A, Brookhaven National Laboratory.
§ Se-Met1–3 data were measured from the same crystal in a multiple-wavelength anomalous diffraction experiment. Dispersive differences (treated as isomorphous differences) for Se-Met2 and Se-Met3 are with respect to Se-Met1. Se-Met4 data were measured from a different crystal and isomorphous differences are with respect to the native data.
‖ Iodine-1, short strand in Fig. 1b is changed to 5′-TAXAACTTCGTAXAGC-3′, where X is 5-iodo-dU; iodine-2, long strand is changed to 5′-ATATGCXAXACGAAGTTAT-3′; iodine-3, long strand is changed to 5′-ATATGCTATACGAAGXXAT-3′.
¶ Phasing power $= \Sigma|F_H|/\Sigma||F_{PH}\text{obs}| - |F_{PH}\text{calc}||$; anomalous phasing power $= \Sigma|F_H''|/\Sigma||\text{AD}_{obs}| - |\text{AD}_{calc}||$, where AD is the anomalous difference.
# *R*-factor $= \Sigma|F_{obs} - F_{calc}|/\Sigma F_{obs}$, where summation is over data used in the refinement; free *R*-factor includes only the 10% of data excluded from all refinements.

Architecture of the Cre–*loxA* complex

The structure of the Cre–*loxA* complex reveals two Cre molecules bound to a single *loxA* site (Fig. 2). Each Cre molecule contacts the outermost 15 base pairs of one *loxA* half-site, which includes the 13-base-pair inverted-repeat sequence and the first two base pairs of the central strand-exchange region. One of the two Cre molecules bound to the *loxA* site has cleaved the DNA substrate to form a covalent 3′-phosphotyrosine linkage. Two antiparallel Cre–*loxA* complexes are associated in the crystal to form a dyad-symmetric recombination synapse (Fig. 5a). A similar protein–protein interface is formed both between Cre molecules bound to the same *loxA* site and between Cre molecules bound to different *loxA* sites in the synapse, generating pseudo-C_4 symmetry in the synaptic assembly. The cleaved *loxA* sites are bent by ~100° at their centres, resulting in a pseudo-fourfold arrangement of *loxA* half-site arms in the synapse that resembles a square planar isomer of a Holliday-junction intermediate[23,24]. The protein:DNA scaffold surrounds a 'strand exchange cavity' which contains the central six bases of the crossover region. Single-stranded DNA segments that are released by the cleavage reaction are not contacted by the recombinase tetramer and have both flexibility and unencumbered access to the active sites of the Cre-*loxA* synapse.

Structure of Cre recombinase

Cre folds into two distinct domains that are separated by a short linker (Fig. 3a, b). The amino-terminal domain (residues 20–129) contains five α-helical segments connected by short loops. Three of the helices (C, D and E) are organized into an antiparallel bundle. Helices A and B are nearly orthogonal to the three-helix bundle, with helix B in close contact with a hydrophobic bundle surface. Helix A is only loosely associated with the rest of the domain. As discussed below, helices A and E are involved in formation of the recombinase tetramer and helices B and D contact the major groove of the *loxA* half-site. The two independent amino-terminal domain conformations in the Cre-*loxA* complex are similar, with an r.m.s. deviation of 0.7 Å for α-carbon atoms.

The larger carboxy-terminal domain of Cre recombinase includes amino acids 132–341 and is primarily helical. A small β-sheet (strands 1–3) packs against one surface of a nine-helix domain (helices F–N). The terminal helix N extends away from the rest of the domain and is involved in intersubunit contacts. The C-terminal domain of Cre recombinase is structurally similar to the lambda integrase and HP1 integrase catalytic domains[3,19]. But unlike the amino-terminal domain, the C-terminal domain differs significantly between the two *loxA*-bound Cre molecules, with the

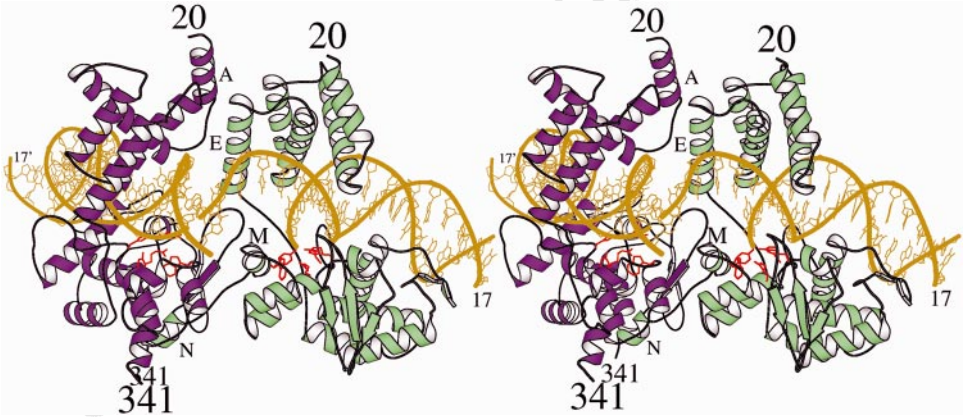
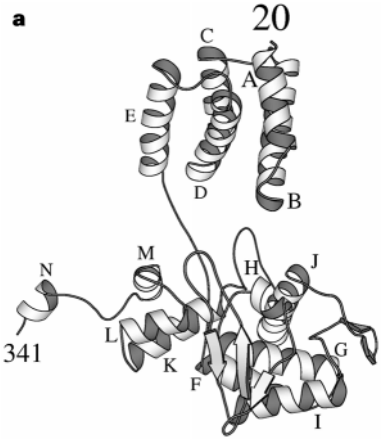


Figure 2 Stereo ribbon model[49] of the Cre–*loxA* complex. The Cre subunit that has cleaved the *loxA* site is in green (right) and the Cre subunit that has not cleaved the *loxA* site is purple (left). Side chains of the active-site residues (Arg 173, His 289, Arg 292, Trp 315 and Tyr 324) are in red. The N- and C-terminal amino-acid numbers, the first and last bases in the uncleaved DNA strand, and selected α-helices are labelled.



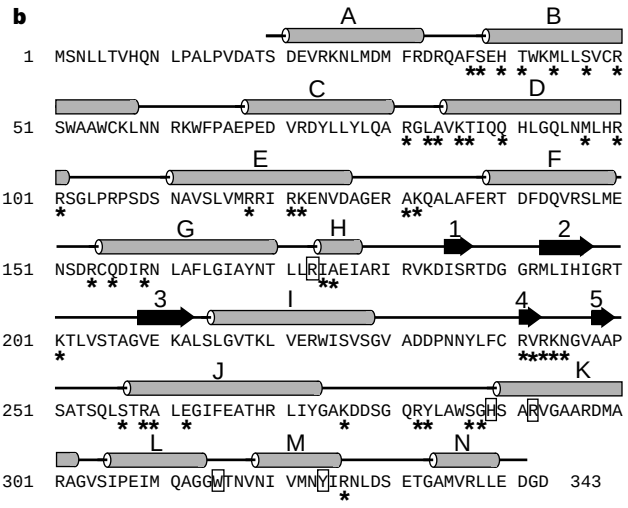


Figure 3 a, Ribbon model[49] of Cre recombinase. α-Helices are labelled A–N and the amino and carboxy termini are indicated by residue numbers. **b**, Secondary structure assignment for Cre recombinase. Active-site residues are boxed; asterisks indicate amino acids that contact DNA (<3.6 Å) in the subunit that has cleaved the *loxA* site.

largest structural deviations coming from the C-terminal segments leading to helix N. The two C-terminal domain structures have r.m.s. deviations of 2.5 Å for all α-carbon atoms and 1.5 Å for α-carbon atoms omitting the final 17 amino acids.

The Cre–*loxA* interface

As shown in Fig. 2, two Cre molecules are bound to the *loxA* site. Each Cre molecule contacts a *loxA* half-site and one of the two molecules has formed a covalent 3′-phosphotyrosine linkage with the DNA. The amino and carboxy-terminal domains of Cre form a clamp around the half-site, making extensive contacts with both the major and minor grooves. This Cre–DNA interface buries ~5,000Å^2 of solvent-accessible surface area for each Cre–half-site interaction. The contacts we observe are in good agreement with the results of DNase and hydroxyl radical footprinting studies using both full-length Cre and a Cre deletion mutant lacking the N-terminal domain[25,26]. The number of electrostatic interactions between Cre and the phosphate backbone of the DNA substrate is striking. A total of 39 contacts to the *loxA* site involving arginine, lysine and histidine side chains are formed by the two Cre subunits. Although the two Cre–DNA interfaces are similar, a clear asymmetry exists that may result in part from conformational changes associated with cleavage of the DNA substrate by only one of the two Cre subunits.

The N-terminal domain contacts the DNA primarily via the orthogonal B and D helices, burying ~1,800 Å^2 of solvent-accessible surface area in the interface formed by the non-cleaving Cre subunit and ~2,300 Å^2 in the interface formed by the cleaving subunit. The crossed B and D helices sit in the major groove, centred at positions 5–6 and 5′–6′ of the *loxA* sites, and extend about three base pairs in each direction. Together, helix B and helix D make three direct contacts to DNA bases. Helix E of the amino-terminal domain in the cleaving Cre subunit is aligned directly over the DNA backbone, where it forms three close electrostatic interactions with phosphates at the concave surface of the bent *loxA* site. This interface is different in the non-cleaving subunit, where helix E is instead positioned between adjacent *loxA* substrates in the synapse (Figs 2, 5a).

The C-terminal domain interacts with the entire 13-base-pair inverted-repeat region of the *loxA* half-site plus the first two base pairs of the strand-exchange region. This interface is more complex, involving the entire face of the domain, with both secondary structural elements and connecting loops interacting with the major groove, the minor groove and the sugar–phosphate backbone. The protein surface area buried at this interface is ~3,000 Å^2 for both the cleaving and non-cleaving subunits, but only a small fraction of this surface involves major-groove interactions. The amino terminus of helix J interacts with the major groove near base pairs 10–11, where Arg 259 forms hydrogen bonds to N7 and O6 of guanine 10 (Gua 10). This is the only direct major-groove contact we observe involving the C-terminal domain. A striking feature of this domain–DNA interface is the number of direct contacts to bases in the minor groove. A short β-strand between helices I and J makes three direct contacts. For example, the side chain of Lys 244 bridges thymine (Thy)-16 O2 and Thy-17 O2. Similarly, the loop between strands 2 and 3 of the β-sheet bridges Gua-4 N3 and Thy-5 O2 with the side chain of Lys 201. A third minor-groove interaction involves Arg 282 in the loop between helices J and K, which forms hydrogen bonds to adenine(Ade)-7 N3.

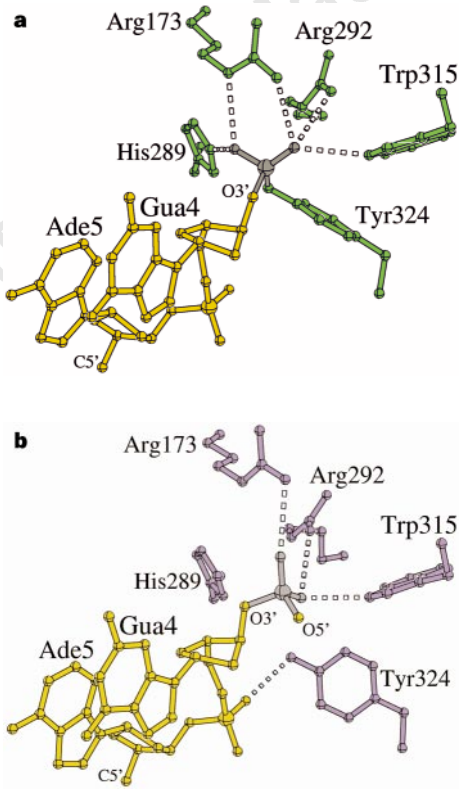


Figure 4 The Cre-*loxA* active sites. Perspective of the active site **a**, where Cre recombinase has cleaved the DNA substrate to form a 3′-phosphotyrosine linkage; and **b**, where Cre has not cleaved the DNA substrate.

Active sites and the phospho-Tyr linkage

The Cre active site contains the conserved catalytic triad residues Arg 173, His 289 and Arg 292, the conserved nucleophilic Tyr 324 and Trp 315. The amino acids in a single active site are all derived from the same subunit, indicating that Cre recombinase does not form a shared active site like Flp recombinase[27]. There are two distinct active sites in the Cre–*loxA* complex, one for each Cre subunit bound to a *loxA* half-site. As the synapse is formed by the association of two Cre–*loxA* complexes, the synapse tetramer contains a total of four active sites, corresponding to the four scissile phosphates in the site-specific recombination reaction. One active site in the Cre–*loxA* complex (two per synapse) contains a 3′-phosphotyrosine DNA–Cre linkage resulting from cleavage of the DNA substrate. This covalently attached phosphate is activated for the strand-exchange step of the reaction, which requires nucleophilic attack of the phosphotyrosine by the 5′-hydroxyl released upon cleavage of the partner substrate (Fig. 1a).

Figure 4a shows the stereochemistry of the active site after cleavage of the DNA substrate. The non-bridging oxygen atoms of the scissile phosphate are coordinated by five hydrogen bonds. Three hydrogen bonds come from the guanidino groups of Arg 173 and Arg 292, which contact the oxygen atoms through their Nε and Nη hydrogen atoms. One oxygen atom also receives a hydrogen bond from the Nε hydrogen of His 289, which is ideally positioned to serve as proton donor to the Tyr 324 leaving group during the strand-transfer reaction. An additional hydrogen bond to the second phosphate oxygen atom comes from the side chain of Trp 315. This tryptophan is not conserved in the integrase family, and a histidine residue is more often found in this position. The side chains of histidine and tryptophan could, however, donate similar hydrogen bonds and this functionality may well be conserved in the integrase family. Together, the active-site His, Arg and Trp side chains form a positively charged 'proton cradle' that appears to be preorganized to accommodate the negative charges of equatorial phosphate oxygen atoms in a trigonal bipyramidal transition state of the phosphoryl transfer reaction.

The other active site in the Cre–*loxA* complex surrounds a scissile phosphate that has not been cleaved by Cre (Fig. 4b). A superposition of the two Cre subunits shows that the catalytic arginine side chains and the Trp-315 side chain occupy similar positions in

this active site. Tyr 324, however, is shifted away from the scissile phosphate with a phosphorus–hydroxyl distance of 5.8 Å. In this conformation, the tyrosine hydroxyl group forms a hydrogen bond with the phosphate group between Gua 4 and Ade 5. This change in structure is partly due to a concerted 3 Å shift in helix M, which contains Tyr 324. His 289 has also shifted away from the scissile phosphate in this active site. We suggest that the stereochemical differences between the two active sites may in part reflect the allosteric consequences of the initial cleavage reaction. The resulting asymmetry is likely to be of mechanistic importance, because the stepwise nature of the integrase family site-specific recombination reaction requires that only one half-site in each substrate be cleaved in preparation for a given exchange of DNA strands.

Interactions in the synaptic complex

The interface between Cre subunits in the synapse tetramer is extensive, with 13,800 Å^2 of solvent-accessible protein surface buried overall, or 3,450 Å^2 buried per subunit. The contacts between Cre molecules bound to the same *loxA* site are similar to the contacts between Cre molecules bound to different *loxA* sites, giving rise to the pseudo-fourfold symmetry of the synapse assembly. The N-terminal domain contributes to this network of interactions primarily with helices A and E, which form a parallel helix–helix interface (Figs 2, 5a). The C terminus of helix E also contacts the β-turn between strands 2 and 3 of the C-terminal domain in the adjacent subunit. The interactions between C-terminal domains in neighbouring subunits are intriguing. The segment following helix M (amino acids 326–333) leads to a short C-terminal helix (N) that buries one face in a hydrophobic pocket of the adjacent subunit (Fig. 2). The hydrophobic pocket is formed partly by the L and M helices, which are integral components of the recombinase active sites. This intermolecular grapple crosslinks Cre subunits in a cyclic manner (Fig. 5c) and may provide a means of communication through which changes in the conformation of helix M and in the segment leading to helix N are relayed to the adjacent subunit. A reciprocal, rather than cyclic, exchange of helices was observed in the structure of the HP1 integrase catalytic domain, where the domains crystallized as dimers that mutually exchanged their C-terminal helix homologues of Cre[19]. RuvA, a protein involved in mediating branch migration of Holliday junctions, uses an intermolecular organization in forming a tetramer that is similar to the cyclic exchange observed in the Cre–*loxA* synapse[28].

The two Cre subunits bound to the *loxA* substrate differ in the conformation of the segment connecting helix M and helix N. Helix M in the subunit that has cleaved *loxA* is shifted by ~3 Å relative to the same helix in the non-covalently bound subunit, in order to form the phosphotyrosine linkage. As a consequence, the segment connecting helices M and N adopts a more extended conformation

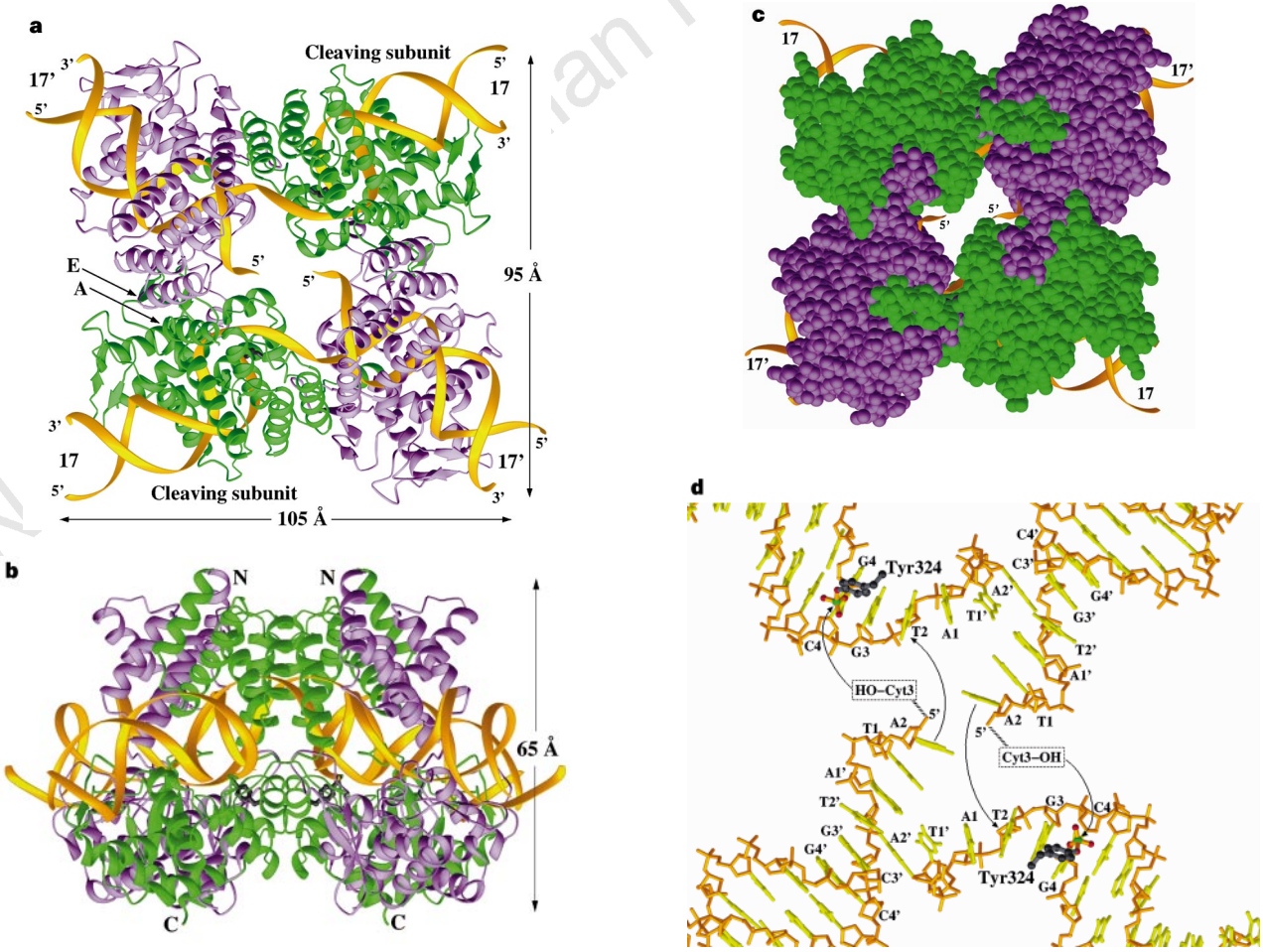


Figure 5 Ribbon and space-filling models[50] of the Cre–*loxA* synapse. **a**, View from the N-terminal domain side of the tetramer, parallel to the dyad of symmetry and to the pseudo-fourfold axis of symmetry. Helices A and E are indicated. **b**, Side view, rotated 90° about a horizontal axis with respect to **a**. Approximate positions of the N and C termini are indicated. Uncleaved DNA strands are drawn as continuous ribbons in **a** and **b**. **c**, Space-filling model of the recombinase tetramer, showing the cyclic interaction between subunits involving the C-terminal segments of Cre; the view is from the C-terminal side of the synapse. **d**, Close-up of the DNA substrate in the same orientation as in **c**, with the recombinase subunits omitted. The direction of nucleophilic attack by missing Cyt 3 and formation of a recombinant base-pair between Ade 2 and Thy 2 in the strand-exchange step of the recombination reaction are indicated by arrows. Tyr 324, visible in **b** and **d**, is in black.

in the cleaving subunit (Fig. 2). In the non-cleaving Cre subunit, the position of helix M causes the M–N linker segment to adopt a more compact conformation involving formation of a short 3_{10} helix and a different set of interactions with the cleaving subunit located across the synapse on the partner substrate (Fig. 5c).

Structure of the synapsed *loxA* DNA

The synapsed loxA sites have the overall appearance of a distorted four-fold symmetric Holliday structure, with arms formed by the duplex inverted repeat segments of the *loxA* half-sites (Fig. 5a, d). However, as the strand-exchange step of the recombination reaction has been blocked, the covalent bonds leading to a Holliday intermediate have not been formed. There is a smooth bend of ~25° within the inverted repeat arms that is present in both the cleaved and uncleaved half-sites. This bend produces a concave DNA surface on one side of the synapse that is evident in Fig. 5b. An increase in the width of the minor groove occurs in the regions involving side chain–minor groove base contacts, but the helical parameters of these duplex segments are otherwise generally consistent with those of B-form DNA[29].

The crossover region in the cleaved *loxA* site is composed of both single- and double-stranded segments (Fig. 5d). In the cleaved strand (top strand in Fig. 1b, c), Gua 4 maintains standard base-pairing geometry in spite of the phosphotyrosine linkage. Cytosine 3 has been released by the cleavage reaction and is missing from the electron density, as expected. Ade 2 and Thy 1 are not base-paired with the uncleaved strand, but instead have moved towards the centre of the strand-exchange cavity, where their weak electron density and high thermal parameters indicate a high degree of mobility. The remaining four nucleotides (Ade 1′ to Cyt 4′) in the cleaved strand are base-paired, with a trend of decreasing mobility and increasing electron density quality nearer to the end of the strand-exchange region. The first two single-stranded bases in the uncleaved strand (Gua 3 and Thy 2) are stacked against neighbouring Cyt 4 and have well-defined conformations. Ade 1 is poorly stacked against its neighbours and provides the nucleation point for a bend in the *loxA* site that leads to a ~100° disposition between half-site arms. This base has a poorly defined conformation in the Cre–*loxA* structure and is the only position within the strand-exchange region of *loxP* that was previously found to be intolerant to mutation[30].

To investigate whether the framework of the Cre–*loxA* synapse has the appropriate geometry to mediate the strand-exchange reaction, we modelled the missing Cyt 3 to the 5′ end of Ade 2 and manually moved the DNA chain towards the phosphotyrosine of the partner substrate (arrows in Fig. 5d). By adjusting their backbone torsion angles, Cyt 3 and Ade 2 could be base-paired with Gua 3 and Thy 2 of the partner substrate, placing the 5′-hydroxyl of Cyt 3 in an appropriate position for nucleophilic attack. Only the torsion angles of Thy 1, Ade 2 and Cyt 3 needed to be changed to adopt this strand-attack geometry. From this analysis, the strand-exchange step of the site-specific recombination reaction could occur in the nucleoprotein architecture of the Cre–*loxA* synapse without major alterations in quaternary structure.

Cre–*loxP* site-specific recombination

Models of the λ integrase family site-specific recombination reaction have until recently been based on the assumption that the branch point of the Holliday junction intermediate is formed at the initial strand-exchange site. To exchange a second set of DNA strands, branch migration and isomerization of the Holliday structure would be required to bring the appropriate pair of cleavage sites into mutual proximity[31]. The requirement for homology between recombining sites has been explained by the need for branch migration of the Holliday junction between cleavage sites, which would stall upon encountering a mismatch[1,32]. In contrast, we find that the four active sites in the Cre–*loxA* synapse are symmetrically positioned around their respective scissile phosphates following the first cleavage step of the recombination reaction. In this arrangement, the distances between adjacent active sites are all similar and the *loxA* half-site arms adopt a pseudosquare planar conformation. We therefore propose that the Holliday junction intermediate adopts a nearly four-fold symmetric structure that is stabilized by Cre–DNA interactions and that the cleavage and strand-exchange steps of the recombination reaction occur within the Cre/DNA framework, with no requirement for large-scale protein–DNA isomerization steps (Fig. 1a). A more recent model for site-specific recombination in the λ integrase family requires only limited branch migration of the Holliday junction intermediate, in which the branch point is formed near the centre of the crossover region[33–35]. Homology between recombining sites is tested in this model by formation of recombinant base pairs during the strand-exchange steps of the reaction. The Cre–*loxA* structure strongly supports this model; the arrows in Fig. 5d indicate how two such recombinant base pairs might be formed.

The stepwise nature of the integrase-family recombination mechanism requires that only one half-site of the DNA substrate be cleaved before strand exchange. What prevents double-strand cleavage? In the synaptic complex, a number of subtle conformational changes involving the C-terminal 40 amino acids may accompany the first strand-cleavage step and prevent attack of the neighbouring half-site through allosteric effects. In the unsynapsed Cre–DNA complex, where two Cre molecules bind to a *loxP* site, cleavage at only one half-site may be ensured by the formation of only a single competent active site. As shown in Fig. 2, only one of the C-terminal segments in a Cre–*loxP* complex would be able to interact with its adjacent Cre subunit, even if the bend in the uncleaved *loxP* site were somewhat different. Without the appropriate intersubunit contacts, the second C-terminal segment may adopt a different structure, or it may not adopt a unique conformation, as seen in the structure of the isolated λ integrase catalytic domain[3]. It is also unclear how the structure of Cre recombinase would change when bound to an isolated half-site substrate, where there would be no intersubunit contacts. This could be important for interpreting experiments that have relied on the structural integrity of the recombinases when bound to modified DNA half-site probes[36].

The three short helices (L–N) that comprise the C-terminal 40 amino acids of Cre also lie at the heart of a mechanistic paradox in the integrase family[36]. In question is the source of the conserved tyrosine that is responsible for strand cleavage at a given half-site. Does it come from the same recombinase molecule that provides the other active site residues for that half-site (*cis* cleavage) or is it from one of the remaining three subunits (*trans* cleavage)? The Cre–*loxA* structure presented here shows that Cre recombinase undergoes *cis* cleavage of the *loxA* site, in disagreement with a recent report[37]. Lambda integrase and the Xer recombinases both cleave their recombination sites in *cis*[38,39]; Flp recombinase cleaves in *trans*[27]. The answer may lie both in the context-dependent nature of the C termini of the recombinases and in subtle structural differences among integrase family members. Based on the Cre–*loxA* complex, *trans* donation of helix M (which contains Tyr 324) to an adjacent active site is plausible: the helical segments are close together and only small changes in structure would be required (Fig. 2). This is particularly feasible for Flp recombinase, which contains a slightly longer linker between helices L and M[40]: indeed, modelling of the Cre–*loxA* structure indicates that, given a longer linker, the M helix could be donated *in trans* and the remaining C-terminal residues could loop back to the same subunit to bury helix N *in cis*. These interactions would generate a '*trans*-cyclic' arrangement in the synapse, a scheme that is similar to the 'asymmetric' mode proposed for Flp recombinase[41].

In summary, our structure of Cre recombinase bound to a suicide DNA substrate provides a model for a λ integrase family site-specific

recombination synapse. The nearly four-fold symmetric tetramer of Cre–DNA half-sites formed by the association of two Cre–*loxA* complexes suggests a mechanism for recombination that does not involve branch migration of the Holliday junction intermediate between cleavage sites or isomerization of the protein–DNA architecture. Intersubunit contacts between C-terminal segments of Cre may regulate the ordered cleavage and strand-exchange steps of the reaction and these regions are likely to adopt conformations that depend on their molecular context. □

Methods

Crystallization of Cre–*loxA*. A Cre expression construct was prepared by polymerase chain reaction (PCR) amplification of the Cre coding region from bacteriophage P1 DNA and ligation into pET21a (Novagen). Both Cre and Se-Met Cre were purified from bacterial lysates on SP-Sepharose, Source-15S (Pharmacia) and CH15 hydroxyapatite (BioRad) columns. Crystals of the Cre–*loxA* complex were grown at 18 °C by the hanging-drop method. Initial drops contained 4–8 μl 50 mM sodium acetate buffer, pH 5, 12% 2-methyl-2,4-pentanediol (MPD), 10 mM $CaCl_2$, 50 μM Cre, 75 μM *loxA* half-site, and reservoirs contained 100 mM acetate, pH 5, 24% MPD, 20 mM $CaCl_2$. Orthorhombic crystals ($a = 107.7, b = 121.0, c = 180.4$ Å; $C222_1$) grew in 3–4 days and reached a maximum size of $0.2 \times 0.2 \times 0.04$ mm after 2–4 weeks.

Structure determination and refinement. Crystals were mounted in loops made of 20-μm nylon filaments and flash-frozen in liquid propane[42]. The asymmetric unit of this crystal form contained two Cre molecules and two *loxA* half-sites; the crystals diffracted well to 3 Å at a temperature of 100 K on laboratory sources and to at least 2.4 Å using synchrotron sources. Data were measured using a rotating anode X-ray source equipped with Charles Supper focusing mirrors and a MAR Research image-plate detector; NSLS X25 data were measured using a MAR Research image-plate detector, and NSLS X4A data were measured using Fuji image plates. All data were measured at 90–100 K and processed as described[42]. Two sets of dispersive differences (with respect to wavelength 1) and anomalous differences from a three-wavelength MAD experiment performed at NSLS X4A were treated as isomorphous and anomalous differences, respectively, and used together with the MIRAS data in a locally modified maximum-likelihood procedure[43] to phase the parent Cre/loxA structure. Modification of the solvent region[44] and phase combination with experimental probability distributions resulted in a clean electron density map with interpretable density for protein residues 21–340 and nearly all of the DNA bases. The two Cre molecules in the asymmetric unit were independently traced and non-crystallographic symmetry averaging was not used at any stage of the structure solution or refinement.

The Cre–*loxA* model was refined using simulated annealing coupled with torsion angle dynamics[45] at 3.0, 2.7 and 2.4 Å resolution, with model adjustments following each refinement round[46]. A bulk solvent correction[47] allowed refinement with no low-resolution cutoff in the reflection data. A new parameter set was used to restrain nucleic acid geometry[48], with no torsional restraints imposed on the phosphotyrosine linkage. The final model includes 644 out of 686 amino acids (residues 20–341 in both Cre molecules), all nucleotides except one of the two overhanging 5′ T residues, and 488 water molecules with unit occupancy and good hydrogen-bonding stereochemistry. All amino acids have (ϕ, ψ) backbone torsion angles in allowed regions of Ramachandran space.

Received 12 June; accepted 15 July 1997.

1. Craig, N. L. The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* **22,** 77–105 (1988).
2. Stark, W. M., Boocock, M. R. & Sherratt, D. J. Catalysis by site-specific recombinases. *Trends Genet.* **8,** 432–439 (1992).
3. Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276,** 126–131 (1997).
4. Argos, P. *et al.* The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* **5,** 433–440 (1986).
5. Abremski, K. E. & Hoess, R. H. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng.* **5,** 87–91 (1992).
6. Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* **58,** 913–949 (1989).
7. Abremski, K., Hoess, R. & Sternberg, N. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell* **32,** 1301–1311 (1983).
8. Abremski, K. & Hoess, R. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.* **259,** 1509–1514 (1984).
9. Sternberg, N., Hamilton, D., Austin, S., Yarmolinsky, M. & Hoess, R. Site-specific recombination and its role in the life cycle of bacteriophage P1. *Cold Spring Harbor Symp. Quant. Biol.* **1,** 297–309 (1981).
10. Sauer, B. Manipulation of transgenes by site-specific recombination: use of Cre recombinase. *Meth. Enzymol.* **225,** 890–900 (1993).
11. Kilby, N. J., Snaith, M. R. & Murray, J. A. Site-specific recombinases: tools for genome engineering. *Trends Genet.* **9,** 413–421 (1993).
12. Tsurushita, N., Fu, H. & Warren, C. Phage display vectors for *in vivo* recombination of immunoglobulin heavy and light chain genes to make large combinatorial libraries. *Gene* **172,** 59–63 (1996).
13. Qin, M., Bayley, C., Stockton, T. & Ow, D. W. Cre recombinase-mediated site-specific recombination between plant chromosomes. *Proc. Natl Acad. Sci. USA* **91,** 1706–1710 (1994).
14. Lakso, M. *et al.* Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89,** 6232–6236 (1992).
15. Betz, U. A., Vosshenrich, C. A., Rajewsky, K. & Muller, W. Bypass of lethality with mosaic mice generated by Cre-loxP-mediated recombination. *Curr. Biol.* **6,** 1307–1316 (1996).
16. Zou, Y. R., Muller, W., Gu, H. & Rajewsky, K. Cre-loxP-mediated gene replacement: a mouse strain producing humanized antibodies. *Curr. Biol.* **4,** 1099–1103 (1994).
17. Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice. *Science* **269,** 1427–1429 (1995).
18. Metzger, D., Clifford, J., Chiba, H. & Chambon, P. Conditional site-spedific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc. Natl Acad. Sci. USA* **92,** 6991–6995 (1995).
19. Hickman, A. B., Waninger, S., Scocca, J. J. & Dyda, F. Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. *Cell* **89,** 227–237 (1997).
20. Subramanya, H. S. *et al.* Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **17,** 5178–5187 (1997).
21. Schultz, S. C., Shields, G. C. & Steitz, T. A. Crystallization of *Escherichia coli* catabolite gene activator protein with its DNA-binding site. The use of modular DNA. *J. Mol. Biol.* **213,** 159–166 (1990).
22. Pargellis, C. A., Nunes-Duby, S., de, V. L. & Landy, A. Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. *J. Biol. Chem.* **263,** 7678–7685 (1988).
23. Sigal, N. & Alberts, B. Genetic recombination: the nature of crossed strand-exchange between two homologous DNA molecules. *J. Mol. Biol.* **71,** 789–793 (1972).
24. Duckett, D. R. *et al.* The structure of the Holliday junction, and its resolution. *Cell* **55,** 79–89 (1988).
25. Hoess, R. H. & Abremski, K. Interaction of the bacteriophage P1 recombinase Cre with the recombining site loxP. *Proc. Natl Acad. Sci. USA* **81,** 1026–1029 (1984).
26. Hoess, R., Abremski, K., Irwin, S., Kendall, M. & Mack, A. DNA specificity of the Cre recombinase resides in the 25 kDa carboxyl domain of the protein. *J. Mol. Biol.* **216,** 873–882 (1990).
27. Chen, J. W., Lee, J. & Jayaram, M. DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell* **69,** 647–658 (1992).
28. Rafferty, J. B. *et al.* Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* **274,** 415–420 (1996).
29. Lavery, R. & Sklenar, H. The definition of generalised helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6,** 63–91 (1988).
30. Hoess, R. H., Wierzbicki, A. & Abremski, K. The role of the loxP spacer region in P1 site-specific recombination. *Nucleic Acids Res.* **14,** 2287–2300 (1986).
31. Stark, W. M., Sherratt, D. J. & Boocock, M. R. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell* **58,** 779–790 (1989).
32. Kitts, P. A. & Nash, H. A. Homology-dependent interactions in phage lambda site-specific recombination. *Nature* **329,** 346–348 (1987).
33. Nunes-Duby, S., Azaro, M. A. & Landy, A. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr. Biol.* **5,** 139–148 (1995).
34. Arciszewska, L. K., Grainge, I. & Sherratt, D. J. Action of site-specific recombinases XerC and XerD on tethered Holliday junctions. *EMBO J.* **16,** 3731–3743 (1997).
35. Azaro, M. A. & Landy, A. The isometric preference of Holliday junctions influences resolution bias by lambda integrase. *EMBO J.* **16,** 3744–3755 (1997).
36. Stark, W. M. & Boocock, M. R. Gatecrashers at the catalytic party. *Trends Genet.* **11,** 121–123 (1995).
37. Shaikh, A. C. & Sadowski, P. D. The Cre recombinase cleaves the lox site in trans. *J. Biol. Chem.* **272,** 5695–5702 (1997).
38. Nunes-Duby, S. *et al.* Lambda integrase cleaves DNA in *cis*. *EMBO J.* **13,** 4421–4430 (1994).
39. Arciszewska, L. K. & Sherratt, D. J. Xer site-specific recombination in vitro. *EMBO J.* **14,** 2112–2120 (1995).
40. Blakely, G. W. & Sherratt, D. J. *Cis* and *trans* in site-specific recombination. *Mol. Microbiol.* **20,** 233–238 (1996).
41. Qian, X. H., Inman, R. B. & Cox, M. M. Protein-based asymmetry and protein-protein interactions in FLP recombinase-mediated site-specific recombination. *J. Biol. Chem.* **265,** 21779–21788 (1990).
42. Van Duyne, G. D., Ghosh, S., Maas, W. K. & Sigler, P. B. Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli*. *J. Mol. Biol.* **256,** 377–391 (1996).
43. Otwinowski, Z. in *CCP4 Proc.* 80-88 (Daresbury Laboratory, Warrington, UK, 1991).
44. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochontrial F1 ATPase. *Acta Crystallogr. D* **52,** 30–42 (1996).
45. Rice, L. M. & Brunger, A. T. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Prot. Struct. Funct. Genet.* **19,** 277–290 (1996).
46. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
47. Jiang, J. S. & Brunger, A. T. Protein hydration observed by X-ray diffraction: solvation properties of penicillopepsion and neuraminidase crystal structures. *J. Mol. Biol.* **243,** 100–115 (1994).
48. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T. & Berman, H. New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr. D* **52,** 57–64 (1996).
49. Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24,** 946–950 (1991).
50. Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24,** 958–961 (1991).

Acknowledgements. We thank L. Berman and Z. Yin for access to and help with the NSLS X25 beamline, W. Hendrickson and C. Ogata for access to and help with the NSLS X4A beamline, J. Ni for assistance with data collection, Y.-M. Zheng for technical support, R. Hoess for valuable discussions, and M. Lemmon, H. Nelson, M. Lewis, H. Lu and K. Ferguson for helpful comments. This work was partly supported by ACS IRG and NCI core support grants through the V. Penn Cancer Center.

Correspondence and requests for materials should be addressed to G.D.V. (e-mail: vanduyne@crystal.med.upenn.edu). Coordinates have been deposited in the protein data bank at Brookhaven National Laboratory with accession code 1CRX.