Published online 1 February 2022

Nucleic Acids Research, 2022, Vol. 50, No. 3 ***1753–1769***
https://doi.org/10.1093/nar/gkac032

Mechanisms of Cre recombinase synaptic complex assembly and activation illuminated by Cryo-EM

Kye Stachowski [1], Andrew S. Norris [1,2], Devante Potter[1], Vicki H. Wysocki [1,2] and Mark P. Foster [1,*]

[1]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210, USA and [2]Resource for Native Mass Spectrometry Guided Structural Biology, The Ohio State University, Columbus, OH 43210, USA

Received September 21, 2021; Revised January 04, 2022; Editorial Decision January 07, 2022; Accepted January 12, 2022

ABSTRACT

Cre recombinase selectively recognizes DNA and prevents non-specific DNA cleavage through an orchestrated series of assembly intermediates. Cre recombines two *loxP* DNA sequences featuring a pair of palindromic recombinase binding elements and an asymmetric spacer region, by assembly of a tetrameric synaptic complex, cleavage of an opposing pair of strands, and formation of a Holliday junction intermediate. We used Cre and *loxP* variants to isolate the monomeric Cre-*loxP* (54 kDa), dimeric Cre_2-*loxP* (110 kDa), and tetrameric Cre_4-$loxP_2$ assembly intermediates, and determined their structures using cryo-EM to resolutions of 3.9, 4.5 and 3.2 Å, respectively. Progressive and asymmetric bending of the spacer region along the assembly pathway enables formation of increasingly intimate interfaces between Cre protomers and illuminates the structural bases of biased *loxP* strand cleavage order and half-the-sites activity. Application of 3D variability analysis to the tetramer data reveals constrained conformational sampling along the pathway between protomer activation and Holliday junction isomerization. These findings underscore the importance of protein and DNA flexibility in Cre-mediated site selection, controlled activation of alternating protomers, the basis for biased strand cleavage order, and recombination efficiency. Such considerations may advance development of site-specific recombinases for use in gene editing applications.

INTRODUCTION

We report Cryo-EM structural studies of DNA recognition and assembly of recombination intermediates by the enzyme Cre (causes recombination). Cre is a tyrosine site-specific DNA recombinase (YSSR), originally discovered in the P1 phage, where Cre functions to ensure proper partitioning of prophage replicons during bacterial cell division (1–3). Exploitation of Cre in biotechnology applications has enabled precise removal or replacement of defective DNA through a highly controlled process that precludes DNA lesions elevating its candidacy as an effective gene editing technology (4–8). Cre excises, exchanges, or inverts double stranded DNA beginning with recognition of two 34-bp asymmetric *loxP* DNA sites. Each *loxP* site comprises two palindromic 13-base pair (bp) recombinase binding elements (RBEs) that are recognized in antiparallel fashion by two Cre proteins (Figure 1A and B). The RBEs are separated by an 8-bp asymmetric spacer that determines the orientation of each site and the outcome of recombination reactions (9–12). Four molecules of Cre assemble with two *loxP* DNA sites, in alternating fashion, and tyrosine residues of each protomer perform a series of strand cleavages, exchanges and ligations, forming covalent 3′-phosphotyrosine and Holliday junction intermediates to generate recombinant products. Cre catalyzes this reaction without the need of additional factors or consumption of ATP, and thus, has emerged as a powerful tool for genome engineering in the laboratory (13). However, expanded use of the technology for targeting arbitrary or asymmetric recognition sites is hindered by our limited understanding of the molecular features that control site selection, DNA cleavage, and recombination (4–8).

Tetrameric Cre-*loxP* structure

Much of our understanding of the structural bases for site selection, assembly, and recombination of *loxP* DNA by Cre comes from a series of crystal structures of similar tetrameric assemblies (a quasi-equivalent trimeric complex has also been reported) (12,14–17). However, until now, high-resolution structural models of assembly precursors have not been available. Analytical ultra-centrifugation (AUC) and mobility shift assays have shown that Cre is monomeric in solution and cooperatively oligomerizes at nanomolar concentrations in the presence of a *loxP* DNA substrate (Figure 1C) (16,18). The available >20 crystal

*To whom correspondence should be addressed. Tel: +1 614 292 1377; Email: foster.281@osu.edu

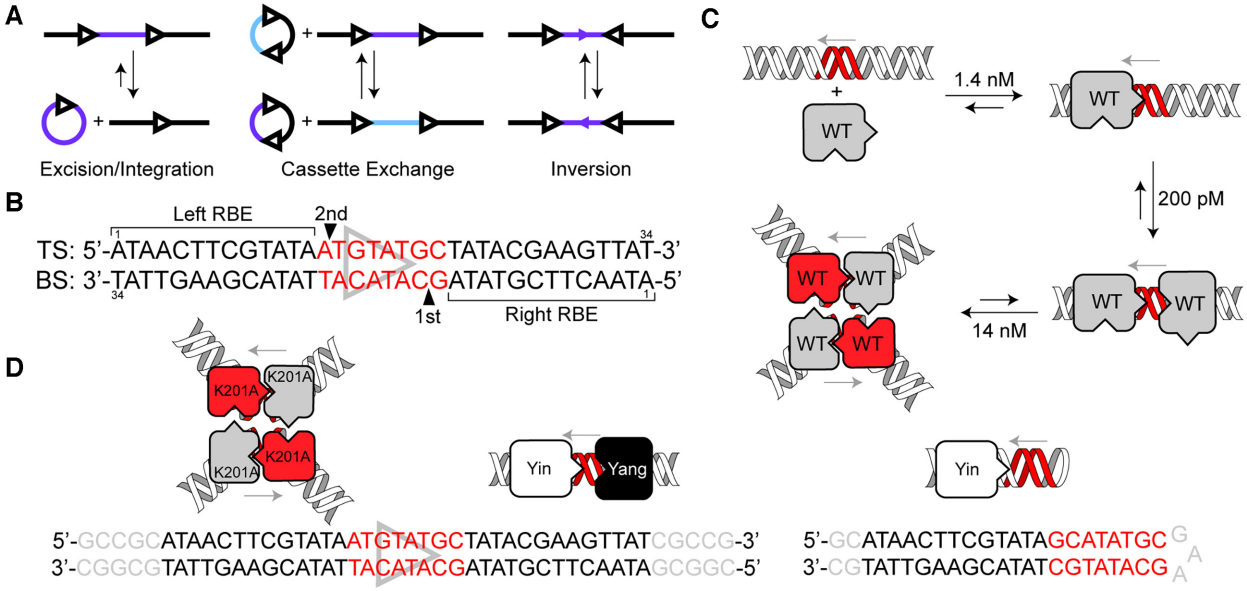

Figure 1. Cre function and assembly intermediates. (**A**) Cre carries out excision, integration, cassette exchange, and inversion reactions on DNA containing pairs of *loxP* sites (black triangles) and manipulated DNA regions (blue/purple). The direction of the arrows indicates the orientation of each asymmetric *loxP* site (13). (**B**) The *loxP* DNA sequence with conventional nomenclature. The spacer sequence (red) is asymmetric (gray triangle) and thus defines the overall orientation of the *loxP* site and determines which are the *left* and *right* RBEs (recombinase binding elements). The *top strand* (TS) runs 5′-3′ from the *left* RBE and has the spacer sequence 5′-ATGTATGC-3′, whereas the *bottom strand* (BS) runs 5′-3′ from the right RBE and has the complementary spacer sequence 5′-GCATACAT-3′. Cre recombines DNA with a bias for cleaving opposing sites on the BS, at the right RBE, with later cleavage of the two TS (filled triangles) (68,72). (**C**) Wild type (WT) Cre assembles in a reversible, stepwise, and cooperative manner to form a Cre-*loxP* monomer, then a Cre_2-*loxP* dimer, followed by antiparallel dimerization of dimers to form a $(Cre_2\text{-}loxP)_2$ homotetramer with each Cre protomer related by pseudo-four-fold symmetry. WT Cre makes and accepts protein–protein contacts (point, pocket) in a circular fashion in tetrameric complexes. Equilibrium constants for each step were established by analytical ultracentrifugation and mobility shift assays (16,18). (**D**) Tetrameric, dimeric and monomeric Cre complexes used in these studies. Tetrameric complexes were prepared using Cre^{K201A} in complex with a full *loxP* site; additional stabilizing GC 'clamps' are shown in gray. Dimeric complexes were assembled using Cre^{Yin} and Cre^{Yang} variants featuring complementary 'point' and 'pocket' mutations, respectively. The palindromic RBE sequences are black, the asymmetric spacer red, and the stabilizing GC clamps in gray. Monomeric complexes were obtained with Cre^{Yin} bound to a DNA hairpin with a single RBE and a symmetric *loxA* spacer (14,34).

structures of Cre or Cre-like recombinases in tetrameric complexes exhibit similar characteristics, as exemplified by the tetrameric, pre-cleavage complex between a cleavage-deficient variant, Cre^{K201A} and *loxP* (2HOI (16); Supplemental Figure S1): (i) The tetramer exhibits pseudo C2-symmetry, in which the *loxP* sites are antiparallel and the protomers bound to the right RBEs are *primed* for DNA cleavage with their tyrosine nucleophile on helix αM positioned within ∼3 Å of the scissile phosphate, while the protomers on the left RBEs are in an *inactive* conformation with their catalytic tyrosines ∼6 Å away from the scissile phosphate. In this bottom strand (BS) complex the *primed* protomers reside on the right RBEs, and the bottom strands are poised for cleavage. (ii) Each *loxP* DNA duplex is asymmetrically bent by ∼108°. (iii) DNA bending is coincident with extensive protein–protein interactions between protomers on the same duplex and across the synapse. These interactions include reciprocal cyclic *trans* docking of the C-terminal helix αN and a loop between β-strands 2 and 3 (the β2–3 loop) in a manner hypothesized to act as allosteric switches (14,19). This structural asymmetry in the C-terminus and inter-protomer interfaces observed in all tetrameric synaptic complexes of Cre suggest that protein flexibility is important for regulating protomer activity (20–26).

loxP site recognition

The structural basis for *loxP* sequence specificity by Cre has been difficult to decipher. Tetrameric crystal structures have revealed that relatively few of the bases in the RBEs or spacer are directly contacted by amino acid sidechains (5,14,27–30). Biochemical studies using variant Cre and *loxP* sequences have identified specificity determinants that are not well explained by the contacts observed in tetrameric structures. These observations led to the conclusion that shape complementarity and water-mediated contacts are at least as important as direct readout in sequence selectivity by Cre (4,14,19,27,31–33). Nevertheless, in the absence of high-resolution structures of assembly intermediates, it has remained unclear whether additional protein–DNA contacts might play important roles in initial steps of site recognition that are not retained in the assembled tetrameric complex. For instance, NMR studies of the catalytic domain of Cre showed that the C-terminal region of the protein containing helix αN docks in a *cis* autoinhibited conformation over the DNA binding site in the absence of DNA, and is displaced upon binding to *loxP* DNA, thereby extending into solution (34). The cryo-EM structures described below of Cre bound to DNAs containing one and two RBEs lend clarity to recognition and progressive assembly of synaptic complexes, and emphasize the roles of shape

recognition, DNA deformation (indirect readout (35–37)), and protein–protein interactions contributing to specificity.

Activation of Cre

Control over the DNA cleavage activity of the four assembled Cre protomers is critical to faithful DNA recombination because activation of adjacent protomers would result in double-strand breaks and failed recombination. The mechanism for this control is particularly intriguing due to the close topological and stereochemical similarity of YSSRs to type IB topoisomerases, whose tyrosine nucleophiles are active to cleave DNA as monomers and exhibit low sequence selectivity (38). Both classes of enzymes sport highly conserved active sites comprised of conserved arginine and lysine residues that coordinate the scissile base and phosphate in similar manners (38). Tetrameric crystal structures of synaptic complexes and Holliday junction intermediates of Cre (and other YSSRs) have invariantly shown pseudo-C2 symmetry with opposing pairs of protomers adopting *inactive* and *primed* conformations, as defined above. Comparison of the *primed* and *inactive* protomers in these structures has drawn attention to the β2–3 loop and C-terminal helices αM and αN (Supplemental Figure S1) (14,27). Solution NMR studies of the C-terminal domain of apo Cre showed flexibility in the β2–3, αJ-K and αM-N loops but not for helix αN (34). However, the αJ-K loop is involved in direct DNA recognition and does not differ between *primed* and *inactive* protomers in crystals, refocusing attention on the β2–3 loop and αM-αN.

New insights from Cryo-EM structures

In the work described below, cryo-EM structures of the Cre assembly intermediates provide important insights into the mechanism of Cre activation. We designed two mutants, Cre^{Yin} and Cre^{Yang}, that when added to a full *loxP* site, in a 1:1:1 fashion, are expected to assemble into a Cre_2-*loxP* dimer that maintains a native interface along the *lox* site but prevents interactions at the tetramer interface (Figure 1D, Supplemental Figure S1, Methods). Addition of Cre^{Yin} to a DNA hairpin containing a single RBE allowed us to isolate the Cre-*loxP* monomer complex. Lastly, we assembled a pre-cleavage Cre_4-$loxP_2$ tetramer using the cleavage deficient Cre^{K201A} variant (16). These new structures illuminate the critical roles of protein and DNA deformation in enabling site recognition, trans-docking of the αN helix, and formation of protein–protein contacts involving the key regulatory elements β2–3 loop, αM and αN (Supplemental Figure S2). These insights simultaneously help us understand site selectivity, controlled activation of alternating protomers, and the basis for the observed bias for strand cleavage order. Additionally, our work highlights the use of conventional defocus methods for high resolution structure determination of low molecular weight (<100 kDa) macromolecular complexes by cryo-EM.

MATERIALS AND METHODS

Computational modeling of Cre mutants

In tetrameric pre-synaptic assemblies with pseudo-C4 symmetry, each Cre protomer can be described as having a 'point' and a 'pocket' that affords specific protein–protein interactions at the interface along the DNA strand and across the synapse (Figure 1C) (14,16,39). Using the crystallographic coordinates of the tetrameric pre-cleavage complex of Cre^{K201A} bound to *loxP* (2HOI), we engineered the variant Cre^{Yin} by mutating the 'pocket' of the protein bound to the *right* RBE of the *loxP* site and deleted the 'point' of the protein bound to the *left* RBE of the loxP site to engineer Cre^{Yang} (Figure 1C and D) (16). The Proteins, Interfaces, Structures and Assemblies (PISA) tool from PDBe was used to identify buried residues and residues involved in hydrogen bonding and salt bridge interactions across the synapse at both of the possible interfaces (Supplemental Figure S1, chains A:H and B:G) (40,41). In the coordinates for 2HOI, residues D33 and R192 on the *primed* protomers (chains B and H), and residues R72 and R119 on the *inactive* protomers (chains A and G) had the most distinct and extensive hydrogen bonding and salt bridge networks: D33 forms salt bridges with R72 and R119, while R192 forms a salt bridge with E308. We then used the mCSM-PPI2 webserver to perform saturating mutagenesis on the A:H and B:G interfaces to predict which mutations of the four residues listed above would provide the most destabilizing effect (42). Informed by this analysis, we selected two mutations: D33A and A36V on helix αA to eliminate favorable electrostatic interactions with nearby residues, and R192A on the β1–2 loop to disfavor the coordination of αN in the binding pocket of Cre^{Yin}. Although varying A36 in helix αA was not predicted by mCSM-PPI2 to be detrimental, we also included in Cre^{Yin} the mutation of A36V, which was previously shown to be defective in synapsis (presumably a steric interference, but structurally unverified) (16). For Cre^{Yang}, we flipped the charges on R72E and R119D to create a largely negative charged interface to disfavor interactions with D33, while L115D was selected to generate similar unfavorable charge-charge interactions and eliminate favorable apolar contacts.

Site-directed mutagenesis, protein expression and purification

Cre mutants were generated from a pET-21a plasmid template (Novagen) containing WT Cre (provided by Gregory Van Duyne, University of Pennsylvania Philadelphia, PA), using the Q5 Site-Directed Mutagenesis Kit (NEB) (14,16). DNA primers (Supplemental Table S1) were obtained from Integrated DNA Technologies (Coralville, IA). Each mutant was transformed into *Escherichia coli* T7 Express LysY/Iq competent cells (NEB) using electroporation in a 1 mm electroporation cuvette. Transformations were plated on Luria-Bertani (LB) (RPI) media plates supplemented with 100 μg/ml carbenicillin (RPI) and 25 μg/ml chloramphenicol (Fisher Scientific) and incubated overnight (16 h) at 37°C. A single, fresh colony was used to inoculate a 50 ml starter culture of LB supplemented with 100 μg/ml carbenicillin and grown overnight for 16 h at 37°C in a shaking incubator at 220 rpm. 10 ml of starter culture was used to inoculate a 1 l culture of LB supplemented with 100 μg/ml of carbenicillin and left to grow in a shaking incubator at 220 rpm until an $O.D._{600}$ of 0.8 was reached (∼3 h). Cultures were then induced with 400 μl of 1M IPTG (RPI) and Cre mutants were expressed for 2 h at 37°C and 220 rpm.

After incubation, cells were pelleted at 4k rpm and 4°C and then frozen at −80°C until purified. [U-^{15}N]-Yang was expressed as stated above, where LB was replaced with M9 minimal media containing 1 g of ^{15}N-ammonium chloride (Cambridge Isotope Labs) as the sole nitrogen source. The expression time for minimal media cultures was lengthened to 4 h.

All proteins (WT, K201A, Yin and Yang) were purified as follows. Frozen pellets were resuspended on ice in lysis buffer, 40 mM Tris, 300 mM NaCl, 5 mM DTT, pH 7.0 (Fisher Scientific/Fisher Scientific/RPI, respectively). A complete protease inhibitor tablet (Roche) was added to the cell suspension. The resuspended cells were lysed using sonication and centrifuged at 23 000 × *g* for 45 min at 4°C. The clarified supernatant was then filtered through a 0.8/0.2 μm syringe filter (Pall) and diluted with Buffer A (40 mM Tris, 100 mM NaCl, pH 7.0), 1:1. The filtrate was loaded onto a SPFF column (GE/Cytiva) for cationic exchange purification and eluted using a gradient of 100% buffer A to 100% buffer B (40 mM Tris, 1M NaCl, pH 7.0). Fractions containing the Cre protein as determined by SDS-PAGE were pooled and diluted with buffer A. The pooled and diluted Cre containing fractions were loaded onto a 5 ml Heparin HP column (GE/Cytiva) and eluted using a gradient of 100% buffer A to 100% buffer B over 12 column volumes. Fractions containing Cre protein were combined and concentrated to 1.5 ml using an Amicon Ultra-10, 10 kDa MWCO spin filter. Concentrated samples were loaded onto a GE Hi-Load Superdex 75 16/600 prep column and eluted using Cryo-EM buffer of 20 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.0 (RPI/Fisher Scientific/Alfa Aesar, respectively). Fractions containing pure protein were pooled and used in further experiments.

Purification of DNA constructs

Single-stranded DNA oligonucleotides were purchased from IDT and shipped dry (Supplemental Table S1). Upon receipt, oligonucleotides were dissolved in water to a concentration of 100 μM. Complementary pairs of dissolved oligonucleotides were combined in equimolar amounts. Concentrated Tris base (pH 7.0) and concentrated NaCl were added to each mixture to obtain a final concentration of 10 mM Tris and 100 mM NaCl. Mixtures were then placed into a 95°C water bath and left to cool overnight. DNA duplex mixtures were loaded onto a 5 ml QHP ion exchange column (GE/Cytiva) with buffer A (10 mM Tris (pH 7.0) and 25 mM NaCl) and the duplexes were eluted with a gradient 0–100% buffer B (10 mM Tris (pH 7.0) and 1 M NaCl) over 10 column volumes. DNA duplexes eluted near 600 mM NaCl. Samples were EtOH-precipitated for further use (43).

Purification of complexes (SEC)

Protein and DNA complexes were formed by addition of 1.2:1 protein to Cre binding sites and were concentrated using an Amicon Ultra-0.5, 3 kDa MWCO spin concentrator yielding 250 μl samples of Cre^{Yin}-*loxA* hairpin, Cre^{Yin}-Cre^{Yang}-*loxP*, and Cre^{K201A}-*loxP* at 50 mg/ml, 3.5 mg/ml and 10 mg/ml, respectively. Yin-*loxA* was purified over a GE Superdex 75 10/300 analytical column, and Cre^{Yin}-Cre^{Yang}-*loxP* and Cre^{K201A}-*loxP* were purified over a GE Superdex 200 Increase 10/300 analytical column, both equilibrated with cryo-EM buffer (20 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, pH 7.0).

Recombination reactions

Recombination reactions were performed in 20 μl with 150 ng of a plasmid containing two *loxP* sites (Addgene Plasmid #26852) and 500 nM WT or variant Cre in recombination buffer (50 mM Tris–Cl, 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.0) at 37°C for 1 hr (16,44). Reactions were subjected to a 70°C water bath for 5 min to deactivate Cre. The samples were then combined with 6× gel loading dye (NEB) and electrophoresed on a 1%, 10 cm TBE/agarose gel (Fisher Scientific) that was impregnated with SYBR safe (Invitrogen) at 120 V for 55 min.

Electrophoretic mobility shift assay

Gel polymerization and gel running buffers were of the same composition (50 mM HEPES and 20 mM Tris, pH 7.0). Native 6% PAGE gels were prepared by first making a gel solution containing 6% acrylamide (29:1) (Fisher Scientific) (16). This solution was then degassed under vacuum while stirring. 12 ml of degassed solution (enough for two gels) was polymerized by mixing with 200 μl of 10% (w/v) ammonium persulfate (Biorad) dissolved in water and 10 μl of TEMED (Acros).

A set of 10 μl reactions containing varying amounts of Cre variants (1–1 000 nM) and 10 nM 5′-Cy5-labeled (bottom strand) *loxP* DNA (IDT), were prepared in binding buffer (20 mM HEPES, 150 mM NaCl, 5 mM MgCl2, 2 mM DTT, pH 7.0) along with 50 μg/ml salmon sperm DNA (Invitrogen) and 100 μg/ml BSA (Thermo Scientific) (16,44). Binding reactions were equilibrated at room temperature for 30 min. 1 μl of 10× EMSA gel loading dye (60% glycerol, 40% binding buffer, 0.1% bromophenol blue) was added to each reaction and then loaded onto a prerun (120 V, 30 min) 6% native PAGE gel using gel running buffer. Samples were electrophoresed at 80 V for 75 min and imaged on an Amersham Typhoon Gel imager.

Native mass spectrometry

Native MS experiments were performed on a Q Exactive Extended Mass Range (EMR) mass spectrometer (Thermo Fisher Scientific) modified to allow for surface-induced dissociation (SID) (45). Stock solutions of proteins and DNA were individually dialyzed into 200 mM ammonium acetate (pH unadjusted) using a Pierce 96-well microdialysis plate, 3.5K MWCO (Themo Fisher Scientific). The concentrations of the dialyzed stock samples were determined using a NanoDrop 2000/2000c spectrophotometer (Thermo Fisher) and their calculated extinction coefficients; extinction coefficients for the proteins were calculated using the ExPASy ProtParam tool and the DNA extinction coefficient was calculated from the sequence using the built-in NanoDrop software (46). For preparation of protein-DNA complexes, stock samples were mixed to their final experimental concentration. Each sample was then injected

into an in-house pulled borosilicate filament capillary (OD 1.0 mm, ID 0.78 mm) and subsequently ionized by nano-electrospray ionization in positive mode. Ion optics (including SID device) were tuned to allow for transfer of complex ions to the mass analyzer without causing dissociation. Varying higher energy collisional dissociation (HCD) voltage was tested and then set to 60 V for optimal transmission and de-adducting of the complex ions. Other parameters included: spray voltage adjusted to 0.6 kV and then held constant, capillary temperature 250°C, trap gas pressure set shifted the UHV sensor up to ∼8E–10 mbar, ion inject time 200 ms, averaged micro scans 5, and resolution 8750 as defined at 200 m/z.

All data were visually examined and then deconvolved using UniDec V4.4 (47). UniDec parameters included charge range of +5 to +25, mass range of 10 000–1 100 000 Da, sample mass every 1 Da, peak full-width at half-maximum (FWHM) 0.85 Th, Gaussian peak shape function, beta 50 (artifact suppression), charge smooth width 1, point smooth width 10, native charge offset range -5 to +5, and peak detection range of 40 Da. These settings worked well, but manual mode was used to specify charge states for known m/z regions. This was beneficial for artifact suppression, particularly for optimizing parameters for a range of species present across the m/z range. The resulting deconvolutions were plotted as the sum normalized relative signal intensities in the form of zero-charge mass spectra.

Cryo-EM grid preparation and data collection

All samples were SEC purified on a GE Superdex 200 Increase 10/300 analytical column before grid preparation. Before plunge freezing of grids, octyl-*n*-glucoside was added to a concentration of 16.8 mM to the Cre^{Yin}-Cre^{Yang}-*loxP* and Cre^{K201A}-*loxP* samples, while CHAPSO was added to a concentration of 4 mM to the Cre^{Yin}-*loxA* hairpin sample. 3 μl of Cre^{Yin}-*loxA* hairpin (7.5 mg/ml), Cre^{Yin}- Cre^{Yang}-*loxP* (5 mg/ml) and Cre^{K201A}-*loxP* (4 mg/ml) were applied to glow discharged (20 mA, 1 min, Pelco easiGlow) Au Quantifoil R1.2/1.3 300 mesh grids (Ted Pella) and incubated for 60 s horizontally. Grids were loaded into Vitrobot Mark IV (Thermo Fisher Scientific) operating at 4°C and 100% humidity and blotted for 4 s with a blot force of 1 before being plunge frozen into liquid ethane.

Data were acquired on a Titan Krios G3i (Thermo Fisher Scientific) operated at 300 keV with a Cs corrector in nanoprobe EFTEM mode (50 μm C2 aperture, 100 μm objective aperture). A Gatan K3 camera, operating in correlated double sampling mode, and a BioContinuum energy filter (15 eV slit width, zero-energy loss) were used to record 45 frame movies with a total dose of 60 $e^-/Å^2$. 3 538 movies (two shots per hole) were collected for the Cre^{Yin}-*loxA* hairpin sample using a defocus range of –1.0 μm to –2.2 μm at a nominal magnification of 105 000× (0.702 Å physical pixel size). For the Cre^{Yin}–Cre^{Yang}-*loxP* sample, 1 527 movies were collected with a tilt of 28° using a defocus range of –1.5 μm to –3.0 μm, and 2 779 movies were collected with no tilt using a defocus range of –1.0 μm to –2.2 μm at a nominal magnification of 81 000× (0.899 Å physical pixel size). For the Cre^{K201A}-*loxP* sample, 906 movies were collected with a tilt of 28° using a defocus range of –1.5 μm to –3.0 μm, and 2 756 movies were collected with no tilt using a defocus range of –1.0 μm to –2.2 μm at a nominal magnification of 81 000× (0.899 Å physical pixel size). An additional 3 330 movies were collected of the Cre^{Yin}-*loxA* complex to obtain more particles for resolution improvement. These were collected with the same settings listed above.

Cryo-EM data processing

All data processing was carried out using CryoSPARC v3.0 (48).

Cre^{Yin}-loxA hairpin. 3 538 movies were patch motion corrected and their CTF parameters were estimated using patch CTF estimation. Micrographs were curated by CTF fits <4 Å and total motion <200 pixels, resulting in a stack of 3 211 micrographs. 300 micrographs were blob picked and 2D classified to yield a set of 47 291 particles that were used to train a TOPAZ model. This model was then used to pick all 3 211 micrographs, yielding 477 224 particles (49). Particles were extracted using a 256 × 256 box subsequently down-sampled to 128 × 128 (1.404 Å/pixel). Initial 2D classification removing false picks or poorly defined classes left 415 307 particles. These particles were then subjected to *ab initio* reconstruction using two classes, of which 284 516 particles were left in the class resembling a monomer bound to a DNA hairpin. These particles were then subjected to three rounds of 3D classification that resulted in a stack of 90 940 particles that upon non-uniform refinement resulted in a map with a gold-standard FSC global resolution of 4.26 Å. Re-centering and re-extraction with a box size of 350 × 350 pixels and down sampled to 256 × 256 pixels (0.959 Å/px) followed by non-uniform refinement resulted in a map with a 4.10 Å resolution using 87 515 particles (50). Additionally, 3 330 movies were collected and processed in the same manner yielding 59 200 more particles. In total, 146 715 particles were used for the reconstruction of the Yin-*loxA* hairpin complex, resulting in a final global resolution of 3.9 Å.

*Cre^{Yin}- Cre^{Yang}-*loxP. 1 527 28°-tilt and 2 779 0°-tilt movies were independently patch motion-corrected and their CTF parameters were estimated using patch CTF estimation. Micrographs were curated by CTF fits <4 Å and total motion less than 200 pixels, resulting in a stack of 1 198 and 2 650 micrographs, respectively. Blob picking of 300 micrographs resulted in 62 749 particles and these particles were used to generate templates for template picking. Template picking all micrographs yielded roughly 700k particles. Concurrently, a TOPAZ model was trained on the same 62k, blob-picked particles and the model was used to pick all micrographs resulting in 1.9M particles (49). These particle stacks were combined, and duplicates were removed to yield a stack of 2.2M particles. Particles were extracted using a 256 × 256 box subsequently down-sampled to 128 × 128 (1.798 Å/pixel). 1.2M particles were then subjected to *ab initio* reconstruction using three classes. Multiple rounds of 3D classification and non-uniform refinement resulted in a stack of 256 734 particles. Particle subtraction of partial density attributed to a fractionally bound third

protein (Supplemental Figure S6) followed by local, non-uniform refinement yielded a map at 4.48 Å (50).

Cre^{K201A}-loxP. 906 28°-tilt and 2 756 0°-tilt movies were independently patch motion corrected and their CTF parameters were estimated using patch CTF estimation. 28°-tilt micrographs were curated by CTF fits <4 Å and total motion <200 pixels, yielding 806 micrographs. Blob picking of 150 micrographs resulted in 22 250 particles and these particles were used to train a TOPAZ model. The TOPAZ model was used to pick all micrographs, resulting in 1.2M picked particles (49). Blob picking of the individual datasets yielded roughly 588k particles. TOPAZ and blob-picked particle sets were combined and duplicates were removed to yield 737 980 particles. 2D classification to remove poorly aligning particles resulted in 608k particles that were further processed using 3D methods. Particles were extracted using a 256 × 256 box (0.899 Å/pixel). 608K particles were then subjected to *ab initio* reconstruction using a single class. A single round of 3D classification with two classes resulted in 315 574 particles that refined to 3.23 Å. These 315K particles were then subjected to 3D Variability Analysis with two modes (51).

Model building

Cre^{Yin}*-loxA hairpin.* Chain A of PDB 2HOI was used as the starting model for the monomer with mutations introduced using PyMol. The hairpin DNA was generated using 3DNA (http://web.x3dna.org/) by compositing a B-form *loxA* RBE and spacer sequence with a GAA hairpin from PDB 1JVE (52,53). The *loxA* hairpin and protein model were rigidly docked into the map using ChimeraX (54,55). The complex was subjected to all-atom refinement in Coot using user-defined restraints (all atoms within 5 Å of one another) (56). Iterative rounds of flexible fitting using the program ISOLDE, implemented in ChimeraX, and real space refinement in Phenix were used to obtain the final monomer model (57,58).

Cre^{Yin}- *Cre*^{Yang}-loxP. The protein chain from the Yin-*loxA* model was used as the starting model for each protein in the dimer with mutations introduced using PyMol. 3DNA was used to generate a B-form *loxP* site that matched our DNA construct (53). Orientation of the asymmetric *loxP* spacer, and densities of the proteins were assigned based on similarity of DNA helical parameters to those of BS complexes in tetrameric complexes from crystallography (2HOI), and Cryo-EM (below). Each DNA and protein model were rigidly docked into the map using ChimeraX (54,55). The complex was then subjected to all-atom refinement in Coot using user-defined restraints (all atoms within 5 Å of one another) (56). Iterative rounds of flexible fitting using the program ISOLDE, implemented in ChimeraX, and real space refinement in Phenix were used to obtain the final dimer model (54,55,57,58).

Cre^{K201A}-loxP. PDB 2HOI was used as the starting model and was rigidly docked into the map using ChimeraX (16,54,55). The DNA ends of each strand were extended using PyMol to match the DNA used in the study. Iterative rounds of flexible fitting using the program ISOLDE, implemented in ChimeraX and real space refinement in Phenix were used to obtain the final tetramer model (57,58).

3DVA was performed in CryoSPARC using two modes, a filter resolution of 4.0 Å and maps displayed in 20 frames (51). ISOLDE was used to fit the consensus model into the first and the last frame output from the 3DVA job (58).

RESULTS

Characterization of isolated Cre assembly intermediates

Wild-type Cre, Cre^{K201A}, Cre^{Yin} and Cre^{Yang} proteins were purified to greater than 95% homogeneity as indicated by band intensities from Coomassie stained SDS-PAGE gels (Figure 2A). WT Cre, Cre^{K201A}, Cre^{Yin} (D33A/A36V/R192A) and Cre^{Yang} (R72E/L115D/R119D) were assayed against a linearized plasmid containing two *loxP* sites in direct repeat (Figure 2B). Cre excises the intervening sequence to produce a circular 1 200 bp product and a linear 4 000 bp product, which can be resolved on a 1% agarose gel (59,60). Cre recombines the substrate as expected (to ∼65% completion due to the competing reverse reaction), whereas cleavage-deficient Cre^{K201A} does not (16). In addition, Cre^{Yin}, Cre^{Yang} and the Cre^{Yin}/Cre^{Yang} combination did not generate recombinant products.

We used EMSA to assay the Cre constructs for their ability to assemble with *loxP* DNA at concentrations approximating the equilibrium dissociation constant for synapsis of WT Cre (16). In this assay, each Cre mutant was combined with a 5′-Cy5-labeled *loxP* DNA and electrophoresed on a 10% native page gel. For Cre^{K201A}, which has been shown to be capable of synapsis but not cleavage (16), we observed bands corresponding to free DNA, singly bound *loxP* DNA (monomers), fully bound *loxP* sites (dimers), and two *loxP* DNA sites coordinated by four Cre proteins (tetramers) (Figure 2C). Using 10 nM *loxP* DNA and various amounts of Cre^{K201A} we observe about 50% dimers and 50% tetramers, consistent with the published synaptic K_D of ∼8 nM; moreover, the monomer to dimer transition is highly cooperative as indicated by a very low population of monomers (16). Next, we tested a Cre deletion mutant (ΔαN, residues 1–330) that has been shown to be synapsis-defective (16). Upon addition of Cre$^{\Delta\alpha N}$ to *loxP* DNA, we can observe that most, if not all, cooperativity is lost, evident from the prominent monomer band, and the formation of tetramers is strongly disfavored (Figure 2C). Although Cre^{Yin} and Cre^{Yang} were designed to not interact favorably, they nevertheless are observed to form homodimers on DNA in a cooperative manner, with a low population of monomeric species; no tetrameric species were observed under these conditions (Figure 2C). When Cre^{Yin}, Cre^{Yang}, and *loxP* DNA were combined, a mixture of dimer species was observed, with more than one band corresponding to the dimer species on DNA. Because the electrophoretic mobilities of the Cre^{Yin}/Cre^{Yang} homo- and heterodimers on DNA were almost indistinguishable, this warranted investigation with higher-resolution analytical tools.

To quantify and characterize the composition of the Cre$^{Yin/Yang}$ heterodimers on DNA, we turned to native mass

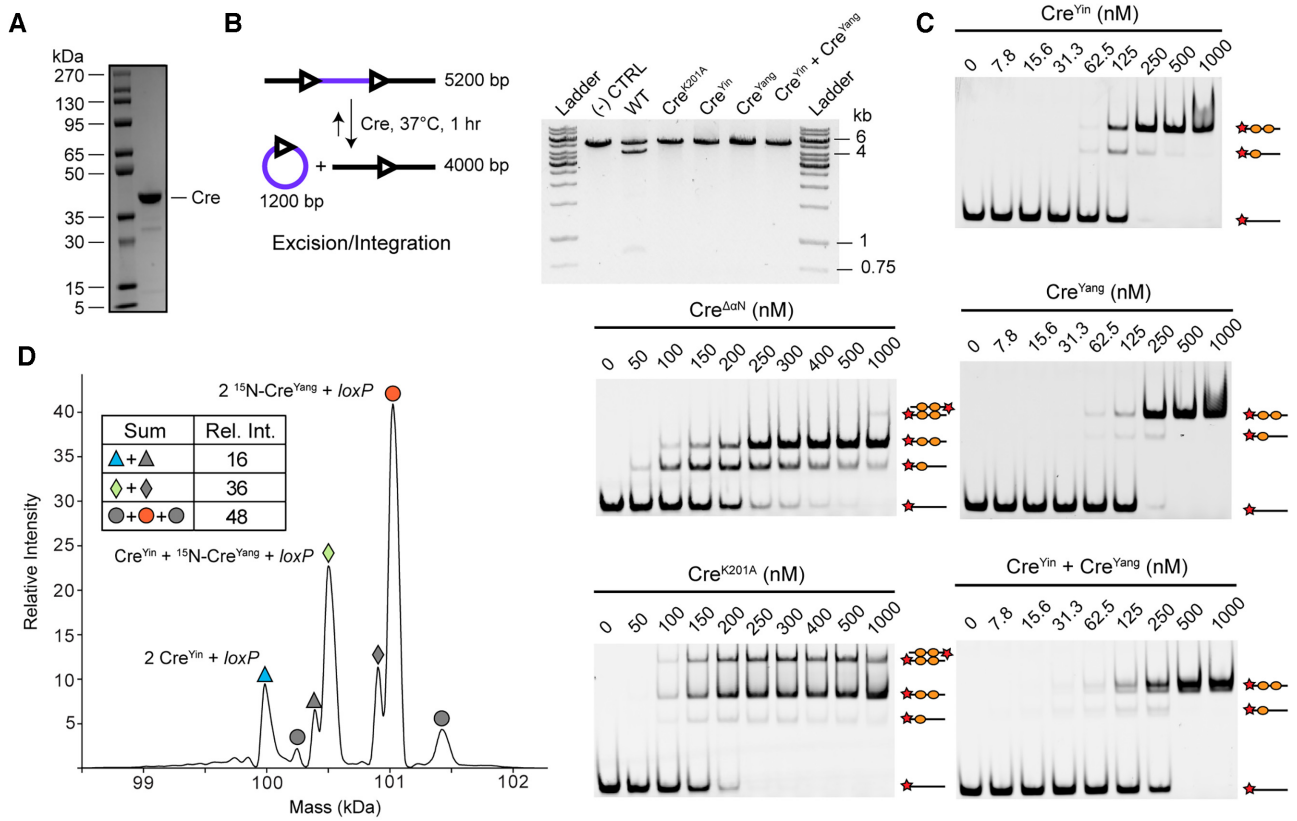

Figure 2. Assembly and recombination of *loxP* DNA by WT Cre and variants Cre^{K201A}, Cre^{Yin}, Cre^{Yang} and $Cre^{\Delta\alpha N}$. (**A**) Representative Coomassie-stained 8–16% SDS-PAGE of SEC purified WT Cre (∼95% purity). (**B**) *In vitro* recombination assay in which 150 ng of a linearized 5.2 kb plasmid with two *loxP* sites in direct repeat is treated with 500 nM Cre at 37°C for 1 h and visualized on a 1% agarose gel. (**C**) Electrophoretic mobility shift assays (EMSA) on 10% polyacrylamide gels, with 10 nM fluorescently labeled (5′-Cy5 [bottom stand]) *loxP* DNA and varying concentrations of $Cre^{\Delta\alpha N}$, Cre^{K201A}, Cre^{Yin}, Cre^{Yang} or Cre^{Yin} + Cre^{Yang}. Unbound DNA and protein-DNA complexes with 1:1, 2:1 and 2:2 stoichiometry are indicated schematically. (**D**) The zero-charge deconvolved ESI mass spectrum of the 2:1 protein:*loxP* complexes formed by Cre^{Yin}, [U-^{15}N]-Cre^{Yang} and *loxP* DNA. Masses correspond to two Cre^{Yin} + *loxP* (blue triangle), Cre^{Yin} + [U-^{15}N]-Cre^{Yang} + *loxP* (green diamond), and two ^{15}N-Cre^{Yang} + *loxP* (orange circle). Gray shapes are assigned as proteoforms of these complexes. The summed relative intensities of the three complexes are indicated in the inset table.

spectrometry (MS). MS uses soft ionization methods and non-denaturing solution conditions to record gas-phase mass spectra that preserve non-covalent interactions reflective of those in solution (61). We first screened each individual component's mass and purity (Supplemental Figure S3, Supplemental Table S2), and subsequently analyzed mixtures of Cre^{Yin}, Cre^{Yang} and *loxP* DNA. To increase the mass separation between Cre^{Yin} (38 307 Da) and Cre^{Yang} (38 342 kDa), we used uniformly ^{15}N-labeled Cre^{Yang} (38 844 Da, if fully labeled); mass analysis showed >95% labeling (Supplemental Table S1). The experimentally determined masses of Cre^{Yin} and Cre^{Yang} matched the theoretical masses, but some minor species were present for ^{15}N-Cre^{Yang} and Cre^{Yin}. These are likely truncations, modifications, or specific salt adducts (Supplemental Figure S3, Supplemental Table S2). The ^{15}N-Cre^{Yang} spectrum contains two charge state distributions with the higher charge state centered around +17 being indicative of unfolded or disordered protein (62). The determined mass for *loxP*-38 (*loxP* site with a GC clamp on each end) matched the theoretical mass. (Supplemental Figure S3, Supplemental Table S2). In all spectra, many of the peaks had additional mass spacing of ∼22 Da attributable to sodium adducts.

Native mass spectra were recorded of samples containing 0.10 μM *loxP* and either (i) 0.50 μM Cre^{Yin}, (ii) 0.50 μM ^{15}N-Cre^{Yang} or (iii) 0.25 μM Cre^{Yin} + 0.25 μM ^{15}N-Cre^{Yang} in a solution of 200 mM ammonium acetate. For each sample, we observed signals corresponding to the molecular weights of complexes with protein:DNA stoichiometries of 1:1 (monomeric complexes) and 2:1 (dimeric complexes) (Supplemental Figure S4). When both Cre^{Yin} and ^{15}N-Cre^{Yang} were mixed with *loxP*, we observed abundant signals corresponding to both homo- and heterodimeric complexes with *loxP* DNA (Cre^{Yin}-Cre^{Yin}, Cre^{Yin}-Cre^{Yang} and Cre^{Yang}-Cre^{Yang}) (Figure 2D). Absent interactions between protomers, we would expect statistical population ratios of 1:2:1 for these complexes, and an additional bias in favor of the heterocomplex due to their complementary protein–protein interfaces. However, the number of gas-phase ions corresponding to Cre^{Yin} + ^{15}N-Cre^{Yang} were only ∼36% of the total ion count; this may reflect imperfect proportionality between ion counts and solution populations (63). Nevertheless, this compositional heterogeneity decreased the number of usable particles in subsequent cryo-EM reconstructions and may also have resulted in averaging of homo- and heteromeric species (see below).

Cryo-EM of Cre assembly intermediates

Three Cre-DNA complexes were examined by Cryo-EM: Cre^{Yin}-*loxA* hairpin (monomer), Cre^{Yin}-Cre^{Yang}-*loxP* (dimer) and Cre^{K201A}-*loxP* (tetramer) were assembled, purified using size exclusion chromatography, vitrified, and imaged using a Titan Krios G3i microscope equipped with a Gatan K3 detector and energy filter. Images were subjected to single particle analysis via 2D classification of motion corrected and CTF estimated particles, followed by 3D classification and non-uniform refinement in cryoSPARC v3.0 (48,50). Additionally, 3D Variability Analysis was performed on the tetramer dataset (51). These resulted in 3D density maps with global resolutions (gold standard FSC cutoff of 0.143) of 3.91 Å for the Cre^{Yin}-loxA monomeric complex (Figure 3A, Supplemental Figure S5, Supplemental Table S3), 4.48 Å for the Cre^{Yin}-Cre^{Yang}-*loxP* dimeric complex (Figure 3B, Supplemental Figure S6, Supplemental Table S3), and 3.23 Å for the K201A-*loxP* tetrameric complex (Figure 3C, Supplemental Figure S7, Supplemental Table S3). Local resolutions varied from 3.3 to 7.1 Å, from 4.1 to 7.0 Å, and from 2.9 to 5.4 Å, respectively. All single particle reconstructions were performed without the use of symmetry. Atomic models were built in an iterative fashion using Coot, ISOLDE and Phenix real space refinement (56–58). Protein residues 1–19 in all complexes had no resolvable density and were omitted from modeling. Data collection, data processing, and model building statistics are summarized in Supplemental Table S1.

Cre^{Yin}-*loxA* half-site map and model

The density map for the Cre^{Yin}-*loxA* complex contains both domains of the protein, their interdomain linker, and the fully resolved *loxA* hairpin (Figure 3A). Helices αB, αD and αJ, participants in all the major groove protein–DNA interactions (Supplemental Figure S9), and catalytic core (Supplemental Figure S9) were of the highest resolution (between 3.3 and 4.0 Å), allowing for accurate sidechain placement into the map (64). Density is present for the β2 and β3 strands that form the anchor for the β2–3 loop, but no robust density was observed for the loop itself (residues 199–207, Figure 4A and B), which includes the active site residue K201. Likewise, clear density was observed for αM and the catalytic tyrosine, placing the Oη of Y324 ∼6 Å from the scissile phosphate (Figure 4F), but no density existed to model αN. Thus residues 199–207 and 330–343 were excluded from the model. The *loxA* hairpin phosphate backbone, major and minor grooves were well resolved, and in the higher resolution regions of the DNA individual ribosyl groups can be visualized. To better visualize the global DNA bend, the *loxA* hairpin model was extended with B-form DNA to contain a second RBE (Figure 5A, monomer, white RBE) (53,65). The *loxA* hairpin DNA exhibits an overall bend of 18° with a maximal bend located at the RBE–spacer junction (Figure 5A and C) (66).

Cre^{Yin}-Cre^{Yang}-*loxP* map and model

Single particle analysis of the Cre^{Yin}-Cre^{Yang}-*loxP* assembly was complicated by the presence of homo- and heterodimers (*loxP* bound by Cre^{Yin}-Cre^{Yin}, Cre^{Yang}-Cre^{Yang} and Cre^{Yin}-Cre^{Yang}), as evidenced by the EMSA and native MS experiments described above. During 3D classification, three populations of dimers emerged. The resolution of the three maps was insufficient to assign the identity of each bound protomer as Cre^{Yin} or Cre^{Yang}. However, in two of the three populations, the density for the N-terminal domain of one protomer was diffuse and displaced from its expected binding site in the major groove, and in the third population the N- and C-terminal domains of both protomers were positioned in the major groove of the *loxP* DNA (Supplemental Figure S6). We assigned this third class as being Cre^{Yin}-Cre^{Yang}-*loxP* heterodimers, and the two poorly-bound classes as homodimers on the expectation that engineered electrostatic repulsion between two Cre^{Yang} protomers, or steric repulsion and loss of salt bridge between two Cre^{Yin} protomers, would disfavor tandem docking of both N-terminal domains in these species. This assignment is also supported by the observation that the relative populations for each dimer class closely matched that of native MS experiments (Figure 2D).

A particle subtraction routine was used to further refine the dimer map. After selecting for fully assembled heterodimers on DNA, it was apparent that the resulting reconstructed particle set had a third protein loosely bound (Supplemental Figure S6). This third protein was faintly visible in 2D classes and in 3D reconstructions as a smearing of density that originated from the end of helix αM of the density assigned to the Cre^{Yang} protomer. Complexes with a third Cre protein bound to Cre_2-*loxP* were not detected by EMSA or native mass spectra; therefore, we used a particle subtraction routine to remove density of the third loosely associated protein; nevertheless, it seems likely that this type of interaction could be important in the context of synapsis of two Cre_2-*loxP* dimers.

In the resulting map, we could identify two Cre protomers (Figure 3B, blue, orange) bound to a full *loxP* DNA site (Figure 3B, magenta). The *loxP* DNA helix was well defined in the map with an asymmetric global bend of ∼48° (Figure 5A and C); this bend is consistent with that measured by phase-sensitive gel mobility shift (67). However, the resolution of the map was insufficient to unambiguously assign the orientation of the asymmetric spacer of the *loxP* DNA; therefore, it is possible that the maps arise from a mixture of DNA orientations. Nevertheless, we modeled the *loxP* DNA with a BS orientation such that the helical deformation parameters (Supplemental Figure S8) of the dimer agree with structural and biochemical data showing the preference for BS complexes (68). Due to the limited resolution of this map, protein modeling was guided by secondary structural elements in the monomer (above) and tetramer (below). Density is clearly visible for the αN helix of the Cre^{Yin} protomer docked into the αN binding pocket of Cre^{Yang}. Compared to crystal structures and the tetramer (below), the αM-αN linker adopts an extended conformation, thereby comprising the only extensive interprotomer interactions in the dimer complex (Supplemental Table S4). The αN helix of Cre^{Yang} was likely removed via particle subtraction due to its association with a third protein, and therefore, residues 328–343 were not modeled. The

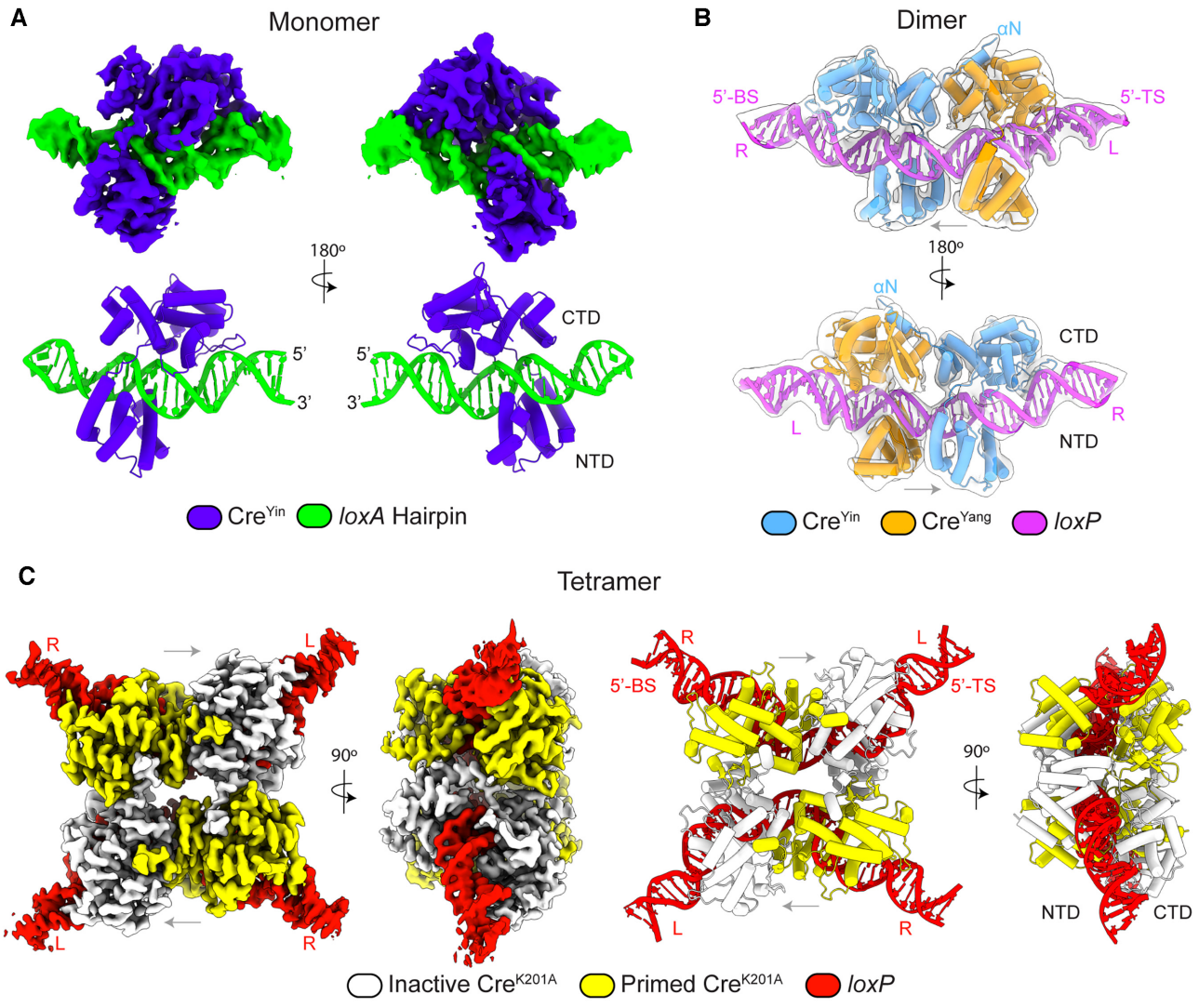

Figure 3. Cryo-EM maps and refined models of Cre-*loxP* assembly intermediates. (**A**) Cre^{Yin}-*loxA* half-site DNA complex (threshold of 0.55); hairpin DNA is green and Cre^{Yin} is blue. (**B**) Cre^{Yin}-Cre^{Yang}-*loxP* hetero-dimer (threshold of 0.5); with Cre^{Yin} in blue, Cre^{Yang} in orange, and the *loxP* DNA in magenta. Cryo-EM map shown in transparent white. The dimer is bent in a direction that favors bottom strand (BS) cleavage, on the right RBE.'R' and 'L' indicate right and left RBEs, respectively. (**C**) BS tetrameric K201A-*loxP* synaptic complex (threshold of 0.4); with *loxP* DNA in red, protomers in *primed* and *inactive* conformations colored yellow and white, respectively. Gray arrows indicate spacer orientation. 'R' and 'L' indicate right and left RBEs, respectively.

beta sheet comprised of strands β1, β2, and β3 of both protomers could be modeled as there was clear density present, but both Cre^{Yin} and Cre^{Yang} lacked robust density for the β2–3 loops, so residues 199–207 were also omitted from modeling (Figure 4A and C).

Tetrameric Cre^{K201A}-*loxP* map and model

Local resolutions ranging from 2.9 to 3.9 Å for much of the tetrameric complex allowed for accurate modeling of each protein and the *loxP* DNA (Supplemental Figure S7). Clearly resolved backbone and sidechain density permitted modeling of residues 20–341 for all four protomers. Sparse density was present for a few residues preceding residue 20, but lack of clear connectivity prevented us from modeling into it (not shown). The *loxP* DNA constitutes the highest resolution portions of the Cre^{K201A}-*loxP* map. For much of the DNA, the density allowed for discrimination of purine and pyrimidine bases, and accurate placement of bases was enabled by clear density for the phosphate backbone, ribosyl groups and the bases themselves. Each *loxP* DNA is highly bent to about 83 degrees (Figure 5A, tetramer), most prominently at the right side of the spacer, which features six consecutive AT or TA base pairs (Figure 5C). Despite this, it is possible that the tetrameric reconstruction arose from averaging of synapses poised for top strand and bottom strand cleavage (Figure 1); if BS and TS cleavage complexes are overall isomorphous, particle classification with the resolution of acquired data would not be sensitive to nucleotide identity within the spacer sequence (69,70). Nevertheless, the tetramer is highly isomorphous to that of the crystallized pre-synaptic complex of Cre^{K201A} bound to *loxP*, with

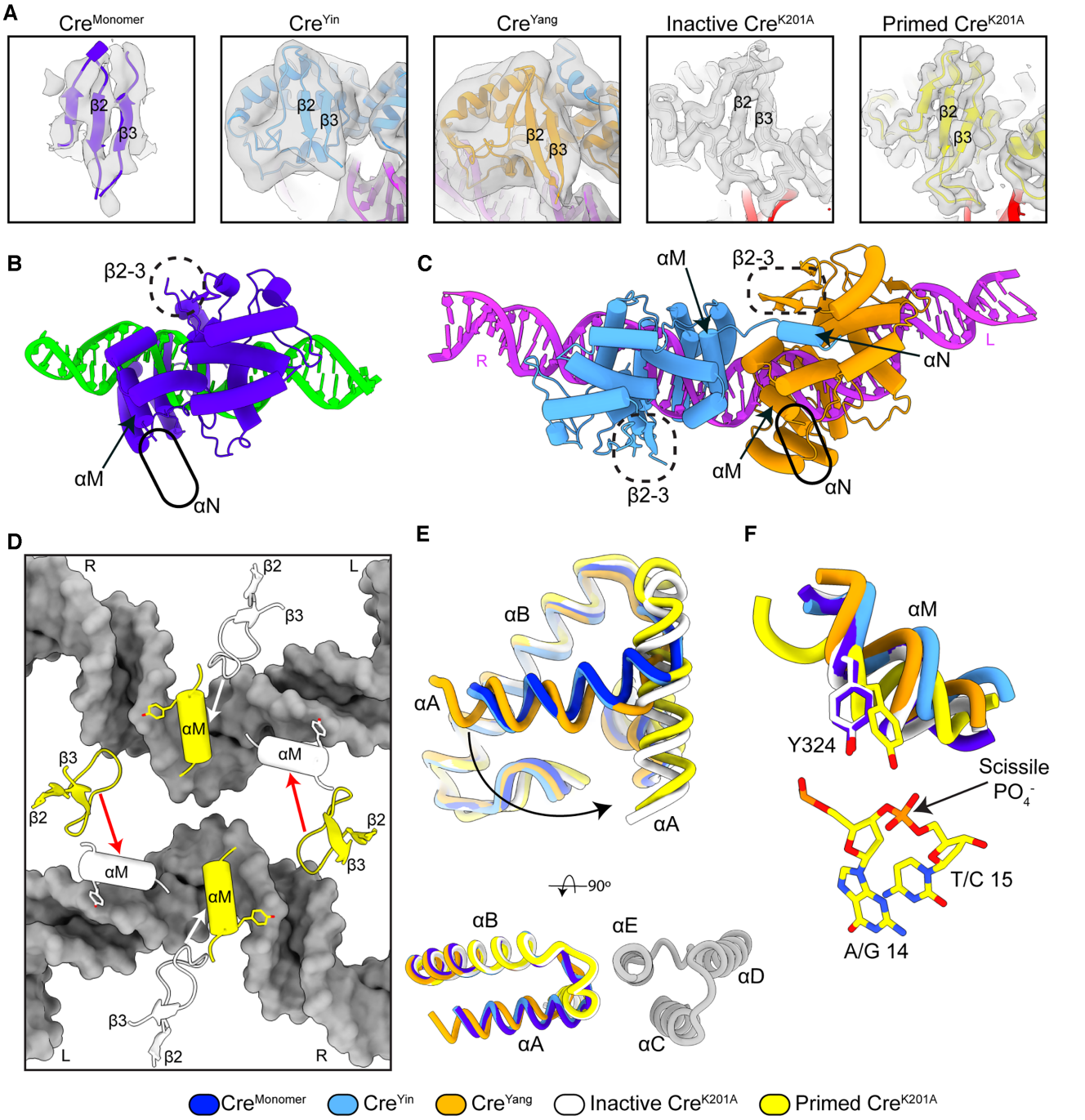

Figure 4. Progressive structural changes that accompany assembly and activation of Cre-*loxP* complexes. (**A**) EM density (gray) and model of the β2–3 strands and loop for each protomer; density for the loop is only observed in the tetrameric complex. (**B**) Monomer and (**C**) dimer atomic models with solid and dashed ellipses highlighting missing αN and β2–3 loop density, respectively. 'R' and 'L' indicate right and left RBEs, respectively. (**D**) Intra- and inter-molecular β2–3 contacts help position αM of primed protomers (yellow) in the tetrameric structure. Close contacts (white arrows) result in activation along the duplex; loose contacts (red arrows) are insufficient to activate protomers across the synapse. *loxP* DNA is gray. 'R' and 'L' indicate right and left RBEs, respectively. (**E**) Selected region of superposition from Figure 5 showing the difference in position of αA between the monomer, dimer, and tetramer protomers. A second protomer from the tetramer is shown in gray where αA docks against αC and αE of the neighboring protomer. (**F**) Selected region of superposition from Figure 5 showing αM and Y324 position in contrast to the scissile phosphate for the monomer, dimer and tetramer protomers.

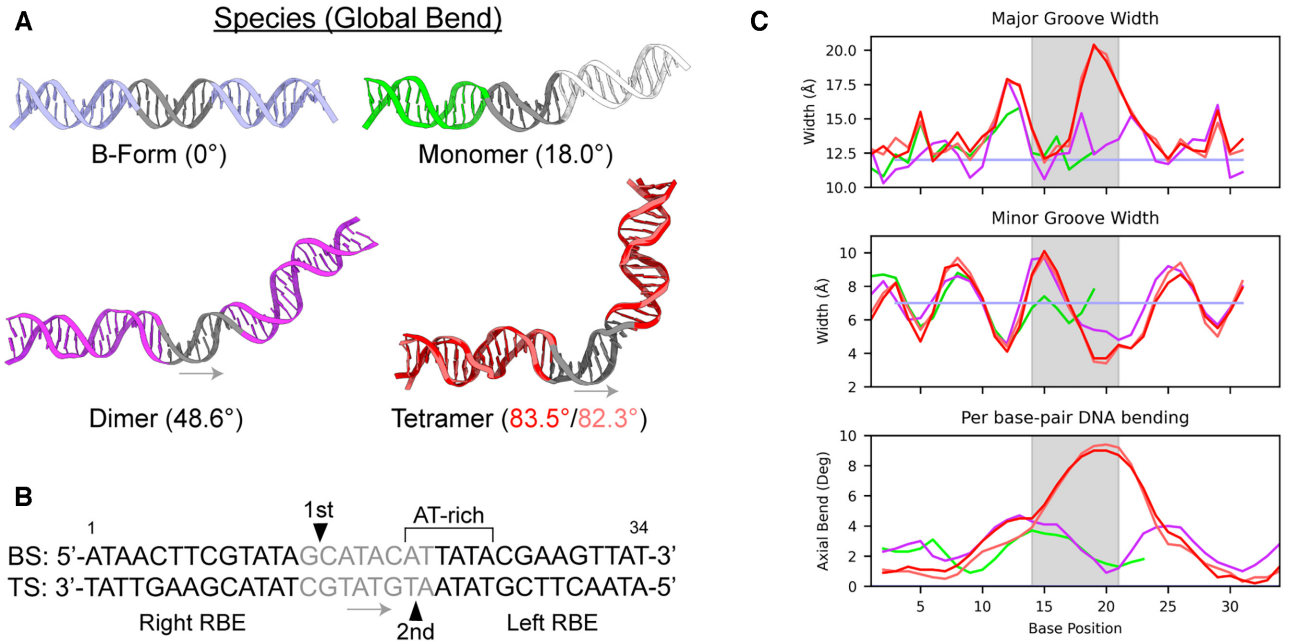

Figure 5. DNA bending and deformations of *loxP* DNA in assembly intermediates. (**A**) Comparison of global bends for B-form (lavender), extended hairpin (green with B-form extension in white), dimer (magenta) and tetrameric (salmon and red) Cre-*loxP* complexes. Spacers are gray and arrows indicate spacer orientation. (**B**) *loxP* DNA sequence. Note: *loxP* orientation is opposite of Figure 1. (**C**) Major and minor groove widths and local, per base-pair DNA bending for each *loxP* DNA substrate were calculated using Curves+ (66). Spacer region shaded in gray.

an overall backbone RMSD of ∼2.1 Å, sharing most features of the active presynaptic complex (16).

This Cryo-EM-derived structure recapitulates key features of the x-ray crystal structure of the K201A tetramer (16). As seen in this crystal structure and others of Cre in tetrameric complexes (12,13,26), each protomer bound to the right RBE adopts a *primed* conformation wherein Y324 is positioned close to the scissile phosphate on the BS, and the Y324 of the protomer bound to the left RBE is far from the scissile phosphate on the TS, adopting an *inactive* conformation (Figure 4F). Primed and inactive conformers in the tetramer are related by a pseudo C2 axis, such that each dimer can be seen as exhibiting half-the-sites activity. The density map was refined with no imposed symmetry, and a separate homogenous refinement in cryoSPARC with C2 symmetry did not yield a map with significant improvements in resolution (∼0.1 Å). Overlay of each Cre_2-*loxP* dimer in the tetramer resulted in an all-atom RMSD of 0.4 Å, similar to that reported in other crystal structures of Cre (14,16,39).

Progressive DNA deformations occur during assembly

Comparative analysis of the assembly intermediates reveals the roles of charge and shape complementarity in recognition of the *loxP* DNA sequence. Protein binding results in marked DNA deformations relative to canonical B-form DNA (Figure 5C) (53,65). Each RBE (Figure 5C, non-shaded regions) experiences similar deformations due to protein binding. Superposition of protomer models by α-carbons, excluding the regions of varied structure (αA, αM, αN and β2–3 loop), yields an overall RMSD of 1.1 Å (Figure 6). In all three structures, helices αJ, αB, and αD are inserted into the major groove of the *loxP* DNA site at bases 6–9, 10–12 and 12–15, respectively (Figure 6). In addition, the β4–5 loop is inserted into the minor groove at bases 1–3 and the αJ-K loop is inserted into the minor groove at bases 9–10 (Figure 6). These interactions result in widening and narrowing of the major and minor grooves that is consistent among the assembly intermediates. The major and minor groove widths within the spacer region (Figure 5C, shaded region) vary between assembly intermediates and follow the global DNA bend that progresses along the assembly path. The protomers in the monomer and dimer complexes, and the inactive protomers in the tetramer, bury roughly 2100 $Å^2$ of surface area at their respective protein–DNA interface. The primed protomers in the tetramer make more extensive DNA contacts burying roughly 2500 $Å^2$ of surface area (Figure 5B) due to engagement of β2–3 loop into the minor groove and increased interactions of side chains from αE with phosphates of bases in the spacer region.

Evolution of PPIs during assembly of synaptic complexes

Assembly of Cre-*loxP* synaptic complexes is accompanied by formation of protein–protein interactions (PPIs) that position motifs important for DNA cleavage by Cre. In the monomer and dimer structures, there is no visible density for the β2–3 loops of each protomer (Figure 4A and B) and two of the three αN helices (Figure 4B and C). In the dimer, docking of αN from Cre^{Yin} on the right RBE into Cre^{Yang} on the left RBE buries ∼690 $Å^2$ of surface area (Supplemental Table S4), but there are no other significant interactions between the protomers. The αN helices and β2–3 loops become structured upon synapsis, with each αN helix docking onto neighboring protomers in a cyclic fashion in

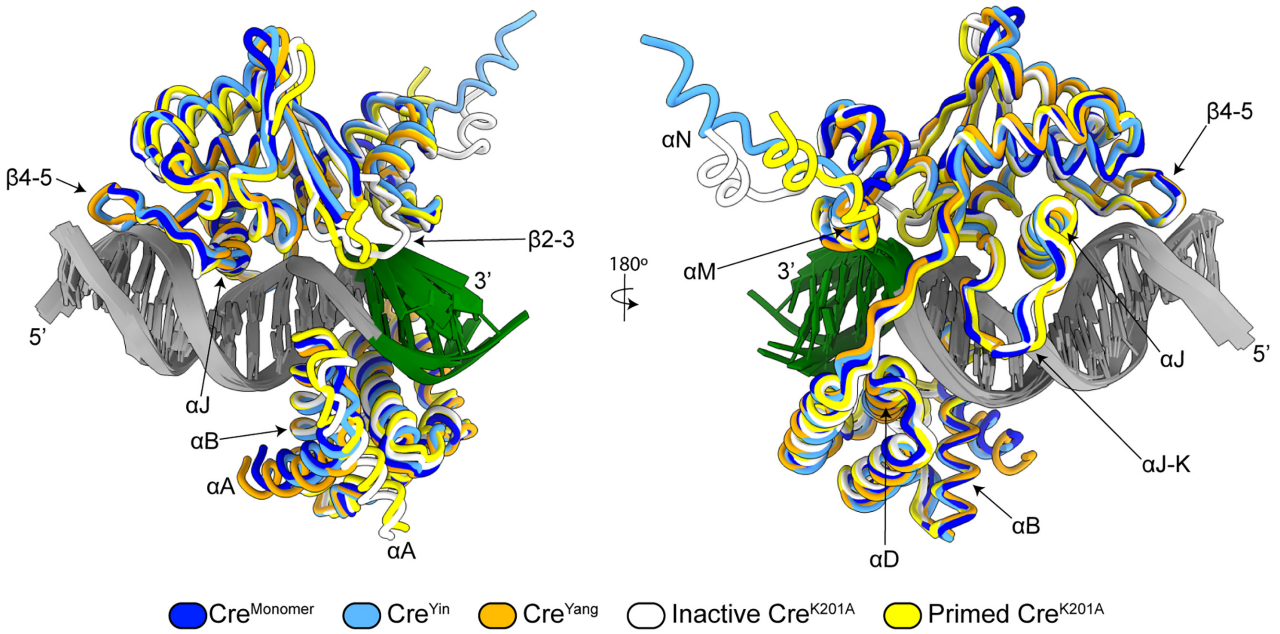

Figure 6. Similarity of site recognition across assembly intermediates. Superposition of the five protein models with important motifs labeled. Superposition performed using all C_αs in α-helices except αA, αM and αN. RBEs are colored gray and the spacer region is colored green.

the tetramer. The β2–3 loops of *primed* protomers on the right RBE insert themselves into the DNA minor groove at the scissile base (Figure 4D). The β2–3 loops originating from the *inactive* protomers on the left RBE are not inserted into the DNA minor groove, and instead pack tightly in a reciprocal fashion against the αM helix of the *primed* protomer whose αN helix is donated to it. Together, the packing of *inactive* protomer β2–3 loops can be seen to confine the αM of each *primed* protomer, while the *primed* protomer does not similarly constrain the αM of *inactive* protomers, allowing those αM-N linkers to adopt extended conformations (Figure 4D, Figure 6). Together with local differences in DNA geometry between the primed and inactive sites, these structural differences position Y324 of αM at the BS scissile phosphate in the *primed* protomer of each right RBE (Figure 4F). Additional structural changes that accompany assembly include the reorientation of αA: in the monomer and dimer structures αA packs against helix αB, whereas it rotates by about 90° to pack against αE of an adjacent protomer when forming the tetrameric complex (Figure 4E).

Conformational sampling of Cre tetrameric complexes

A range of conformations were revealed in the Cre^{K201A}-*loxP* complexes. 3D variability analysis (3DVA) (51) was carried out on the tetramer cryo-EM dataset using two principal components, of which PC1 yielded continuous heterogeneity associated with changes in the degree of *loxP* DNA bending (Supplemental Movie S1). We used ISOLDE to flexibly fit the consensus model into the start and end conformations of PC1 (Figure 7A) to better interpret the structural changes connecting the first and last frames (58). The bending angle of the DNA varies by about 10° over the course of the modeled heterogeneity (Figure 7b) (66). In the first frame (F000), the *loxP* DNA is bent at about 75°, and there is no clear density for either inactive or primed β2–3 loops (Figure 7B and C). The angle of DNA bending increases to about 85° in the last frame (F019) as the protomers rotate along the *loxP* DNA interface, and clear density for the β2–3 loop is observed (Figure 7B and C). Increased DNA bending and rigidification of the β2–3 loop is thus coincident with positioning of the primed protomer's Y324 near the scissile phosphate (Figure 7D).

DISCUSSION

Cryo-EM offers new insights into Cre assembly and function

These cryo-EM studies of Cre-*loxP* assembly intermediates provide new insights into mechanisms of site recognition, half-the-sites activity, strand cleavage order, and concerted strand cleavage. The cryo-EM structure of tetrameric Cre^{K201A}-*loxP* complex reaffirms the key attributes of an active, pre-cleavage complex revealed by prior crystallographic studies of the Cre^{K201A}-*loxP* complex and lends new insights into concerted conformational changes required for activation, DNA bending, and maintenance of half-sites activity (16). The structure of the Cre_2-*loxP* dimer supports long standing hypothesis of initial regulation of half-sites activity and order of strand exchange. The structure of the Cre-*loxA* monomer (54 kDa) shows similar protein–DNA contacts to those of tetrameric proteins and provides a structural explanation for its lack of topoisomerase activity.

Monomeric Cre specifically binds and recognizes a single RBE but cannot cleave it

To prevent intermolecular protein–protein interactions meditated through helix αN, we assembled complexes of Cre^{Yin} (which bears a mutation at its αN docking site) on

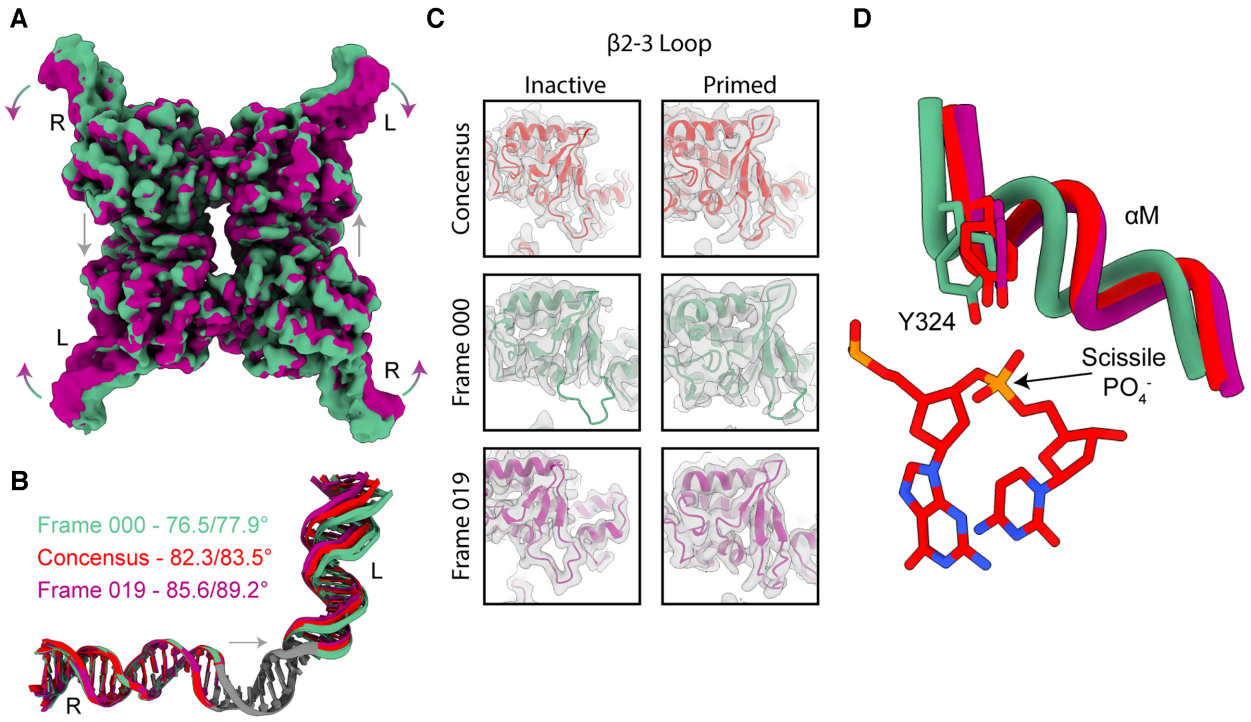

Figure 7. Flexibility in protein–protein interfaces and differing degrees of protomer activation in 3D Variability Analysis of the Cre^{K201A}-*loxP* dataset. (**A**) Density maps filtered at 4 Å for frame 000 (green) and frame 019 (purple). Arrows indicating the interpolated movement between the start and end frames. (**B**) Superposition of modeled DNAs from frame 000 (green), frame 019 (purple), and the consensus refinement (red). (**C**) Density and models (chain A, inactive; chain B, primed) for the β2–3 loop for frame 000, frame 019 and consensus refinements. (**D**) Models (chain B) from consensus (red), frame 000 (green) and frame 019 (purple) refinements showing the relationship between Y324 and the scissile phosphate.

a *loxA* hairpin substrate comprising a single RBE and symmetrized 'right' spacer (Figure 1). In the cryo-EM density map of the Cre-*loxP* half-site complex, the αJ-K loop is ordered, consistent with a role in DNA binding. However, no density is observed for helix αN and the β2–3 loop, presumably due to the absence of stabilizing protein–protein interactions. This is consistent with their role in mediating protein–protein interactions with neighboring protomers and structural variability in the *inactive* and *primed* conformations. Indeed, the 3.9 Å structure is highly congruent with the *inactive* protomer in the synaptic complex, with an overall backbone RMSD of ∼1 Å (Figure 6), and Y324 is far from the scissile phosphate (Figure 4F). Since K201 resides at the tip of the β2–3 loop, the active site is not fully formed at this step of assembly. The monomeric Cre-*loxP* complex reveals the same major and minor groove protein–DNA contacts observed in protomers of fully assembled synaptic complexes (Supplemental Figure S9), indicating that sequence specific recognition of the RBE is realized upon a single Cre protein binding to a *loxP* half-site. Comparison of individual protein–phosphate backbone contacts observed in the monomer (Cre-*loxA* complex) and in the active and inactive protomers of the tetrameric complex underscores the role of both base-specific contacts and shape recognition (indirect readout (35,36)) in site selection (Supplemental Figure S9). Thus, the cryo-EM structure of the monomer is consistent with the observed site-specific binding of Cre and the lack of topoisomerase activity due to absence of a fully formed active site from insufficient protein–protein interactions (13,26,38,71).

Half-sites activity and order of strand exchange inferred from pre-synaptic Cre-*loxP* dimers

The use of Cre^{Yin}/Cre^{Yang} pairs allowed for isolation and characterization of pre-synaptic Cre_2-*loxP* complexes featuring a native protein–protein dimer interface (Figures 3B and 4C). A defining feature of dimer assembly is *trans* docking by the αN of Cre^{Yin} onto its docking site on Cre^{Yang}, accompanied by an increase in DNA bending and increased DNA deformations of bases proximal to and in the spacer (Figure 5A and C). Once two Cre protomers are assembled onto a *loxP* site, directional DNA bending results from the αN of one protomer docking into its site on the adjacent protomer. Remembering that directionality of the *loxP* site is determined by the asymmetric spacer sequence, in a BS complex (Figure 4C) the *loxP* DNA is favorably bent within the AT-rich region to allow the αN helix of the protomer bound to the right RBE (blue) to dock into the protomer positioned on the left RBE (orange); the αN of the recipient protomer then extends out into solution in the opposite direction as the DNA bend. Thus, in the case of synapsis by two BS dimers *only the protomers positioned to cleave the BS (*Figure 4C, blue) *can be activated for cleavage* through packing with the β2–3 loops of the adjacent (orange) protomers. Conversely, if the *loxP* is bent in the opposite direction, the αN from the protomer on the left RBE will dock on the protomer on the right RBE, and the dimer will be poised for TS cleavage.

Productive synapsis occurs when two similarly bent dimers encounter each other in an antiparallel orientation

(69,70). In this model, tetrameric TS and BS complex formation is determined by the probability of a Cre_2-*loxP* dimer forming bends that favor BS or TS cleavage, and then encountering another similarly bent dimer. Formation of BS-BS synapses would result in bottom strand first cleavage, TS–TS synapses could cleave the top strands, whereas assembly of BS–TS complexes would result in unproductive, or mutagenic recombination events (73). Assuming similar efficiency for recombination by either BS–BS or TS–TS synaptic complexes, an ∼80% preference for BS strand first recombination would be explained by a small 15% bias (kT ln (0.65/0.35) ∼0.6 kT) toward formation of BS-bent dimers [$0.65 \times 0.65 \div (0.65 \times 0.65 + 0.35 \times 0.35) = 78\%$] (72,74). An 88% BS preference (74) would correspond to a bend bias of ∼1 kT.

The asymmetry of this protein–protein interaction also illuminates the basis for half-the-sites activity of the enzyme (14,19). In tetrameric complexes, the packing of β2–3 of one protomer (Figure 4D, white) onto the αM of an adjacent protomer (Figure 4D, yellow) activates it (white arrows) for cleavage. However, in the other (synaptic) interface, the packing of the β2–3 loops with the αM of their adjacent protomers (red arrows) is insufficient to activate that pair of protomers for cleavage. Thus, the dynamics and conformational state (Figure 4A–D and Figure 7C) of the β2–3 loop can be seen as acting as a switch driven by protein–protein interactions, that results in activation of alternating active sites, thereby preventing double strand breaks. Therefore, the structure of the Cre_2-*loxP* dimer provides new insights into the structural basis for cooperative DNA binding (by *trans* αN docking), for the observed BS/TS cleavage preferences (biased spacer bending), and half-the sites activity due to poorly formed protein–protein interfaces that fail to assemble a complete active site at either scissile phosphate.

A model for assembly and activation of Cre recombinase

The new structures support a model in which protein and DNA dynamics play important roles in Cre-*loxP* synaptic complex assembly and activation (Figure 8): (i) Sequestration of helix αN over the C-terminal domain DNA binding surface prevents premature oligomerization of Cre protomers, while flexibility in the linker facilitates C-clamp formation by the independently folding N- and C-terminal domains (34). A single Cre protein binds to an RBE and extends a flexible helix αN into solution. Incomplete formation of the Cre active site, due to insufficient protein–protein interactions to stabilize the β2–3 loop and helices αM and αN, results in a loosely bound complex incapable of premature DNA cleavage, while enabling DNA sequence scanning and intersegmental transfer (75,76). (ii) The αN helix, extended and flexibly sampling multiple conformations, uses a fly-casting mechanism to capture a second Cre protomer on an adjacent RBE of the *loxP* site. Biased bending of the asymmetric spacer of doubly bound *loxP* DNA allows sampling of TS and BS complexes, each of which could be reversibly stabilized by docking of helix αN from one protomer into its site on the other protomer in the dimer. Conformational sampling is enabled by inherent flexibility of the DNA, and of the protein: the β2–3 loop, helix α, as well as helix αA, which result in ill-formed protein–protein interactions in the dimer. (iii) Pairing of antiparallel duplexes with matching bends is mediated by fly-casting by the free αN of each dimer to result in synapsis, wherein each synaptic complex is primed for either BS- or TS-first cleavage. Consistent with the observation of a third protomer recruited to dimer complexes, this model poses that synapsis occurs in stepwise fashion: (i) capturing of one protomer in a dimer by the free αN of another dimer complex, thereby tethering two dimer complexes together; (ii) concerted dynamics that result in full capture of the second dimer. (iii) Productive synapsis of two like dimers is accompanied by additional DNA bending, and assembly of active sites through stabilization of protein–protein interfaces by rearrangements of αA, β2–3 and αM–N. Conformational sampling in the tetrameric complex, described by 3DVA, suggests that synapsis is not coincident with activation, and hints at a propensity for the conformational changes that follow initial DNA cleavage. Stochastic and concerted conformational sampling involving the flexible protein–protein interfaces would also facilitate strand exchange and the subsequent isomerization step that exchanges *primed* and *inactive* protomer conformations.

Insights into *loxP* site recognition

The potential for repurposing Cre recombinase for lesion-free DNA recombination in a cellular context has motivated broad efforts to understand the biochemical basis for its sequence specificity (12,13,77). Specificity could be achieved at the level of highly selective assembly of Cre tetramers on target sites, or by much higher recombination efficiency at those sites (i.e. k_{cat}/K_M). Optimizing these properties in a rational way requires knowledge of the structural features that drive both activities.

Recombination assays with WT Cre and variants have shown that DNA sequence specificity is driven by some amino acid residues that in tetrameric crystal structures do not make direct contacts to DNA bases in the RBEs, as well as those that do (5–7,20,28–30,33,78–82). The structures of the Cre-*loxP* monomer, dimer and tetramer feature the same sets of contacts with the RBEs (from helices αB, αD, αJ, the αJ-K loop and the β4–5 loop; Supplemental Figure S9), and similar deviations from canonical B-DNA helical parameters (Figures 5 and 6; Supplemental Figure S8). These observations suggest that no major RBE recognition elements evolve during the assembly process, underscoring the dual importance of direct and indirect readout.

Despite almost no direct interaction between Cre and bases in the spacer DNA, its sequence is nevertheless an important determinant for recombination by Cre (31,83,84). *In vitro* and *in vivo* experiments have shown that even single base pair changes to the spacer region can reduce recombination efficiency to greater than 90% by wild-type Cre (83). Thus, it is reasonable that these changes might alter the structure or rigidity of the spacer region and thus its ability to adopt cleavage competent conformations (33,82).

Nevertheless, directed evolution experiments have produced Cre variants capable of efficiently recombining non-cognate sites (5,7,78,80,82). These variants contain mutations mapping primarily to protein–protein interfaces in tetrameric complexes, suggesting that such mutations in

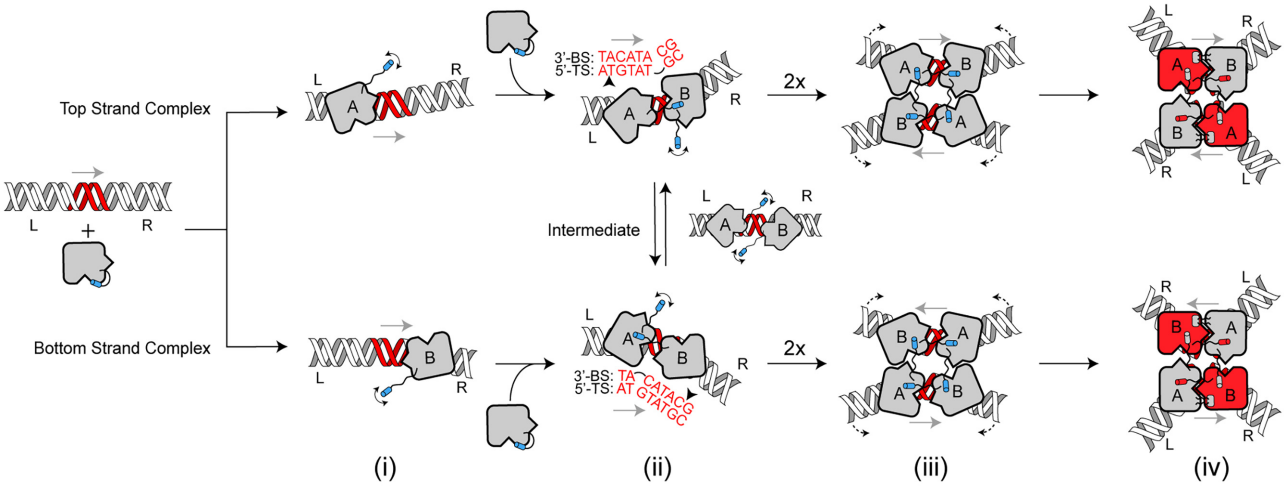

Figure 8. Assembly pathway of Cre recombinase on *loxP* DNA. (**i**) Binding of the first Cre protomer to an RBE dislodges αN from its autoinhibited conformation and induces a slight bend (∼18°) in the DNA; 'A' and 'B' identify protomers bound to the left or right RBEs ('L', 'R'), respectively. (**ii**) Fly-casting by the extended and flexible αN (blue cylinders) captures a second Cre protomer on an adjacent RBE; this is accompanied by further DNA bending (to ∼50°) to form top strand (TS) or bottom strand (BS) dimer complexes. Isomerization between TS and BS complexes can occur through reversible undocking of the αN to form an extended intermediate; DNA re-bending and subsequent αN docking generates either TS or BS dimers. (**iii**) Synapsis of two similarly bent Cre_2-*loxP* dimers (TS or BS), facilitated by fly-casting by the free αN on each dimer, results in initial tetrameric complexes that undergo further conformational sampling (dotted arrows) to form either TS or BS synaptic tetramers. (**iv**) The fully formed synaptic tetramers feature ∼83° bends in the DNA and alternating pairs of *primed* and *inactive* protomers; packing of the β2–3 loop (small, rounded rectangles) across the spacer activates the recipient protomers (red) to cleave either the TS, or BS. Additional conformational sampling of the synaptic complexes (evident in 3DVA analysis) may help promote cleavage and enable subsequent recombination steps. Gray arrows indicate the direction of the asymmetric spacer sequence (red). Cleavage sites on the top and bottom strands (TS, BS) are indicated by black arrowheads.

early rounds of selection might function by relaxing specificity for these altered sites, whereby cleavage-competent protein–protein and protein–DNA conformations can be sampled (5,6,33,82). This idea is supported by the observations of the variable DNA distortions and protein–protein interfaces observed between the dimeric complex and in the 3DVA series for the tetrameric complex. Thus, the balance between protein–protein interactions and DNA flexibility seems to be an important determinant for selecting DNA sites.

CONCLUSION

We used cryo-EM to determine the structures of three Cre-*loxP* complexes along the path of site selection and synapsis to form recombinogenic tetramers. From these structures, it is evident that in addition to direct recognition of the palindromic *loxP* sites, DNA plasticity, evolution of protein–protein interactions, and protein–DNA conformational sampling are coupled to the process by which Cre selectively assembles on and is poised to cleave *loxP* DNA. These considerations may help guide future studies aimed at developing Cre/*lox* variants targeting alternative sequences.

DATA AVAILABILITY

Atomic models are available in the PDB for the monomer, dimer and tetramer at the following accession codes, respectively: 7RHY, 7RHZ, and 7RHX. Cryo-EM maps are available in the EMDB for the monomer, dimer and tetramer at the following accession codes, respectively: EMD-24471, EMD-24472, EMD-24470. Data not deposited into the PDB or EMDB are available upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Greg Van Duyne for providing Cre expression vectors and encouragement and Dr Aparna Unnikrishnan for preparation of Cre samples. Electron microscopy was performed at the Center for Electron Microscopy and Analysis (CEMAS) at The Ohio State University. We thank Dr Yoshie Narui (CEMAS) for training and support with Cryo-EM experiments and access to computational resources. Lastly, we thank members of the Foster and Wysocki labs for helpful discussions and support throughout this work.

FUNDING

National Institute of Health [R01 GM122432 to M.P.F., P41 GM128577 to V.H.W., S10 OD023582 to J.Jackman].
Conflict of interest statement. None declared.

REFERENCES

1. Sadowski,P. (1986) Site-specific recombinases: changing partners and doing the twist. *J. Bacteriol.*, **165**, 341–347.
2. Kilby,N.J., Snaith,M.R. and Murray,J.A.H. (1993) Site-specific recombinases: tools for genome engineering. *Trends Genet.*, **9**, 413–421.
3. Sternberg,N. and Hamilton,D. (1981) Bacteriophage pl Site-specific recombination I. Recombination between loxP sites. *J. Mol. Biol.*, **150**, 467–486.
4. Saraf-Levy,T., Santoro,S.W., Volpin,H., Kushnirsky,T., Eyal,Y., Schultz,P.G., Gidoni,D. and Carmi,N. (2006) Site-specific

recombination of asymmetric lox sites mediated by a heterotetrameric cre recombinase complex. *Bioorg. Med. Chem.*, **14**, 3081–3089.
5. Buchholz,F. and Stewart,A.F. (2001) Alteration of cre recombinase site specificity by substrate-linked protein evolution. *Nat. Biotechnol.*, **19**, 1047–1052.
6. Santoro,S.W. and Schultz,P.G. (2002) Directed evolution of the site specificity of cre recombinase. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 4185–4190.
7. Karpinski,J., Hauber,I., Chemnitz,J., Schäfer,C., Paszkowski-Rogacz,M., Chakraborty,D., Beschorner,N., Hofmann-Sieber,H., Lange,U.C., Grundhoff,A. *et al.* (2016) Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high specificity. *Nat. Biotechnol.*, **34**, 401–409.
8. Abi-Ghanem,J., Chusainow,J., Karimova,M., Spiegel,C., Hofmann-Sieber,H., Hauber,J., Buchholz,F. and Pisabarro,M.T. (2013) Engineering of a target site-specific recombinase by a combined evolution-and structure-guided approach. *Nucleic Acids Res.*, **41**, 2394–2403.
9. Shaw,D., Serrano,L. and Lluch-Senar,M. (2021) Lox'd in translation: contradictions in the nomenclature surrounding common lox-site mutants and their implications in experiments. *Microbiology*, **167**, 000997.
10. Abremski,K., Hoess,R. and Sternberg,N. (1983) Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell*, **32**, 1301–1311.
11. Sternberg,N. and Hamilton,D. (1981) Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J. Mol. Biol.*, **150**, 467–486.
12. van Duyne,G.D. (2015) Cre recombinase. *Microbiol. Spectr*, **3**, MDNA3–0014–2014.
13. Meinke,G., Bohm,A., Hauber,J., Pisabarro,M.T. and Buchholz,F. (2016) Cre recombinase and other tyrosine recombinases. *Chem. Rev.*, **116**, 12785–12820.
14. Guo,F., Gopaul,D.N. and Van Duyne,G.D. (1997) Structure of cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.
15. Gopaul,D.N., Guo,F. and Van Duyne,G.D. (1998) Structure of the holliday junction intermediate in Cre-loxP site-specific recombination. *EMBO J.*, **17**, 4175–4187.
16. Ghosh,K., Guo,F. and Van Duyne,G.D. (2007) Synapsis of loxP sites by cre recombinase. *J. Biol. Chem.*, **282**, 24004–24016.
17. Woods,K.C., Martin,S.S., Chu,V.C. and Baldwin,E.P. (2001) Quasi-equivalence in site-specific recombinase structure and function: crystal structure and activity of trimeric cre recombinase bound to a three-way lox DNA junction. *J. Mol. Biol.*, **313**, 49–69.
18. Ringrose,L., Lounnas,V., Ehrlich,L., Buchholz,F., Wade,R. and Stewart,A.F. (1998) Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. *J. Mol. Biol.*, **284**, 363–384.
19. Guo,F., Gopaul,D.N. and van Duyne,G.D. (1999) Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Nat. Acad. Sci. U.S.A.*, **96**, 7143–7148.
20. Zhang,C., Myers,C.A., Qi,Z., Mitra,R.D., Corbo,J.C. and Havranek,J.J. (2015) Redesign of the monomer-monomer interface of cre recombinase yields an obligate heterotetrameric complex. *Nucleic Acids Res.*, **43**, 9076–9085.
21. Aihara,H., Kwon,H.J., Nunes-Düby,S.E., Landy,A. and Ellenberger,T. (2003) A conformational switch controls the DNA cleavage activity of λ integrase. *Mol. Cell*, **12**, 187–198.
22. Subramaniam,S., Tewari,A.K., Nunes-Duby,S.E. and Foster,M.P. (2003) Dynamics and DNA substrate recognition by the catalytic domain of lambda integrase. *J. Mol. Biol.*, **329**, 423–439.
23. Serre,M.C., el Arnaout,T., Brooks,M.A., Durand,D., Lisboa,J., Lazar,N., Raynal,B., van Tilbeurgh,H. and Quevillon-Cheruel,S. (2013) The carboxy-terminal αN helix of the archaeal XerA tyrosine recombinase is a molecular switch to control site-specific recombination. *PLoS One*, **8**, e63010.
24. Hickman,A.B., Waninger,S., Scocca,J.J. and Dyda,F. (1997) Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. *Cell*, **89**, 227–237.
25. Kazmierczak,R.A., Swalla,B.M., Burgin,A.B., Gumport,R.I. and Gardner,J.E. (2002) Regulation of site-specific recombination by the C-terminus of λ integrase. *Nucleic Acids Res.*, **30**, 5193–5204.
26. van Duyne,G.D. (2001) A structural view of Cre-loxP site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 87–104.
27. Ennifar,E., Meyer,J.E.W., Buchholz,F., Stewart,A.F. and Suck,D. (2003) Crystal structure of a wild-type cre recombinase-IoxP synapse reveals a novel spacer conformation suggesting an alternative mechanism for DNA cleavage activation. *Nucleic Acids Res.*, **31**, 5449–5460.
28. Wierzbicki,A., Kendall,M., Abremski,K. and Hoess,R. (1987) A mutational analysis of the bacteriophage pl recombinase cre. *J. Mol. Biol.*, **195**, 785–794.
29. Hartung,M. and Kisters-Woike,B. (1998) Cre mutants with altered DNA binding properties. *J. Biol. Chem.*, **273**, 22884–22891.
30. Kim,S.-T., Kim,G.-W., Lee,Y.-S. and Park,J.-S. (2001) Characterization of Cre±loxP interaction in the major groove: hint for structural distortion of mutant cre and possible strategy for HIV-1 therapy. *J. Cell. Biochem.*, **80**, 321–327.
31. Missirlis,P.I., Smailus,D.E. and Holt,R.A. (2006) A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination. *BMC Genomics*, **7**, 73.
32. Abi-Ghanem,J., Samsonov,S.A. and Pisabarro,M.T. (2015) Insights into the preferential order of strand exchange in the Cre/loxP recombinase system: impact of the DNA spacer flanking sequence and flexibility. *J. Comput. Aided Mol. Des.*, **29**, 271–282.
33. Baldwin,E.P., Martin,S.S., Abel,J., Gelato,K.A., Kim,H., Schultz,P.G. and Santoro,S.W. (2003) A specificity switch in selected cre recombinase variants is mediated by macromolecular plasticity and water. *Chem. Biol.*, **10**, 1085–1094.
34. Unnikrishnan,A., Amero,C., Yadav,D.K., Stachowski,K., Potter,D. and Foster,M.P. (2020) DNA binding induces a cis-to-trans switch in Cre recombinase to enable intasome assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 24849–24858.
35. Otwinowski,Z., Schevitz,R.W., Zhang,R.G., Lawson,C.L., Joachimiak,A., Marmorstein,R.Q., Luisi,B.F. and Sigler,P.B. (1988) Crystal structure of trp represser/operator complex at atomic resolution. *Nature*, **335**, 321–329.
36. Koudelka,G.B., Mauro,S.A. and Ciubotaru,M. (2006) Indirect readout of DNA sequence by proteins: the roles of DNA sequence-dependent intrinsic and extrinsic forces. *Prog. Nucleic Acid Res. Mol. Biol.*, **81**, 143–177.
37. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
38. Yang,W. (2010) Topoisomerases and site-specific recombinases: similarities in structure and mechanism. *Crit. Rev. Biochem. Mol. Biol.*, **45**, 520–534.
39. Van Duyne,G.D. (2001) A structural view of Cre-loxP site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 87–104.
40. Krissinel,E. and Henrick,K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
41. Evgeny,K. (2010) Crystal contacts as nature's docking solutions. *J. Comput. Chem.*, **31**, 133–143.
42. Rodrigues,C.H.M., Myung,Y., Pires,D.E.v and Ascher,D.B. (2019) mCSM-PPI2: predicting the effects of mutations on protein–protein interactions. *Nucleic Acids Res.*, **47**, 338–344.
43. Green,M.R. and Sambrook,J. (2018) Precipitation of DNA with ethanol. *Cold Spring Harb. Protoc.*, **2016**, 1116–1121.
44. Ghosh,K. and van Duyne,G.D. (2002) Cre-loxp biochemistry. *Methods*, **28**, 374–383.
45. Vanaernum,Z.L., Gilbert,J.D., Belov,M.E., Makarov,A.A., Horning,S.R. and Wysocki,V.H. (2019) Surface-Induced dissociation of noncovalent protein complexes in an extended mass range orbitrap mass spectrometer. *Anal. Chem.*, **91**, 3611–3618.
46. Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D. and Bairoch,A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31**, 3784–3788.
47. Marty,M.T., Baldwin,A.J., Marklund,E.G., Hochberg,G.K.A., Benesch,J.L.P. and Robinson,C.V. (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Anal. Chem.*, **87**, 4370–4376.

48. Punjani,A., Rubinstein,J.L., Fleet,D.J. and Brubaker,M.A. (2017) CryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods*, **14**, 290–296.
49. Bepler,T., Morin,A., Rapp,M., Brasch,J., Shapiro,L., Noble,A.J. and Berger,B. (2019) Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods*, **16**, 1153–1160.
50. Punjani,A., Zhang,H. and Fleet,D.J. (2020) Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods*, **17**, 1214–1221.
51. Punjani,A. and Fleet,D.J. (2021) 3D variability analysis: resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.*, **213**, 107702.
52. Ulyanov,N.B., Bauer,W.R. and James,T.L. (2002) High-resolution NMR structure of an AT-rich DNA sequence. *J. Biomol. NMR*, **22**, 265–280.
53. Li,S., Olson,W.K. and Lu,X.-J. (2019) Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res.*, **47**, 26–34.
54. Goddard,T.D., Huang,C.C., Meng,E.C., Pettersen,E.F., Couch,G.S., Morris,J.H. and Ferrin,T.E. (2018) UCSF chimerax: meeting modern challenges in visualization and analysis. *Protein Sci.*, **27**, 14–25.
55. Pettersen,E.F., Goddard,T.D., Huang,C.C., Meng,E.C., Couch,G.S., Croll,T.I., Morris,J.H. and Ferrin,T.E. (2021) UCSF chimerax: structure visualization for researchers, educators, and developers. *Protein Sci.*, **30**, 70–82.
56. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**, 486–501.
57. Afonine,P.V., Poon,B.K., Read,R.J., Sobolev,O.V., Terwilliger,T.C., Urzhumtsev,A. and Adams,P.D. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D: Struct. Biol.*, **74**, 531–544.
58. Croll,T.I. (2018) ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D: Struct. Biol.*, **74**, 519–530.
59. Lin,C.C. and Potter,C.J. (2016) Editing transgenic DNA components by inducible gene replacement in drosophila melanogaster. *Genetics*, **203**, 1613–1628.
60. 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.
61. Leney,A.C. and Heck,A.J.R. (2017) Native mass spectrometry: what is in the name?*J. Am. Soc. Mass. Spectrom.*, **28**, 5–13.
62. Santambrogio,C., Natalello,A., Brocca,S., Ponzini,E. and Grandori,R. (2019) Conformational characterization and classification of intrinsically disordered proteins by native mass spectrometry and charge-state distribution analysis. *Proteomics*, **19**, e1800060.
63. Erba,E.B. and Zenobi,R. (2011) Mass spectrometric studies of dissociation constants of noncovalent complexes. *Annu. Rep. Progr. Chem. C*, **107**, 199–228.
64. Gibb,B., Gupta,K., Ghosh,K., Sharp,R., Chen,J. and van Duyne,G.D. (2010) Requirements for catalysis in the cre recombinase active site. *Nucleic Acids Res.*, **38**, 5817–5832.
65. Lu,X.J. and Olson,W.K. (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.*, **3**, 1213–1227.
66. Lavery,R., Moakher,M., Maddocks,J.H., Petkeviciute,D. and Zakrzewska,K. (2009) Conformational analysis of nucleic acids revisited: curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
67. Lee,L., Chu,L.C.H. and Sadowski,P.D. (2003) Cre induces an asymmetric DNA bend in its target loxP site. *J. Biol. Chem.*, **278**, 23118–23129.
68. Ghosh,K., Lau,C.K., Gupta,K. and van Duyne,G.D. (2005) Preferential synapsis of loxp sites drives ordered strand exchange in cre-loxp site-specific recombination. *Nat. Chem. Biol.*, **1**, 275–282.
69. Scheres,S.H.W. (2016) Processing of structurally heterogeneous Cryo-EM data in RELION. In: *Methods in Enzymology*. Academic Press Inc., Vol. **579**, pp. 125–157.
70. Serna,M. (2019) Hands on methods for high resolution cryo-electron microscopy structures of heterogeneous macromolecular complexes. *Front. Mol. Biosci.*, **6**, 33.
71. Grindley,N.D.F., Whiteson,K.L. and Rice,P.A. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem.*, **75**, 567–605.
72. Pinkney,J.N.M., Zawadzki,P., Mazuryk,J., Arciszewska,L.K., Sherratt,D.J. and Kapanidis,A.N. (2012) Capturing reaction paths and intermediates in Cre-loxP recombination using single-molecule fluorescence. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 20871–20876.
73. Martin,S.S., Pulido,E., Chu,V.C., Lechner,T.S. and Baldwin,E.P. (2002) The order of strand exchanges in Cre-LoxP recombination and its basis suggested by the crystal structure of a Cre-LoxP holliday junction complex. *J. Mol. Biol.*, **319**, 107–127.
74. Lee,L. and Sadowski,P.D. (2003) Sequence of the loxP site determines the order of strand exchange by the cre recombinase. *J. Mol. Biol.*, **326**, 397–412.
75. Tafvizi,A., Mirny,L.A. and Van Oijen,A.M. (2011) Dancing on DNA: kinetic aspects of search processes on DNA. *ChemPhysChem*, **12**, 1481–1489.
76. Berg,O.G., Winter,R.B. and von Hippel,P.H. (1981) Diffusion-Driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
77. Bogdanove,A.J., Bohm,A., Miller,J.C., Morgan,R.D. and Stoddard,B.L. (2018) Engineering altered protein-DNA recognition specificity. *Nucleic Acids Res.*, **46**, 4845–4871.
78. Eroshenko,N. and Church,G.M. (2013) Mutants of cre recombinase with improved accuracy. *Nat. Commun.*, **4**, 2509.
79. Bessen,J.L., Afeyan,L.K., Dančík,V., Koblan,L.W., Thompson,D.B., Leichner,C., Clemons,P.A. and Liu,D.R. (2019) High-resolution specificity profiling and off-target prediction for site-specific DNA recombinases. *Nat. Commun.*, **10**, 1937
80. Rüfer,A.W. and Sauer,B. (2002) Non-contact positions impose site selectivity on cre recombinase. *Nucleic Acids Res.*, **30**, 2764–2771.
81. Sarkar,I., Hauber,I., Hauber,J. and Buchholz,F. (2007) HIV-1 proviral DNA excision using an evolved recombinase. *Science*, **316**, 1912–1915.
82. Meinke,G., Karpinski,J., Buchholz,F. and Bohm,A. (2017) Crystal structure of an engineered, HIV-specific recombinase for removal of integrated proviral DNA. *Nucleic Acids Res.*, **45**, 9726–9740.
83. Lee,G. and Saito,I. (1998) Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene*, **216**, 55–65.
84. Sheren,J., Langer,S.J. and Leinwand,L.A. (2007) A randomized library approach to identifying functional lox site domains for the cre recombinase. *Nucleic Acids Res.*, **35**, 5464–5473.