Structural basis of DNA recombination catalysis and regulation by ϕC31 integrase

Yunyuan E. Sun[1,2], Louie Aspinall[1,a], Agnel P. Joseph[3], Sean D. Colloms[1*], W. Marshall Stark[1*], Laura Spagnolo[1,2*]

[1]School of Molecular Biosciences, University of Glasgow, University Avenue, Glasgow G12 8QQ, United Kingdom
[2] Research Complex at Harwell, Rutherford Appleton Laboratory, Didcot OX11 0FA, United Kingdom
[3] Scientific Computing Department, UKRI Science and Technology Facilities Council, Research Complex at Harwell, Didcot, United Kingdom
[*] Corresponding authors
[a] Current address: Astbury Centre for Structural Biology, University of Leeds, United Kingdom

Abstract
Large serine integrases catalyse the insertion (integration) and excision of bacteriophage genomes into and from the genomes of their bacterial hosts via site-specific recombination. The recombination reaction is notable for its striking directionality: integrase recognises and efficiently recombines a ~50-bp attachment site in the phage DNA (*attP*), and a ~40-bp site in the bacterial genome (*attB*), but is inactive on the product sites *attL* and *attR* unless another phage-encoded protein, the Recombination Directionality Factor (RDF) is present, when the reaction is reversed. Serine integrases are versatile tools for applications in experimental biology, biotechnology, synthetic biology and gene therapy. Here we report structures of ϕC31 integrase, the most-studied and most widely used member of the serine integrase family, in complexes with its DNA recombination sites. Our structures, shed light on how the C-terminal coiled-coil motif facilitates recombination and directionality control.

Introduction

Bacteriophage integrases catalyse the insertion (integration) of circular double-stranded phage DNA into the genome of the bacterial host by site-specific recombination, allowing the phage to enter lysogeny. The integrase from bacteriophage ϕC31 belongs to the family of serine recombinases[1,2] and promotes site-specific recombination between the phage recombination site *attP* and the *attB* site on its *Streptomyces* host's genome. This reaction can be reconstituted *in vitro* using purified integrase and substrates containing *attP* and *attB* sites. Recombination starts with integrase dimers recognizing and binding to *attP* and *attB* sites. Dimer-bound *attP* and *attB* then interact via protein-protein intereactions to form a synaptic complex in which *attP* and *attB* are held together on the outside of an integrase tetramer. Integrase cleaves both sites at 2-nucleotide staggered positions near the centre of both sites using a catalytic serine residue. The reaction proceeds by exchange of half sites and re-ligation to form two recombinant sites, *attL* and *attR* (Figure 1A). The reaction is very efficient and unidirectional; *attL* and *attR* sites do not recombine with each other in the presence of integrase on its own so that integration is essentially irreversible. Remarkably however, the presence of an additional phage-encoded protein, the Recombination Directionality Factor (RDF), transforms the activity of integrase so that it promotes efficient *attL* × *attR* recombination, and *attP* × *attB* recombination is inhibited[3]. In general, RDF expression is activated on entry into the phage lytic cycle, to allow integrase to excise the provirus genome from the bacterial chromosome[1].

Serine integrases are thought to share a strand exchange mechanism called subunit rotation with all other members of the serine recombinase family[2,4,5]. After double strand cleavage in the synaptic complex, it is proposed that one half of the complex rotates through 180° relative to the other half and the cleavage steps are reversed to form recombinant sites (Figure 1B).The catalytic domain of ϕC31 integrase (NTD, residues 1-163) is at the extreme N-terminal end of the 605-amino acid polypeptide chain, and is related by sequence similarity to the catalytic domains of the other serine recombinases. The catalytic domain is followed by a central so-called "recombinase domain" (RD, residues 164-344, also known as DNA-binding domain 1) which is characteristic of the "large" serine recombinases including serine integrases, and finally a large C-terminal "zinc ribbon domain" (ZD, residues 355-605, also know as DNA-binding domain 2) (Figure 1B). The recombinase domain and zinc ribbon domain are involved in specific recognition of the *attP* and *attB* recombination sites. Within the zinc ribbon domain is a coiled-coil structural motif (CC, residues 450-518) which plays crucial roles in synapsis and directionality control [1,6]. The ϕC31 integrase sequence can be aligned along its entire length with the sequences of other serine integrases, revealing a number of insertions in the ϕC31 sequence between conserved sequence elements, compared to shorter members of the group[7].

Crystal structures of a C-terminal fragment of the LI serine integrase consisting of the RD and ZD bound to its cognate attP DNA revealed that the CC motif has a very flexible connection with the rest of the zinc ribbon domain[6]. These structures allowed Rutherford *et al.* to model the structures of the synaptic intermediates in serine integrase-mediated recombination and to hypothesize roles for the coiled-coil motif in recombination directionality[8].

We have now determined high-resolution cryoEM structures of the full-length ϕC31 integrase protein, assembled in complexes with its recombination sites. Integrase forms a tetrameric complex with attP and attB, but a locked/inactive complex with attL DNA. These

assemblies reveal a striking structural plasticity for ϕC31 integrase monomers on DNA; the integrase subunits adopt very different conformations on *attB*, *attP* and *attL* sites. Integrase binding of *attB* and *attP* sites in the minor groove is different, and the ZD binding site is closer to the centre of *attB* than it is in *attP*. The synaptic complex structures reveal that the recombining sites are held together by interactions of both the catalytic domains and the CC structural motifs. The parts of the CC motif closer to the DNA in the synaptic complex make distinct interactions with the other parts of their subunit when bound at the *attP* and the *attB* sites. The interface between the CC motifs of the two *attB*-bound integrase subunits with those bound to *attP* is formed by an asymmetric 'top to bottom' interaction of residues at the tip of the CC. Surprisingly, this interaction is inverted ('bottom to top') in our structure of the post-recombination *attL*-integrase dimer complex.

Results

1. ϕC31 integrase forms tetrameric synaptic complexes with attB and attP but a dimeric complex with attL

For DNA recombination to occur, two DNA *att* sites must come together in a synaptic complex. Each *att* site is bound by two integrase monomers. Pairs of these dimer-DNA complexes can then interact to form a synaptic complex containing four integrase subunits and two DNA *att* sites. Previous studies have shown that all the *att* sites bind integrase dimers with similar affinity, but *attP* and *attB* complexes interact selectively with each other to form the productive synaptic *attP*-*attB* complex in which recombination can take place, whereas all other pairs of *att* sites form less stable synaptic complexes that are catalytically inactive[1,9,10].

In preliminary analysis prior to our structural studies, we first assembled ϕC31 integrase separately with oligonucleotides containing *attB*, *attP*, or *attL* sites, and purified the respective complexes (which we refer to as Int-*attB*, Int-*attP*, and Int-*attL*) using size exclusion chromatography. These purified complexes were then analysed either alone or in combination by size exclusion chromatography coupled to multi-angle laser light scattering (SEC-MALLS) experiments to determine their solution stoichiometry. For these experiments we used a catalytically inactive integrase mutant carrying the S12A mutation[9].

First, we mixed equal amounts of the reconstituted Int-*attB* and Int-*attP* complexes and analysed the mixture by SEC-MALLS using a Superose 6 column. A single peak was observed with an elution volume of 13.8 ml. The measured M_W (weight-averaged molar mass) at the peak elution position was 327 kDa, matching well to the expected mass of 348 kDa for the full tetrameric synaptic complex (integrase tetramer = 278 kDa, *attB* = 33 kDa, *attP* = 37 kDa) (Figure 1C). Next, we examined the unmixed Int-*attB* and Int-*attP* samples. The Int-*attB* complex eluted as a single peak at 13.7 ml, with a measured M_W of 352 kDa. Likewise, the Int-*attP* complex eluted as single peak at 13.7 ml, with a measured M_W of 352 kDa. These results indicate a synaptic pairing of two integrase-bound *attP* (or *attB*) sites, which was not expected from earlier gel electrophoresis-based studies[9], suggesting that the interaction is strong enough to hold the identical sites together in SEC but not PAGE experimental conditions. Taken together, these data show that ϕC31 integrase can readily synapse *attB* and *attP* DNA sites.

Complexes of integrase with *attL* and *attR* interact in the presence of the RDF protein to form a synaptic complex that promotes 'excisive' *attL* × *attR* recombination[1]. However, the

attL and *attR* DNA sequences are very similar except at the central 2 bp, so are predicted to behave similarly prior to catalysis of strand exchange[11]. It was previously shown that ϕC31 *attL* and *attR* sites bind integrase to form dimer complexes, but do not interact to form synaptic complexes in the absence of the RDF gp3[9], whereas *attL-attL*,*attR-attR, and attL-attR* synaptic complexes can be formed in the presence of integrase and gp3[12]. In agreement with these results, an Int-*attL* complex eluted from a Superose 6 column as a single peak at 15.2 ml, with its measured M_W of 171 kDa matching well to the expected mass of 176 kDa for an integrase dimer bound to a single *attL* site (Figure 1C).

Both the N-terminal catalytic domain and the CC motif of serine integrases are known to be capable of multimerization and potentially contribute to synapsis of *att* sites[13-17]. To investigate the roles of these structural units in ϕC31 integrase-mediated synapsis, we generated two deletion mutants, removing either the N-terminal catalytic domain (ΔNTD, residues 1-140 removed) or the CC motif (ΔCC, residues 450-515 removed). SEC-MALLS analysis showed that the ΔNTD mutant readily formed synaptic tetramers with *attB* and *attP* (Figure 1D). Deleting the CC motif resulted in integrases that can still complex with *attB* and *attP* sites to form dimeric Int:DNA complexes, but are unable to form tetrameric assemblies (Figure 1E). Taken together, these results indicate that the CC motif of ϕC31 integrase is mainly responsible for synapsis, in agreement with previous mutational studies[16].

2. Structure of ϕC31 integrase bound to attB and attP sites

To further understand how ϕC31 integrase synapses *attB* and *attP* sites, we carried out single-particle cryogenic electron microscopy (cryoEM) experiments on the purified S12A mutant bound to a 53-bp *attB* site (*attB*53) and a 60-bp *attP* site (*attP*60) (Figure 2A).

Image processing in RELION 5.0[18] revealed that the particles of the tetrameric synaptic complex exhibited a high degree of structural heterogeneity, which hindered the reconstruction of high-resolution maps. Separating the particles into distinct structural classes showed that the main structural movements occurred between the *attB*-bound dimer and the *attP*-bound dimer within the full synaptic tetramer, thereby moving the *attB* and *attP* DNA duplexes relative to each other (Figure 2B). The motion appears to be continuous, with some particles adopting a more parallel arrangement for their two DNA duplexes, whilst in others, the duplexes form an angle of up to 30 degrees. To improve map resolution, we carried out masked 3D classifications and refinements[19] by applying a mask that included signals from either the *attB*- or the *attP*-bound dimer subcomplex (Figure 2B). The applied masks restricted image alignments and classifications to signals from within the mask, effectively ignoring the conformational heterogeneity in the unmasked area. This approach resulted in high-resolution maps for both the *attB*- and *attP*-bound dimer subcomplexes, which refined to 3.2 and 3.3 Å, respectively, allowing for the building and refinement of atomic models.

Atomic models were built into map regions with good local density, which for both reconstructions included the *att* site DNA and the DNA-interacting protein regions (Figure 2C, 2D, S3D, S4D). In the *attP*-bound dimer structure, residues 140-586 from each monomer are modelled, including the C-terminal portion of αE (residues 140-163), the entire recombinase domain (RD, residues 164-345), the zinc ribbon domain (ZD) except the residues 587-605 at the extreme C-terminus, the CC motif (residues 450-518), and the RD-ZD linker (residues 345-354). Densities for the two N-terminal catalytic domains of the dimer complexes are

missing, likely due to molecular motion. Similar protein regions are modelled for the two *attB*-bound monomers (residues 140-449 and 519-589); Except that the CC motifs in both attB-bound monomers (residues 450-518) could not be modelled due to the lack of density. Atomic models were built for the entire *attB*53 DNA, and all but 3 bp at each end of the *attP*60 DNA. For both reconstructions, the map resolution at the central TT-dinucleotide is sufficiently high to distinguish the left half-site from the right half-site, allowing DNA orientation to be assigned (Figure S7). The orientation of *attB*53 was also clear because there are 3 more base-pairs outside the *att* site at the left end than at the right end (Figure 2A). The density for these base-pairs are resolved the map.

Since both *attB* and *attP* DNA sequences are pseudo-palindromic, we numbered each base-pair according to its position from the centre, such that the central 2 bp are numbered 1 and -1 (Figure 2A). The overall binding of ϕC31 integrase to the two *attP* half-sites is similar to that seen in the crystal structure of the LI integrase bound to a single *attP* half-site (PDB ID code 4KIS[6]). The αE, RD, RD-ZD linker, and ZD of both proteins contact similar *attP* base-pair positions, which for ϕC31 integrase are positions -/+1-3, -/+4-15, -/+17-27, and -/+13-18, respectively (Figure 2E, 2F, S8A, S8B). On *attP*, the RD and ZD are bound on the same side of DNA, making mainly major groove contacts, whereas αE and the RD-ZD linker make predominantly minor groove contacts (Figure 2C). Notably, ϕC31 integrase makes numerous additional DNA contacts compared to the LI integrase, involving several of its longer intradomain loops within the RD and ZD. These include the αF-αG loop, the αG-αH loop, and the β8-β9 loop in the RD, which contact base-pairs -/+15, -/+6-9, and -/+7-8 respectively, and the β14-β15 loop in the ZD, which contacts base pairs -/+23-25.

It was previously hypothesised that the key difference between the complexes of serine integrases with *attP* and *attB* binding is in the positions of the ZD-DNA interactions[6,7]. Our structures confirm this hypothesis for ϕC31 integrase, showing that whereas the *attB*-bound and *attP*-bound RDs (RD^B and RD^P) contact base-pairs at similar distances from the centres of the sites,, the *attB*-bound ZD (ZD^B) contacts are 5 bp closer to the centre than the attP-ZD contacts(-/+12-22) (Figure 2E, 2F, S8). This difference moves ZD^B to the opposite side of the DNA double helix from its position in the *attP* complex (Figure 2C, 2D), which has functional implications for synapse assembly. The RD^B-ZD^B loop also adopts a different binding mode to its RD^P-ZD^P counterpart. The RD^P-ZD^P loop binds along the minor groove formed by base pairs -/+15-18, whilst the RD^B-ZD^B loop binds across the minor groove over the phosphate backbones of both DNA strands (Figure 2E), contacting base pairs -/+12-14, and thereby positions ZD^B closer to the centre of the site (Figure 2E, S8). Interestingly, instead of being bound by the RD^B-ZD^B loop, the minor groove at *attB* base-pairs -/+16-17 is occupied by residues 323-327 of the RD αF-αG loop (Figure 2E).

Structural comparisons of ZD-DNA binding of all four monomers reveal that a region in the β14-β15 loop, comprising residues 568-573, is bound differently in one of the monomers (at the *attB* left half-site). This loop is inserted into the minor groove at base pairs -/+23-25 of both *attP* half-sites,but remains outside the minor groove at the *attB* left half-site (base-pairs 18-20) (Figure 2F). This difference might be due to the very different DNA sequence at this position in the *attB* left half-site (CGG) from the other three half-sites (GTA, CTA, and GTA for *attP* left, *attP* right, and *attB* right half-sites respectively). A previous study has shown that the asymmetry of the *attB* sequence at base-pair -/+19 is crucial for ϕC31 integrase activity[20], suggesting that the alternative binding to the *attB* left half-site that we observe may have biological implications and warrant further investigation.

3. Structure of the full synaptic complex

To visualise how the full synaptic complex assembles, we employed an image processing approach involving iterative rounds of 3D classifications without masking sub-regions, resulting in a reconstruction of the complete Int-*attB*-*attP* tetrameric synaptic complex at a resolution of 7.5 Å (Figure 3A top, 3B, S2A). Regions with good map density include all the modelled regions in the *attB*- and *attP*-bound dimer structures, and some of the unmodelled regions, namely the CC motifs from *attB*-bound monomers (CC^B, residues 450-517), and the N-terminal ends of all four αE helices (residues 129-139), allowing us to build and refine atomic models for these regions (Figure 3B bottom). Briefly, model building involved first docking and flexibly fitting the two high-resolution *attB*- and *attP*-bound dimer structures using TEMPy-ReFF[21]. The two CC^B motifs were then docked using EM placement in CCP-EM[22], using the *attP*-bound CC (CC^P) structure as a starting model. Finally, the four αE helices were extended from D140 in the dimer structures to their N-terminal residue V129, using residues 129-150 from the crystal structure of the ϕC31 integrase catalytic domain as a starting model (PDB ID 4BQQ). Detailed model building and refinement procedures are described in Materials and Methods. Density for the remaining part of the catalytic domain (residues 1-128) is visible at the centre of the synaptic space adjacent to the four αE helices. However, significant molecular motion appears to occur in this region, as indicated by the increased noise level in the map, which hindered the docking of atomic models. As a result, these areas are represented as transparent surfaces in Figure 3B. Density for the three extra base-pairs at the left end of the *attB*53 site helped to determine the direction of *attB*53 DNA, but the map resolution is not sufficient to distinguish the direction of the *attP*60 DNA. Given that only the 'parallel' orientation is predicted to yield productive complexes[1], we have modelled *attB* and *attP* in the parallel configuration, but it remains unclear whether the antiparallel complex is also present.

Our structure of the synaptic tetramer shows that the *attB* and *attP* duplexes are brought into proximity by integrase, and are separated by a distance of approximately 50 Å (Figure 3B). The complex exhibits imperfect twofold rotational symmetry, with the symmetry axis passing through the central two base-pairs of both *attB* and *attP*. The two *att* duplexes are aligned in approximately the same plane, with the scissile phosphodiester moieties of all four DNA strands facing inward toward the centre of the complex. Both DNA duplexes are slightly bent, so that their central regions are 48 Å apart (measured between *attB* and *attP* scissile phosphates) and their ends are 77 Å apart. The bend angle of the *attB* duplex is greater than that of *attP*.

The *attP* and *attB* DNA duplexes are held together by several interactions between *attP*- and *attB*-bound monomers. These include the tetrameric core formed by the four N-terminal catalytic domains (NTDs) at the centre of the synaptic complex, and a pair of CC^B-CC^P dimers on either side of this tetrameric core (Figure 3B). The CC^B-CC^P dimer assembles in a directional 'top-to-bottom' arrangement similar to that seen in the crystal structure of the isolated CC motif from the LI integrase[13]. The interface involves residues 468-497 located at the tip of each CC, with the 'bottom' face of CC^P contacting the 'top' face of CC^B (Figure 3C). This interaction is stabilised by a negatively charged patch on αL and a positively charged patch on αM (Figure 3C). Notably, CC^B and CC^P adopt markedly different trajectories; CC^P is positioned close to the *attP* DNA, whereas CC^B extends away from the *attB* DNA (Figure 3B). This difference is due to distinct conformations of the connecting loops between the αK and αL helices (residues 447-450), which we call the CC hinge region. The difference rotates the

CC around αK by approximately 90°, and moves the tip region of the CC by 100 Å (Figure 3D). We call these two conformational states "open" and "closed" (for the *attB*- and *attP*-bound monomers respectively).

Previous structural studies have shown that the catalytic domains of serine recombinases form a tetramer within the synapse with 222 symmetry[23-28]. The αE helices play an important role in interactions within this tetramer. Crystal structures show that the angles made by the αE helices with each other can vary, likely represent different rotational states during subunit rotation, however, the structure of the pre-rotational stage remained elusive[25,29]. Our structure, of a pre-cleavage ϕC31 integrase-*attP*-*attB* synaptic complex, shows that the four αE helices pack in a similar way to that seen in small serine recombinases, using residues 129-140 (the three N-terminal helical turns of αE) (Figure 3F). However, there are several differences compared to existing structures of serine recombinase tetramers. The interface formed by ϕC31 integrase in our synaptic complex structure is smaller (3 helical turns) than that observed in other recombinases (5 turns in γδ resolvase and TP901 integrase tetramers; PDB ID code 1ZR4 and 3BVP respectively), and 6 turns in the Sin resolvase and Gin invertase tetramers (PDB ID code 3PKZ and 3UJ3 respectively). Also, the *att* sites in our structure are positioned further apart (48 Å between *attB* and *attP* site scissile phosphates)than in the γδ resolvase-DNA structure 1ZR4 (38.8 Å) (Figure 3F, 3G).

4. Molecular details at the CC-CC interface

Our structure of the full synaptic complex showed that synapsis is mediated by a pair of CC^B-CC^P dimers that associate through residues at the tip region of each CC. However, due to the low map resolution (7.5-Å) we were unable to confidently capture high-resolution details in our structural model, such as the involvement of specific amino acid side chains. To improve the map resolution, we employed a similar approach to that which generated the high-resolution dimer structures; we carried out masked classifications and refinements using a mask covering a single *attP*-bound monomer and the interacting CC^B (for details see Materials and Methods, and Figure S5). This approach resulted in a map with a resolution of 3.8 Å, providing more detailed insights into the binding interface (Figure 4A). We were able to build atomic models for the entire CC^P, and residues 465-509 in CC^B which contain the full interacting region. In our EM map, we observe strong density for the bulky side chain from Y475 on αL in CC^P, which packs against the top face of αM in CC^B (Figure 4A). The side chain of D477 located on the top face of CC^B can be seen interacting with the bottom face of αM in CC^P, a region rich in positively charged residues (Figure 4A).

To validate this interface, we generated two point mutants of the S12A integrase, Y475H and D477K. These two mutants had previously been identified by mutagenesis screens to be defective in recombination both in E. coli cells and *in vitro*[15]. Indeed, SEC-MALLS analysis showed that both S12A/Y475H and S12A/D477K mutants bound *att* sites as dimers, but were unable to make synaptic *attB* × *attP*, *attB* × *attB*, or *attP* × *attP* complexes (Figure 4E, F).

We reasoned that in the observed arrangement where the bottom face of CC^P contacts the top face of CC^B, the Y475H mutation in CC^P (on the bottom face) would disrupt the interface, whilst the Y475H mutation in CC^B might have no effect. To test this hypothesis, we purified the integrase S12A and S12A/Y475H complexes with *attB* and *attP*. Integrase S12A-*attB* and S12A-*attP* complexes both eluted as synaptic complexes, with elution volumes at

~13.5 ml, whereas integrase S12A/Y475H bound to both *attB* and *attP* to form dimer complexes with an elution volume around 15.5 ml (Figure 4H, I). When equimolar amounts (by *att* sites) of integrase S12A-a*ttB* complex and integrase S12A/Y475H-*attP* complex were mixed, we observed two peaks eluting at 13.5 ml and 15.5 ml, consistent with the original complexes, and suggesting that no new complexes were formed (Figure 4H). However, when the integrase S12A-*attP* complex was mixed with integrase S12A/Y475H-*attB* complex, only a single peak eluting at 13.5 ml was observed (consistent with synaptic tetramer) (Figure 4I). A peak at 15.5 ml (for the originally added S12A/Y475H-*attB* dimer complex) was absent, suggesting that this complex had associated with integrase S12A-*attP* dimer complexes formed by partial dissociation of the S12A-*attP* synaptic tetramers, to form an *attP*-*attB* synaptic complex. These results are consistent with the 'top to bottom' interaction of the CC-motifs observed in our cryoEM structure.

5. The CC motif hinge region is important for recombination directionality control

Interestingly, masked classification of the *attP*-bound integrase subunit yielded two distinct conformational states, which differed mainly in the relative positioning between the RD and ZD domains. We termed these State 1 and State 2, with State 1 being the conformation of the structure described above. State 1 refined to 3.8 Å and showed good density for the CC^B, whereas State 2 refined to 3.5 Å and showed particularly good local resolution for the hinge region that joins αK of the ZD and the first helix αL of the CC (Figure 4A, B top and middle).

Superposition of the two maps and models at the RD reveal that the ZD and CC^P in State 2 are shifted downwards by approximately 5 Å compared to their positions in State 1 (Figure 4C top and middle). The CC is thus shifted downwards by 2 helical turns in State 2, such that αI from the RD packs against αL of the CC motif at the first helical turn in State 2, but at the third helical turn in State 1 (Figure 4C bottom).

In our structure of the synaptic tetramer, we observe a short loop region (residues 448-451) that connects αK on the ZD to the first helix (αL) of the CC motif at its base. This region acts as a flexible hinge, adopting two distinct states for the *attP*- and *attB*-bound monomers within the synaptic tetramer. We call these states “open” and “closed” to reflect the different conformations of the hinge. As mentioned above, our *attP*-bound particles in State 2 resulted in a 3.5 Å resolution map with good side chain density at the *attP*-bound CC motif hinge region, and revealed interactions that stabilise the closed hinge. In particular, E449 (located within the hinge region) extends its side chain into the space between αK and αL, where it forms a salt bridge with the side chain from R457 (located on the second helical turn of αL). This interaction resembles a latch and likely stabilises the ‘closed’ conformation of the *attP*-bound CC motif. An E449R integrase mutant has been shown previously to promote *attL* x *attR* recombination in the absence of RDF, unlike wild-type integrase[15]. To examine this further, we carried out SEC-MALLS on the S12A/E449R integrase, and confirmed that it can readily form *attL* x *attL* synaptic complexes, whilst also retaining the ability to synapse *attB* x *attB*, *attP* x *attP*, and *attB* x *attP* (Figure 4G). This result highlights the importance of the CC motif hinge region in regulating synapsis and recombination directionality.

6. Structure of the post-integration attL-bound complex

To gain structural insights into the product of *attP* × *attB* recombination, we determined the cryoEM structure of the "post-recombination" complex of *attL* bound by an integrase S12A dimer at 6-Å resolution (Figure 5). The *attL* sequence used in our experiment consists of a 30-bp *attB* half-site and a 30-bp *attP* half-site . Our map showed good density throughout, for both the N-terminal catalytic domains, allowing us to build models for nearly all protein and DNA regions. Model building involved using the structures of *attB*- and *attP*-bound monomers from the *attB*- and *attP*-bound dimer structures, and the crystal structure of a ϕC31 NTD dimer (PDB ID code 4BQQ) as starting models.

It has been hypothesised that the CC motifs can form 'intrasite' dimers within *attL* (and *attR*), potentially preventing them from participating in synaptic interactions and thereby ensuring the *attL* and *attR* sites cannot recombine in the absence of RDF[6,7,13]. Our structure of the *attL*-bound complex confirms the presence of an intrasite CC dimer. Unlike the monomers bound to *attB* and *attP*, the ZD regions of the two *attL*-bound monomers are positioned on the same side of the DNA, with their CC motifs extended towards the centre of the complex where they meet (Figure 5A). The binding interface consists of the same residues as observed in the CC^B-CC^P synaptic dimer (Figure S1C), but the binding direction is reversed, to a bottom-to-top arrangement; that is,the bottom face of the CC motif bound at the *attB*-derived half-site of *attL* now interacts with the top face of the CC motif at the *attP*-derived half-site (Figure 5B, S1C).

This intrasite CC motif interaction is accompanied by changes in the trajectories of both CC motifs compared to their *attP*-bound and *attB*-bound counterparts in the synaptic complex, along with corresponding conformational changes in their respective CC-hinge regions. Notably, the trajectory of CC^P in the *attL*-bound complex resembles the 'closed' state observed in the synaptic CC^P, but is rotated approximately 80° around its long axis, reorienting its 'top' face toward the DNA and CC^B. An even more pronounced change is seen in *attL*-bound CC^B, which transitions from the 'open' conformation seen in the synaptic complex to one resembling the 'closed' conformation characteristic of the two CC^Ps (Figure S1B, E, F).

A mutation within the CC-hinge region, E449K, has been previously shown to disrupt directionality control and promote *attL* x *attR* recombination in the absence of RDF[15]. We observed that in the 'closed' conformations of the hinge region, the Cα of this residue is positioned such that its side chain can make electrostatic interactions with the side chain of R457 located on αL (10.3, 11.8, 12.3 Å), but the two residues are positioned too far apart to interact in the 'open' state (Cα of E449 and R457 in the synaptic complex *attB*-bound monomer are 13.8 Å apart). Thus, it is highly likely the E449-R457 interaction only contributes to the stability of the 'closed' conformation, which is adopted by both CC motifs in the *attL*-bound monomers when forming the intrasite CC dimer. Another feature of the 'closed' CC trajectory is that, when bound to an *attP* half-site, αL of the CC motif is placed against and interacts with the β9-αI loop of the RD from the same subunit. This intra-monomer CC-RD interface is mediated by an acidic patch on αL formed by E452, E456, and E463, packed against a basic patch on the β9-αI loop involving residues R311, R314, and R320 (Figure 5C). Remarkably, all three acidic residues involved in this CC-RD interface have previously been identified to be important for directionality control, with charge-swapping mutations E452K, E456K, and E463K all shown to promote *attL* x *attR* recombination in the absence of RDF[15]. Together, these observations highlight the level of regulation of recombination directionality and underscore the important role played by the trajectory of the CC element.

Our structure of the *attL* complex also reveals the arrangement of the two NTDs, including the cores of the catalytic domains (residues 1-128) and their αE helices (residues 129-163). In this complex, αE of the monomer on the *attP*-derived half-site forms a continuous helix, similar to its *attP*-bound counterpart in the synaptic complex (Figure S1A, S1G). In contrast, the αE helix of the monomer on the *attB*-derived half-site is kinked in its central region, around resides R138-K152 (Figure 4D, S1B, S1G). The two NTDs form a dimer, with the dimer interface comprised of αD and the N-terminal portion of αE from each monomer (Figure 5E). Structural comparisons show that this NTD dimer interface in the *attL*-bound complex is identical to one observed previously in the 2.15 Å crystal structure of the isolated NTD (PDB ID 4BQQ) (Figure 5E). The higher-resolution crystal structure provides additional detail, revealing that the side chain of I136 on αE packs against αD of the same monomer, and that the dimer interface is stabilised by hydrophobic interactions involving I103 on αD and I133 on αE from each monomer (Figure 5F).

The NTD dimer observed in our structure appears to be in an inactive state, as indicated by the considerable distances between the S12A Cα atoms of each monomer and their respective scissile phosphates, which is 20 Å and 35 Å for the monomers bound at the *attP*- and *attB*-derived half-sites respectively (Figure 5G). This arrangement is consistent with the observation that *attL*-bound complexes cannot undergo recombination in the absence of RDF. Additionally, structural comparisons reveal that, in contrast to the inactive dimer on attL, where I133 contacts the monomer bound to the adjacent half-site, the side chain of I133 in the synaptic complex engages in a synaptic interaction, contacting αE of the monomer that is bound across the synaptic space to the partner *att* site (Figure 5H). These findings suggest that substantial structural rearrangement within the NTD is required upon synapsis.

7. Catalytic domains undergo conformational change upon synapsis

In our map of the synaptic complex, we observed well-defined density that enabled accurate modelling of the tetrameric assembly formed by the four αE helices. In contrast, the region corresponding to the remainder of the catalytic domain exhibited weaker and more diffuse density, characteristic of disordered regions, which prevented structural modelling (Figure 3B). This apparent disorder may result from molecular motion causing the NTDs to move relative to the rest of the complex. To investigate this motion and improve map resolution, we carried out multibody refinement analysis[30] in RELION, a method designed to analyse molecular motion by treating the full complex as several distinct independently moving bodies, allowing for higher resolution reconstructions and the analysis of their relative motions. We divided the complex into two masked bodies: one encompassing the central NTD tetramer and another covering the remaining parts of the complex (Figure 6A). The resulting reconstruction showed substantial map improvement for the NTD tetramer; the map density became stronger and more continuous (Figure 6B). Analysis of the relative motions between the two bodies revealed that 67% of the variance is explained by the first three principal motions (Figure 6C), all of which involved twisting motions. In principal motion 1, the NTDs rotate around the central twofold symmetry axis of the complex (Figure 6D, Video S1). In principal motions 2 and 3, the NTDs rotate around axes perpendicular to that in principal motion 1 (Video S2, Video S3, Figure S9). However, although this analysis improved the map quality of the NTD tetramer, the interpretability remained insufficient for the rigid body fitting of NTD monomers.

To build a model of the fully assembled NTD tetramer, we used the structure of an NTD monomer in our *attL*-bound structure (the monomer bound to the *attP*-derived half-site),

and aligned its αE helix to those in the synaptic complex. The resulting model showed substantial steric clashing between the core regions of the NTDs (residues 1-128) from a pair of rotating dimers, suggesting that conformational changes within the NTD are necessary to avoid these clashes (Figure 6E). To explore this potential conformational change, we used AlphaFold3[31] to model the assembled structure of four ϕC31 integrase NTD monomers (residues 1-163). The resulting structural model predicted a tetrameric assembly similar to those previously observed in serine recombinase crystal structures (Figure 6F). Notably, the αE tetrameric arrangement in this predicted model closely matched that observed in our structure of the synaptic complex. Structural comparisons between individual monomers revealed that the monomers in the AlphaFold3-predicted model adopt a different trajectory for their NTD core region, where it is rotated approximately 70° around αE compared to that of the monomers in the *attL*-bound structure (Figure 6G). This movement is accompanied by a conformational change at the αE-β5 loop that connects αE to the NTD core region (Figure 6G). Conformational changes at the same loop have previously been shown to switch the NTD between an inactive and an active state in the small serine recombinases γδ and Sin resolvase[23-25,27]. Indeed, structural comparison shows that the AlphaFold3-predicted ϕC31 integrase monomer adopts a similar NTD core trajectory to those in the activated γδ and Sin resolvases (Figure 6G). To test whether the conformation of the AlphaFold3-predicted monomer is permitted in our synaptic structure, we superimposed both structures at αE residues 129-145. The resulting model no longer shows any clashing between the pair of rotating dimers (Figure 6H). Furthermore, the active site serines of both monomers have been moved adjacent to their respective scissile phosphates, with the Cα to phosphorus atom distance being 4.6 Å and 7.3 Å respectively, suggesting that the modelled structure could represent the architecture of a 'ready-for-cleavage' state.

Discussion

The site-specific recombination system of bacteriophage ϕC31 was the first to be shown to use a serine recombinase (integrase) to promote DNA strand exchange. The 'large serine integrase' group of recombinases, of which ϕC31 integrase is the founder member, now numbers many thousands of proteins, mostly identified from 'database mining'[32,33]. The efficiency and remarkable unidirectionality of recombination by the large serine integrases (see Introduction) has led to great interest in their application as versatile tools for DNA manipulations, including DNA fragment assembly, creation of genetic switches, circuits and logic systems, genome engineering, the permanent genomic recording of cellular events[34]. Gene therapy, and animal transgenesis[35,36]. A structure-based understanding of the mechanisms of these enzymes is essential for full development of these potential applications.

To date, X-ray crystal structures of fragments of large serine integrases have been published, including the structures of an LI integrase subunit bound to a DNA half-site[6], and these have been used to build models of reaction intermediates (refs). However, there are no published structures of intermediates including the full DNA recombination (*att*) sites and full-length integrase proteins. Here, we used single-particle cryo-EM to solve structures of a synaptic complex comprising a ϕC31 integrase tetramer bound to *attB* and *attP* DNA sites,

and an autoinhibited complex of an integrase dimer bound to an *attL* site. The integrase in our structures is the full-length wild-type protein except for a mutation of the catalytic serine residue (S12A) rendering integrase inactive for catalysis of DNA strand cleavage and rejoining. As a result, our *attP-attB* structure is also the first of a synaptic intermediate of any serine recombinase in a pre-cleavage state, immediately prior to catalysis of strand exchange.

Our structure of integrase bound to *attP* and *attB* in a synaptic complex (Figure 2) reveals an integrase tetramer formed by interactions of the NTDs at the centre of the complex, with the *attP* and *attB* DNA duplexes on the outside of the tetramer, similar to solved structures of γδ resolvase synaptic intermediates and in accord with published models of the integrase synaptic complex (refs). Each *att* site is bound roughly symmetrically by two integrase subunits, with their NTDs at the site centre, flanked by the RDs, and with the ZDs at the ends of the sites (Figure 2, 3). Further bridges between the two sites are established by interactions of pairs of CC motifs. The CC motif shows shows significant conformational plasticity; *attB*-bound CC motifs adopt an 'open' conformation (Figure 2D-F), allowing bridging interactions with the CC motifs of the integrase subunits bound to *attP*.In contrast, the CC motifs of the *attP*-bound subunits adopt a 'closed' conformation, with the basal parts of the CC motif contacting the body of the ZD. The paired CC motifs form two 'braces', one on either side of the NTD tetramer. The electron density of the NTDs was insufficient for us to model them directly from the cryoEM data. but we can estimate their positions using existing crystallographic data of the small serine recombinase γδ resolvase (1ZR4 [37]) and a partial structure of ϕC31 integrase protein (4BQQ), in which the NTD is well resolved (Figure 3). This modelling indicates that surfaces of the NTDs are quite close to the paired CC motif braces. It is thus feasible that contacts of the CC motifs with the NTDs might have a regulatory role, analogous to regulatory mechanisms observed in other serine recombinase systems[25,38,39].

Our data give good resolution of the full lengths of the E-helices at their N-termini, showing that they interact with each other in a manner quite similar to that seen in the γδ resolvase synaptic complex structure 1ZR4, forming a flat interface between the 'rotating pairs' of subunits that would participate in a subunit rotation mechanism of DNA strand exchange. However, the extent of overlap between the E-helices bound at *attP* and those bound at *attB* is only about two helical turns (residues x - y), whereas the overlap is greater in the 1ZR4 structure (about 5 turns). Furthermore, the two *att* sites are significantly further apart in our structure (x A between scissile phosphodiesters [or some other reference point in the DNA?]), compared to 1ZR4 (y A). These differences lead us to speculate that our structure shows an early stage on the reaction pathway, where the NTDs have made initial contacts but must undergo further conformational changes prior to catalysis, including more extensive interdigitation of the E-helices which would bring the *att* site DNA molecules closer together and into the recombinase active sites as in the 1ZR4 structure.

The only parts of the subunits of the integrase synaptic tetramer that cross the putative rotation interface are the β-loops at the N-termini of the RDs (Figure 3). These loops would clash during rotation unless they changed conformation. We thus speculate that completion of 180° rotation, bringing the cleaved DNA half-sites into the recombinant arrangement, might be coupled to movement of these loops and other conformational changes of integrase that promote ultimate conversion to the inactive recombinant dimer-bound *attL* and *attR* products (see below). We note however that this loop in ϕC31 integrase is much longer than it is in other large serine integrases[7].

Our structure of an integrase dimer bound to *attL*, one of the product sites from *attP* × *attB* recombination (Figure 5) , reveals a 'double lock' preventing any further recombination of this site. Firstly, as predicted by the modelling of Rutherford *et al.*[6], the CC motifs of the two integrase subunits on *attL* reach towards and interact with each other, making them unavailable for synaptic interactions with another integrase-bound *att* site. The CC motif bound at the *attP*-derived half-site of *attL* is in a similar conformation to that seen in the *attP*-bound subunits of the synaptic complex, but the hinge region conformation of the subunit at the *attB*-derived half-site bound subunit is substantially different from that in the synaptic complex (Figure 4), reorienting the CC motif so that it crosses the centre of the site to contact the other CC motif bound at the '*attP*' half-site. The interacting pair of CC motifs is positioned asymmetrically with respect to the centre of *attL*, on the same side as the E-helix of the integrase subunit at the *attP*-derived half-site but further away from the DNA. Secondly, although the NTDs interact to form a dimer, they are positioned asymmetrically with respect to the *attL* DNA and the active site residues are far from the scissile phosphodiesters, so a very substantial conformational change would be required before any catalysis of strand cleavage and rejoining could occur. Notably, the NTD of the subunit bound to the '*attP*' half-site of *attL* is in a similar conformation to the NTDs bound at *attP* in the *attP*-*attB* synaptic complex, whereas the NTD at the '*attB*' half-site is substantially reconfigured, most strikingly in its E-helix which is kinked in two places. Binding of RDF (gp3) at the base of the *attL*-bound integrase CC motifs is proposed to promote dissociation and re-orientation of the CC motifs to make them available for inter-site (synaptic) pairing (refs.) . Our structure suggests that the RDF-promoted conformational changes will be mainly in the subunit bound at the *attB*-derived half-site. RDF binding might also be needed to induce the NTDs of the *attL* complex to adopt a 'recombination-ready' conformation in a synaptic tetramer with the NTDs from a partner site (*attR* in the natural system). The E449K mutation, which allows *attL* x *attL* synapsis in the absence of RDF as we observed in our SEC-MALLS experiments, might facilitate adoption of an extended position of the CC motif by mutual positive charge repulsion with basic residues in the adjacent K-helix. Full understanding of the mechanism by which RDF binding switches the directionality of integrase-mediated recombination will require structures of complexes including the RDF (gp3).

Methods

Protein expression and purification

All ϕC31 integrase proteins used in this study were expressed in *E. coli* using synthetic codon-optimised coding sequences that were cloned into a pET28a expression vector. All constructs had an N-terminal His_6 tag to allow affinity purification using Nickel-NTA resin columns. Constructs were expressed in *E. coli* strain BL21 LOBSTR (DE3) grown in LB media. Single colonies were used to inoculate an overnight pre-culture grown at 37 °C, shaken at 180 rpm, which was subsequently used to seed main cultures at a 1:200 ratio. Seeded cultures were then grown at 37 °C, shaken at 180 rpm, and protein expression was induced by the addition of 0.5 mM IPTG to log phase cultures prior to subsequent growth overnight at 18 °C.

Cells were resuspended 1:5 (w/v) in lysis buffer (20 mM HEPES, pH 7.4, 1 M NaCl, 2 mM $MgCl_2$, 10 mM imidazole, 10% (v/v) glycerol, 0.1% (v/v) Tween 20, 0.5 mM TCEP,1x cOmplete mini protease inhibitor tablet, 1x benzonase (Sigma), then lysed by sonication, and centrifuged (Sorvall F21-8x50y, 20 000 rpm, 30 min). Clarified lysate was then passed through a 1 ml HisTrap NiNTA column to capture His-tagged protein. The column was then washed extensively with 30 column volumes (CV) of IMAC buffer A (20 mM HEPES, pH 7.4, 1 M NaCl, 2 mM $MgCl_2$, 10 mM imidazole, 0.5 mM TCEP), before being eluted using an imidazole gradient of 10-300 mM over 10 CV. Fractions were then analysed by SDS-PAGE. Fractions containing pure integrase were concentrated in a Amicon® 50 kDa MWCO spin concentrator, then injected and further purified by size-exclusion chromatography using a Superose 6 10/300 column equilibrated in 20 mM HEPES, 150 mM NaCl, 2 mM MgCl2 and 0.1 mM TCEP (pH 7.4). Fractions containing pure integrase protein were then pooled and concentrated to 10-30 mg/ml, flash frozen and stored at -80 ºC.

Assembly and purification of protein-DNA complexes

Integrase-DNA complexes were assembled fresh before SEC-MALLS and cryo-EM experiments. To assemble complexes, stoichiometric ratios of integrase and *att*-site duplexes were mixed and incubated at 30 °C for 15 minutes to allow complex formation. The mixture was spun at 15 000 rpm at 4 °C for 10 minutes to remove precipitate, then injected onto a Superose 6 column equilibrated in 20 mM HEPES, 150 mM NaCl, 2 mM $MgCl_2$ and 0.1 mM TCEP (pH 7.4) to purify the protein-DNA complex.

SEC-MALLS

SEC-MALLS was used to determine the solution molecular weight of integrase-DNA complexes. Samples (100 µL) were applied to a Superose 6 10/300GL column mounted on an AktaPure 25 (Cytiva), equilibrated in 20 mM HEPES, 150 mM NaCl, 2 mM $MgCl_2$ and 0.1 mM TCEP (pH 7.4) at a flow rate of 0.5 mL/min. The scattered light intensity and protein concentration of the column eluate were recorded using a Wyatt Dawn HELEOS-II 18-angle light scattering detector and a Wyatt Optilab T-rEX differential refractive index detector (ΔRI) (dn/dc = 0.180), respectively. The weight-averaged molecular mass of material contained in chromatographic peaks was determined from the combined data from both detectors using the software ASTRA.

Cryo-EM Sample preparation

Cryo-EM grids were prepared using purified integrase carrying the inactivating S12A mutation. Prior to grid making, stoichiometric ratios of integrase and *att* site DNA duplexes were mixed, and the resulting complex was purified using a Superose 6 10/300 column equilibrated in 20 mM HEPES, 150 mM NaCl, 2 mM $MgCl_2$ and 0.1 mM TCEP (pH 7.4). 4 µl of purified integrase-*attB*-*attP* complex (0.25 mg/ml) or integrase-*attL* complex (0.3 mg/ml) was applied onto a freshly glow discharged Quantifoil R 1.2/1.3 copper 300 mesh grid. The sample was then blotted for 2 s at a blot force of 6 and vitrified by plunging into liquid ethane using a Vitrobot Mark IV (FEI Company) operated at 4 °C and 100% humidity.

Cryo-EM data collection

An integrase-*attB*-*attP* synaptic complex dataset was collected on a FEI Titan Krios transmission electron microscope operated at 300 keV, equipped with a Falcon IV direct electron detector with a Selectris X energy filter (ThermoFisher Scientific). The dataset for the integrase-*attL* complex was collected on a FEI Glacios transmission electron microscope operated at 200 keV, equipped with a Falcon IV direct electron detector (ThermoFisher Scientific).

Automated data acquisition was carried out using EPU software (FEI) at a nominal magnification of 120 000×, resulting in calibrated pixel size of 1.171 Å and 1.192 Å for the Krios- and Glacios-collected dataset respectively, with a defocus range between -0.7 to -2.3 micron. Multiframe movie stacks were collected, totalling 10,235 for the integrase-*attB*-*attP* dataset and 15,814 for the integrase-*attL* datataset. Movies from both datasets contained 60 frames at a total dose of ~60 e^-/Å². Detailed data collection optic settings and image processing parameters are summarised in Table 1.

Cryo-EM image processing

All image processing was done in RELION[40,41] v.5.0. Motion correction was carried out using RELION's own implementation and the resulting micrographs were contrast transfer function (CTF) corrected using CTFFIND-4.1[42].

Integrase-attB-attP data set

An ab-initio model of the full integrase-*attB*-*attP* synaptic complex was generated with auto-picked particles using RELION's Laplacian-of-Gaussian (LoG) filter based procedure. The model was then used as a 3D reference to autopick from all micrographs, yielding 1,921k particles. Subsequent processing revealed a high degree of conformational heterogeneity among the particles, primarily due to a partial dissociation of the synaptic CC-CC interactions, causing the *attB*- and *attP*-bound dimers to move in relation to one another. As a result, five separate reconstructions were generated from five distinct particle sets: full tetrameric synaptic complex (resolution 7.5 Å), *attB*-bound dimer subcomplex (3.2 Å), *attP*-bound dimer subcomplex (3.3 Å), *attP*-bound monomer, CC motif-up conformation (3.8 Å), and *attP*-bound monomer CC motif-down conformation (3.5 Å). All reported resolutions were calculated using the gold standard[43] Fourier shell correlation (FSC) 0.143 criterion.

For the reconstruction of the full synaptic complex, the 1,921k autopicked particles were classified into four 3D classes, carried out with a global mask covering the entire tetrameric synaptic complex. The best resulting class contained 725k particles, and was

further classified into five classes. These five classes showed substantial movement of the two *attB* and *attP* DNA in relation to one another. In one class, the two DNA arrange in a more parallel manner, whereas in other classes they are form an angle of up to 35°. It appears the 'parallel' class represents a fully assembled complex, whereas the others are partially dissociated, because the latter all show density for only one of the two CC-CC synaptic dimer pairs. Particles from the 'parallel' class (138k particles) were selected and were subjected to two more rounds of 3D classification, with each round selecting the most 'parallel' class. The final particle class contained 11k particles, and resulted in a 7.5 Å reconstruction. This map was used for model building of the full synaptic complex.

The 7.5-Å map was further analysed using Multibody refinement[40,41] to characterise motion of the NTD tetramer. The map was divided into two bodies using two masks: one mask covered the region of the NTD tetramer, and another mask that covered the rest of the complex. The resulting reconstructions of the two bodies revealed significant improvement in the map quality for the NTD tetramer. Principal motion analysis showed the first three eigenvectors explain 67% of variance in the data, with 27% for component 1, 22% for component 2 and 18% for component 3.

To generate high-resolution reconstructions of the subregions, we carried out focused refinement by using local masks to restrict image alignments to individual subregions, allowing us to effectively ignore any conformational heterogeneity outside the masked regions. First, we re-picked particles from all micrographs, using the 7.5-Å reconstruction of the full synaptic tetramer as a 3D reference, resulting in 6,091k total particles picked. One round of 2D classification was then carried out to remove junk particles, resulting in 2,633k good particles selected. These were then 3D classified into 3 classes, and two of the resulting classes, totalling 1,850k particles, showed good density were selected for further processing. The 1,850k selected particles were subjected 3D refinement and produced a 6.3-Å reconstruction. Hereon, local masks were applied to generate high-resolution maps of individual subregions. Multiple rounds of classification and signal subtraction masking each *attP*- and *attB*-bound dimer subcomplexes resulted in particle classes of 275k and 404k respectively, and refined to 3.2 Å and 3.3 Å each. Atomic models were built into these two maps. To generate high-resolution reconstruction of the CC-CC interface, the 1,850k selected particles (from the 6.3-Å map) were subjected to one round of 3D classification masking the *attB*-bound dimer, followed by a second round masking the *attP*-bound dimer, resulting in 226k good particles that showed good CC-CC density in the reconstruction. A local mask covering a single *attP*-bound monomer and the CC from its partner *attB* was then applied, yielding a class of 58k particles that showed strong density for the *attB* CC. This produced a 3.8 Å reconstruction, and is used for model building of the *attP*-bound monomer bound to *attB* CC.

Integrase-attL complex data set

For the integrase-*attL* complex data set, RELION LoG-autopicked particles were used to generate an ab-initio model which was then used as a 3D reference to autopick from all micrographs, resulting in 9,473k picked particles. Particles were then split into 10 subsets, and subjected to three rounds of 3D classification. We observed disorder in the N-terminal catalytic domain from both monomers, and selected particle classes that exhibited better density for this region. This resulted in a class of 219k particles, refined to 4.7 Å resolution. To further improve the resolution at the N-terminal domains, we carried out signal subtraction masking the N-terminal domains and their surrounding regions, and performed 3D classification without angular search. This resulted in a class of 13.7k particles that showed significant improvement

in the density, revealing secondary structural elements that allowed confident fitting of a model into both monomer density.

All maps used for model building were post processed in RELION to obtain resolution estimates carried out using the gold standard FCS procedure[43]. Local resolution estimates were calculated in RELION and displayed onto the final maps using ChimeraX[44,45]. Data processing workflows are summarised in Figures S2, S3, S4, S5 and S6.

Model building and refinement

Atomic models were first built into the high-resolution focused refined maps of the *attB*- and *attP*-bound subregions (map resolution 3.2 Å and 3.3 Å, respectively). Alphafold3 (AF3)[46] was used to generate a model each for the *attP*- and *attB*-bound dimer. The AF3 models showed good map agreement for the individual domain structures of RD and ZD. DNA binding positions for each of the *attP*-bound monomers are also in good agreement with map density, but for the *attB*-bound monomers, ZD-DNA binding position is incorrectly modelled, and was adjusted manually in ChimeraX[45]. The DNA trajectories for both *attB* and *attP* were also manually adjusted in ChimeraX[45] and in Coot[47]. Model regions that lacked density were removed, these were the NTD residues 1-139 and residues 587-605 at the C-terminus, for all four monomers. Residues 450-518 (CC motif) were removed for both the *attB*-bound monomers, and base-pairs -/+28-30 for *attP*60 DNA. Iterative manual refinement in Coot followed by automated refinement using Refmac-Servalcat and Phenix[48] were carried out. External restraints from the AF3 model generated using ProSMART and LibG[49] were used in refinement. Many of these computations were run using the CCP-EM v2 software interface (unpublished). The final refined models are used as starting models in the model building for the remaining maps.

To model the 3.8-Å map of the region consisting a single *attP*-bound monomer interacting with its partner CC^B, a single *P'*-bound monomer from the refined structure of the *attP*-bound dimer was first docked into the map. CC^B residues 465-509 was then docked into the density, using the same residues in CC^P as a starting model. The model was then manually refined in Coot, followed by automated refinement in Servalcat.

To model the fully assembled synaptic complex, the high-resolution refined models for the *attB*- and *attP*-bound subregions are first docked and flexibly fitted into the 7.5-Å map using TEMPy-ReFF[21]. The map density revealed additional protein regions that are absent from the high-resolution models; these are residues 129-139 (N-terminal portion of αE helices) from all four monomers, and 450-518 (the CC motif) from the two *attB*-bound monomers. Models for the two *attB*-bound CC were docked using EM placement in the CCP-EM[22] package, using the structure of the *attP*-bound CC as a starting model. The four αE helices were extended from residue D140 to their N-terminal residue V129, using the αE region (residues 129-150) in the ΦC31 integrase crystal structure 4BQQ as a starting model. Automatic model refinement was then performed using Refmac-Servalcat, with external restraints from the AF3 model generated using ProSMART and LibG[49]

To model the *attL*-bound complex, the *B*- and *P'*-bound subregion are docked into the 6.0-Å map in ChimeraX[45], and manually adjusted in Coot with ProSMART and LibG model restraints. The missing regions, residues 1-128 from both NTDs, were docked into the density using the crystal structure 4BQQ, which shared the nearly identical NTD dimer structure. The completed model was refined in Phenix using inbuild protein and nucleic acid restraints.

Acknowledgements

We acknowledge support from BBSRC grant BB/W017571/1 to LS, MWS and SDC and of a College of Medicine Veterinary Medicine and Life Sciences PhD studentship to LS. Preliminary data were collected at the Scottish Centre for Macromolecular Imaging (SCMI), University of Glasgow, funded on grants MC_PC_17135 (MRC) and H17007 (SFC). We acknowledge Diamond Light Source for access and support of the cryo-EM facilities at the UK national electron Bio-Imaging Centre (eBIC), proposal BI37630. We acknowledge help for preliminary cryo-EM data collection from Dr James Streetley (SCMI) and Dr Michal Klosowski (Research Complex at Harwell), and assistance for cryo-EM data collection at eBIC from Dr Parijat Majumder.

References

1 Smith, M. C. M. Phage-encoded Serine Integrases and Other Large Serine Recombinases. *Microbiol Spectr* **3**, doi:10.1128/microbiolspec.MDNA3-0059-2014 (2015).

2 Stark, W. M. The Serine Recombinases. *Microbiol Spectr* **2**, doi:10.1128/microbiolspec.MDNA3-0046-2014 (2014).

3 Stark, W. M. Cutting out the phiC31 prophage. *Mol Microbiol* **80**, 1417-1419, doi:10.1111/j.1365-2958.2011.07699.x (2011).

4 Olorunniji, F. J. *et al.* Gated rotation mechanism of site-specific recombination by varphiC31 integrase. *Proc Natl Acad Sci U S A* **109**, 19661-19666, doi:10.1073/pnas.1210964109 (2012).

5 Bai, H. *et al.* Single-molecule analysis reveals the molecular bearing mechanism of DNA strand exchange by a serine recombinase. *Proc Natl Acad Sci U S A* **108**, 7419-7424, doi:10.1073/pnas.1018436108 (2011).

6 Rutherford, K., Yuan, P., Perry, K., Sharp, R. & Van Duyne, G. D. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res* **41**, 8341-8356, doi:10.1093/nar/gkt580 (2013).

7 Van Duyne, G. D. & Rutherford, K. Large serine recombinase domain structure and attachment site binding. *Crit Rev Biochem Mol Biol* **48**, 476-491, doi:10.3109/10409238.2013.831807 (2013).

8 Rutherford, K. & Van Duyne, G. D. The ins and outs of serine integrase site-specific recombination. *Curr Opin Struct Biol* **24**, 125-131, doi:10.1016/j.sbi.2014.01.003 (2014).

9 Smith, M. C. *et al.* Synapsis and DNA cleavage in phiC31 integrase-mediated site-specific recombination. *Nucleic Acids Res* **32**, 2607-2617, doi:10.1093/nar/gkh538 (2004).

10 Li, H., Sharp, R., Rutherford, K., Gupta, K. & Van Duyne, G. D. Serine Integrase attP Binding and Specificity. *J Mol Biol* **430**, 4401-4418, doi:10.1016/j.jmb.2018.09.007 (2018).

11 Ghosh, P., Bibb, L. A. & Hatfull, G. F. Two-step site selection for serine-integrase-mediated excision: DNA-directed integrase conformation and central dinucleotide proofreading. *Proc Natl Acad Sci U S A* **105**, 3238-3243, doi:10.1073/pnas.0711649105 (2008).

12 Khaleel, T., Younger, E., McEwan, A. R., Varghese, A. S. & Smith, M. C. A phage protein that binds phiC31 integrase to switch its directionality. *Mol Microbiol* **80**, 1450-1463, doi:10.1111/j.1365-2958.2011.07696.x (2011).

13 Gupta, K., Sharp, R., Yuan, J. B., Li, H. & Van Duyne, G. D. Coiled-coil interactions mediate serine integrase directionality. *Nucleic Acids Res* **45**, 7339-7353, doi:10.1093/nar/gkx474 (2017).

14 Rowley, P. A. & Smith, M. C. Role of the N-terminal domain of phiC31 integrase in attB-attP synapsis. *J Bacteriol* **190**, 6918-6921, doi:10.1128/JB.00612-08 (2008).

15 Rowley, P. A., Smith, M. C., Younger, E. & Smith, M. C. A motif in the C-terminal domain of phiC31 integrase controls the directionality of recombination. *Nucleic Acids Res* **36**, 3879-3891, doi:10.1093/nar/gkn269 (2008).

16 McEwan, A. R., Rowley, P. A. & Smith, M. C. DNA binding and synapsis by the large C-terminal domain of phiC31 integrase. *Nucleic Acids Res* **37**, 4764-4773, doi:10.1093/nar/gkp485 (2009).

17 Mandali, S. & Johnson, R. C. Control of the Serine Integrase Reaction: Roles of the Coiled-Coil and Helix E Regions in DNA Site Synapsis and Recombination. *J Bacteriol* **203**, e0070320, doi:10.1128/jb.00703-20 (2021).

18 Kimanius, D. *et al.* Data-driven regularization lowers the size barrier of cryo-EM structure determination. *Nature Methods* **21**, 1216-1221, doi:10.1038/s41592-024-02304-8 (2024).

19 Scheres, S. H. Processing of Structurally Heterogeneous Cryo-EM Data in RELION. *Methods Enzymol* **579**, 125-157, doi:10.1016/bs.mie.2016.04.012 (2016).

20 Gupta, M., Till, R. & Smith, M. C. Sequences in attB that affect the ability of phiC31 integrase to synapse and to activate DNA cleavage. *Nucleic Acids Res* **35**, 3407-3419, doi:10.1093/nar/gkm206 (2007).
21 Beton, J. G., Mulvaney, T., Cragnolini, T. & Topf, M. Cryo-EM structure and B-factor refinement with ensemble representation. *Nature communications* **15**, 444, doi:10.1038/s41467-023-44593-1 (2024).
22 Burnley, T., Palmer, C. M. & Winn, M. Recent developments in the CCP-EM software suite. *Acta Crystallogr D Struct Biol* **73**, 469-477, doi:10.1107/s2059798317007859 (2017).
23 Li, W. *et al.* Structure of a Synaptic γδ Resolvase Tetramer Covalently Linked to Two Cleaved DNAs. *Science (New York, N.Y.)* **309**, 1210-1215, doi:doi:10.1126/science.1112064 (2005).
24 Kamtekar, S. *et al.* Implications of structures of synaptic tetramers of gamma delta resolvase for the mechanism of recombination. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 10642-10647, doi:10.1073/pnas.0604062103 (2006).
25 Rice, P. A. Serine Resolvases. *Microbiol Spectr* **3**, Mdna3-0045-2014, doi:10.1128/microbiolspec.MDNA3-0045-2014 (2015).
26 Yuan, P., Gupta, K. & Van Duyne, G. D. Tetrameric structure of a serine integrase catalytic domain. *Structure (London, England : 1993)* **16**, 1275-1286, doi:10.1016/j.str.2008.04.018 (2008).
27 Keenholtz, R. A., Rowland, S. J., Boocock, M. R., Stark, W. M. & Rice, P. A. Structural basis for catalytic activation of a serine recombinase. *Structure (London, England : 1993)* **19**, 799-809, doi:10.1016/j.str.2011.03.017 (2011).
28 Ritacco, C. J., Kamtekar, S., Wang, J. & Steitz, T. A. Crystal structure of an intermediate of rotating dimers within the synaptic tetramer of the G-segment invertase. *Nucleic Acids Res* **41**, 2673-2682, doi:10.1093/nar/gks1303 (2013).
29 Trejo, C. S., Rock, R. S., Stark, W. M., Boocock, M. R. & Rice, P. A. Snapshots of a molecular swivel in action. *Nucleic Acids Res* **46**, 5286-5296, doi:10.1093/nar/gkx1309 (2018).
30 Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. W. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861, doi:10.7554/eLife.36861 (2018).
31 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500, doi:10.1038/s41586-024-07487-w (2024).
32 Yarnall, M. T. N. *et al.* Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nat Biotechnol* **41**, 500-512, doi:10.1038/s41587-022-01527-4 (2023).
33 Durrant, M. G. *et al.* Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. *Nat Biotechnol* **41**, 488-499, doi:10.1038/s41587-022-01494-w (2023).
34 Stark, W. M. Making serine integrases work for us. *Curr Opin Microbiol* **38**, 130-136, doi:10.1016/j.mib.2017.04.006 (2017).
35 Karow, M. & Calos, M. P. The therapeutic potential of PhiC31 integrase as a gene therapy system. *Expert Opin Biol Ther* **11**, 1287-1296, doi:10.1517/14712598.2011.601293 (2011).
36 Hollis, R. P. *et al.* Phage integrases for the construction and manipulation of transgenic mammals. *Reprod Biol Endocrinol* **1**, 79, doi:10.1186/1477-7827-1-79 (2003).
37 Li, W. *et al.* Structure of a synaptic gammadelta resolvase tetramer covalently linked to two cleaved DNAs. *Science* **309**, 1210-1215, doi:10.1126/science.1112064 (2005).
38 Montano, S. P. *et al.* Structural basis for topological regulation of Tn3 resolvase. *Nucleic Acids Res* **51**, 1001-1018, doi:10.1093/nar/gkac733 (2023).
39 Liu, T. A., Fang, Y. X., Yin, H. & Jing, Z. Z. A possible way that phiC31 integrase regulates the recombination direction. *Genet Mol Res* **14**, 507-514, doi:10.4238/2015.January.26.4 (2015).
40 Nakane, T. & Scheres, S. H. W. Multi-body Refinement of Cryo-EM Images in RELION. *Methods Mol Biol* **2215**, 145-160, doi:10.1007/978-1-0716-0966-8_7 (2021).

41 Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**, doi:10.7554/eLife.36861 (2018).
42 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).
43 Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of molecular biology* **333**, 721-745, doi:10.1016/j.jmb.2003.07.013 (2003).
44 Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci* **32**, e4792, doi:10.1002/pro.4792 (2023).
45 Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci* **27**, 14-25, doi:10.1002/pro.3235 (2018).
46 Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500, doi:10.1038/s41586-024-07487-w (2024).
47 Casanal, A., Lohkamp, B. & Emsley, P. Current developments in Coot for macromolecular model building of Electron Cryo-microscopy and Crystallographic Data. *Protein Sci* **29**, 1069-1078, doi:10.1002/pro.3791 (2020).
48 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877, doi:10.1107/s2059798319011471 (2019).
49 Nicholls, R. A., Tykac, M., Kovalevskiy, O. & Murshudov, G. N. Current approaches for the fitting and refinement of atomic models into cryo-EM maps using CCP-EM. *Acta Crystallogr D Struct Biol* **74**, 492-505, doi:10.1107/S2059798318007313 (2018).

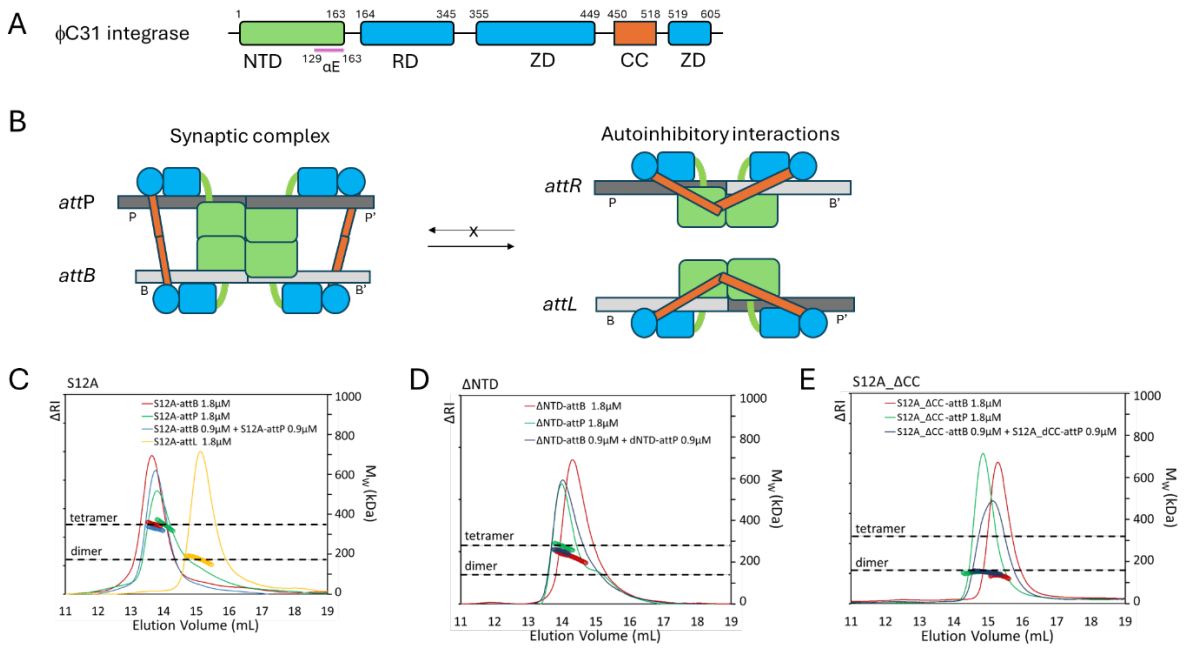

Figure 1. ϕC31 integrase forms tetrameric synaptic complexes with *attB* and *attP* DNA through its CC motif

(A) Domain architecture of ϕC31 integrase. NTD: N-terminal catalytic domain; αE: conserved helix important for oligomerisation, DNA binding and subunit rotation during recombination; RD: recombinase domain; ZD: zinc ribbon domain, CC: coiled-coil motif. (B) Model for the DNA recombination reaction scheme catalysed by large serine integrases. (C-E) SEC-MALLS analysis of full-length ϕC31 integrase (C) and variants with domain deletions (D, E) complexed with *attB* (red), *attP* (green), *attL* (yellow), or *attB-attP* mix (blue). The protein elution profile is shown as elution volume (x-axis) against differential refractive index (ΔRI) (y-axis). Circular markers represent the calculated weight averaged molar mass (M_W). ΔNTD, residues 1-140 removed; ΔCC, residues 450-515 removed. M_w values for the expected dimer and tetramer are marked by dashed lines.

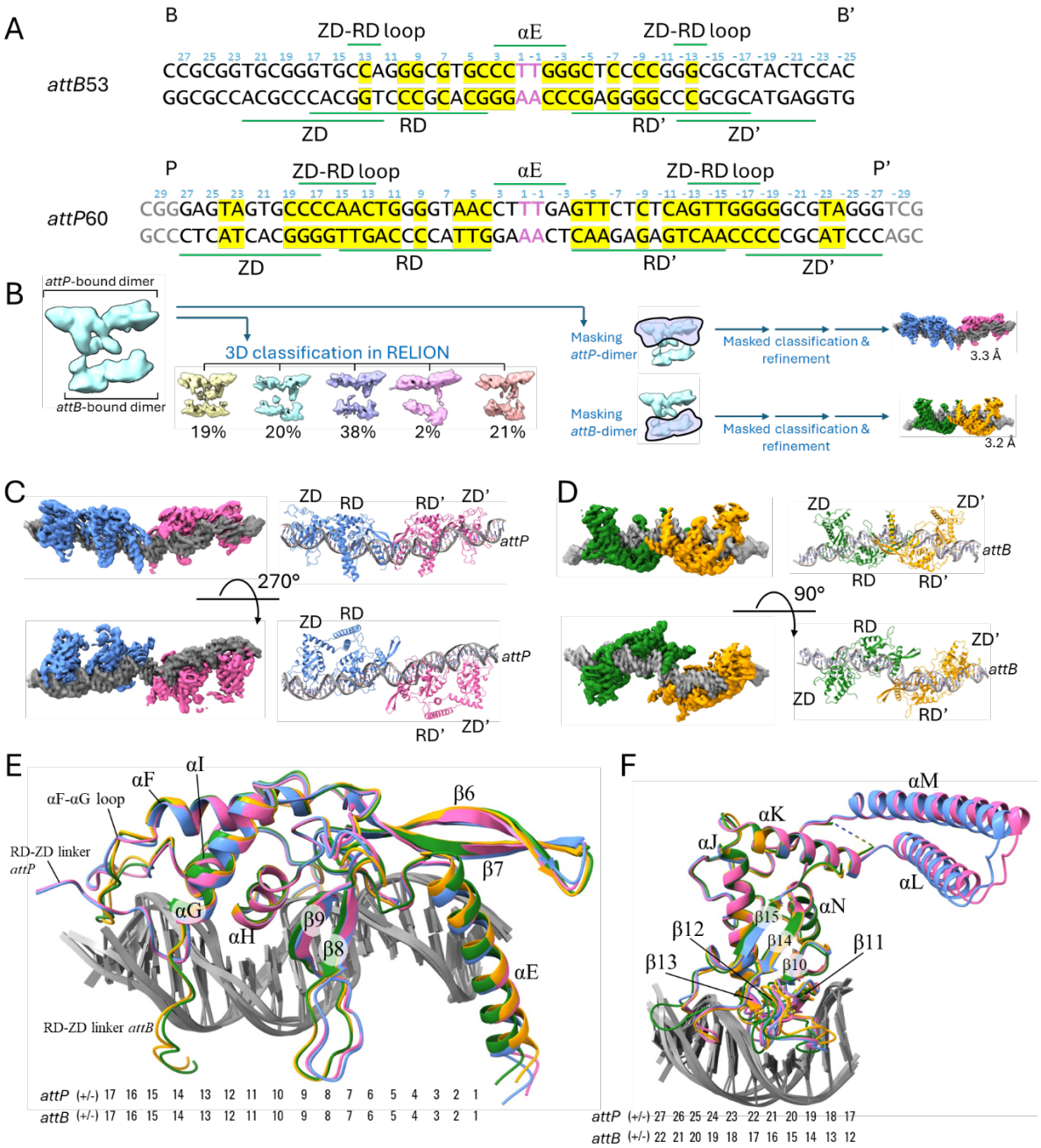

Figure 2. High-resolution structures of the *attB*- and *attP*-bound dimer regions
(A) Sequences of the 53-bp *attB*53 and 60-bp *attP*60 DNA oligonucleotides used in the cryo-EM experiments. The central 2 bp of the *att* sites are coloured in pink. Base-pairs conserved between the left and right half-sites are highlighted in yellow. DNA base-pairs are numbered sequentially from the centre, with positive and negative numbering for the left and right half-site respectively. Solid lines mark the base-pair positions bound by each DNA-binding domain and motif, as observed in the cryo-EM structure. Base-pairs in *attP* that are unmodelled due to poor density are shown in grey letters. (B) Summary of the cryo-EM data processing procedures that produced the high-resolution dimer maps. (left) Classification of entire particles revealed a high degree of conformational heterogeneity that prevented high-resolution reconstruction. (right) A mask (transparent blue) was applied to either the *attP*- or *attB*-bound dimer, which restricted image alignment and classifications to map regions within the mask, ignoring the conformational heterogeneity in the unmasked area. This procedure gave *attB*- and *attP*-bound dimer maps at 3.2 Å and 3.3 Å resolution, respectively. (C) Cryo-EM reconstruction of the *attP*-bound dimer at 3.3 Å resolution (left panel), and the refined

atomic model (right panel) shown in cartoon representation. The *attP*-bound monomer is coloured in blue, the *attP'*-bound monomer is in pink, and *attP* DNA is in dark grey. (D) Cryo-EM reconstruction (left panel) of the *attB*-bound dimer at 3.2 Å resolution, and the refined atomic model (right panel) shown in cartoon representation. The *attB*-bound monomer is in green, the *attB'*-bound monomer is in orange, and the *attB* DNA is in light grey. (E, F) Structural comparisons of RD-DNA binding (D) and ZD-DNA binding (E) by all four monomers. The monomers and DNA are coloured as in (B) and (C). Structures are aligned at their Cα atoms from the protein regions in RD and ZD. Helices and sheets have been labelled. The *attB* and *attP* DNA base-pairs are numbered below the structures, as in (A). All resolution calculations are based on the gold standard Fourier shell correlation (FSC) 0.143 criterion. The structural images were generated in ChimeraX[45].

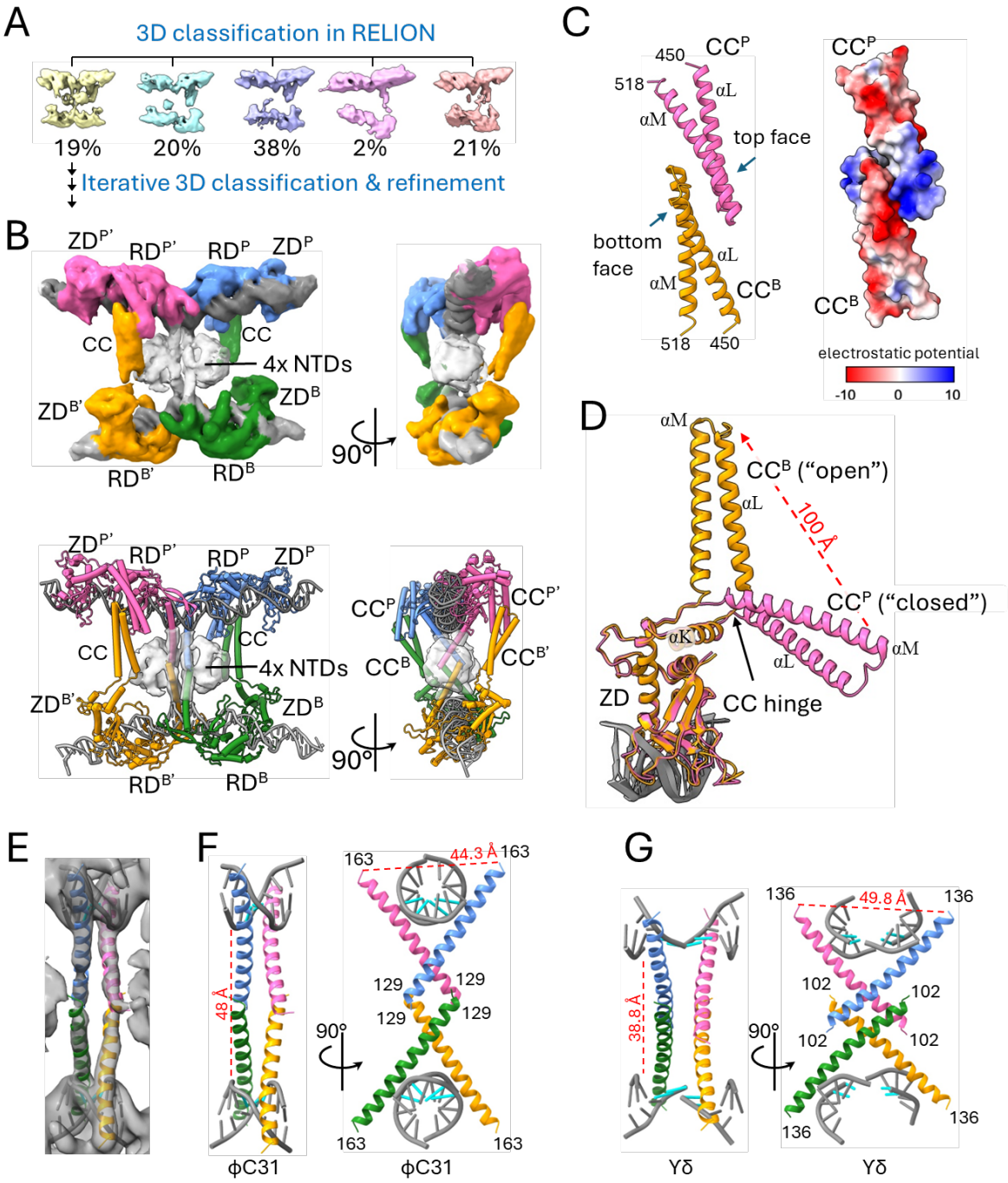

Figure 3. Structure of the *attB* x *attP* synaptic tetramer
(A) Graphical representation of the cryo-EM image processing scheme that resulted in the 7.5 Å map of the full synaptic tetramer. 3D classification yielded particle classes with high conformational heterogeneity that differed in their relative positioning of the *attB* and *attP* DNA. Multiple rounds of 3D classifications were carried out to isolate particles that produced the best resolution map. These particles belonged to a class with the nearest to parallel arrangement of the *attB* and *attP* DNA double helices (see text for details). (B) Cryo-EM map (top) and the refined model shown in cartoon representation (bottom) of the pre-integration *attP* x *attB* synaptic complex. The *attB* and *attP* oligonucleotides are coloured in light and dark grey, respectively. The integrase subunits and the DNA are coloured as in Figure 2. The map density corresponding to the tetrameric NTD core (residues 1-128) is shown as a grey transparent surface. These residues are unmodelled due to lower local map resolution in this region. (C) Details of the CC^B-CC^P interaction shown in cartoon representation (left) and in

surface view coloured by electrostatic potential (right). (D) Structural comparison showing the different CC motif trajectories adopted by the *attB*- and *attP*-bound monomers. Structures have been aligned using the Cα atoms in the ZD. The black arrow marks the hinge region (residues 447-450) involved in movements of the CC motif. (E) Density fit for the four αE helices and the central 8 bp, with the experimental map shown in grey transparent surface. (F, G) Structural comparisons of the αE helices synaptic tetramer (uncleaved DNA) and the equivalent region of the γδ resolvase-DNA synaptic complex structure 1ZR4 (in a strand exchange intermediate state with double-strand cleaved DNA). The central 2 bp of the recombination sites are coloured in cyan. The distances between the scissile phosphates in the two structures are indicated in red, and the distances between the C-terminal residues of the αE helices (measured between their Cα atoms; L163 in ϕC31 integrase and K136 in γδ resolvase). Structural images were generated in ChimeraX[45].

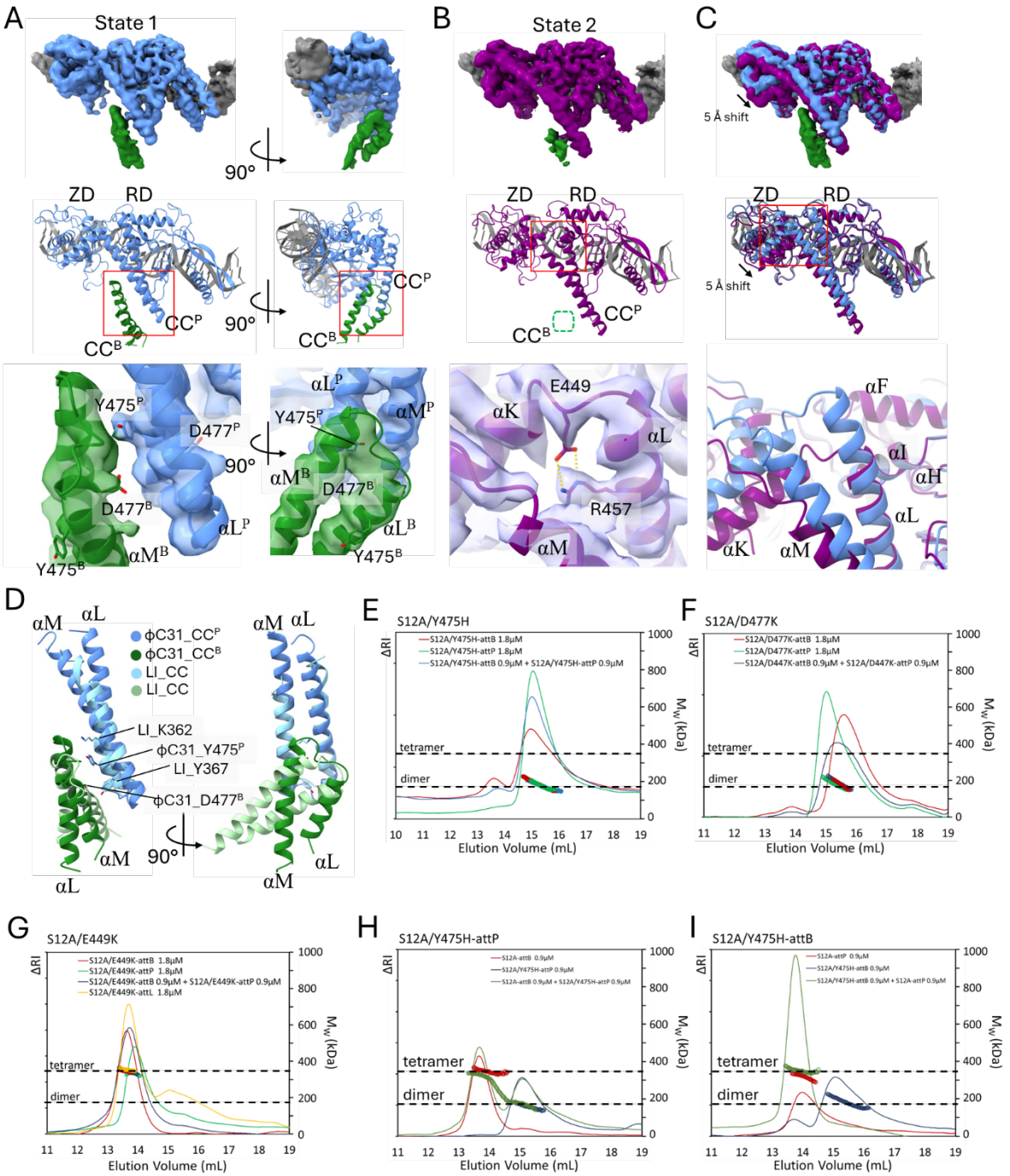

Figure 4. Structures of the CC motifs reveal mechanisms for synapsis and directionality control

(A) Cryo-EM reconstruction of a focus-refined map masking a single *attP*-bound monomer and the CC motif from the interacting *attB* monomer at 3.8 Å resolution. The map (top panel) and model (middle panel) are coloured as follows: *attP*-bound monomer blue, *attB* monomer CC motif green, DNA grey. Red boxes mark the enlarged regions shown in the bottom panels. (bottom) Close-up view of the 'top-to-bottom' CC-CC interface showing experimental density as transparent surface and the model in cartoon view. CC-CC interface residues targeted by mutagenesis are shown as sticks. (B) Cryo-EM reconstruction resulting from a separate set of particles in a different particle class refined to 3.5 Å resolution. These particles adopt a different conformational state compared to the ones shown in (A), and the two conformations are designated state 1 and state 2. Top and middle panels show the map and models respectively. The colour scheme is the same as in (A) except that the *attP*-bound monomer is

shown in purple. Density for the *attB*-bound CC motif in state 2 is weak and did not allow confident modelling, so it is represented as a dotted box in the middle panel. The bottom panel shows a close-up view of the CC-hinge region that connects αK to αL, with the experimental map density shown as transparent surface. Side chains from E449 and R457 are shown as sticks. This pair of side chains forms a salt bridge that stabilises the 'closed' hinge. (C) Structural comparisons of the two conformational states shown in (A) and (B). The RDs of the maps and models have been aligned. The ZD (including the CC motif) of state 2 is shifted downwards by approximately 5 Å (indicated by the black arrow). The bottom panel shows a close up view at the RD-CC interface. The CC motif of state 2 is shifted downwards by 2 helical turns, such that αI (RD) packs against αL (CC motif) at its third helical turn in state 1, but at the first helical turn in state 2. (D) Structural comparisons of the CC-CC dimer from the ϕC31 integrase synaptic complex and LI integrase (PDB ID code 5U96). Side chains that block dimerisation in ϕC31 and LI integrases[13] are shown in sticks. Structural images were generated in ChimeraX[45]. (E-I) SEC-MALLS analysis of ϕC31 integrase mutants Y475H, D477K, and E449K.

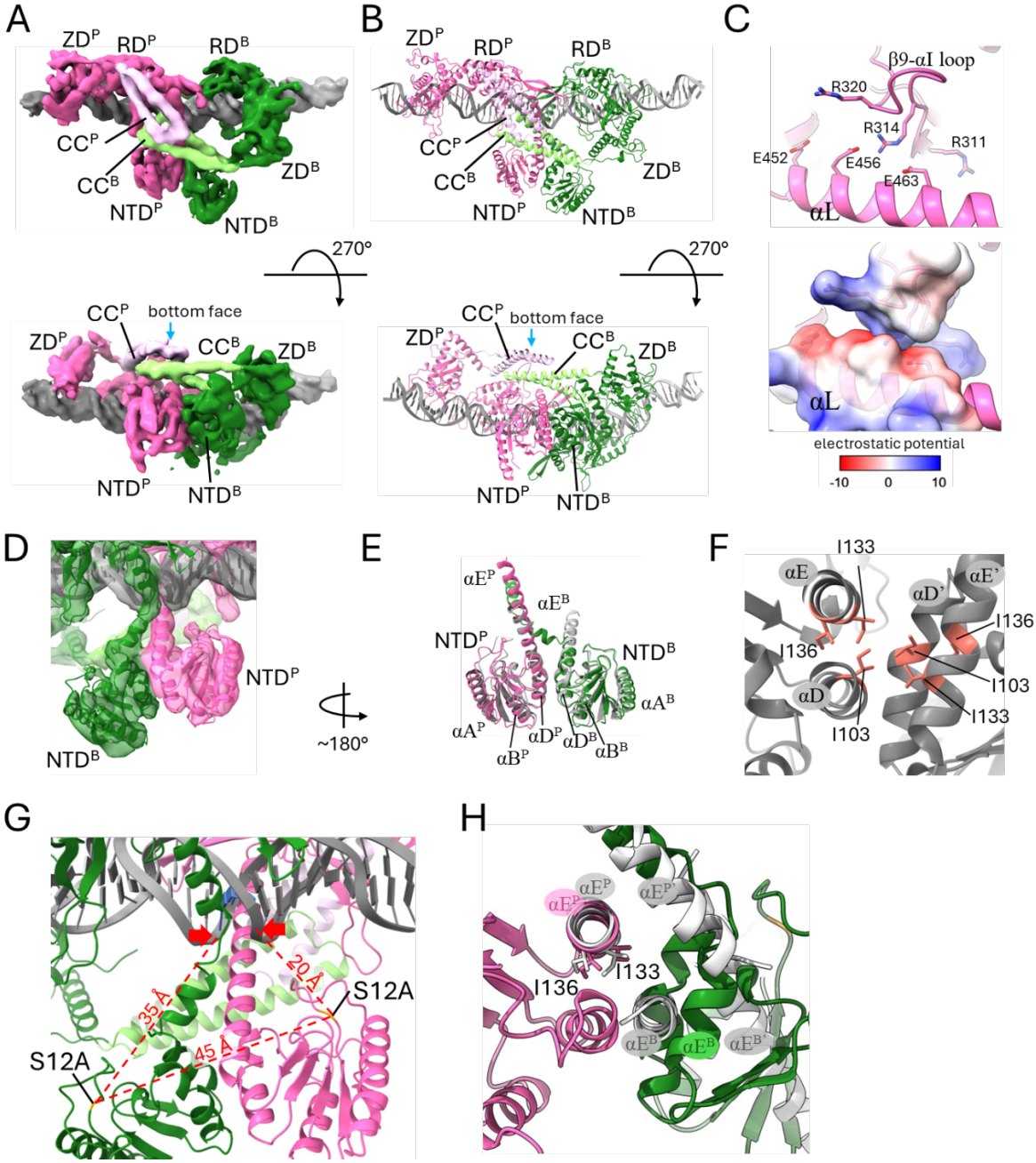

Figure 5. Structure of the *attL*-bound dimer complex

(A, B) Map and model of the *attL*-bound dimer complex at 6.0 Å resolution. NTD: N-terminal catalytic domain; RD: recombinase domain; ZD: zinc ribbon domain, CC: coiled-coil motif. The *attL* site is a product of *attP* x *attB* recombination and thus comprises an *attP* half-site joined to an *attB* half-site; B and P superscripts denote the *attP*- and *attB*-half site bound monomer respectively. Blue arrows mark the 'bottom face' of the CC^P motif. (C) Structural details of the CC^P-RD^P interface, formed between the β9-αI loop (residues 310-322) and the N terminal region of αL (residues 452-463). (top) Cartoon representation showing the interface, with acidic side chains that are mutated in hyperactive integrase mutants[14] basic side chains on the β9-αI loop shown as sticks. (D) Density fit of the NTDs. (E) Structural comparisons of the *attL*-bound NTD dimer and that observed in the NTD+RD crystal structure (PDB ID code 4BQQ). (F) Close up view of the dimer interface observed in 4BQQ. Residues at the interface mentioned in the text are coloured pink and shown as sticks. (G) Close up view of the *attL*-bound NTD dimer and the DNA at the centre of *attL*. The central 2 base pairs of *attL* are in blue. The active site S12A residues are in orange. Red arrows mark the positions of the two

scissile phosphates. The distances between each S12A residue (α-carbon atom) and its scissile phosphate (P atom), and the distance between the two S12A residues are indicated on the red dashed lines. (H) Structural comparison of the dimeric αE arrangement in the *attL*-bound complex and the tetrameric αE arrangement in the synaptic structure. The αE of the monomer bound to the *attP*-derived half-site of *attL* is superimposed on the *attP*-bound αE of the synaptic tetramer structure. The *attB*- and *attP*-half-site bound monomers of the *attL* complex are in green and pink respectively, and the four αE's of the synaptic tetramer are in pale grey. Different protein interfaces are used to dimerise *attL*-bound NTDs and to assemble the synaptic tetramer.

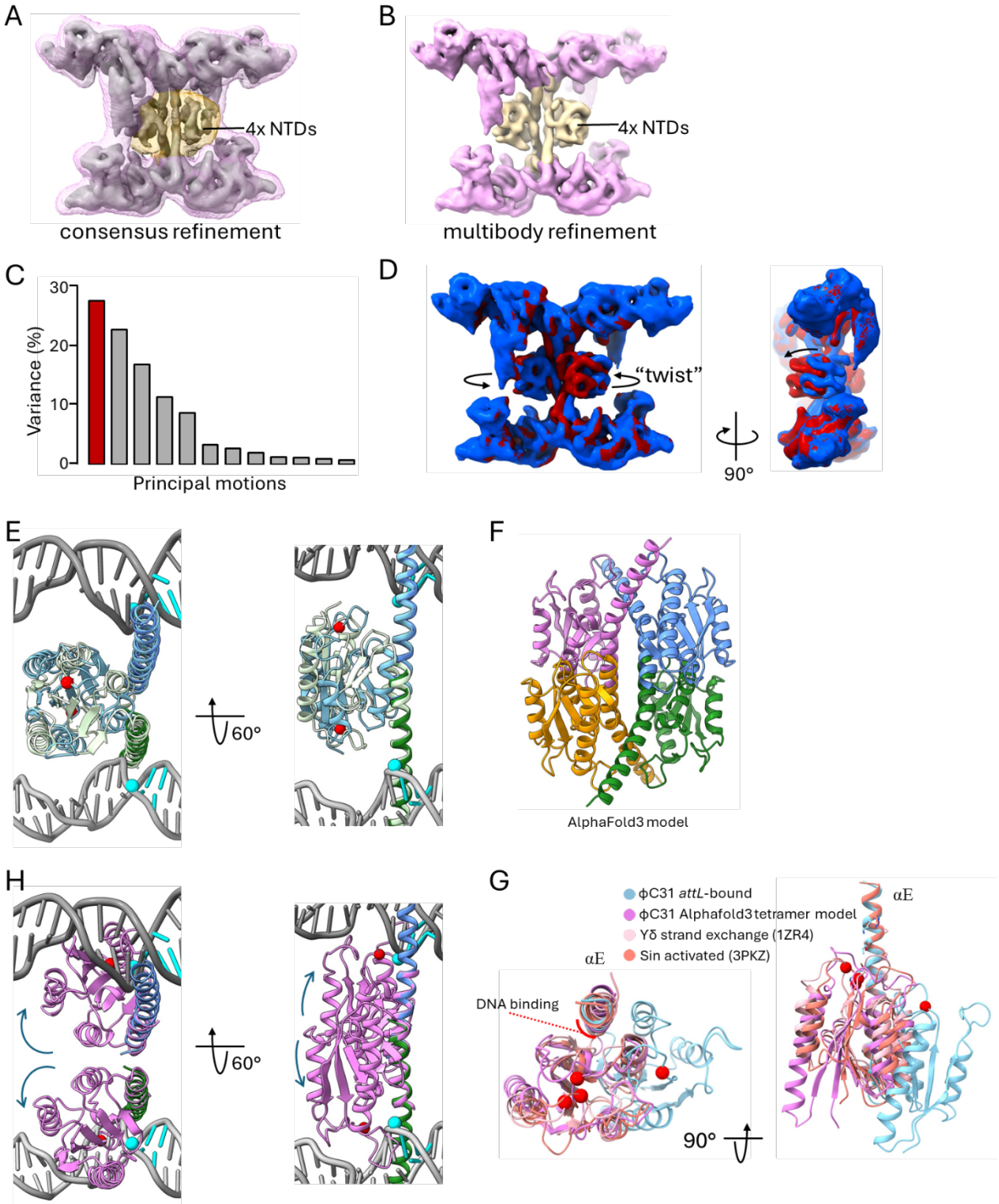

Figure 6. Conformational changes move the NTDs towards the scissile phosphates upon synapsis

(A) The map of the full synaptic tetramer complex, with the two body masks used in multibody refinement superimposed on top and shown as transparent surfaces. (B) The resulting density from multibody refinement coloured the same as their masks in (A). (C) The contributions of each principal motion to the total variance. The first principal motion, for which the maps at the conformational extremes are shown in panel D, is highlighted in red. (D) Motion in the first principal motion. The NTD tetramer is seen rotating about the twofold symmetry axis of the complex. (E-G) Modelling of the NTD assembly within the *attP-attB* synaptic complex. (E) Two copies of the *attL*-bound NTD (light green and blue) are aligned with two αE helices from a 'rotating dimer' of the synaptic structure (dark green and blue), by superimposing the Cα atoms of residues 129-163. The resulting model displays substantial steric clashes in the NTD core

region. *attB* and *attP* DNA are coloured in light and dark grey respectively, with the central 2 bp coloured in cyan. The Cα atoms of the active site S12A residues are shown as red spheres. The phosphorus atoms of the scissile phosphates are shown as cyan spheres. (F) AlphaFold3 predicted structure of the tetramer formed by the NTDs (residues 1-163). (G) Structural superposition, at their αE helices, of NTD monomers from the structure of the *attL*-bound dimer, the AlphaFold3 predicted tetramer, the γδ resolvase synaptic intermediate tetramer (1ZR4), and the activated Sin invertase tetramer (3PKZ), respectively. Two orthogonal views are shown. The Cα atoms of the active site serines are shown as red spheres. The side of the αE helix involved in DNA binding is marked by a curved red line, with amino acid side chains involved in DNA minor groove interaction shown as sticks. The trajectory of the NTD core in the AlphaFold model is rotated about αE relative to that in the *attL*-bound monomer, and resembles those seen in activated small serine resolvases. (H) The same modelling exercise as in (E) but performed with the monomer from the AlphaFold-predicted NTD tetramer, whose trajectory avoids the previously observed steric clashes, and positions the two active site serine residues adjacent to their respective scissile phosphates. Arrows show the potential trajectory change for the NTD core regions.

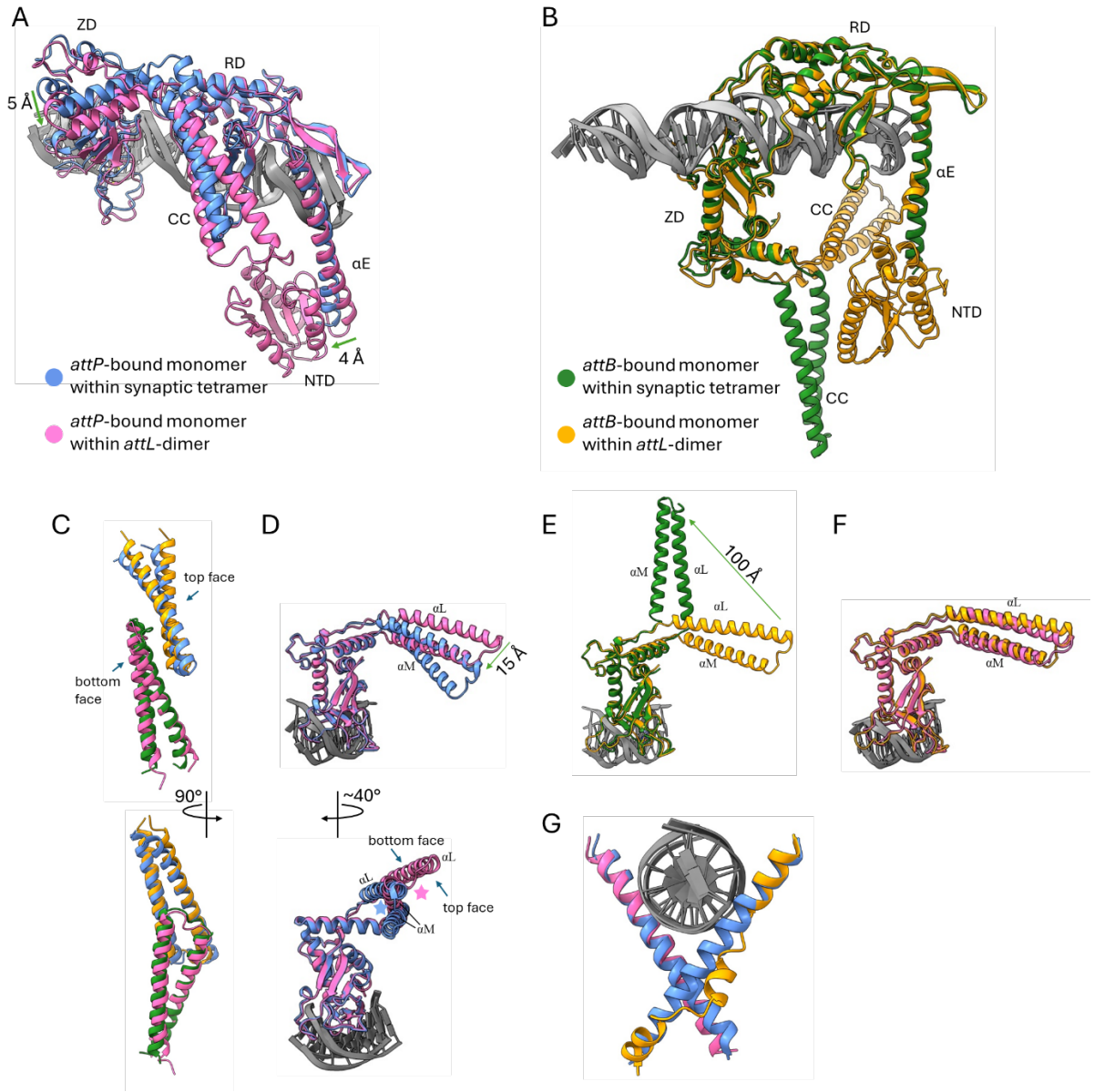

Figure S1. Structural comparisons between monomers within the *attL*-bound dimer and the synaptic tetramer.

(A) The *attP*-bound monomer within the synaptic tetramer (blue) compared to the monomer bound to the *attP*-derived half-site of *attL* (pink). Structures were superposed using the Cα atoms of the RD. (B) *attB*-bound monomer within the synaptic tetramer (green) compared to the monomer bound to the *attB*-derived half-site of *attL* (orange). (C) Comparison of the CC motif dimer interaction in the synaptic complex and the CC^B-CC^P interaction in the *attL* complex. (D-F) Comparison of the ZD (including the CC motif) bound to *attB* and *attP* within the synaptic tetramer and in the *attL*-bound dimer. Structures were superposed using the Cα atoms of the ZD domain, excluding those in the CC motif. Blue and pink stars mark the binding positions of CC^B for the synaptic and *attL*-bound complexes respectively. (G) Structural comparisons of the *attP*-bound αE helices within the synaptic tetramer and the αE helices in the *attL*-bound integrase dimer structure. Colour codes in C-G are the same as in (A) and (B). NTD: N-terminal catalytic domain; RD: recombinase domain; ZD: zinc ribbon domain, CC: coiled-coil motif. Green arrows represent structural movements observed.

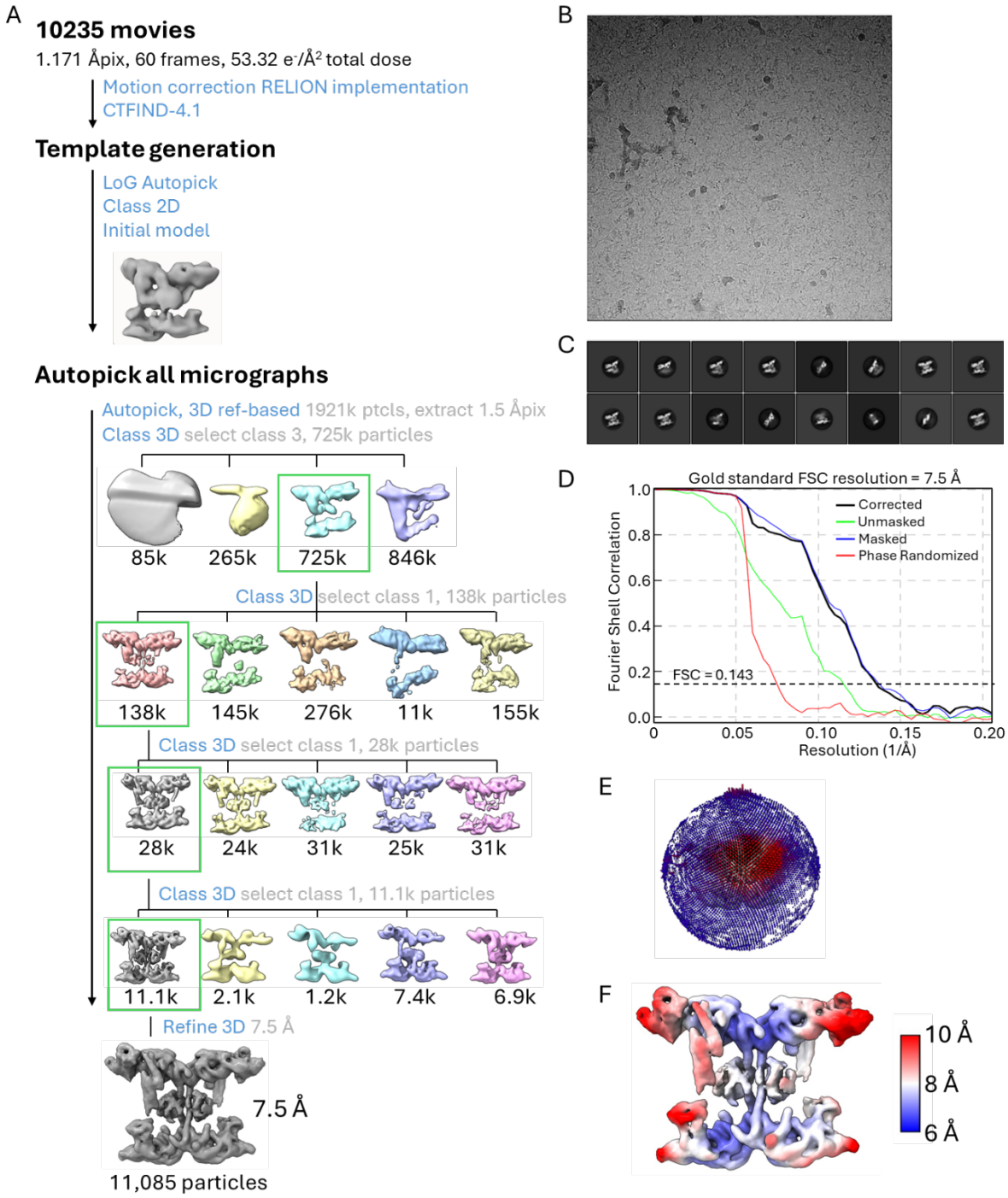

Figure S2. Cryo-EM data processing of the synaptic tetramer reconstruction.
(A) Schematics of the image processing workflows carried out using RELION 5.0. 3D classification and refinement jobs were done using a global mask that included signals from the entire synaptic complex. Green boxes mark the classes of particles selected for further processing. The final reconstruction used for model building was obtained using 11,085 particles and had an overall resolution of 7.5 Å as determined by the Fourier shell correlation (FSC) 0.143 criterion. (B) A representative cryo-EM micrograph. (C) 2D class averages showing the various projection views. (D) FSC plots for the final reconstruction shown in (A). (E) Euler angle distribution of particles used in the final reconstruction. (F) Local resolution assessment of the final reconstruction.

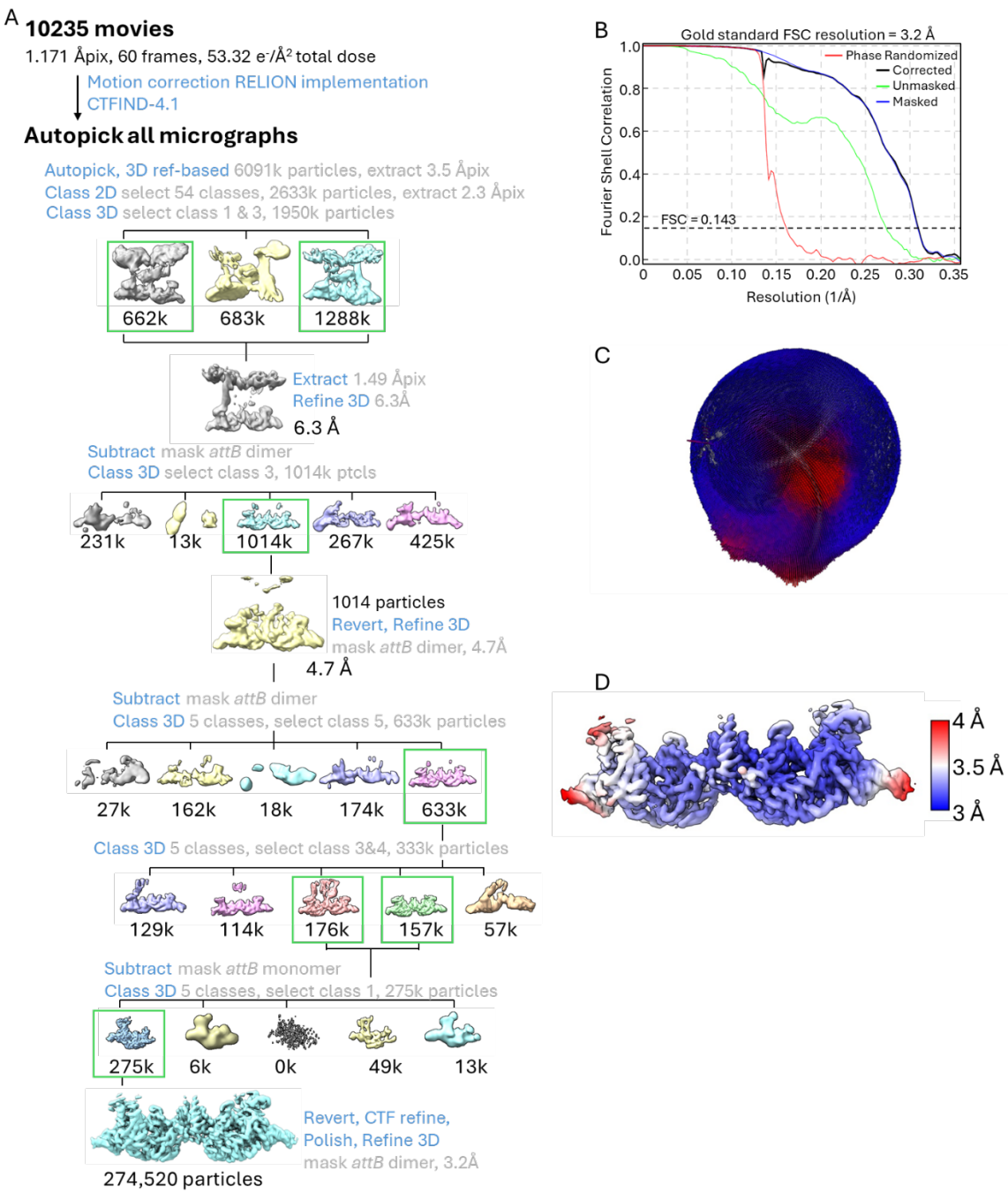

Figure S3. Cryo-EM data processing for the *attB*-bound dimer reconstruction.
(A) Schematics of the image processing workflows carried out using RELION 5.0. 3D classification and refinement jobs in the later stages were done using a mask that only included signals from the *attB*-bound dimer. Green boxes mark the particle classes selected for further processing. The final reconstruction used for model building was obtained using 274,520 particles and had an overall resolution of 3.2 Å as determined by the Fourier shell correlation (FSC) 0.143 criterion. (B) FSC plots for the final reconstruction shown in (A). (C) Euler angle distribution of particles used in the final reconstruction. (D) Local resolution assessment of the final reconstruction.

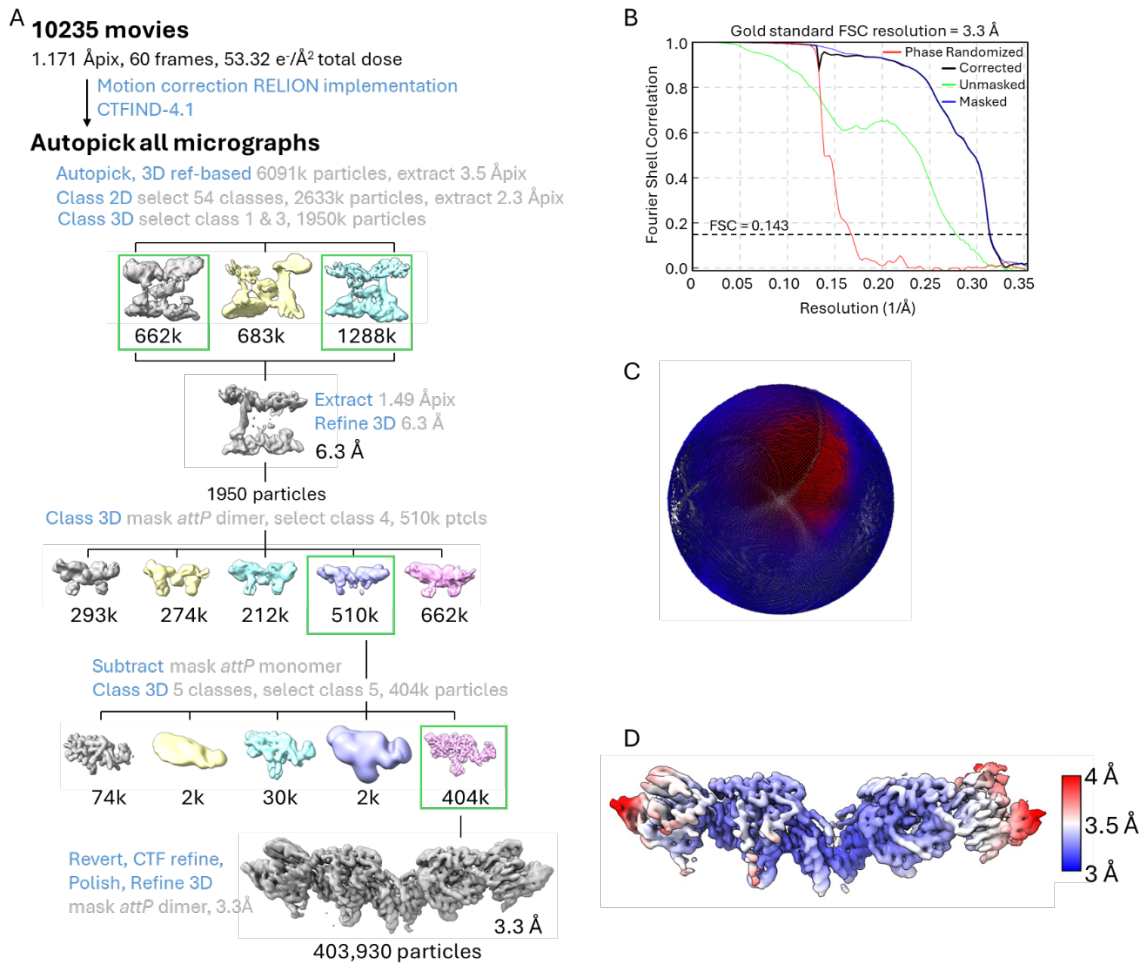

Figure S4. Cryo-EM data processing for the *attP*-bound dimer reconstruction.
(A) Schematics of the image processing workflows carried out using RELION 5.0. 3D classification and refinement jobs in the later stages were done using a mask that only included signals from the *attP*-bound dimer. Green boxes mark the particle classes selected for further processing. The final reconstruction used for model building was obtained using 403,930 particles and had an overall resolution of 3.3 Å as determined by the Fourier shell correlation (FSC) 0.143 criterion. (B) FSC plots for the final reconstruction shown in (A). (C) Euler angle distribution of particles used in the final reconstruction. (D) Local resolution assessment of the final reconstruction.

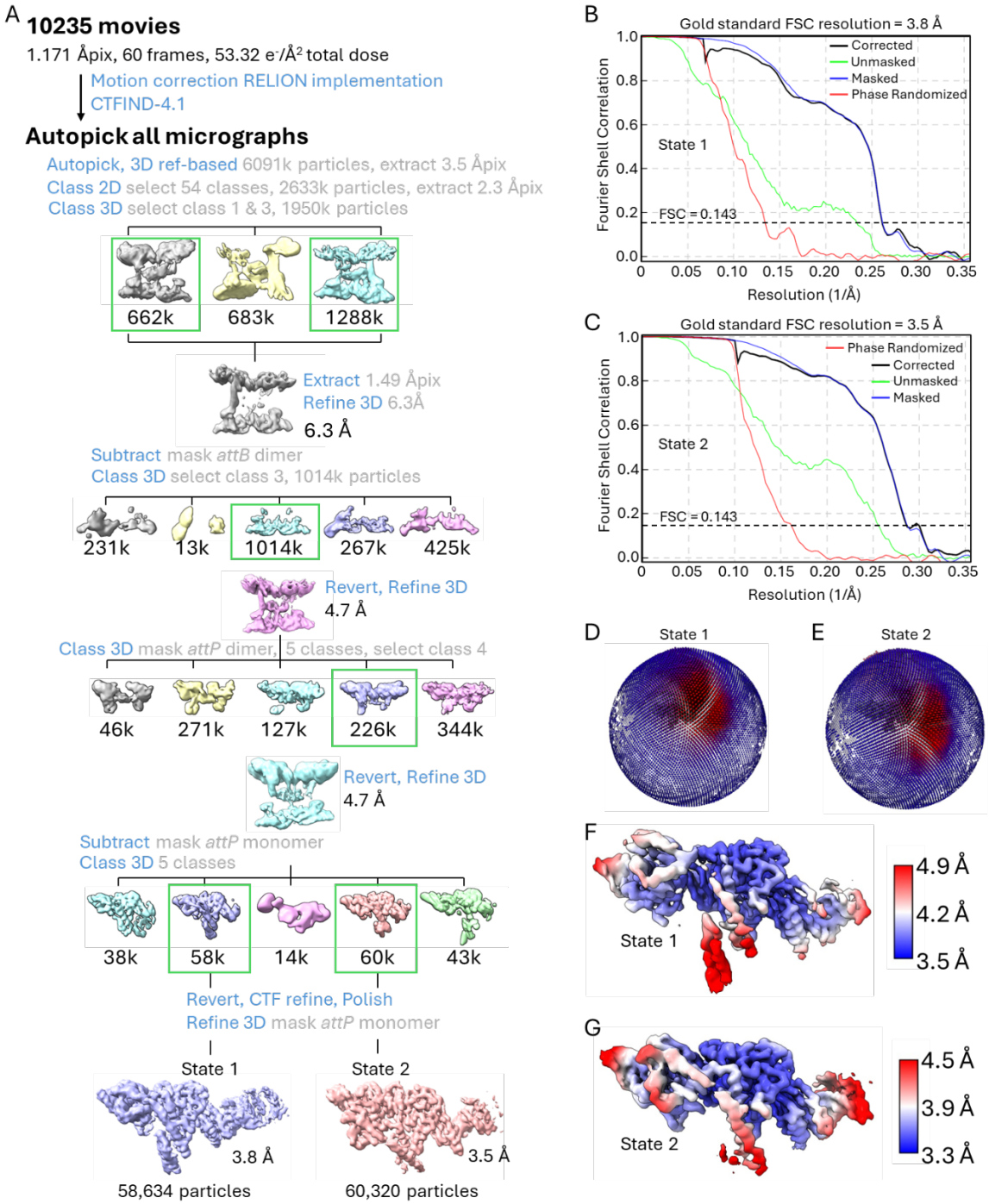

Figure S5. Cryo-EM data processing for the state 1 and state 2 *P*-bound monomer reconstructions.

(A) Schematics of the image processing workflows carried out using RELION 5.0. Green boxes mark the particle classes selected for further processing. Particle subtraction and 3D classification using a protein mask that covered a single *attP*-bound monomer had yielded particles belonging to two conformational states, designated state 1 and state 2 respectively. The final reconstructions for state 1 and state 2 were obtained using 58,634 and 60,320 particles respectively, and had an overall resolution of 3.8 Å and 3.5 Å respectively. (B-C) FSC plots for the final reconstructions for state 1 and state 2 respectively. (D-E) Euler angle distribution of particles used in the final reconstructions for state 1 and state 2 respectively. (F-G) Local resolution assessment of the final reconstructions for state 1 and state 2 respectively.

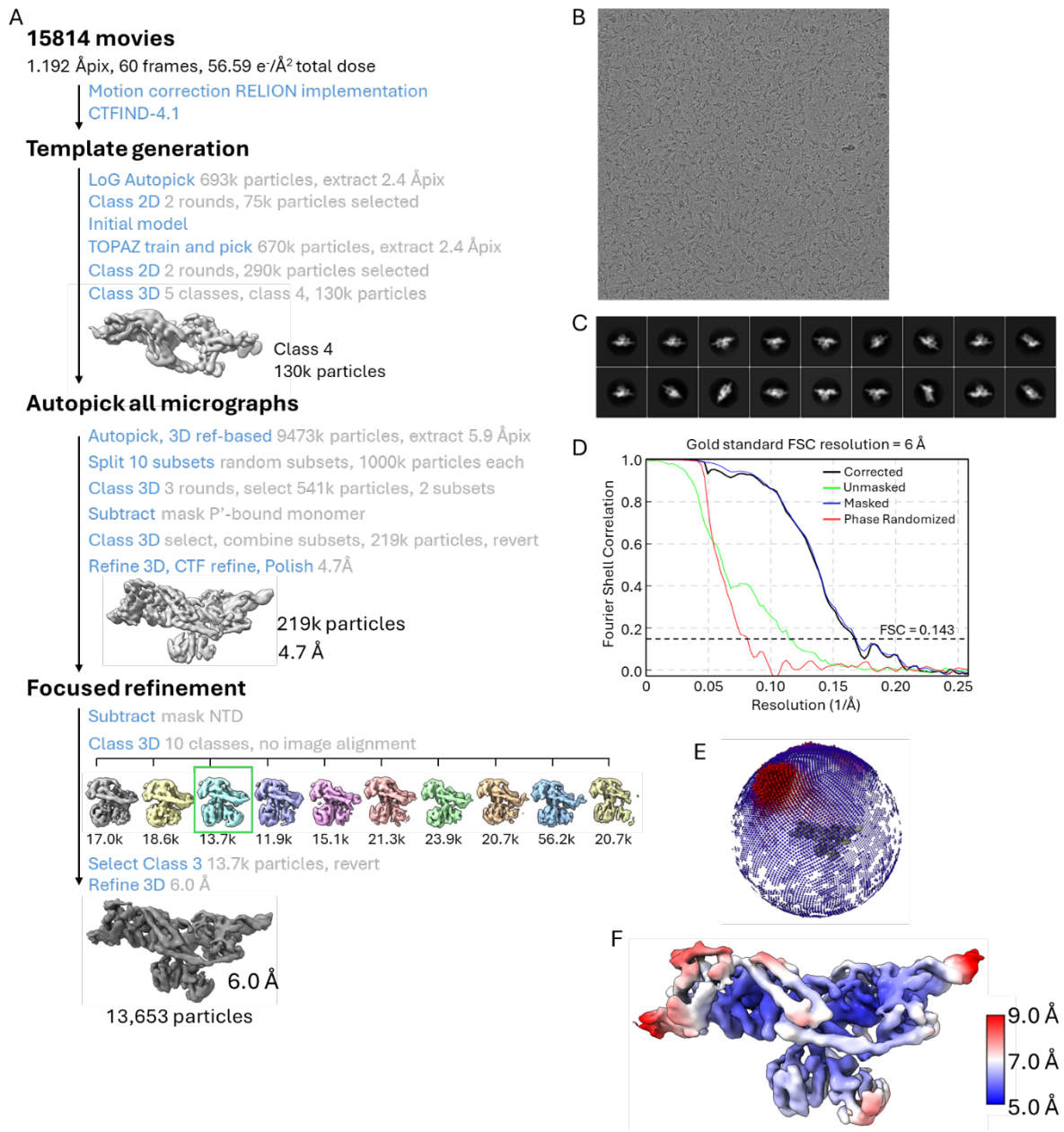

Figure S6. Cryo-EM data processing of the *attL*-bound dimer.
(A) Schematics of the image processing workflows carried out using RELION 5.0. The green box marks the class of particles selected to generate the final reconstruction, which was obtained using 13,653 particles and had an overall resolution of 6.0 Å as determined by the Fourier shell correlation (FSC) 0.143 criterion. (B) A representative cryo-EM micrograph. (C) 2D class averages showing the various projection views. (D) FSC plots for the final reconstruction shown in (A). (E) Euler angle distribution of particles used in the final reconstruction. (F) Local resolution assessment of the final reconstruction.

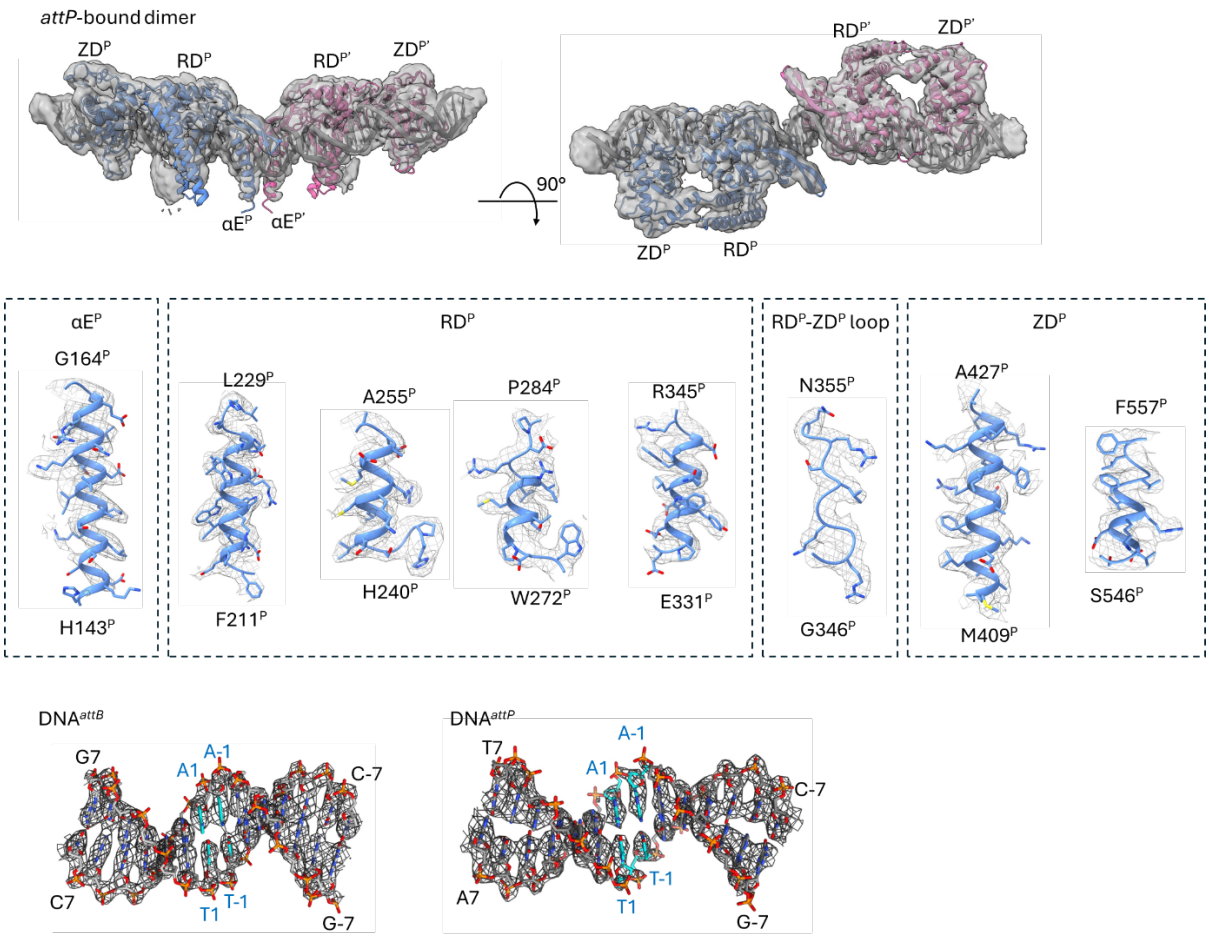

Figure S7. Cryo-EM density.
Models are shown as cartoon and sticks and the EM density is shown as a grey surface or mesh.

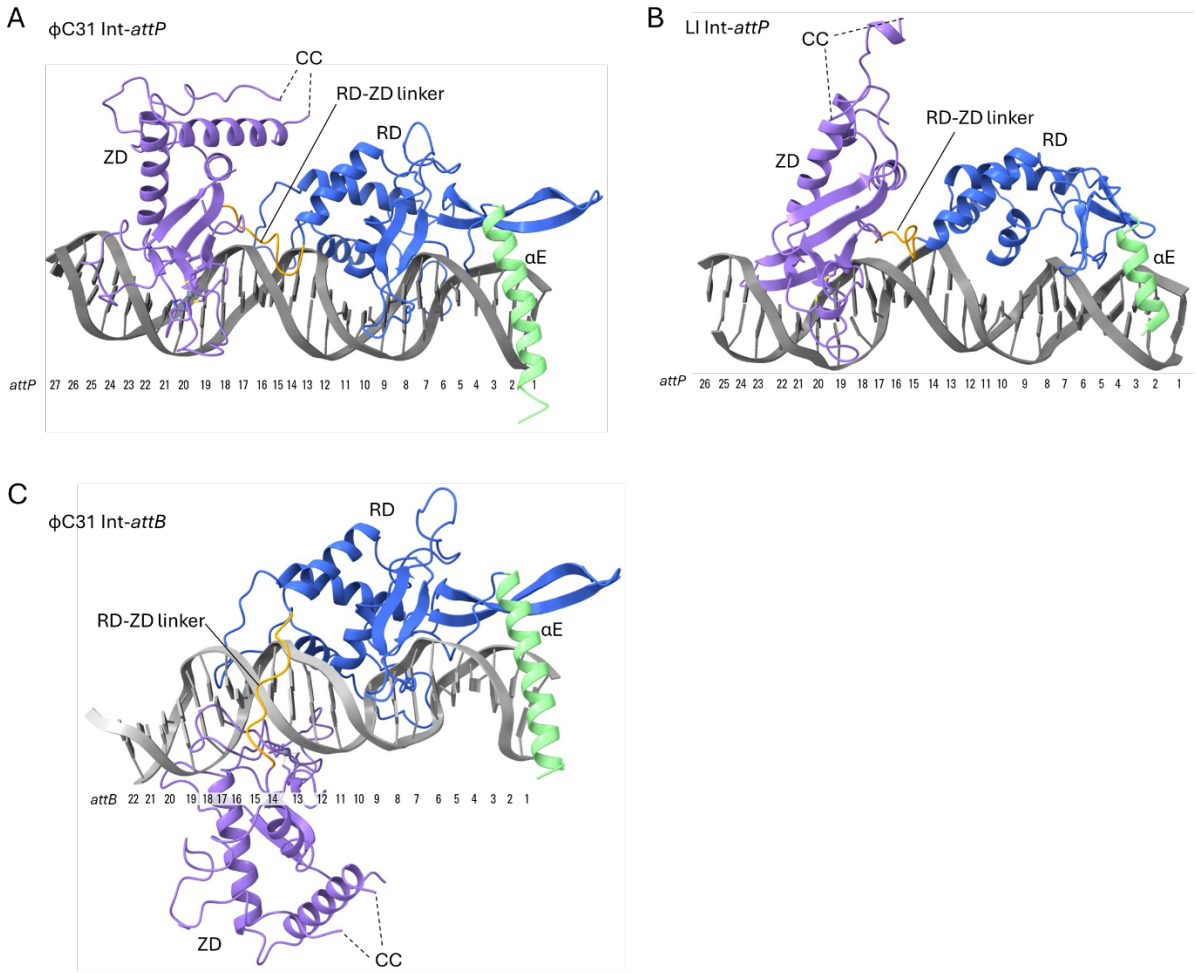

Figure S8. Structural comparisons of *att* site binding between ϕC31 and LI integrases. (A) Structure of ϕC31 integrase bound to an *attP* half-site in the synaptic complex. (B) Structure of LI integrase bound to an *attP* half-site (PDB ID code 4KIS[6], chain A). (C) Structure of ϕC31 integrase bound to an *attB* half-site in the synaptic complex. Structures are shown in cartoon representations and are coloured based on structural features. For clarity, the CC motifs are not shown. The *att* site DNA base-pairs are numbered as shown below the structures, using the same numbering scheme as in Figure 2A.

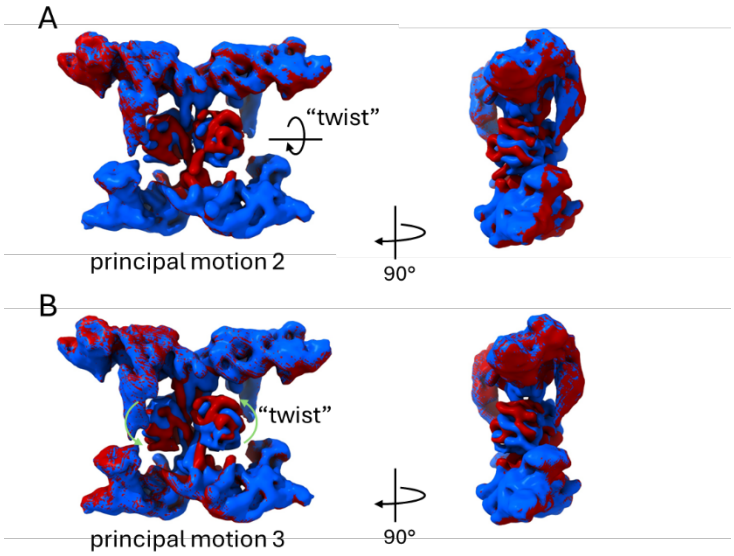

Figure S9. Principal motion analysis of the NTD tetramer within the synaptic complex. Maps at the two conformational extremes are aligned and coloured in blue and red respectively, for the second and third principal motion shown in Figure 6C.

	Synaptic tetramer	*attB*-bound dimer	*attP*-bound dimer	*attP*-bound monomer state 1	*attP*-bound monomer state 2	*attL*-bound dimer
Movies collected	10235	10235	10235	10235	10235	15814
Magnification	120k	120k	120k	120k	120k	120k
Voltage (kV)	300	300	300	300	300	200
Electron exposure (e-/Å^2)	55.32	55.32	55.32	55.32	55.32	56.59
No. of frames	60	60	60	60	60	60
Exposure time (s)	11.2	11.2	11.2	11.2	11.2	12.0
Defocus range (μm)	0.7-2.3	0.7-2.3	0.7-2.3	0.7-2.3	0.7-2.3	0.7-2.3
Pixel size (Å)	1.171	1.171	1.171	1.171	1.171	1.192
Initial particle images	1,921,071	6,090,789	6,090,789	6,090,789	6,090,789	9,473,033
Final particle images	11,085	274,520	403,930	58,634	60,320	13,653
Map resolution (Å)	7.49	3.20	3.29	3.53	3.82	5.98
FCS threshold	1.43	1.43	1.43	1.43	1.43	1.43

Table 1. Cryo-EM data collection statistics.