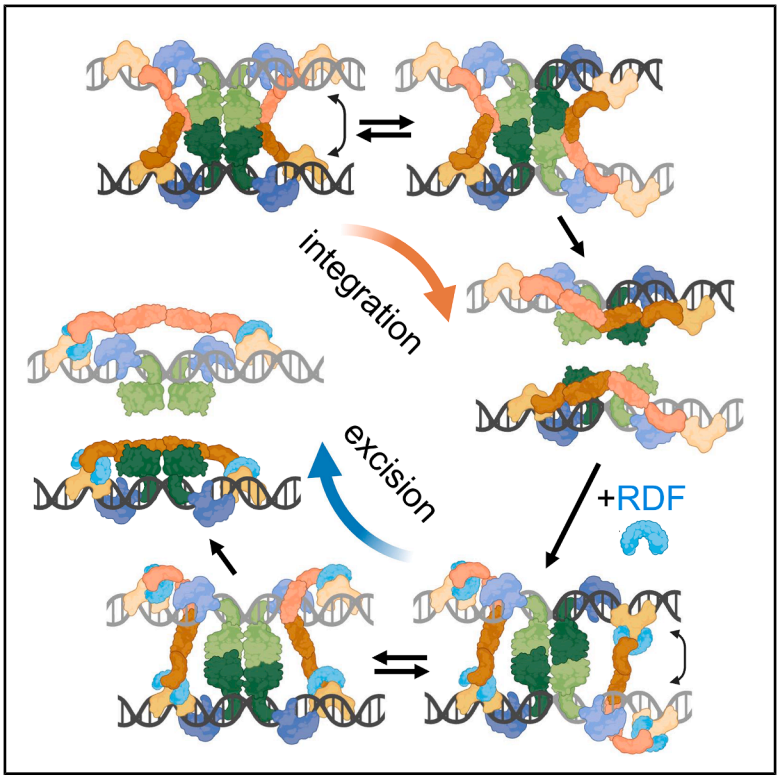
Molecular Cell

Article

Structural basis of directionality control in large serine integrases

Graphical abstract



Authors

Heewhan Shin, Ying Pigli, Tania Peña Reyes, ..., James R. Fuller, Femi J. Olorunniji, Phoebe A. Rice

Correspondence

f.j.olorunniji@ljmu.ac.uk (F.J.O.), price@uchicago.edu (P.A.R.)

In brief

Shin et al. present cryo-EM structures of the SPβ integrase along the integration and excision pathways that provide mechanistic insights addressing the long-standing questions of how large serine integrases enforce highly directional DNA recombination and how the cognate recombination directionality factor leads to a complete reversal of directionality.

Highlights

- Cryo-EM snapshots during and after both integrative and excisive recombination
- RDF-driven remodeling of the complexes dictates excision versus integration pathways
- Directionality in both pathways stems from conformational trapping of DNA-bound dimers
- Structures explain selective synapsis of appropriate substrate DNAs for each pathway

Shin et al., 2026, Molecular Cell *86*, 1–17
November 19, 2026 © 2026 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.molcel.2026.09.028

CellPress

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

Molecular Cell

CellPress
OPEN ACCESS

Article

Structural basis of directionality control in large serine integrases

Heewhan Shin,[1] Ying Pigli,[1] Tania Peña Reyes,[1] Adebayo J. Bello,[2] Oluwateniola T. Taiwo-Aiyerin,[2] James R. Fuller,[1] Femi J. Olorunniji,[2,*] and Phoebe A. Rice[1,3,*]

[1]Department of Biochemistry & Molecular Biology, The University of Chicago, Chicago, IL 60637, USA
[2]School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Liverpool L3 3AF, UK
[3]Lead contact
*Correspondence: f.j.olorunniji@ljmu.ac.uk (F.J.O.), price@uchicago.edu (P.A.R.)
https://doi.org/10.1016/j.molcel.2026.09.028

SUMMARY

Large serine integrases catalyze unidirectional insertion of large DNA payloads and, in the presence of a cognate recombination directionality factor (RDF), catalyze unidirectional excision. Understanding how directionality is enforced has been hampered by a lack of structural information. Here, we report 8 structures of SPβ integrase-DNA complexes along the integrative (−RDF) and excisive (+RDF) reaction pathways at resolutions extending to 3.15 Å. These include tetrameric intermediates before and after strand exchange and product-bound dimers for both pathways. We find that recombination-induced conformational changes and RDF-mediated repositioning of the integrase's coiled-coil subdomain (1) dictate which pairs of DNA sites can be assembled into active synaptic complexes and (2) dictate which product complexes will be conformationally locked, preventing back reactions. Critically, we find that the synaptic complex in which excision occurs is fundamentally different from that in which integration occurs. These insights provide a framework for engineering improved genome editing tools.

INTRODUCTION

Large serine integrases (LSIs; also called large serine recombinases) and their cognate recombination directionality factors (RDFs) mediate site-specific DNA recombination reactions that move the phages or mobile genetic elements that encode them into and out of host bacterial genomes.[1] Despite extensive biochemical and engineering studies, structural information has been limited, leaving many unanswered questions regarding how they recognize different attachment (*att*) sites, how their conformations shift along the reaction pathway, and, most importantly, how directionality is dictated.[2–7] The work presented here elucidates a unique and elegant system for control of recombination directionality of both the integration and excision reactions.

Several features make these systems useful as genetic tools.[8,9] The unidirectional integration reaction requires only integrase and two DNA sites: a ~50–54 bp phage DNA site (*attP*) and a ~40–46 bp bacterial site (*attB*).[10–12] The products are two hybrid sites (*attL* and *attR*) flanking the integrated phage DNA (Figure 1A). Binding of the RDF to the integrase triggers the reverse (excision) reaction and inhibits the integration reaction.[11–16] Unlike insertions mediated by DDE-family transposases or CRISPR-Cas, no DNA repair is needed after LSI-mediated reactions because all broken phosphodiester bonds are religated (Figure 1B). The large number of known integrases provides an array of sequence specificity options, and new tools for inserting *att* sites into heterologous genomes have streamlined the use of LSIs in genome editing.[17–24]

Serine recombinases share overlapping biological roles with tyrosine recombinases, but their sequences, structures, and mechanisms for DNA strand exchange and directionality control are unrelated. Some tyrosine recombinases such as Cre and Flp require only simple DNA sites but catalyze bidirectional reactions. Others such as phage lambda integrase catalyze unidirectional insertion but require a long *attP* site (240 bp) and host proteins.[25] Furthermore, their RDFs (xis) bind DNA as well as the integrase protein, whereas the RDFs for LSIs only bind the integrase protein.[25] Nevertheless, we show here that serine integrases share a fundamental concept with tyrosine integrases: the use of structurally different synaptic complexes for integrative and excisive recombination.

LSIs share a conserved catalytic domain and strand exchange mechanism with smaller serine-family recombinases such as resolvases and invertases.[1,26–28] Within a synaptic complex, which contains a recombinase tetramer and two partner DNAs, a catalytic serine residue from each subunit attacks a DNA phosphate, displacing a 3′ OH and generating double-strand breaks with 2-nt 3′ overhangs. In this intermediate state, each 5′ end is covalently linked to a protein subunit (Figures 1B and S1), and the two

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

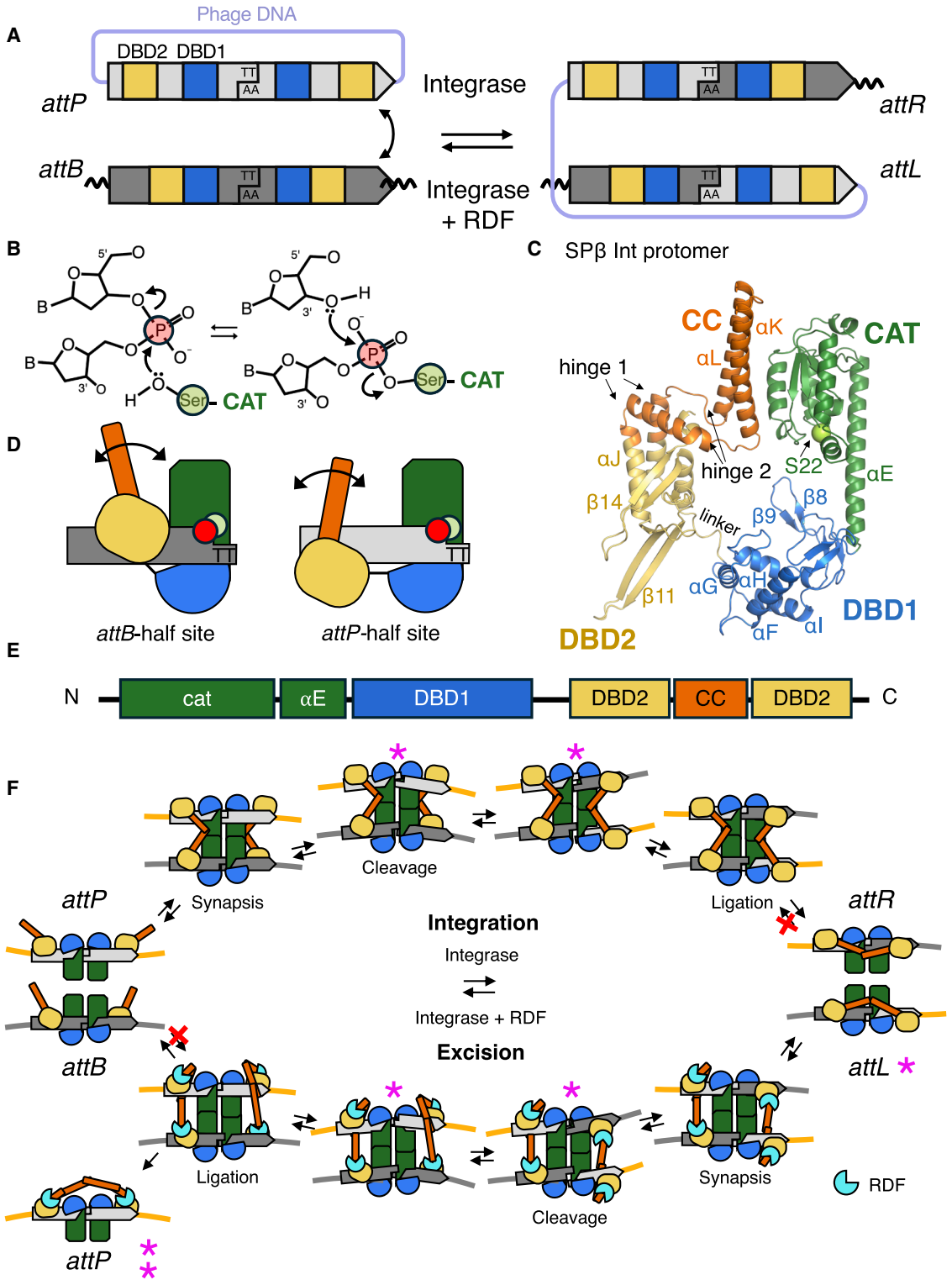


Figure 1. LSI-mediated DNA recombination

(A) Schematic reaction diagram. The integrase unidirectionally recombines a specific site in the phage (*attP*; light gray) with one in the host chromosome (*attB*; dark gray). After DNA cleavage (which creates central 2 nt overhangs), exchange of the right halves of the sites (as shown), and religation, the prophage is inserted within the host chromosome, flanked by hybrid sites (*attL* and *attR*). Motifs recognized by the integrase's two DNA-binding domains, shown in blue (DBD1) and yellow (DBD2), are spaced differently in *attP* and *attB*. The RDF protein triggers the integrase to catalyze the reverse reaction and inhibits the forward reaction.

(legend continued on next page)

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

CellPress
OPEN ACCESS

halves of the synaptic complex can swivel relative to one another about a central flat, hydrophobic interface to reposition the broken ends.[29] Following a 180° rotation, the chemical reverse of the cleavage reaction religates the DNA in the recombinant configuration.[29–31] However, if the 2-nt overhangs (the "central dinucleotides") do not base pair after 180° rotation, the complex can continue swiveling and eventually religate substrate DNAs.[32–35]

A key control point for serine recombinase regulation is the transition from inactive dimeric protein-DNA complexes to active tetrameric synaptic complexes. Interactions among catalytic domains are central to both oligomeric forms. However, tetramer formation entails large conformational changes that license catalytic activity and create the central flat interface about which rotation occurs. In the absence of additional tetramer-stabilizing factors, the dimer conformation is strongly favored.[36–38] The additional factors that control this conformational equilibrium vary widely among different subfamilies of serine recombinases.[1,27,39,40] For many small serine recombinases, the tetramer is stabilized by additional DNA-binding proteins that form a topologically defined complex.[41–45] As described below, for LSIs, the tetramer is stabilized by an additional domain of the integrase protein itself.

The forward progression of serine recombinase-catalyzed reactions is not driven by changes in chemical bond energy but rather by other energetic differences between the substrates and products.[26] For small serine resolvases, an overall negative ΔG is provided by the accompanying change in DNA topology.[26,46] The unidirectionality of LSI-catalyzed reactions does not rely on DNA topology and therefore must depend on differences between the substrate and product protein-DNA complexes. This is feasible in a thermodynamic framework because, unlike true catalysts that cannot change chemical equilibria, serine recombinases generally remain bound to their products and are therefore components of the reaction itself. However, a lack of 3D structural information has hampered our understanding of how LSIs apply this principle to accomplish directionality in general, as well as how RDFs reverse it.

The modular domain structure of LSIs facilitates stabilization of substrate-bound tetramers and control of reaction direction. All serine recombinases share a catalytic domain ("CAT"), followed by a long helix ("helix E") that is sometimes considered to be part of the catalytic domain. Small serine recombinases carry a single DNA-binding domain, usually at the C terminus, but at the N terminus for some serine transposases.[47] LSIs contain two DNA-binding domains that follow helix E, DBD1 and DBD2, also referred to as a recombinase domain (RD) and a zinc ribbon domain (ZD), respectively (Figures 1C–1E).[5,10] Inserted within DBD2, and flanked by flexible hinges, is a coiled coil (CC) with a hydrophobic tip. As described below, interactions between CC tips can mediate either activating inter-dimer interactions or inhibitory intra-dimer interactions, depending on the context—that is, depending on which DNA sites the proteins are bound to and the presence or absence of the RDF.[2,5,48–52]

Differentiating substrates from products requires differences in the partner DNA sites that alter the bound proteins' conformations. For LSIs, the integration substrates *attP* and *attB* are each functionally symmetric (except for their central dinucleotides) but differ from one another in length and sequence. The critical difference between *attP* and *attB* sites was deduced from the structure of the DBDs of LI integrase bound to half of an *attP* site.[48] The authors noted that the motifs recognized by DBD2 are 5 bp, or half a helical turn, closer to the center of *attB* sites than of *attP* sites.[5] Furthermore, they proposed that this positioning of DBD2s allows the CCs within them to (1) promote synapsis of *attP* and *attB* sites via inter-dimer "handshakes" between CCs and (2) after subunit rotation, swap interaction partners to lock the dimer-bound *attL* and *attR* products via intra-dimer "handcuffing" (Figure 1F, top half).

RDF binding was proposed to unlock the *attL*- and *attR*-bound dimers, allowing re-formation of an active synaptic complex.[5,51] In agreement with this idea, known RDFs share a common predicted binding site near the base of the CC motif of their cognate LSIs.[53–55] However, this model for RDF function lacks a structural underpinning and struggles to explain the unidirectionality of excision—that is, it does not explain why the RDF not only activates *attL* × *attR* recombination but also inhibits *attP* × *attB* recombination.

To understand the structural basis of directionality control, and the role of the RDF in the process, we determined 8 cryo-electronmicroscopy (cryo-EM) structures of *B. subtilis* phage SPβ integrase ("Int", also known as SprA) in the presence and absence of its cognate RDF (also known as SprB)[56] and with different *att* sites, at resolutions from 3.15 to 7.18 Å. These structures include synaptic tetramers for both the integrative and excisive pathways, as well as product dimers for both pathways. These are the first experimental views of LSI-RDF interactions. The synaptic tetramers were captured in the covalent intermediate state, when the complex was free to swivel and, for each pathway, both the pre- and post- strand exchange states were populated. Our results provide direct evidence to support and expand upon prior models for directionality control in

(B) The catalytic domain ("CAT"; green) is covalently linked to the 5′ ends of the cleaved DNA intermediate through a phosphoserine linkage. The phosphorus atom is indicated by a red circle, and Cα of Ser 22 is indicated by a yellow-green circle here and throughout the figures.
(C) Ribbon diagram of one SPβ integrase ("Int") monomer isolated from the synaptic complex.
(D) Cartoons of a LSI bound along an *attB*-half-site (left) and an *attP*-half-site (right). The arrow indicates mobility of the CC motif.
(E) Domain architecture of LSI, which consists of catalytic domain (green), DBD1 (blue), DBD2 (yellow), and a coiled-coil motif (CC; orange) that is inserted within DBD2.
(F) Pathways for integrative (upper) and excisive (lower) recombination. Magenta asterisks denote cryo-EM structures reported here. Inter-dimer CC-CC interactions stabilize the allowed synaptic complexes, while intra-dimer CC-CC interactions trap products, inhibiting the reverse reactions (red X's). The integration pathway follows that proposed by Rutherford and Van Duyne,[5] but the CC motifs are redrawn to match the experimental data presented here. In the excision pathway, structures described here show that the RDF redirects the trajectories of the CCs so that a different synaptic complex (between *attL* and *attR*) and a different product dimer (*attP*-bound) are stabilized.

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

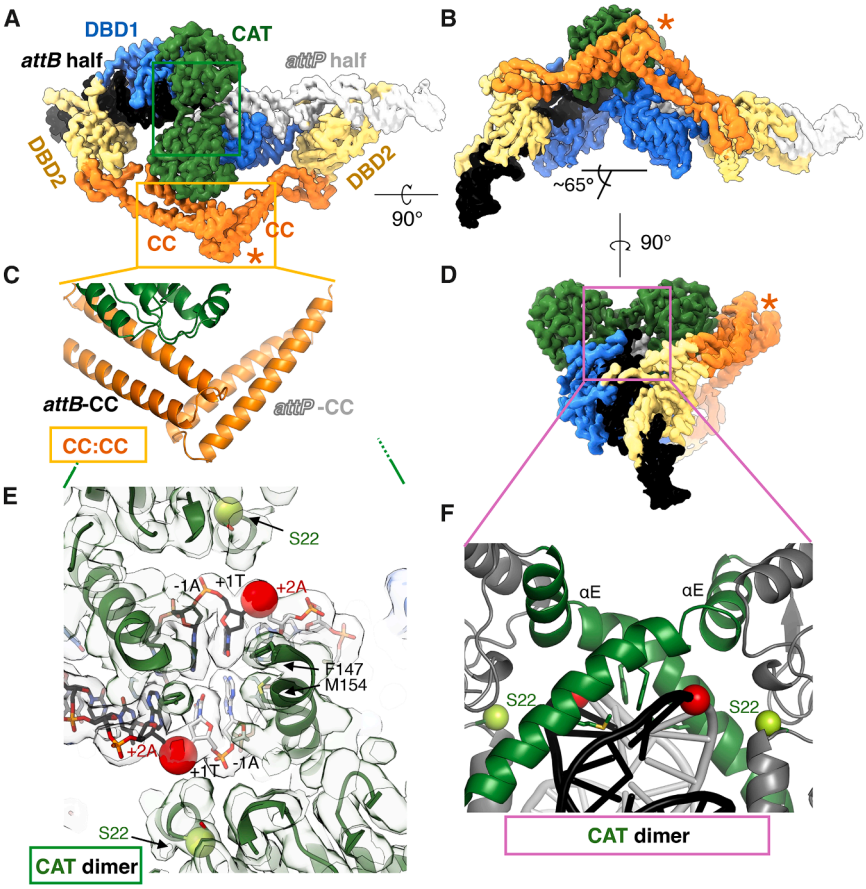


Figure 2. Overview of the *attLsym*-bound SPβ Int dimer

(A) Composite cryo-EM map of the complex, colored as in Figure 1. Orange asterisk highlights the CC-CC interactions.

(B) Orthogonal view of the same map as in (A).

(C) Ribbon diagram view of the handcuffing CC-CC interaction oriented as in (A) (orange box).

(D) The same map as in (A) and (B), from an orthogonal view.

(E) Close-up view near the central dinucleotide region, oriented as in (A) (green box), with the map in transparent gray. The catalytic S22s (yellow-green spheres) are distant from the scissile phosphates (red spheres). Residues F147 and M154 intercalate between bases +1T and +2A.

(F) The bent E-helices shown in close-up view oriented as in (D) (purple box). The proteins are gray except for the E-helices (green). F147 and M154 are shown as sticks.

LSI-mediated integrative recombination and provide a revised model for unidirectionality of the excision reaction.[2,49,52,57] We find that product dimers in both pathways, not just the integrative one, are handcuffed by CC-CC interactions (Figure 1F). Furthermore, we show that the prevailing model for RDF function is incorrect: rather than simply unlocking *attL*- and *attR*-bound dimers, RDF binding redirects the CC to dictate a very different synaptic complex for the excisive versus the integrative reaction.

RESULTS

Eight structures illuminate integrative and excisive recombination

To obtain structures along the integrative pathway, we used wild-type (WT) SPβ Int. For excisive pathway structures, we used the previously described method of fusing the RDF to the C terminus of Int with a flexible linker between the two,[58] which prevented loss of the small RDF protein. The DNA duplexes used carried the WT *att* sequences, except for the central dinucleotides, as described below. For most structures, dA_8 was added to the ends of *attP*-derived half-sites to facilitate distinguishing them from *attB*-derived half-sites. This "tag" proved unnecessary but was visible in the maps even though not contacted by the protein, presumably due to the rigid nature of A tracts.[59,60] We numbered all DNAs such that the central dinucleotides are −1 and +1.

For product-bound dimer complexes containing uncleaved *attL* or *attP*, the central dinucleotide was symmetrized (AT rather than AA, denoted as *attLsym* and *attPsym*; see STAR Methods for sequences). This facilitated biochemical experiments: either relative orientation of *attPsym* and *attBsym* can be recombined by the integrase, while two copies of *attLsym* can be recombined by the integrase-RDF fusion. The only functional difference between WT *attL* and *attR* sites is their asymmetric central dinucleotide (Figure S2).[61]

We also trapped synaptic complexes in the double-strand-break intermediate state in which Ser 22 of each subunit is covalently attached to a 5′ DNA end. To do so, we changed the central dinucleotide of both strands to TT (*attPmm*, *attBmm*, and *attLmm*; Figure S1), based on observations that serine recombinases can cleave but not religate substrates with central mismatches.[32] These samples yielded structures of the pre- and post-rotation states for both the integrative and the excisive synaptic complexes, as well as two unplanned additional structures, as described below.

The attL-bound product dimer: A 4 bp shift between half-sites allows CC-CC handcuffing

We focus first on the *attL*-bound Int dimer structure because that map extended to the highest resolution (3.15 Å) (Figures 2 and S3 and Table 1). This complex provides a model for both products of integrative recombination, because *attL* and *attR* each contain one *attB*- and one *attP*-derived half-site.

The primary difference between the *attB* and *attP* sides of the complex is that DBD2 lies 4 bp closer to the center on the *attB* half-sites. The overall architecture of the *attP* side of the complex resembles that of the LI integrase-*attP* half-site complex structure.[48] However, although a 5 bp shift of DBD2 between *attP* and *attB* sites was predicted for many LSIs,[3,5,51,54] the density

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

Table 1. Cryo-EM data collection, refinement, and validation statistics

	SPβ Int pre-rotation state (EMD-47288) (PDB 9DXH)	SPβ Int post-rotation state (EMD-47289) (PDB 9DXJ)	SPβ Int-*attPmm* tetramer (EMD-47290) (PDB 9DXK)	SPβ Int *attLsym* dimer (EMD-72552) (PDB 9Y66)	SPβ Int-RDF pre-rotation state (EMD-47284) (PDB 9DXD)	SPβ int-RDF post-rotation state (EMD-47286) (PDB 9DXF)	SPβ Int-RDF *attPmm* dimer (EMD-47287) (PDB 9DXG)	SPβ Int-RDF *attPsym* dimer (EMD-47284) (PDB 9Y6V)
Data collection and processing								
Magnification	81,000	81,000	81,000	81,000	81,000	81,000	81,000	81,000
Voltage (kV)	300	300	300	300	300	300	300	300
Electron exposure (e-/Å2)	65	65	65	65	65	65	65	65
Defocus range (μm)	−0.9 to −1.9	−0.9 to −1.9	−0.9 to −1.9	−0.9 to −1.9	−0.9 to −1.9	−0.9 to −1.9	−0.9 to −1.9	−0.9 to −1.9
Pixel size (Å)	1.065	1.065	1.065	1.065	1.065	1.065	1.065	1.065
Symmetry imposed	C1	C1	C1	C1	C1	C1	C1	C1
Initial particle images (no)	1,364,105	1,364,105	1,364,105	860,317	2,114,654	2,114,654	2,114,654	162,334
Final particle images (no)	13,917	27,865	92,812	342,692	550,169	248,916	54,766	24,616
Map resolution (Å)	7.18	6.92	4.81	3.3	4.82	5.00	6.61	6.9
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Refinement								
Model resolution (Å)	7.34	7.44	7.37	3.8	4.23	4.22	12.70	7.10
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Map sharpening B factor (Å2)	−368.8	−438.8	−229.6	−127.8	−321.1	−395.5	−505.3	−620.6
Non-hydrogen atoms	42438	42440	43689	21563	46399	46360	23758	23118
Protein residues	2134	2134	2140	1072	2366	2364	1182	1182
Nucleotide	232	232	268	124	232	232	134	114
Ligands	ZN: 4	ZN: 4	ZN: 4	ZN: 2	ZN: 4	ZN: 4	ZN: 2	ZN: 2
Protein	208.26	212.19	596.91	139.24	72.31	71.31	1333.16	617.12
Nucleotide	235.45	237.11	598.59	110.12	68.69	68.69	71.64	523.94
Bond lengths (Å)	0.006	0.005	0.006	0.004	0.005	0.005	0.004	0.003
Bond angles (°)	0.724	0.684	0.757	0.675	0.633	0.633	0.757	0.594

(*Continued on next page*)

CellPress
OPEN ACCESS

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

Table 1. *Continued*

	SPβ Int pre-rotation state (EMD-47288) (PDB 9DXH)	SPβ Int post-rotation state (EMD-47289) (PDB 9DXJ)	SPβ Int-*attPmm* tetramer (EMD-47290) (PDB 9DXK)	SPβ Int *attLsym* dimer (EMD-72552) (PDB 9Y66)	SPβ Int-RDF pre-rotation state (EMD-47284) (PDB 9DXD)	SPβ int-RDF post-rotation state (EMD-47286) (PDB 9DXF)	SPβ Int-RDF *attPmm* dimer (EMD-47287) (PDB 9DXG)	SPβ Int-RDF *attPsym* dimer (EMD-47284) (PDB 9Y6V)
MolProbity score	2.50	2.38	2.39	1.52	1.86	1.90	2.36	2.23
Clashscore	28.11	22.46	25.88	5.47	7.96	8.63	22.42	14.07
Poor rotamers (%)	0.00	0.00	0.00	0.20	0.00	0.00	0.00	1.94
Favored (%)	89.46	90.50	91.93	96.53	93.39	93.23	91.23	94.89
Allowed (%)	10.21	9.17	7.93	3.47	6.41	6.69	8.77	5.11
Disallowed (%)	0.33	0.33	0.14	0.00	0.00	0.09	0.00	0.00

map was clear, and the outer portions of the SPβ binding site sequences align better if a 4 bp shift is assumed (Figure S4). Because of this shift, DBD2 of the *attP*-bound subunit docks onto same side of the DNA as DBD1, whereas for the *attB*-bound subunit it docks onto the opposite side. Furthermore, the two copies of DBD2 are positioned on the same face of the complex.

The *attL*-bound dimer map clearly shows the product-stabilizing intra-dimer CC-CC interaction (Figures 2 and S3) first proposed by Rutherford and Van Duyne.[5,51] This interaction is facilitated by the relative placement of the two DBD2s, which the CCs are inserted into. The DBD2-CC junctions are flexible, but the CCs themselves are well ordered. As seen in structures of the isolated CCs of LI integrase, the two tips interact in a front-to-back manner, rather than a 2-fold-symmetric manner (Figures S5A and S5B).[2] The *attB*-bound subunit's CC is sandwiched between that subunit's catalytic domain and the *attP*-bound subunit's CC (Figures 2 and S5B), an interaction that was unexpected.

The *attL*-bound dimer is clearly in an inactive state because the two active site serines (S22) are 9–10 Å from the phosphorus atoms that they would attack in a DNA cleavage reaction (Figures 2E and 2F). The E-helices are bent and interact with one another in a largely antiparallel fashion that differs from the more nearly parallel interaction seen in DNA-bound dimers of small serine recombinases.[36,43,44]

The DNA in this complex is bent at the center by ~65°, reflecting two kinks flanking the central dinucleotide with large roll angles between adjacent base pairs (Figures 2B, 2E, 2F, and 3). The C-terminal portion of helix E, which connects the catalytic domain to DBD1, docks into the minor groove, widens it, and inserts the hydrophobic side chains of F147 and M154 into the kink. Similar groove-widening by helix E, sometimes including hydrophobic intercalation, was reported for small serine recombinases.[43,44] These interactions likely contribute to indirect readout of the DNA sequence near the cleavage sites, as they do in LI integrase.[57] These helix E-DNA interactions are similar on both sides of the *attL*-bound dimer but are not maintained in the complexes with cleaved DNA reported below.

Partially modular protein-DNA interactions

Details of the Int—DNA interactions are described for the *attL*-bound dimer because comparison with additional structures showed that most of the Int-DNA interactions depend only on whether a subunit is bound to an *attP*- versus an *attB*-derived half-site, rather than on the larger context of the complex (e.g., cleaved versus uncleaved, +/−RDF or pre- versus post-rotation). The sole exception is helix E, which interacts differently with cleaved versus uncleaved DNA. Four different protein segments interact closely with the DNA: the C-terminal portion of helix E, DBD1, the linker between DBDs, and DBD2 (Figures 3A–3E and S6; Video S1). Int-DNA interactions are only partially modular: while most local interactions are similar on *attB* and *attP* half-sites, the outer portion of DBD1 and the interdomain linker interact differently with *attB* versus *attP* DNA.

The inner portion of DBD1 binds base pairs 4–7 of both types of half-sites. The identity of these bases is conserved among all four half-sites (Figure S4). Of particular interest is the long β-hairpin loop between αH and αI that makes multiple contacts with the phosphodiester backbone, inserts Y254 into the minor

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

CellPress
OPEN ACCESS

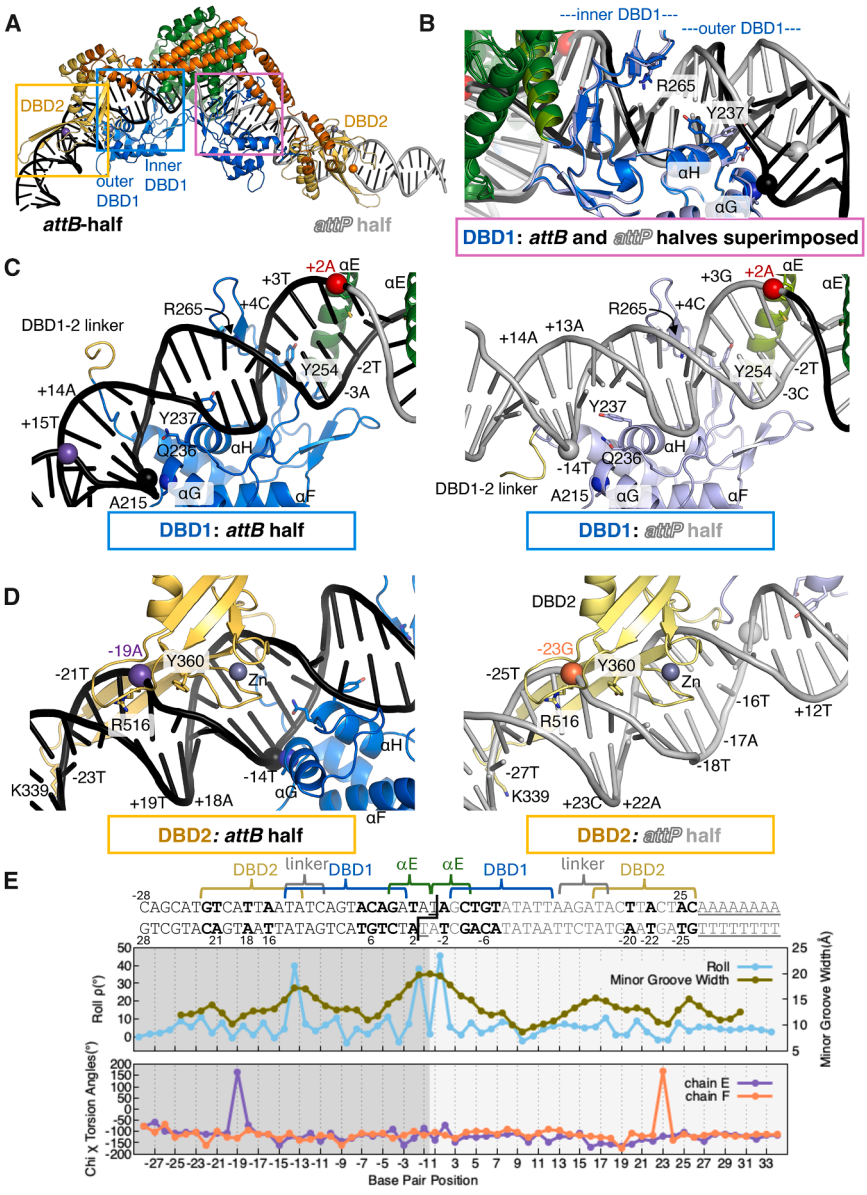


Figure 3. Partially modular protein-DNA interactions

(A) Ribbon diagram of the SPβ Int dimer-*attLsym* complex oriented as in Figure 2B. Boxes indicate close-up views in parts (B)–(D).

(B) Close-up view of the DBD1 interactions. The *attP* side (lighter blue and gray) is oriented as in (A) (purple box), and the *attB*-side DBD1 and DNA (darker blue and black) are superimposed on it. Due to the divergence in DNA paths, Y237 adopts different rotamers, and the backbone NH of A215 (blue sphere at the end of helix G) only interacts with the phosphate of T-14 on the *attB*-side (black sphere versus gray sphere).

(C) Close-up views of DBD1 and DNA on both sides, oriented as in (A) (blue box).

(D) Close-up views of DBD2 and DNA on both sides, oriented as in (A) (yellow box). Binding of the DBD2 domains induces an unusual DNA kink at nucleotides −19/−23, indicated by spheres in colors corresponding to χ angle plot in (E). See Figure S6 for details of the DBD1-DBD2 linker-DNA interactions.

(E) Sequence and structural parameters of *attL-sym*. Underlined nucleotides differ from the WT *attL* sequence, and bold nucleotides are identical in all four half-sites (see alignment in Figure S4). Interacting protein domains are noted above the sequence, and structural parameters are plotted below (roll angle between successive base pairs, minor groove widths, and χ torsion angles). DBD2 is 4 bp further from the center on the *attP* side versus the *attB* side.

groove and R265 into the major groove, where it makes a sequence-specific bidentate interaction with the conserved G6. The flat guanidinium group of R265 is positioned to stack against T5 with its protruding methyl, forming an interaction similar to that seen in the TG- and 5-methyl-CG specific interactions of other proteins.[62] This loop varies widely among LSIs, and its major groove-interacting tip is absent in LI integrase, where DBD1 lacks significant sequence specificity.[57]

The outer portion of DBD1 interacts differently with the two types of half-sites. It contains a helix-turn-helix motif (αF, αG, and αH). Even though such motifs often display strong preferences for specific DNA sequences, the bases that this one contacts are variable. Helix H docks into the major groove, but only Q236 and Y237 extend toward the unique edges of the bases themselves. The tip of Q236 is within H-bonding distance of a base on the *attB* side but not on the *attP* side, and the side chain of Y237 adopts different rotamers and contacts different bases (Figures 3B and 3C). On the *attB* site, a kink between base pairs 13 and 14 facilitates H-bonding between a backbone amide at the N terminus of αG and phosphate −14 (blue and black spheres, respectively, in Figures 3B and 3C). On the *attP* side, the major groove is wider at this point, the minor groove narrower, and the phosphodiester backbone is pulled away from αG. This correlates with a more AT-rich sequence in *attP* than *attB* (Figure 3E): AT-rich sequences intrinsically favor a narrowed minor groove.[63]

The DBD1-DBD2 linker also interacts differently with the two types of half-sites (Figure S6). On the *attB* side of the complex, L302 and L303 of the linker partially intercalate between base pairs 13 and 14, inducing and/or stabilizing a nearly 40° roll angle that greatly widens the minor groove (Figures 3E and S6A). The linker also interacts with a widened minor groove on the *attP* side, but the DNA is more smoothly bent by only ∼ 30°. Furthermore, rather than L302 and L303, residues P304, K306, and F309 interact most closely with the floor of the minor groove (Figure S6B). On the shorter *attB* side, the linker-binding region overlaps with the outer portion of the DBD1 binding site.

The local interactions between DBD2 and DNA are similar on both types of *att* sites despite the overall 4 bp shift (Figure 3D).

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

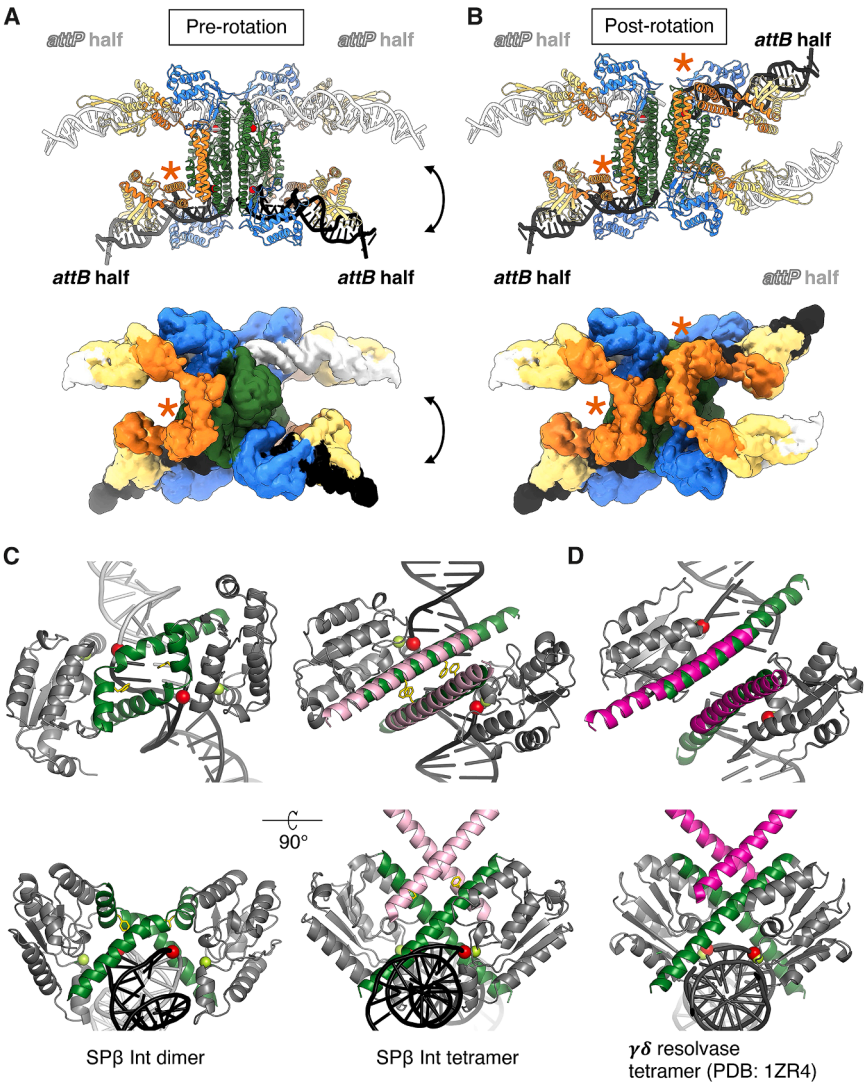


Figure 4. Integrative synaptic complex structures and comparison of E-helices in dimeric and synaptic complexes

(A) Ribbon diagram (top) and cryo-EM map (bottom) of the SPβ Int *attP*:*attB* synaptic complex in the pre-rotation state, captured in the covalent intermediate state. Colors are as in previous figures, and orange asterisks mark approximately where CC-CC interactions occur. See also the "side" view in Figure 5C.

(B) Ribbon diagram (top) and cryo-EM map (bottom) of the same complex in the post-rotation state. The entire right half of this complex has rotated ~180° relative to the one in part (A).

(C) Comparison of the E-helices in the dimeric (left column) and synaptic complexes (right column) in the top (top row) and side (bottom row) views. The dimer's two E-helices are green; the additional two E-helices from the tetramer are pink. The remainder of the dimer's CAT domains are shown in gray. F138, which repacks between forms, is shown as yellow sticks. Spheres indicate the scissile phosphates (red) and catalytic S22 residues (yellow-green).

(D) Ribbon diagram of the γδ resolvase tetramer[29] (PDB: 1ZR4; also in the covalent intermediate state) shown in the same orientations and color scheme as in (C).

Direct H-bonding interactions between protein side chains and DNA bases are sparse, implying that sequence recognition may rely heavily on indirect readout of sequence-dependent variations in the structure and flexibility of the DNA. The most prominent DNA-binding feature of DBD2 is a long beta hairpin that lies in the major groove and makes multiple contacts with the phosphodiester backbone, ending with K339 contacting the 5′ phosphate of T-23 (*attB* side)/T-27 (attP side). This hairpin is unusually long in SPβ Int (Figure S7) and, therefore, its cognate *attB* and *attP* sites are longer than those for many LSIs: 46 and 54 bp, respectively. A second beta hairpin places R516 in the minor groove, where it donates H-bonds to two consecutive bases. These contacts by themselves are not sequence-specific because A, T, C, and G all have an H-bond acceptor in a similar location. However, together with the other beta hairpin and Y360, they form part of a grip across the DNA backbone that induces an unusual kink centered on A-19 (*attB* side)/G-23 (*attP* side). Integrase activity depends not only on DNA affinity but on distinguishing *attP* from *attB*,[64] which in turn depends on proper placement of DBD2. Figure S4C shows that placement of DBD2 probably depends not only on the presence of conserved sequence motifs in the correct place but also on their absence from the wrong place.

The integrative synaptic complex trapped in the swiveling state

Integrative synaptic complexes were formed by mixing SPβ Int with *attPmm* and *attBmm*. These substrates were paired and cleaved by Int, but not religated, and therefore trapped the complex in the swiveling-competent covalent intermediate state. Two classes of particles contain Int tetramers along the integrative reaction pathway: one in the pre-rotation cleaved-substrate conformation, synapsing an *attP* and an *attB* site, and one in the post-rotation not-yet-ligated-product state in which one half of the complex is rotated by ~180° relative to the other (Figure 4). While there are no covalent bonds between the two half-complexes to restrain rotation, there are likely energy wells at the 0° and 180° positions that are deep enough to favor these positions but shallow enough to allow rotation driven by thermal energy. The global resolutions for these maps were 7.18 and 6.92 Å, respectively, and local resolution extended to 4.69 Å (Figure S8).

Helix E is straight in these complexes and forms the central platform for rotation, as seen previously for small serine recombinases in the active tetrameric form (Figure 4).[29,65] Although most of the Int-DNA interactions with individual half-sites are similar to those seen in the *attL*-Int dimer complex, helix E and the catalytic domains form very different interactions with the DNA and with

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

CellPress
OPEN ACCESS

each other. Helix E is still docked into the minor groove, but due to DNA cleavage the DNA kinking and intercalation of the F147 and M154 side chains is no longer seen (Figure S9).

The integrative synaptic complex is stabilized by inter-dimer CC-CC interactions

In these synaptic complexes, the CC motifs, which are clearly visible in the maps, form inter- rather than intra-dimer interactions. The two cleaved DNAs lie on the outside of the synaptic complexes and are held together not only by tetramerization of the catalytic domains but also by inter-dimer handshakes between the CC motifs (Figures 4 and S8). The CC of each *attP*-bound subunit is sandwiched between the catalytic domain and the CC of its *attB*-bound partner. These tetramer-buttressing interactions occur on opposite sides of the pre-rotation complex but on the same side of the post-rotation complex. As discussed below, this geometry prevents formation of handcuffing intra-dimer CC-CC interactions before rotation and facilitates them after rotation.

An unexpected third class of particles contained a tetramer of Int synapsing two *attP* sites, yielding a 4.81 Å map (Figures S8 and S10). The presence of multiple *attP* sites in a cell is unlikely in a biologically relevant context, and these tetramers are unlikely to allow strand exchange because the CC-CC interactions occur diagonally, which is incompatible with rotation. We confirmed that SPβ Int-*attPmm* complexes do elute as tetramers in size exclusion chromatography (SEC) (Figure S11). Similar observations were recently reported for φC31 integrase.[66]

Conformational changes may be needed after rotation to bring the 3′ OH groups close enough for religation and to fully assemble the active site for catalysis. In all the synaptic complexes described here, the two halves of each cleaved DNA are roughly co-linear, but the phosphoserines are ∼24 Å apart (measured between P atoms), ∼10 Å further than they would be in intact B-form DNA. Furthermore, the catalytically important conserved arginines are not well ordered in these structures.[67,68] The proposed conformational changes may involve rigid-body motion of the DNA half-sites as well as adjustments within the catalytic domains. The *attB* half-sites in our synaptic complexes tilt away from the catalytic domains, as does the DNA in most serine recombinase complexes,[28,29,36,43,44] but the *attP* half-sites tilt toward the catalytic domains. This may reflect an intrinsic feature of *attP*, or it may reflect the inter-dimer CC-CC interactions pulling the *attP* DNA toward its *attB* partner. 3D variability analysis and 3D flexible reconstruction showed some flexibility in the relative orientation of the DNAs (Video S2).

The RDF dictates a different synaptic complex for excision

Three structures were determined from samples of the SPβ Int-RDF fusion protein in complex with *attLmm*, also trapped in the covalent intermediate state (Figures 5A, 5B, and S12): synaptic tetramers in the pre- and post-rotation states and a smaller class of particles containing dimers bound to unligated product *attP* sites. We could not identify *attB*-bound dimers. The global resolution was 4.82 Å for the pre-rotation state (with local resolution extending to 4.16 Å), 5.00 Å for the rotated state, and 7.1 Å for the *attP*-bound dimer.

Comparison of the integrative and excisive synaptic complexes shows a major and unexpected difference: the CC-CC interactions that stabilize them occur on opposite sides (Figure 5C versus 5D; see Figure S13 for maps). Nevertheless, the local interactions between the tips of the CCs are quite similar in both cases, except that "front" versus "back" is switched—that is, without the RDF, the front of each *attB*-bound subunit's CC interacts with the back of an *attP*-bound subunit's CC, whereas with the RDF, the backs of the *attB* CCs interact with the fronts of the *attP* CCs (Figure S5). Overall, RDF binding to Int does not simply disrupt the handcuffed *attL*- and *attR*-bound dimers that are the product of integration—it stabilizes a very different trajectory of the CC and promotes the formation of a different synaptic complex for excision.

The RDF changes the pivot point for CC flexibility

The SPβ RDF is a small protein that binds both DBD2 and the DBD2-proximal portion of the CC, tying them together (Figure 5D; Video S3). Comparison of all the DBD2-CC conformations seen with and without the RDF shows that there are two hinges between DBD2 and the CC tip (Figure S14): hinge 1, located at the CC-DBD2 junction and hinge 2 further along the CC. Within each set of structures (−RDF versus +RDF), hinge 1 remains rigid, while hinge 2 serves as the pivot point for CC motion. RDF binding to the inter-hinge region changes the conformation of hinge 1, which repositions hinge 2 and consequently changes the spatial range accessible to the CC tips. In general, in the absence of the RDF, hinge 2 lies in front of or behind the DNA, but in the presence of the RDF, it lies above or below the DNA (Figures 5C, 5D, S15, and S16). In *attL*- and *attR*-bound Int-RDF substrate complexes, the RDF-stabilized position of hinge 2 would prevent intra-dimer CC-CC handcuffing. However, within the synaptic tetramer, the subunit rotation that converts substrates to products changes the relative positioning of the *attP*-bound subunits' hinge 2s from diagonally opposed across the tetramer to close proximity, which facilitates handcuffing of the *attP*-bound product (see below). Although we could not determine an *attB*-bound product structure, Figure S17 suggests that it is probably also handcuffed.

We also noted that the RDF appears to destabilize the integration-competent trajectory of the CC while stabilizing the excision-competent trajectory (Figure S17). In the absence of the RDF, the helical inter-hinge segment (amino acids [aa] 381–397) packs against DBD2 in the same location where a loop of the RDF binds, effectively shielding a small hydrophobic region. This shielded pocket includes the extreme C-terminal region of Int (aa 530–545), notably F532, which is essential for catalyzing excision reaction but largely dispensable for integration.[69] Additionally, the CC-proximal ends of helices J (aa 376–379) and M (aa 474–484) within DBD2 are unwound in the presence of the RDF and form part of the inter-hinge region. Both the experimental maps and AlphaFold models support this helical extension in the absence of the RDF. The endpoints

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

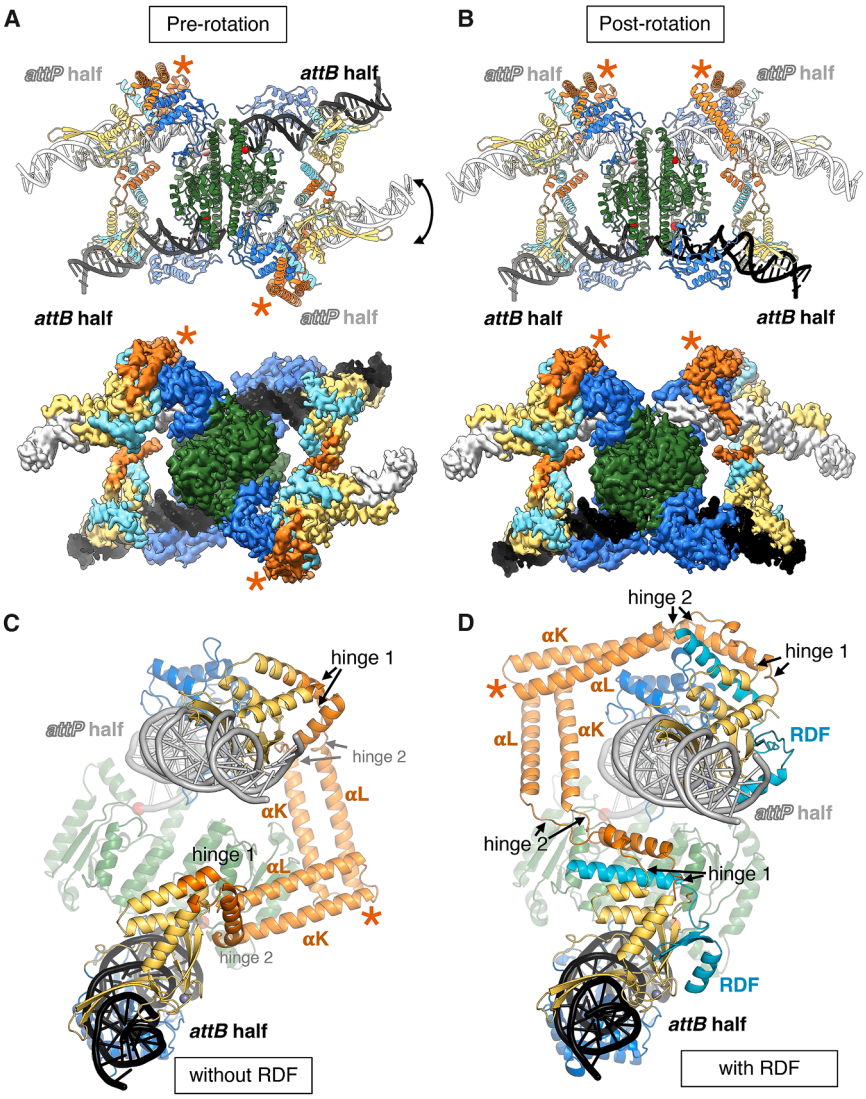


Figure 5. The RDF redirects the CC subdomain, dictating a different synaptic complex for excision

(A) Ribbon diagram (top) and cryo-EM map (bottom) of the SPβ Int-RDF fusion synaptic complex in the pre-rotation state, captured in the covalent intermediate state. Structures are colored as in previous figures, with the RDF in turquoise. The orange asterisks indicate where the CC tips from the subunits bound to the upper and lower DNAs meet (see also the side view in D).

(B) Ribbon diagram (top) and cryo-EM map (bottom) of the same complex in the post-rotation state.

(C) DNA-end view of half of an integrative synaptic complex in the absence of the RDF, colored as in previous figures. The orange asterisk marks where two CC motifs meet.

(D) A similar view of half an excisive synaptic complex in the presence of the RDF. See Figure S13 for experimental maps.

of these two helices are also variable among the 4 copies in the asymmetric unit of the crystal structure of the LI integrase DNA-binding domains.[48]

RDF binding displaces the inter-hinge helical segment (aa 381–397) by inserting its C-terminal helix and the loop that precedes it into the DBD2 hydrophobic region. This displacement destabilizes the helix M extension, exposing hydrophobic residues such as I472 and V477 to form new contacts with F40 of the RDF, thereby stabilizing the repositioned hinge 2 conformation. Concurrently, repositioning of the integrase C terminus allows F532 and F534 to engage in close hydrophobic contacts with RDF residues including L37 and F40 (with additional support from Y41 and F44), stabilizing the new conformation of hinge 1. These interactions between DBD2 and the RDF align with functional data[69] from Abe et al., who described the Int C terminus as a potential component of a ''molecular toggle switch'' controlling recombination directionality.

CC-CC handcuffing of the attP-bound integrase-RDF dimer

Although a low-resolution structure of an *attP*-bound dimer structure was derived from the *attLmm*-SPβ Int-RDF sample, the central mismatches trapped it in the cleaved state rather than the final religated state. We therefore directly formed product dimer complexes with uncleaved DNA by mixing the SPβ Int-RDF fusion protein with *attPsym* and *attBsym* substrates. Under these conditions, only an *attP*-bound dimer species was observed (Figure S18), which yielded a structure at an overall resolution of 7.1 Å. No CC-CAT interactions are seen in either of these dimers.

In both *attP*-bound dimers, the two copies of hinge 2 lie on the top face of *attP*, placing the CCs and their tips in close proximity and enabling intra-dimer CC-CC interactions that likely drive or stabilize tetramer-to-dimer dissociation (Figure 6A). Figure 6B compares the relevant half of a post-rotation tetramer in both reaction pathways to the final product dimer. In both cases, hinge 2 is placed to facilitate handcuffing of the product.

Although both the *attLmm*-derived *attP* dimer and *attPsym*-bound dimer bind essentially identical *attP* half-sites, they reveal markedly different degrees of CAT domain ordering and positioning: the CAT domains are much better ordered when the DNA is uncleaved (*attPsym*, Figures 6Aand S18) than when it is cleaved (*attLmm-derived*, Figures 6A and S12). In the best-resolved 3D class, the dimer formed by the two CAT domains closely resembles the unusual (for serine recombinases) CAT domain dimer seen in the handcuffed *attLsym*-bound dimer in the absence of the RDF (Figure 6B). Thus, while the dimer containing cleaved DNA samples a highly dynamic ensemble of CAT

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

CellPress
OPEN ACCESS

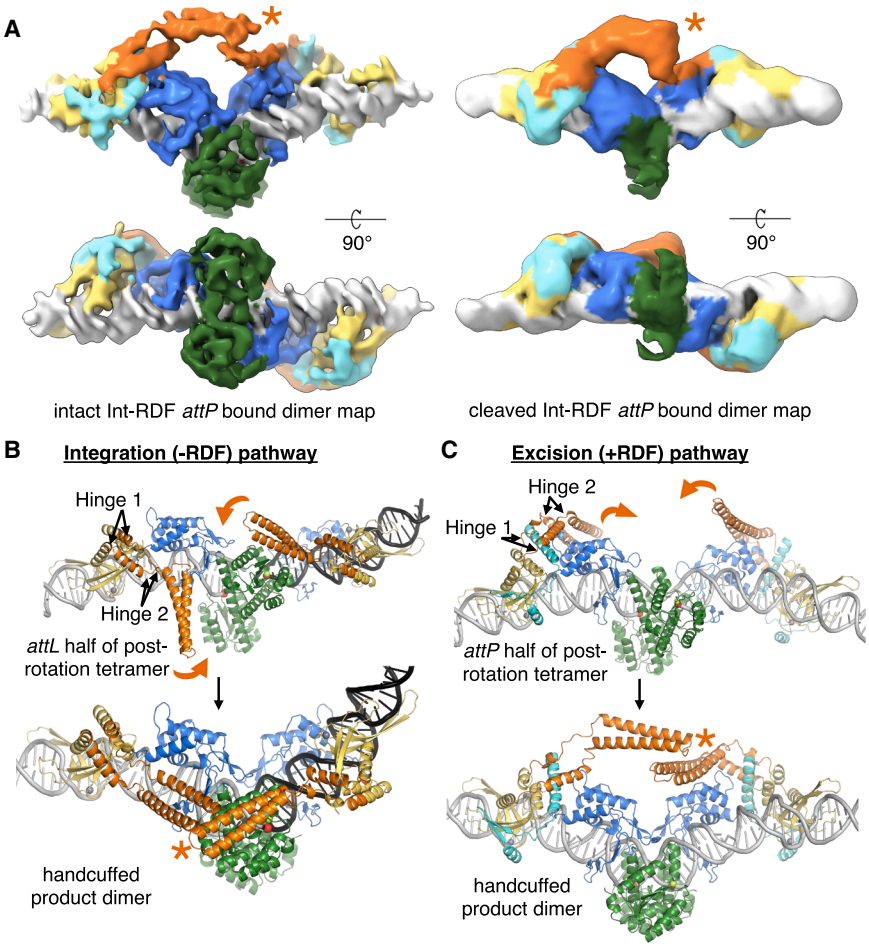


Figure 6. CC:CC handcuffing traps product dimers for both pathways

(A) Orthogonal views of the cryo-EM maps of SPβ Int-RDF dimers bound to intact (left column) and cleaved (right column) *attP* DNA.

(B) Comparison of one half of the post-rotation integrative synaptic complex (top) with the corresponding handcuffed product dimer (bottom). Arrows show the CC motions presumed to occur in the transition between states. See Figure S15 for more detail.

(C) Similar comparison of one half of the post-rotation excisive synaptic complex (top) with the corresponding handcuffed product dimer (bottom). See Figure S16 for more detail.

conformations, the dimer containing uncleaved DNA is biased toward an open, E-helix bent, and catalytically inactive state.

Mutational analysis reveals the RDF-CC interface is essential for redirecting CC trajectory

We employed an *E. coli*-based promoter inversion assay to investigate the functional importance of Int-RDF interactions (Figure S19).[53,70] Int flips a promoter flanked by *attP* and *attB* sites to drive GFP rather than RFP expression, while the Int-RDF fusion acts on the *attL* and *attR* sites in the product plasmid, flipping the promoter back to RFP expression. Use of the Int-RDF fusion avoids artifacts due to unequal protein expression or weakened protein-protein affinity. Figure S19B shows that the WT proteins drove recombination to completion after only 2 h of expression. Deletion of the N-terminal helix of the RDF (Int-RDF ΔN-ter helix 1–13), which interacts only with DBD2, had no effect on RDF function in this assay. In contrast, the RDF's C-terminal helix (Int-RDF ΔC-ter helix 37–58), which packs between DBD2 and the inter-hinge region of the CC, is essential: when this helix was deleted, the fusion protein acted as a slow version of WT Int without RDF, eventually converting all substrates to the *attL*/att*R* configuration. Similar results were obtained for two less drastic changes in the C-terminal helix: pairwise mutations L37D/F40D and A45D/L49D. The first pair flanks Int F532, which was previously shown to be important for directionality control[69] as discussed above, and the second pair interacts directly with the inter-hinge segment. These data support the hypothesis that restraining the inter-hinge region of the CC is critical to RDF function.

We also tested the effects of lengthening the hinges in Int. Hinge 1 was lengthened by inserting GGSGSSG into both the outgoing and returning chains (Figure S19C), and hinge 2 was similarly lengthened by inserting GGSGSG (see Figure S19C). The effects of elongating hinge 1 were minor: both integrative and excisive reactions were slower, and the Int-RDF fusion was less cleanly switched—excisive recombination was still favored but not as decisively. Elongating hinge 2 had no detectable effect on Int or Int-RDF function in our assays. Although elongating the CC of LI integrase by ~two helical turns weakened recombination efficiency and directionality,[3] our results suggest that the most important function of the RDF is to change the pivot point of the CC rather than simply to restrain how far it can extend.

Additional protein-protein interactions

The SPβ synaptic complexes include additional protein-protein interactions that were not predicted from partial structures and previous modeling (Figure S20). In all tetramers, E258 of the long DNA-binding loop of one subunit's DBD1 could form a salt bridge with the catalytic domain of the subunit bound to the DNA on the opposite side of the synapse, potentially helping to stabilize the synaptic complex. Contacts between the CC and CAT domains occur only in the absence of the RDF. The details vary between the tetramer and dimer states, but P54 (CAT domain) and F426 (CC) are included in both sets of contacts. Interestingly, that region of the catalytic domain is involved in activating contacts from other proteins or DNAs in small serine recombinases.[27] In contrast, only when the RDF is present, the CC from the *attP*-bound subunits contacts a small hydrophobic patch on DBD1 of the same subunit. Similar CC-DBD1 contacts

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

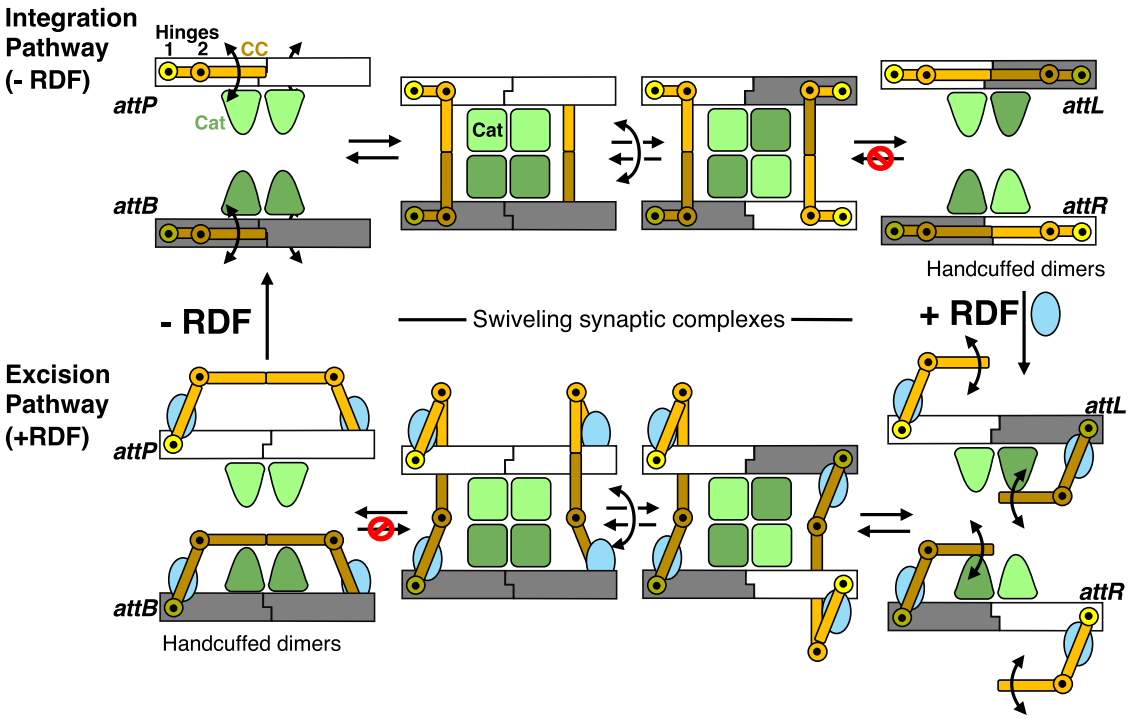


Figure 7. Schematic model of how CC-CC interactions and the RDF mediate directionality control

Only the catalytic domain (green), CC (orange), hinges (black dots in yellow and orange circles), RDF (blue), and DNA (*attP* in white and *attB* in gray) are shown. DNA-binding domains are not shown for simplicity. Both hinge 1 and hinge 2 are critical to directionality. The location of hinge 1 depends only on whether the protein is bound to an *attP*- or *attB*-derived half-site. The RDF causes rotation about hinge 1, which repositions hinge 2. Within each pathway, the pivot point for the CC is hinge 2.

are made by both *attP*-bound subunits within the tetramer and by one subunit within the *attP*-bound dimer.

We tested the effects of mutating these contacts in a promoter-flipping assay (Figure S20). E258A decreased excision but did not detectably affect integration, perhaps because the CC-CAT contacts formed in the absence of the RDF also stabilize CAT domain positioning. Because F426 also participates in CC-CC contacts, we could only test the importance of the integration-specific CC-CAT contacts by mutating the CAT domain. A triple mutation was deleterious in both reaction directions, suggesting that it affected catalytic activity, but it did not completely ablate excision. Most mutations of the excision-specific DBD1-CC contacts affected excision more strongly than integration, in agreement with those contacts only occurring in the presence of the RDF. Our results support the idea that some mutations may affect the two reaction pathways differently.

DISCUSSION

Biology often requires that site-specific DNA recombination reactions proceed to completion even when the net number of phosphodiester bonds does not change. Multiple solutions to this conundrum have evolved. For tyrosine-family integrases, additional DNA-binding proteins set up a synaptic complex from which recombination is downhill in terms of conformational energy.[25] The XerC/XerD tyrosine recombinase pair cooperates with the FtsK translocase to resolve chromosome dimers or, to resolve plasmid dimers, exploits additional DNA-binding proteins and supercoiling.[71] Small serine recombinases that resolve replicon fusions also employ additional DNA-binding proteins to sense and be driven by DNA topology.[46] In contrast, unidirectional recombination by LSIs requires neither a particular DNA topology nor additional proteins.

Understanding the control of reaction directionality for the LSIs has been a long-standing challenge. Our series of SPβ Int and Int-RDF complex structures can be assembled into a movie of an LSI "in action" (Video S4). This work provides direct evidence for an updated model, diagrammed in Figures 1F and 7, that agrees with current models for integrative recombination[51] but fundamentally changes the model for excisive recombination. Unidirectionality in both pathways is explained by the interplay among 3 key features: the CC, the RDF, and the difference between *attP*- and *attB*-half-sites. These control the oligomeric state of the integrase and consequently the conformation and activity of the catalytic domain.

For both reaction pathways, the CC-CC interactions stabilize the synaptic complex, then stabilize or "handcuff" at least one, if not both, product dimers. Dimer handcuffing prevents reaction products (*attL* and *attR* for integration and *attP* and *attB* for excision) from forming synaptic tetramers, thus preventing the back reaction. Integration and excision are driven forward by similar mechanistic logic but occur within structurally different

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

complexes. In the initial substrate state of both types of synaptic complex, the critical 2nd hinges that are the CCs' pivot points are located such that the CC tips can only interact with subunits bound to a partner duplex, thus stabilizing synapsis. However, rotation to the product state brings the 2nd hinges of subunits bound to a single product duplex close enough that they can swap interaction partners to handcuff the product dimer.

The RDF not only unlocks the handcuffing of the *attL*- and *attR*-bound dimers that are the products of integrative recombination (as previously proposed) but also changes the location of the CCs' second hinges. The RDF therefore favors a synaptic complex with different geometry from that used in integrative recombination and allows trapping of the *attP*-bound dimers (and probably *attB*-bound dimers) that are the products of excisive recombination. The use of structurally different synaptic complexes for integration versus excision is conceptually similar to the regulation of phage lambda integrase (a tyrosine recombinase) even though the two systems differ wildly in mechanism and complexity.[25,26]

The work presented here provides a framework for engineering LSI-RDF pairs for use as genetic switches and genome editing tools. We suggest that optimization of reaction efficiency in a particular pathway could be streamlined by focusing on protein segments that adopt different interactions and/or conformations in the excisive versus integrative pathways. The sequences of those segments probably reflect evolutionary compromises between pathways—not only for efficiency in both directions but also for inhibition of the "wrong" reaction in both directions. They are therefore likely hotspots for mutations that optimize reaction efficiency in one direction without regard for the other, as seen in the clustering of efficiency mutations in CC region reported in[72] and for mutations that loosen directionality control, as seen in CC region in Rowley et al.[52]

Integrase variants with altered sequence specificity are also of interest to the engineering community. Two features of SPβ Int that are of particular interest are the β8-9 hairpin extending from the inner portion of DBD1 and the β11-12 hairpin of DBD2, which dock into the major groove and contact conserved sequence motifs. Furthermore, these hairpins are likely to be amenable to mutation because they protrude from the core DBD folds. Variability in hairpin length among LSIs (Figure S7) may contribute to differences in the stringency of sequence specificity and in overall *att* site length. For example, SPβ Int-DNA contacts extend across 46 and 54 bp of *attB* and *attP*, respectively—longer than is often assumed to be sufficient for LSI binding, while DBD1 of LI integrase, which has a short β8-9 hairpin, has little sequence specificity.[57]

Integrative-pathway structures have recently been reported for Bxb1 and φC31 integrases,[66,73] which confirm that, like SPβ Int, they follow the generalized model for integrative recombination proposed by Rutherford and Van Duyne, in which CC-CC interactions stabilize the initial synaptic tetramer. For φC31, RDF-bound structures and handcuffed product dimers were also reported. Although the φC31 RDF is structurally very different from the SPβ RDF, its mechanism is conceptually similar: it binds to the inter-hinge region of the CC, dictates formation of a different synaptic complex between *attL* and *attR*, and allows handcuffing of an excision product dimer. Furthermore, in the SPβ and φC31 integration product dimers, the CCs contact different regions of the catalytic domains. Structural differences among dimers are permissible because they are not the catalytically active species, allowing for conformational variation without compromising the overall recombination pathway.

Beyond flexibility within the protein framework, SPβ Int also differs slightly from φC31 and Bxb1 integrases in that DBD2 is shifted by 4 rather than 5 bp on *attP* versus *attB* half-sites. Similar variations probably occur for other LSIs, given the variability in DBD1-2 linkers, inter-hinge regions and predicted CC length (e.g., Figure S7). Functionality only requires that *attP* can be differentiated from *attB* and that the appropriate CC-CC interactions can form throughout both recombination pathways. These observations suggest that bioinformatic analyses of *att* sites should consider the shift in DBD2-binding motifs as a variable.

Additional differences between synaptic tetramer structures for φC31 and SPβ integrases may reflect different reaction stages. For φC31, an active site mutant was used to trap the pre-DNA-cleavage state, whereas for SPβ mismatched central dinucleotides were used to trap the post-cleavage state. The φC31 tetramer displays a smaller crossing angle between E-helices at the center of the tetramer, a less extensive interface between the *attP*- and *attB*-bound subunits, and a greater distance between the two DNA duplexes. There may be multiple sub-states along the pathway from initial synapsis to the swiveling-competent cleaved state, with catalysis itself occurring somewhere between those two points. If the catalytic sites were always fully assembled and active, the system would risk (1) premature DNA cleavage before verification of a proper *attP* and *attB* pairing and (2) hydrolysis of the phosphoserine linkage during subunit rotation. The latter consideration explains why the catalytic arginines were not well ordered around the phosphoserine in the SPβ tetramers.

For LSI-catalyzed DNA recombination to proceed efficiently, the reaction pathway must avoid deep energy wells before the final handcuffed product dimer state is reached. This may be achieved by balancing favorable macromolecular interactions with conformational strain, including the bending of the DNA. For example, we noted an unusual tilt of the *attP* DNAs toward the catalytic domains in post-cleavage SPβ tetramers that was not seen in pre-cleavage φC31 tetramers. We propose that there is a balance between having all four DBD2s bound to DNA with both CC-CC pairs interacting across the synapse versus conformational strain in the DNA near the active site. Such a balance would explain the difficulties reported in obtaining pre-cleavage φC31 particles with well-ordered CC-CC pairs on both sides of the tetramer. DNA cleavage may relieve some conformational strain, driving the tetramer to the cleaved and swiveling state. Religation of the DNA in the product configuration would re-introduce the strain, but would also favor dissociation into dimers, which are then trapped by CC-CC handcuffing. In contrast, religation in the substrate configuration would re-introduce the strain with no energetic compensation.

Overall, the LSI reaction pathway is remarkably simple in its protein and DNA requirements, yet entails a delicate energetic balance among numerous protein-protein and protein-DNA interactions to guide the reaction forward. The protein-protein interactions must be strong enough to stabilize key intermediates

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

and enforce directionality, yet transient enough to avoid kinetic trapping prior to product formation. Directionality thus arises from dynamic control of conformational strain, DNA bending, and selective interaction switching rather than from any single dominant interaction. Together, this architecture enables tuning of LSIs to combine stringent reaction control with catalytic efficiency and provides a general framework for understanding and ultimately engineering directionality in complex DNA recombination systems.

Limitations of the study

While we have visualized the key steps along both the integration and excision reaction pathways, there may have been additional conformations not analyzed because they were present as a small percentage of the particles in the cryo-EM data—e.g., rotational states of the tetramers other than 0° and 180°. Additionally, we do not yet fully understand the conformational dynamics of the CAT domains, which were visualized before and after phosphotransfer, but not during. Finally, although specific CC-CAT contacts are observed during the integrative (but not excisive) pathway, their potential roles in catalytic activation in the tetramer state and inhibition in the dimer state require more exploration.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Phoebe Rice (PRice@uchicago.edu).

Materials availability

This study did not generate new unique reagents. All plasmids generated in this study are available from the lead contact with an appropriate material transfer agreement.

Data and code availability

- All imaging data have been deposited in the Mendeley Data repository at the following link: (https://doi.org/10.17632/9cd374hrv5.1).
- The models generated in this study were deposited at the Protein Data Bank (PDB) under the accession codes PDB: 9DXH, 9DXJ, 9DXK, 9Y66, 9DXD, 9DXF, 9DXG, and 9Y6V. The raw cryo-EM images have been deposited in the Electron Microscopy Public Image Archive (EMPIAR) under codes EMPIAR-12957, EMPIAR-12972, EMPIAR-12973, and EMPIAR-12974. The cryo-EM maps have been deposited in the Electron Microscopy Data Bank (EMDB) under codes EMDB: EMD-47288, EMD-47289, EMD-47290, EMD-72552, EMD-47284, EMD-47286, EMD-47287, and EMD-72632.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank the UChicago Research Computing Center (RCC) for services using the Beagle3 computing cluster, funded by the NIH 1S10OD028655-01. Cryo-EM datasets were collected at the University of Chicago Advanced Electron Microscopy Core Facility (RRID: SCR_019198). This publication is primarily supported by the National Science Foundation and UK Research and Innovation (collaborative grant NSF/BIO 2223480 and UKRI/BBSRC BB/X012085/1 to P.A.R. and F.J.O.) and a Wellcome Trust Accelerator Award (316534/Z/24/Z) to F.J.O. Funding for open access charge: UChicago incentive funds.

AUTHOR CONTRIBUTIONS

Conception and design and methodology, P.A.R., F.J.O., and H.S.; funding acquisition, P.A.R. and F.J.O.; data acquisition, H.S., A.J.B., J.R.F., Y.P., O.T.T.-A., and T.P.R.; data analysis, P.A.R., F.J.O., and H.S.; drafting and revision, P.A.R., F.J.O., and H.S.

DECLARATION OF INTERESTS

P.A.R., F.J.O., and H.S. are listed as inventors on a patent application related to this work (application no. [PCT/US26/33512], filed on June 19, 2026, assigned to the University of Chicago). T.P.R. is currently at the University of Virginia, although the work reported in this manuscript was done at the University of Chicago.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - *Escherichia coli* strains
- METHOD DETAILS
  - Cloning and plasmid construction
  - Protein expression and purification
  - *In vivo* integrase mediated inversion assay
  - *In vitro* recombination and binding assays
  - Sample preparation for single particle cryo-EM
  - Data collection for single particle cryo-EM
  - Image processing
  - Model building, refinement and validation
- QUANTIFICATION ANT STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2026.09.028.

Received: February 12, 2026
Revised: July 29, 2026
Accepted: September 18, 2026

REFERENCES

1. Smith, M.C.M. (2015). Phage-encoded Serine Integrases and Other Large Serine Recombinases. Microbiol. Spectr. *3*, 4. https://doi.org/10.1128/microbiolspec.MDNA3-0059-2014.
2. Gupta, K., Sharp, R., Yuan, J.B., Li, H., and Van Duyne, G.D. (2017). Coiled-coil interactions mediate serine integrase directionality. Nucleic Acids Res. *45*, 7339–7353. https://doi.org/10.1093/nar/gkx474.
3. Mandali, S., and Johnson, R.C. (2021). 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. https://doi.org/10.1128/JB.00703-20.
4. Thorpe, H.M., Wilson, S.E., and Smith, M.C.M. (2000). Control of directionality in the site-specific recombination system of the *Streptomyces* phage ϕC31. Mol. Microbiol. *38*, 232–241. https://doi.org/10.1046/j.1365-2958.2000.02142.x.
5. Van Duyne, G.D., and Rutherford, K. (2013). Large serine recombinase domain structure and attachment site binding. Crit. Rev. Biochem. Mol. Biol. *48*, 476–491. https://doi.org/10.3109/10409238.2013.831807.
6. McMahon, S.A., McEwan, A.R., Smith, M.C.M., and Naismith, J.H. (2013). Protein crystal structure of the N-terminal and recombinase domains of

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

the Streptomyces temperate phage serine recombinase, fC31 integrase. https://doi.org/10.2210/pdb4BQQ/pdb.

7. Yuan, P., Gupta, K., and Van Duyne, G.D. (2008). Tetrameric Structure of a Serine Integrase Catalytic Domain. Structure *16*, 1275–1286. https://doi.org/10.1016/j.str.2008.04.018.
8. Olorunniji, F.J., Rosser, S.J., and Stark, W.M. (2016). Site-specific recombinases: molecular machines for the Genetic Revolution. Biochem. J. *473*, 673–684. https://doi.org/10.1042/BJ20151112.
9. Stark, W.M. (2017). Making serine integrases work for us. Curr. Opin. Microbiol. *38*, 130–136. https://doi.org/10.1016/j.mib.2017.04.006.
10. Smith, M.C.M., and Thorpe, H.M. (2002). Diversity in the serine recombinases. Mol. Microbiol. *44*, 299–307. https://doi.org/10.1046/j.1365-2958.2002.02891.x.
11. Khaleel, T., Younger, E., McEwan, A.R., Varghese, A.S., and Smith, M.C.M. (2011). A phage protein that binds φC31 integrase to switch its directionality. Mol. Microbiol. *80*, 1450–1463. https://doi.org/10.1111/j.1365-2958.2011.07696.x.
12. Ghosh, P., Wasil, L.R., and Hatfull, G.F. (2006). Control of Phage Bxb1 Excision by a Novel Recombination Directionality Factor. PLoS Biol. *4*, e186. https://doi.org/10.1371/journal.pbio.0040186.
13. Smith, M.C.M., Brown, W.R.A., McEwan, A.R., and Rowley, P.A. (2010). Site-specific recombination by φC31 integrase and other large serine recombinases. Biochem. Soc. Trans. *38*, 388–394. https://doi.org/10.1042/BST0380388.
14. Bibb, L.A., and Hatfull, G.F. (2002). Integration and excision of the Mycobacterium tuberculosis prophage-like element, ϕRv1. Mol. Microbiol. *45*, 1515–1526. https://doi.org/10.1046/j.1365-2958.2002.03130.x.
15. Bibb, L.A., Hancox, M.I., and Hatfull, G.F. (2005). Integration and excision by the large serine recombinase ϕRv1 integrase. Mol. Microbiol. *55*, 1896–1910. https://doi.org/10.1111/j.1365-2958.2005.04517.x.
16. Breüner, A., Brøndsted, L., and Hammer, K. (1999). Novel Organization of Genes Involved in Prophage Excision Identified in the Temperate Lactococcal Bacteriophage TP901-1. J. Bacteriol. *181*, 7291–7297. https://doi.org/10.1128/JB.181.23.7291-7297.1999.
17. Blanch-Asensio, A., Grandela, C., Brandão, K.O., de Korte, T., Mei, H., Ariyurek, Y., Yiangou, L., Mol, M.P.H., van Meer, B.J., Kloet, S.L., et al. (2022). STRAIGHT-IN enables high-throughput targeting of large DNA payloads in human pluripotent stem cells. Cell Rep. Methods *2*, 100300. https://doi.org/10.1016/j.crmeth.2022.100300.
18. Gomide, M.S., Sales, T.T., Barros, L.R.C., Limia, C.G., de Oliveira, M.A., Florentino, L.H., Barros, L.M.G., Robledo, M.L., José, G.P.C., Almeida, M.S.M., et al. (2020). Genetic switches designed for eukaryotic cells and controlled by serine integrases. Commun. Biol. *3*, 255. https://doi.org/10.1038/s42003-020-0971-8.
19. Merrick, C.A., Zhao, J., and Rosser, S.J. (2018). Serine Integrases: Advancing Synthetic Biology. ACS Synth. Biol. *7*, 299–310. https://doi.org/10.1021/acssynbio.7b00308.
20. Guiziou, S., Maranas, C.J., Chu, J.C., and Nemhauser, J.L. (2023). An integrase toolbox to record gene-expression during plant development. Nat. Commun. *14*, 1844. https://doi.org/10.1038/s41467-023-37607-5.
21. Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. *38*, 824–844. https://doi.org/10.1038/s41587-020-0561-9.
22. Durrant, M.G., Fanton, A., Tycko, J., Hinks, M., Chandrasekaran, S.S., Perry, N.T., Schaepe, J., Du, P.P., Lotfy, P., Bassik, M.C., et al. (2023). Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. Nat. Biotechnol. *41*, 488–499. https://doi.org/10.1038/s41587-022-01494-w.
23. Yarnall, M.T.N., Ioannidi, E.I., Schmitt-Ulms, C., Krajeski, R.N., Lim, J., Villiger, L., Zhou, W., Jiang, K., Garushyants, S.K., Roberts, N., et al. (2023). Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. Nat. Biotechnol. *41*, 500–512. https://doi.org/10.1038/s41587-022-01527-4.
24. Vo, P.L.H., Ronda, C., Klompe, S.E., Chen, E.E., Acree, C., Wang, H.H., and Sternberg, S.H. (2021). CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. Nat. Biotechnol. *39*, 480–489. https://doi.org/10.1038/s41587-020-00745-y.
25. Van Duyne, G.D., and Landy, A. (2024). Bacteriophage lambda site-specific recombination. Mol. Microbiol. *121*, 895–911. https://doi.org/10.1111/mmi.15241.
26. Grindley, N.D.F., Whiteson, K.L., and Rice, P.A. (2006). Mechanisms of Site-Specific Recombination. Annu. Rev. Biochem. *75*, 567–605. https://doi.org/10.1146/annurev.biochem.73.011303.073908.
27. Rice, P.A. (2015). Serine Resolvases. Microbiol. Spectr. *3*. MDNA3-0045-2014. https://doi.org/10.1128/microbiolspec.MDNA3-0045-2014.
28. Kamtekar, S., Ho, R.S., Cocco, M.J., Li, W., Wenwieser, S.V.C.T., Boocock, M.R., Grindley, N.D.F., and Steitz, T.A. (2006). Implications of structures of synaptic tetramers of gamma delta resolvase for the mechanism of recombination. Proc. Natl. Acad. Sci. USA *103*, 10642–10647. https://doi.org/10.1073/pnas.0604062103.
29. Li, W., Kamtekar, S., Xiong, Y., Sarkis, G.J., Grindley, N.D.F., and Steitz, T.A. (2005). Structure of a Synaptic γδ Resolvase Tetramer Covalently Linked to Two Cleaved DNAs. Science *309*, 1210–1215. https://doi.org/10.1126/science.1112064.
30. Dhar, G., Sanders, E.R., and Johnson, R.C. (2004). Architecture of the Hin Synaptic Complex during Recombination: The Recombinase Subunits Translocate with the DNA Strands. Cell *119*, 33–45. https://doi.org/10.1016/j.cell.2004.09.010.
31. Stark, W.M., Sherratt, D.J., and Boocock, M.R. (1989). Site-Specific Recombination by Tn3 Resolvase: Topological Changes in the Forward and Reverse Reactions. Cell *58*, 779–790. https://doi.org/10.1016/0092-8674(89)90111-6.
32. Stark, W.M., Grindley, N.D.F., Hatfull, G.F., and Boocock, M.R. (1991). Resolvase-catalysed reactions between res sites differing in the central dinucleotide of subsite I. EMBO J. *10*, 3541–3548. https://doi.org/10.1002/j.1460-2075.1991.tb04918.x.
33. Olorunniji, F.J., Buck, D.E., Colloms, S.D., McEwan, A.R., Smith, M.C.M., Stark, W.M., and Rosser, S.J. (2012). Gated rotation mechanism of site-specific recombination by ϕC31 integrase. Proc. Natl. Acad. Sci. USA *109*, 19661–19666. https://doi.org/10.1073/pnas.1210964109.
34. Bai, H., Sun, M., Ghosh, P., Hatfull, G.F., Grindley, N.D.F., and Marko, J.F. (2011). Single-molecule analysis reveals the molecular bearing mechanism of DNA strand exchange by a serine recombinase. Proc. Natl. Acad. Sci. USA *108*, 7419–7424. https://doi.org/10.1073/pnas.1018436108.
35. Cadden, G.M., Schloetel, J.G., McKenzie, G., Boocock, M.R., Magennis, S.W., and Stark, W.M. (2024). Direct observation of subunit rotation during DNA strand exchange by serine recombinases. Nat. Commun. *15*, 10407. https://doi.org/10.1038/s41467-024-54531-4.
36. Yang, W., and Steitz, T.A. (1995). Crystal Structure of the Site-Specific Recombinase γδ Resolvase Complexed with a 34 bp Cleavage Site. Cell *82*, 193–207. https://doi.org/10.1016/0092-8674(95)90307-0.
37. Mouw, K.W., Steiner, A.M., Ghirlando, R., Li, N.S., Rowland, S.J., Boocock, M.R., Stark, W.M., Piccirilli, J.A., and Rice, P.A. (2010). Sin Resolvase Catalytic Activity and Oligomerization State are Tightly Coupled. J. Mol. Biol. *404*, 16–33. https://doi.org/10.1016/j.jmb.2010.08.057.
38. Arnold, P.H., Blake, D.G., Grindley, N.D.F., Boocock, M.R., and Stark, W.M. (1999). Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. EMBO J. *18*, 1407–1414. https://doi.org/10.1093/emboj/18.5.1407.
39. Johnson, R.C. (2015). Site-specific DNA Inversion by Serine Recombinases. Microbiol. Spectr. *3*, 199–236. https://doi.org/10.1128/microbiolspec.MDNA3-0047-2014.

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

40. McLean, M.M., Chang, Y., Dhar, G., Heiss, J.K., and Johnson, R.C. (2013). Multiple interfaces between a serine recombinase and an enhancer control site-specific DNA inversion. eLife *2*, e01211. https://doi.org/10.7554/eLife.01211.

41. Sarkis, G.J., Murley, L.L., Leschziner, A.E., Boocock, M.R., Stark, W.M., and Grindley, N.D.F. (2001). A Model for the Γδ Resolvase Synaptic Complex. Mol. Cell *8*, 623–631. https://doi.org/10.1016/S1097-2765(01)00334-3.

42. Dhar, G., McLean, M.M., Heiss, J.K., and Johnson, R.C. (2009). The Hin recombinase assembles a tetrameric protein swivel that exchanges DNA strands. Nucleic Acids Res. *37*, 4743–4756. https://doi.org/10.1093/nar/gkp466.

43. Montaño, S.P., Rowland, S.J., Fuller, J.R., Burke, M.E., MacDonald, A.I., Boocock, M.R., Stark, W.M., and Rice, P.A. (2023). Structural basis for topological regulation of Tn3 resolvase. Nucleic Acids Res. *51*, 1001–1018. https://doi.org/10.1093/nar/gkac733.

44. Mouw, K.W., Rowland, S.J., Gajjar, M.M., Boocock, M.R., Stark, W.M., and Rice, P.A. (2008). Architecture of a Serine Recombinase-DNA Regulatory Complex. Mol. Cell *30*, 145–155. https://doi.org/10.1016/j.molcel.2008.02.023.

45. Rowland, S.J., Boocock, M.R., and Stark, W.M. (2006). DNA bending in the Sin recombination synapse: functional replacement of HU by IHF. Mol. Microbiol. *59*, 1730–1743. https://doi.org/10.1111/j.1365-2958.2006.05064.x.

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

47. Boocock, M.R., and Rice, P.A. (2013). A proposed mechanism for IS607-family serine transposases. Mobile DNA *4*, 24. https://doi.org/10.1186/1759-8753-4-24.

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

49. Fogg, P.C.M., Younger, E., Fernando, B.D., Khaleel, T., Stark, W.M., and Smith, M.C.M. (2018). Recombination directionality factor gp3 binds ϕC31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. Nucleic Acids Res. *46*, 1308–1320. https://doi.org/10.1093/nar/gkx1233.

50. McEwan, A.R., Rowley, P.A., and Smith, M.C.M. (2009). DNA binding and synapsis by the large C-terminal domain of φC31 integrase. Nucleic Acids Res. *37*, 4764–4773. https://doi.org/10.1093/nar/gkp485.

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

52. Rowley, P.A., Smith, M.C.A., Younger, E., and Smith, M.C.M. (2008). A motif in the C-terminal domain of ϕC31 integrase controls the directionality of recombination. Nucleic Acids Res. *36*, 3879–3891. https://doi.org/10.1093/nar/gkn269.

53. MacDonald, A.I., Baksh, A., Holland, A., Shin, H., Rice, P.A., Stark, W.M., and Olorunniji, F.J. (2024). Variable orthogonality of serine integrase interactions within the ϕC31 family. Sci. Rep. *14*, 26280. https://doi.org/10.1038/s41598-024-77570-9.

54. Shin, H., Holland, A., Alsaleh, A., Retiz, A.D., Pigli, Y.Z., Taiwo-Aiyerin, O.T., Peña Reyes, T., Bello, A.J., Quan, J., Tang, W., et al. (2025). Identification of cognate recombination directionality factors for large serine recombinases by virtual pulldown. Nucleic Acids Res. *53*, gkaf691. https://doi.org/10.1093/nar/gkaf691.

55. Alsaleh, A., Holland, A., Shin, H., Reyes, T.P., Baksh, A., Taiwo-Aiyerin, O.T., Pigli, Y., Rice, P.A., and Olorunniji, F.J. (2025). Large serine integrases utilise scavenged phage proteins as directionality cofactors. Nucleic Acids Res. *53*, gkaf050. https://doi.org/10.1093/nar/gkaf050.

56. Abe, K., Takamatsu, T., and Sato, T. (2017). Mechanism of bacterial gene rearrangement: SprA-catalyzed precise DNA recombination and its directionality control by SprB ensure the gene rearrangement and stable expression of spsM during sporulation in *Bacillus subtilis*. Nucleic Acids Res. *45*, 6669–6683. https://doi.org/10.1093/nar/gkx466.

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

58. Olorunniji, F.J., McPherson, A.L., Rosser, S.J., Smith, M.C.M., Colloms, S.D., and Stark, W.M. (2017). Control of serine integrase recombination directionality by fusion with the directionality factor. Nucleic Acids Res. *45*, 8635–8645. https://doi.org/10.1093/nar/gkx567.

59. Nelson, H.C.M., Finch, J.T., Luisi, B.F., and Klug, A. (1987). The structure of an oligo(dA).oligo(dT) tract and its biological implications. Nature *330*, 221–226. https://doi.org/10.1038/330221a0.

60. Mitchell, J.S., Glowacki, J., Grandchamp, A.E., Manning, R.S., and Maddocks, J.H. (2017). Sequence-Dependent Persistence Lengths of DNA. J. Chem. Theory Comput. *13*, 1539–1555. https://doi.org/10.1021/acs.jctc.6b00904.

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

62. Liu, K., Xu, C., Lei, M., Yang, A., Loppnau, P., Hughes, T.R., and Min, J. (2018). Structural basis for the ability of MBD domains to bind methyl-CG and TG sites in DNA. J. Biol. Chem. *293*, 7344–7354. https://doi.org/10.1074/jbc.RA118.001785.

63. Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein–DNA recognition. Nature *461*, 1248–1253. https://doi.org/10.1038/nature08473.

64. Singh, S., Ghosh, P., and Hatfull, G.F. (2013). Attachment Site Selection and Identity in Bxb1 Serine Integrase-Mediated Site-Specific Recombination. PLoS Genet. *9*, e1003490. https://doi.org/10.1371/journal.pgen.1003490.

65. Trejo, C.S., Rock, R.S., Stark, W.M., Boocock, M.R., and Rice, P.A. (2018). Snapshots of a molecular swivel in action. Nucleic Acids Res. *46*, 5286–5296. https://doi.org/10.1093/nar/gkx1309.

66. Sun, Y.E., Aspinall, L., Joseph, A.P., Colloms, S.D., Stark, W.M., and Spagnolo, L. (2025). Structural basis of unidirectional DNA recombination by the ϕC31 serine integrase. Preprint at bioRxiv. https://doi.org/10.1101/2025.05.02.651858.

67. Olorunniji, F.J., and Stark, W.M. (2009). The catalytic residues of Tn3 resolvase. Nucleic Acids Res. *37*, 7590–7602. https://doi.org/10.1093/nar/gkp797.

68. Keenholtz, R.A., Mouw, K.W., Boocock, M.R., Li, N.S., Piccirilli, J.A., and Rice, P.A. (2013). Arginine as a General Acid Catalyst in Serine Recombinase-mediated DNA Cleavage. J. Biol. Chem. *288*, 29206–29214. https://doi.org/10.1074/jbc.M113.508028.

69. Abe, K., Takahashi, T., and Sato, T. (2021). Extreme C-terminal element of SprA serine integrase is a potential component of the "molecular toggle switch" which controls the recombination and its directionality. Mol. Microbiol. *115*, 1110–1121. https://doi.org/10.1111/mmi.14654.

70. Olorunniji, F.J., Lawson-Williams, M., McPherson, A.L., Paget, J.E., Stark, W.M., and Rosser, S.J. (2019). Control of ϕC31 integrase-mediated site-specific recombination by protein trans-splicing. Nucleic Acids Res. *47*, 11452–11460. https://doi.org/10.1093/nar/gkz936.

71. Fournes, F., Campos, M., Cury, J., Schiavon, C., Pagès, C., Touchon, M., Rocha, E.P.C., Rousseau, P., and Cornet, F. (2025). The pathway to resolve dimeric forms distinguishes plasmids from megaplasmids in Enterobacteriaceae. Nucleic Acids Res. *53*, gkae1300. https://doi.org/10.1093/nar/gkae1300.

72. Fanton, A., Bartie, L.J., Martins, J.Q., Tran, V.Q., Goudy, L., Kernick, C., Durrant, M.G., Wei, J., Armour-Garb, Z., Pawluk, A., et al. (2025). Site-specific DNA insertion into the human genome with engineered

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

recombinases. Nat. Biotechnol. https://doi.org/10.1038/s41587-025-02895-3.

73. Soma, T., Hiraizumi, M., Fell, C.W., Tagliaferri, D., Lequyer, J., Okazaki, S., Isayama, Y., Kato, K., Sapkota, S., Ramani, H., et al. (2026). Structure and engineering of the large serine recombinase Bxb1 for gene integration. Mol. Cell *86*, 2247–2262.e7. https://doi.org/10.1016/j.molcel.2026.05.018.

74. Summers, D.K., and Sherratt, D.J. (1988). Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the cer site. EMBO J. *7*, 851–858. https://doi.org/10.1002/j.1460-2075.1988.tb02884.x.

75. Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. *177*, 4121–4130. https://doi.org/10.1128/jb.177.14.4121-4130.1995.

76. Cohen, S.N., Chang, A.C.Y., Boyer, H.W., and Helling, R.B. (1973). Construction of Biologically Functional Bacterial Plasmids In Vitro. Proc. Natl. Acad. Sci. USA *70*, 3240–3244. https://doi.org/10.1073/pnas.70.11.3240.

77. Morin, A., Eisenbraun, B., Key, J., Sanschagrin, P.C., Timony, M.A., Ottaviano, M., and Sliz, P. (2013). Collaboration gets the most out of software. eLife *2*, e01456. https://doi.org/10.7554/eLife.01456.

78. Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332. https://doi.org/10.1038/nmeth.4193.

79. Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., and Berger, B. (2019). Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. Nat. Methods *16*, 1153–1160. https://doi.org/10.1038/s41592-019-0575-8.

80. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). CryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

81. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

82. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr., D *60*, 2126–2132. https://doi.org/10.1107/S0907444904019158.

83. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr., D *66*, 213–221. https://doi.org/10.1107/S0907444909052925.

84. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315. https://doi.org/10.1002/pro.3330.

85. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/jcc.20084.

86. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.

87. Croll, T.I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallogr. D Struct. Biol. *74*, 519–530. https://doi.org/10.1107/S2059798318002425.

88. Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J.M., Sorzano, C.O.S., and Vargas, J. (2021). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Commun. Biol. *4*, 874. https://doi.org/10.1038/s42003-021-02399-1.

89. Schrödinger, L.L.C. (2025). The PyMOL Molecular Graphics System, Version 3.1.3. https://www.pymol.org.

90. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675. https://doi.org/10.1038/nmeth.2089.

91. Punjani, A., and Fleet, D.J. (2021). 3D Variability Analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. J. Struct. Biol. *213*, 107702. https://doi.org/10.1016/j.jsb.2021.107702.

92. Punjani, A., and Fleet, D.J. (2023). 3DFlex: determining structure and motion of flexible proteins from cryo-EM. Nat. Methods *20*, 860–870. https://doi.org/10.1038/s41592-023-01853-8.

93. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E.coli, strain: BL21(DE3)pLysS	Novagen	Cat#69451-3
E.coli, strain: DH5α	Thermo Fisher	Cat#1826017
E.coli, strain DS941	Summers and Sherratt et al.[74]	N/A
Chemicals, peptides, and recombinant proteins		
Fluorinated octyl maltoside (FOM)	Anatrace	Cat#O310F
SPβ Integrase	This paper	N/A
SPβ Integrase (L197A, L206A, L223A)	This paper	N/A
SPβ Integrase (L197R, L206R, L223R)	This paper	N/A
SPβ Integrase (E258A)	This paper	N/A
SPβ Integrase (M409A)	This paper	N/A
SPβ Integrase (M409R)	This paper	N/A
SPβ Integrase (T49A, I53D, P54D)	This paper	N/A
SPβ Integrase-RDF	This paper	N/A
SPβ Integrase-RDF (L197A, L206A, L223A)	This paper	N/A
SPβ Integrase-RDF (L197R, L206R, L223R)	This paper	N/A
SPβ Integrase-RDF (E258A)	This paper	N/A
SPβ Integrase-RDF (M409A)	This paper	N/A
SPβ Integrase-RDF (M409R)	This paper	N/A
SPβ Integrase-RDF (T49A, I53D, P54D)	This paper	N/A
SPβ Integrase-RDF (ΔN-ter helix 1-13)	This paper	N/A
SPβ Integrase-RDF (ΔN-ter helix 37-58)	This paper	N/A
SPβ Integrase-RDF (L37D, F40D)	This paper	N/A
SPβ Integrase-RDF (A45D, L49D)	This paper	N/A
SPβ Integrase (elongated hinge 2)	This paper	N/A
SPβ Integrase-RDF (elongated hinge 2)	This paper	N/A
SPβ Integrase (elongated hinge 1)	This paper	N/A
SPβ Integrase-RDF (elongated hinge 1)	This paper	N/A
Deposited data		
Cryo-EM map of SPβ Int pre-rotation state	This paper	EMD-47288
Coordinates of SPβ Int pre-rotation state	This paper	PDB: 9DXH
Cryo-EM map of SPβ Int post-rotation state	This paper	EMD-47289
Coordinates of SPβ Int post-rotation state	This paper	PDB: 9DXJ
Cryo-EM map of SPβ Int *attP*mm tetramer	This paper	EMD-47290
Coordinates of SPβ Int *attP*mm tetramer	This paper	PDB: 9DXK
Cryo-EM map of SPβ Int *attL*sym dimer	This paper	EMD-72552
Coordinates of SPβ Int *attL*sym dimer	This paper	PDB: 9Y66
Cryo-EM map of SPβ Int RDF pre-rotation state	This paper	EMD-47284
Coordinates of SPβ Int RDF pre-rotation state	This paper	PDB: 9DXD
Cryo-EM map of SPβ Int RDF post-rotation state	This paper	EMD-47286

(*Continued on next page*)

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Coordinates of SPβ Int RDF post-rotation state	This paper	PDB: 9DXF
Cryo-EM map of SPβ Int RDF *attP*mm dimer	This paper	EMD-47287
Coordinates of SPβ Int RDF *attP*mm dimer	This paper	PDB: 9DXG
Cryo-EM map of SPβ Int RDF *attP*sym dimer	This paper	EMD-72632
Coordinates of SPβ Int RDF *attP*sym dimer	This paper	PDB: 9Y6V
attLsym bound Int raw cryoEM data	This paper	PDB: 9Y66; EMPIAR-12972
Int-RDF raw cryoEM data	This paper	PDB: 9DXD, 9DXF, 9DXG; EMPIAR-12973
Int-RDF raw cryoEM data	This paper	PDB: 9Y6V; EMPIAR-12957
Int raw cryoEM data	This paper	PDB: 9DXH, 9DXJ, 9DXK; EMPIAR-12974
All other source data	Mendeley	Mendeley Data: https://doi.org/10.17632/9cd374hrv5.1
Oligonucleotides		
*attLsym*_top: CATGTCATTAATATCA GTACAGATATAGCTGTATA TTAAGATACTTACTACA TATCTAAAAAAAAAA	Integrated DNA Technologies	N/A
*attLsym*_bot: TTTTTTTTTTAGATATGTAG TAAGTATCTTAATATACAGCT ATATCTGTACTGATATTAATGACATG	Integrated DNA Technologies	N/A
*attLsym*_top for cryoEM: GCAGCATGTCATTAATATCA GTACAGATATAGCTGTATATTAAGAT ACTTACTACAAAAAAAA	Integrated DNA Technologies	N/A
*attLsym*_bot for cryoEM: TTTTTTTTGTAGTAAG TATCTTAATATACAGCTATATCTGTACT GATATTAATGACATGCTGC	Integrated DNA Technologies	N/A
*attRsym*_top: AAAAAAAAAAAGAAAAGTAGT AAGTATCTTAAAAAACAGATATAGCT GTATCTCCTGTGAACACAAT	Integrated DNA Technologies	N/A
*attRsym*_bot: ATTGTGTTCACAGGAGATACAGCTA TATCTGTTTTTTAAGATCATTACTA CTTTTCTTTTTTTTTTT	Integrated DNA Technologies	N/A
*attLmm*_top: CATGTCATTAATATCAGTACAG ATTTAGCTGTATATTAAGATAC TTACTACAAAAAAAA	Integrated DNA Technologies	N/A
*attLmm*_bot: TTTTTTTTGTAGTAAGTATCT TAATATACAGCTTTATCTGTA CTGATATTAATGACATG	Integrated DNA Technologies	N/A
*attPmm*_top: AAAAAAAAGGTAGTAAGTAT CTTAAAAAACAGATTTAGCTGTATA TTAAGATACTTACTAC	Integrated DNA Technologies	N/A
*attPmm*_bot: GTAGTAAGTATCTTAATATAC AGCTTTATCTGTTTTTTAAGAT ACTTACTACTTTTTTTTT	Integrated DNA Technologies	N/A

(*Continued on next page*)

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
*attPsym*_top: AAAAAAAAAAAGGTAGTAAGT ATCTTAAAAAACAGATATAGCTG TATATTAAGATACTTACTACATA	Integrated DNA Technologies	N/A
*attPsym*_bot: TATGTAGTAAGTATCTTAATAT ACAGCTATATCTGTTTTTTAAGATA CTTACTACCTTTTTTTTTTT	Integrated DNA Technologies	N/A
*attBmm*_top: CATGTCATTAATATCAGTACAG ATTTAGCTGTATCTCCTGTGAAC ACAATAAAAAAAA	Integrated DNA Technologies	N/A
*attBmm*_bot: TTTTTTTATTGTGTTCACAGGAG ATACAGCTTTATCTGTACTG ATATTAATGACATG	Integrated DNA Technologies	N/A
*attBsym*_top: GCAGCATGTCATTAATATC AGTACAGATATAGCTGTATC TCCTGTGAACACAATAAAAAAAAAAAAA	Integrated DNA Technologies	N/A
*attB*sym_bot: TTTTTTTTTTTTTATTGTGT TCACAGGAGATACAGCTAT ATCTGTACTGATATTAATGACATGCTGC	Integrated DNA Technologies	N/A
5'fluorescein labeled *attB*_top for recombination: /56-FAM/, CATGTCATTAATATCAGTA CAGATAAAGCTGTATCTCC TGTGAACACAAT	Integrated DNA Technologies	N/A
Unlabeled *attB*_bot for recombination assay: ATTGTGTTCACAGGAGATACAGCTTTA TCTGTACTGATATTAATGACATG	Integrated DNA Technologies	N/A
56-FAM-SPBc_attLsym2_top: /56-FAM/, CATGTCATTAATA TCAGTACAGATATAGCTGTATATT AAGATACTTACTACATATCTA AAAAAAAAA	Integrated DNA Technologies	N/A
56-FAM-SPBc_attLsym2_bot: /56-FAM/, TTTTTTTTGTAGTAAGT ATCTTAATATACAGCTTTATCTGTAC TGATATTAATGACATG	Integrated DNA Technologies	N/A
Unlabeled attLsym2_top: CATGTCATTAATATCAGTA CAGATATAGCTGTATATTAAGATACT TACTACATATCTAAAAAAAAAA	Integrated DNA Technologies	N/A
Unlabeled attLsym2_bot: TTTTTTTTGTAGTAAGTATCTTAATA TACAGCTTTATCTGTACTGATATTA ATGACATG	Integrated DNA Technologies	N/A
56-FAM-SPBc_attRsym2_top: /56-FAM/, AAAAAAAAAAAGAAAAGTAGTAAGTA TCTTAAAAAACAGATATAGCTGTATC TCCTGTGAACACAAT	Integrated DNA Technologies	N/A
56-FAM-SPBc_attRsym2_bot: /56-FAM/, ATTGTGTTCACAGGAGATACAGCTA TATCTGTTTTTTAAGATCATTACTAC TTTTCTTTTTTTTTT	Integrated DNA Technologies	N/A

(*Continued on next page*)

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
pUC19	NEB	Cat#N3041L
pET-28a(+)	Novagen	Cat#69864-3
pBAD33	Guzman et al.[75]	N/A
pSC101	Cohen et al.[76]	N/A
pTPR11_SPβ_LR	This study	N/A
pTPR12_SPβ_PB	This study	N/A
Software and algorithms		
SBGrid	Morin et al.[77]	https://sbgrid.org/
MotionCor2	Zheng et al.[78]	https://github.com/singleparticle/MotionCor2
Topaz	Bepler et al.[79]	https://github.com/tbepler/topaz
cryoSPARC 4.6.0	Punjani et al.[80]	https://cryosparc.com
AlphaFold 2 (AF2)	Jumper et al.[81]	https://www.alphafold.ebi.ac.uk/
Coot 0.9.8.8	Emsley and Cowtan[82]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Phenix 1.21	Adams et al.[83]	https://www.phenix-online.org/
Molprobity	Williams et al.[84]	https://github.com/rlabduke/MolProbity
UCSF Chimera 1.17.3	Petterson et al.[85]	https://www.cgl.ucsf.edu/chimera/
UCSF ChimeraX 1.4	Pettersen et al.[86]	https://www.cgl.ucsf.edu/chimerax/
ISOLDE	Croll et al.[87]	https://tristanic.github.io/isolde/
deepEMhancer 0.14	Sanchez-Garcia et al.[88]	https://github.com/rsanchezgarc/deepEMhancer
PyMOL 3.1.3	Schrödinger[89]	https://www.pymol.org
BioRender	BioRender	https://biorender.com
ImageJ 1.54g	Schneider et al.[90]	https://imagej.org
Other		
Quantifoil R1.2/1.3 300-mesh, Cu	Electron Microscopy Sciences	Cat#Q3100CR1.3
CFlat R4.0/1.0 300-mesh, Cu grids	Electron Microscopy Sciences	Discontinued

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Escherichia coli strains

E.coli pET28-a(+) (Novagen) was used for plasmid construction, while *E.coli* BL21(DE3)pLysS (Novagen) was used for protein expression. *E.coli* DS941[74] was used for in vivo integrase mediated inversion assays.

METHOD DETAILS

Cloning and plasmid construction

The DNA sequence of SPβ Integrase was optimized for expression in *E. coli* and cloned between the NdeI and XhoI sites of pET-28a(+) (Novagen). The proteins expressed from this vector have an N-terminal hexahistidine tag followed by a thrombin cleavage site (the tag was not cleaved off, but was not visible in the density maps). The DNA sequence encoding the SPβ integrase–RDF fusion protein was made by joining the coding sequences for SPβ integrase and its RDF via a 54 bp DNA linker sequence between the SpeI and XhoI sites in pET28-a(+). The DNA linker sequence is translated to a flexible 18-residue linker (TSGSGGSGGSGGSGRSGT) between the C terminal residue of SPβ integrase and the second amino acid residue in the RDF. Mutants of Integrase and Integrase-RDF were generated using site-directed mutagenesis, and all constructs were verified by whole plasmid sequencing of the expression vectors.

Protein expression and purification

Expression plasmids containing SPβ Int or SPβ Int-RDF were used to transform *E. coli* BL21(DE3)pLysS competent cells. The transformed cells were grown at 37 °C in LB broth supplemented with 50μg/ml Kanamycin and 100 μM $ZnSO_4$, until OD_{600} reached 0.7.

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

Protein expression was induced with isopropylβ-D-1-thiogalactopyranoside (IPTG) at final concentration of 0.5 mM, and the cultures were grown overnight at 20 °C. The cells were pelleted by centrifugation at 8,000 r.p.m. in a Fiberlite F9-6X1000LEX rotor for 7 minutes at 4 °C and stored at -20 °C.

Cell pellets were resuspended in 50ml per liter of original culture of Ni A buffer (50mM Sodium Phosphate, 1M NaCl, 5% Glycerol, 1mM Tris(2-carboxylethyl) phosphine (TCEP), pH 7.5) plus 1 tablet of complete Mini EDTA-free protease inhibitor cocktail (Roche) and 200 μg/ml lysozyme, then lysed by ultrasonication in 3 1-minute pulses in an ice bath. The lysate was clarified by centrifugation at 18,000 r.p.m. in an SS-34 rotor for 1 hour at 4 °C. The supernatant was collected, filtered and then loaded on a 5ml Ni Sepharose HisTrap HP affinity column (Cytiva). The column was washed with 95% Ni A / 5% Ni B buffer (Ni B = Ni A plus 500 mM Imidazole pH 7.5) until the baseline was flat, then eluted by a gradient from 5% to 100% Ni B) over 40' with 2ml/min and 2ml/fraction. The protein peak fractions were analyzed on a Novex 8-16% Tris-Glycine Minigel (Invitrogen). The cleanest fractions were pooled and dialyzed into Ni A buffer at 4 °C O/N, then rechromatographed on the Ni column as before. The best fractions were pooled, then diluted into Heparin A buffer (20mM MES, 5% Glycerol, 1mM TCEP, pH 5.5) to a final salt concentration of 200-300mM, and then loaded onto a 20ml Heparin column (Heparin FF, 16/10, Cytiva), washed with 15% Heparin B buffer (Heparin A buffer plus 2M NaCl, pH5.5) and eluted with a gradient from 15% to 75% B over 90' at 2ml/min and 2ml/fraction. The cleanest peak fractions were and concentrated to a small volume, then dialyzed into 20 mM Tris-Cl pH 8, 200 mM NaCl, 2 mM TCEP, and 20% glycerol. The protein concentration was determined by A_{280} and the presence of contaminating endonucleases was ruled out by incubating the sample with supercoiled pUC19 and 10mM $MgCl_2$ followed by agarose gel electrophoresis.

For the fluorescently labeled oligos, some of the oligos listed above were custom ordered with 5' fluorescein label or 3' Cy5 label from Integrated DNA technologies (IDT). Complementary oligos were mixed at a final concentration of 1 mM in TE buffer containing 100 mM NaCl and annealed by incubating at 80 °C for 20 minutes. Afterward, the water bath was turned off and samples were left in it to cool to room temperature.

In vivo integrase mediated inversion assay

The design and protocol for the *in vivo* recombination assay were similar to those previously described.[53,70] The substrate for *in vivo* intramolecular recombination (inversion) is a reporter plasmid (kanamycin resistance) in which *attP*/*attB* or *attR*/*attL* sites are in inverted orientation and flank a constitutive promoter that drives either expression of GFP or RFP (Figure S19A). To assay recombination activities *in vivo*, *E. coli* DS941 cells were cotransformed with a substrate plasmid and an arabinose-inducible expression vector, pBAD33 (chloramphenicol resistance) containing the gene of interest (Int, Int-RDF fusion or its mutants cloned between NdeI and XhoI sites). To do so, 2μl of each plasmid at 20ng/μl was incubated with 100ul chemically competent DS941 cells on ice for 5', then heat shocked at 37 °C for 5', then iced for 2'. 200μl LB+0.2% glucose was then added, cells were allowed to recover 37 °C for 2 hours, then added to 10ml LB+0.2% glucose+50μg/ml kanamycin (kan) +30μg/ml chloramphenicol (cam) and grown overnight at 37 °C. All liquid cultures for these experiments were shaken at 200 rpm. Glucose was added to suppress protein expression, and we noted that the overall experiment worked best if the overnight culture reached an OD_{600} of > 1.5 – if not it was best to restart with freshly made competent cells.

200μl of the overnight culture was used to inoculate 10ml LB + 0.2% Glucose + kan (50 μg/ml) + cam (30μg/ml), then grown at 37 °C for 1.5 hours, at which point OD_{600} was ~0.4-0.5. Cells were spun down to remove the glucose, then resuspended in 10ml LB + 0.2% L-Arabinose + kan (50 μg/ml) + cam (30 μg/ml) and incubated at 37 °C for 2 hours or overnight, after which the plasmids were recovered by miniprep. 1μl of miniprepped plasmid was then used to transform 50ul chemically competent DS941 cells: 5' on ice, 5' heat shock at 37 °C and 2' on ice then recovery in 500μl LB at 37C for 90'. We choose to isolate and then retransform the product plasmids to separate out individual copies of the substrate plasmid, which carries a pSC101 origin and can be present in multiple copies per cell. Two different volumes of each culture (20μl and 100μl) were then plated LB + Kan (50 μg/ml) agar plates and incubated at 37 °C overnight. Plates were visualized in a chemi-doc (Bio-Rad, Hercules, CA, USA) with blue epi illumination and a 530nm±28nm bandpass filter for GFP visualization and red epi illumination with a 695nm±55nm bandpass filter for RFP. Exposure times were set at 0.025s and 0.25s respectively.

For quantification shown in Figure S19, a custom Python script was used to classify colonies as red or green. Briefly, green and red colonies were segmented by hue, saturation, and value (HSV) color thresholding, and detected independently. The automated counts for red or green were then reviewed and manually corrected before final counts were recorded for each plate. For each plate, the number of green and red colonies was calculated as a proportion of the total colony count on that plate. All mutants were tested on both reporter plasmids as an internal control, and all individual assays were repeated at least twice. For the elongated-hinge mutants, proportions were averaged across replicates, which followed the protocol described above. The mean proportions of red and green colonies (±SD) were plotted for each condition. The initial assays for the other mutants were carried out with a slightly different protocol (cells were switched back to glucose and then grown longer before miniprepping, etc) and therefore, while they gave very similar results, their quantification was not averaged with that shown in Figure S19.

A slightly simplified protocol was used to generate the data shown in (Figure S20). Overnight cultures of each recombination strain were grown at 37 °C in LB media to mid-log phase (approximately 2 hours) with kanamycin (50 μg/ml) and chloramphenicol (25 μg/ml) selections. Integrase or integrase-RDF expression was induced for 2 hours by addition of 0.2% arabinose after which cultures were

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

diluted 1:1000 in fresh LB media containing 0.2% glucose to repress further integrase expression. The cultures were grown for further 16 hours with shaking (200 rpm) at 37 °C. Cultures were diluted 1/100, 000 in fresh LB media and spread on LB agar plates containing kanamycin (50 μg/ml).

Due to the thickly plated colonies, the overall green and red colony areas were quantified using ImageJ.[90] Color thresholding was used to isolate green and red colonies separately. The background outside each selected colony population was cleared, and the resulting images were converted to 8-bit grayscale. Thresholding was then applied to generate binary images in which the selected colonies appeared as white pixels. The total white-pixel area (px^2) was measured separately for the green and red colonies. The fraction of green and red colony areas was then calculated for each plate. At least two independent experiments were performed for each construct (three for the 3L and M409 mutants, both with and without the RDF), and the mean and standard deviation of the calculated red and green fractions were plotted for comparison.

In vitro recombination and binding assays

For binding and recombination assays, protein and DNA were used in ratio 2:1 (final concentrations: DNA, 0.2 μM; protein, 0.4 μM). The reactions were performed in a buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 % glycerol, 50 μg/ml BSA, 0.5 mM TCEP. Samples were incubated at 37 °C for 1 hr.

After the incubation, the samples were divided into two sets. To assess recombination, 10 % of Sodium Dodecyl Sulfate (SDS) and 6X DNA loading dye (6X Gel loading dye, purple—New England Biolabs) were added, resulting in final concentrations of 1.1% SDS and 1X loading dye. Both sets of samples were separated on Novex 6% TBE gels (Invitrogen) in 1X TBE running buffer at 160V for 45 minutes or until the dye front reached near the bottom of the gel. The gels were visualized using a Chemidoc gel imager (Bio-Rad).

Sample preparation for single particle cryo-EM

Purified and concentrated SPβ integrase (Int) or Int-RDF fusion proteins in storage buffer (20 mM Tris-Cl pH 8, 200 mM NaCl, 2 mM TCEP, and 20 % glycerol) were mixed with annealed mismatched or symmetrized dsDNA substrates to assemble protein-DNA complexes. These complexes contained either covalently trapped reaction intermediates with cleaved central dinucleotides or product dimer complexes with uncleaved central dinucleotides.

To generate and visualize covalently trapped intermediates, Int was mixed with *attPmm and attBmm*, whereas Int-RDF was mixed with *attLmm*. For the product dimer complexes, Int was mixed with *attLsym*, and Int-RDF was mixed with *attPsym* and *attBsym*. In all preparations, the final protein and DNA concentrations were 80 μM to 40 μM, respectively, in low salt buffer (20 mM Tris-Cl pH 8, 100mM NaCl, and 0.125mM TCEP).

Samples were incubated at 37 °C for 1 hour and purified using size-exclusion chromatography (SEC) on a Superose 6 10/300 GL column (Cytiva) pre-equilibrated with low salt buffer (20mM Tris-Cl pH 8, 100mM NaCl, and 0.125mM TCEP). Fractions containing the target complexes were pooled and concentrated to 5mg/ml using using Amicon Ultra Centrifugal Filter (30kDa MWCO, MilliporeSigma).

Immediately prior to blotting and vitrification, samples were adjusted to a final concentration of 2mg/ml with low salt buffer (20mM Tric-Cl pH 8, 100mM NaCl, and 0.125mM TCEP) and supplemented with fluorinated octyl maltoside (FOM, Anatrace) to a final concentration of 0.02 % (w/v). Then, 3.5 μl of the sample was applied to a freshly glow discharged CFlat 4.0/1.0 300-mesh grid or Quantifoil 1.2/1.3 300-mesh grid, blotted for 5.5 seconds with blot force set to 0, and plunged into liquid ethane using a Vitrobot™ MK IV (FEI) at 18 °C and 100 % relative humidity.

Data collection for single particle cryo-EM

Imaging of particles was performed using a Titan Krios G3i (Thermo Fisher Scientific) operating at 300 kV and equipped with a Gatan K3 direct detection camera. Data collection was conducted at the Advanced Electron Microscopy Facility, the University of Chicago. Each movie consisted of 54 frames with a total exposure of approximately 65 electrons per $Å^2$. The defocus range was set at -0.9 μm to -1.9 μm. Additional acquisition and imaging parameters are detailed in Table 1.

Image processing

All data processing steps were conducted using cryoSPARC (v4.6.0)[80] and are summarized in Figures S3, S8, S12, and S18. For both datasets, micrograph movies were aligned, summed, and dose-weighted; contrast transfer function (CTF) parameters were estimated using patch based CTF estimation from cryoSPARC. Micrographs with significant non-vitreous ice contamination or outliers in motion correction and CTF estimation were excluded.

Initial particle picking was performed using the blob particle picker to generate two-dimensional (2D) classifications. After a second round of 2D classification, selected classes showing intact complexes were used as templates for automated particle picking. False-positive or poorly defined particles were filtered out after additional rounds of 2D classification. The remaining particles were used to train a Topaz[79] model, which were applied to the combined micrographs for refined particle picking.

Particles classes showing clear secondary structures were iteratively classified, and three to five three-dimensional (3D) ab initio models were generated with C1 or C2 symmetry. Classes representing intact synaptic complexes were selected and further refined through heterogenous and non-uniform refinement steps. Resolutions for both Int and Int-RDF datasets were improved through focused refinement steps, where particles from both pre- and post-rotated classes were combined, resulting in the alignment of

Please cite this article in press as: Shin et al., Structural basis of directionality control in large serine integrases, Molecular Cell (2026), https://doi.org/10.1016/j.molcel.2026.09.028

the half complex. Reference-based motion correction was applied, followed by particle subtraction to remove the unaligned part of the other half complex, and the final set of particles were used for local refinement.

To refine flexible regions of the complexes, built-in 3D variability analysis[91] and 3D flex reconstruction[92] from cryoSPARC[80] were performed to address disordered regions and deconvolute the heterogeneous conformations of the complex into top three components ranked by significance in conformational variability. Using the outputs of these two analytical programs, we generated movies (see Videos S2, S5, S6, S7, S8, S9, S10, and S11) depicting the various conformational variabilities in the consensus maps and in the focused maps within UCSF Chimera (v.1.17.3).[85] Four maps from the last stage of the non-uniform refinement cycle, corresponding to pre- and post-rotated states in both the Int and Int-RDF datasets, were used as representative consensus maps. Two focused refined half-complex maps were then combined using the corresponding consensus map as a guide to generate composite maps for these rotational states. The final maps were sharpened using cryoSPARC's built-in map sharpening tools, and the composite maps for pre- and post- rotated states were further processed with deepEMhancer (v.0.14).[88]

Model building, refinement and validation

Alphafold2-multimer[81,93] was used to create the starting model of the synaptic complex. Specifically, the complex was predicted in overlapping fragments to allow flexibility in relative domain arrangements. These fragments were subsequently positioned and aligned according to the cryo-EM map using ChimeraX (v.1.4).[86] Overlapping regions were manually corrected, adjusted, and refined in COOT (v.0.9.8.8).[82] All of the DNA could be modeled except for the 3' most of the two overhanging Ts created by the DNA cleavage reaction. While most protein residues were built, some regions containing flexible loops, the 20aa long N-terminal tag, 28aa C-terminal residues and the linker connecting the Int and RDF were not built because they were disordered in the maps. The register of the main chain was carefully checked and corrected based on bulky residues.

B-form DNA starting models for the DNA substrates were built in ChimeraX and fitted into the cryo-EM map with local restraints using COOT.[82] ISOLDE (v.1.4),[87] embedded in UCSF ChimeraX,[86] was used to introduce phosphoserine linkages between the protein and DNA, and to correct most of peptide bond geometry, rotamer, and Ramachandran outliers.

Manual model building and refinement were performed iteratively using COOT,[82] ISOLDE,[87] and PHENIX (v.1.21)[83] real-space refinement. Both unsharpened and sharpened maps, as well as 3D-flex reconstructed maps, were used in conjunction with previously solved truncated structures and Alphafold2[81,93] predicted models to guide the model building process. Geometry restraints for zinc-binding sites were applied based on a zinc-binding model to ensure accurate representation of these regions.

Molecular graphics and structural analyses were conducted using UCSF ChimeraX and PyMOL (v. 3.1.3).[89]

QUANTIFICATION ANT STATISTICAL ANALYSIS

All in vitro and in vivo biochemical data were generated from independent biological replicates. Data were analyzed using Matplotlib or Gnuplot. No data were excluded from the analyses. No statistical methods were used to determine sample size. No blinding or randomization methods were used during these experiments.