



HAL
open science

Structural basis for broad DNA-specificity in integron recombination.

Douglas Macdonald, Gaëlle Demarre, Marie Bouvier, Didier Mazel, Deshmukh N. Gopaul

► **To cite this version:**

Douglas Macdonald, Gaëlle Demarre, Marie Bouvier, Didier Mazel, Deshmukh N. Gopaul. Structural basis for broad DNA-specificity in integron recombination.. *Nature*, 2006, 440 (7088), pp.1157-62. 10.1038/nature04643 . pasteur-00140781

HAL Id: pasteur-00140781

<https://pasteur.hal.science/pasteur-00140781v1>

Submitted on 10 Apr 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Saturday, Jan 28th, 2006

Structural basis for broad DNA specificity in integrin recombination

Douglas MacDonald¹, Gaëlle Demarre², Marie Bouvier², Didier Mazel² and Deshmukh N. Gopaul¹

¹Laboratoire de Biochimie et Biophysique des Macromolécules, Département de Biologie Structurale et Chimie. ²Unité Postulante Plasticité du Génome Bactérien, CNRS URA 2171, Département Structure et Dynamique des Génomes, Institut Pasteur, 75724, Paris, France

Lateral DNA transfer, the movement of genetic traits between bacteria, has a profound impact on genomic evolution and speciation. The efficiency with which bacteria incorporate genetic information reflects their capacity to adapt to changing environmental conditions. Integron integrases are proteins that mediate site-specific DNA recombination between a proximal primary site (*attI*) and a secondary target site (*attC*) found within mobile gene cassettes encoding resistance or virulence factors. The lack of sequence conservation among *attC* sites has led to the hypothesis that a sequence-independent structural recognition determinant must exist within *attC*. Here we report the crystal structure of an integron integrase bound to an *attC* substrate. The structure shows that DNA target site recognition and high-order synaptic assembly are not dependent on canonical DNA but on the position of two flipped-out bases that interact in *cis* and in *trans* with the integrase. These extrahelical bases, one required for recombination *in vivo*, originate from folding of the bottom strand of *attC* due to its imperfect internal dyad symmetry. The mechanism reported here supports a new paradigm for how sequence-degenerate single-stranded genetic material is recognized and exchanged between bacteria.

Nature has evolved several classes of enzymes that mediate the capture and spread of genetic information between bacteria¹. The efficiency by which these genetic traits are utilized is dependent on the rate of donor DNA delivery, their genomic incorporation, and level of gene expression. Novel traits are incorporated by several methods, among these recombination by site-specific recombinases play an important

role². These enzymes require short matching DNA sequences between the donor and genomic DNA which undoubtedly reduces the rate of DNA assimilation among highly diverse genetic populations.

Integron integrases (IntI) are site-specific recombinases that form a subclass within the tyrosine recombinase family due to the presence of a unique insertion that is required for activity³. Unlike other members of the family such as bacteriophage P1 Cre and yeast Flp proteins^{4,5}, IntI can mediate the exchange of DNA between two architecturally distinct sites even though homology predicts only one DNA binding domain⁶. The mechanism by which IntI achieves this dual site-specificity is unknown. IntIs catalyze an insertion between a primary recombination site (*attI*), located within a DNA element called an integron, and secondary target sites (*attC*) located within mobile gene cassettes⁷. These insertions are balanced by the excision of gene cassettes that occurs between two *attC* sites (Fig. 1a).

The *attI* site, like other site-specific recombinase binding sites, contains a core of short symmetrical dyad sequences at its recombination crossover point as well as two upstream secondary sites that are potentially either regulatory or repressive towards illegitimate recombination^{8,9}. The *attC* sites in contrast display poor sequence conservation and vary in length (57 to 141 bp), containing only short regions of sequence similarity at their boundaries. These conserved regions are separated by a stretch of imperfect internal dyad symmetry⁷.

One might have expected that excision of a gene cassette would occur via the classic model of Holliday junction (HJ) formation and resolution using two duplex *attC* sites as observed with Cre-mediated loxP recombination¹⁰ (Fig. 1b). However, we have

proposed an alternative pathway for IntI recombination that involves hairpin substrates formed from the bottom strand of the *attC* site in which the HJ intermediate is resolved by yet unknown cellular factors or perhaps DNA replication^{7,11-16} (Fig. 1c). The transition to these complex substrates requires ssDNA stages that can potentially be generated through either transformation or conjugation¹⁷.

To address the structural attributes of our model, we have determined the crystal structure of IntI from *Vibrio cholerae* (VchIntIA) bound to a complex substrate (*attC*) that was derived from the bottom strand of a *V. cholerae* repeated sequence (VCRs). VCRs are the *attC* sites of superintegrons¹⁷. The structure shows that the site of recombination along the DNA backbone is determined by the position of two pre-existing extrahelical bases that act not only as molecular markers, positioning the integrase along the DNA, but also mediate the high-order assembly of the synaptic complex. The remainder of the protein-DNA interface is composed almost entirely of non-specific protein to DNA phosphate interactions. This structural mechanism of recognition and assembly allows a greater diversity of genetic traits to be captured and exchanged during lateral DNA transfer thus increasing the rate of bacterial speciation.

A Folded Single-Stranded Recombination Substrate

The DNA construct for co-crystallization was based on the conserved features of the predicted secondary structures of numerous *attC* bottom strands. (Fig.1d, SF1a). The resulting bulged duplex, VCR_{bs}, formed discrete complexes with VchIntIA in electrophoretic mobility shift assays (EMSA, see below). In addition SDS-PAGE analysis of equilibrium mixtures of VchIntIA-VCR_{bs} complexes revealed a band with

reduced mobility representing ~5% of the total protein. Mass spectrometry confirmed this slower migrating species to be covalently linked VchIntIA-VCR_{bs} molecules. This suggested the nucleophilic tyrosine (Tyr 302) within the integrase had formed the characteristic phosphoprotein intermediate.

The structure of the non-covalent VchIntIA-VCR_{bs} complex was determined to 2.8 Å using phases obtained from a single wavelength anomalous diffraction (SAD) experiment on selenomethionine-labeled protein. These phases resulted in a readily interpretable electron density map. Model building and refinement allowed for rapid convergence to $R_{\text{cryst}} = 0.234$ and $R_{\text{free}} = 0.262$ for 48-2.80 Å resolution data and good geometry. Data collection, phasing statistics and refinement results are summarized in Supplementary Table S1.

Architecture of the VchIntIA-VCR_{bs} Excision Complex

The VchIntIA-VCR_{bs} complex contains four VchIntIA molecules bound to two antiparallel VCR_{bs} duplexes (Fig. 2a). This constitutes a recombination synapse representing the step preceding first strand cleavage in an *attC* x *attC* cassette excision reaction (Fig. 1c step 2). Two of the VchIntIA molecules (subunits A and C) have Tyr 302 adjacent (~3 Å) to the DNA backbone and thus attacking the scissile phosphate between nucleotides A14' and C15' on strand 1 (Fig. SF2a). The equivalent tyrosines in the B/D subunits are ~7 Å away (Fig. SF2b). The amino acids in each active site are all derived from the same subunit i.e. provided in *cis*. The extrahelical base T12" is stabilized by *cis*-interactions within subunits B/D and is important for DNA site recognition (see below). The extrahelical base G20" is buried in a deep hydrophobic

pocket, located in subunits A/C bound to the other VCR_{bs} duplex, thus forming a set of *trans*-interactions which hold the synaptic complex together. The protein-protein interfaces between subunits bound to the same VCR_{bs} differ from those bound in *trans* across the synapse. This arrangement yields an overall synapse that is only two-fold symmetric.

Structure of VchIntIA Recombinase

VchIntIA folds into two distinct domains (Fig. SF3a). The N-terminal domain (residues 1-85) contains four α -helices (α A- α D) organized as helix turn helix motifs nearly orthogonal to each other pair-wise (α A- α B, α C- α D), thus resembling the corresponding λ Int and XerD^{18,19} folds. Helices α B and α D contact the major groove of the VCR_{bs}, whilst helix α A is involved in inter-subunit contacts across the synaptic complex (Fig. 2, 3). The four separate N-terminal domains have a low root mean square deviation (~ 0.5 Å) between them.

The C-terminal domain (residues 105-320) contains the characteristic insertion (residues 192-210, Fig. SF3b) found only in IntIs. The α I₂ helix within this essential region³, as discussed below, plays an important role in synapse formation. The rest of the C-terminal domain is structurally similar to other tyrosine recombinase family members^{10,18,20-22}.

Substrate Recognition: An Adaptive Molecular Switch

VchIntIA binds its *attC* substrate as a dimer. However, unlike other tyrosine recombinases it does not form symmetrical protein-DNA contacts (Fig. SF4). The

attacking subunit (green, Fig. 3) positions Tyr 302 to engage the scissile phosphate, and to trigger first strand cleavage and transfer. The N- and C-terminal domains of this subunit wrap around the DNA forming a clamp burying 4,000 Å² of accessible surface area. Two base contacts are made at this half-site both via Lys 160. The positioning of this invariant catalytic residue is key, as it may play a role in the protonation of the 5'-hydroxy leaving group after strand cleavage²³. The remainder of this interface is characterized by protein-DNA backbone phosphate contacts.

The non-attacking integrase subunit (magenta, Fig. 3) forms two distinct protein-DNA contact points burying 5,000 Å² of accessible surface area. One of these contacts contains the β-4,5 hairpin interacting with the flipped out nucleotide T12", dictates the position of the integrase dimer along the DNA (Fig. 4a, b). This extrahelical base is inserted between two stacked histidines (His 240 and His 241), invariant among IntIs, at one end and a highly conserved proline (Pro 232) at the other end, to form a tight non-polar nucleotide protein interface. Interruption of this interface alters VchIntIA's ability to bind the VCR_{bs} duplex *in vitro* as discussed below.

The larger protein-DNA contact of this non-attacking subunit forms a DNA footprint that is roughly a mirror image of the attacking subunit interface covering half a helical turn of DNA, and sharing many of the equivalent non-specific protein-DNA phosphate contacts (Fig. SF4). Interestingly, this subunit while centered on G20" does not make direct contacts with this extrahelical base. Instead this flipped out base binds within a deep hydrophobic pocket located in the attacking subunit (C) across the synaptic interface (Fig. 2b).

IntIs must recognize two architecturally distinct sites during cassette integration (*attI* x *attC*). To achieve this dual site-specificity, we propose that IntIs have developed a molecular switch that allows sequence degenerate binding (*attC*), as seen in the VchIntIA-VCR_{bs} complex reported here, as well as sequence dependent binding during *attI* site recognition. The molecular switch that adapts to these two different modes of binding is the β -4,5 hairpin. During VCR_{bs} (*attC*) binding the β -4,5 hairpin is found in two environments corresponding to each unique half site along the DNA. In the non-attacking subunit the β -4,5 hairpin recognizes the molecular marker T12" (Fig. 4a,b), whilst in the attacking one it does not contact the DNA (Fig. 4c). We suggest that in the attacking interface, the position of the β -4,5 may have rotated away from the minor groove relative to its position when IntIs are in complex with an *attI* site, possibly due to the *trans*-interaction mediated by G20" (Fig. SF5). During *attI* binding the β -4,5 hairpin would make contacts with the minor groove analogous to the equivalent disposition of this element in the Cre-lox system¹⁰. The increased DNA footprint for *attI* site binding (Fig. SF4) relative to the *attC* interface reported here supports such a hypothesis⁸.

Synaptic Assembly: A Multiple Structure Problem

Assembly of a tyrosine recombinase synapse is mediated by a highly-specific set of inter-subunit protein-protein interactions, induced by DNA target capture. This stepwise assembly ensures that strand exchange takes place only between appropriate DNA substrates. How can IntIs, which bind a vast array of *attC* sequences, guarantee competent assembly of their DNA excision synapse? The strategy they have adopted is to use an invariant flipped out DNA base, acting as a linch-pin, to position the two dimers

in the synapse (Fig. 4 c,d). This flipped out base (G20") guarantees proper geometrical assembly by inserting itself within the hydrophobic pocket, created by Trp 157 and 219, across the synaptic interface. This interaction allows helix αI_2 from the attacking subunit(s) to form several important DNA contacts in *trans* holding the synapse together.

These *trans*-interactions mediated by G20" result in only a two-fold symmetric synapse which may inhibit HJ isomerization as discussed below. Subunits bound to the same VCR_{bs} bury only 1,600 Å² of solvent-accessible protein surface, by docking their C-terminal helix (αN) of the attacking subunit in the neighboring non-attacking subunit^{10,24} (Fig. 2b, 3). The spacing and geometry of the bases in the central region of the DNA duplex preclude intersubunit interaction between the N-termini on the same VCR_{bs} (Fig. 2a, 3). The more extensive intersubunit interface across the synapse buries 4,000 Å² of accessible surface area. These interfacial contacts are due namely the C-terminal helix (αN) exchange, helix αA in the attacking subunits contacting the helix-turn-helix region of the αC - αD helices within the non-attacking subunits across the synapse, and the unusual *trans*- interactions mediated by extrahelical base G20".

Integron Site-Specific Recombination

Recently, exploiting conjugation as a medium to exclusively deliver single stranded *attC* substrates, we demonstrated that the bottom strand of *attC* recombined with a resident *attI* at a rate of 1000-fold higher than the comparable top strand of *attC*¹². Disruption of the postulated¹³⁻¹⁶ secondary structure of *attC* affected recombination. Based on these results we proposed a recombination model (Fig. 1c). The structure of the

VchIntIA-VCR_{bs} complex presented here provides a structural basis for IntI mediated site-specific recombination using the bottom strand of *attC* as a substrate.

The biological relevance of the observed structural aspects of our model, i.e. the *cis*- and *trans*-interactions observed in the VchIntIA-VCR_{bs} complex, were investigated using a combination of EMSA and *in vivo* excision assays¹³. Four discrete bands are observed with wild type VchIntIA-VCR_{bs} (SF6a), potentially corresponding to the occupancy of the four available half sites. Deletion of T12" from VCR_{bs} significantly reduced the overall amount of complex formation as judged by EMSA (Fig. SF6a), with however, only a 5-fold drop in excision frequency (Fig. SF6c). Mutagenesis of His 240 (H240V), which interacts with T12" in the crystal, did not allow us to further investigate this interaction since the mutant protein formed aggregates with VCR_{bs} (Fig. SF6b, lane 5). Interestingly deletion of G20" from VCR_{bs} affected both the intensity and the migration pattern, with a quantitative increase of the fastest migrating band, (Fig. SF6a). This suggests a higher occupancy of only one half site and change in the effective radius/molecular weight or nature of higher order complexes. This same mutated sequence tested *in vivo* resulted in a 10,000-fold drop in excision activity (Fig. SF6c). Preliminary *in vivo* mutagenesis experiments of key residues (W157I and W219I), also produced a large reduction (1000-fold) in excision frequency. These results suggest that G20" plays a central role in maintaining an active VchIntIA-VCR_{bs} synapse, but that other factors namely C-terminal helix exchange and B/C and A/D interfaces also contribute to the assembly of the tetrameric synapse.

A folded single-strand *attC* substrate during integron recombination necessitates that second strand cleavage and transfer is down-regulated relative to most members of

the tyrosine recombinase family. A second round of cleavage and transfer does not lead to the excision/insertion of the gene cassettes but only to rearrangements within the *attC* sites (Fig. SF7). This regulation likely occurs at the HJ isomerization step and hence perturbing the next stage of the reaction in IntI recombination²⁵. Studies based on λ , XerC/D, Flp and Cre systems support models of HJ isomerization that involve only limited branch migration resulting in subtle movements of the quaternary structure^{10,20,26-28}. The two isomeric HJ intermediates have structures that are similar except for an exchange in the roles of their DNA strands²⁹. In this model the protein-protein interface between subunits is maintained. Small changes between subunits, which result in the loss of interfacial binding energy, are regained by reciprocal changes in symmetrically related interfaces. These HJ intermediates thus have similar free energies which explains the lack of bias observed in their resolution in the Cre and Flp systems³⁰. These reciprocal inter-subunit changes require a pseudo four-fold symmetric synapse. We suggest that since the VchIntIA-VCR_{bs} synapse is only two-fold symmetric and thus contains non-equivalent subunit interfaces (Fig. 2a), that the reciprocal changes needed to maintain the iso-energetic intermediates are lost. This may lead to an increased energy barrier for isomerization and thus produce a larger population of HJ intermediates that have not isomerized i.e. with a bias to revert to the original substrates. However this group of stalled HJ intermediates needs then to be resolved via other cellular processes (Fig. 1c).

In addition the rotation ($\sim 15^\circ$) of the C-terminal domains within the non-attacking subunits could also reduce the rate of second strand cleavage (Fig. SF8). This movement is a consequence of binding the extrahelical base T12". It results in the translation of

helix α M and repositions nucleophilic Tyr 302 away from the DNA backbone to a distance of ~ 7 Å in the non-attacking subunits.

Increased Genomic Diversity via Broad DNA Specificity

Several mechanisms mediate genetic exchange between prokaryotes, but the introduction of foreign DNA does not guarantee its assimilation. Processes such as DNA recombination are needed if persistence as an episome is not established. These recombination systems, however, place restrictions on the type of DNA that can be transferred. Homologous recombination primarily rearranges sequences among closely related taxa and is unlikely to allow for the introduction of novel traits³¹. Site-specific recombination by contrast can mediate the exchange of more diverse traits but normally these systems are still constrained by the need for the necessary core-sites for recognition and catalysis². By contrast, IntIs have evolved an ingenious way of allowing highly diverse genetic traits, containing little sequence homology at their DNA target sites, to be recognized and assimilated into the integron element. Although the positioning of DNA bases in an extrahelical geometry for genomic repair is well documented³², the use of extrahelical bases for DNA recognition of single stranded DNA as well as assembly of higher-ordered protein DNA complexes is gathering more support^{12,33,34}. We suspect that since most of the DNA exchanged via lateral transfer is in the single-stranded form, be it through natural transformation or conjugation³⁵, that this may have contributed to the development of the mechanism reported here.

Fig. Legends

Fig. 1, Pathways of IntI mediated cassette excision. **a**, Integrons contain a gene, *IntI*, encoding a tyrosine recombinase and an adjacent recombination site, *attI*. Gene cassettes (ORFs) are flanked by secondary sites, *attCs*. IntI recombines *attI* and *attC* during integration and two *attC* sites during excision. P_i , P_c promoters for *IntI* and gene cassettes, respectively; DR1, DR2 directly repeated accessory binding sites; L, R binding sites within the core region of *attI*; ORF, open reading frame; L', L'' inner repeats; R', R'' flanking repeats. **b**, Excision via the classic tyrosine recombinase model. Each duplex *attC* site (1b.1) is bound by two IntI molecules to form an antiparallel recombination synapse (1b.2). Tyr 302 cleavage forms covalent 3'phosphotyrosine intermediates (1b.3). The free 5'-hydroxyl groups attack their partner substrates yielding a HJ intermediate (1b.4), which isomerizes (1b.5) before undergoing a second round of cleavage and strand-exchange reactions to yield the recombinant products^{5,6} (1b.6). **c**, Proposed IntI excision via a single-stranded DNA substrate pathway. The bottom strand of the integron element, produced via conjugation or transformation, folds upon itself to yield an active stem-loop substrate (1c.1). Two IntI molecules bind each folded *attC* site to form an antiparallel recombination synapse (1c.2). The attack and strand exchange steps proceed in a similar fashion to 1b.3-4, however, the HJ intermediate requires cellular components in order to be resolved¹² (1c.5-6). The reaction intermediate shown 1c.2 represents the VchIntIA-VCR_{bs} structure described here. IntI molecules colored green and magenta, respectively, are potentially active or non-active for cleavage. **d**, DNA sequence, VCR_{bs} used to form VchIntIA- DNA co-crystals. Yellow boxes highlight the inner (L' and L'')

and flanking (R' and R'') repeats. The nucleotides, T12'' (red) and G20'' (blue), have an extrahelical geometry upon folding of *attC* bottom strands (Fig. SF1).

Fig. 2, Architecture of the VchIntIA-VCR_{bs} synapse. **a**, N-terminal view of the complex. Four VchIntIA molecules bind two antiparallel VCR_{bs} duplexes to form the active synapse. The extrahelical base T12'' (red) is stabilized by *cis*-interactions and is involved with DNA site recognition (Fig. 5a,b). Extrahelical base G20'' (blue) is buried in subunits that are bound to the other VCR_{bs} duplex forming a set of *trans*-interactions (Fig. 4c,d). The non-symmetric interfaces between VchIntIA molecules yield a two-fold symmetric synapse. **b**, Orthogonal view with respect to **a**. The C-terminal helices (N) bury one face in a hydrophobic pocket of the adjacent subunit in a cyclic manner (N_A→B, N_B→C, ..etc).

Fig. 3, Stereo model of half the VchIntIA-VCR_{bs} synapse. The attacking VchIntIA subunit is shown in green (right) and the non-attacking VchIntIA subunit in magenta (left). While both subunits use their N- and C-terminal domains to encompass the DNA forming a clamp, they do not share equivalent protein-DNA contacts (Fig. SF4). In the non-attacking subunit the β 4,5 hairpin contacts the DNA via the extrahelical base T12''. Tyr 302 from both subunits is shown in pink.

Fig. 4, Cis- and trans- extrahelical interactions. **a**, Interaction with β-4,5 from a non-attacking subunit and the extrahelical base T12''. **b**, Detailed view of the *cis*-interaction depicted in **a**. Base T12'' (red) forms a protein-nucleic interface by stacking between His

240 and Pro 232. Experimental electron densities after phase modification are shown (1.5- σ contour level). **c**, Ribbon diagram showing G20" (blue) binding in *trans*- between Trp 157 & Trp 219. The β -4,5 hairpin from the attacking subunit does not contact the DNA. **d**, Detailed view of the *trans*-interactions shown in **c**. The NH₂ group of G20" forms hydrogen bonds with the O γ of Glu 145 and N1 of Trp 157 within the attacking subunits. Non-specific contacts are made with the adjoining non-attacking subunit (magenta). A 2.8 Å 2F_o-F_c simulated annealed omit electron density map (1.6- σ contour level) is shown. G20" and all protein residues labeled in **d** were omitted during map calculation.

Methods

Structure determination and refinement. The preparation of VchIntIA and the DNA used to form cocrystals is described in the Supplementary Information. X-ray diffraction data to 2.8 Å was collected at beamline ID14-4 of the European Synchrotron Radiation Facility (ESRF). A wavelength of 0.9793 Å, corresponding to the SeMet Peak, was used for data collection at 100K on a single cocrystal. X-ray data were processed with XDS³⁶ and CCP4³⁷ program suites. Phases for the VchIntIA-VCR_{bs} complex were determined by using the single wavelength anomalous dispersion (SAD) method with selenomethionine (SeMet) labeled protein. The position of substructure atoms, initial phases and solvent flattening were determined using BnP³⁸. The atomic model was fitted into electron density using the program O³⁹. The asymmetric unit contained four VchIntIA molecules and two antiparallel VCR_{bs} duplexes. Residues 1 and 305-310 in subunit A and 305-307 in subunit C were not fit into the density. For the DNA helices,

nucleotides 1-4 and 36-40 of chain E, nucleotides 1-5 and 40-43 of chain F, nucleotides 1-5 and 35-40 of chain G, and nucleotides 1-6 and 39-43 of chain H were not fit due to poor electron density. DNA chains E,G correspond to strand 1, DNA chains F,H correspond to strand 2 in Figure 1d. The four protein subunits and DNA duplexes were all independently modeled and refined with CNS⁴⁰ using maximum likelihood target using amplitudes and phase probability distribution with alternate cycles of manual density fitting. The structural similarity of subunits A/C and B/D warranted the use of non-crystallographic symmetry restraints in the final rounds of refinement. This resulted in a $R_{\text{cryst}} = 0.234$ and $R_{\text{free}} = 0.262$ at 2.8 Å, with good geometry for bond lengths and angles. All amino acids have (ϕ, ψ) backbone torsion angles in allowed regions of Ramachandran space. Statistics for the final model are given in Supplementary Table S1.

VCR_{bs} and VchIntIA mutagenic analysis. Details of EMSA and preparation of mutants is described in the Supplementary Information. The *in vivo* excision assay has been previously described¹³. Briefly, the assay measures the frequency at which a synthetic integron gene cassette, the reporter gene (*lacI*^q), is excised from its two flanking recombination sites, *attC*_{aadA7} and the VCR_{2/1}. In all experiments symmetrical mutations were introduced to both the VCR_{2/1} and *attC*_{aadA7} sites when possible. The equivalent base to T16" is not present within *attC*_{aadA7} (see Fig. SF1a). This protocol was followed since previous experiments showed mutations introduced in either site independently led to a slight decrease in deletion frequency, while mutations introduced at identical positions in tandem resulted in considerably larger reductions in recombination frequency. This suggests the presence of a wild type *attC* site assists in integrase complex formation between a mutated site, likely in a cooperative fashion.

This observation may be similar to the formation of the *attC* x *attI* synapse, where the extrahelical base G20" within the *attC* site drives formation of the active synapse (unpublished observations). Mutations in the *attC_{aadA7}* and VCR_{2/1} sites were introduced as described previously¹³.

The figures were produced with PYMOL (<http://www.pymol.org>)

Supplementary Information is linked to the online version of the paper at www.nature.com/nature. A movie showing the final model of the VchIntIA-VCR_{bs} synaptic complex is available in Supplementary Information.

Acknowledgements: This work was supported by lab startup funds from the Pasteur Institute to D.N.G., research grants from CNRS URA 2185 to D.N.G., research grants from the Institut Pasteur and from CNRS URA 2171 and GDR 2157 on transposable elements to D. Mazel, and a French Ministry of Research (MENESR) Grant to D. Mazel and D.N.G. D.M. is Florence Gould Scholar funded by the Pasteur Foundation of New York. G. D. and M. B. are doctoral fellows from the Fondation pour la Recherche Médicale and MENESR respectively. We also thank G. Liou for critical review of this manuscript.

Author Contributions: D.M. carried out the *in vitro* assays, structural determination and wrote the paper with editing from D.N.G. D. Mazel provided the conceptual framework for the substrate and invaluable background. G.D. and M.B. performed the *in vivo* excision frequency assays. All authors discussed the results and commented on the manuscript.

Author Information: Coordinates and structure factors are deposited in the PDB under accession code 2A3V. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare that they have no

competing financial interests. Correspondence and requests for materials can be addressed to D.N.G. (gopauld@pasteur.fr)

1. Kazazian, H. H., Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626-32. (2004).
2. Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M. *Mobile DNA II* (ASM press, Washington, DC, 2002).
3. Messier, N. & Roy, P. H. Integron integrases possess a unique additional domain necessary for activity. *J Bacteriol* **183**, 6699-706. (2001).
4. Chen, Y. & Rice, P. A. New insight into site-specific recombination from Flp recombinase-DNA structures. *Annu Rev Biophys Biomol Struct* **32**, 135-59. Epub 2003 Feb 11. (2003).
5. Van Duyne, G. *A Structural View of Tyrosine Recombinase site-specific Recombination* (ed. Nancy L. Craig, R. C., Martin Gellert & Alan M. Lambowitz) (ASM Press, Washington D.C., 2002).
6. Azaro, M. A. & Landy, A. *Lambda Integrase and the Lambda Int Family* (ed. Nancy L. Craig, R. C., Martin Gellert & Alan M. Lambowitz) (ASM Press, Washington D.C., 2002).
7. Stokes, H. W., O'Gorman, D. B., Recchia, G. D., Parsekhian, M. & Hall, R. M. Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol Microbiol* **26**, 731-45. (1997).
8. Gravel, A., Fournier, B. & Roy, P. H. DNA complexes obtained with the integron integrase IntI1 at the attI1 site. *Nucleic Acids Res* **26**, 4347-55. (1998).
9. Stark, W. M., Boocock, M. R. & Sherratt, D. J. Site-specific recombination by Tn3 resolvase. *Trends Genet* **5**, 304-9. (1989).
10. Guo, F., Gopaul, D. N. & van Duyne, G. D. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40-6. (1997).
11. Hansson, K., Skold, O. & Sundstrom, L. Non-palindromic attI sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Mol Microbiol* **26**, 441-53. (1997).
12. Bouvier, M., Demarre, G. & Mazel, D. Integron cassette insertion: a recombination process involving a folded single strand substrate. *Embo J* **8**, 8 (2005).
13. Biskri, L., Bouvier, M., Guerout, A. M., Boisnard, S. & Mazel, D. Comparative study of class 1 integron and *Vibrio cholerae* superintegron integrase activities. *J Bacteriol* **187**, 1740-50. (2005).
14. Rowe-Magnus, D. A., Guerout, A. M., Biskri, L., Bouige, P. & Mazel, D. Comparative analysis of superintegrons: engineering extensive genetic diversity in the Vibrionaceae. *Genome Res* **13**, 428-42. (2003).
15. Johansson, C., Kamali-Moghaddam, M. & Sundstrom, L. Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res* **32**, 4033-43. Print 2004. (2004).

16. Francia, M. V., Zabala, J. C., de la Cruz, F. & Garcia Lobo, J. M. The IntI1 integrin integrase preferentially binds single-stranded DNA of the attC site. *J Bacteriol* **181**, 6844-9. (1999).
17. Mazel, D., Dychinco, B., Webb, V. A. & Davies, J. A distinctive class of integron in the *Vibrio cholerae* genome. *Science* **280**, 605-8. (1998).
18. Subramanya, H. S. et al. Crystal structure of the site-specific recombinase, XerD. *Embo J* **16**, 5178-87. (1997).
19. Aihara, H., Kwon, H. J., Nunes-Duby, S. E., Landy, A. & Ellenberger, T. A conformational switch controls the DNA cleavage activity of lambda integrase. *Mol Cell* **12**, 187-98. (2003).
20. Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. & Rice, P. A. Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell* **6**, 885-97. (2000).
21. Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* **276**, 126-31. (1997).
22. Hickman, A. B., Waninger, S., Scocca, J. J. & Dyda, F. Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. *Cell* **89**, 227-37. (1997).
23. Krogh, B. O. & Shuman, S. Catalytic mechanism of DNA topoisomerase IB. *Mol Cell* **5**, 1035-41. (2000).
24. Rafferty, J. B. et al. Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. *Science* **274**, 415-21. (1996).
25. Gopaul, D. N. & Duyne, G. D. Structure and mechanism in site-specific recombination. *Curr Opin Struct Biol* **9**, 14-20. (1999).
26. Arciszewska, L. K., Grainge, I. & Sherratt, DJ. Action of site-specific recombinases XerC and XerD on tethered Holliday junctions. *Embo J* **16**, 3731-3743 (1997).
27. Biswas, T. et al. A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature* **435**, 1059-66. (2005).
28. Nunes-Duby, S. E., Azaro, M. A. & Landy, A. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr Biol* **5**, 139-48. (1995).
29. Gopaul, D. N., Guo, F. & Van Duyne, G. D. Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *Embo J* **17**, 4175-87. (1998).
30. Meyer-Leon, L., Inman, R. B. & Cox, M. M. Characterization of Holliday structures in FLP protein-promoted site-specific recombination. *Mol Cell Biol* **10**, 235-42. (1990).
31. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299-304. (2000).
32. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357-369 (1994).
33. Val, M. E. et al. The single-stranded genome of phage CTX is the form used for integration into the genome of *Vibrio cholerae*. *Mol Cell* **19**, 559-66. (2005).

34. Ronning, D. R. et al. Active site sharing and subterminal hairpin recognition in a new class of DNA transposases. *Mol Cell* **20**, 143-54. (2005).
35. Llosa, M., Gomis-Ruth, F. X., Coll, M. & de la Cruz Fd, F. Bacterial conjugation: a two-step mechanism for DNA transport. *Mol Microbiol* **45**, 1-8. (2002).
36. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795-800 (1993).
37. COLLABORATIVE COMPUTATIONAL PROJECT, N. The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst.* **D50**, 760-763 (1994).
38. Weeks, C. M., Blessing, R.H., Miller, R., Mungee, R., Potter, S.A., Rappleye, J., Smith, G.D., Xu, H., & Furey, W. Towards automated protein structure determination: BnP, the SnB-PHASES interface. *Z. Kristallogr.* **217**, 686-693 (2002).
39. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47**, 110-9. (1991).
40. Brunger, A. T. e. a. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54 (Pt5)**, 905-921 (1998).

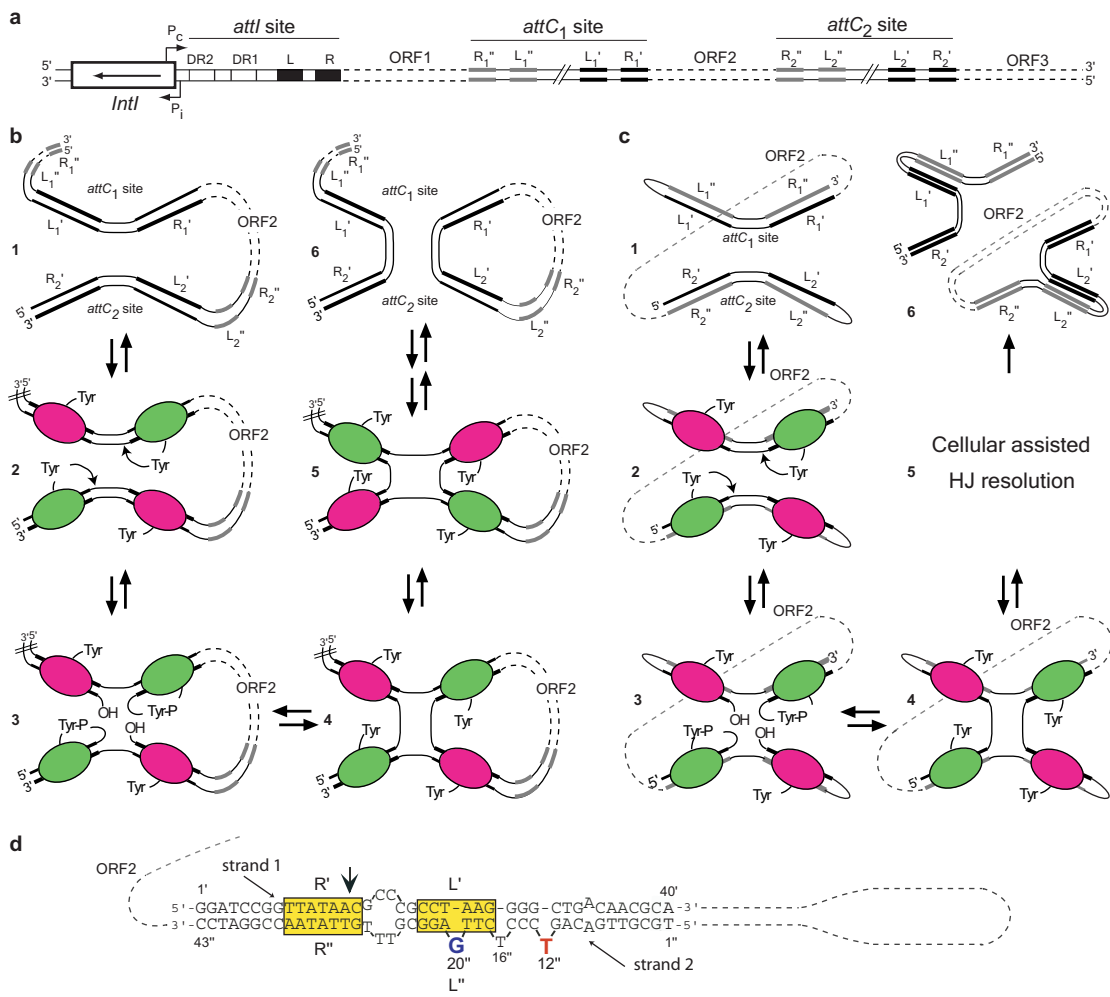


Figure 1

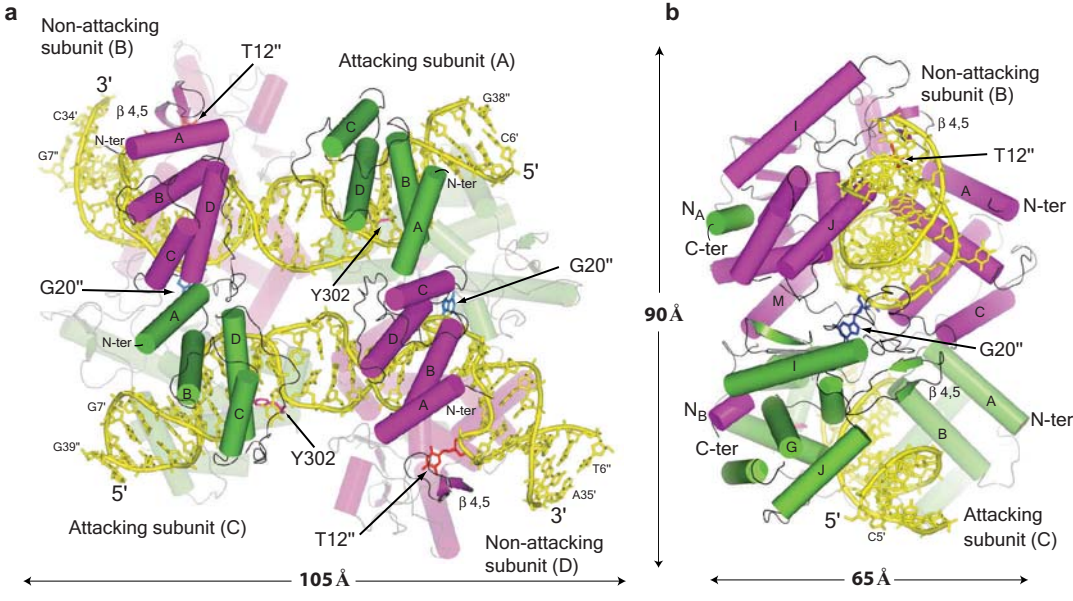


Figure 2

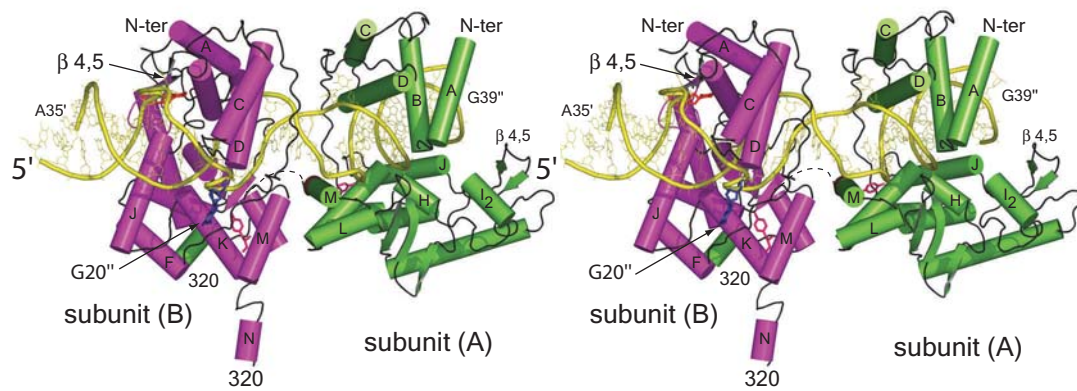


Figure 3

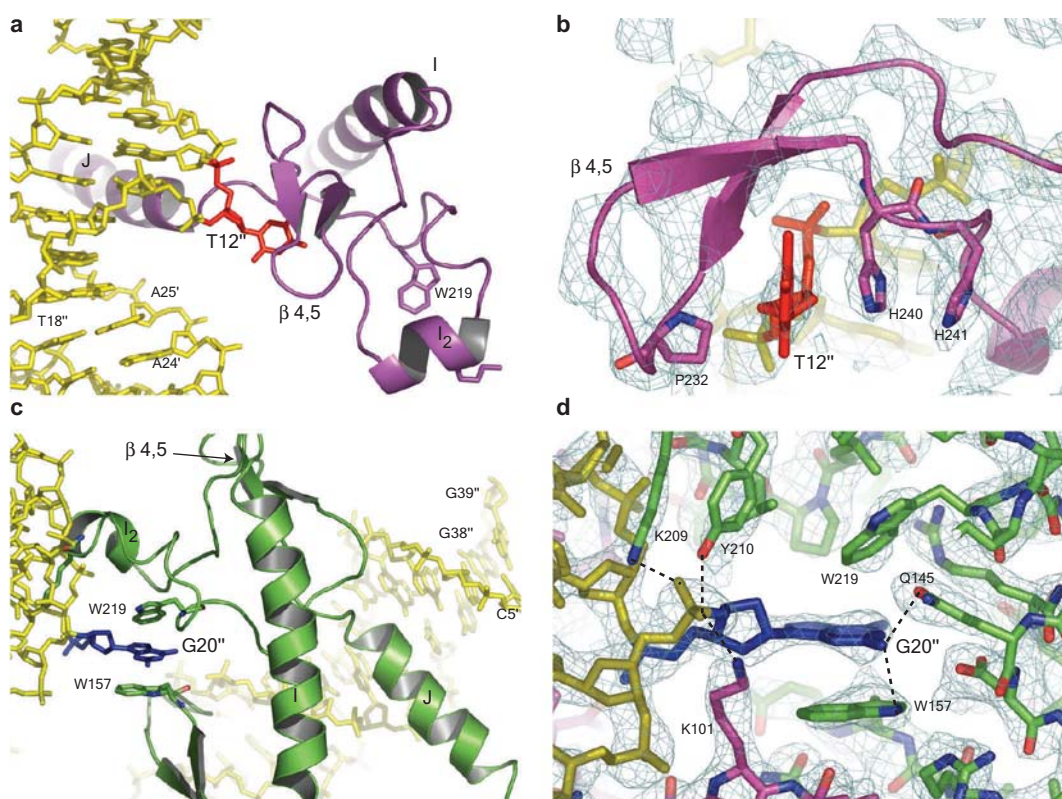


Figure 4

SUPPLEMENTARY INFORMATION

For MacDonald et al.

Crystallization of VchIntIA-VCR_{bs}. The gene for VchIntIA was amplified by polymerase chain reaction (PCR) from *V. cholerae* genomic DNA. The resulting coding region was ligated into a pET (Novagen) derived plasmid, downstream to a hexahistidine tag fused to a tobacco etch virus (TEV) protease site. Native, mutant and Se-Met VchIntIA were all purified from bacterial lysates on HiTrapTM Chelating HP (Amersham Biosciences) followed by cleavage with TEV protease. Proteins lacking a His-tag were separated from those with a N-terminal hexahistidine using a second HiTrapTM Chelating HP column and then purified to homogeneity with a Heparin HP column (Amersham Biosciences).

The DNA used to make the VCR_{bs} and a suicide substrate, by introducing a nick 3' to C15' on strand 1, were synthesized using standard phosphoramidite chemistry leaving the 5'-dimethoxytrityl (DMT) group attached after DNA synthesis. The DMT group was removed and the oligonucleotides purified using standard reverse phase chromatography by the manufacturer (Proligo, France). These purified DNA oligos were then applied to a TSK-Gel DEAE-5PW column (Tosoh Corporation) at 50°C and eluted with an increasing concentration of ammonium chloride. Appropriate fractions were concentrated and ethanol precipitated. All oligos and DNA duplexes were checked by denaturing and non-denaturing gel electrophoresis for purity.

Crystal trials using the vapor diffusion method at 18°C for both VCR_{bs} and the suicide substrate were undertaken. However, only crystals of the non-covalent VchIntIA-VCR_{bs} complex were obtained. These crystals were extensively washed and assayed

using SDS-PAGE with silver staining to confirm that no covalent molecules were present. Crystals were obtained by taking 2 μ l of VchIntIA-VCR_{bs} (2-4 mg/ml based on protein concentration), 10 mM HEPES (pH 7.4), 150 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 100 μ M VchIntIA and 100 μ M VCR_{bs} with 2 μ l of reservoir solution containing 50 mM MES (pH 6.4), 0.2 M ammonium acetate, 10 mM CaCl₂ and 8-10% PEG 4000. Orthorhombic crystals (a =149.9, b =170.2, c =209.4 Å; C222₁) grew within 3-5 days.

Electrophoretic mobility shift assays (EMSA). Strand 1 of the VCR_{bs} and the top strand of the random control (5'-TACGTCTACTGGGCTACTGATCGAGTTCCTGGCAAGCTGA-3') were 5' labeled with infrared dye 700 (IRD700, Li-Cor, Inc) during chemical synthesis. These oligodeoxynucleotides were annealed with equal molar ratios of strand 2 of VCR_{bs} or the appropriate mutant. For the random control the bottom strand was 5'-TCAGCTTCCAGGGACACTACGATCGTCAGCCCAGTAGACGTA-3' (see Fig. SF1b). A typical 20 μ l binding reaction contained 50 nM duplex DNA, 0-200 nM VchIntA, 50 mM HEPES (pH 7.5), 200 mM NaCl, 5% glycerol, 5 mM DTT, 0.5% (v/v) Tween-20, 50 μ g/ml BSA and 10 μ g/ml poly (dI·dC). Incubation for binding was for 30 min at 18°C and then 3 μ l of loading dye (0.25% bromophenol blue, 40% sucrose) was added. Separation took place on a 5% non-denaturing polyacrylamide gel at 18°C in 0.5X TBE buffer.

Generation of mutant VchIntIA. Mutants H35V, W157I, W219I and H240V used for EMSA analysis were generated using the QuikChange[®] Site-Directed Mutagenesis Kit (Stratagene). All mutants were sequenced and assayed for correct molecular weight by mass spectrometry.

Supplementary Information Figure Legends

Fig. SF1. Predicted secondary structures of *attC* bottom strands. **a,** Predicted secondary structures of *attC* bottom strands. The bottom strands from various *attC* sites were folded using the program mfold¹. Extrahelical bases that are in equivalent positions to T12" and G20" are highlighted in red and blue, respectively. The *attC* sites that have been tested for cassette excision competency are marked with an asterisk.

b, Construction of a random sequence version of the substrate. A random nucleotide sequence was used to generate a geometric mimick of the VCR_{bs}. This construct presumably disrupts the conserved core sequence (AAC) required for binding.

Fig. SF2. Stereo view of the VchIntIA-VCR_{bs} active sites. **a,** View of the attacking subunit where Tyr 302 is at a distance of ~ 3 Å from the DNA phosphate backbone. **b,** The non-attacking subunit where Tyr 302 is ~ 7 Å away from the DNA backbone and the catalytic Lys 160 is no longer positioned in the minor groove. The scissile phosphates are shown in red and distances < 3.5 Å, are depicted by dotted lines. A $2F_o - F_c$ simulated annealing omit map calculated at a resolution of 2.8 Å (1.2σ contour level) using the final model omitting from all four VchIntIA subunits the active site residues (R135, K160, R270, H271, H293 & Y302) and the four DNA basepairs (A13'-G16', T28"-T31"; attacking & A21"-C24", G20'-T23'; non-attacking) adjacent to Tyr 302 is shown.

Fig SF3. Structure of VchIntIA. **a,** Stereo model of VchIntIA from an attacking subunit. α -helices are labeled A-N and the positions of the amino and carboxy termini

are indicated. The I₂ helix is highlighted in yellow. **b**, Sequence alignment of VchIntIA and Cre recombinase with their corresponding secondary structure elements². The I₂ helix (yellow) is an insertion in the canonical fold. Active-site residues are boxed and the three positions within VchIntIA that significantly differ (~3Å), as determined by a difference distance matrix (DDM) analysis³, in the attacking and non-attacking subunits are underlined.

Fig. SF4. Schematic representation of the protein-DNA contacts between VchIntIA and VCR_{bs}. All hydrogen bonding protein contacts <3.5 Å are shown, with *trans*-interactions highlighted in bold. Protein-phosphate contacts are depicted by magenta circles and the position of base-specific hydrogen bonding is shown in green. Locations where modified nucleotides within IntI1's *attI* site interfere with protein binding have been overlaid on the equivalent bases of VCR_{bs} (red arrow heads). This suggests a larger DNA footprint for *attI* binding relative to *attC*. Contacts that are equivalent between the attacking and non-attacking interfaces are denoted by (Sym).

Fig. SF5. Proposed movement of the β 4,5 hairpin. **a**, An attacking subunit (green) from the VchIntIA-VCR_{bs} complex was superposed on the Cre subunit that has cleaved its DNA {Guo, 1997 #35} (cyan) by performing a least square fit analysis using equivalent Cα atoms within their active sites (VchIntIA:Cre, R135:R173, H267:H289,R270:R292 and H293:W315). The resulting overlay reveals that the DNA structure of the VCR_{bs} (yellow) and loxA (light brown), in the proximity of their

respective scissile phosphates (red), is similar. In the absence of the extrahelical base G20" (blue), such as found in a duplex *attI* recombination site, the interface between the β 4,5 hairpin and the I₂ helix within the attacking subunits may be disrupted causing the rotation of this hairpin back into the minor groove as seen in the Cre-loxA structure. Tyr 302 (pink) and Tyr 324 (cyan) from the VchIntIA and Cre recombinase molecules, respectively, are shown. **b**, Similar overlay except the non-attacking VchIntIA subunit (magenta) and the Cre monomer (cyan) that has not cleaved its DNA². Helices J from both integrases are shown in the major groove of the DNA.

Fig. SF6. VCR_{bs} and VchIntIA mutagenic analysis. **a**, EMSA of VCR_{bs} mutants.

Lane 1, complex formation with the VCR_{bs} used to form co-crystals. Lane 2-4, deletion of the extrahelical bases found within VCR_{bs}. Notice that removing G20" results in protein-DNA complexes that migrate faster relative to the other DNA sequences. Lane 5, a randomly chosen DNA sequence that contains extrahelical bases and the mismatched region, between the R & L boxes, in the same positions as VCR_{bs} (see Fig. SF1b for its predicted secondary structure) **b**, EMSA analysis of VchIntIA mutants. Lane 2, mutation of His 35, which interacts weakly with T16" in the VchIntIA-VCR_{bs} synapse, does not affect complex formation. Lane 3, mutation of Trp 157 which forms the lower half of the hydrophobic pocket that G20" intercalates into (see Fig. 4d) reduces protein binding. Trp 157 is conserved (~80%) among IntIs. Lane 4-5, mutation of Trp 219 and His 240, which are invariant among IntIs, yield proteins possibly changed in their stability, folding or solubility characteristics. W, aggregated DNA in well; C, protein-DNA complexes; F, free DNA. The EMSA depicted here were performed at 200 nM

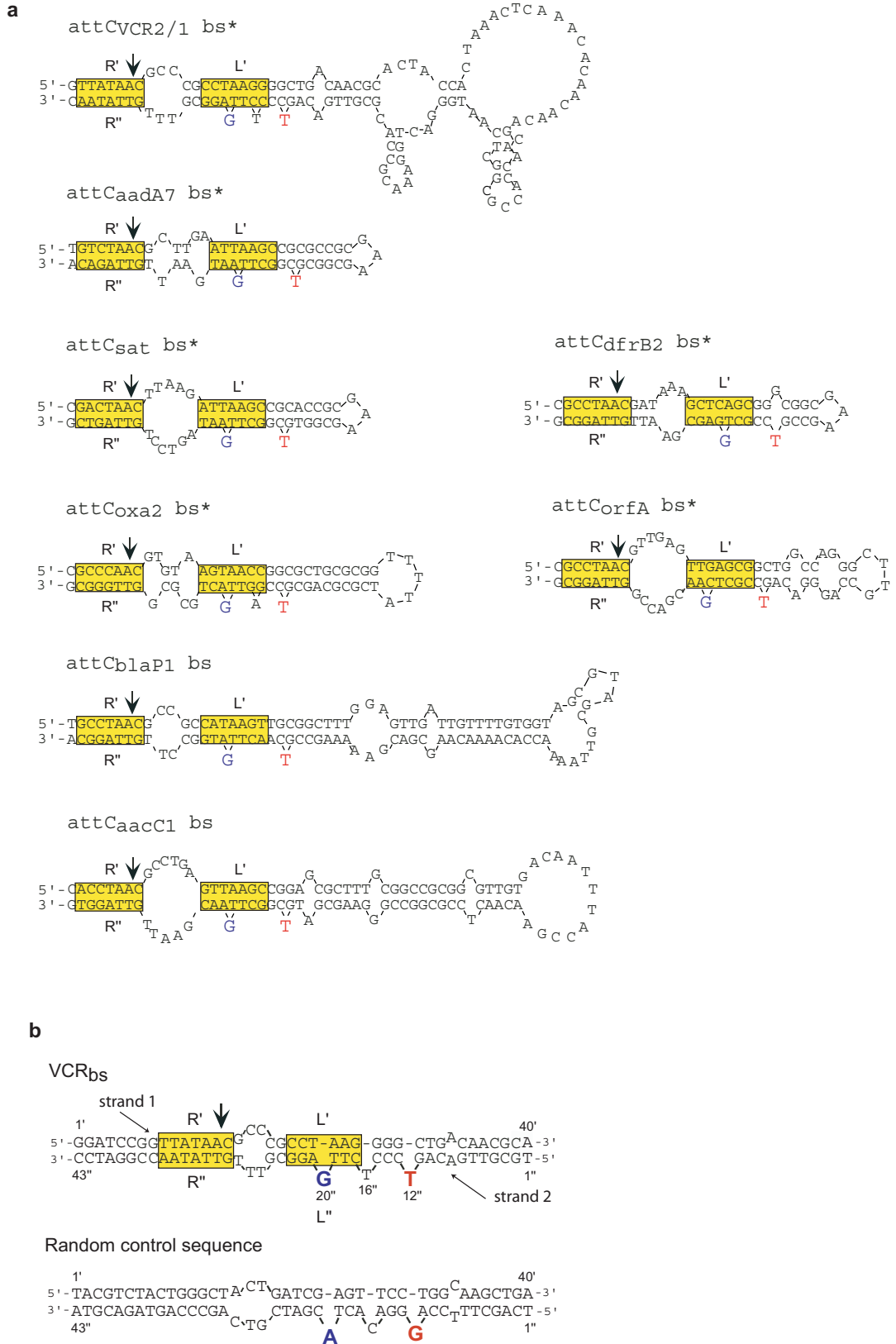
VchIntIA, see Supplementary Information for details. **c**, *In vivo* cassette deletion frequencies for the excision of a reporter gene (*lacI*^q) from its two flanking recombination sites, *attC*_{aadA7} and the VCR_{2/1}. Symmetrical DNA mutations were introduced into both *attC* sites when possible (see Methods). Removal of T12" results in a 5-fold drop in deletion frequency compared to 10,000-fold decrease when G20" is removed. This deletion (G20") renders the gene cassette defective for excision. The asterisk indicates that no recombinant colonies were detected. Error bars indicate standard deviations between three independent trials.

Fig. SF7. Proposed IntI double strand exchange pathway via a single-stranded DNA substrate. **a**, The integron element consists of a gene, *IntI*, encoding a tyrosine recombinase and an adjacent recombination site, *attI*. The gene cassettes (ORFs) when present are flanked by a secondary recombination site, *attC*. IntI recombinase *attI* and *attC* during integration and two *attC* sites during excision. P_i, P_c promoters for *IntI* and inserted gene cassettes, respectively; DR1, DR2 directly repeated accessory recombinase binding sites; L, R recombinase binding sites within the core region of *attI*; ORF, open reading frame; L', L" inner repeats; R',R" flanking repeats. **b**, The bottom strand of the integron element, produced via conjugation or transformation, folds upon itself due to the inverted repeat character of *attC* to yield an active substrate (1). Two IntI molecules bind each folded *attC* site to form a recombination synapse (2). Tyr 302 cleaves each substrate in an antiparallel arrangement (3) and the released 5'-hydroxyl groups become nucleophiles in strand-transfer reactions to form a HJ intermediate (4). Isomerization of the HJ intermediate (5) is followed by a second round of cleavage (6) and strand-

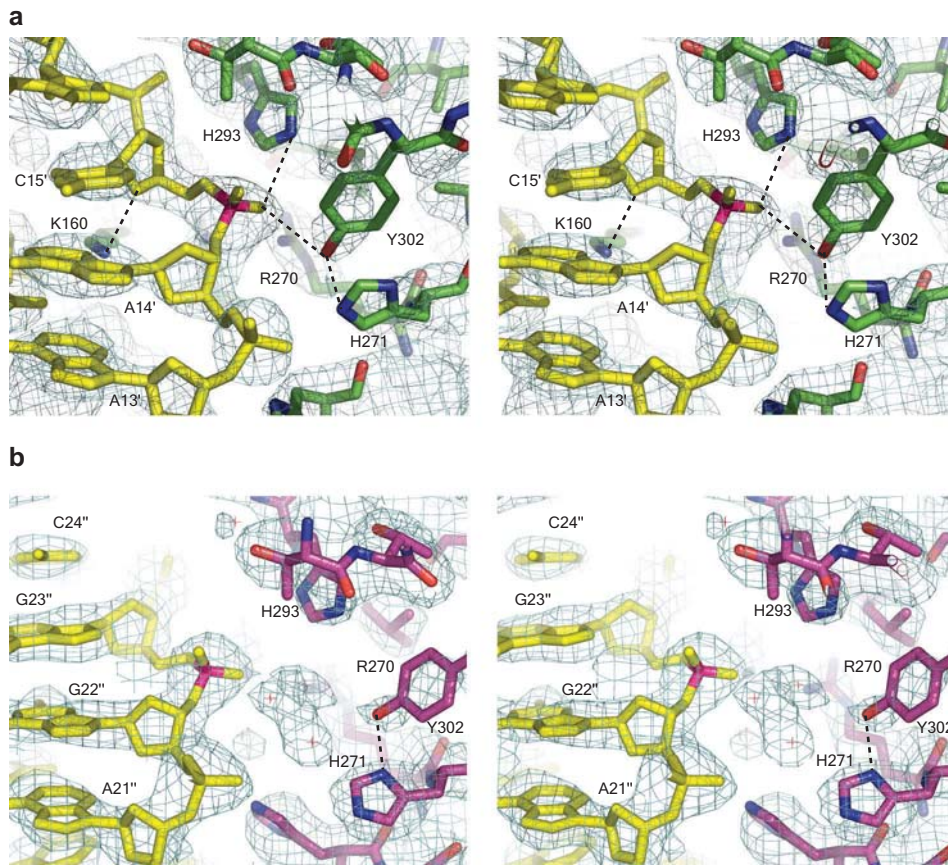
exchange reactions (7) to yield the recombinant products (8). This pathway does not result in excision of the gene cassette (ORF2), but only the swapping of the inner repeat sequence (L", L') between the two *attC* sites.

Fig. SF8. Rotation of the C-terminal domain within the non-attacking subunits. a, An attacking subunit (A) bound to its half site was superimposed on a non-attacking subunit (B) bound to its half site by performing a least square fit analysis with the C α atoms from the N-terminal domains (residues 4-84) of both subunits. The resulting ribbon model diagram shows that subunit B (magenta), which is non-active for cleavage, has rotated $\sim 15^\circ$ relative to subunit A (green) due to its binding of the extrahelical base T12'' (red). **b,** View, rotated 180° about a vertical axis relative to **a**. The rotation of the C-terminal domain of subunit B results in the translation of its M helix, which contains Tyr 302, away from the DNA backbone to a distance of $\sim 7 \text{ \AA}$ compared to $\sim 3 \text{ \AA}$ in subunit A. Both DNA half sites associated with the attacking (A) and the non-attacking (B) subunits are shown and have similar structures at their corresponding scissile phosphates (red). The N-terminus for both subunits has been omitted in **b** for clarity.

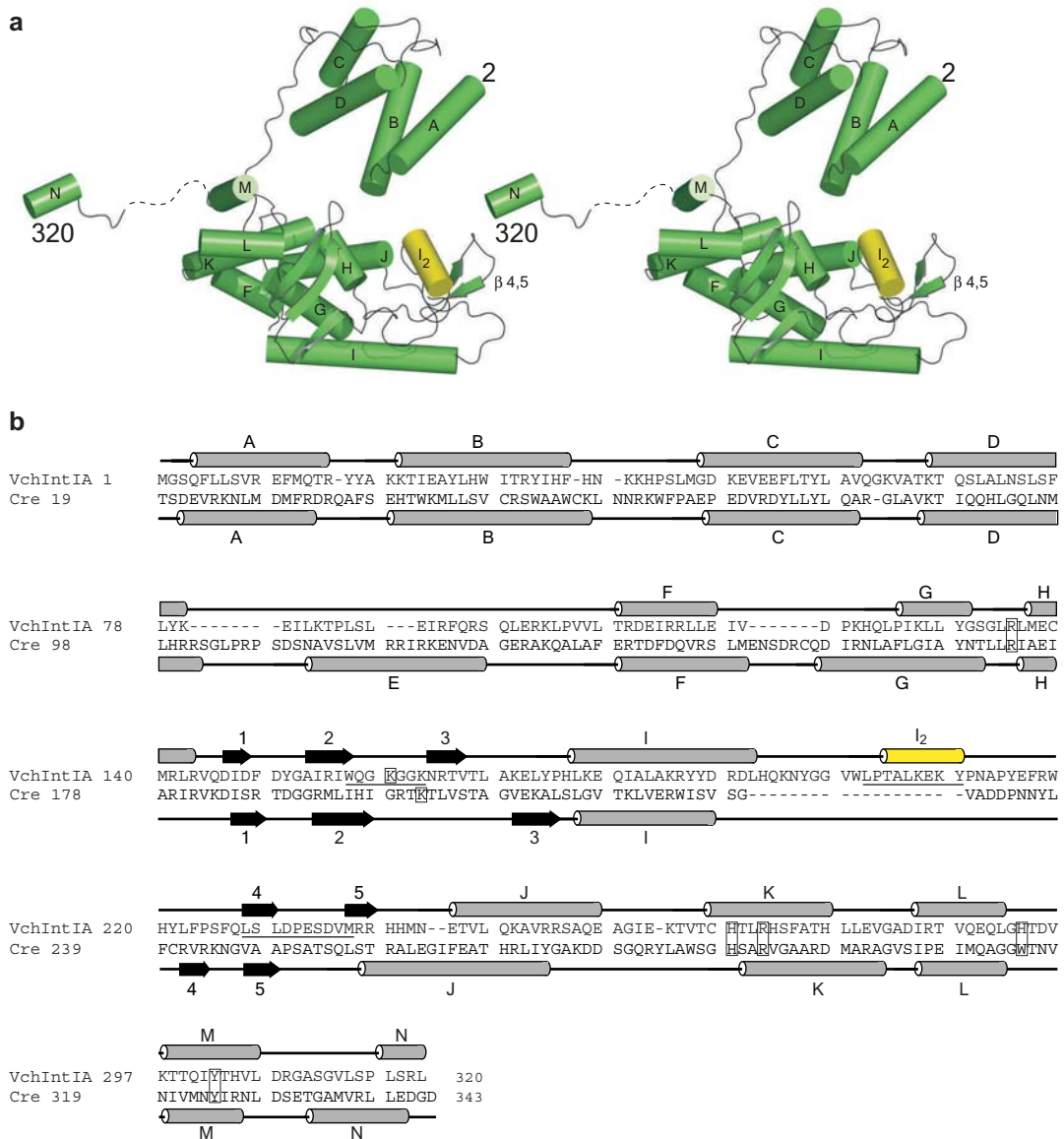
1. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-15. (2003).
2. Guo, F., Gopaul, D. N. & van Duyne, G. D. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40-6. (1997).
3. Richards, F. M. & Kundrot, C. E. Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure. *Proteins* **3**, 71-84. (1988).



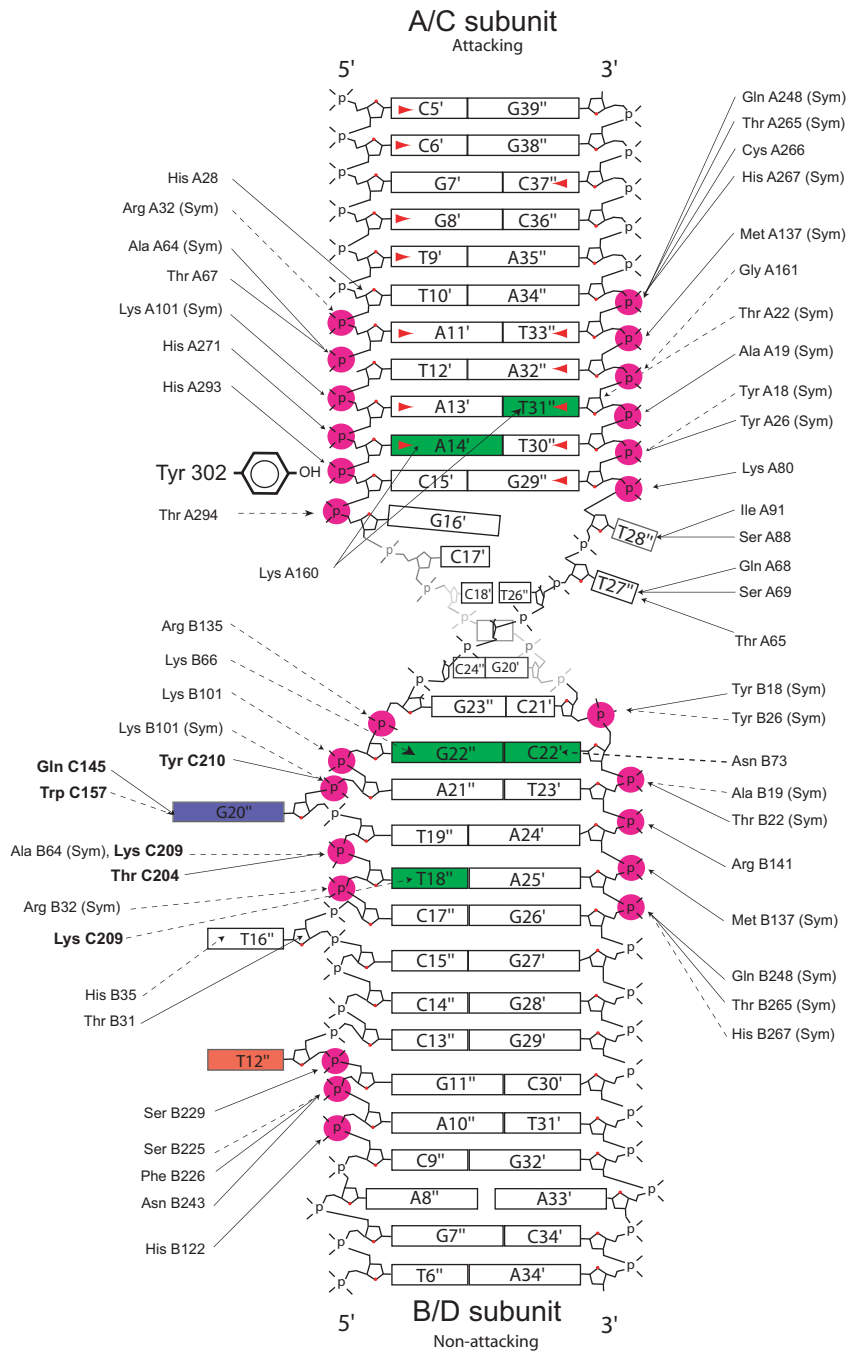
Supplementary Figure 1



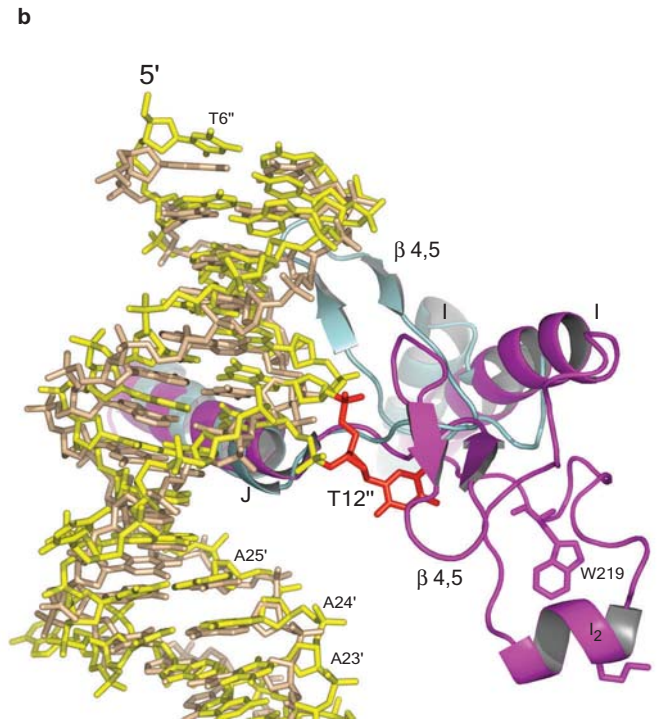
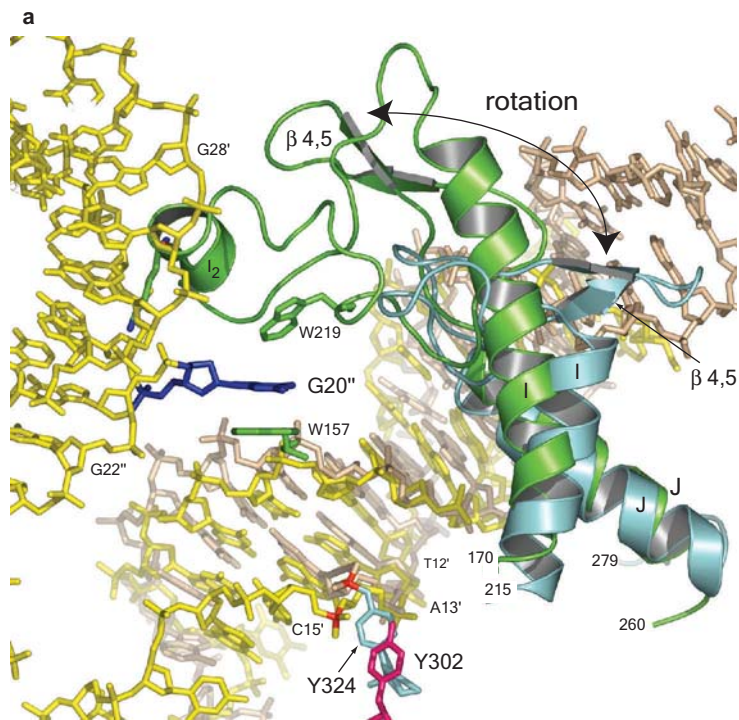
Supplementary Figure 2



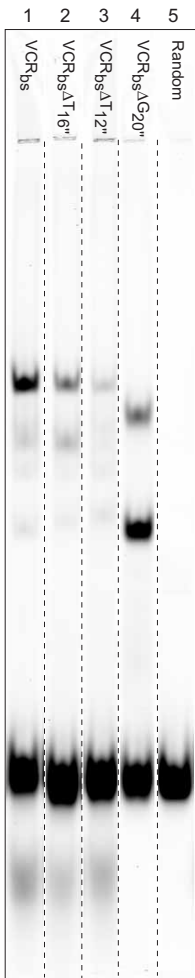
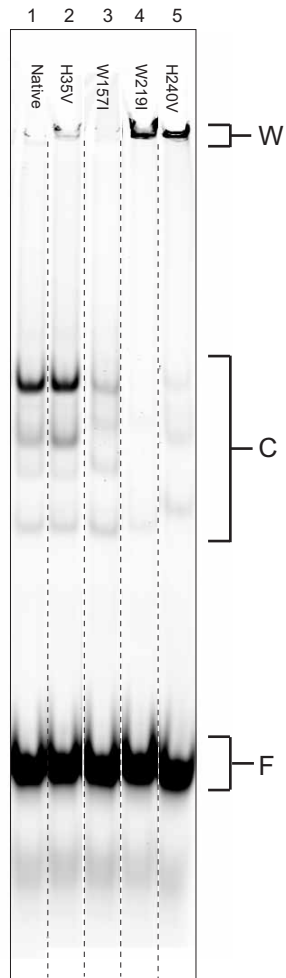
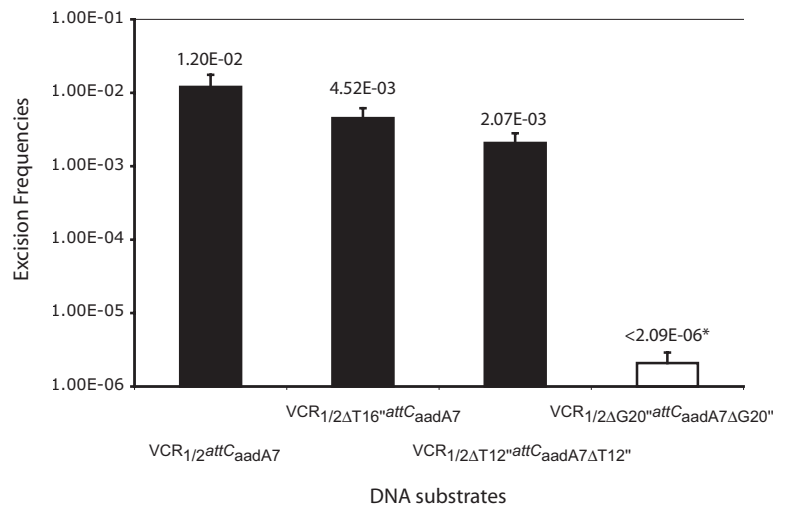
Supplementary Figure 3



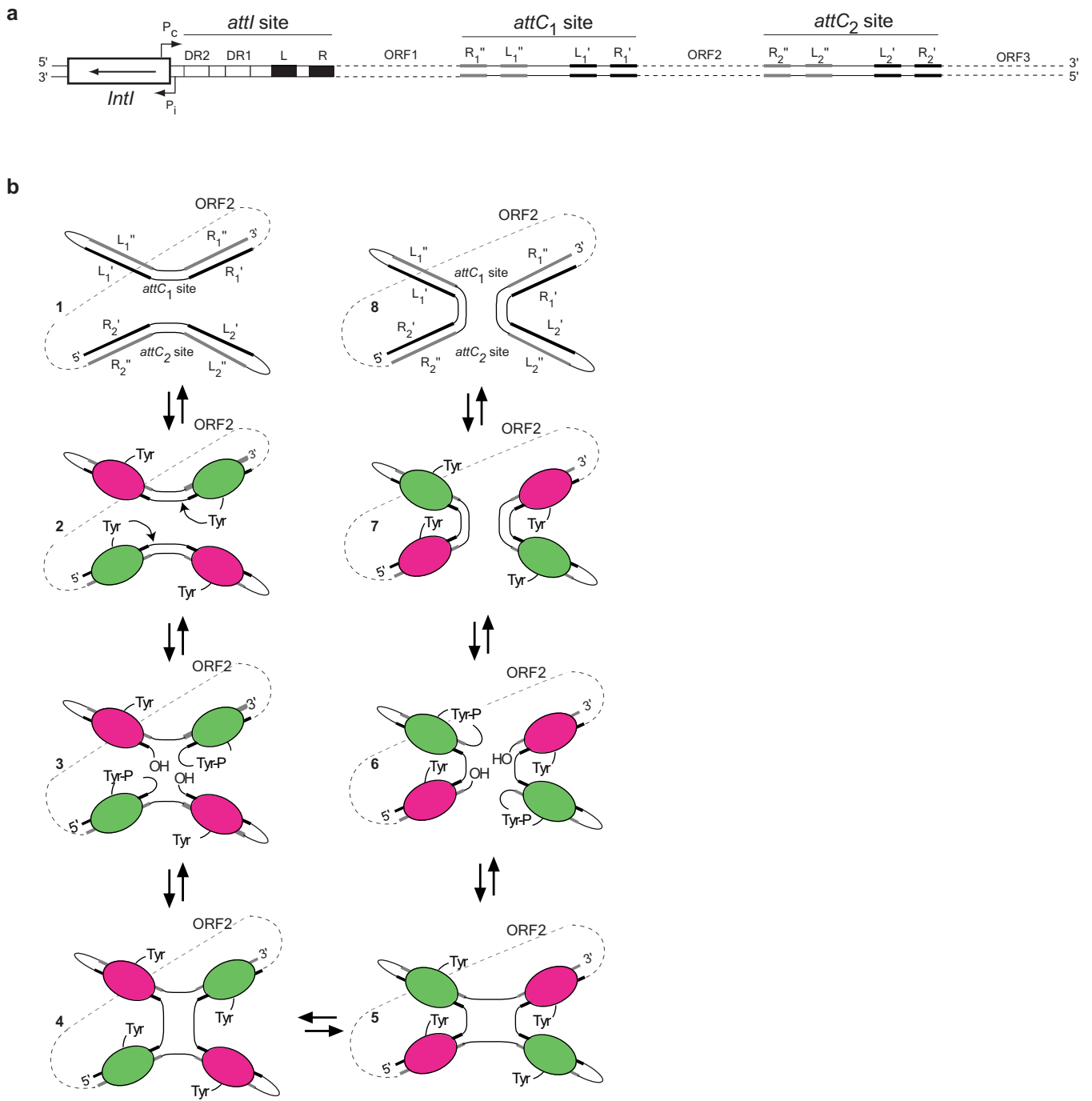
Supplementary Figure 4

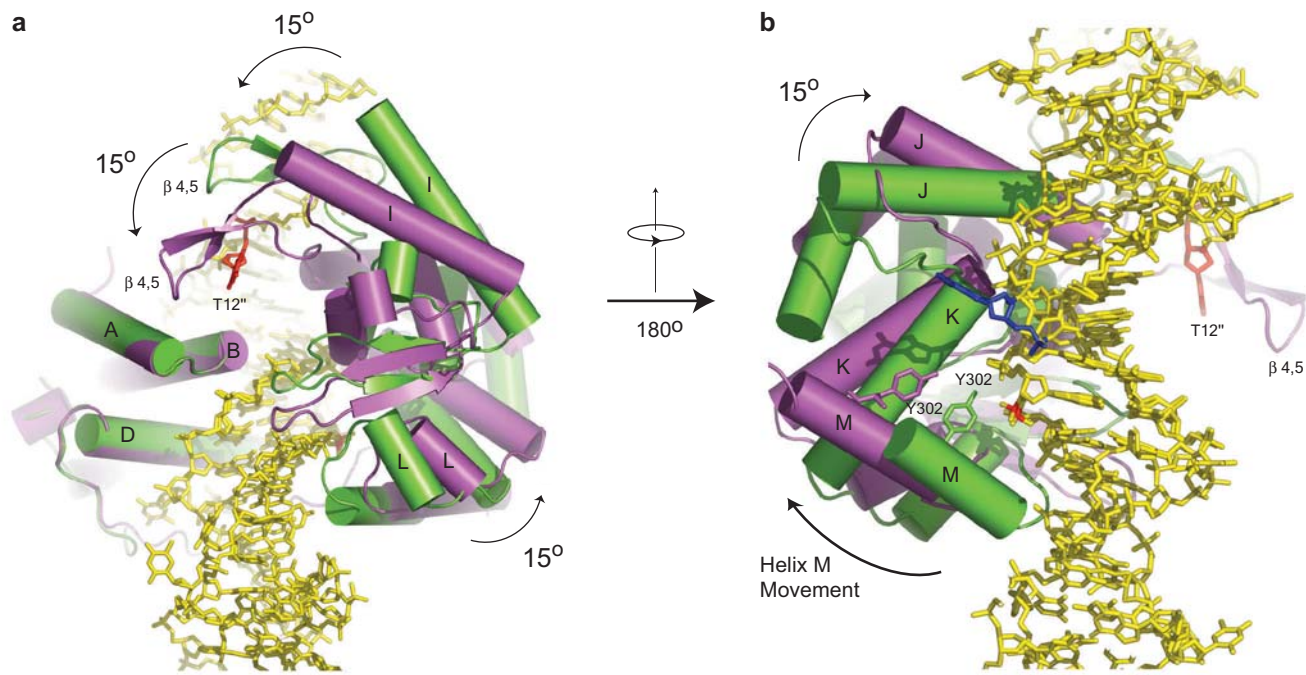


Supplementary Figure 5

a Mutant DNA EMSA**b Mutant Protein EMSA****c**

Supplementary Figure 6





Supplementary figure 8

Supplementary Table S1 Data collection and refinement statistics

	VchIntIA-VCR _{bs}
Data collection	
Space group	C222 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	149.9, 170.2, 209.4
α , β , γ (°)	90, 90, 90
Resolution (Å)	44.6-2.8 (2.97-2.80) *
<i>R</i> _{merge}	0.089 (0.38)
<i>I</i> / σ <i>I</i>	14.2 (3.3)
Completeness (%)	97.4 (91.4)
Redundancy	3.8 (3.8)
Refinement	
Resolution (Å)	44.6-2.8
No. reflections	63,682
<i>R</i> _{work} / <i>R</i> _{free}	0.234/0.262
No. atoms	13,236
Protein	10,498
DNA	2,568
Water	170
Average B-factors (Å ²)	
Protein	61.8
DNA	60.8
R.m.s deviations	
Bond lengths (Å)	0.009
Bond angles (°)	1.39

*Highest resolution shell is shown in parenthesis.