Published online 20 September 2007 *Nucleic Acids Research, 2007, Vol. 35, No. 19* ***6475–6489***
doi:10.1093/nar/gkm709

Identification of key structural determinants of the IntI1 integron integrase that influence *attC* × *attI1* recombination efficiency

Gaëlle Demarre[1], Clara Frumerie[1], Deshmukh N. Gopaul[2] and Didier Mazel[1,*]

[1]Unité Plasticité du Génome Bactérien, CNRS URA 2171 and [2]Laboratoire de Biochimie et Biophysique des Macromolécules, CNRS URA 2185, Institut Pasteur, 25 rue du Dr Roux, 75724, Paris 75724, France

Received May 25, 2007; Revised and Accepted August 27, 2007

ABSTRACT

The integron platform codes for an integrase (IntI) from the tyrosine family of recombinases that mediates recombination between a proximal double-strand recombination site, *attI* and a single-strand target recombination site, *attC*. The *attI* site is only recognized by its cognate integrase, while the various tested *attCs* sites are recombined by several different IntI integrases. We have developed a genetic system to enrich and select mutants of IntI1 that provide a higher yield of recombination in order to identify key protein structural elements important for *attC* × *attI1* recombination. We isolated mutants with higher activity on wild type and mutant *attC* sites. Interestingly, three out of four characterized IntI1 mutants selected on different substrates are mutants of the conserved aspartic acid in position 161. The IntI1 model we made based on the VchIntIA 3D structure suggests that substitution at this position, which plays a central role in multimer assembly, can increase or decrease the stability of the complex and accordingly influence the rate of *attI* × *attC* recombination versus *attC* × *attC*. These results suggest that there is a balance between the specificity of the protein and the protein/protein interactions in the recombination synapse.

INTRODUCTION

Integrons are the primary system for antibiotic-resistance gene capture and dissemination among Gram-negative bacteria. They are natural cloning and expression devices that incorporate open reading frames (ORF) and convert them to functional genes (1). The integron platform codes for an integrase (IntI) that catalyses site-specific recombination between a proximal primary recombination site, *attI* and a secondary target, called an *attC* site. The *attC* site is normally found associated with a promoterless single ORF, and the ORF-*attC* structure is termed a gene cassette. Upon recombination between *attI* and *attC*, the gene gains a promoter and is expressed. IntI belongs to the family of tyrosine recombinases (Y-recombinases), of which phage λ integrase (Intλ) is the paradigm. Their C-terminal catalytic domain contains the conserved amino acids Arg_{146}–Lys_{171}–His_{277}–Arg_{280}–His/Trp_{303} and the nucleophilic tyrosine, Tyr_{312} (IntI1 nomenclature). However, an additional insertion of ca. 25 amino acids, carrying an additional α-helix [I2 in (2)], conserved within the IntI (3–5) is also present.

The site-specific recombination mechanism involves a tetramer of proteins of which only two subunits are active per cleavage round. Their catalytic tyrosine forms a reversible protein-DNA intermediate that is either resolved back to substrate when the free 5′-hydroxyl attacks in *cis*, or if its the invading DNA strand, a Holliday junction (HJ) is formed. For most of the Y-recombinases, the locus of these reactions is a simple site composed of a pair of 9- to 13-bp inverted binding sites, separated by a 6- to 8-bp overlap region.

The integron recombination sites, *attI* and *attC*, do not preserve this structure. The *attC* sites are composed of two simple sites at the boundaries of an imperfect palindromic region, whilst the *attI* can vary in sequence and is devoid of a palindromic sequence. We have demonstrated that IntI1 recombines the single bottom strand of the *attC* sites (bs *attC*), folded in an imperfect hairpin structure with extrahelical bases, and the double-stranded (ds) *attI1* substrate (2,6). Cassette insertion at the *attI1* by IntI1 results in a HJ intermediate. However, because of the single-stranded (ss) structure of the *attC* site, the HJ cannot be resolved by a second-strand exchange, which would result in a linearization of the replicon carrying the integron. We have proposed that a replication

*To whom correspondence should be addressed. Tel: +33 1 40 61 32 84; Fax: +33 1 45 68 88 34; Email: mazel@pasteur.fr
Present address:
Gaëlle Demarre, Laboratory of Biochemistry and Molecular Biology, NCI, 37/6044, NIH, Bethesda, MD 20892-4260, USA

step could resolve the HJ, which thus regenerates the initial *attI* site and integrates the gene cassette (6).

The 3D structure of the *Vibrio cholerae* integron integrase, VchIntIA, in complex with a bs *attC* DNA substrate (2) showed some dissimilarity with Cre. Several secondary structural elements have evolved to provide essential interactions (2) between particular amino acids and the folded bs *attC* extrahelical bases.

IntIs are strictly specific for their cognate *attI* sites but are able to recombine *attC* sites of very different sequences and structures with comparable efficiency, as illustrated by VchIntIA (7), IntI2*179E (8)) and IntI3 (9). IntI1 efficiently recombines *attC* sites with the *attI1* site but not with the *attIVch* (7), *attI2* (8) or *attI3* (9) sites. Several studies show that IntI1 (7,10–12), IntI2 (8) and VchIntIA (7) are all able to recombine various *attC* sites. This is still puzzling, as *in vivo* and *in vitro* data do not completely overlap. Indeed, while IntI1 is able to recombine single-strand DNA folded in an imperfect hairpin structure and double-strand DNA with high efficiency *in vivo*, its *in vitro* affinity toward ds *attI1* is low compared to its affinity for bs *attC* (6,13).

The tolerance of IntI for *attC* sites is remarkable compared to other Y-recombinases. Structural analysis of the VchIntIA complex suggests that the tolerance relies in part on recognition of the extrahelical bases, which directs binding to the bs *attC* site and high order complex assembly (2). However, the crystal structure does not give a dynamic view of the recombination reaction and structural data for *attI* site recognition are still very loose (14,15). Here, we present a study performed with the aim of identifying key structural elements in IntI, more specifically in the IntI1 paradigm, involved in the *attC* × *attI* recombination reaction efficiency.

We chose to study the IntI specificity and efficiency in the context of the *attC* × *attI* integrative reaction, by mutagenesis of the integrase and selection of the mutants showing a higher recombination activity, a strategy successfully used to study the flipase (FLP) specificity on its FRT sites (16). Hence, we performed an extensive mutagenesis of *intI1* by mutagenic PCR and generated a 10^5 independent clone library, with the aim of selecting for the more active mutants. In order to facilitate the structural analyses, we choose to perform all selections using the *attC* site from the *V. cholerae* superintegron cassettes, named VCR (for *V. cholerae* repeats), or its mutated derivatives, as the structural data previously obtained corresponded to complexes of VchIntIA with this specific *attC* site. We realized three independent enrichment assays. The first was made using a wild-type VCR. The other two assays were done with two different VCR mutants, VCR_{INV} and VCR_{GTT}. VCR_{INV} is a VCR derivative in which the unpaired central spacer sequences are inverted, since structural data suggested that this region could direct the bending of the bs *attC* tertiary structure, and as such play an important role for recombination. In VCR_{GTT}, the three extrahelical bases, G, T and T [G20″, T16″ and T12″ respectively, (2)], which are specifically recognized by the integrase, are complemented. IntI1 and VchIntIA show 65% similarity and in order to interpret the different properties shown by these mutants, we further modeled the IntI1 3D structure based on the VchIntA crystal structure (2).

MATERIALS AND METHODS

Bacterial strains, plasmids and media

Bacterial strains and plasmids are described in Tables 1 and 2. *Escherichia coli* strains were grown in Luria Bertani broth (LB) at 37°C. Antibiotics were used at the following concentrations: ampicillin (Ap), 100 μg/ml, chloramphenicol (Cm), 25 μg/ml, kanamycin (Km), 25 μg/ml. Diaminopimelic acid (DAP) was supplemented when necessary to a final concentration of 0.3 mM. IPTG (isopropyl-beta- D-thiogalactopyranoside) and L-arabinose were added at, respectively, 0.8 mM and 2 mg/ml final concentration. Chemicals were obtained from Sigma-Aldrich (France).

Polymerase chain reaction (PCR) procedures

PCR for plasmid assembly used the *Pfu* DNA polymerase (Promega). For PCR reactions followed by topo cloning GoTaq polymerase (Promega) was used. Other PCR reactions used the PCR Reddy mix (Abgene, UK). All were used according to the manufacturer's instructions. PCR primers listed in Table 3 were obtained from Proligo (France).

Construction of pSU38::*lacZ*:: $attC_{aadA7}$-*catT4*-VCR_2

Each independent element ($attC_{aadA7}$, *catT4* and VCR_2) was assembled in pUC19 (Table 2). The primers aada7-3 and aada7-4 (Table 3) were phosphorylated by T4 polynucleotide kinase, annealed and cloned in pUC19 digested by EcoRI and BamHI. The *catT4* gene was amplified using primers cat-del-1 and cat-del-2 (Table 3) from pSU18 (17), digested by BamHI and BglII and cloned in pUC19::$attC_{aadA7}$ digested by the same enzymes. Finally, the synthetic cassette was assembled by cloning

Table 1. Bacterial strains used in this study

E.coli strains	Description/Relevant characteristics	Reference
DH5α	(F⁻) *supE44 ΔlacU169 (Φ80lacZΔM15)ΔargF hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Laboratory collection
TG1 (F⁻)	(F⁻) *supE hsdΔ5 thi Δ(lac-proAB)*	Laboratory collection
β2163	(F⁻) RP4-2-Tc::Mu *ΔdapA ::(erm-pir)* [Km ^{R} Em ^{R}]	(29)
KS272	F⁻ ΔlacX74, galE, galK, thi rpsL (StrA) ΔphoA (PvuII)	Ghigo J.M. unpublished data
B462	DH2 lacIQ, thi1, relA1, supE44, endA1, recA1, hsdR17, gyrA462, zei298::Tn10	(20)
BL21(DE3)pLysS	*E.coli* B strain containing the T7 polymerase and plasmid pLysS expressing T7 lysozyme	(30)
ω198	B462 (pSU38ccdB-attI1)	This study

the BglII/HindIII fragment from p453 containing VCR_2 (7) in the pUC19::$attC_{aadA7}$-*catT4*, digested by the same enzymes, leading to the p1187 (pUC19::$attC_{aadA7}$-*catT4*-VCR_2).

Table 2. Plasmids used in this study

Plasmids	Description	Reference
pUC19	ori_{ColE1} $[Ap]^R$	(31)
pSU18	ori_{p15A} $[Cm]^R$	(17)
pSU38	ori_{p15A} $[Km]^R$	(17)
pRS551	ori_{ColE1} $[Ap]^R$	(18)
pBAD18	ori_{ColE1} $[Ap]^R$	(32)
p112	pTRC99A::intI1, ori_{ColE1} $[Ap]^R$	(33)
p453	ori_{p15A} $[Cm]^R$	(7)
p1187	pUC19::$attC_{aadA7}$- *catT4*-VCR_2	This study
p1266	pSU38△, ori_{p15A} $[Km]^R$	This study
p1370	pSU38△::*lacZ*, ori_{p15A} $[Km]^R$	This study
p1372	pSU38△::*lacZ*::$attC_{aadA7}$-*catT4t*-VCR_2, ori_{p15A} $[Km]^R$	This study
p755	pSU38△::*attI1*, ori_{p15A} $[Km]^R$	(7)
p1880	pSW23T::$VCR_{2/1}$B, $ori_{R6K\gamma}$ $[Cm]^R$	(7)
p2677	pSU38-ccdB-*attI1*, ori_{p15A} $[Km]^R$	This study
p734	pBAD18::*intI1*, ori_{ColE1} $[Ap]^R$	This study
p4067	pBAD18::*intI1_P109L*, ori_{ColE1} $[Ap]^R$	This study
P4130	pBAD18::*intI1_D161G*, ori_{ColE1} $[Ap]^R$	This study
p4634	pET3a::*HisInt* $[Ap]^R$	This study
p4635	pET3a::$HisInt_{D161G}$ $[Ap]^R$	This study
p4636	pET3a::$HisInt_{P109L}$ $[Ap]^R$	This study

The *lacZ*α gene of pSU38 was replaced by the entire *lacZ* gene. First, *lacZ*α was deleted by inverse PCR. pSU38 was amplified with primers PSDMF and PSDH3 (Table 3), digested by MfeI and religated, leading to p1266. *lacZ* was amplified from pRS551 [Table 2, (18)], with primers LZFT-1 and LZFT-2 (Table 3), digested by EcoRI and HindIII and cloned in p1266, digested with the same enzymes, leading to p1370 (Table 2).

Finally, the EcoRI/SphI fragment from p1187, containing the synthetic cassette, was cloned in p1370 digested by EcoRI and SphI, generating pSU38△::lacZ::$attC_{aadA7}$-*catT4*-VCR_2 (p1372).

Construction of pETHis-*intI1*, pETHis-$intI1_{P109L}$ and pETHis-$intI1_{D161G}$

The IntI1 genes from plasmid p734, p4067 and p4130 (Table 2), were amplified with GoTaq polymerase (Promega) using primers HisIntI1 (Table 3) which contains a NdeI site and a tag coding for 10 histidines downstream of the ATG start codon of the *intI1* gene, and BamH1Int-rev (Table 3). The different amplicons were cloned into pTopo 2.1 (Invitrogen). Correct clones were selected after sequencing, and the various *intI* genes were transferred into vector pET3a using the NdeI and BamHI unique sites.

Table 3. Oligonucleotides

Oligonucleotides	Sequences
EMSA experiments	
VCRWT	GGATCCGGTTATAACGCCCGCCTAAGGGGCTGACAACGCGAAGCGTTGAC AGTCCCTCTTGAGGCGTTTGTTATAACAGATCTGAATTC
VCRDG	GGATCCGGTTATAACGCCCGCCTAAGGGGCTGACAACGCGAAGCGTTGAC AGTCCCTCTTAGGCGTTTGTTATAACAGATCTGAATTC
VCRG	GGATCCGGTTATAACGCCCGCCTCAAGGGGCTGACAACGCGAAGCGTTGA CAGTCCCTCTTGAGGCGTTTGTTATAACAGATCTGAATTC
VCRDGTT	GGATCCGGTTATAACGCCCGCCTAAGGGGCTGACAACGCGAAGCGTTGAC AGCCCCTTAGGCGTTTGTTATAACAGATCTGAATTC
VCRGTT	GGATCCGGTTATAACGCCCGCCTCAAGAGGGACTGACAACGCGAAGCGTTGAC AGTCCCTCTTGAGGCGTTTGTTATAACAGATCTGAATTC
VCRCOMP	GGATCCGGTTATAACAAACGCCTAAGGGGCTGACAACGCGAAGCGTTGAC AGTCCCTCTTGAGGCGTTTGTTATAACAGATCTGAATTC
VCRINV	GGATCCGGTTATAACTTTCGCCTAAGGGGCTGACAACGCGAAGCGTTGAC AGTCCCTCTTGAGGCGGCCGTTATAACAGATCTGAATTC
Plasmids construction	
aada7-3	AATTCGGTTATAACAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCTT AATTCAAGCGTTAGATGG
aada7-4	GATCCCATCTAACGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCGGCTT GAATGAATTGTTATAACCG
cat-del-1	CGCGGATCCACTGCCTTAAAAAAATTACGCCCCGCCCTGCC
cat-del-2	GGAAGATCTAAAAAAAAGCCCGCTCATTAGGCGGGCTCGAATAAATACCT GTGACGG
PSDMF	GGCTCAATTGGCATGCCTGCAGGTCGACTCTA
PSDH3	GGCTCAATTGAAGCTTGTCCGGCGGTGCTTTTGCCGTTA
LZFT-1	GTGAATTCCCGCGCACTGGCCGTCGTTTTACAACG
LZFT-2	GCGGATCCAAGCTTACCGAACATCCAAAAGTTTGTGT
HisIntI1	GCCATATGCATCATCATCATCATCATCATCATCATCATAAAACCGCCACTG CGCCG
BamH1Int-rev	GCGGATCCCTACCTCTCACTAGTGAGGGG
Reverse	AGCGGATAACAATTTCACACAGGA
Forward	CGCCAGGGTTTTCCCAGTCAC

Extensive random mutagenesis of IntI1

The *intI1* mutants library was generated by mutagenic PCR, using the 'gene morph PCR mutagenesis' Kit from Stratagene. As the mutation frequency directly relies on the DNA matrix quantity, we used three different conditions, in order to theoretically generate 1–10 mutations per kilobase, according to the manufacturer's instructions. *intI1* was amplified from p112 (Table 2) with primers reverse and forward (Table 3), digested by EcoRI and HindIII and cloned at the same sites of pBAD18, leading to pBAD18::*intI1**. The library products of the three reaction conditions were then introduced in TG1(F^-) cells by electroporation. This gave a total of 10^5 independent clones.

Integration assay

The integration assay derives from the *in vivo* recombination assay we previously developed (6). It uses conjugation to deliver one of the recombination substrates into a recipient cell expressing IntI1 under the control of the P_{BAD} promoter in the pBAD18 vector (Table 2) and carrying the partner recombination substrate on a pSU38 plasmid derivative. The recombination sites provided by conjugation were carried on pSW suicide vectors (Table 2). Briefly, the donor strain β2163 [*dapA*−, *pir*+] carries an RP4(IncPα) conjugation system integrated into its chromosome and can sustain pSW replication through the expression of a chromosomally integrated *pir* gene. The recipient is ω198 [*dapA*+, *pir*−] a B462 derivative which contains the pSU38-*ccdB-attI1* [Km^R] (Table 2) and either pBAD18::$intI1_{WT}$ or the mutant population of pBAD18::*intI1** (19). As pSW replication absolutely requires the Π protein, the number of recipients expressing the pSW marker directly reflects the frequency of cointegrate formation between the conjugated pSW plasmid and the target replicon. Conjugation was performed as previously described (6). The integration activity using this assay was calculated as the ratio of transconjugants expressing the pSW marker, Cm resistant colonies, over the total number of recipient Ap and Km resistant colonies.

Selection assay by enrichment

This assay is based on the repetition of cycles of selection for intI1* mutants active on a given substrate, in a competitive set up. The selection is exerted on their overall propensity to perform the *attC* × *attI* integrative activity in the recipient strain. One cycle consists of four steps (Figure 1), in which all mutants are subjected to selection all together. The two first steps consist of the integration assay after suicide conjugation in ω198, as described

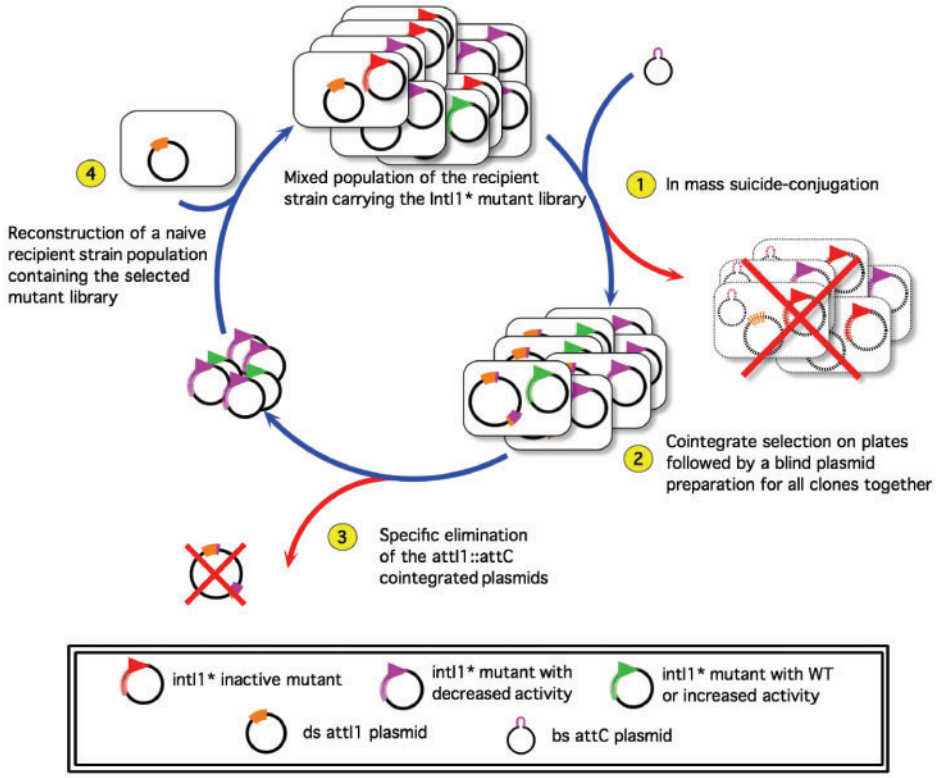

Figure 1. Schematic representation of the enrichment cycle of the conjugation–recombination assay. Briefly, the intI1 mutant population (IntI1*) is carried in a recipient strain also carrying a *attI1* plasmid (pSU38△::*attI1*) and subjected to an *en masse* suicide conjugation with a suicide plasmid injecting the *attC* bs (ppSW23T::$Vcr_{2/1}B$) or its mutated derivatives. Inactive intI1 mutants are symbolized by red dotted arrows, intI1 mutants showing a decreased activity are symbolized by purple dotted arrows, while intI1 WT or mutants with WT or increased activity are symbolized by green dotted arrows, as depicted in the cartouche. The four different steps of the cycles are described in the Materials and Methods section, they will result in an enrichment of the intI1* population in mutants showing and increased activity, compared to the WT protein, on a given substrate.

above, followed by the selection of cointegrates. IntI1 mutant genes are expressed under the control of AraC from pBAD18::*intI1**. In ω198, the *attI1* site is carried by a pSU38 derivative, which constitutively expresses the F toxin CcdB, pSU38-*ccdB-attI1*. This strain also carries the gyrA462 mutation which confers resistance to CcdB (20).

In a third step, all the selected cointegrates carrying recipient cells are gathered, and the whole plasmid population extracted. The selected pBAD18::*intI1** population is then purified from the cointegrates produced by recombination between incoming pSW::*attC* and the pSU38-*ccdB-attI1* by both a digestion with BglII and KpnI, which only cleave in the pSW and the pSU38, and transformation in DH5α, which is naturally sensitive to CcdB. This double process is essential to guarantee the elimination of all cointegrates formed after recombination and that the (19) transformants obtained contain only the pBAB18::*intI1** as if a pSU38-*ccdB* derivative survived the digestion step, it would kill the transformed DH5α. The transformants are once again gathered, subjected to a plasmid maxipreparation, which is used to transform the CcdB resistant ω198, generating a new recipient cell population (step 4 in Figure 1), that is used for the reiteration of step 1. The evolution of the library was followed by sequencing of the pBAD18::*intI1** in 50 randomly chosen clones after every other cycle.

Deletion assay

The deletion assay is based on the integrase ability to delete a synthetic cassette, *catT4-attC*$_{VCR2}$, by mediating recombination between the attC$_{aadA7}$ and *attC*$_{VCR2}$ sites. The synthetic cassette disrupts a copy of the *E. coli lacZ* gene carried on a pSU38 (pSU38△::lacZ::*attC*$_{aadA7}$-*catT4-attC*$_{VCR2}$) and prevents the lacZ productive translation by introducing a stop codon in the reading frame, 111 bp downstream of the ATG codon. IntI mediated deletion of the cassette restores a single reading frame allowing functional β-galactosidase synthesis. IntI1 WT and mutants were expressed from the pBAD18::*intI1* plasmid used in the enrichment assay. Recombinants were screened on S-Gal/LB agar blend from Sigma. The deletion frequency was measured by determining the ratio of [Lac$^+$] colonies (black) to [Lac$^-$] colonies (white). The assay was performed in strain KS272, an *E. coli* deleted of the *lac* operon (Table 1).

Protein purification

Bacterial strain BL21(DE3)plysS was transformed with the pET-derived plasmids expressing His-IntI1$_{WT}$, His-IntI1$_{P109L}$ and His-IntI1$_{D161G}$, respectively. The protein purification was carried out as described by Johansson and collaborators (21), with the following minor modifications. The colonies from each plate were harvested and grown overnight in 25 ml of LB containing Ap and Cm. After overnight incubation the cells were harvested by centrifugation and resuspended in 25 ml of fresh media and inoculated in 2500 ml LB supplemented with the appropriate antibiotics. At log phase the cells were put on ice in a 4°C cold room for ~1 h for cooling, then induced with 0.8 mM IPTG and incubated overnight at 16°C. The cells were harvested by centrifugation and resuspended in 1 × HisTrap buffer (GE Healthcare) supplemented by protease inhibitor cocktail tablets, 1 tablet/50 ml, (Roche), imidazole to a final concentration of 10 mM (GE Healthcare) and 1/10 the volume of 1% Triton X-100 (SIGMA). Cells were then disrupted by sonication and the supernatant collected after centrifugation with a Sorvall centrifuge (rotor SS-34) for 30 min at 12 000 r.p.m. The purification was carried out in the cold room, over 1 ml HisTrap FF crude columns according to manufacturer's recommendations (GE Healthcare) using 200 mM imidazole in the 1× HisTrap wash buffer. The proteins were eluted with 1× HisTrap buffer containing 800 mM imidazole. The proteins were estimated to ~80% purity after electrophoresis and the concentration determined using the Bradford assay (22) with BSA as a standard.

Electromobility shift assay

EMSA with ss DNA. The purified proteins were incubated with DNA oligonucleotides corresponding to the bottom strand of the *attC*$_{VCR2}$ site and of the various mutant sites (Table 3). In the synthetic oligonucleotides, the sequence corresponding to the central loop of the hairpin in the *attC*$_{VCR2}$ site, shown not to influence the recombination frequency by the wt IntI1 (Bouvier *et al*., to be published), was removed. Each reaction contained 50 ng of sonicated salmon sperm DNA, 12 mM HEPES–NaOH pH 7.7, 12% glycerol, 4 mM Tris–HCl pH 8.0, 60 mM KCl, 1 mM EDTA, 0.06 μg/μl BSA, 1 mM DTT, 200 mM imidazole, 5 mM phosphate, 125 mM NaCl, 0.6 pmol P^{32} labeled probe and 1.5 μM protein, in a final volume of 20 μl. For each probe a negative control reaction without protein was also performed. Samples were incubated at 30°C for 20 min before loading onto a precooled 5% native polyacrylamide gel (30% Acrylamide/Bis Solution, 29 : 1 Bio-Rad), with 1× TBE as buffer. Gels were run at 20 mA with water as cooling until the bromo-phenol blue (BPB) had reached ~5 cm. Identical batch of labeled probes were used in the experiments shown and gels were run on the same day. The corresponding gels in Figures S3–S5 were performed as above with 100 ng of poly dI-dC as unspecific competitor and ran until the BPB had reached ~11 cm in the gels. The amount of DNA used and the protein concentrations are indicated in the figure legend.

EMSA with attI1. *attI1* was amplified by PCR from plasmid pSU38::*attI1* using primers forward and reverse (Table 3), generating a 300 bp fragment. Each reaction contained 50 ng poly dI-dC, 12 mM HEPES–NaOH pH 7.7, 12% glycerol, 4 mM Tris–HCl pH 8.0, 60 mM KCl, 1 mM EDTA, 0.06 μg/μl BSA, 1 mM DTT, 200 mM imidazole, 5 mM phosphate, 125 mM NaCl, 0.001 pmol P^{32} labeled probe and the amount of protein indicated in the Figure 5 legend, in a final volume of 20 μl. Gels were run at 20 mA with water as cooling and 0.5× TBE as running buffer, for ~4 h.

The gels were dried prior to autoradiography and visualized using a phosphor imager 445 SI (Molecular dynamics).

Quantifications were made using the Image Gauge version 4.0.0 (FUJI Photo Film Co. Ltd.), manually defining the lanes and using the automatic peak search function. All other functions were at default values.

RESULTS

Development of an enrichment selection assay

IntI1 has a relaxed specificity, in the sense that it is able to recombine sites that have not only remote sequence similarities but also huge structural differences—the *attI1* site seems to be recognized under canonical ds form whilst *attC* sites are recognized as ss secondary structures. Moreover, the activity of IntI1 is low compared to other Y-recombinases (6). In order to determine if a correlation between specificity and activity could exist, in particular, if a more active IntI1 protein would also be more specific, we performed an extensive mutagenesis of the protein, generating a large library of mutants collectively designated as IntI1*, and developed an enrichment selection assay to find more active mutants of IntI1. In this assay, the mutants are selected on their recombination activity for the *attC* integration at *attI1*. The enrichment is done by the repetition of cycles, including the selection step. In theory with such a strategy, the most active IntI1* are positively selected and concentrated after each cycle. The assay is described in detail in materials and methods and in Figure 1.

IntI1 mutant library

The IntI1* library was generated by error prone PCR. A library of 10^5 clones was created, using three different mutagenesis conditions to theoretically generate 1–7 mutations per kilobase. Sequencing of 50 clones showed that the library contained 88% mutants with 1–9 amino-acid substitutions, nucleotide (nt) deletions and/or insertions. In addition, mutations were found scattered over the length of the *intI1* gene (see Supplementary Figure S1).

Enrichment capacity

The enrichment capacity of the assay was tested by competition between wild-type IntI1 ($IntI1_{WT}$) and several IntI1 mutants with different relative activities. These IntI1* mutants were chosen among the 50 clones sequenced after the first mutagenesis round, described above. The R280 catalytic residue mutant, $IntI1_{R280S}$, was inactive; whilst mutants $IntI1_{L87V}$ and $IntI1_{R25H–R263H}$ were respectively 10- and 100-fold less active than $IntI1_{WT}$ (data not shown). Competitions were achieved by mixing 1, 10, 50 or 90% of recipient strain containing $IntI1_{WT}$ and respectively 99, 90, 50 or 10% of recipient strain containing the different IntI1*, at the conjugation step of the first cycle. After each enrichment cycle, $IntI1_{WT}$/IntI1* ratio was determined by PCR on the colonies obtained after the recombination selection. The discrimination between $IntI1_{WT}$ and IntI1* was permitted by the addition of a specific silent nucleotide sequence downstream of the $IntI1_{WT}$ gene in pBAD18::$intI1_{WT}$, allowing specific PCR amplification. One cycle of selection was sufficient to completely eliminate the inactive $IntI1_{R280S}$ mutant, irrespectively of the initial $IntI1_{WT}$/$IntI1_{R280S}$ proportions (Table S1). With the 100-fold less active $IntI1_{R25H–R263H}$ mutant, we observed a massive proportion decrease after the first cycle and complete disappearance after the second enrichment cycle (Table S1). In the case of the moderately affected $IntI1_{L87V}$ mutant (which retained 10% of the $IntI1_{WT}$ activity), the $IntI1_{WT}$ proportion increased more progressively (Table S1). In the control competition, when two $IntI1_{WT}$ were mixed in 50:50 ratio, this ratio was conserved for the two cycles (Table S1).

Selection for mutants with a higher recombination activity on a wild-type *attC* site, $attC_{VCR2}$

We pooled the 10^5 mutants and subjected them *en masse* to the enrichment strategy using $attC_{VCR2}$ (hereafter called VCR_{WT}). We followed the enrichment by sequencing 50 randomly chosen clones at different cycles. As expected from the elimination of both the inactive mutants and those showing a reduced activity in comparison to $IntI1_{WT}$, the mutant proportion drastically decreased from 88% to 21%, during the first cycles of the enrichment (Figure S2). Subsequently, the mutant proportion increased until four mutants, $IntI1_{D161E}$, $IntI1_{D161V}$, $IntI1_{A4V–G322E}$ and $IntI1_{L287M}$ became predominant. These mutants altogether represented 70% of the mutants in the library after 20 cycles (Figure S2).

Selection for mutants with a higher recombination activity on mutated $attC_{VCR2}$ sites

The *attC* bs are folded as imperfect hairpins with extrahelical structures. They all show an unpaired central segment between the R and L boxes, and two to three bases have no complement in the palindrome forming the hairpin (Figure 2A). These extrahelical elements were shown to be important for *attC* × *attC* recombination activity (2) and *in vitro* binding (2,21). The $IntI1_{WT}$ recombination activity between *attI1* and the VCR_{WT} mutants, VCR_{INV} and VCR_{GTT}, where these imperfections have been modified, is especially low. VCR_{INV} is a VCR derivative in which the sequences, which form the unpaired central spacer located between the sequences bound to the integrase monomers, are inverted (Figure 2B). Structural data suggested that this unpaired central spacer could direct the bending of the bs *attC* tertiary structure, and as such play an important role for recombination. In VCR_{GTT} the three extrahelical bases, G, T and T [G20″, T1″ and T12″, respectively, (2)] are complemented (Figure 2B). While $IntI1_{WT}$ activity on VCR_{WT} is $\sim 10^{-1}$, VCR_{INV} and VCR_{GTT} are recombined with activities of 1.10×10^{-4} and 3.76×10^{-5}, respectively (Figure 3). The IntI1 mutant library was used in two independent enrichment assays with these two VCR mutated sites. In each assay, we found that a single mutant represented more than 60% of the population sequenced after six cycles. These mutants were $IntI1_{P109L}$ in the library selected with VCR_{INV} and $IntI1_{D161G}$ in the library selected with VCR_{GTT}.

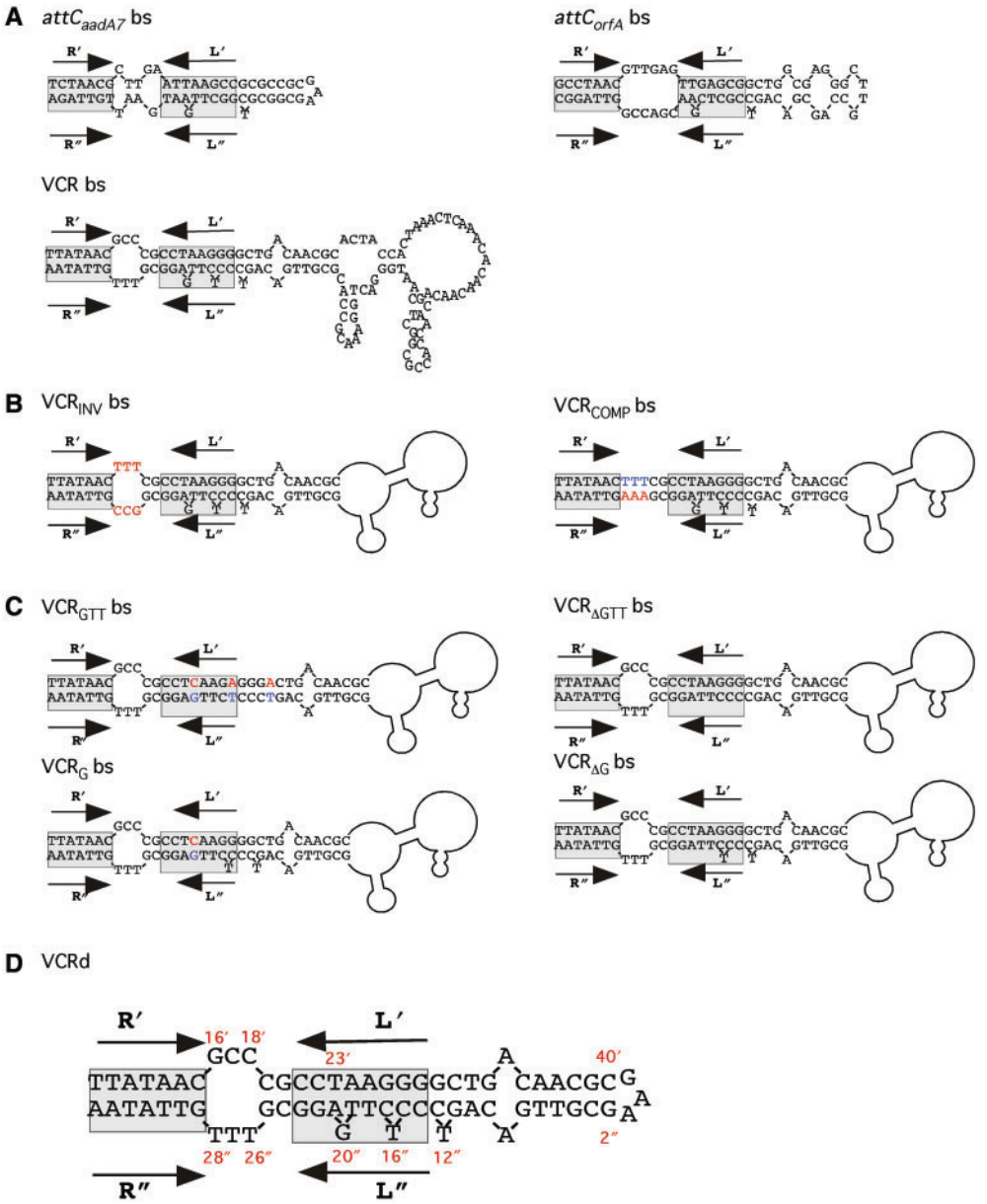

Figure 2. Proposed secondary structures of the *attC* sites bs studied in this work. Secondary structures were predicted using the MFOLD online interface at the Pasteur Institute. The inverted repeats L′ and L″, R′ and R″ are indicated with black arrows. The putative IntI1 binding domains, as defined by Bouvier *et al.* (2006), are marked with grey boxes. (**A**) Secondary structure of $attC_{aadA7}$, VCR and $attC_{ofrA}$ WT sites. (**B**) and (**C**) Secondary structure of VCR mutants. Mutations are shown in red and in blue when they involve complementary mutations. The internal central loop of the palindrome is shown by a black line mimicking the secondary structure predicted by MFOLD and shown in panel A. (**D**) VCRd, a VCR derivative used in EMSA experiments, in which the central and imperfect part of the palindrome was replaced by GAA, numbering of the different nt according to our published work (2) is given in red above or below the cognate nt.

In vivo integration and deletion activities of the different mutants

In vivo activities of mutants $IntI1_{D161E}$ and $IntI1_{D161V}$. As mutants at position 161 were also found in the selection on the VCR_{GTT} mutant site, we focused our investigations on the two D161 mutants, $IntI1_{D161E}$ and $IntI1_{D161V}$.Their activities were tested *in vivo*, by integration and deletion as described in materials and methods. To study the specificity of the mutants, we chose to test recombination using two different *attC* sites: VCR_{WT}, used for the enrichment, and $attC_{aadA7}$, which has a different secondary structure (Figure 2A). The deletion assay involves the same *attC* sites, VCR_{WT} and $attC_{aadA7}$. $IntI1_{D161E}$ and $IntI1_{D161V}$ were found to be more active than $IntI1_{WT}$ in the integration assay with both *attC* sites, and also in the deletion assay (Figure 3).

In vivo $IntI1_{P109L}$ activities. $IntI1_{P109L}$ recombines VCR_{INV} with an activity of 1.6×10^{-2}, 100-fold higher than $IntI1_{wt}$ (Figure 3). This mutant is also more active than $IntI1_{wt}$, on VCR_{WT}, $attC_{aadA7}$ and $attC_{orfA}$, in which

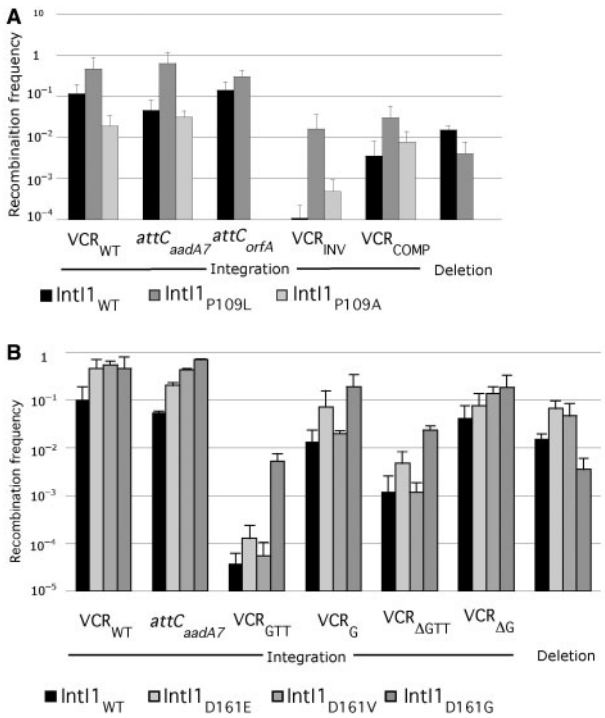

Figure 3. Integration and deletion activities of $IntI1_{WT}$ and the IntI1 mutants, on different recombination substrates. Integration assays are between different WT or mutant *attC* sites and the *attI1* site, while the deletion assays are between the *attCaadA7* and the $attC_{VCRWT}$ sites. Recombination frequencies (vertical axis of histogram) correspond to the average of three independent trials. Error bars show standard deviations. (**A**) Mutant $IntI1_{P109L}$. (**B**) Comparison of $IntI1_{P109L}$ and $IntI1_{P109A}$ mutants. (**C**) D161 Mutants.

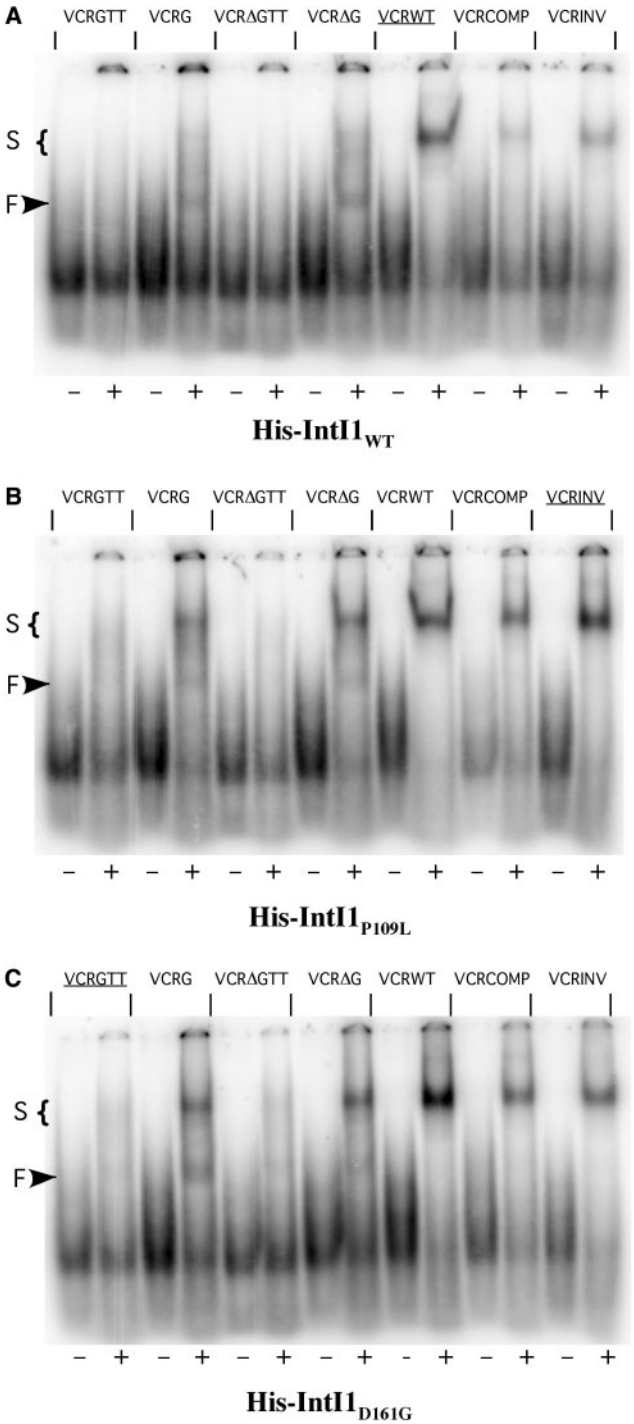

Figure 4. Gel retardation of bs VCRd derivatives by (**A**) His-$IntI1_{WT}$ (**B**) His-$IntI1_{P109L}$ and (**C**) His-$IntI1_{D161G}$. The substrate that the respective mutant was selected on is underlined. 0.6 pmol of each probe and 1500 nM of protein were used in each reaction. In the lanes marked with –, no protein was added. Faster (F) and slower (S) complexes are indicated by arrows.

the unpaired central segment between the R and L boxes has a different organization (Figure 2A). Moreover, when tested with VCR_{COMP}, a VCR mutant in which the unpaired central segment is complemented to eliminate the bulge (Figure 2B) as integration substrate, its activity was found to be 4-fold higher than $IntI1_{WT}$ (3.0×10^{-2} and 7.62×10^{-3}, respectively, Figure 3). Conversely, its deletion activity was found to be 6-fold lower than the WT protein (Figure 3).

Another mutant in position P109, $IntI1_{P109A}$, was isolated in the randomly chosen clones sequenced at different stages during the enrichment of the library on VCR_{WT}. The activity of this mutant is closer to $IntI1_{WT}$ activity on the tested substrates and inferior to the activity of $IntI1_{P109L}$ (Figure 3).

Finally, we looked for the ability of $IntI1_{P109L}$ to restore WT activity on the other VCR mutant substrate used in this study, VCR_{GTT}, where the mutations affect the extrahelical bases. Interestingly, we found that $IntI1_{P109L}$ had a very low activity, similar to that of $IntI1_{WT}$ on this substrate (1.6×10^{-5} and 3.7×10^{-5}, respectively.).

In vivo IntI1D$_{161G}$ activities. $IntI1_{D161G}$ recombines VCR_{GTT} with an activity of 4.75×10^{-3}, 500-fold higher than $IntI1_{wt}$ (Figure 3). The VCR_{wt} integration frequency catalysed by $IntI1_{D161G}$ is also higher than that of $IntI1_{wt}$ on the same substrate (Figure 4). Furthermore, we tested this mutant integration activity with different VCR mutants where: (i) only the extrahelical G is complemented, VCR_G, (ii) the three extrahelical bases are deleted, $VCR_{\triangle GTT}$ and (iii) only the extrahelical G is deleted,

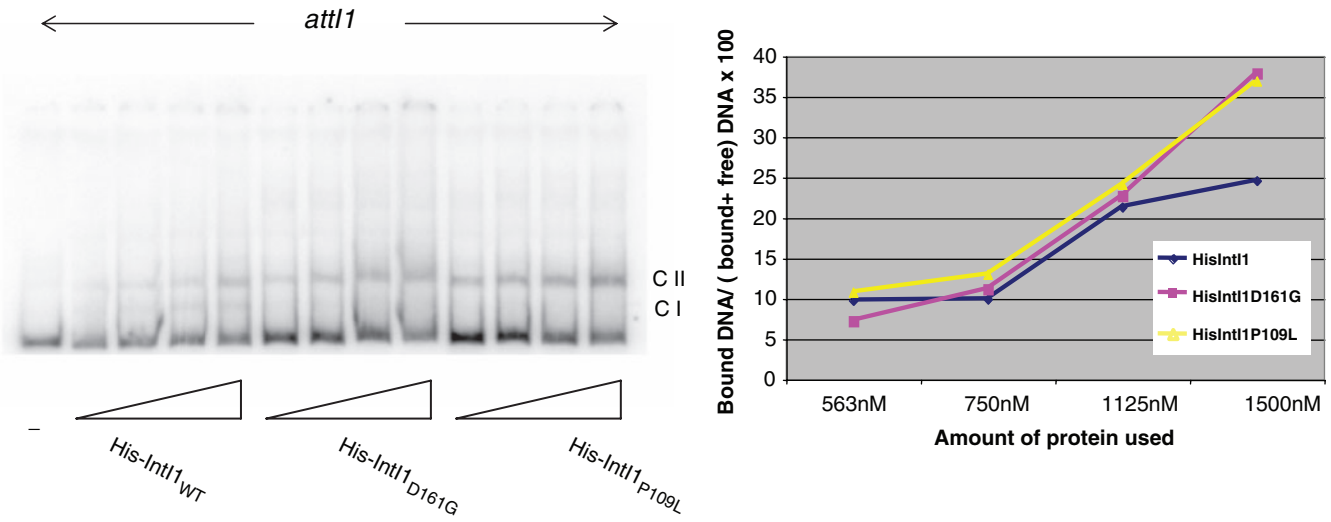

Figure 5. Gel retardation of ds *attI1* by His-IntI1$_{WT}$, His-IntI1$_{D161G}$ and His-IntI1$_{P109L}$ (**A**) and its quantification (**B**). (A) The amount of protein used in the EMSA were respectively of 563 nM, 750 nM, 1125 nM and 1500 nM, for each His-IntI1 derivative. 0.001 pmol of ds *attI1* probe was used in each reaction. (B) Fraction (×100) of the labeled DNA substrate bound by the His-IntI1 derivative in complexes I or II/labeled (bound + free) DNA, detected by automatic peak search of Image Gauge.

VCR$_{\triangle G}$ (Figure 2C). In all cases, IntI1$_{D161G}$ was found to be 10- to 100-fold more active than IntI1$_{WT}$ (Figure 3). The recombination of these mutant substrates was further tested with the D161 mutants selected on VCR$_{WT}$, IntI1$_{D161E}$ and IntI1$_{D161V}$. The recombination activity of these mutants on the mutated substrates was more similar to that of IntI1$_{wt}$ than to IntI1$_{D161G}$, although IntI1$_{D161E}$ showed a 2- and 5-fold higher recombination rate than IntI1$_{wt}$ on the VCR$_{GTT}$ and VCR$_{\triangle GTT}$ substrates, respectively (Figure 3).

Interestingly, as for IntI1$_{P109L}$, the deletion activity of IntI1$_{D161G}$ was found to be 10-fold lower than the WT protein (Figure 3). The IntI1$_{D161G}$ activity on VCR$_{INV}$ is as low as the IntI1$_{wt}$ activity on the same substrate (6.8×10^{-4} and 1.1×10^{-4}, respectively).

DNA Binding capacity of His-IntI1$_{wt}$, His-IntI1$_{P109L}$ and His-IntI1$_{D161G}$

One possible explanation for the higher integration activity of the mutants would be that they recognize the DNA substrates better than IntI1$_{wt}$. Hence, because of having a greater binding capacity for the sites, they are also able to mediate integration more efficiently. Therefore, the binding capacity of IntI1$_{wt}$, IntI1$_{P109L}$ and IntI1$_{D161G}$ was investigated using EMSA. IntI1$_{D161G}$ was chosen for purification over IntI1$_{D161E}$ and IntI1$_{D161V}$, since it was the mutant in position D161 that had the highest integration activity on all the assayed substrates. To perform the EMSA, IntI1$_{WT}$, IntI1$_{P109L}$ and IntI1$_{D161G}$ were produced in His-tagged versions and were purified by metal affinity chromatography on Ni-NTA.

EMSA with the VCR sites. The different VCRs used in the conjugation assays were tested in EMSA as well. Single-stranded oligonucleotides of ~90 bases, containing a shorter version of the VCR (VCRd) were used as substrate. In this substrate, the central part of the site that corresponds to the imperfect matches of the palindrome, namely the loop structure, was removed and substituted with GAA (Figure 2D). This modification has been shown to have no effect on the *in vivo* recombination frequency. The recombination capacity of the corresponding VCRd mutant substrates VCRd$_{INV}$ and VCRd$_{GTT}$, was confirmed, using the conjugation assay with the His-tagged IntI1$_{wt}$, IntI1$_{P109L}$ and IntI1$_{D161G}$ proteins (data not shown). At the same time, the His-tagged mutants were shown to be recombinogenic. The basal T7 promoter leak was sufficient to confer a recombination frequency of 1/100th of the frequency obtained in arabinose-induced conditions from the pBAD::*intI1*. Furthermore, the relative activity of the His derivatives were similar to those of the non His-tagged proteins (data not shown). This was expected as a large N-terminal extension such as the MBP-tag had been previously shown to leave the recombination activity of IntI1 unaffected (23).

His-IntI1$_{WT}$, His-IntI1$_{P109L}$ and His-IntI1$_{D161G}$ were assayed for binding with all the VCRd substrates. The results from the EMSA experiments can be seen in Figure 4 and in Figures S3–S5. In presence of the His-IntI1, distinct bands corresponding to different protein–DNA complexes can be seen. The faster migrating complex is always weak, if detectable. The slower migrating complex is a double band that can be separated with an increased migration time (Figure S3). The band intensities on the gels in the supplementary figures were quantified as described in materials and methods. The results were used as a complement in the interpretation of the gels (Figures S3–S5).

Interestingly, one can see with His-IntI1$_{WT}$ that there is a rough correlation between binding to a given substrate and *in vivo* recombination rates, i.e. those bound the most poorly, being those that recombined at the lowest rate,

as for example $VCRd_{GTT}$. There is however an exception with $VCRd_{INV}$, which if bound only slightly less than $VCRd_{WT}$ *in vitro*, is recombined with a 1000-fold lower frequency.

His-$IntI1_{P109L}$ and His-$IntI1_{D161G}$ both bind the $VCRd_{wt}$ with a better efficiency than His-$IntI1_{WT}$ (Figure 4 and Figure S3). All versions of the His-IntI1 proteins bind to the $VCRd_{G}$ and $VCRd_{\triangle G}$ substrates. Nevertheless, with these substrates, the slower migrating complex only weakly observed with the wt protein, is more prominent with the His-$IntI1_{P109L}$ and His-$IntI1_{D161G}$ mutants (Figure 4). In fact, the His-$IntI1_{wt}$ binding to the $VCRd_{G}$ and $VCRd_{\triangle G}$ substrates is too low to be detected by Image Gauge (data not shown). The binding to the two substrates that lack the extrahelical nt in the L box stem, $VCRd_{GTT}$ and $VCRd_{\triangle GTT}$ is when distinguishable very poor for all three proteins (Figure S5).

For the substrates $VCRd_{COMP}$ and $VCRd_{INV}$, both the mutant proteins recognize these substrates slightly better than His-$IntI1_{WT}$ (Figure 4 and Figure S4). All proteins bind $VCRd_{INV}$ slightly better than $VCRd_{COMP}$, where the largest difference is for His-$IntI1_{wt}$ (Figure 4). In summary, the EMSA experiments show that both the mutants bind slightly better than the WT protein to the substrates $VCRd_{wt}$, $VCRd_{INV}$, $VCRd_{COMP}$, $VCRd_{G}$ and $VCRd_{\triangle G}$; with, in addition, an increment in the slower complex formation for both mutants with the last two substrates.

EMSA with the attI1 site. The common DNA substrate in the different enrichment experiments was ds *attI1*. His-$IntI1_{WT}$, His-$IntI1_{D161G}$ and His-$IntI1_{P109L}$ were assayed for their binding to the *attI1* site (Figure 5). $IntI1_{WT}$ has been shown to have less affinity for ds *attI1* site than for bs *attC* sites (6). This was true also for the His-$IntI1_{WT}$ and the mutants. While the protein to DNA ratio in the experiments with the *attC* bs is 50:1, the protein to DNA ratio for the *attI1* shifts is almost 600 times higher. However, the EMSA experiments show that both mutant proteins bind *attI1* differently than His-$IntI1_{WT}$. Two distinct complexes can be seen with the WT protein, complex I and complex II (Figure 5). With the two mutant proteins, complex II is the most distinct (Figure 5). The amount of shifted DNA is slightly (up to 1.5×) higher for both mutants compared to His-$IntI1_{WT}$.

DISCUSSION

IntI1 is considered as the paradigm of the Y-recombinases subfamily of the integron integrases. The integron integrases have the rare ability to recombine ss and ds DNA, the bs *attC* and *attI*, respectively. It is extremely interesting to investigate what in its structure allows such different behaviors. Unfortunately, IntI1 has proven difficult to crystallize and its structural studies have been limited to a few directed mutagenesis studies (4,23). We recently reported the 3D structure of the *V. cholerae* superintegron integrase VchIntIA (2), which shares 45% identity and 65% similarity with IntI1 (24), and have been shown to have similar recombination properties, especially in terms of their *attC* substrate spectrum (7). The crystal structure of VchIntIA in complex with bs *attC* sites of the VCR type, which corresponds to the cassette deletion reaction, has provided several answers to the questions raised by the dual specificity of IntI, but many aspects of the mechanism are still elusive. In this study, we chose to look closer at the relationship between the structure and the properties of IntI1, by selecting after an extensive mutagenesis, mutants of the protein more active in the *attI1* × *attC* integration reaction on either WT or mutated *attC* substrates. In order to facilitate the analysis of the results, we choose to perform the selection using a VCR *attC* site, or its mutated derivatives. Indeed, as the available 3D structure corresponds to complex of VchIntIA with this specific *attC* site, this option should limit structural speculations to the parallel modeling of IntI1 structure based on the one of VchIntIA rather than to any difference in the structure of the *attC*.

The hypermutagenic PCR method we used generated a library of 10^5 independent clones, showing a broad spectrum of mutations and a large proportion of mutants. As it seemed unreasonable to analyse their recombination properties one by one, we decided to adopt an enrichment-by-cycling strategy and perform an *en masse* selection, analogous to the *in vitro* approaches (25,26).

We postulated that for plasmids expressing a protein variant, producing an increment in the efficiency of *attC* × *attI1* recombination would ultimately generate a larger number of progeny carrying a cointegrate in our suicide conjugation assay. Based on the initial competition trials where IntI1 variants with 10% or 1% residual activity were efficiently eliminated in competition with IntI1 WT (Table S1), we applied the cycling of selective steps protocol to the library. The principle of the enrichment assay is based on the controlled repetition of steps and on the advantageous possibility of changing or modifying each partner of the recombination reaction independently (IntI, *attI* or *attC*) (Figure 1). The system is highly effective since we observed that inactive mutants were eliminated after the first cycle, whilst others either decreased or augmented proportionally to the number of cycles, depending on their relative lower or higher recombination activity.

Although the library we constructed had in theory a large redundancy for a 337 amino acid protein, the low probability of multiple mutations in the same codon favored dispersed amino acid changes involving a single nucleotide switch. The analysis of 50 randomly chosen clones by sequencing, showed that our library contained many multiple mutants with 2–9 substitutions (Figure S1). With a size of 10^5 independent clones, our library covered a certain set of these possible multiple mutants, but still a minor fraction of all possible ones.

Initially, using WT *attI1* and a WT *attC* site of the VCR family, four mutants became predominant after 20 cycles. Two of them, $IntI1_{D161E}$ and $IntI1_{D161V}$, were shown to be slightly more efficient than $IntI1_{wt}$ in integration via *attC* × *attI1* recombination using two structurally different WT *attC* sites, VCR_{WT} and $attC_{aadA7}$.

They were also more efficient for deletion through *attC* × *attC* recombination.

Subsequently, two VCR mutant sites, VCR_{GTT} and VCR_{INV}, both altered in the extrahelical structures of the bs *attC* site (Figure 2B and C) were used. This resulted in the selection of $IntI1_{D161G}$ and $IntI1_{P109L}$, respectively. $IntI1_{P109L}$, was more active for integration than $IntI1_{WT}$ on the different WT *attC* substrates, but showed a decrease in deletion efficiency. An alanine at this position has no effect on the protein activity (Figure 3). Interestingly, the residue found at the homologous position to P109 is a leucine in VchIntIA (Figure 6B) and several other *Vibrio* IntIA (data not shown).

$IntI1_{D161G}$, is more active for integration than $IntI1_{WT}$ with all the WT and VCR_{GTT} related substrates, while its deletion activity through *attC* × *attC* recombination is about 10-fold lower than that of $IntI1_{wt}$. This substitution by a glycine residue brings the integration efficiency with the VCR (mutagenized in its extrahelical bases) to a level similar to that of $IntI1_{WT}$ on VCR_{WT}. Among the different amino acid substitutions identified at position 161, glycine is the only one having such a phenotype (Figure 3).

It is surprising that these mutants were not selected on VCR_{WT} as we started the selection cycling from the same library. During the 20 cycles of the enrichment assay on VCR_{WT}, we recloned the entire library into a new vector every third cycle in order to avoid mutations accumulation in the plasmid carrying *intI1** that could confer an indirect higher recombination frequency; i.e. mutations in the inducible promoter, its activator or in the replication origin of the plasmid. Despite our efforts to avoid it, it is possible that in these manipulations, some mutants might have been lost.

The low number of mutants with a higher activity we isolated with this method, the small number of positions touched by these mutations and the relative small increment in activity obtained (2- to 3-fold), suggests that the activity of IntI1 is not limited by the structure of the integrase. Indeed, we found only two positions (109 and 161) where substitutions slightly increase IntI1 activity. The two other mutants selected in the VCR_{WT} enrichment, L287M and A4V-G322E, show an even smaller increment (data not shown). In addition, none of these mutants showed a narrowing of the *attC* site recognition spectrum, as their recombination increment was not limited to the VCR site used for the selection, but was also effective with the shorter and unrelated $attC_{aadA7}$ site. This suggests that the IntI1 activity is not limited by a necessary structure/function compromise to recognize the many structurally different *attC* sites. The IntI1 structure is also constrained by the various interactions in which the protein is involved: (i) the interaction with ss substrates, (ii) the interaction with the ds substrate, (iii) the monomer folding and (iv) the interactions between the protein subunits. It is probable that the natural availability, or the production, of the ss form of the *attC* substrates inside a cell is the major limitation of the recombination, rather than the catalytic properties of the integrase itself.

DNA binding aptitude of the different IntI1 mutants

The results of the different EMSA experiments clearly show that one cannot reduce the effect of the selected protein mutations to a site recognition improvement. Indeed, both the $IntI1_{D161G}$ and $IntI1_{P109L}$ mutants bound the bs *attC* and ds *attI1*substrates only slightly better than $IntI1_{wt}$ (Figure 4 and Figure S3), hinting that the recognition of both substrates through specific binding is secondary in terms of selective forces in the assay, even if these increments could account for some of the observed increase in recombination frequency.

As mentioned in the results, a site such as VCR_{INV} which is recombined with a 1000-fold lower frequency than the VCR_{WT}, is efficiently bound *in vitro* and produce the same complexes as the WT site. This suggests that in this case, the reaction is affected at a later stage, either at the bs *attC* × ds *attI* synapse assembly or the cleavage activation. Moreover, if there is in general a rough correlation between the binding to a given substrate and its *in vivo* recombination rates, i.e. those bound the most poorly, being those recombined at lowest rate, as for example $VCRd_{GTT}$, this is not the case with the IntI1 mutants selected. Indeed, the very weak binding to the $VCRd_{GTT}$ and the $VCRd_{\triangle GTT}$ observed for the WT protein is conserved with both mutants, even if one of these, $IntI1_{D161G}$ which was selected on the first site, recombine them at 140-fold and 20-fold higher rates, respectively.

Another case of differences in the recombination of a given *attC* substrate that do not directly correlate with the binding, is that of $IntI1_{D161G}$ which binds VCR_{INV} better than VCR_{GTT} while its *in vivo* integration activity is 8-fold higher with the latter (Figure 4, Figures S3 and S5).

These discrepancies confirm that DNA binding in itself is not the unique determinant for an efficient recombination, and suggest that the selection with these specific mutants sites mainly acts on the aptitude to form a proficient synaptic complex, gathering the four subunits and the two different substrates, the ss *attC* and ds *attI*, and/or for the cleavage activation.

3D modeling of the IntI1 mutants

To get a better understanding of the mutants properties, we made a 3D model of $IntI1_{wt}$, based on the crystal structure of VchIntIA in complex with VCRbs (2). IntI1 and VchIntIA share 45% identity and 65% similarity. The model of the IntI1 molecule was built for both the cleaving and non-cleaving subunits. Regions that were not visible in electron density in the VchIntIA structure were not considered. The model shows the spatial positioning of the residues of IntI1 relative to their alignment with VchIntIA. For the sake of clarity, in the following part of the discussion, IntI1 amino acids numbering are always followed by a parenthesis giving the corresponding residue and its numbering in VchIntIA. A schematic picture of the *attC* × *attI1* synaptic complex and a secondary structure alignment between IntI1 and VchIntI1A are presented in Figure 6.

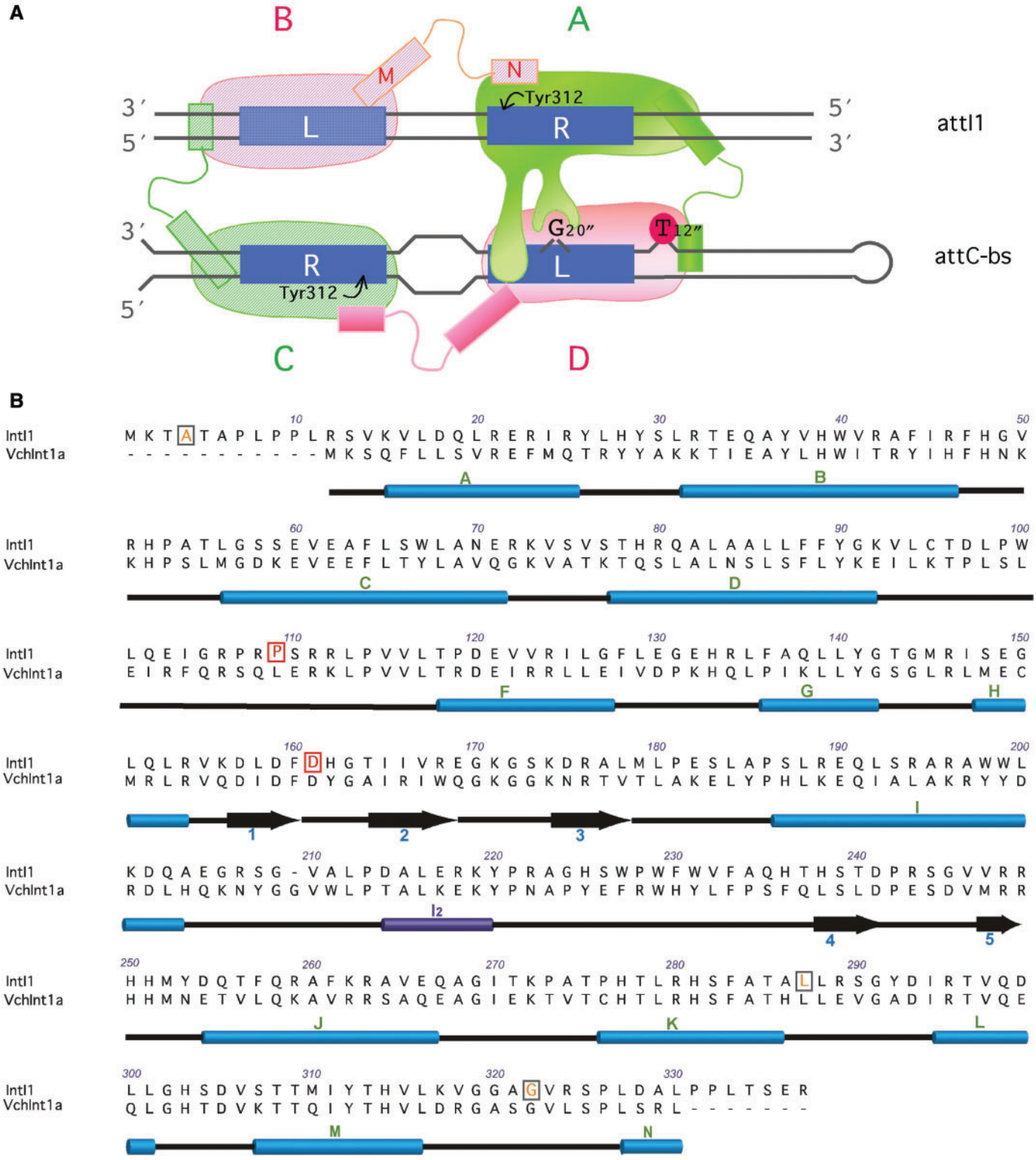

Figure 6. (**A**) A schematic picture of the IntI1 synapse with *attC* × *attI1*. Subunits B and C are striped since the positioning of these molecules cannot be predicted from the crystal structure of the VchIntIA, made from an *attC* × *attC* recombination complex. (**B**) Alignment of the protein sequences from IntI1 and VchIntA. The residues carried by the mutants selected by the enrichment assay are in frame. Those selected on the mutated $attC_{VCR}$ substrates are in bold, red and with a red frame. The secondary structure of VchIntIA with the β-strands (arrows) and the α-helices (cylinders) is presented below the sequence alignment.

Position 109. P109 is located in the linker between two helixes (D and F, see Figure 6) which functions as a hinge between the N-terminus and the C-terminus of the protein (Figure 7). A change from a sterically constrained proline (and helix breaker, see below) to a leucine, would give a higher flexibility to this linker region, which could help to accommodate the distortion of the substrate caused by the exchange of the unpaired central segment, and lead to a higher recombination frequency with this specific substrate. A correlate to this observation is the absence of prolines in this segment of the primary structure in the case of Cre. Indeed, the equivalent region in Cre carries α-helix E, which is involved in extensive inter subunit contacts with α-helix A for binding DNA in *cis*. The requirement for structured secondary elements in this region would favor ds type DNA binding, with 6–8 bp spacing. In the integron integrases (IntI1, VchIntIa) which bind non-canonical ss DNA with a bulge, the presence of two prolines (P107, P109) and one proline (P86) respectively, maintains this region in an unstructured state thus providing the necessary flexibility required for recognition and binding of such substrates (Figure 7B). The observation that the C-terminal part of the linker, which contains this residue, is in close proximity to the DNA of the central unpaired segment (T26″–T28″ and facing nt G16′–C18′, see figure 2D), strengthen the hypothesis that the higher flexibility in this region could enhance the formation of a productive synapse in the case of a modified central segment, be it inverted or fully annealed. It is noticeable that the P109L substitution does not affect the recombination of the substrates mutated in the extrahelical bases T12″, T16″ and G20″, which are located in remote regions of the substrate. On the other hand, the P109L substitution has a negative effect on the $attC_{WT} \times attC_{WT}$ recombination, as attested by the 6-fold decrease of the deletion frequency (Figure 3). Interestingly, however, this also suggests that the higher flexibility introduced in the P109L protein is prejudicial for the synaptic complex stabilization formed in the bs *attC* × bs *attC* reaction.

Position 161. The integron integrases alignment shows that the amino acid residue found at this position is in most cases an aspartate or a glutamate. In both IntI1 and VchIntIA, it is an aspartate, D161 and D150, respectively. The 3D model reveals that it is situated in a key area at the interface between the A and D (and the C and B) molecules in the protein–DNA synapse (Figure 7C). Indeed, across the D161 residue of subunit A [$D161_A$] two residues are found within distance for hydrogen bonding or ionic interactions, protruding on the D molecule, $K272_D$ (K262) and $E121_D$ (E110). $K272_D$ (K262) is preceded by the helix J that is in contact with the major groove (Figure 7C). Another attribute of this area is its proximity to the entry point of the D subunit C-terminal helix N (Figure 7C), which has a stabilizing function on the synaptic complex, and postulated to participate in the regulation of the cleavage step in the case of the Cre recombinase (27). D161 is located in the loop between the β-sheets 1 and 2. β-sheet 2 contains one of the two residues, R168 (W157), which interact with

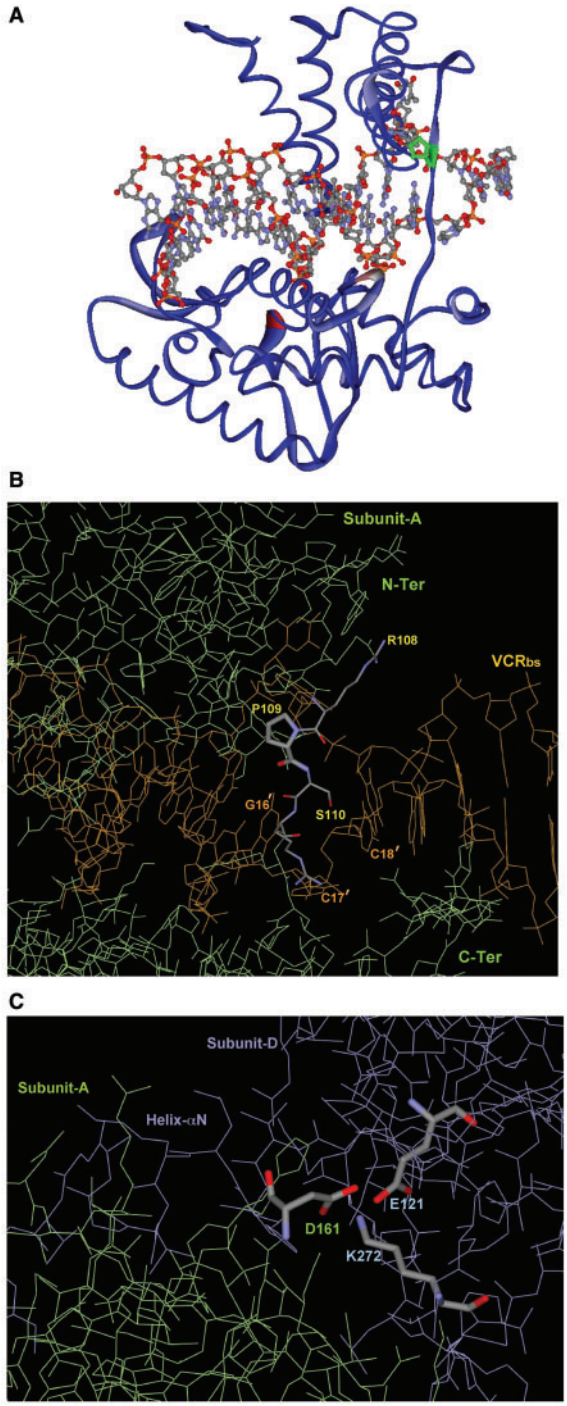

Figure 7. Model of IntI1 based on VchIntIa-VCRbs x-ray structure (2). (**A**) The monomer corresponding to subunit A is shown, positioned relative to the VCRbs DNA. The position of P109 is shown in green, in the linker region spanning the N-to-C terminal domains of the integron integrase. (**B**) Close up of the P109L mutant in the linker region of the intI1 model. P109 is located in a key position to influence the flexibility of the N-ter to C-ter domain hinge. When compared to Cre, this region harbors an α-helix (α-E), thus much more structured that in the integron integrases. It is also in the viscinity of the bulge due to the 3 bp mismatch of the DNA substrate. (**C**) Close up of the D161 mutant in the intI1 model. The region is located at the interface between trans-bound monomers. Interfacial contacts between $D161_A$ and $E121_D$ or $K272_D$ help to maintain these molecules closer, hence reinforcing the synapse. This region is the entry point of the domain swapped α-N helix between the D and A subunits.

the extrahelical base G20″ (2) and stabilize the synaptic complex. It is remarkable that three of the selected mutants carry a mutation of the D161 residue. The substitution of D161 by a glycine, does not lead to the same phenotype as its substitution by a valine or a glutamate (Figure 3). Glycine has no lateral chain and can radically change the structure of the loop between β-sheets 1 and 2, while the lateral chain size of a valine or a glutamate should conserve the global loop structure. These results suggest that this loop, and D161 in particular, are very important for the inter subunit protein/protein interaction. Mutations made of the next amino acid in this loop H162, further illustrates the importance of the loop, where substitutions to a E, G or A, decrease the recombination frequency about 5-fold, and to a K, by 100-fold (data not shown). However, no gain-of-function mutations were found at this position.

A structural trade-off to guaranty productive recombination

As illustrated by Grindley and colleagues in their review (28), the outcome of the recombination is determined by the conformation of the DNA in the initial synaptic complex. The initial recombinase-bound duplex bending would decide which strand would be cleaved first. bs *attC* site secondary structure does not offer the two alternate bending possibilities. The unpaired central segment imposes a structural bending, visible in the 3D structure of the crystallized complex, which shows an elbow in the bs *attC* site due to the unpaired central segment (2). Moreover, the interactions in *cis* and *in trans* with the extrahelical bases do not allow HJ isomerization (2), such that only one conformation is compatible with recombination. Hence, only one strand is cleaved and this is always the same strand. Therefore, one can consider that the unpaired central segment and the extrahelical bases directly or indirectly determine the outcome of recombination. This is supported by our results, as modifications of the central segment structure or mutations of the extrahelical bases directly affect the $IntI1_{WT}$ recombination activity. The $IntI1_{D161G}$ and $IntI1_{P109L}$ mutants restore recombination activity on the mutated substrate used for their respective selection. It is interesting to note that both show the same *in vivo* properties—a higher activity in the *attI* × *attC* reaction, and a lower activity for the *attC* × *attC* reaction—while *in vitro*, they both show a slight increment in their affinity for both the bs *attC* and the ds *attI*. According to structural modeling, both mutations would lead to an increment in the relative flexibility of the integrase (intra, and inter subunit), and modify the protein/protein or the protein/DNA interactions involved in the integration reaction synapse. These ultimately lead to an increase in the recombination frequencies. On the other hand, the correlated relative decrease in the *attC* × *attC* recombination frequency suggests that an increase in protein flexibility is deleterious for the assembly or the stability of synaptic complexes exclusively involving the folded bs *attC*. An explanation could be that complex formation, which involves a flexible bs *attC* and a stiffer ds *attI,* is facilitated with an IntI1 mutant having an improved flexibility, whilst in the bs *attC* × bs *attC* complexes, both substrates are flexible and the resulting synapse become less stable than in the bs *attC* × ds *attI* reaction. The extrahelical G is involved in the stability of the synapse, via its interaction *in trans* with the active subunits (2). This particular conformation allows the correct protein/protein interaction between the subunits bound to the same type of substrate. A higher flexibility in IntI1 structure would allow for better *attI1* binding but would no longer provide this *in trans* interaction that stabilizes the synapse. In the integration process, the loss of flexibility would be compensated for by the gain of *attI* binding, while in the deletion reaction, the loss of flexibility would affect both IntI1 dimer-*attC* complexes. One can imagine that in this case, the gain in binding capacity we observe *in vitro* is not sufficient to compensate for the flexibility loss.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors acknowledge Dr D. Rowe-Magnus for critical reading of the manuscript. This study was carried out with financial assistance from the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS-URA 2171 and 2185) and the EU (STREP CRAB, LSHM-CT-2005-019023, and NoE EuroPathoGenomics, LSHB-CT-2005-512061). Funding to pay the Open Access publication charges for this article was provided by Institut Pasteur.

Conflict of interest statement. None declared.

REFERENCES

1. Mazel,D. (2006) Integrons: agents of bacterial evolution. *Nat. Rev. Microbiol.*, **4**, 608–620.
2. MacDonald,D., Demarre,G., Bouvier,M., Mazel,D. and Gopaul,D.N. (2006) Structural basis for broad DNA specificity in integron recombination. *Nature*, **440**, 1157–1162.
3. Nunes-Duby,S.E., Kwon,H.J., Tirumalai,R.S., Ellenberger,T. and Landy,A. (1998) Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.*, **26**, 391–406.
4. Messier,N. and Roy,P.H. (2001) Integron integrases possess a unique additional domain necessary for activity. *J. Bacteriol.*, **183**, 6699–6706.
5. Nield,B.S., Holmes,A.J., Gillings,M.R., Recchia,G.D., Mabbutt,B.C., Nevalainen,K.M. and Stokes,H.W. (2001) Recovery of new integron classes from environmental DNA. *FEMS Microbiol. Lett.*, **195**, 59–65.
6. Bouvier,M., Demarre,G. and Mazel,D. (2005) Integron cassette insertion: a recombination process involving a folded single strand substrate. *EMBO J.*, **24**, 4356–4367.
7. Biskri,L., Bouvier,M., Guerout,A.M., Boisnard,S. and Mazel,D. (2005) Comparative study of class 1 integron and *Vibrio cholerae* superintegron integrase activities. *J. Bacteriol.*, **187**, 1740–1750.
8. Hansson,K., Sundstrom,L., Pelletier,A. and Roy,P.H. (2002) IntI2 integron integrase in Tn7. *J. Bacteriol.*, **184**, 1712–1721.
9. Collis,C.M., Kim,M.J., Partridge,S.R., Stokes,H.W. and Hall,R.M. (2002) Characterization of the class 3 integron and the site-specific recombination system it determines. *J. Bacteriol.*, **184**, 3017–3026.

10. Collis,C.M. and Hall,R.M. (1992) Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J. Bacteriol.*, **174**, 1574–1585.
11. Collis,C.M., Recchia,G.D., Kim,M.J., Stokes,H.W. and Hall,R.M. (2001) Efficiency of recombination reactions catalyzed by class 1 integron integrase IntI1. *J. Bacteriol.*, **183**, 2535–2542.
12. Biskri,L. and Mazel,D. (2003) Erythromycin esterase gene ere(A) is located in a functional gene cassette in an unusual class 2 integron. *Antimicrob. Agents Chemother.*, **47**, 3326–3331.
13. Francia,M.V., Zabala,J.C., de la Cruz,F. and Garcia-Lobo,J.M. (1999) The IntI1 integron integrase preferentially binds single-stranded DNA of the *attC* site. *J. Bacteriol.*, **181**, 6844–6849.
14. Collis,C.M., Kim,M.J., Stokes,H.W. and Hall,R.M. (1998) Binding of the purified integron DNA integrase Intl1 to integron- and cassette-associated recombination sites. *Mol. Microbiol.*, **29**, 477–490.
15. Gravel,A., Fournier,B. and Roy,P.H. (1998) DNA complexes obtained with the integron integrase IntI1 at the attI1 site. *Nucleic Acids Res.*, **26**, 4347–4355.
16. Voziyanov,Y., Konieczka,J.H., Stewart,A.F. and Jayaram,M. (2003) Stepwise manipulation of DNA specificity in Flp recombinase: progressively adapting Flp to individual and combinatorial mutations in its target site. *J. Mol. Biol.*, **326**, 65–76.
17. Bartolome,B., Jubete,Y., Martinez,E. and de la Cruz,F. (1991) Construction and properties of a family of pACYC184-derived cloning vectors compatible with pBR322 and its derivatives. *Gene*, **102**, 75–78.
18. Simons,R.W., Houman,F. and Kleckner,N. (1987) Improved single and multicopy lac-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.
19. Morabito,S., Tozzoli,R., Caprioli,A., Karch,H. and Carattoli,A. (2002) Detection and characterization of class 1 integrons in enterohemorrhagic. *Escherichia coli. Microb. Drug Resist.*, **8**, 85–91.
20. Bernard,P. and Couturier,M. (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.*, **226**, 735–745.
21. Johansson,C., Kamali-Moghaddam,M. and Sundstrom,L. (2004) Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res.*, **32**, 4033–4043.
22. Bradford,M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.*, **72**, 248–254.
23. Gravel,A., Messier,N. and Roy,P.H. (1998) Point mutations in the integron integrase IntI1 that affect recombination and/or substrate recognition. *J. Bacteriol.*, **180**, 5437–5442.
24. Mazel,D., Dychinco,B., Webb,V.A. and Davies,J. (1998) A distinctive class of integron in the *Vibrio cholerae* genome. *Science*, **280**, 605–608.
25. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
26. Clackson,T., Hoogenboom,H.R., Griffiths,A.D. and Winter,G. (1991) Making antibody fragments using phage display libraries. *Nature*, **352**, 624–628.
27. Guo,F., Gopaul,D.N. and van Duyne,G.D. (1997) Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature*, **389**, 40–46.
28. Grindley,N.D., Whiteson,K.L. and Rice,P.A. (2006) Mechanisms of site-specific recombination. *Annu. Rev. Biochem.*, **75**, 567–605.
29. Demarre,G., Guerout,A.M., Matsumoto-Mashimo,C., Rowe-Magnus,D.A., Marlière,P. and Mazel,D. (2005) A new family of mobilizable suicide plasmids based on the broad host range R388 plasmid (IncW) or RP4 plasmid (IncPa) conjugative machineries and their cognate *E. coli* host strains. *Res. Microbiol.*, **156**, 245–255.
30. Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.*, **185**, 60–89.
31. Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103–119.
32. 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.
33. Rowe-Magnus,D.A., Guerout,A.-M., Ploncard,P., Dychinco,B., Davies,J. and Mazel,D. (2001) The evolutionary history of chromosomal super-integrons provides an ancestry for multi-resistant integrons. *Proc. Natl Acad. Sci. USA*, **98**, 652–657.