Published online 13 November 2009

Nucleic Acids Research, 2010, Vol. 38, No. 2 ***559–569***
doi:10.1093/nar/gkp990

The relaxed requirements of the integron cleavage site allow predictable changes in integron target specificity

Clara Frumerie[1], Magaly Ducos-Galand[1], Deshmukh N. Gopaul[2] and Didier Mazel[1,*]

[1]Institut Pasteur, Unité Plasticité du Génome Bactérien, CNRS URA 2171 and [2]Unité d'Immunologie Structurale, CNRS URA 2185, 25 rue du Dr Roux, 75724, Paris, France

Received August 10, 2009; Revised October 14, 2009; Accepted October 15, 2009

ABSTRACT

Integrons are able to incorporate exogenous genes embedded in mobile cassettes, by a site-specific recombination mechanism. Gene cassettes are collected at the *attI* site, via an integrase mediated recombination between the cassette recombination site, *attC*, and the *attI* site. Interestingly, only three nucleotides are conserved between *attC* and *attI*. Here, we have determined the requirements of these in recombination, using the recombination machinery from the paradigmatic class 1 integron. We found that, strikingly, the only requirement is to have identical first nucleotide in the two partner sites, but not the nature of this nucleotide. Furthermore, we showed that the reaction is close to wild-type efficiency when one of the nucleotides in the second or third position is mutated in either the *attC* or the *attI1* site, while identical mutations can have drastic effects when both sites are mutated, resulting in a dramatic decrease of recombination frequency compared to that of the wild-type sites. Finally, we tested the functional role of the amino acids predicted from structural data to interact with the cleavage site. We found that, if the recombination site triplets are tolerant to mutation, the amino acids interacting with them are extremely constrained.

INTRODUCTION

Integrons are genetic elements commonly found in gram-negative bacteria either as sedentary chromosomal components or embedded in transposons (1,2). They have the capability to incorporate promoterless gene cassettes site-specifically in proximity to a promoter (P_C) and thereby allow the expression of the gene carried by the cassette (1). The integron platform consists of a tyrosine recombinase encoding gene, the promoter P_C, and the *attI* insertion site where the cassette integration occurs. We recently reported that integrase expression is triggered during the SOS response (3). Integrative recombination takes place between a gene cassette recombination site, the *attC* site (or 59 bp element) which can greatly vary in length and sequence, and the *attI* site of the integron platform (1). Recombination can also occur via two flanking *attC* sites (*attC* × *attC* recombination), resulting in excision of a gene cassette from the integron, or between two *attI* sites, a reaction whose biological relevance is not yet known (4–6). Integrons are classified according to the integrase they encode, with two paradigmatic systems, the class 1 integron for the mobile integrons and the *Vibrio cholerae* superintegron for the sedentary ones (1). Previous work on class 1 integrons, defined by the encoded IntI1 integrase, has shown that only the *attC* site bottom strand was recombined after folding of the single-stranded (ss) DNA over the palindromic sequence that defines the *attC* site (7). The ss folding generates the proper substrate for recombination and the recombined strand selection is directed by unpaired nucleotides (nt) (8,9) (Figure 1). The *attI1* site, however, is in canonical double stranded form during the process (7). Additionally, IntI1 accepts several distinct *attC* sites as substrate, while it only efficiently recognizes its own *attI1* site (5,10,11). The recognition characteristics of IntI1 mean that very little sequence identity is necessary between *attC* and *attI1*. In particular, and in contrast to most other tyrosine recombinase substrates which require a common 6–8 nt core sequence, only three completely conserved nts, GTT, are observed (1,12–14) (Figures 1 and 2).

The crystal structure of the integron integrase VchIntIA, from the *V. cholerae* superintegron, bound to the ss *attC* in the synaptic complex substrate representing

*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:
Clara Frumerie. Department of Molecular Biology and Functional Genomics, Stockholm University, Svante Ahrenius väg 18F, 106 91 Stockholm, Sweden.

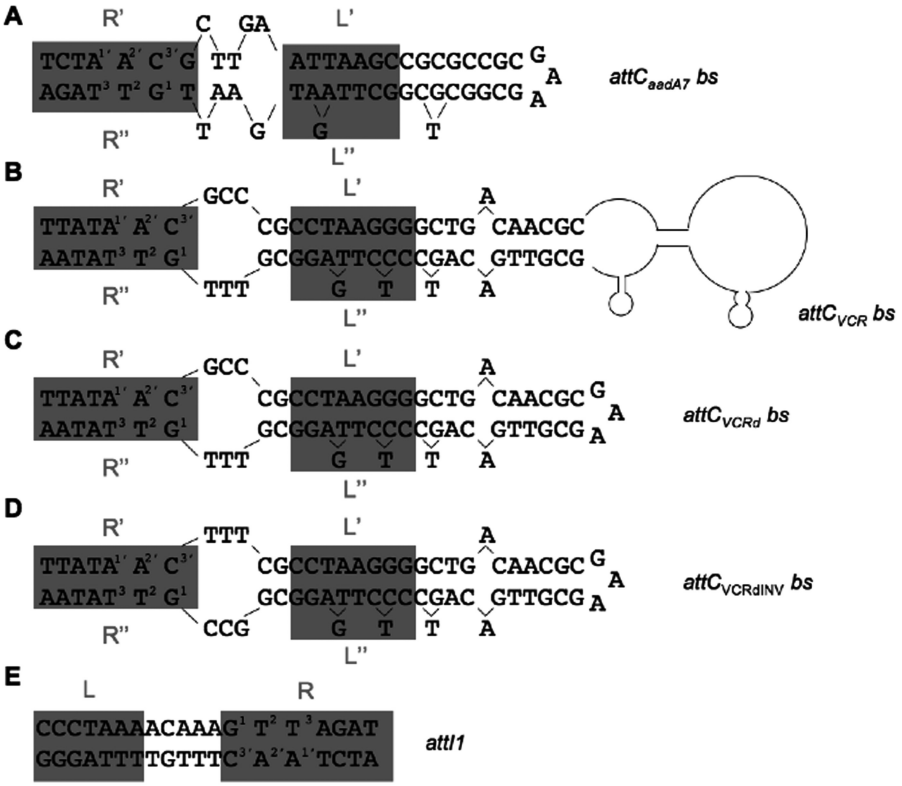

Figure 1. Sequence of the substrates used in this study. Bottom strand sequence and folding of the $attC_{aadA7}$ (**A**), $attC_{VCR}$ (**B**), $attC_{VCRd}$ (**C**), and $attC_{VCRdINV}$ (**D**). (**E**) Sequence of the double strand attI1 site core. The cleavage site, G1T2T3/$A^{1'}A^{2'}A^{3'}$, is indicated for each substrate.

the excising complex (*attC* × *attC* recombination), has been solved (8). This crystal structure has been used to model the highly similar (65% identity) IntI1 (15). These structures show no base specific interaction between the first nt of the cleavage site and the integrase (8). This led to the hypothesis that the conserved G^1 in the $G^1T^2T^3$ triplet is conserved because of the requirement for identity between the sites, rather than the properties of the base itself. Furthermore, analysis of the VchIntIA structure and our IntI1 model reveals that the only amino acid (aa) that shows base specific interactions with the three conserved nt of the cleavage site, is the conserved lysine aa K171 (K160, in VchIntIA numbering), which interacts with the last T of the GTT triplet and the middle nt A of the opposite strand (Figure 2B). The remainder of the aa, which are known or predicted to interact with the cleavage site, do so with the phosphate backbone only, and many of these interactions are weak (more than 3.5 Å away) (8). In this work, the cleavage site sequence has been dissected to elucidate the importance of the respective nt of the GTT triplet in the recombination reaction, using the cleavage site of $attC_{aadA7}$ (Figure 1). We show that the first nt is responsible for the docking and that the base in itself is indifferent to the cleavage and strand-exchange reaction. This is consistent with the requirement for micro-homology readout in tyrosine recombinase site specific recombination (16), with the novelty that in class 1 integron, *attC* × *attI1* recombination, only one-nt is required for this. Furthermore, we found that mutations in the second and third nt in either the *attC* or the *attI1* site, still yields full recombination products, whereas the same mutation can be detrimental when present on both sites simultaneously.

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 (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 0.8 mM. IPTG (isopropyl-β-D-thiogalactopyranoside) was added at 0.8 mM final concentration. Chemicals were from Sigma.

PCR procedures

PCR reactions for topo cloning were performed using GoTaq polymerase (Promega). PCR reactions for site-directed mutagenesis (SDM) used the Pfu turbo polymerase (Stratagene). The primers listed in Table 3, were from Sigma-Aldrich (France).

Plasmid constructions

The plasmids with mutations in the $G^1T^2T^3$ cleavage sites were constructed by assembling two long partially hybridizing primers (Table 3) for the different

Table 1. Bacterial strains used in this study

Escherichia coli strains	Pertinant feature(s)	Origin or reference
DH5α	(F$^-$) *supE44 ΔlacU169 (Φ80lacZΔM15) ΔargF hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Laboratory collection
UB5201	F-pro met recA56 gyrA [Nal]^R	(25)
β2163	(F$^-$) RP4-2-Tc::Mu *ΔdapA*::(*erm-pir*) [Km^R Em^R]	(26)
BL21(DE3)pLysS	*E. coli* B strain containing the T7 polymerase and plasmid pLysS expressing T7 lysozyme	(27)

Table 2. Plasmids used in this study

Plasmids	Description	Origin or reference
p1266	pSU38Δ*lacZα*, ori_{p15A} [Km]^R	(15)
p112	pTRC99A::*intI1*, ori_{ColE1} [Ap]^R	(28)
p4634	pET3a::*HisInt* [Ap]^R	(15)
p10000	pET3a::*HisInt* $_{Y37F}$[Ap]^R	This work
p4632	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{wt}$. *attI1* starts –81nt from the cross-over point	This work
p4867	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteATT}$	This work
p5060	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteCTT}$	This work
p5059	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteTTT}$	This work
p4761	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteGAT}$	This work
p4669	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteGGT}$	This work
p4671	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteGTA}$	This work
p5019	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteGTC}$	This work
p4674	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteGAA}$	This work
p5020	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteGGG}$	This work
p7817	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1_{cleavagesiteGTG}$	This work
p929	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1long_{wt}$	(7)
p4394	pSU38Δ*lacZα*, ori_{p15A} [Km]^R_$attI1long_{cleavagesiteGCT}$	This work
p3030	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7wt}$	(7)
p8144	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteG1T2T3 - A1'T2'C3'}$	This work
p4741	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteATT}$	This work
p6751	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteCTT}$	This work
p6009	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteTTT}$	This work
p4986	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteGAT}$	This work
p4397	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteGCT}$	This work
p4739	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteGGT}$	This work
p4740	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteGTA}$	This work
p6787	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteGTC}$	This work
p4738	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteGAA}$	This work
p5017	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteGGG}$	This work
p7524	pSW23T, oriT_RP4 [Cm]^R_$attC_{aadA7cleavagesiteGTG}$	This work
p4907	pSW23T, oriT_RP4 [Cm]^R_$attC_{VCRd}$	(9,15)
p4836	pSW23T, oriT_RP4 [Cm]^R_$attC_{VCRdbulgeinversed}$	(9,15)
p7138	pSW23T, oriT_RP4 [Cm]^R_$attC_{VCRdbulgeinversedcleavagesiteTTT}$	This work
p7137	pSW23T, oriT_RP4 [Cm]^R_$attC_{VCRdcleavagesiteTTT}$	This work
p8141	pSW23T, oriT_RP4 [Cm]^R_$attC_{VCRdcleavagesiteGTC}$	This work
p8140	pSW23T, oriT_RP4 [Cm]^R_$attC_{VCRdcleavagesiteGCT}$	This work

recombination sites harboring, an EcoR1 and a BamH1 site at their respective ends. These were then subjected to PCR, cloned into vector pCR2.1 TOPO (Invitrogen) and transformed into Top10 competent cells (Invitrogen). After sequencing, clones carrying the desired recombination site were cleaved with EcoR1 and BamH1 and ligated into vectors pSU38Δ (p1266) for the *attI1* sites, and pSW23T (26) for the *attC* sites. The $attC_{aadA7CTT}$ was generated within the sites Sal1 and BamH1.

pET3a_HisIntI1$_{Y37F}$ was constructed using SDM with pET_HisIntI1 as a template (15).

Integration assay

The integration assay was performed as described (7) with the following modifications. After concentration of the recipient and donor cells on a 0.45 μm filter, the bacteria were incubated at 37°C for 5 h on LB plates with DAP and plated on the appropriate antibiotics. The recombination frequency was calculated as the ratio of Cm and nalidixic acid (antibiotic resistance markers of the donor plasmid and the recipient strain UB5201) resistant cells over the number of Km and Ap (antibiotic resistance markers of the plasmids carrying the *attI1* site and *intI1*) resistant cells. The correct recombination site was verified by PCR using at least eight randomly picked colonies. The number of correct clones thus identified was multiplied by the value for the original frequency to eliminate the background colonies. For clones with efficient recombination such as the wild-type (wt), the number of correct clones was usually 8/8, however, with clones having poor

Table 3. Oligonucleotides used in this study

Oligos	Sequences
Plasmid construction	
R1attI1wt	GAATTCAGCGCGTTACGCCGTGGGTCGATGTTTGATGTTATGGAGCAGCAACGATGTTACGCAGCAGGGCAGTCGCC
H1attI1wt	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTAACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{ATT}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTAATTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{CTT}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTAAGTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{TTT}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTAAATTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{GAT}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTATCTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
attI1$_{GCT}$sdm	CGCCCTAAAACAAAGCTAGGCATCACAAAGTACAG
attI1$_{GCT}$sdm	CTGTACTTTGTGATGCCTAGCTTTGTTTTAGGGCG
H1attI1$_{GGT}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTACCTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{GTA}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTTACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{GTC}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTGACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{GTG}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTCACTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{GAA}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTTTCTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
H1attI1$_{GGG}$	GGATCCTGTTGGTCACGATGCTGTACTTTGTGATGCCTCCCTTTGTTTTAGGGCGACTGCCCTGCTGCGTAACATCG
R1attC$_{ATT}$	GAATTCATGTCTAATAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{ATT}$	GGATCCATGTCTAATGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
H1attC$_{CTT}$	GGATCCATGTCTAAGAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
SalI1 attC$_{CTT}$	CCGCGGATGTCTAAGGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{TTT}$	GAATTCATGTCTAAAAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{TTT}$	GGATCCATGTCTAAAGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{GAT}$	GAATTCATGTCTATCAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{GAT}$	GGATCCATGTCTATCGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{GCT}$	GAATTCATGTCTAGCAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{GCT}$	GGATCCATGTCTAGCGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{GGT}$	GAATTCATGTCTACCAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{GGT}$	GGATCCATGTCTACCGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{GTA}$	GAATTCATGTCTTACAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{GTA}$	GGATCCATGTCTTACGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{GTC}$	GAATTCATGTCTGACAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{GTC}$	GGATCCATGTCTGACGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{GTG}$	GAATTCATGTCTCACAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{GTG}$	GGATCCATGTCTCACGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{GAA}$	GAATTCATGTCTTTCAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{GAA}$	GGATCCATGTCTTTCGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
R1attC$_{GGG}$	GAATTCATGTCTCCCAATTCATTCAAGCCGACGCCGCTTCGCGGCGCGGCT
H1attC$_{GGG}$	GGATCCATGTCTCCCGCTTGAATTAAGCCGCGCCGCGAAGCGGCGTCG
IntI1Y37Fbot	GGCACGAACCCAGTGGACAAAAGCCTGTTCGGTTCGTAAGC
IntI1Y37Ftop	GCTTACGAACCGAACAGGCTTTTGTCCACTGGGTTCGTGCC
PCR	
MFD	CGCCAGGGTTTTCCCAGTCAC
SW23begin	CCGTCACAGGTATTTATTCGGCG

recombination efficiency the number of correct clones could be as little as 1/8, where the others had recombined into secondary sites. The recombination frequencies presented in Table 4 are the average of at least three experiments, normalized to that obtained with the wt. The wt frequency is 2.2×10^{-3}, except for mutations GTT→GCT, which were performed with plasmid p929 containing a longer *attI1* fragment than p4632, and the wt recombination frequency was found to be 3.45×10^{-2}.

Protein purification

Protein purification was carried out as described previously (15).

EMSA

Each reaction contained 50 ng polyDI–polyDC, 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 the amount of protein indicated in the figure, in a final volume of 20 μl. The samples were incubated 30°C for 10 min before loaded to a 5% native polyacrylmide gel (30% Acrylamide/Bis Solution, 29:1 BioRad), with 1× TBE as buffer. The gel was run at 20 mA with room temperature water as cooling and 0.5× TBE as running buffer, for ~2 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.

Cleavage assay

The cleavage assay was performed as above until the incubation step. Afterwards, denaturing stop solution was added, and the sample loaded onto a 5% denaturing

Table 4. Recombination frequencies of the different cleavage site mutants after normalization with the wt $attC_{aadA7} \times attI1$ recombination frequency or wt $attC_{VCRd} \times attI1$ recombination frequency

A	Mutations in nucleotide nr 1 of the $G^1T^2T^3$ cleavage site[a]					
	$attC_{aadA7}/attI1\mathbf{ATT}$	3.4×10^{-3} (2.3×10^{-3})	$attC_{aadA7}\mathbf{ATT}/attI1$	3.1×10^{-3} (4.1×10^{-3})	$attC_{aadA7}\mathbf{ATT}/attI1\mathbf{ATT}$	1 (6.4×10^{-1})
	$attC_{aadA7}/attI1\mathbf{CTT}$	1.5×10^{-4} (2.5×10^{-4})	$attC_{aadA7}\mathbf{CTT}/attI1$	1.5×10^{-3} (1.7×10^{-3})	$attC_{aadA7}\mathbf{CTT}/attI1\mathbf{CTT}$	9.7×10^{-1} (3.4×10^{-1})
	$attC_{aadA7}/attI1\mathbf{TTT}$	none detected	$attC_{aadA7}\mathbf{TTT}/attI1$	2.1×10^{-4} (2.1×10^{-4})	$attC_{aadA7}\mathbf{TTT}/attI1\mathbf{TTT}$	4.4×10^{-2} (4.3×10^{-2})
B	Mutations in nucleotide nr 2 of the $G^1T^2T^3$ cleavage site					
	$attC_{aadA7}/attI1\mathbf{GAT}$	4.6 (3.8)	$attC_{aadA7}\mathbf{GAT}/attI1$	1.9 (0.9)	$attC_{aadA7}\mathbf{GAT}/attI1\mathbf{GAT}$	3.2 (2.7)
	$attC_{aadA7}/attI1\mathbf{GCT}$	2.1 (1.39)	$attC_{aadA7}\mathbf{GCT}/attI1$	1.5×10^{-1} (1×10^{-1})	$attC_{aadA7}\mathbf{GCT}/attI1\mathbf{GCT}$	4.1×10^{-2} (1.9×10^{-2})
	$attC_{aadA7}/attI1\mathbf{GGT}$	9.6×10^{-1} (5.7×10^{-1})	$attC_{aadA7}\mathbf{GGT}/attI1$	1 (5.8×10^{-1})	$attC_{aadA7}\mathbf{GGT}/attI1\mathbf{GGT}$	7.1×10^{-1} (4.3×10^{-1})
	$attC_{VCRd}/attI1\mathbf{GCT}$	8.7×10^{-1} (1.4×10^{-1})	$attC_{VCRd}\mathbf{GCT}/attI1$	4.8×10^{-1} (2.7×10^{-2})	$attC_{VCRd}GCT/attI1\mathbf{GCT}$	3.6×10^{-2} (2.7×10^{-2})
C	Mutations in nucleotide nr 3 of the $G^1T^2T^3$ cleavage site					
	$attC_{aadA7}/attI1\mathbf{GTA}$	1.3 (1.1)	$attC_{aadA7}\mathbf{GTA}/attI1$	5.6×10^{-1} (3.6×10^{-1})	$attC_{aadA7}\mathbf{GTA}/attI1\mathbf{GTA}$	5.6×10^{-2} (1.5×10^{-2})
	$attC_{aadA7}/attI1\mathbf{GTC}$	1.9 (1.5)	$attC_{aadA7}\mathbf{GTC}/attI1$	1 (7.3×10^{-1})	$attC_{aadA7}\mathbf{GTC}/attI1\mathbf{GTC}$	5.6×10^{-3} (2.5×10^{-3})
	$attC_{aadA7}/attI1\mathbf{GTG}$	9.4×10^{-1} (5.1×10^{-1})	$attC_{aadA7}\mathbf{GTG}/attI1$	7.3×10^{-2} (4.1×10^{-2})	$attC_{aadA7}\mathbf{GTG}/attI1\mathbf{GTG}$	1.2×10^{-2} (9.8×10^{-3})
	$attC_{VCRd}/attI1\mathbf{GTC}$	2.6×10^{-1} (2.1×10^{-1})	$attC_{VCRd}\mathbf{GTC}/attI1$	4.4×10^{-2} (2.4×10^{-2})	$attC_{VCRd}\mathbf{GTC}/attI1\mathbf{GTC}$	4.8×10^{-5} (2.4×10^{-4})
D	Mutations in nucleotides nr 2 and 3 of the $G^1T^2T^3$ cleavage site					
	$attC_{aadA7}/attI1\mathbf{GAA}$	7.5×10^{-1} (2.8×10^{-1})	$attC_{aadA7}\mathbf{GAA}/attI1$	4.4×10^{-1} (3.7×10^{-1})	$attC_{aadA7}\mathbf{GAA}/attI1\mathbf{GAA}$	1.2×10^{-3} (1.3×10^{-3})
	$attC_{aadA7}/attI1\mathbf{GGG}$	1.1×10^{-1} (3.4×10^{-2})	$attC_{aadA7}\mathbf{GGG}/attI1$	7.3×10^{-3} (5.5×10^{-3})	$attC_{aadA7}\mathbf{GGG}/attI1\mathbf{GGG}$	1.6×10^{-5} (1.1×10^{-5})

[a]The experiments were performed with $attC_{aadA7}$wt and mutated *attI1*, mutated $attC_{aadA7}$ and *attI1*wt and finally with mutated $attC_{aadA7}$ and mutated *attI1*, for each of the mutations described.
Standard deviations are presented in parenthesis.

polyacrylmide gel with 1% of SDS. SDS (0.1%) was added in the sample buffer. In the reactions, 30 pmol of the appropriate protein was used. The covalently bound DNA–protein complex will migrate at a slower rate relative to free DNA and make a shift on the gel.

RESULTS

The first-nt of the GTT triplet requires homology between the *attC* and *attI1* for effective recombination

In the literature, the sequence of the cleavage site is denoted GTT while in fact cleavage occurs on the opposite strand, between the A and the C of AAC. Thus, the wt site will here be described as GTT (or $G^1T^2T^3$) and the mutants made for this study will be identified as mutations thereof. The opposite strand will be denoted $A^{1'}A^{2'}C^{3'}$ (Figure 1).

Because of the ss folded structure of the *attC* site, two complementary mutations are necessary at this site to ensure the complementarity of the folded strand (Figure 1). The recombination frequency between an $attC_{aadA7}$ site with a non-complementary cleavage site $G^1T^2T^3$ – $A^{1'}T^{2'}C^{3'}$ and the $attI1_{wt}$, is 6×10^{-6} while the wt ($G^1T^2T^3$–$A^{1'}A^{2'}C^{3'}$) $attC_{aadA7wt} \times attI1_{wt}$ recombination frequency is 2.2×10^{-3}. The chance of the same spontaneous mutations in 3 nt of the *attI1* and *attC* sites at once, is very low, $(1/4 \times 10^{-9})^3$, assuming a probability for a base pair to be mutated to 10^{-9} per bacterial generation (17). This means that the need for complementary nts upon strand exchange in order to achieve effective recombination is highly conserved and to find the corresponding match of a mutated nt in the facing substrate mutated as well is quite unlikely. We constructed *attC* site mutants, where we modified both complementary positions of the site. Recombination experiments testing mutations of the $G^1/C^{3'}$ in only one or the other of the two partner sites, the $attC_{aadA7}$ and $attI1_{wt}$, sites yield a frequency that is 10^3–10^4 times lower than the frequency obtained with two wt sites (Table 4). However, when performed with the partner site carrying the same mutation the recombination frequency is fully restored. The reason for this becomes evident when looking at the schematic picture of the recombination event in Figure 2A. In order for the hybridization of the exchanging DNA strand to occur on the synapsed DNA molecule, complementarity is mandatory.

However, the mutation $G^1 \rightarrow T^1$ differs from this pattern, in the sense that it is still 20 times lower compared to the wt, when both the $attC_{aadA7}$ and *attI1* sites carry this mutation (Table 4). The sequence of the $attC_{aadA7}$ substrate, exposes three possible Ts as docking sites for the invading $A^{3'}$ from the *attI1* DNA molecule (Figure 2A), since after cleavage, two Ts from the bulge of

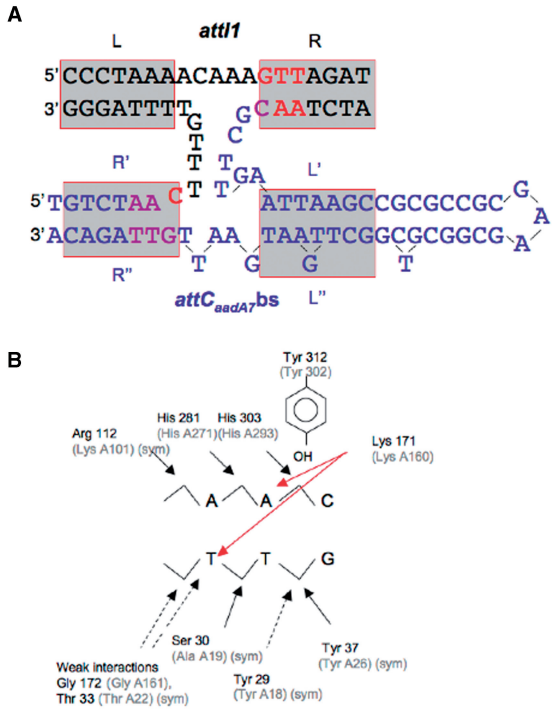

Figure 2. (**A**) Holliday Junction with the sequence of *attI1* and $attC_{aadA7}$bs. The *attI1* sequence is represented in black, with the cleavage site in red and the $attC_{aadA7}$bs sequence is in blue, with the cleavage site in purple. The presumed IntI1 binding sites are boxed. (**B**) Amino acid interactions of IntI1 and VchIntIA with the cleavage site, based on the crystal structure of the VchIntIA synaptic complex. The amino acids of IntI1 are written in black and the corresponding aa of VchIntIA is in grey in parenthesis. 'A' means that the amino acid is from the active subunit and 'sym' means that the interactions of this amino acid from active and the non-active subunits are equivalent (8).

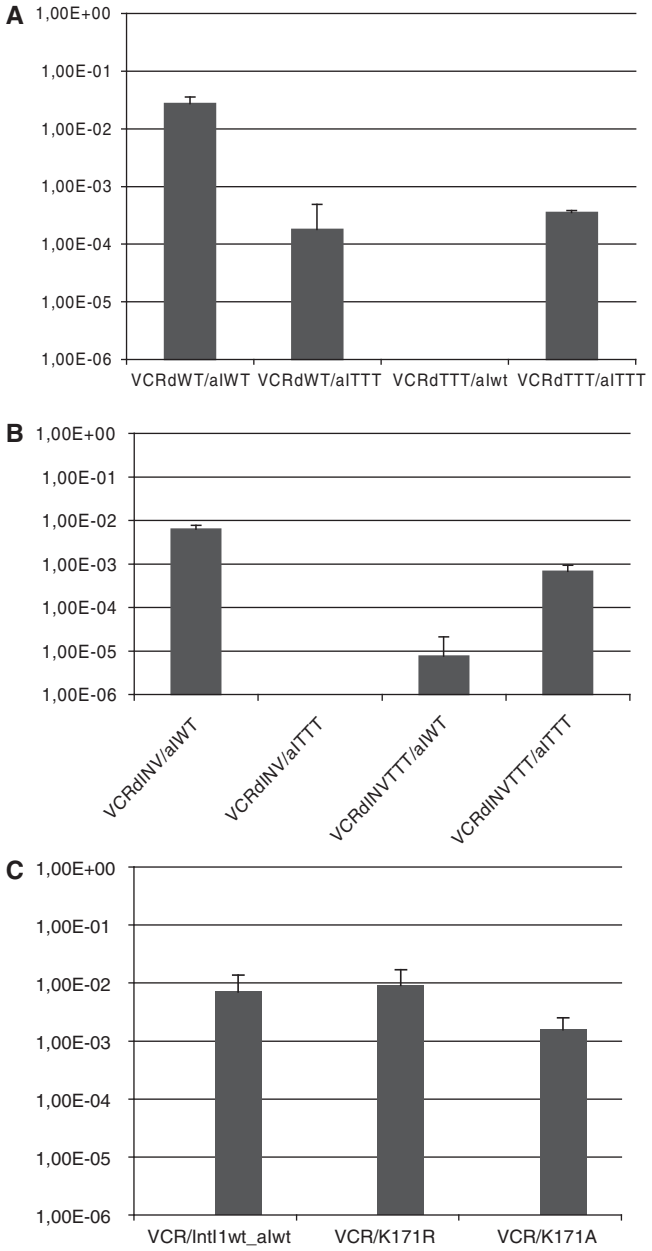

Figure 3. (**A**) Integration assay frequencies of recombination of $attC_{VCRd}$ mutant sites with $attI1_{wt}$ and $attI1_{TTT}$, by the mutant $IntI1_{P109L}$. (**B**) Integration assay frequencies of recombination of $attC_{VCRdINV}$ mutant sites with $attI1_{wt}$ and $attI1_{TTT}$, by the mutant $IntI1_{P109L}$. (**C**) Integration frequency of $attC_{VCR}$ and $attI1_{wt}$ by the integrase mutants $IntI1_{K171R}$ and $IntI1_{K171A}$.

$attC_{aadA7}$ will be liberated. Docking at the wrong position by the invading strand would prevent ligation of the two DNA strands and therefore inhibit the resolution of the Holliday junction. We hypothesize that the cleavage site $G^1 \rightarrow T^1$ is poor because of the additional possibilities for docking. To verify this, we inversed the bulge of a modified $attC_{VCR}$ substrate and used the mutant $IntI1_{P109L}$ for recombination. The $IntI1_{P109L}$ mutant has been shown to recombine $attC_{VCR}$ mutant substrates with an inversed bulge, $attC_{VCRINV}$, efficiently, although not as well as it recombines the wt substrate (15). The inversion has a drastic effect on recombination with the $IntI1_{wt}$ protein (9,15). Using the modified $attC_{VCR}$ substrate, $attC_{VCRd}$ (Figure 1) (9,15), we were able to show that the possibility of docking at many places most likely plays a role in the decrease in activity of the $G^1 \rightarrow T^1$ substrate. Indeed, $G^1 \rightarrow T^1$ mutation in both the $attC_{VCRd}$ and *attI1* sites ($attC_{VCRdTTT}/attI1_{TTT}$, Figure 3A) leads to a 100-fold decrease of the recombination compared to the frequency of the $attC_{VCRd}/attI1_{wt}$ substrates. On the other hand when using the substrate with the inversed bulge, $attC_{VCRdINV}$ (Figure 1), we found that introduction of the same $G^1 \rightarrow T^1$ mutation in the two partner site ($attC_{VCRdINVTTT}/attI1_{TTT}$, Figure 3B), the recombination frequency decreases by 10-fold only (compared with the frequency of the $attC_{VCRdINV}/attI1_{wt}$ substrates). In this case, the $attC_{VCRdINV}$ site will allow only one pairing possibility for the mutant T^1, on the invading strand, carrying mutation $A^{1'}$.

The second nt of the GTT triplet

The outcome of the mutations in T^2 is interesting for several reasons. In contrast to mutations in G^1, the ($T^2 \rightarrow A^2$, $T^2 \rightarrow C^2$ or $T^2 \rightarrow G^2$) mutations with their complementary positions modified, do not follow the same

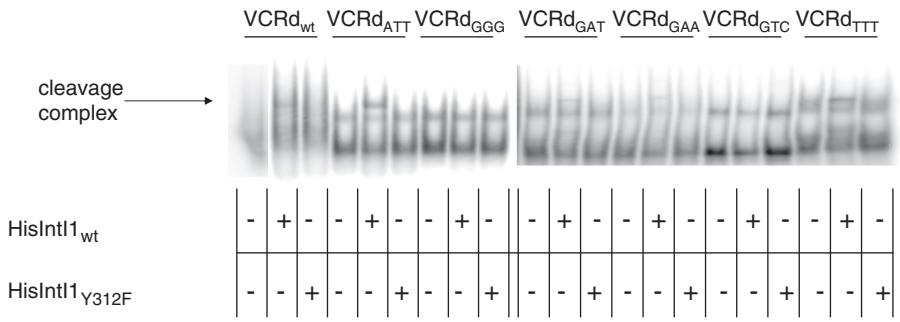

Figure 4. Cleavage assay of different $attC_{VCRd}$ site mutants by HisIntI1$_{wt}$. In the different VCR$_{dXXX}$ tested, the XXX nucleotides indicate the mutations made in the $attC_{VCRd}$ cleavage site, the wild-type being GTT. The cleavage defective mutant HisIntI1$_{Y312F}$ was used as a control. In the assay 30 pmol of HisIntI1$_{wt}$ or HisIntI1$_{Y312F}$ were used. Addition of either of the HisIntI1 proteins is indicated by a '+' sign in the corresponding row.

general pattern (Table 4). The few significant variations from the wt substrates are both increase and decrease in the recombination efficiency. In general, the second position of the $G^1T^2T^3$ triplet is quite tolerant to mutations in $attC \times attI1$ recombination. It is thus evident that $attC \times attI1$ recombination is not dependent on the identity of the second nt of the cleavage site between the two substrates. With the mutation $T^2 \rightarrow A^2$ we see an increase in recombination frequency, while the only large decrease in recombination frequency in this group is a 30-fold drop in recombination between $attC_{aadA7GCT} \times attI1_{GCT}$. The recombination frequency of $attC_{aadA7wt} \times attI1_{GCT}$ is unaffected, while $attC_{aadA7GCT} \times attI1_{wt}$ shows an ~6-fold decrease. We confirmed these results to be due to cleavage and not to a disruption of the $attC_{aadA7}$ site due to the mutation, by using the $attC_{VCRd}$ substrate carrying the same mutations, in the corresponding recombination assays with $attI1_{GCT}$ and $attI1_{wt}$ (Table 4).

Mutations of the third nt in the GTT triplet, or of the second and third nts at once indicate a higher sensitivity to G or C than A mutations

Mutations in T^3 are tolerable in $attC \times attI1$ recombination only when either the *attI1* or the *attC* site carries the mutation. When both of the sites are mutated we see a large decrease in recombination efficiency. Recombination with $attC_{aadA7GTA}/attI1_{GTA}$ is reduced 17-fold, while the same mutation in either *attC* or *attI1* by itself brings a recombination rate close to the wt one, or at most reduced by 2-fold. With the mutation $T^3 \rightarrow C^3$, we see a 150-fold decrease with $attC_{aadA7GTC}/attI1_{GTC}$, while either one of the mutations independently tested recombine at wt levels. This result was confirmed using an $attC_{VCRdGTC}$ substrate with $attI1_{GTC}$ and $attI1_{wt}$, which generated similar results (Table 4).

The same pattern, but even more drastic, can be seen when both T^2 and T^3 are mutated at once. $T^2T^3 \rightarrow A^2A^3$ drops 700-fold when the mutation is present both in the $attC_{aadA7}$ and the *attI1* site, from close to wt levels when only one site is carrying the mutation. The recombination frequency with the mutation $T^2T^3 \rightarrow G^2G^3$ drops by eight when the mutation is carried in the *attI1* site, by 160 when the mutation is present in the $attC_{aadA7}$ site only, and by more than 10^4 when recombination is performed with $attC_{aadA7GGG}/attI1_{GGG}$.

The recombination frequencies obtained with mutations in T^3 or in T^2 and T^3 at once, follow the same general pattern, where IntI1 generally accepts one substrate with a mutation as recombination target. When both DNA targets are mutated, it has a dramatic impact on the recombination effect.

In vitro cleavage of $attC_{VCRd}$ sites with mutations in the cleavage site

We were interested to see if the results from the integration assay correlated with the proteins ability to recognize and cleave *attC* substrates carrying the corresponding mutations. Synthetic oligonucleotides of the $attC_{VCRd}$ site with mutations in the cleavage site were labeled with P^{32} and incubated with purified HisIntI1$_{wt}$ protein or a cleavage deficient mutant HisIntI1$_{Y312F}$ In the cleavage assay we look at a cleavage product from a corresponding $attC \times attC$ complex. As one can see from Figure 4, the substrates $attC_{VCRdGTC}$ and $attC_{VCRdGGG}$, which have a drastic impact on the recombination frequency are not detectably cleaved *in vitro* by the wt protein. With the substrate $attC_{VCRdGAA}$, there is some *in vitro* cleavage, even though this mutation in the $attC_{aadA7}$ site shows a significant decrease in $attC \times attI$ recombination frequency (Table 4).

Mutations of the IntI1 residues interacting with the cleavage site and their effects on recombination

The results from the mutations in the third nt could be explained by the specific interactions of the aa K171$_{IntI1}$ from the attacking subunits with $A^{2'}$ and T^3 (Figure 2B). We therefore assayed the mutants IntI1$_{K171R}$ and IntI1$_{K171A}$, in our $attC \times attI1$ recombination assay and found a 5-fold decrease in the recombination frequency with the mutant IntI1$_{K171A}$, compared to IntI1$_{wt}$, and no significant variation for IntI1$_{K171R}$ (Figure 3C).

In order to establish the importance of the non-specific interactions with the GTT triplet, mutations in the conserved cleavage-site interacting aa between IntI1 and VchIntIA, were made. H303$_{IntI1}$ (H293$_{VchIA}$) which is a

member of the conserved HRHY motif of tyrosine recombinases (18), or $Gly172_{IntI1}$ ($Gly161_{VchIntIA}$) and $Thr33_{IntI1}$ ($Thr22_{VchIntIA}$), which are more than 3.5 Å away from the phosphate backbone (8) were not investigated. Instead we looked at the mutations $Y37F_{IntI1}$ ($Y26_{VchIntIA}$), $H281I_{IntI1}$ ($H271_{VchIA}$) and $Y29F_{IntI1}$ ($Y26_{VchIntIA}$), residues that are conserved between IntI1 and VchIntIA but not in other tyrosine recombinases. $IntI1_{Y29F}$ had a recombination efficiency identical to the wt IntI1. However, we found a drastic decline of the recombination frequency, more than 10^5-fold, for the $IntI1_{Y37F}$ and $IntI1_{H281I}$ mutants (data not shown). We were particularly interested in the $IntI1_{Y37F}$ mutant. The superposition of the Int11 model (15), shows that the $Y37F_{IntI1}$ ($Y26_{VchIntIA}$), is in close proximity to the GTT strand of the substrate (<5.0 Å). We decided to investigate the impact that this residue could have on the reaction, considering that the active site tyrosine ($Y312_{IntI1}$) is positioned in a similar fashion with respect to the phosphate backbone but close to the cleaved C|AA strand of the DNA, and that for a successful recombination reaction event to occur, there is a shared participation in the nature of the nts, and the surrounding protein environment. The region of cleavage is close to the unpaired segment of the substrate (the so-called central bubble) and therefore suffers from non-stacking disorder (Figure 1). After cleavage and departure of the |OH–C exchanged strand, one hypothesis would be that the $Y37F_{IntI1}$ ($Y26_{VchIntIA}$) hydroxyl could help stabilizing this part of the DNA (the GTT) through H-bond interaction with its backbone phosphate groups and thus preventing it from having alternative positions or trajectories due to flexibility (8). In support of this it was observed that the cleaved, assembled suicide substrate version of the VCRbs substrate does not crystallize with VchIntIA, whilst the LoxS–Cre complex does (19), in the latter this central region is fully base paired. This stability of the GTT is key since an incoming OH–C* strand from the facing partner DNA would be arriving to read-out base matching and ligate to form the Holliday junction intermediate. As the histidine-tagged $IntI1_{wt}$ had been shown to be recombinogenic (15), we created a his-tagged derivative of mutant $IntI1_{Y37F}$ ($HisIntI1_{Y37F}$, and purified it. This protein was found to have a poor DNA binding (Figure 5). However, in EMSA experiments with mixed portions of $HisIntI1_{wt}$ and $HisIntI1_{Y37F}$, the mutant and the wt protein were found to bind cooperatively to DNA (Figure 5).

It is possible that the mutation $Y37F_{IntI1}$ leads to a looser interaction with, or a poorer recognition of the DNA which destabilizes the synaptic complex. The Y37 residue is positioned quite close (<5 Å) to the GTT in the Int1 Model (15) with the possibility of interacting through H-bonds with the phosphate backbone, or ribosyl moiety of the G. Its replacement with a phenylalanine eliminates this possibility, therefore explaining the resulting drop in recombination frequency by a factor of 10^5. Addition of the wt protein compensates for this loss of stabilizing interaction, if a shared complex is formed on the DNA. In this hypothesis, the wt partner on the GTT side will provide the specificity through H-bonds for initial binding, whilst the Y37 mutant once recruited in the *cis* side complex, goes on to form the known c-termini alpha-helix swap interaction in *trans*, as seen in the case of Cre and other recombinases, to generate the tetrameric synaptic complex. This increased stabilization prevents the wt from getting released, thus showing an overall increased binding.

DISCUSSION

Our goal was to understand what governs the absolute conservation of the $G^1T^2T^3$ triplets observed in the integron recombination sites, as all integrons characterized so far show the same conserved sites. We have shown that the $G^1T^2T^3$ identity requirement between the *attI* and *attC* cleavage sites is not directed by an absolute requirement to obtain a successful recombination event (except for G^1), or homology readout, since the cleavage sites are quite far away from each other in the synapse (Figure 2A). The mutations in the first nt pair $G^1/C^{3'}$ support our hypothesis that there is only one nt responsible for the docking during the strand exchange event. In the integrative reaction of integrons, the linker regions, that are the regions between the two integrase binding sites of the *attC* and *attI* sites have no identity, in contrast to other tyrosine recombinase systems, such as for example Cre (20) (Figure 2A).

The results obtained in the experiments where the second nt is mutated are extremely interesting, and in particular the observed increase in recombination frequency with the $T^2 \rightarrow A^2$ mutations (Table 4). It has been reported that the sequence GAT is a preferred secondary integration site over a GTT, in the absence of *attI1* (21). In addition, in their *attI1* study, Hansson and collaborators found that this $T^2 \rightarrow A^2$ substitution in one of the cleavage sites was the mutation that affected the recombination frequency to the least, (40% of the wt frequency) (6). The high tolerance for mutations in T^2 in $attC_{aadA7} \times attI1$ recombination contrasts with the effect of the same mutations in one of the partner substrates in $attI1 \times attI1$ recombination, implying different recombination determinants for the two reactions (6).

There seem to be several explanations for the results obtained with mutations in the third or in the second and third positions at once, of the cleavage site, $T^3/A^{1'}$ or $T^2T^3/A^{1'}A^{2'}$. The decrease in recombination frequency is 10–100-fold greater when Ts were mutated to G or C rather than to A. The mutations $T^3 \rightarrow A^3$ or $T^2T^3 \rightarrow A^2A^3$ should disturb neither the pairing of the DNA, nor its interactions with the catalytic subunits. The AT stacked structure differs somewhat from mixed AT–GC regions and could contribute to the reduction in recombination frequency (22). The AT rich tracks indeed have a propensity to have lower step sizes. They typically lack the amino exocyclic groups that exist in Guanine which through steric clashes, promote higher step sizes. Additionally, AT tracks exhibit a higher propeller twist, thus allowing these regions to have altered curvatures (23). Distorted geometry in one of the two partner sites appears to be tolerated but only on one of the two recombination

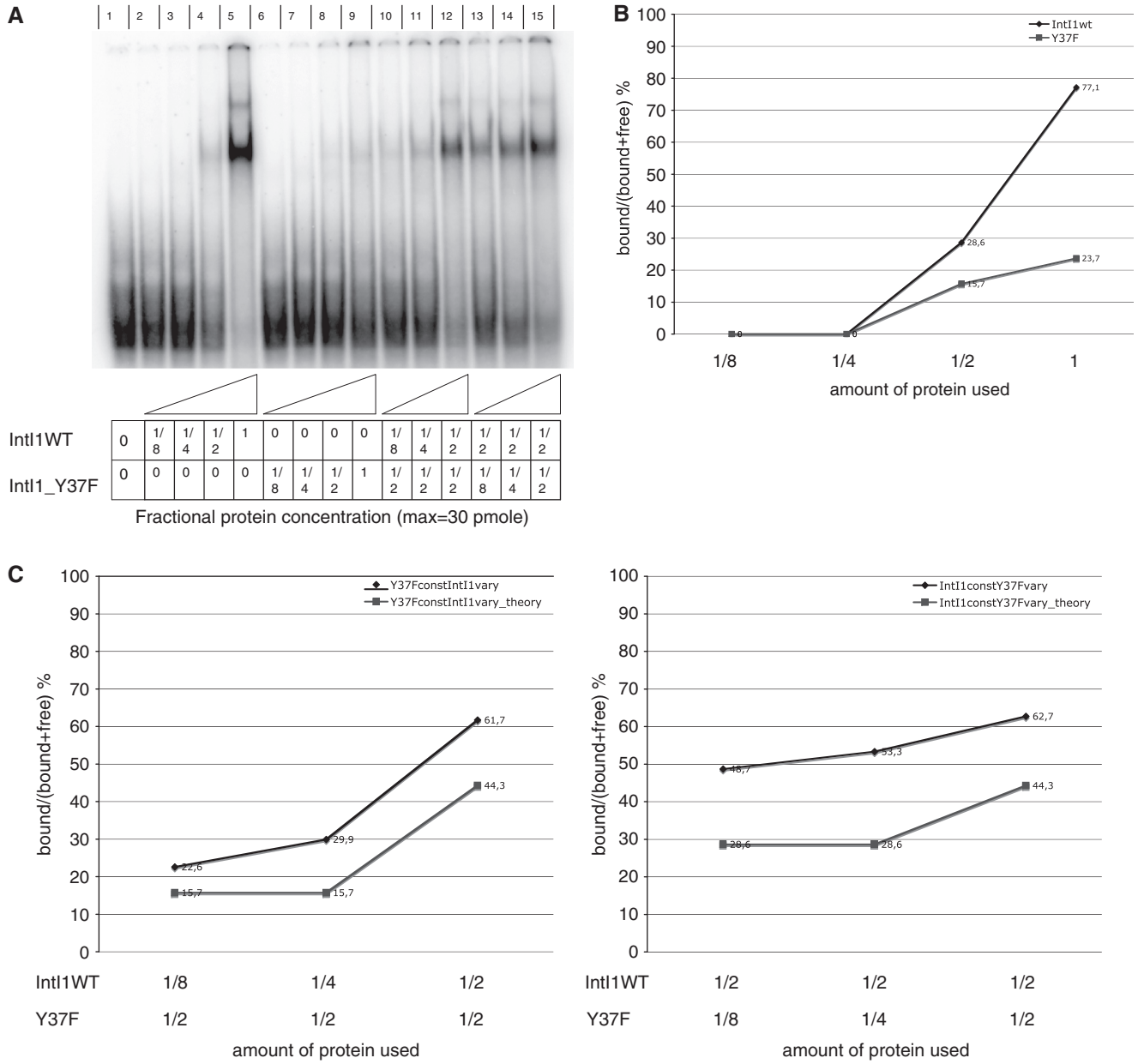

Figure 5. (**A**) EMSA experiment with $HisIntI1_{wt}$ and $HisIntI1_{Y37F}$ with the $attC_{VCRdwt}$ substrate. The amount of protein used is indicated in the picture. One-eighth corresponds to 3.75 pmol, 1/4 to 7.5 pmol, 1/2 to 15 pmol and 1 corresponds to 30 pmol. In lane 10–12, $HisIntI1_{wt}$ is kept constant at 15 pmol and an increasing amount of $HisIntI1_{Y37F}$ is added, in lane 13–15 $HisIntI1_{Y37F}$ is kept constant and $HisIntI1_{wt}$ added in an increasing amount as indicated in the picture. (**B**) Graph showing the quantification of bound/(bound + free) product of $HisIntI_{wt}$ (lanes 2–5) and $HisIntI1_{Y37F}$ (lanes 6–9). (**C**) Graphs showing the quantification of the bound/(bound + free) product of constant $HisIntI1_{wt}$ and variable $HisIntI1_{Y37F}$ or constant $HisIntI1_{Y37F}$ and variable $HisIntI1_{wt}$. The curves are plotted together with the expected percent binding of the proteins, calculated by adding the binding of the corresponding concentration of the free proteins in lanes 2–9, if the binding of one protein was independent of the presence of the other.

substrates. The curvature imposed on the purine-rich side of the two strands and the increased propeller twist of the AT base pairs may significantly modify the ability of the integrase to recognize and bind to the DNA in this region.

The aa interacting with the cleavage site are conserved between VchIntIA and IntI1, with two exceptions, $S30_{IntI1}$–$A19_{VchIntIA}$ and $R112_{IntI1}$–$K101_{VchIntIA}$ (Figure 2B). The protein–DNA interface is mainly due to non-specific interactions between the protein and the DNA backbone. There is only a single specific interaction: $K171_{IntI1}$ ($K160_{VchIntIA}$), which interacts specifically with nts T^{3} and $A^{2'}$ flanking the scissile phosphate (8). The results from the mutations in the third nt could partly be explained by the specific interactions of the aa K171 from the attacking subunits. In the VchIntIA–VCRbs substrate crystal structure (8) this residue is seen to interact with the A14′ and T30 bases. The corresponding nts in the IntI1 substrate are A2′ and T3. The interaction is based on hydrogen bonds, with a 2.9 Å distance between the side-chain of K171 (Nz atom) with the A2′

N3 atom, and by a similar interaction (3.4 Å) between the same Nz atom and the O_2 atom of the T3 base. Mutations in the third nt could abolish these bonds. We assayed the mutants $IntI1_{K171R}$ and $IntI1_{K171A}$, in our *attC* × *attI1* recombination assay and found a 5-fold decrease in the recombination frequency with the mutant $IntI1_{K171A}$, compared to $IntI1_{wt}$, and no significant variation for $IntI1_{K171R}$ (Figure 3C). This suggests that the NZ-N3 bound is conserved in IntI1-*attC* interaction and plays a role in the recombination. When both $attC_{aadA7}$ and *attI1* are mutated there will be no specific interactions between IntI1 and the cleavage site. This also explains why the recombination frequency is lower when both T^2 and T^3 are mutated, since K171 interacts with both the second position of the cleaved strand, $A^{2'}$ and with T^3 on the opposite strand. Although the base at the $A^{2'}$-T^3 pair is not involved in the strand exchange, the reversal of the curvature due to the introduction of purines on the non-exchanging strand seems to play a critical role in reducing recombination frequency. This was supported by *in vitro* cleavage experiments with $attC_{VCRd}$ probes carrying some of the mutations from the recombination experiments. Mutations $G^1 \rightarrow A^1$, $G^1 \rightarrow T^1$, $T^2 \rightarrow A^2$ and $T^2\ T^3 \rightarrow A^2A^3$, were all cleaved, while no detectable cleavage product was generated with the mutations $T^3 \rightarrow C^3$ and $T^2\ T^3 \rightarrow G^2\ G^3$ (Figure 4). IntI1 synaptic complex modeling suggests more specific interactions with G^1 and T^3 than for VchIntIA (24). This would further explain why mutations are more drastic with T^3 in our experiments, than in T^2. For G^1, however, these interactions do not seem to be essential for recombination.

The investigations of the conserved aa Y37 and H281, which are involved in interactions with the cleavage-site, indicate the importance of the non-specific interactions with this site. The mutant $HisIntI1_{Y37F}$ was found to bind DNA poorly. The equivalent residue in VchIntIA, Y26, is part of a hydrophobic core composed of (F12, Y18 and F77, VchIntIA coordinates). It seems likely that the protein–DNA interactions in the cleavage site help orient the DNA in the correct position for cleavage and recombination. It is possible that the decrease in recombination frequency with the $IntI1_{Y37F}$ mutant is an effect of an unstable protein–DNA complex.

The conservation of an efficient recombination level when either the $attC_{aadA7wt}$ site or *attI1* is mutated indicates that T^2T^3 in the $G^1T^2T^3$ site was optimized for cleavage by the IntI1 protein rather than for homology requirements. Our results suggest that the $G^1T^2T^3$ conservation is likely driven by the aptitude to exchange cassettes between the different integrons, more than by mechanistic grounds. We demonstrate that the recombination system of integrons class 1 has more relaxed requirements than previously thought. Regardless of the substitution, one mutation in the second or third nt of either *attC* or *attI* was highly tolerated by the integrase protein, and a recombination frequency compared to that of the wt was obtained. In addition these studies, by showing that the first nt of the conserved triplet can be substituted without harm, open the way for the utilization of the integron recombination machinery in synthetic biology. This shows that recombination of sequence 'à la carte' is certainly an objective that can be reached starting from the integron system. Indeed, we had previously established which of the secondary structure features are directing the recombined strand recognition (9). Now we show that to accomplish a successful recombination, in contrast to other known tyrosine recombination systems, the conserved triplet of the Recombinase binding element sequence is highly tolerant for substitutions and that the only absolute requirement is identity of the first nt in the two partner sites.

ACKNOWLEDGEMENTS

The authors thank Céline Loot for providing the purified $IntI1_{Y312F}$ protein.

FUNDING

Institut Pasteur, the Centre National de la Recherche Scientifique, the French National Research Agency (ANR-08-MIE-016); EU (STREP CRAB, LSHM-CT-2005-019023, and NoE EuroPathoGenomics, LSHB-CT-2005-512061); Fondation pour la Recherche Médicale (programme 'équipe FRM' 2007). Funding for open access charge: 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. Boucher,Y., Labbate,M., Koenig,J.E. and Stokes,H.W. (2007) Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends Microbiol.*, **15**, 301–309.
3. Guerin,E., Cambray,G., Sanchez-Alberola,N., Campoy,S., Erill,I., Da Re,S., Gonzalez-Zorn,B., Barbe,J., Ploy,M.C. and Mazel,D. (2009) The SOS response controls integron recombination. *Science*, **324**, 1034.
4. Cameron,F.H., Groot Obbink,D.J., Ackerman,V.P. and Hall,R.M. (1986) Nucleotide sequence of the AAD(2″) aminoglycoside adenylyltransferase determinant aadB. Evolutionary relationship of this region with those surrounding aadA in R538-1 and dhfrII in R388. *Nucleic Acids Res.*, **14**, 8625–8635.
5. Collis,C.M., Kim,M.J., Stokes,H.W. and Hall,R.M. (2002) Integron-encoded IntI integrases preferentially recognize the adjacent cognate attI site in recombination with a 59-be site. *Mol. Microbiol.*, **46**, 1415–1427.
6. Hansson,K., Skold,O. and Sundstrom,L. (1997) Non-palindromic attl sites of integrons are capable of site-specific recombination with one another and with secondary targets. *Mol. Microbiol.*, **26**, 441–453.
7. 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.
8. 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.
9. Bouvier,M., Ducos-Galand,M., Loot,C., Bikard,D. and Mazel,D. (2009) Structural features of single-stranded integron cassette *attC* sites and their role in strand selection. *PLoS Genet.*, **5**, e1000632.
10. 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.
11. Hansson,K., Sundstrom,L., Pelletier,A. and Roy,P.H. (2002) IntI2 integron integrase in Tn7. *J. Bacteriol.*, **184**, 1712–1721.

12. Hallet,B. and Sherratt,D.J. (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.*, **21**, 157–178.
13. Nunes-Duby,S.E., Yu,D. and Landy,A. (1997) Sensing homology at the strand-swapping step in lambda excisive recombination. *J. Mol. Biol.*, **272**, 493–508.
14. Stokes,H.W., O'Gorman,D.B., Recchia,G.D., Parsekhian,M. and Hall,R.M. (1997) Structure and function of 59-base element recombination sites associated with mobile gene cassettes. *Mol. Microbiol.*, **26**, 731–745.
15. Demarre,G., Frumerie,C., Gopaul,D.N. and Mazel,D. (2007) Identification of key structural determinants of the IntI1 integron integrase that influence attC × attI1 recombination efficiency. *Nucleic Acids Res.*, **35**, 6475–6489.
16. Rajeev,L., Malanowska,K. and Gardner,J.F. (2009) Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol. Mol. Biol. Rev.*, **73**, 300–309.
17. Lewin,B. (2000) In Lewin,B. (ed.), *Genes VII*. Oxford University Press Inc., NY, pp. 118–148.
18. 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.
19. Guo,F., Gopaul,D.N. and Van Duyne,G.D. (1999) Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Natl Acad. Sci. USA*, **96**, 7143–7148.
20. Van Duyne,G.D. (2001) A structural view of cre-loxp site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 87–104.
21. Francia,M.V., de la Cruz,F. and Garcia Lobo,J.M. (1993) Secondary-sites for integration mediated by the Tn21 integrase. *Mol. Microbiol.*, **10**, 823–828.
22. MacDonald,D., Herbert,K., Zhang,X., Pologruto,T. and Lu,P. (2001) Solution structure of an A-tract DNA bend. *J. Mol. Biol.*, **306**, 1081–1098.
23. Rice,P.A. and Correll,C.C. (2008) *Protein-Nucleic Acid Interactions: Structural Biology*. Thomas Graham House. Science Park, Cambridge CB4 OWF, UK.
24. Johansson,C., Boukharta,L., Eriksson,J., Aqvist,J. and Sundstrom,L. (2009) Mutagenesis and Homology Modeling of the Tn21 Integron Integrase IntI1 (dagger). *Biochemistry*, **48**, 1743–1753.
25. Martinez,E. and de la Cruz,F. (1990) Genetic elements involved in Tn*21* site-specific integration, a novel mechanism for the dissemination of antibiotic resistance genes. *EMBO J.*, **9**, 1275–1281.
26. 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 (IncPα) conjugative machineries and their cognate *E. coli* host strains. *Res. Microbiol.*, **156**, 245–255.
27. 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.
28. 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-resitant integrons. *Proc. Natl Acad. Sci. USA*, **98**, 652–657.