



HAL
open science

Reprogrammation de la spécificité de recombinaison d'une intégrase phagique

Kevin Debatisse

► **To cite this version:**

Kevin Debatisse. Reprogrammation de la spécificité de recombinaison d'une intégrase phagique. Biochimie [q-bio.BM]. INSA de Toulouse, 2022. Français. NNT : 2022ISAT0009 . tel-04213583

HAL Id: tel-04213583

<https://theses.hal.science/tel-04213583>

Submitted on 21 Sep 2023

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

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



THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Institut National des Sciences Appliquées de
Toulouse

Présentée et soutenue par
Kevin DEBATISSE

Le 27 janvier 2022

**Reprogrammation de la spécificité de recombinaison d'une
intégrase phagique**

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et
Bioingenieries**

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :

TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering

Thèse dirigée par
Pascal LE BOURGEOIS

Jury

M. François-Xavier BARRE, Rapporteur
Mme Céline LOOT, Rapporteur
Mme Mireille ANSALDI, Rapporteur
M. Pascal LE BOURGEOIS, Directeur de thèse
M. François CORNET, Président

MERCI !

Je souhaite tout d'abord remercier Céline Loot, François-Xavier Barre et Mireille Ansaldi d'avoir accepté d'évaluer ces travaux de thèse et François Cornet d'avoir accepté d'être président du jury.

Je remercie également Claire Le Marrec et Philippe Rousseau d'avoir fait partie de mon comité de thèse, et pour leur bienveillance ainsi que pour les différents conseils qui ont été bénéfique au projet et qui m'ont permis de grandir scientifiquement. Je tiens particulièrement à te remercier Philippe, pour les conseils dans et hors du comité de thèse, notamment pour ton aide et tes nombreuses astuces pour le retard sur gel ! Cela m'a grandement aidé et sans ton aide, je suis sûr que les gels ne seraient pas si jolis !

Je tiens particulièrement à remercier mon encadrant de thèse Pascal Le Bourgeois. Voilà quasiment 4 ans depuis mon arrivée en stage de M2 que l'on travaille ensemble. Je tiens vraiment à te remercier pour m'avoir formé et soutenu pendant toutes ces années ! J'ai vraiment beaucoup appris à tes côtés, que ce soit scientifiquement mais aussi humainement. Je te remercie aussi pour ta confiance, et pour tout tes conseils, dès qu'il y avait un souci, je savais que je pouvais compter sur toi. Je trouve que l'on forme un bon duo ! Je te suis vraiment extrêmement reconnaissant de m'avoir proposé cette aventure qu'est une thèse en 2018 et d'avoir réussi à me passionner pour cette protéine mais aussi pour la recombinaison spécifique de site !

Ensuite je souhaite remercier notre chef d'équipe Muriel Cocaign-Bousquet. Pour moi tu as toutes les qualités nécessaires lorsque l'on gère une équipe. Tu mets énormément d'importance sur l'humain, et sur l'ambiance, et cela se ressent lorsque l'on travaille dans cette équipe. Je te dis donc un grand MERCI pour la chance que j'ai eu d'avoir pu travailler et m'épanouir dans cet environnement. Je remercie donc grandement tous les membres passés et présents de l'EAD4 qui participent à l'ambiance fantastique de cette équipe. Je vous remercie pour votre disponibilité, vos conseils et bien entendu votre bonne humeur et convivialité ! Que ce soit des pauses-café, aux moments conviviaux (escape-game, laser-game, nettoyage du laboratoire.... Oui à l'EAD4 un nettoyage du laboratoire peut être un moment sympathique... !).

Vous avez fait en sorte que le quotidien soit toujours très facile à vivre, même dans les périodes moins marrantes, comme le confinement avec notre petit rendez-vous du lundi matin sur Zoom ! ou bien mon « télé-anniversaire » sur Zoom lors du premier confinement.

Seb, je te remercie pour tout. Tu es vraiment quelqu'un de bienveillant, et c'est un vrai plaisir d'avoir pu travailler à tes côtés. J'ai adoré les moments passés à la paillasse (Je ne sais pas si tu te souviens encore de ce que c'est qu'une paillasse !), que ce soit à la MRV, ou bien après le déménagement à TBI, on a pu rester co-paillasse ! Je me souviens de ces moments au laboratoire où nous étions tous les 4 avec Marjorie et Manon. Je te remercie également pour ta disponibilité, dès que j'avais une question, tu étais toujours là pour y répondre. Durant ces 3 années de thèse, j'ai passé un sacré bout de temps au labo, et c'est toujours plus facile quand on sait que l'on pourra y passer un bon moment, même si les résultats ne sont pas toujours là ! Brice, je te remercie aussi pour ces pauses café, ces blagues (que je ne suis pas sûr de pouvoir écrire dans mes remerciements !). Tu participes à cette bonne ambiance à l'EAD4 et je t'en suis vraiment reconnaissant ! Merci Mamie-Line ! Tu es vraiment une personne extraordinaire. Toujours disponible, bienveillante, ça a été un plaisir de partager ces 3 années de thèse à tes côtés ! Merci aussi pour tous les conseils et de m'avoir permis d'enseigner. Tu es vraiment une enseignante impliquée ! (Je regrette de n'avoir jamais pu être ton élève !). Merci Laurence, pour ces pauses du midi, pour les conseils lors des réunions d'équipe et pour la bienveillance. Je remercie aussi Marie-Pierre, sans toi toutes les thèses de l'équipe seraient beaucoup plus compliquées ! Merci pour tout ce travail que tu fournis à l'équipe, tu es un peu la « maman du labo » de tous les nouveaux arrivants dans l'équipe. Je te remercie pour ta gentillesse, ta bienveillance et pour toutes les pauses du midi. Cela a été un véritable plaisir de travailler à tes côtés ! Je te remercie un peu moins pour les danses de la victoire lorsque Toulouse gagnait contre Clermont lors des phases finales. Je n'ai malheureusement pas eu la chance de te rendre ces danses ! Je remercie aussi Sandrine, pour ton implication à certains moments du projet, mais aussi pour ta bonne humeur. C'était vraiment un plaisir de partager des repas à tes côtés, et tu fais partie des rares personnes à avoir pratiqué le télétravail avant que cela ne devienne à la mode.. ! Je remercie aussi Valérie, pour ta disponibilité et ta bienveillance ! Je n'ai malheureusement pas eu la chance de travailler à tes côtés, j'ai réussi à éviter tout ce qui est HPLC et fermentation... du moins pour le moment ! Je souhaite aussi remercier Adilya, ça a été un vrai plaisir de travailler à tes côtés, tu es vraiment quelqu'un de bienveillante, tu nous manques à l'EAD 4 ! Enfin, merci Nathalie, la petite dernière à nous avoir rejoint lors de ma dernière année de thèse !

Je souhaite aussi remercier les non-permanents (pour ne pas dire « jeunes » ! et vexer les personnes du paragraphe précédent !), passés ou encore présents dans l'équipe. Merci Huong ! je suis arrivé en stage quand tu étais en 3^{ème} année de thèse, tu as pu me donner de nombreux conseils, que je n'ai malheureusement pas tous écouté ! Mais que j'ai pu transmettre à la nouvelle génération... ! Je me souviens quand tu étais en train de rédiger ton manuscrit et que je t'embêtais avec la partie remerciement... C'est mon tour.. le temps passe vite ! Merci Manon, ma co-bureau préférée, on a pu passer toute notre thèse ensemble, notre duo a survécu au déménagement ! Je te remercie de m'avoir accueilli dans cette équipe, c'est difficile de citer un moment particulier tellement il y en a eu ! Merci pour tous ces bons moments au labo, mais surtout (ou surtout) pour tous les bons moments hors-labo, que ce soit les nocturnes du marché, les soirées dans les apparts, nos vacances d'été avec Marjo.. ! Merci d'être devenue une véritable amie. MAC attends son kéké... Merci Marie pour m'avoir aussi accueilli dans l'équipe quand je suis arrivé en stage. Merci pour cette soirée madeleine... qui a forgé le début d'une très belle amitié ! Merci pour ces moments au laboratoire.. (à écouter du fatal bazooka sous le PSM), et aux moments hors-labo. Tu es aussi devenue une véritable amie. Merci ma DJ, j'ai eu un véritable coup de cœur amical lorsque tu es arrivée au labo, tu es vraiment une fille géniale, compétente, toujours de bonne humeur. Tu n'es restée que 6 mois dans l'équipe, mais tu es aussi devenue une véritable amie, j'attends toujours que tu viennes au laboratoire avec ton petit gadget sur le vélo ! Merci Marjo pour tous ces bons moments au laboratoire, que ce soit à la paillasse, ou bien dans les bureaux. Merci pour toutes ces pause-café, et d'avoir réussi à me supporter pendant la phase de rédaction (et aussi pendant la thèse en fait !) ! Tu es vraiment quelqu'un de géniale, et je te remercie aussi pour tous ces moments hors-labo, que ce soit des soirées raclette (on fait la fête entre amis !), des moments au bar, à nos vacances d'été avec Manon. Merci toi aussi d'être devenue une véritable amie ! Merci Mathilde, on a pu passer 1 an dans le même bureau et tu es vraiment quelqu'un d'adorable (et je ne dis pas ça parce que tu étais la seule avec moi dans ce bureau !). On a passé des moments sympas au labo et au bureau, et cela m'a permis de passer ma dernière année de thèse dans une très bonne ambiance ! Je te souhaite le meilleur de ton côté. Je remercie aussi Théo et Antoine, je garde de très bons souvenirs de cette période, que ce soit lors du déménagement du laboratoire avec Théo (ou plutôt inspecteur Pigaglio !), ou bien les tacos XXL et la chasse de Pokémon avec Antoine ! Je remercie également Jade, arrivée lors de l'année covid ! Ça a été un vrai plaisir de te connaître. Je vous souhaite à toi et Antoine un bon courage pour votre thèse même si je ne me fais pas trop de soucis pour vous ! Je remercie aussi Maryse et Romain que j'ai eu la chance d'encadrer, vous avez été super !

Enfin, merci Pedro et Charlotte, Pedro je te souhaite bon courage pour les 2 années qu'il te reste, ça va le faire ! Charlotte bon courage pour la rédaction.. ! Une période pas forcément facile, mais très enrichissante ! Finally, Fan, thanks for these 3 years ! We almost started our thesis at the same time, thanks for all of these good times in the lab, and I wish you the best for the writing part !

Il est difficile de remercier tout le monde sans oublier quelqu'un, je remercie donc chaleureusement toute les personnes du laboratoire avec qui j'ai pu avoir la chance d'interagir, dans et hors du labo (les fameux « jeudi dubli »). Enfin, je souhaite remercier Michou, Vinc' et Axel, trois copains depuis le lycée ! On a tous choisi des orientations différentes, mais on finit toujours par se retrouver. Je vous remercie pour votre soutien ! Particulièrement toi Vinc', tu as toujours sur être là, dans les bons et les mauvais moments. En espérant vous revoir rapidement tous les 3 !

Pour finir, merci Maman pour ton soutien, mais aussi pour tout ce que tu as pu faire pour moi. C'est grâce à toi si j'en suis arrivé là. Je pense que tous les enfants doivent dire ça de leur mère, mais t'es une maman en or, même si je ne te le dis pas assez ! Merci d'être venu me voir sur Toulouse, même si généralement, dès que tu viens, la France est confinée juste après ! Merci à mon frère, Lucas, qui sera toujours là pour moi. Cela m'a fait plaisir de te faire visiter Toulouse et que Méli-Mélo découvre enfin les trains sous terre ! (Oui c'est le métro).

Ps1 : Je ne remercie pas eurofins

Ps2 : Je remercie IDT

Résumé

Le bactériophage mv4 est un virus bactérien tempéré capable d'intégrer par recombinaison spécifique de site son génome dans l'extrémité 3' d'un ARN de transfert (ARNt^{SER}) du chromosome de *Lactobacillus delbrueckii* ssp. *bulgaricus*. Cette recombinaison est médiée par l'enzyme ^{mv4}Int, une protéine appartenant à la sous-famille des intégrases hétérobivalentes des recombinases à tyrosine. Différents travaux avaient montré que le système ^{mv4}Int/*attP*/*attB* présentait des particularités par rapport aux systèmes classiques, tel que celui du bactériophage lambda d'*Escherichia coli* : i) la réaction de recombinaison ne nécessitait pas de facteurs d'hôte, ii) les sites de recombinaison phagiques (*attP*) et bactériens (*attB*) possédaient une architecture atypique, avec des régions cœur dépourvues de site de fixation de ^{mv4}Int, une taille de la région d'échange des brins (région « overlap », O) ainsi qu'une organisation des sites bras inhabituelles, et iii) un site *attB* d'une taille de 16 pb, le plus petit site décrit dans la littérature. Ces travaux de thèse ont eu pour objectif de caractériser précisément les régions cœur atypiques des sites de recombinaison *attB* et *attP*. Une étude exhaustive de ces sites a été réalisée par différents tests *in vitro* et a permis de réinterpréter leurs caractéristiques. Nos résultats montrent i) que le site *attB* possède une taille de 21 pb, ii) que les régions cœurs ont en réalité une organisation similaire aux systèmes classiques, avec la présence de régions répétées inversées (B et B' pour *attB*, C et C' pour *attP*) servant de sites de fixation de ^{mv4}Int et entourant une région d'échange de brins (O) de 7 pb, et iii) que le nombre et la localisation des sites bras de fixation (sites P) de l'intégrase ^{mv4}Int correspond également aux systèmes classiques avec quatre sites, deux sur le bras gauche (P1, P2) et deux sur le bras droit (P'1, P'2). En revanche, ^{mv4}Int se distingue des intégrases hétérobivalentes classiques par une plus grande tolérance vis-à-vis de la nature des nucléotides constituant ces régions cœur de 21 pb, permettant de définir une séquence nucléotidique consensus de chaque région B, B', C et C', ainsi que l'absence de contrainte dans la région d'échange de brins.

Abstract

The phage mv4 is a temperate bacterial virus able to integrate by site-specific recombination its genome into the 3' end of a transfer RNA (^{SER}tRNA) of the *Lactobacillus delbrueckii* ssp. *bulgaricus* chromosome. This recombination is mediated by ^{mv4}Int, a protein belonging to the heterobivalent integrase subfamily of tyrosine recombinases. Several studies had shown that the ^{mv4}Int/*attP*/*attB* system presented particularities compared to conventional systems such as the *Escherichia coli* lambda phage : i) the recombination reaction did not require host factors, ii) the phage (*attP*) and bacterial (*attB*) recombination sites had atypical architectures, with core regions lacking ^{mv4}Int binding sites, with a size of the strand exchange region (“overlap” region, O) and organization of arm-type sites unusual, and iii) an *attB* site with a size of 16 bp, the smallest site described in the literature. The purpose of this thesis was to precisely characterize the atypical core regions of the *attB* and *attP* recombination sites. An exhaustive study of these sites was carried out by different *in vitro* experiments and made it possible to reinterpret their characteristics. Our results show i) that the *attB* has a size of 21 bp, ii) that the core regions have an organization similar to the classical heterobivalent Tyr-recombinase systems, with the presence of inverted repeat regions (B and B' for *attB*, C and C' for *attP*) used as ^{mv4}Int core-binding sites and surrounding a 7-bp strand-exchange region (O), and iii) the number and location of ^{mv4}Int arm-type binding sites (P sites) also correspond to classical systems with four sites, two on the left arm (P1, P2) and two on the right arm (P'1, P'2). In contrast, ^{mv4}Int differs from classical heterobivalent integrases by a greater tolerance regarding the nature of the nucleotides corresponding to the 21-bp *attP* and *attB* core regions, making it possible to define a consensus nucleotidic sequences for each B, B', C and C' region, as well as lack of constraint into the overlap region.

Abréviations :

AA : Acide-aminé

Pour les acides aminés :

D : Aspartate

E : Glutamate

H : Histidine

K : Lysine

R : Arginine

W : Tryptophane

Y : Tyrosine

ADN : Acide désoxyribonucléique

A : Adénine

C : Cytosine

G : Guanine

T : Thymine

B : (C/G/T)

D : (A/G/T)

H : (A/C/T)

N : (A/C/G/T)

R : puRine (A/G)

V : (A/C/G)

W : Weak (A/T)

Y : pYrimidine (C/T)

ARNt : ARN de transfert

attB : Bacterial attachment site

attC : Cassette attachment site

attI : Integron attachment site

attL : Left attachment site

attP : Phage attachment site

attR : Right attachment site

BET : Bromure d'éthidium

CB : Core-Binding

CC : Coiled-Coil

CDS : Coding DNA Sequence

Cer : ColE1 Resolution
Cre : Cyclization recombination
C-ter : C-terminal
CTnDOT : Conjugative Transposon from *Bacteroides thetaiotaomicron* DOT
CTXΦ : Cholera Toxine phage
Cy3 : Cyanine 3
Dif : Deletion Induced Filamentation
DTT : Dithiothréitol
EDTA : EthylèneDiamineTétraAcétique
EHB : ExtraHelical Bases
Fis : Factor for Inversion Stimulation
Flp : Flipase
FRT : Flp Recognition Target
H-NS : Histone-like Nucleoid-Structuring
HTH : Hélice-tour-hélice
ICE : Integrative and Conjugative Element
IHF : Integration Host Factor
^x*Int* : intégrase de l'élément génétique mobile X
IPTG : Isopropyl β-D-1 ThioGalactopyranoside
KOPS : FtsK Oriented Polar Sequence
LB : Lysogenic Broth
LE/RE : Left Element / Right Element
loxP : locus of crossing over in P1
LTR : Long-Terminal Repeat
MI : Mobile Integron
NLS : Nuclear Localization Signal
N-ter : N-terminal
ORF : Open Reading Frame
PCR : Polymerase Chain Reaction
PEG : Polyéthylène glycol
Psi : pSC101 Stabilized Inheritance
Psi-Blast : Position-Specific Iterative Basic Local Alignment Search Tool
RD : Recombinase domain
RDF : Recombination Directionality Factor, Facteur directionnel de la recombinaison

RMCE : Recombination-Mediated Cassette Exchange

SCI : Sedentary Chromosomal Integron

SLiPE : Substrate-Linked directed Protein Evolution

T_m : Melting Temperature, Température de fusion

TLCΦ : Toxin-Linked Cryptic phage

VTS : Variable Terminal Structure

UCS : Unpaired Central Segment

XafT : Xer Activation Factor of TLC

Xer : Chromosomally Encoded Recombinase

ZD : Zinc Ribbon domain

Sommaire

Etude bibliographique..... 1

I) Recombinases à sérine..... 7

I.1) Rôles biologiques des recombinases à sérine	9
I.2) Domaines protéiques des recombinases à sérine	9
I.3) Mécanisme de recombinaison des recombinases à sérine	11
I.4) Les petites recombinases à sérine	13
I.4.1) Les résolvases à sérine	13
I.4.2) Les invertases à sérine	15
I.5) Les grandes recombinases à sérine	17

II) Recombinases à tyrosine..... 25

II.1) Rôles biologiques des recombinases à tyrosine.....	25
II.2) Organisation des sites de recombinaison des recombinases à tyrosine	27
II.3) Organisation des domaines protéiques des recombinases à tyrosine.....	27
II.4) Structure des recombinases à tyrosine.....	29
II.5) Mécanisme de recombinaison	33
II.5.1) Coupure de brin	33
II.5.2) La réaction d'échange de brin	35
II.6) Cre et Flp, modèles des recombinases à tyrosine simple.....	35
II.6.1) La recombinase Cre	35
II.6.2) La recombinase Flp	37
II.7) Le système XerC/D, modèle des recombinases utilisant des protéines accessoires	39
II.7.1) La recombinaison aux sites chromosomiques <i>dif</i>	41
II.7.2) La recombinaison aux sites plasmidiques <i>cer</i> et <i>psi</i> : l'utilisation d'un filtre topologique	43
II.7.3) Les éléments génétiques mobiles exploitant le système XerC/D	45
II.7.3.1) Les IMEXs de type CTX Φ	47
II.7.3.2) Les IMEXs de type VGJ Φ	49
II.7.3.3) Les IMEXs de type TLC Φ	49
II.8) Les intégrons : un mécanisme de recombinaison atypique	51
II.8.1) Le système de recombinaison des intégrons.....	53
II.8.1.1) Le site <i>attI</i>	53
II.8.1.2) Le site <i>attC</i>	55
II.8.1.3) L'intégrase IntI	59
II.8.2) Les réactions de recombinaison des intégrons	59
II.8.2.1) La recombinaison <i>attC/attI</i>	59
II.8.2.1) La recombinaison <i>attC/attC</i>	61
II.8.2.3) La recombinaison <i>attI/attI</i>	61

II.9) Le paradigme des recombinases hétérobivalentes : l'intégrase λ	63
II.9.1) Composants du système de recombinaison	63
II.9.1.1) Les sites de recombinaison.....	63
II.9.1.1.1) Les sites de fixation de λ Int.....	65
II.9.1.1.2) Les sites de fixation IHF.....	67
II.9.1.1.3) Les sites de fixation de l'excisionase Xis et de la protéine Fis.....	67
II.9.1.2) Les protéines impliquées dans la recombinaison	69
II.9.1.2.1) L'intégrase λ	69
II.9.1.2.2) IHF.....	73
II.9.1.2.3) L'excisionase Xis	75
II.9.1.2.4) Fis	77
II.9.2) La réaction de recombinaison.....	79
II.9.2.1) Le Complexe intégratif.....	83
II.9.2.2) Le Complexe d'excision.....	83
II.9.2.3) La mise en place de l'intasome : un phénomène conservé parmi les systèmes hétérobivalents.....	85
II.9.2.3.1) Les systèmes HP1, L5 et KplE1 : une mise en place de l'intasome différent du modèle λ	89
II.9.2.3.2) La formation de l'intasome : un mécanisme partagé avec les ICEs utilisant une intégrase hétérobivalente	95
II.10) Les recombinases spécifiques de site : de précieux outils d'ingénierie génétique	105
III) Le système de recombinaison mv4, un système de recombinaison hétérobivalent atypique	111
Chapitre 1 : Caractérisation de l'intégrase mv4 (mv4Int) et mise en place des outils d'étude.....	123
I) Redéfinition de la séquence de l'intégrase mv4Int	125
II) Caractérisation des protéines impliquées dans la recombinaison.....	127
II.1) Mise en place du système de recombinaison <i>in vitro</i>	127
II.2) Etude de l'importance de l'ORF-8 dans le système de recombinaison mv4 Int/ <i>attP/attB</i>	129
II.3) Etude de l'effet des modifications K/A et Y/F sur la protéine mv4 Int	131
III) Purification de l'intégrase mv4Int.....	133
III.1) Purification de l'intégrase native mv4 Int et de l'intégrase étiquetée 6xHis- mv4 Int	133
III.2) Stabilité des protéines purifiées et optimisation des conditions de stockage de l'intégrase	135
IV) Optimisation du système de recombinaison <i>in vitro</i> lors de l'utilisation de sites non naturels.....	137
Chapitre 2 : Caractérisation des sites de recombinaison du système mv4.....	143

I) Stratégie expérimentale de l'étude des sites de recombinaison et caractérisation des bornes du site <i>attB</i>.....	149
I.1) Utilisation de banques aléatoires des sites <i>attB</i> et <i>attP</i>	149
I.2) Caractérisation des limites du site <i>attB</i> par l'utilisation de banques aléatoires	151
II) Etude de la région centrale (région « Overlap »), région d'échange des brins.....	155
II.1) Redéfinition de la taille de la région « Overlap ».....	155
II.2) Caractérisation des contraintes de la région centrale.....	159
III) Caractérisation des contraintes nucléotidiques s'exerçant dans les régions droites des sites <i>attB</i> (B') et <i>attP</i> (C')	163
IV) Caractérisation des contraintes nucléotidiques s'exerçant dans les régions gauches des sites <i>attB</i> (B) et <i>attP</i> (C).....	165
IV.1) Caractérisation des contraintes à gauche de la région centrale du site <i>attB</i>	165
IV.2) Caractérisation des contraintes à gauche de la région centrale du site <i>attP</i>	169
V) Etude de la présence des sites cœur de fixation de mv^4Int par retard sur gel.....	171
V.1) Mise au point de la technique de retard sur gel	171
V.2) Les sites cœur B, B', C et C' sont des sites de fixation de mv^4Int	175
V.3) Caractérisation de la capacité de fixation de mv^4Int sur les sites uniques B, B', C et C'	177
VI) Caractérisation des sites bras de fixation de mv^4Int par retard sur gel.....	181
VII) Caractérisation des motifs B, B', C et C' par séquençage NGS.....	189
Chapitre 3 : Etude de l'étendue de la flexibilité de la recombinaison médiée par mv^4Int	203
I) Caractérisation du facteur stimulant la réaction de recombinaison lors de l'utilisation de substrats modifiés.....	205
II) Reprogrammation de la réaction de recombinaison vers d'autres séquences d'ARNt^{SER}	209
II.1) Reprogrammation de la réaction de recombinaison <i>in vitro</i>	209
II.2) Reprogrammation de la réaction de recombinaison <i>in vivo</i>	215
III) Reprogrammation de la réaction de recombinaison dans le gène <i>lacZ</i> de la bactérie <i>E. coli</i>.....	215
Discussions générales, conclusion et perspectives	223
Matériels et méthodes.....	243
I) Amorces, plasmides, souches et conditions de culture	245
II) Manipulation de l'ADN.....	245
III) Production extrait cellulaire	247
III.1) Surproduction des protéines recombinantes chez <i>E. coli</i>	247
III.2) Extrait cellulaire d' <i>E. coli</i> et <i>L. lactis</i>	247
IV) Purification des protéines	247

IV.1) Purification de ^{mv4} Int.....	247
IV.2) Purification de 6xHis-HU _{Lla}	249
V) Recombinaison <i>in vitro</i>	253
VI) Construction et analyse des banques aléatoires <i>attB</i> et <i>attP</i>	253
VI.1) Construction de la souche NEB5- α <i>RepA</i> ⁺	253
VI.2) Construction des banques plasmidiques <i>attB</i> et <i>attP</i>	255
VI.3) Recombinaison <i>in vitro</i> et séquençage des banques aléatoires	257
VII) Analyse bioinformatique du séquençage NGS	257
VIII) Retard sur gel	259
VIII.1) Amorces	259
VIII.2) Interaction et électrophorèse des complexes ADN-protéines	259
IX) Recombinaison <i>in vivo</i>	261
IX.1) Recombinaison <i>in vivo</i> chez <i>L. lactis</i>	261
IX.2) Recombinaison <i>in vivo</i> chez <i>E. coli</i>	261
Références bibliographiques.....	263
Annexes	299

Etude bibliographique

La recombinaison génétique est un mécanisme naturel commun à toutes les formes de vies impliquant un échange d'information génétique entre deux génomes ou entre deux molécules d'ADN. La recombinaison est centrale au métabolisme de l'ADN en assurant la conservation de l'intégrité génomique des organismes, le brassage génétique favorisant le maintien de la diversité génétique et donc l'évolution des différentes espèces. La recombinaison génétique est médiée par trois principaux systèmes de recombinaison : la recombinaison homologue, la transposition et la recombinaison spécifique de site.

La recombinaison homologue (ou recombinaison générale) nécessite des substrats partageant une identité de séquence significative (généralement > 40-50 pb) (Singer *et al.*, 1982; Watt *et al.*, 1985; King and Richardson, 1986) et peut être réalisée n'importe où dans le chromosome tant que l'homologie est maintenue entre les substrats. Ce processus nécessite de nombreuses protéines, dont la protéine essentielle RecA chez les organismes procaryotes, des cofacteurs énergétiques comme de l'ATP, et une activité de synthèse de l'ADN.

La transposition est un mécanisme de transfert d'information génétique par la transmission d'éléments génétiques mobiles, les transposons. Chez les bactéries, cette mobilité peut s'effectuer par 2 modes de transposition : la transposition conservative et la transposition répllicative. La sélection de site est relativement aléatoire, et diffère parmi les transposons : certains comme le phage Mu s'intègre dans de nombreux sites (Bukhari and Zipser, 1972) alors que d'autres comme le transposon Tn7 s'intègre dans un nombre restreint de sites (Lichtenstein and Brenner, 1982). Ce processus utilise une protéine généralement codée par le transposon, la transposase qui est regroupée en 4 familles : les transposases DDE, les transposases HUH à ADN simple-brin, les transposases à sérine et les transposases à tyrosine. Ces protéines ne nécessitent pas ou peu d'homologie entre les sites de recombinaison et une activité de réplication/réparation de l'ADN peut être associée à la réaction.

La recombinaison spécifique de site est un processus d'échange de brin d'ADN entre deux séquences d'ADN cibles précises médié par une famille de protéines spécifiques appelées recombinases spécifiques de site, impliquant parfois l'utilisation de protéines accessoires supplémentaires.

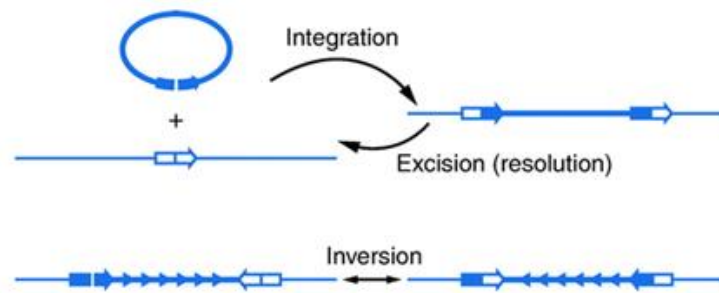


Fig. 1 Les trois possibilités de la recombinaison spécifique de site en fonction de l'orientation des sites de recombinaison (Grindley *et al.*, 2006).

La réaction de recombinaison nécessite au minimum les deux sites de recombinaison et la recombinase capable de couper et de relier les brins d'ADN tout en conservant le lien phosphodiester énergétique (Grindley *et al.*, 2006). Cette réaction de recombinaison est dite conservative car il n'y a ni dégradation, ni synthèse d'ADN, ni nécessité d'un cofacteur énergétique tel que l'ATP. En effet, lors des étapes de clivage et de ligature des brins d'ADN, un intermédiaire covalent recombinase-ADN est formé entre l'acide aminé catalytique (sérine ou tyrosine) de la recombinase et le phosphate scissile de l'ADN, l'apparition et la disparition de cet intermédiaire étant énergétiquement neutre (Grindley *et al.*, 2006). Ubiquitaires chez les bactéries, prévalentes au sein des Archaea et rarement retrouvées chez les eucaryotes (Smith and Thorpe, 2002), les recombinases spécifiques de site sont responsables de diverses fonctions importantes comme l'intégration et l'excision d'éléments génétiques mobiles, la régulation de l'expression de gènes, la ségrégation et la résolution de dimères de chromosomes et de certains plasmides...

En fonction de l'orientation des sites de recombinaison, trois résultats de recombinaison sont possibles : l'intégration d'ADN, l'excision (ou résolution) d'ADN, et l'inversion d'ADN (Fig. 1). L'intégration d'ADN résulte de la recombinaison spécifique de site entre deux sites de recombinaison situés sur deux molécules d'ADN différentes et dont au moins une molécule doit être circulaire. Lorsque les sites de recombinaison sont positionnés sur la même molécule d'ADN, le résultat de la réaction dépend de l'orientation de ces sites : s'ils sont situés en orientation directe, une réaction d'excision est effectuée créant une molécule d'ADN circulaire, alors que si ces deux séquences sont en orientation inverse, une inversion de l'ADN entre ces deux sites est effectuée. Un exemple de réaction de recombinaison provient de la réaction d'intégration et d'excision des bactériophages.

Les bactériophages sont des virus infectant spécifiquement les bactéries. Lors de l'infection d'une cellule hôte, la plupart des bactériophages procèdent à l'établissement d'un cycle lytique, où le virus va détourner la machinerie cellulaire de l'hôte afin de pouvoir répliquer son génome, d'assembler de nombreuses particules virales, et de permettre sa dissémination dans l'environnement par lyse de son hôte. Cependant, lorsque les conditions de croissance ne sont pas favorables, certains bactériophages, appelés bactériophages tempérés, vont être capable de mettre en place un cycle lysogénique aboutissant à l'intégration du génome viral dans le chromosome de leur bactérie hôte. Une fois intégré, le génome du bactériophage, appelé prophage, est répliqué passivement en même temps que le chromosome de la cellule hôte.

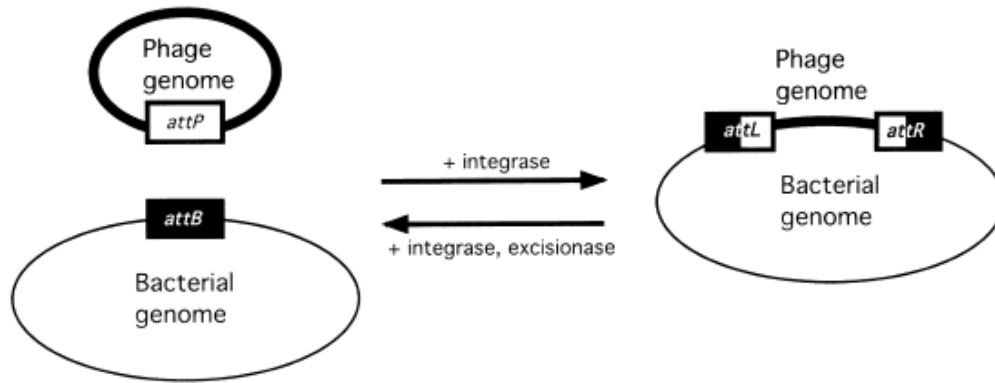


Fig. 2 Réaction d'intégration et d'excision d'un génome de bactériophage (Groth and Calos, 2004). En présence de l'intégrase, et dans certains cas de facteur d'hôte, la réaction d'intégration a lieu et aboutit à la formation de deux sites hybrides, *attL* et *attR*, encadrant le prophage. En présence de l'intégrase, de l'excisionase et dans certains cas de facteur d'hôte, la réaction d'excision a lieu, restaurant le génome viral et le chromosome bactérien.

Cette intégration du génome virale est médiée par une recombinase spécifique de site codée par le bactériophage : l'intégrase (Groth and Calos, 2004). Cette protéine effectue une recombinaison spécifique de site entre une séquence d'ADN portée par le génome du bactériophage, le site *attP* (« phage attachment site ») et une séquence d'ADN située sur le chromosome bactérien, le site *attB* (« bacterial attachment site ») (Fig. 2). En présence de l'intégrase, et dans certains cas de facteurs d'hôte, la réaction d'intégration a lieu aboutissant à la formation de deux sites hybrides *attL* (« left attachment site ») et *attR* (« right attachment site »), encadrant le prophage. Une caractéristique importante des réactions médiées par les intégrases est l'unidirectionnalité. En effet, la réaction d'excision n'est pas l'inverse de la réaction d'intégration puisqu'elle fait intervenir une protéine supplémentaire codée par le bactériophage : une protéine RDF (« Recombination Directionality Factor ») (Lewis and Hatfull, 2001). Cette réaction de recombinaison entre les sites *attL* et *attR* en orientation directe permet l'excision du bactériophage

Les recombinases spécifiques de sites appartiennent à une grande famille de protéine qui peut se diviser en deux sous-familles en fonction du mécanisme de recombinaison impliqué, et de l'acide aminé catalytique : les recombinases à sérine et les recombinases à tyrosine (Grindley *et al.*, 2006).

I) Recombinases à sérine

Historiquement, les recombinases à sérine étaient connues sous le nom de « famille resolvase/invertase » (Stark, 2014). Les premières études de la resolvase du transposon $\gamma\delta$ ont permis de révéler des différences majeures par rapport au mécanisme des recombinases de la famille de l'intégrase λ (λ Int), notamment au niveau des sites de coupure de l'ADN (Reed and Grindley, 1981). Quelques années plus tard, il a été montré que ces recombinases formaient un intermédiaire covalent protéine-ADN grâce à un acide aminé de type sérine (Hatfull and Grindley, 1986; Klippel *et al.*, 1988), se différenciant ainsi des recombinases de la famille λ Int qui formaient un intermédiaire protéine-ADN grâce à une tyrosine (Pargellis *et al.*, 1988). Dans les années 1990, les familles de recombinases spécifique de site ont donc été nommées « recombinases à sérine » et « recombinases à tyrosine » (Stark, 2014).

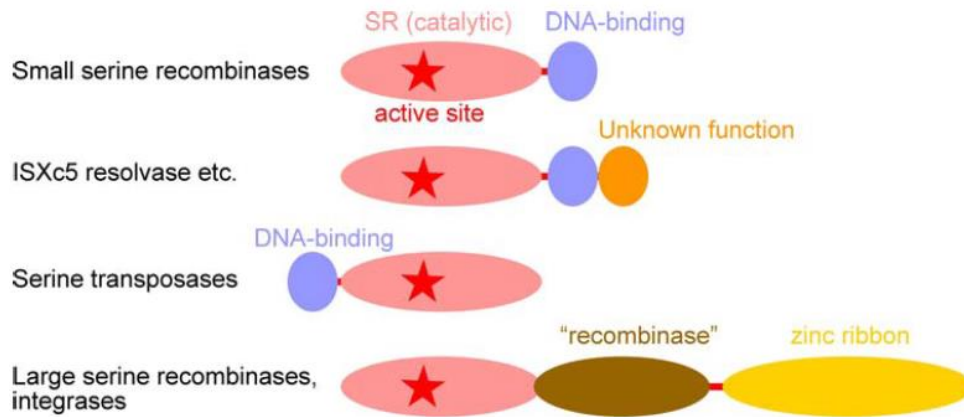


Fig. 3 Organisation des domaines protéiques des recombinases à sérine (Stark, 2014). En plus du domaine catalytique situé dans la partie N-ter (ovale rose) et contenant le site actif (étoile rouge), les recombinases contiennent d'autres domaines de nature ou localisation différente selon leur classe : les petites recombinases à sérine ont un petit domaine de liaison à l'ADN de type hélice-tour-hélice situé en C-ter (cercle bleu) ; les recombinases de type resolvase ISXc5 possèdent en plus un domaine C-ter de fonction inconnue (cercle orange) ; les transposases à sérine possèdent un domaine de liaison à l'ADN de type hélice-tour-hélice situé en N-ter ; les grandes recombinases à sérine possèdent de multiples domaines en C-ter, comprenant un « Recombinase Domain » (forme ovale marron) et un « Zinc-Ribbon Domain » (forme ovale jaune).

I.1) Rôles biologiques des recombinases à sérine

Cette large famille de protéines exerce différentes fonctions essentielles à la biologie d'Archaea, des bactéries et d'éléments génétiques mobiles qui les infectent (bactériophages, plasmides, transposons...). En effet, les résolvases de transposon comme la résolvasse $\gamma\delta$ ou la résolvasse Tn3 permettent la résolution de cointégrats de transposon (Reed, 1981a), une étape essentielle à la transposition. De leur côté, les invertases Hin et Gin, grâce à leur capacité d'inverser spécifiquement un fragment d'ADN, sont responsables des phénomènes de variations de phase des antigènes des flagelles chez *Salmonella enterica* et de la variation des fibres de queue du bactériophage Mu, permettant à ce phage d'infecter de nouveaux hôtes (Kamp *et al.*, 1978; Silverman and Simon, 1980). Enfin les intégrases comme les intégrases ϕ C31, Bxb1, permettent l'intégration et l'excision du génome de bactériophages tempérés dans le chromosome de leur bactérie hôte (Kuhstoss and Rao, 1991; Kim *et al.*, 2003).

I.2) Domaines protéiques des recombinases à sérine

Les recombinases à sérine constituent une famille très hétérogène de protéine (Fig. 3), notamment au niveau de leur taille, avec des protéines de 180 à 700 acides aminés (AA). Cette famille de protéine peut être divisée en deux sous-familles : les petites recombinases à sérine (180 à 200 AA) et les grandes recombinases à sérine (de 400 à 700 AA) (Stark, 2014). La majorité des recombinases à sérine ont un domaine catalytique situé dans la partie N-terminale (N-ter) abritant la sérine catalytique. Ce domaine est très constant entre les protéines (environ 150 AA), et plusieurs AA sont hautement conservés. En revanche, il y a une forte différence au niveau du domaine C-terminal (C-ter) entre les petites et grandes recombinases à sérine. En effet, les petites recombinases à sérine ont un domaine C-ter de petite taille constitué d'un seul domaine de liaison à l'ADN de type hélice-tour-hélice (HTH) alors que les grandes recombinases à sérine possède un domaine C-ter beaucoup plus complexe, comprenant différentes régions de liaison à l'ADN associées à des fonctions de régulation (Stark, 2014).

Actuellement, 4 classes de recombinases à sérine sont retrouvées en fonction de l'arrangement de leurs domaines protéiques (Fig. 3) :

- Les petites recombinases à sérine (comprenant majoritairement des invertases et des résolvases), qui possèdent un site catalytique en N-ter et un domaine court en C-ter de liaison à l'ADN de type HTH.

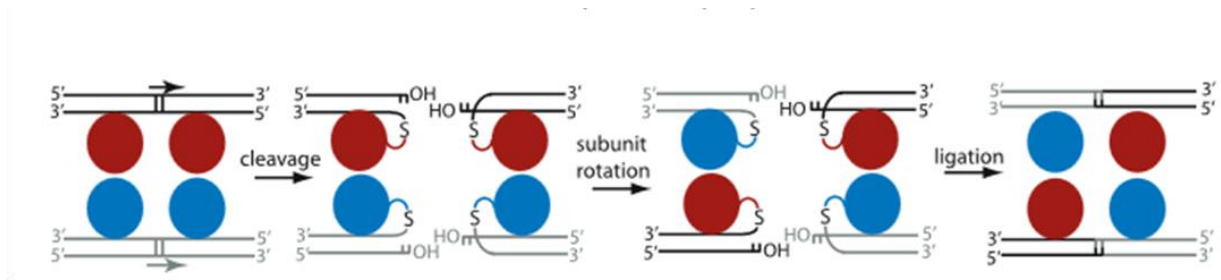


Fig. 4 Mécanisme de recombinaison des recombinases à sérine (Brown *et al.*, 2011). Un dimère de recombinases se fixe sur chaque site de recombinaison, et l'appariement de ces deux sites forme le complexe synaptique. La formation de ce complexe active le tétramère catalytique, qui effectue une coupure double-brin formant les intermédiaires 5'phosphosérine et les groupements 3'OH libres. L'échange de brin s'effectue par un mécanisme de « rotation des sous-unités », où deux sous-unités de recombinases liées de manière covalente à un demi-site de substrat d'ADN différents sont échangées par une rotation de 180° l'une par rapport à l'autre. Les extrémités clivées sont ensuite reliées par ligature, aboutissant à la formation des produits de recombinaison. Les cercles bleus et rouges représentent les recombinases liées à leur site de recombinaison. La lettre S représente la sérine catalytique.

- Les recombinases de type résolvasse ISXc5, qui possèdent les mêmes domaines que les petites recombinases à sérine, mais un domaine supplémentaire de fonction inconnu est retrouvé C-ter.
- Les transposases à sérine possédant un domaine de liaison à l'ADN de type HTH situé en N-ter et un domaine catalytique en C-ter.
- Les grandes recombinases à sérine (comprenant majoritairement des intégrases) possédant un domaine catalytique en N-ter et un domaine complexe en C-ter.

I.3) Mécanisme de recombinaison des recombinases à sérine

Les différentes données biochimiques et l'analyse de données topologiques des produits de recombinaison suggèrent un mécanisme de recombinaison par rotation à 180°. Ce phénomène a pu être observé en temps réel lors d'une expérience en molécule unique pour l'intégrase Bxb1 (Bai *et al.*, 2011). Plus récemment, des structures cristallines montrent deux états rotationnels d'un tétramère de résolvasse Sin dans le même cristal, renforçant l'hypothèse du mécanisme rotationnel (Trejo *et al.*, 2018).

Ce mécanisme de recombinaison est commun à toutes les recombinases à sérine et produit une cassure simultanée des 4 brins par la sérine catalytique (Fig. 4). Tout d'abord, un dimère de recombinases se lie à chaque site de recombinaison puis le complexe synaptique est formé par la réunion de ces deux sites. La formation de ce complexe active les 4 sous-unités de recombinase, qui clivent les 4 brins d'ADN en même temps, formant les coupures cohésives de 2 pb, caractéristiques des recombinases à sérine. Cette coupure aboutit à la formation d'un lien 5'phosphosérine et à des groupements hydroxyles libres 3'OH. L'échange de brin se produit ensuite par un processus connu sous le nom de « rotation des sous-unités » par lequel deux sous-unités protéiques liées à l'ADN effectuent une rotation de 180° l'une par rapport à l'autre. Les extrémités d'ADN peuvent ensuite être reliées, aboutissant à la formation des produits de recombinaison (Grindley *et al.*, 2006).

I.4) Les petites recombinaisons à sérine

Les résolvasés et invertasés à sérine ont été les premières recombinaisons à sérine à être étudiées. Ces deux types de recombinaisons sont proches puisqu'elles ont une organisation identique des domaines protéiques (Fig. 3), présentent 35% d'identité de séquence sur la totalité de la séquence protéique dont 41% d'identité sur le domaine catalytique N-ter (Johnson, 2015).

Les résolvasés, dans la majorité des cas, effectuent une recombinaison sur deux sites de recombinaison uniquement en orientation directe, permettant l'excision du fragment d'ADN se trouvant entre ces deux sites. De leur côté, les invertasés ne permettent la recombinaison qu'entre deux sites de recombinaison en orientation inverse, aboutissant à l'inversion du fragment d'ADN compris entre ces deux sites. L'activité catalytique de ces recombinaisons est donc hautement régulée afin que la réaction de recombinaison ne se fasse que si le complexe synaptique est correctement formé, favorisant fortement une réaction intramoléculaire par rapport à une réaction intermoléculaire (Stark, 2014). La formation correcte de ce complexe implique une structuration particulière, que ce soit par la présence de sites accessoires de la recombinaison ou bien par l'utilisation de protéines accessoires.

Pour les résolvasés, la formation de ce complexe synaptique est dépendante de la présence de sites accessoires, et parfois par l'utilisation de protéines accessoires comme la protéine HU (« Histone U ») (Rowland *et al.*, 2002). Dans le cas des invertasés, ce complexe synaptique est dépendant de la protéine Fis (« Factor for Inversion Stimulation »), en présence de son site de fixation « DNA enhancer », et parfois de protéines accessoires comme la protéine HU (Johnson, 2002).

I.4.1) Les résolvasés à sérine

Les paradigmes des résolvasés à sérine sont les résolvasés du transposon Tn3 et du transposon $\gamma\delta$ de la bactérie *Escherichia coli*. Les résolvasés effectuent leur recombinaison au niveau de deux sites *res* présents en orientation directe sur une molécule d'ADN surenroulée car seule cette orientation permet la formation d'un complexe synaptique fonctionnel.

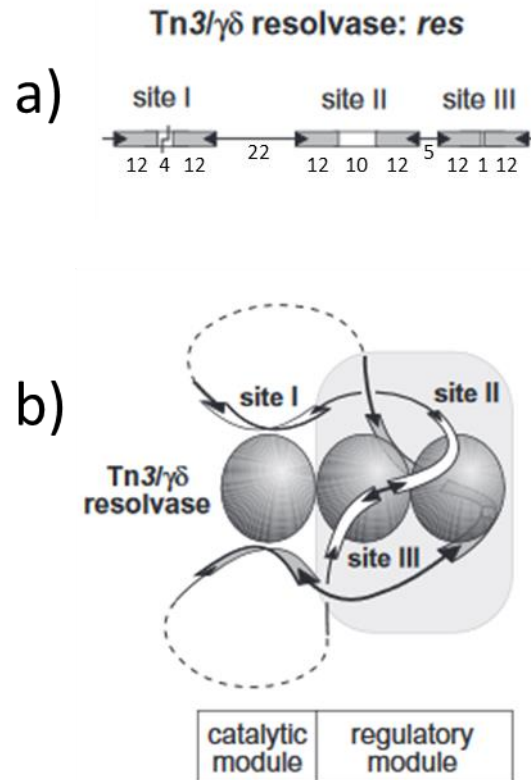


Fig. 5 Structuration des sites de recombinaison « res » et du complexe synaptique (Adapté de Rowland *et al.*, 2006). a) Structure du site *res* du système Tn3/γδ. Le site *res* du système de recombinaison Tn3/γδ présente trois sites de fixation (rectangles), chacun étant capable de lier un dimère de recombinases. L'orientation de la séquence reconnue par la resolvase est indiquée par une flèche noire. La cassure double-brin, l'échange et la ligature des brins s'effectuent au site I et les sites II et III sont des sites régulateurs permettant de stimuler et de diriger la réaction de recombinaison. b) Structuration de l'ADN lors de la formation du complexe synaptique par les résolvases Tn3/γδ. Les tétramères de recombinases sont représentés par des ovals gris. Les tétramères de résolvases situés sur le site II et III permettent l'assemblage du complexe synaptique, ainsi que l'appariement et l'activation du tétramère catalytique présent sur le site I.

Ces sites *res* contiennent chacun un site d'échange de brin et des éléments régulateurs (sites accessoires) permettant la formation d'un filtre topologique biaisant la réaction de recombinaison vers une réaction intramoléculaire entre deux sites en orientation directe (Rice, 2015). Un filtre topologique est un mécanisme où les contraintes structurelles globales des ADN substrats limitent la formation d'une synapse productive à un arrangement particulier des sites de recombinaison (Gellert and Nash, 1987). Les sites *res* sont des sites complexes de 114 pb comportant 3 sites de fixation de la résolvasse (Fig. 5a). Chaque site *res* permet la fixation d'un dimère de résolvasse, le site I étant le site d'échange de brin, et les sites II et III étant des sites accessoires permettant de stimuler et diriger la réaction de recombinaison. Chaque site de fixation est composé d'une paire de séquence de 12 pb en orientation inverse reconnue par la résolvasse, espacées par des fragments de tailles différentes (4 pb, 10 pb et 1 pb, respectivement au site I, II et III). La présence des sites accessoires est strictement nécessaire à la recombinaison car le dimère de résolvasse fixé sur le site I est incapable de former le complexe synaptique. Chaque site *res* est occupé par 3 dimères de résolvasse et, au total, 12 sous-unités de résolvasse sont retrouvées dans le complexe nucléoprotéique. Le complexe résolvasse-sites accessoires (site II-III) permet le contrôle et l'assemblage du complexe synaptique par l'appariement et l'entrelacement de ces sites accessoires (« module régulateur ») (Fig. 5b). La formation de ce complexe synaptique favorise ainsi l'appariement et l'activation des deux dimères de résolvasse liés au site I, site d'échange de brin. Le tétramère catalytique activé peut ensuite réaliser les étapes de coupures, d'échanges et de ligatures des brins d'ADN (Grindley, 2002; Grindley *et al.*, 2006; Olorunniji and Stark, 2010). L'utilisation de 12 sous-unités de résolvasse n'est pas une caractéristique propre aux résolvasse à sérine puisque la résolvasse Sin, recombinase provenant de la bactérie *Staphylococcus aureus*, n'utilise que 8 sous-unités de résolvasse mais nécessite une protéine accessoire supplémentaire, la protéine HU, pour former un complexe synaptique productif (Rowland *et al.*, 2002).

I.4.2) Les invertases à sérine

Un des systèmes d'inversion le mieux caractérisé est le système Hin de la bactérie *S. enterica*. L'invertase Hin effectue une recombinaison sur un ADN surenroulé comportant deux sites de 26 pb, les sites *hixL* et *hixR*, composés de deux séquences inversées imparfaitement répétées. Cette réaction nécessite une troisième séquence d'ADN de 69 pb situé en *cis* c'est-à-dire contenu dans la même molécule d'ADN substrat : la séquence « DNA enhancer ».

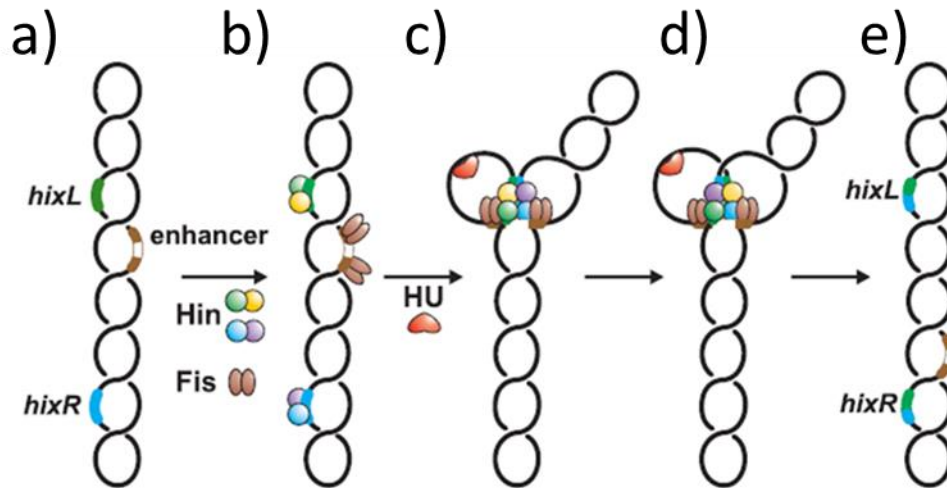


Fig. 6 Modèle de la voie de recombinaison de l'invertase Hin (Adapté de McLean *et al.*, 2013). a-b) La réaction de recombinaison débute par la liaison d'un dimère d'invertases (cercles vert-jaune et bleu-violet) aux deux sites de recombinaison *hixL* (en vert) et *hixR* (en bleu), ainsi que par la liaison de deux dimères de Fis (ovales marrons) sur la séquence « DNA-enhancer » (rectangle marron-blanc). c) La protéine HU se fixe sur l'ADN et aide à la formation du complexe synaptique permettant l'appariement des sites de recombinaison. Des interactions entre Fis et le tétramère induisent un changement de conformation protéique du tétramère d'invertases, devenant catalytiquement actif. d) Le tétramère clive les sites *hix* et l'échange de brin a lieu par rotation des sous-unités violette et jaune. e) Le tétramère catalyse la ligature des fragments d'ADN, aboutissant à la formation des produits de recombinaison.

Cette séquence « DNA enhancer » est fonctionnelle à n'importe quelle position et dans n'importe quelle orientation dans l'ADN substrat (Johnson and Simon, 1985). En plus de la recombinase, la réaction nécessite 2 facteurs d'hôtes supplémentaires : la protéine Fis et la protéine HU (Johnson *et al.*, 1986). En absence de Fis, l'activité de recombinaison *in vitro* et *in vivo* est diminuée d'un facteur 1000 indiquant un rôle crucial de cette protéine pour la recombinaison (Johnson *et al.*, 1986, Johnson *et al.*, 1988). Fis se fixe à deux domaines différents dans la séquence « DNA enhancer » et, en absence de cette protéine, cette séquence d'ADN n'a plus d'effet sur la réaction (Bruist *et al.*, 1987). De son côté, la protéine HU exerce un rôle accessoire sur la recombinaison, améliorant l'activité de recombinaison d'un facteur 10 en fonction de la localisation de la séquence « DNA enhancer ». Si cette séquence se trouve éloignée des sites de recombinaison, la protéine HU n'a plus d'effet sur la réaction (Johnson *et al.*, 1986).

A la différence des sites de recombinaison *res* des résolvas, les sites des invertases à sérine sont simples et permettent la fixation d'un dimère d'invertases. Comme les résolvas, les invertases doivent former un complexe synaptique permettant l'activation catalytique du tétramère de recombinases. Ce complexe synaptique nécessite la présence de 4 sous-unités d'invertases, des deux sites de recombinaison, de Fis et de sa séquence « DNA enhancer » ainsi que de la protéine HU (Fig. 6). Un dimère d'invertases se fixe sur chaque site de recombinaison et deux dimères de Fis se fixent sur la séquence « DNA enhancer » (Fig. 6a ; Fig. 6b) permettant l'appariement des deux sites de recombinaison dans un complexe synaptique au niveau de cette séquence (Fig. 6c). La protéine HU se fixe à l'ADN et facilite l'appariement des sites de recombinaison en induisant une courbure et la formation d'une boucle d'ADN (Fig. 6c). Les dimères de Fis interagissent avec le tétramère de recombinases, favorisant le changement de conformation du tétramère sous sa forme active. Le tétramère peut ensuite cliver les sites de recombinaison et l'échange de brin a lieu par « rotation des sous-unités » (Fig. 6d). Enfin, le tétramère permet la religature des fragments d'ADN, aboutissant à la formation des produits de recombinaison (Fig. 6e) (Johnson, 2015).

1.5) Les grandes recombinases à sérine

Les grandes recombinases à sérine sont composées d'un domaine N-ter hautement similaire à celui des petites recombinases à sérine, et d'un domaine C-ter beaucoup plus important et complexe.

Etude bibliographique

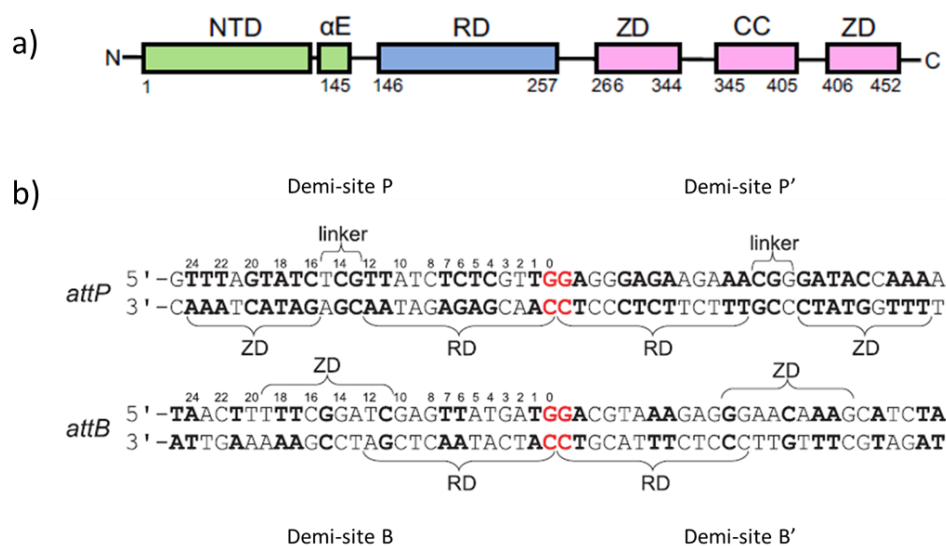


Fig. 7 Organisation des domaines protéiques des intégrases à sérine et des sites de recombinaison *attB* et *attP* (Adapté de Li *et al.*, 2018 ; Rutherford *et al.*, 2013). a) Schéma présentant l'organisation structurale de l'intégrase I1. Le domaine catalytique N-ter (rectangle vert) est relié au domaine C-ter grâce à une hélice αE (carré vert). Le domaine C-ter est composé d'un domaine RD (rectangle bleu) et d'un domaine ZD au sein duquel se trouve un domaine CC (rectangles rose). b) Représentation des différents motifs reconnus sur les sites *attB* et *attP*. Les motifs reconnus par les différents domaines de l'intégrase sont indiqués par les accolades. Le motif ZD est décalé de 5 pb sur le site *attB*. Les résidus identiques entre chaque demi-site sont représentés en gras et le dinucléotide centrale où a lieu la coupure et l'échange de brin est indiqué en rouge.

Ce groupe de protéine est majoritairement composé d'intégrases comme l'intégrase du bactériophage ϕ C31, mais peut également correspondre à des transposases comme la transposase *tpnX* du transposon *Tn4451* de *Clostridium perfringens*. Les intégrases et transposases de ce groupe partagent la même organisation du domaine C-ter (Smith, 2015). Jusqu'au milieu des années 1990, toutes les intégrases de bactériophages connues étaient des recombinases à tyrosine appartenant à la famille λ Int. La découverte de l'intégrase du bactériophage ϕ C31 infectant *Streptomyces* a permis de découvrir une nouvelle classe d'intégrase, appartenant à la famille des recombinases à sérine (Kuhstoss and Rao, 1991).

A la différence des intégrases à tyrosine, la réaction de recombinaison des intégrases à sérine ne nécessite que la recombinase et ses deux sites de recombinaison, en absence de facteur d'hôte ou d'une topologie particulière de l'ADN (Thorpe and Smith, 1998). De plus, à la différence des résolvases et des invertases, ces protéines sont capables d'effectuer une recombinaison inter- ou intramoléculaire entre deux sites de recombinaison en orientation directe ou inverse. Les intégrases à sérine sont composées (Fig. 7a) d'un domaine catalytique N-ter d'environ 130 AA contenant la sérine catalytique, et d'une hélice- α (α E) permettant la connexion au domaine C-ter de 300-500 AA. Le domaine C-ter est complexe et est composé d'un domaine « Recombinase Domain » (RD), d'un « linker » flexible et d'un domaine « Zinc Ribbon Domain » (ZD). Dans le domaine ZD, un domaine « Coiled-Coil » (CC) est retrouvé. Les sites de recombinaison *attP* et *attB* des intégrases à sérine sont courts (de 40 à 50 pb) et différents, avec le site *attP* minimum généralement plus grand d'environ 10 pb que le site *attB* minimum (Rutherford *et al.*, 2013). Ces sites contiennent chacun une séquence inversée imparfaitement répétée permettant la fixation d'un dimère d'intégrases.

Le mécanisme de recombinaison des intégrases à sérine a été méconnu pendant de nombreuses années, notamment parce que les sites *attP* et *attB* présentent une faible identité de séquence suggérant un mécanisme différent de celui des résolvases et des invertases. Malgré cette faible similarité de séquences entre les ADN substrats, l'intégrase est capable de reconnaître et de fixer ces sites avec une affinité similaire, suggérant l'existence d'un motif de l'ADN reconnu par la protéine sur chacun de ces sites (Thorpe *et al.*, 2000; Ghosh *et al.*, 2005a). En 2013, la structure cristalline d'une intégrase d'un phage de *Listeria innocua*, l'intégrase LI (Rutherford *et al.*, 2013), fixée à un demi-site *attP* a grandement facilité la compréhension du mécanisme de recombinaison des intégrases à sérine et a permis la définition des modèles d'intégration et d'excision.

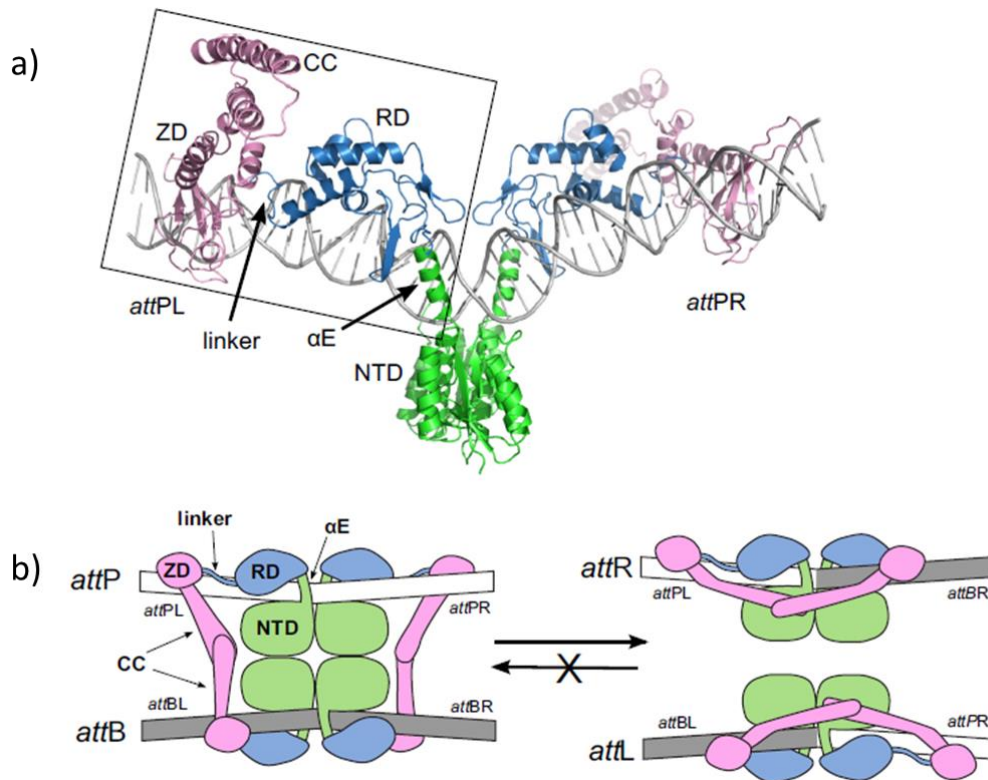


Fig. 8 Structure des domaines protéiques des intégrases à sérine et modèle de recombinaison (Adapté de Li *et al.*, 2018).
a) Modèle structural du complexe intégrase-*attP*. La région encadrée correspond à la structure cristalline du domaine C-terminal de l'intégrase LI fixée à un demi-site *attP* (Rutherford *et al.*, 2013). Cette structure a été dupliquée et, la structure du domaine N-ter et de l'hélice αE étant conservé entre les petites et grandes recombinases à sérine, la structure du complexe $\gamma\delta$ -resolvase/*res* site I a été utilisée comme guide afin de construire le complexe intégrase-*attP*. **b)** Représentation schématique de la réaction d'intégration. Un dimère d'intégrase se fixe sur *attP* et *attB* et les sites sont associés afin de former le complexe synaptique. Le tétramère d'intégrases effectue la réaction de coupure et d'échange de brin afin de former les sites *attL* et *attR*. La réaction d'excision est impossible en absence de la protéine RDF.

Dans cette étude, il est montré que le domaine RD contacte les 13 nucléotides adjacents au dinucléotide central du site, que le « linker » est en contact avec les 3 bases suivantes et que le domaine ZD effectue des interactions spécifiques avec un motif de 9 pb sur le demi-site *attP* (Fig. 7b, Fig. 8a). Dans cette structure, le domaine CC ne semble pas impliqué dans la fixation de l'intégrase sur le site *attP* et cela a été démontré par le fait qu'une version Δ CC de l'intégrase se fixe avec la même affinité sur son site *attP* que l'intégrase sauvage (Rutherford *et al.*, 2013). Bien qu'il y ait une différence de taille entre les sites minimums *attB* et *attP*, leur comparaison montre une similitude de séquence au niveau du site reconnu par le domaine RD, suggérant que ce domaine se fixerait de manière équivalente sur le site *attB*. En revanche, les séquences aux extrémités des deux sites présentent peu de similarité, suggérant l'absence d'un motif ZD sur le site *attB*. Cependant, en regardant attentivement au niveau des sites *attB*, le motif ZD a pu être identifié, celui-ci étant décalé de 5 pb vers l'intérieur de ce site (Fig. 7b) (Rutherford *et al.*, 2013). Ce décalage du motif ZD semble partagé entre les sites *attB* d'autres intégrases à sérine et a pour conséquence un chevauchement des séquences reconnues par le domaine ZD et le « linker », obligeant ce dernier à adopter une conformation différente par rapport à sa conformation sur le site *attP* (Fig. 8b) (Mandali and Johnson, 2021; Smith, 2015; Van Duyne and Rutherford, 2013). Ce changement de structuration sur le site *attB* implique que le domaine CC adopte une position et une trajectoire différente par rapport à celui des intégrases fixées sur le site *attP* (Rutherford and Van Duyne, 2014).

Les intégrases à sérine sont caractérisées par l'importante unidirectionnalité de leurs réactions de recombinaison. En effet, ces protéines sont capables de se fixer sous forme de dimères sur leur site *attB*, *attL*, *attP* et *attR*, mais seule la combinaison *attB* et *attP* permet à l'intégrase de former un complexe synaptique (Thorpe *et al.*, 2000; Ghosh *et al.*, 2005a). Les modèles structuraux des intégrases fixées sur leur site *attB* et *attP* suggèrent que les domaines CC pourraient avoir un rôle dans la régulation de cette directionnalité (Fig. 8b). En effet, d'après ces modèles, les domaines CC pourraient interagir entre les sous-unités fixées sur le demi-site P et B, et les sous-unités fixées sur le demi-site P' et B'. Ces interactions entre domaines CC ne semblent pas possible dans le cas de complexe synaptique *attB/attB* et *attP/attP* et pourrait expliquer le fait que l'intégrase soit capable d'effectuer une recombinaison uniquement entre un site *attB* et un site *attP* (Rutherford *et al.*, 2013).

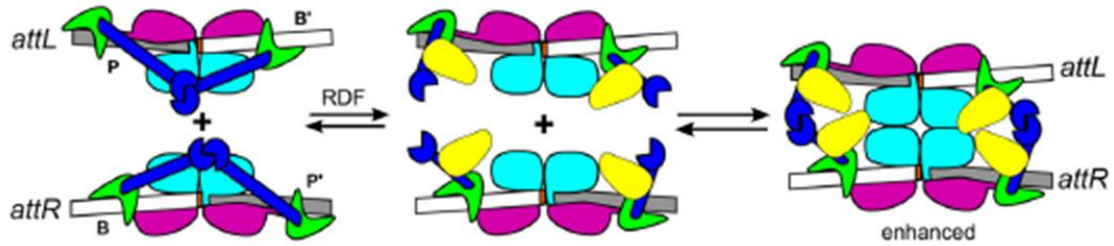


Fig. 9 Modèle de la voie de recombinaison excisive des intégrases à sérine (Rutherford and Van Duynne, 2014). Le domaine catalytique est représenté en bleu clair et le domaine RD est représenté en violet. L'arrangement des domaines ZD (vert) sur les sites *attL* et *attR* favorise une interaction intramoléculaire entre les domaines CC (bleu foncé) inhibant la réaction d'excision. La fixation de la protéine RDF (jaune) à la base des domaines CC des sous-unités d'intégrases induit une réorientation de ces domaines qui favorise une interaction intermoléculaire entre les domaines CC formant ainsi le complexe synaptique.

Une nouvelle structure cristalline a permis de montrer l'interaction entre deux domaines CC d'un dimère d'intégrases LI, renforçant le modèle de recombinaison des intégrases à sérine. De plus, dans cette étude, le domaine CC favorise la dimérisation des intégrases en solution, expliquant pourquoi l'intégrase se fixe sur ses sites d'attachement sous forme de dimère (Gupta *et al.*, 2017). Lors de la réaction d'intégration, un dimère d'intégrases se fixe sur le site *attP* et un dimère sur le site *attB*. Ces deux complexes ont une conformation différente à cause du positionnement différent du motif reconnu par le domaine ZD. Le complexe synaptique est ensuite formé par interaction entre les domaines catalytique N-ter des intégrases, et interaction entre les domaines CC entre les sous-unités d'intégrase (Fig. 8b) (Rowley and Smith, 2008; Rowley *et al.*, 2008; Rutherford *et al.*, 2013). La formation de ce complexe active le tétramère catalytique, permettant la réaction de coupure et d'échange de brin, formant ainsi les produits de recombinaison *attL* et *attR*. L'arrangement des différents domaines ZD fixés sur les sites *attL* et *attR* favorise une interaction intramoléculaire entre les domaines CC, inhibant ainsi la réaction d'excision (Fig. 8b ; Fig. 9) (Rutherford *et al.*, 2013; Van Duyne and Rutherford, 2013; Gupta *et al.*, 2017).

La réaction d'excision est quant à elle dépendante de protéines RDF (« Recombination Directionality Factor »). Les RDF Gp3 (ϕ C31) et Gp47 (Bxb1) ont été identifiés et ces protéines ne se fixent pas à l'ADN, indiquant un mode de fonctionnement différent des excisionases des systèmes de la famille λ Int (Ghosh *et al.*, 2006; Khaleel *et al.*, 2011). L'intégrase A118, intégrase du phage A118 de *Listeria monocytogenes* est une intégrase présentant 98% d'identité avec l'intégrase LI et dont le RDF a récemment été identifié, la protéine Gp44. Cette protéine inhibe la réaction d'intégration en se fixant sur le domaine C-ter de l'intégrase A118, au niveau de la base du domaine CC, et active la réaction d'excision par une deuxième interaction *via* son domaine N-terminal (Mandali *et al.*, 2017). Les différents résultats suggèrent un modèle où les deux contacts effectués par la protéine Gp44, induisent une réorientation des domaines CC, contrôlant ainsi la directionnalité de la recombinaison. Des résultats similaires ont été obtenus avec le RDF Gp3 du système de recombinaison ϕ C31 (Fogg *et al.*, 2018). Ainsi, dans la réaction d'excision, l'ajout de la protéine RDF pourrait interrompre les interactions intramoléculaires en se fixant à la base des domaines CC des différentes intégrases, et induire une réorientation de ces domaines. Ce changement d'orientation pourrait permettre une interaction intermoléculaire de ces domaines donnant lieu à la formation du complexe synaptique *attL/attR*, activant ainsi le tétramère catalytique pour les réactions de coupure et de religature des brins (Fig. 9) (Fogg *et al.*, 2018; Mandali *et al.*, 2017; Rutherford and Van Duyne, 2014).

II) Recombinases à tyrosine

La première recombinase à tyrosine à avoir été purifiée et étudiée fut l'intégrase du bactériophage λ (Nash, 1974a). Des années plus tard, il a été montré que λ Int produisait un intermédiaire covalent intégrase-ADN grâce à un acide aminé de type tyrosine (Pargellis *et al.*, 1988), devenant ainsi le membre fondateur de la famille des recombinases à tyrosine, originellement appelée « λ Int family ». Depuis, de nombreuses protéines appartenant à cette famille ont pu être découvertes. En effet, en 2012 une étude identifiait plus de 1300 recombinases à tyrosine parmi les génomes présents dans les bases de données (Van Houdt *et al.*, 2012).

II.1) Rôles biologiques des recombinases à tyrosine

Les recombinases à tyrosine sont retrouvées dans les trois domaines du vivant où l'on observe la plupart de ces protéines parmi les procaryotes et les bactériophages. Ces protéines sont aussi retrouvées chez les levures (Jayaram *et al.*, 2015) et d'autres études ont également identifié des recombinases à tyrosine chez les rétrotransposons de nombreux eucaryotes et parmi un groupe de transposon à ADN eucaryote : les cryptons (Poulter and Butler, 2015). Au sein de ces différents organismes, ces protéines effectuent une variété de réactions de réarrangement d'ADN permettant l'intégration et l'excision d'éléments génétiques mobiles (bactériophages, éléments mobiles conjugatifs, intégrons...), la résolution de multimères de chromosomes et de plasmides, le contrôle du nombre de copies de plasmides, le contrôle de l'expression de gènes... (Grindley *et al.*, 2006).

En plus de λ Int, différentes recombinases (Cre du bactériophage P1, Flp du plasmide 2 μ m de *Saccharomyces cerevisiae* et XerC/D d'*E. coli*) ont été intensivement étudiées au niveau biochimique et structural, et sont devenues des modèles pour les recombinases à tyrosine, permettant de grandes avancées dans la compréhension du mécanisme de recombinaison de ces protéines (Van Duyne, 2002). En fonction de la complexité du système de recombinaison, il est possible de classer les recombinases à tyrosine en trois groupes. Le premier groupe rassemble les systèmes dits « simples » et représentés par les systèmes Cre/*loxP* et Flp/*FRT*. Ces systèmes ne nécessitent pas de topologie particulière de l'ADN substrat et les réactions de recombinaison ne nécessitent ni protéine ni séquence d'ADN accessoire.

Etude bibliographique

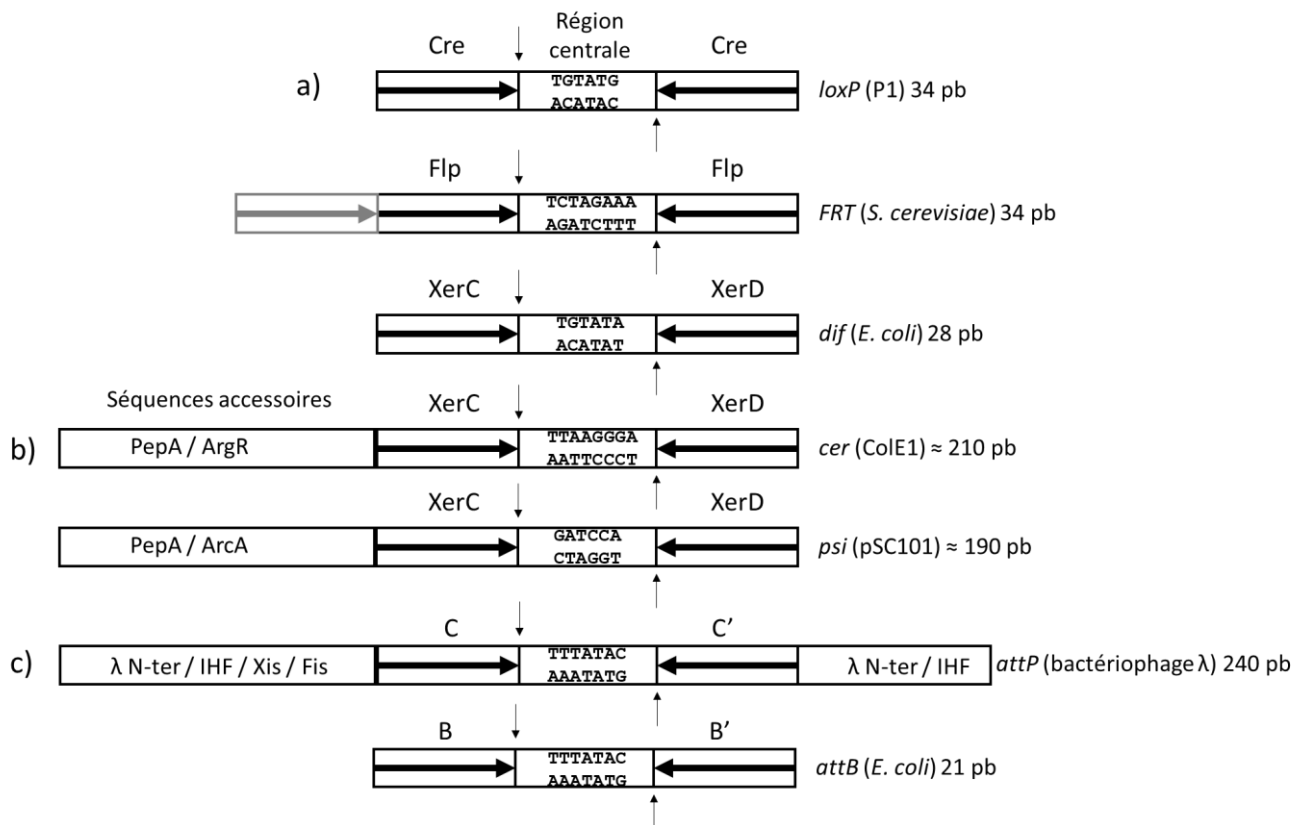


Fig. 10 Organisation des sites de recombinaison des recombinases à tyrosine. Représentation schématique des sites de recombinaison des recombinases Cre, XerC, XerD, Flp et λ Int. **a)** Représentation schématique des sites de recombinaison minimaux (sites cœur) des systèmes Cre, Flp et XerC/D. Les flèches horizontales indiquent les sites de fixation cœur (séquences inversées imparfaitement répétées), encadrant la région centrale. Cette région centrale a une taille de 6 à 8 pb en fonction des recombinases. Les sites de coupures sont représentés par les flèches noires verticales, bordant la région centrale. Le troisième site de fixation de Flp, dispensable à la réaction de recombinaison est indiqué en gris. **b)** Exemple de sites de recombinaison complexe du système XerC/D. Les sites plasmidique *cer* et *psi* sont composés d'une région cœur, et d'une séquence ADN accessoire permettant la fixation des protéines PepA et ArgR ou PepA et ArcA. **c)** Sites de recombinaison du système λ Int. Le site *attB* est simple et ressemble aux sites de recombinaison des systèmes Cre et Flp. Le site *attP* est complexe. Il est composé d'une région cœur et de deux régions bras (P et P'). Le bras P permet la fixation des protéines IHF, Xis, Fis et du domaine supplémentaire N-terminale de l'intégrase λ . Le bras P' permet la fixation du cofacteur IHF et du domaine supplémentaire N-terminale de l'intégrase λ .

Le deuxième groupe comprend les recombinases qui nécessitent des protéines accessoires pour effectuer leur réaction. Le système le plus étudié de ce groupe est le système XerC/D d'*E. coli* qui, en fonction de l'ADN substrat utilisé nécessite soit la protéine FtsK, soit les protéines PepA et ArgR ou les protéines PepA et ArcA. L'originalité de ce système provient de l'utilisation de deux recombinases différentes, XerC et XerD, pour catalyser l'intégralité de la réaction de recombinaison. Le troisième groupe est composé des recombinases dites hétérobivalentes, c'est-à-dire ayant la capacité de reconnaître deux régions d'ADN différentes. Ces protéines nécessitent aussi la présence de protéines accessoires et le paradigme de ce système de recombinaison est l'intégrase du bactériophage λ .

II.2) Organisation des sites de recombinaison des recombinases à tyrosine

Par définition, la taille minimum des séquences d'ADN permettant la fixation des recombinases et l'échange de brin correspondent à ce qui est appelé « sites cœur ». Ces sites cœur sont composés de deux séquences d'ADN inversées imparfaitement répétées permettant la fixation d'un dimère de recombinases, et encadrant la région centrale, lieu de coupure et d'échange des brins (Fig. 10a) (Van Duyne, 2002). En fonction du système de recombinaison, les séquences de fixation de la protéine ont une taille de 9 à 13 pb, et la région centrale a une taille de 6 à 8 pb. Les réactions de coupure et d'échange de brins ont lieu à la jonction des sites de fixation des recombinases et l'extrémité 5' de la région centrale. Dans les systèmes « simple », comme le système *Cre/loxP*, la région centrale est asymétrique et définit l'orientation du site sur l'ADN. Dans le cas de systèmes plus complexes, le site de recombinaison nécessite, en plus du site cœur, une séquence d'ADN accessoire permettant la fixation d'autres protéines, ou d'autres domaines de la recombinase. C'est le cas du système XerC/D sur les sites plasmidiques *cer* et *psi* (Fig. 10b), et du site *attP* du bactériophage λ (Fig. 10c).

II.3) Organisation des domaines protéiques des recombinases à tyrosine

Les membres de la famille des recombinases à tyrosine partagent peu de similarité de séquences et seulement quelques résidus hautement conservés sont retrouvés dans le domaine catalytique. Le premier alignement de 7 séquences de recombinases à tyrosine (Argos *et al.*, 1986) a permis l'identification de 3 résidus conservés dans le domaine catalytique C-ter : H-R-Y (Histidine-Arginine-Tyrosine).

Etude bibliographique

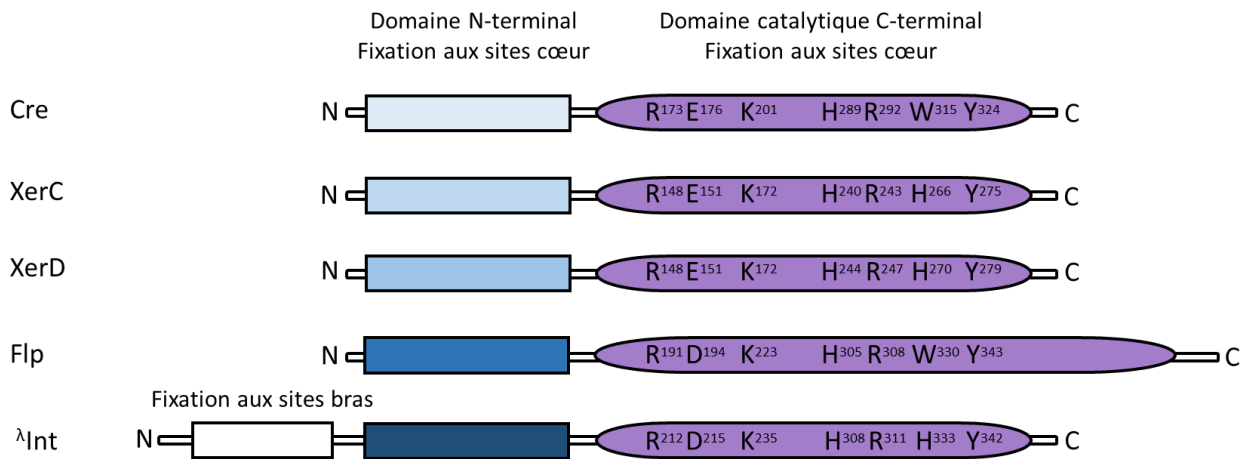


Fig. 11 Organisation des domaines protéiques des recombinases à tyrosine. Représentation schématique des recombinases Cre, XerC, XerD, Flp et λ Int. Les recombinases à tyrosine sont composées d'un domaine N-terminal variable, permettant la fixation aux sites cœur et d'un domaine catalytique conservé contenant le motif R-D/E-K-H-R-H/W-Y. L'intégrase λ , modèle des recombinases hétérobivalentes, se différencie des autres types de recombinases à tyrosine par la présence d'un domaine supplémentaire du côté N-terminal, permettant la fixation aux sites bras présent sur les séquences d'ADN accessoires.

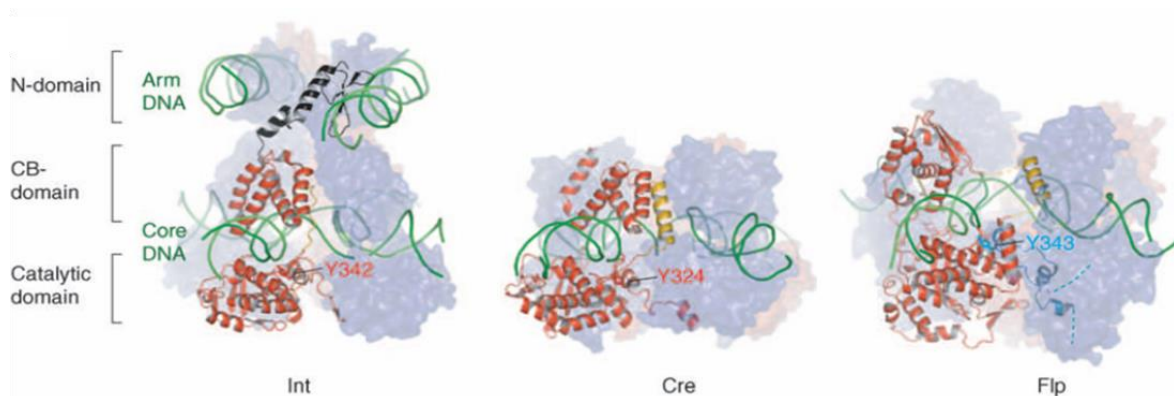


Fig. 12 Comparaison de la structure d'un monomère actif de λ Int, Cre et Flp fixés sur leurs ADN substrats (Radman-Livaja et al., 2006). Les sites actifs de chaque recombinase possèdent une tyrosine nucléophile, indiquée en rouge lorsqu'elle est fournie *in cis* et en bleu lorsqu'elle est fournie *in trans* par un monomère voisin. Les hélices- α rouges représentent les domaines N-ter et C-ter de Cre et Flp, et les domaines « Core-binding » et catalytique de λ Int. Le linker entre le domaine N-ter et C-ter est représenté en jaune et présente une hélice- α additionnelle chez les recombinase Cre et Flp impliquée dans des interactions intermoléculaires. Le domaine supplémentaire en N-ter de l'intégrase λ est représenté en noir. Les trois recombinases adoptent une structure de type « C-shaped clamp » entourant l'ADN.

Des années plus tard, un nouvel alignement de 28 séquences protéiques (Abremski and Hoess, 1992) a mis en évidence la conservation d'un deuxième résidu arginine dans le domaine C-ter : R-H-R-Y. A la fin des années 1990, l'émergence des données structurales ainsi que l'alignement de 105 séquences de recombinases à tyrosine (Nunes-Düby *et al.*, 1998), a permis d'ajouter deux résidus hautement conservés au motif catalytique : un résidu lysine (K) et un résidu histidine ou tryptophane (W) en fonction de la recombinase : R-K-H-R-H/W-Y. Enfin, une analyse Psi-Blast (« Position-Specific Iterative Basic Local Alignment Search Tool ») a comparé le domaine C-ter de la recombinase Cre (Gibb *et al.*, 2010) avec 12000 séquences protéiques et indique qu'un résidu supplémentaire serait hautement conservé dans cette poche catalytique: un résidu chargé négativement, aspartate (D) ou glutamate (E) en fonction de la recombinase.

Les recombinases à tyrosine sont généralement composées de deux domaines protéiques impliqués dans la fixation au site cœur : un domaine N-terminal variable, permettant des interactions protéine-protéine et un domaine catalytique C-terminal contenant le motif catalytique (Fig. 11) (Grindley *et al.*, 2006). Dans le cas des recombinases hétérobivalentes, comme l'intégrase λ , un domaine supplémentaire de fixation à l'ADN est retrouvé à l'extrémité N-terminale.

II.4) Structure des recombinases à tyrosine

La cristallographie aux rayons X a permis d'obtenir une multitude d'informations sur la structure et sur le mécanisme de recombinaison des recombinases à tyrosine. Les structures des complexes synaptiques de Cre, Flp et λ Int sont disponibles, ainsi que les structures des domaines C-ter de λ Int, HP1 Int et de la protéine entière XerD en absence de ligand (Biswas *et al.*, 2005; Chen *et al.*, 2000; Gopaul *et al.*, 1998; Guo *et al.*, 1997; Guo *et al.*, 1999; Hickman *et al.*, 1997; Kwon *et al.*, 1997; Subramanya *et al.*, 1997). La structure du complexe ADN-topoisomérase I humaine, et du domaine C-ter de la topoisomérase du virus de la vaccine présentent une structure du site catalytique similaire à celui des recombinases à tyrosine et ont permis de faciliter la compréhension du mécanisme de coupure et de ligature de l'ADN (Cheng *et al.*, 1998 ; Redinbo *et al.*, 1998). Malgré la faible similarité de séquence au niveau protéique, les recombinases à tyrosine semblent toutes adopter une structure de type « C-shaped clamp », entourant la molécule d'ADN (Fig. 12). Le domaine N-ter interagit avec le sillon majeur de l'ADN, proche de la région centrale, et sur la face opposée de l'ADN le domaine C-terminal interagit avec les sillons mineur et majeur (Grindley *et al.*, 2006).

A l'intérieur du domaine catalytique, les résidus R-K-H-R-H/W-Y sont regroupés dans une poche catalytique et permettent les réactions de coupure et de ligature des brins. L'architecture du domaine catalytique est partagée entre les recombinases à tyrosine et les topoisomérases de type IB (Sherratt and Wigley, 1998). Ainsi, les topoisomérases de type IB partagent 5 des 6 résidus conservés du motif catalytique, dont la tyrosine nucléophile : RKRHY (Cheng *et al.*, 1998). Parmi les recombinases à tyrosine, la coupure de brin peut être réalisée de deux manières. Dans la majorité des cas, dont Cre et λ Int, les résidus formant la poche catalytique sont apportés par le même monomère et la coupure réalisée par ce site actif est donc appelée coupure en « *cis* » (Fig. 12) (Guo *et al.*, 1997; Nunes-Düby *et al.*, 1994). Dans le cas de la recombinase Flp, le site actif est formé par l'introduction de la tyrosine nucléophile portée par l'hélice M d'un monomère à l'intérieur d'une poche catalytique contenant les autres résidus d'un deuxième monomère. Dans ce cas, la coupure réalisée par la recombinase Flp est dite en « *trans* » (Chen *et al.*, 1992).

Au niveau structural, les domaines catalytiques des recombinases adjacentes interagissent par un échange d'hélices (ou de feuillettes en fonction de la recombinase), insérant leur segment C-terminal dans le domaine catalytique d'une sous-unité adjacente. Cet échange n'est pas réciproque entre les monomères liés sur un même site, mais se fait de manière cyclique entre les 4 monomères de recombinase formant le complexe synaptique. Cette interaction entre recombinases est essentielle car elle constitue un moyen de communication entre les protéines du complexe, permettant l'activation et l'inactivation de l'activité catalytique des recombinases et garantissant que seul un des deux monomères fixés sur le site soit en conformation active (Grindley *et al.*, 2006). Ce contrôle de l'activité s'effectue de plusieurs façons en fonction des recombinases. Chez la recombinase Flp, les monomères actifs permettent un positionnement précis de l'hélice porteuse de la tyrosine nucléophile (en *trans*), tandis que les monomères inactifs positionnent la tyrosine à distance du site actif (Chen *et al.*, 2000). Chez l'intégrase λ , la régulation se fait aussi par le positionnement de la tyrosine nucléophile (en *cis*). Parmi les intégrases actives, des interactions stabilisent le positionnement de la tyrosine, alors que chez les intégrases inactives, ces interactions sont perturbées, et l'hélice portant la tyrosine est désordonnée (Aihara *et al.*, 2003). Enfin, chez la recombinase Cre, la lysine catalytique joue un rôle important. Parmi les recombinases actives, cette lysine se trouve près du phosphate scissile, alors que chez les recombinases inactives, cette lysine est déplacée hors du site actif (Van Duyne, 2001).

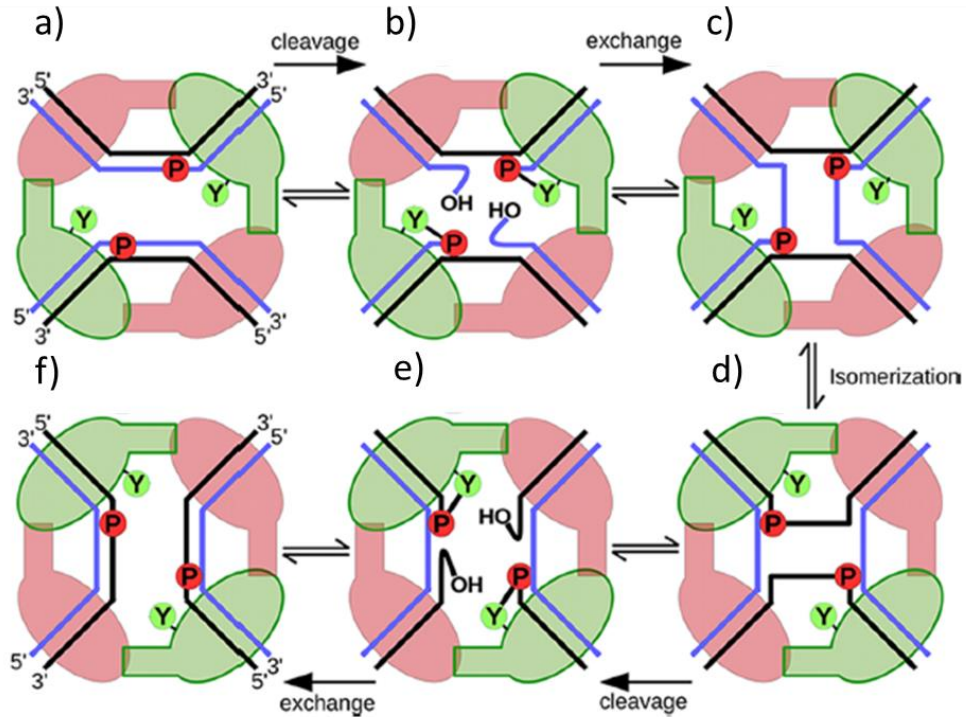


Fig. 13 Représentation schématique du mécanisme de recombinaison des recombinases à tyrosine (Adapté de Meinke *et al.*, 2016). La recombinase fixée à chaque demi-site est représentée soit sous forme active (formes vertes) soit sous forme inactive (formes roses). La tyrosine nucléophile est indiquée par la lettre Y et le phosphate scissile activé est représenté par la lettre P. Les protomères de recombinase communiquent à leur voisin *en trans* grâce à leur extrémité C-terminale. **a)** Le complexe synaptique se forme et est composé de deux ADNs substrats et d'un tétramère de recombinases. **b)** La réaction de recombinaison commence par l'attaque des phosphates scissiles par les tyrosines nucléophiles portées par les recombinases actives, aboutissant à la formation d'un lien 3' phospho-tyrosine et d'une extrémité 5' OH libre. **c)** L'échange de brin a lieu lorsque l'extrémité 5'OH attaque la liaison 3' phosphotyrosine voisine aboutissant à la formation d'une jonction de Holliday. **d)** Le complexe s'isomérisé ce qui a pour conséquence de rendre les monomères actifs inactifs et *vice versa*. **e)** Les tyrosines catalytiques attaquent les phosphates scissiles, formant de nouveau un lien 3' phosphotyrosine et libérant une extrémité 5'OH libre. **f)** L'extrémité 5'OH libre attaque la liaison phosphotyrosine voisine, aboutissant à la formation des produits de recombinaison.

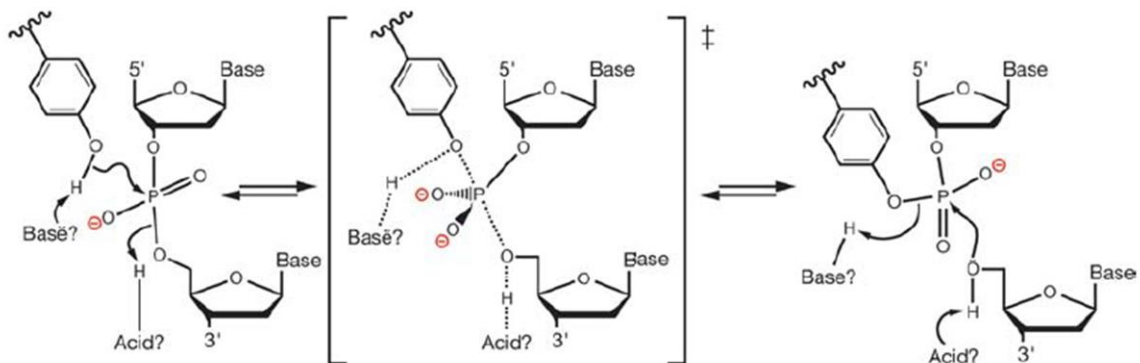


Fig. 14 Réaction de coupure de brins réalisée par les recombinases à tyrosine (Grindley *et al.*, 2006). Cette réaction utilise une catalyse acide-base. Dans la réaction, la base activerait la tyrosine nucléophile par déprotonation de son groupement hydroxyle et l'acide est responsable de la protonation du groupe 5'O libre de l'ADN. Cette réaction implique la formation d'un intermédiaire pentavalent du phosphate.

II.5) Mécanisme de recombinaison

La réaction de coupure est effectuée par un mécanisme similaire à celui utilisé par les topoisomérases de type IB. Le mécanisme de recombinaison commence par la formation d'un complexe synaptique composé de deux sites de recombinaison et d'un tétramère de recombinases (Fig. 13a). Les tyrosines nucléophiles attaquent les phosphates scissiles (Fig. 13b), entraînant la formation d'une jonction de Holliday (Fig. 13c). Ce complexe s'isomérise et permet l'activation des recombinases inactives et l'inactivation des recombinases actives (Fig. 13d). Ensuite, les mêmes étapes de coupure et d'échange de brins (Fig. 13e) ont lieu, aboutissant à la formation des produits de recombinaison (Fig. 13f).

II.5.1) Coupure de brin

Le mécanisme de coupure et de ligature des brins implique la formation d'un intermédiaire 3'-phosphotyrosine qui est conservé chez les recombinases à tyrosine et partagé avec les topoisomérases de type IB. Cette réaction consiste en une attaque nucléophile du phosphate scissile par la tyrosine catalytique. Il a été proposé que la réaction de transestérification utilise une catalyse acide-base. Dans la réaction de coupure de brin, la base accepte un proton de la tyrosine et un acide donne un proton au groupe 5'O libre. Dans la réaction de ligature des brins, les rôles sont inversés et l'acide accepte un proton du groupement 5'OH et la base donne un proton à la tyrosine (Fig. 14) (Van Duyne, 2001).

Les données structurales, aidées par des analyses mutationnelles, ont fourni des informations sur le rôle des différents résidus compris dans le motif catalytique : R-D/E-K-H-R-H/W-Y. Dans la topoisomérase du virus de la vaccine, la lysine a été montrée comme étant l'acide catalytique responsable de la protonation du groupe 5'O libre de l'ADN (Krogh and Shuman, 2000) et un rôle identique a été attribué à la lysine de Cre quelques années plus tard (Ghosh *et al.*, 2005a). Dans la réaction, la tyrosine est le nucléophile et les deux arginines semblent avoir un rôle important dans la stabilisation des charges du phosphate (Van Duyne, 2015). Chez la recombinase Flp, le premier résidu histidine pourrait être la base générale qui accepte un proton de la tyrosine, permettant son activation en nucléophile (Whiteson *et al.*, 2007). Chez la recombinase Cre, ce résidu pourrait aussi être la base générale et participer à la stabilisation des charges du phosphate, alors que les résidus glutamate et tryptophane ont plutôt un rôle structural du site actif (Van Duyne, 2015).

II.5.2) La réaction d'échange de brin

La réaction d'échange de brin s'effectue dans la région centrale, où une identité de séquence entre les deux sites de recombinaison était supposée. Au départ, cette nécessité d'identité entre les sites impliquait un mécanisme de migration de la branche de la jonction de Holliday au travers de la région centrale, en échangeant une base l'une après l'autre (Kitts and Nash, 1987). Cependant, ce mécanisme a été invalidé par différentes données biochimiques et structurales. A la place, un mécanisme de « strand-swapping isomerisation » a été proposé dans le système λ , puis dans les autres systèmes (Nunes-Düby *et al.*, 1995a). Dans ce mécanisme, les 2-3 bases adjacentes à la coupure vont se désappairer de leur brin complémentaire et être échangées au brin complémentaire du substrat partenaire, permettant la formation de la jonction de Holliday. Ensuite, une étape d'isomérisation du complexe a lieu et ce changement de conformation permet l'activation des recombinases inactives et l'inactivation des recombinases actives. Cette isomérisation semble être accompagnée d'une migration de la branche limitée en fonction de la taille de la région centrale. Elle est de 0-1, 1-2 et 2-3 bases pour des régions centrales de 6 pb (Cre), 7 pb (λ Int) et 8 pb (Flp), respectivement (Voziyanov *et al.*, 1999). Ensuite, le deuxième échange de brin est réalisé par les recombinases permettant la résolution de la jonction de Holliday et la formation des produits de recombinaison. Ce mécanisme de « strand-swapping isomerisation » nécessite aussi une identité nucléotidique parfaite entre les régions centrales des sites de recombinaison, et il a été pensé pendant de nombreuses années que l'identité dans la région centrale était une règle d'or parmi les recombinases à tyrosine (Grindley *et al.*, 2006). Cependant, la découverte de nouvelles recombinases à tyrosine a montré que certaines protéines étaient capables d'effectuer leur réaction de recombinaison avec de nombreux mésappariements dans la région centrale, et que la règle d'une identité parfaite n'était pas généralisable à l'ensemble de la famille des recombinases à tyrosine (Rajeev *et al.*, 2009).

II.6) Cre et Flp, modèles des recombinases à tyrosine simple

II.6.1) La recombinase Cre

Cre (« Cyclization recombination ») est une recombinase à tyrosine de 38 kDa (Sternberg *et al.*, 1986), composée d'un domaine N-terminal de 130 AA et d'un domaine C-terminal de 211 AA. Cette protéine est codée par le bactériophage P1, infectant la bactérie *E. coli*.

Le bactériophage P1 diffère de la plupart des bactériophages tempérés car le génome viral n'est pas intégré dans le chromosome bactérien, mais se réplique comme un plasmide dans les cellules lysogènes. Le rôle de Cre dans le cycle de vie de ce virus est de circulariser le génome linéaire après infection dans la cellule hôte (Hochman *et al.*, 1983) et de résoudre les dimères de chromosomes P1 créés lors de la réplication par recombinaison homologue, favorisant une ségrégation stable des unités génomiques de P1 chez les cellules filles, assurant ainsi sa maintenance dans les cellules lysogènes (Van Duyne, 2015).

La recombinaison s'effectue au niveau des sites *loxP* (« locus of crossing over in P1 ») qui sont des sites courts de 34 pb, composés de deux séquences inversées imparfaites répétées de 14 pb permettant la fixation des recombinases séparées par une région centrale de 6 pb (Fig. 10) (Hoess *et al.*, 1982). Cre est capable d'effectuer une recombinaison réversible intramoléculaire, entre des sites en orientation directe (excision) ou en orientation inverse (inversion), et est aussi capable de recombinaison réversible intermoléculaire. Cette recombinaison ne nécessite pas de topologie particulière de l'ADN (ADN surenroulé ou linéaire), et ne nécessite pas de protéines accessoires (Abremski and Hoess, 1984). Ainsi, Cre est devenu un modèle parmi les recombinases à tyrosine simple car la réaction ne nécessite que la recombinase, et les deux ADN substrats.

II.6.2) La recombinase Flp

Flp (« flipase ») est une recombinase à tyrosine de 45 kDa (Broach and Hicks, 1980), composée d'un domaine N-ter de 13 kDa et d'un domaine C-ter de 32 kDa. Cette protéine est codée par le plasmide 2 μ m de la levure *S. cerevisiae* et permet l'amplification de ce plasmide en inversant un segment d'ADN encadré par deux sites *FRT* (« Flp Recognition Target »), initiant la réplication en cercle roulant en modifiant la direction de la fourche de réplication et permettant la formation de multimères de plasmides 2 μ m. Enfin, Flp catalyse la recombinaison entre deux sites *FRT* en orientation directe, permettant la résolution des multimères de plasmides en monomères. Chez la levure, Flp est donc utilisée comme une invertase et comme une résolvasse (Sadowski, 1995).

Les sites minimum *FRT* ont une structuration similaire aux sites *loxP* du système Cre. Ces sites de 34 pb sont composés de deux séquences inversées imparfaitement répétées de 13 pb permettant la fixation de la recombinase et encadrant une région centrale de 8 pb (Fig. 10).

Une particularité des sites *FRT* provient de la présence d'un troisième site de fixation de Flp, en orientation directe, non essentiel pour les réactions de recombinaison *in vivo* et *in vitro* (Sadowski, 1995). Tout comme Cre, Flp est capable de réaliser des inversions-délétions intra- et intermoléculaires réversibles. De plus, elle ne nécessite pas de topologie particulière de l'ADN substrat, ni de cofacteur, faisant également de cette protéine un modèle des recombinases à tyrosine simple (Jayaram *et al.*, 2002).

II.7) Le système XerC/D, modèle des recombinases utilisant des protéines accessoires

Les protéines XerC et XerD (« Chromosomally Encoded Recombinase ») sont des recombinases à tyrosine découvertes chez *E. coli* et présentes dans la majorité des espèces. Elles sont impliquées dans la ségrégation du chromosome bactérien et de divers plasmides en résolvant les dimères d'ADN en monomères. La particularité de ce système provient de l'utilisation de deux recombinases différentes pour catalyser la réaction de recombinaison. Cette propriété n'est pas unique au système XerC/D parmi les recombinases à tyrosine, puisque le système d'inversion de gènes régulant l'expression des fimbriae de type I chez *E. coli* utilise les recombinases FimB et FimE (Klemm, 1986). Des homologues au système XerC/D sont retrouvés pratiquement chez toutes les bactéries et Archaea possédant un chromosome circulaire (Duggin *et al.*, 2011; Kono *et al.*, 2011; Recchia and Sherratt, 1999). Cependant, ce système de résolution de dimères de chromosome n'implique pas toujours deux recombinases différentes, notamment chez les lactocoques et les streptocoques où cette recombinaison est effectuée par une seule recombinase à tyrosine, XerS (Le Bourgeois *et al.*, 2007), et chez un sous-groupe de ϵ -protéobactéries incluant le genre *Helicobacter* et *Campylobacter* avec la protéine XerH (Carnoy and Roten, 2009).

Les protéines XerC et XerD sont des protéines de 298 AA partageant 37% d'identité de séquence (Blakely *et al.*, 1993). XerC a été la première des deux protéines à être découverte et fût impliquée dans la stabilité des plasmides à copies multiples en résolvant les dimères de plasmides en monomères par recombinaison au niveau d'un site spécifique *cer* (« ColE1 Resolution ») présent dans les plasmides ColE1 (Colloms *et al.*, 1990; Summers and Sherratt, 1984). Peu de temps après, cette protéine a été impliquée dans la résolution de dimères de chromosome en agissant au niveau des sites *dif* (« Deletion Induced Filamentation ») (Blakely *et al.*, 1991; Kuempel *et al.*, 1991). XerD a été découverte quelques années plus tard, ainsi que son rôle dans la recombinaison aux sites *cer* et *dif* (Blakely *et al.*, 1993; Lovett and Kolodner, 1991).

Etude bibliographique

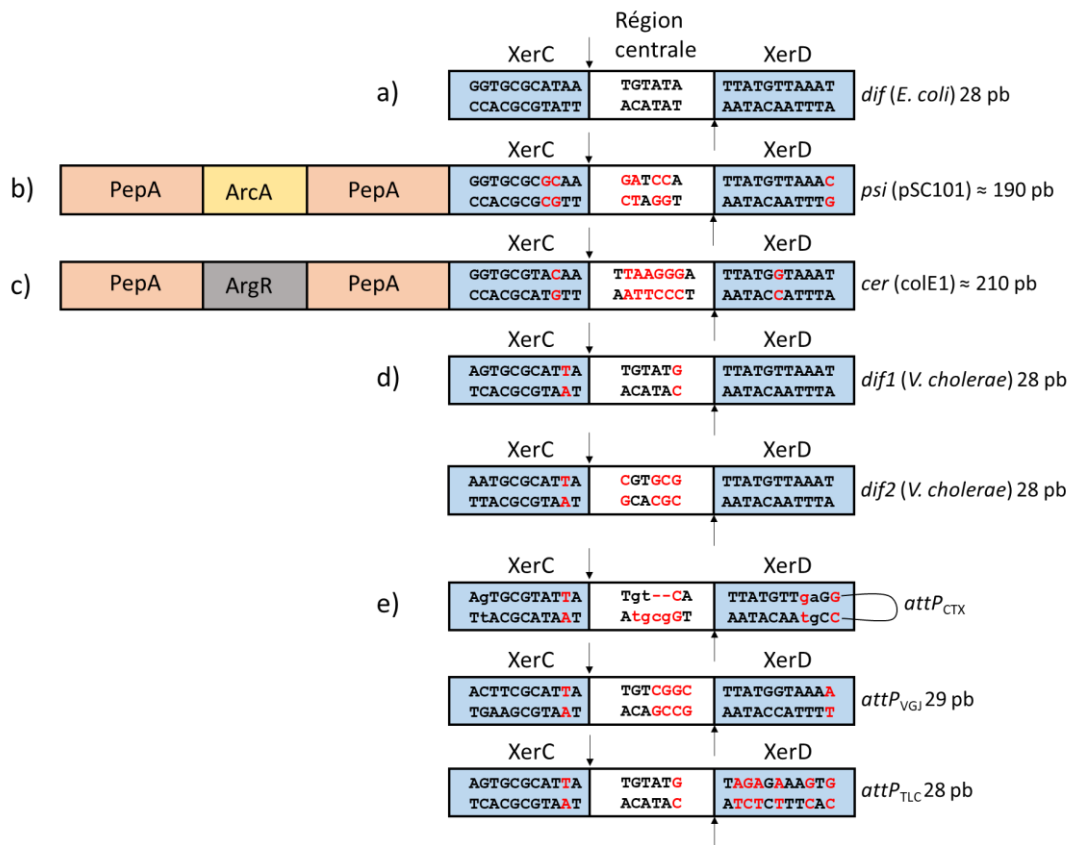


Fig. 15 Représentation schématique des sites de recombinaison du système XerC/D (Adapté de Midonet and Barre, 2014). Les rectangles bleus représentent les sites de fixation des recombinases XerC et XerD dans la région cœur. Les bases différentes de la séquence consensus issue de l'alignement de 715 sites *dif* sont représentées en rouge (Midonet and Barre, 2014). Les bases en minuscules sont des bases non Watson-Crick. Les rectangles orange, jaune, et gris représentent les sites de fixation des protéines accessoires PepA, ArcA, et ArgR, respectivement. Les sites de coupures sont représentés par les flèches noires verticales, encadrant la région centrale. **a)** Le site *dif* est simple et ressemble à la structuration des sites *loxP* et *FRT*. **b)** Les sites *psi* et *cer* sont composés d'une région cœur et d'une séquence d'ADN supplémentaire permettant la fixation des protéines accessoires. **c)** Le site *cer* se différencie par une région centrale plus grande, de 8 pb, par rapport à la région centrale canonique de 6 pb. **d)** Représentation schématique du site *dif1* du chromosome 1 et du site *dif2* du chromosome 2 de la bactérie *V. cholerae*. **e)** Représentation des sites *attP* des éléments génétiques mobiles intégratifs exploitant le système XerC/D. Le site *attP_{CTX}* présente une architecture atypique, issue du repliement d'un ADN simple brin en tige-boucle.

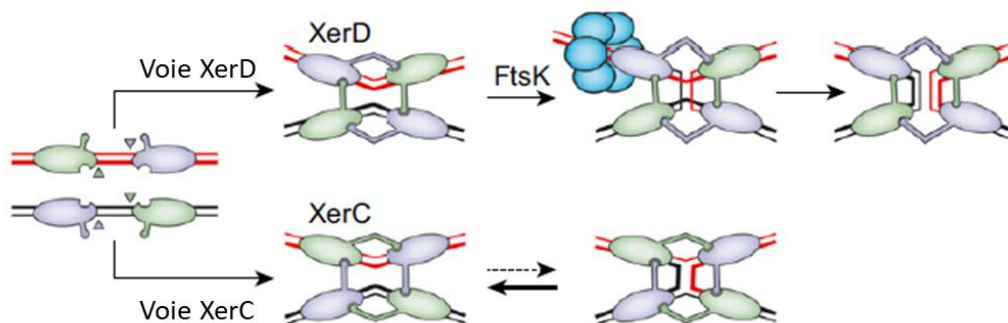


Fig. 16 Représentation schématique de la recombinaison du système XerC/D sur les sites *dif* (Adapté de Midonet and Barre, 2016). Les recombinases XerC et XerD sont représentées en vert et en gris, respectivement. L'hexamère de la protéine FtsK est représenté en bleu. Au niveau des sites *dif*, deux voies de recombinaison sont possibles : la voie XerC, qui clive le brin supérieur, mais cette jonction de Holliday ne peut pas être résolue par XerD qui est inactive et est donc résolue par un nouvel échange des brins supérieurs par la protéine XerC, redonnant les substrats initiaux. Dans la voie XerD, FtsK interagit directement avec la protéine XerD, permettant son activation. La protéine XerD activée permet la formation d'une jonction de Holliday par l'échange des brins inférieurs, qui est résolue par XerC en produits de recombinaison.

La résolution de dimères de plasmides par le système XerC/D a mené à la découverte du site *psi* (« pSC101 Stabilized Inheritance ») présent dans le plasmide pSC101 de la bactérie *Salmonella typhimurium* (Cornet *et al.*, 1994). La recombinaison Xer illustre comment un système de recombinaison spécifique de site peut être adapté pour accomplir différentes fonctions biologiques (résolution de dimères de chromosome, résolution de dimères de plasmide, intégration de bactériophage).

II.7.1) La recombinaison aux sites chromosomiques *dif*

La résolution de dimères de chromosomes est un processus important permettant une hérédité stable du matériel génétique. Chez la bactérie *E. coli*, les dimères de chromosomes sont résolus en monomères grâce à une recombinaison spécifique de site médiée par les recombinases XerC et XerD. Cette recombinaison est sous le contrôle d'une protéine ancrée au septum, la translocase à ADN ATP-dépendante, FtsK (Aussel *et al.*, 2002). Cette protéine permet la coordination de la division cellulaire avec la ségrégation des chromosomes. FtsK induit la translocation du chromosome, de façon orientée, en se fixant sur des petites séquences d'ADN, les sites KOPS (« FtsK Oriented Polar Sequence ») (Bigot *et al.*, 2006; Pease *et al.*, 2005). Ce contrôle directionnel guide la translocation du chromosome vers les sites *dif*, permettant leur association à l'intérieur du dimère de chromosomes. FtsK permet aussi l'activation de la recombinase XerD par contact direct. Le site *dif*, localisé dans la région terminale du chromosome a une organisation simple, ressemblant aux sites *loxP* et *FRT* (Fig. 15a). Il a une taille de 28 pb et est composé de deux sites de fixation de 11 pb, le site gauche étant le site de fixation de la recombinase XerC et le site droit étant le site de fixation de la recombinase XerD, encadrant une région centrale de 6 pb.

Les protéines XerC/D sont capables de recombinaison intramoléculaire sur des sites *dif* en orientation directe ou inverse et de recombinaison intermoléculaire. Cependant, en absence de la protéine FtsK, XerC initie la recombinaison par un clivage du brin supérieur, mais la jonction de Holliday formée ne peut pas être résolue en produit car la recombinase XerD est inactive et la jonction est donc résolue en substrat initial (Fig. 16). En présence de FtsK, le domaine γ de FtsK active la recombinase XerD (Grainge *et al.*, 2011) qui initie la formation de la jonction de Holliday par un clivage du brin inférieur (Aussel *et al.*, 2002). Cette jonction de Holliday est ensuite résolue en produit par la recombinase XerC permettant la résolution du dimère de chromosomes et la ségrégation des chromosomes dans les cellules filles (Fig. 16).

II.7.2) La recombinaison aux sites plasmidiques *cer* et *psi* : l'utilisation d'un filtre topologique

La formation de multimères de plasmides a un effet négatif sur la stabilité plasmidique en abaissant le nombre d'unités plasmidiques effectuant la ségrégation et, ainsi, augmentant la fréquence de perte de plasmides. Par conséquent, de nombreux plasmides possèdent un gène codant une résolvasse permettant la maintenance de leur état monomérique comme par exemple le système ParA/*res* du plasmide RK2 (Eberl *et al.*, 1994) ainsi que la famille des β recombinases dont les plus étudiées sont la protéine β du plasmide pSM19035 de *Streptococcus pyogenes* (Rojo and Alonso, 1994) et la protéine Res β du plasmide pAM β 1 d'*Enterococcus faecalis* (Petit *et al.*, 1995). Cependant, certains plasmides ne possèdent pas leur propre système de résolution de dimères et exploitent les recombinases Xer de leur hôte. C'est le cas des plasmides ColE1 (Colloms *et al.*, 1990) et des plasmides pSC101 (Cornet *et al.*, 1994). Ainsi, ces systèmes utilisent des séquences et des protéines accessoires comme filtre topologique afin de contrôler la réaction de recombinaison, permettant à ces plasmides de favoriser la réaction de résolution de dimères, tout en évitant leur formation.

Les sites *psi* (plasmide pSC101) et *cer* (plasmide ColE1) présentent des sites de fixation de XerC et XerD similaires à ceux du site *dif* encadrant une région centrale canonique de 6 pb, et une région centrale atypique de 8 pb respectivement (Fig. 15b ; Fig. 15c). En revanche, ces sites possèdent des séquences d'ADN accessoires de 160 pb et 180 pb, permettant la fixation des protéines PepA et ArcA, et PepA et ArgR respectivement (Alen *et al.*, 1997; Colloms *et al.*, 1996, Colloms *et al.*, 1998; Stirling *et al.*, 1988). Ces protéines et séquences accessoires forment un filtre topologique biaisant la réaction de recombinaison en impliquant la formation d'un complexe synaptique particulier dont la seule réaction de recombinaison possible est une recombinaison intramoléculaire entre deux sites en orientation directe, présent sur un ADN surenroulé (Colloms *et al.*, 1997).

Au niveau du site *psi*, une fois fixées, les protéines accessoires activent la recombinaison du système Xer : XerC réalise le premier échange de brin afin de former une jonction de Holliday, qui est résolue par XerD permettant la résolution du dimère de plasmides (Colloms *et al.*, 1996). Lorsque cette séquence accessoire est placée à côté du site de fixation de XerD, XerD réalise le premier échange de brin formant la jonction de Holliday, qui sera résolue par XerC. Cette séquence accessoire dicte donc l'ordre d'échange des brins et influence l'activité ou l'inactivité des recombinases Xer pour le premier échange de brin.

Cela suggère que les contraintes topologiques imposées par les protéines et séquences accessoires transforment le complexe synaptique en un complexe adapté afin que la protéine XerC réalise le premier échange de brin, et que la jonction de Holliday formée est dans une conformation adaptée pour que la recombinase XerD effectue le deuxième échange de brin (Bregu *et al.*, 2002). Au niveau du site *cer*, la recombinase XerC réalise le premier échange de brin, mais la jonction de Holliday ne peut pas être résolue par la recombinase XerD. L'incapacité de cette jonction de Holliday à être résolue semble liée à la structure atypique de la région centrale qui est de 8 pb. *In vivo*, la jonction de Holliday est résolue par des processus intracellulaires, indépendant du système XerC/D et il est proposé que cette jonction soit résolue par réplication, menant à la production de deux monomères de plasmide et d'un dimère de plasmides (Midonet and Barre, 2014).

II.7.3) Les éléments génétiques mobiles exploitant le système XerC/D

La majorité des éléments génétiques mobiles intégratifs (bactériophages, éléments mobiles conjugatifs, transposons...) codent leur propre recombinase permettant leur intégration dans le génome de leur bactérie hôte. Cependant, certains éléments génétiques mobiles intégratifs dépourvus de gènes codant une intégrase détournent la machinerie Xer bactérienne pour réaliser leur intégration dans le génome hôte. Ces éléments génétiques ont été nommés les IMEXs (« Integrative Mobile Element using Xer ») (Das *et al.*, 2013).

Afin de pouvoir utiliser la machinerie Xer, ces éléments possèdent un site « *attP* » similaire au site chromosomique *dif* (ayant le rôle d'un site *attB*). Après intégration des IMEXs, le site « *attR* » devient le nouveau site *dif* utilisé par la bactérie pour résoudre les dimères de chromosomes. Les IMEXs peuvent au moins être classés en trois catégories selon leur site *attP*. La bactérie *Vibrio cholerae* est une bactérie possédant deux chromosomes dont chaque chromosome présente un site *dif* unique ressemblant fortement au site *dif* de la bactérie *E. coli* (Fig. 15d). Cette bactérie modèle est infectée par des IMEXs appartenant à chacune des catégories et la plupart des informations obtenues sur les mécanismes moléculaires d'intégration des IMEXs proviennent d'études chez cet organisme. Par conséquent, les trois classes d'IMEXs actuellement caractérisées ont été nommées par rapport au modèle le plus étudié chez *V. cholerae* soit, CTX Φ , VGJ Φ et TLC Φ (Das *et al.*, 2013).

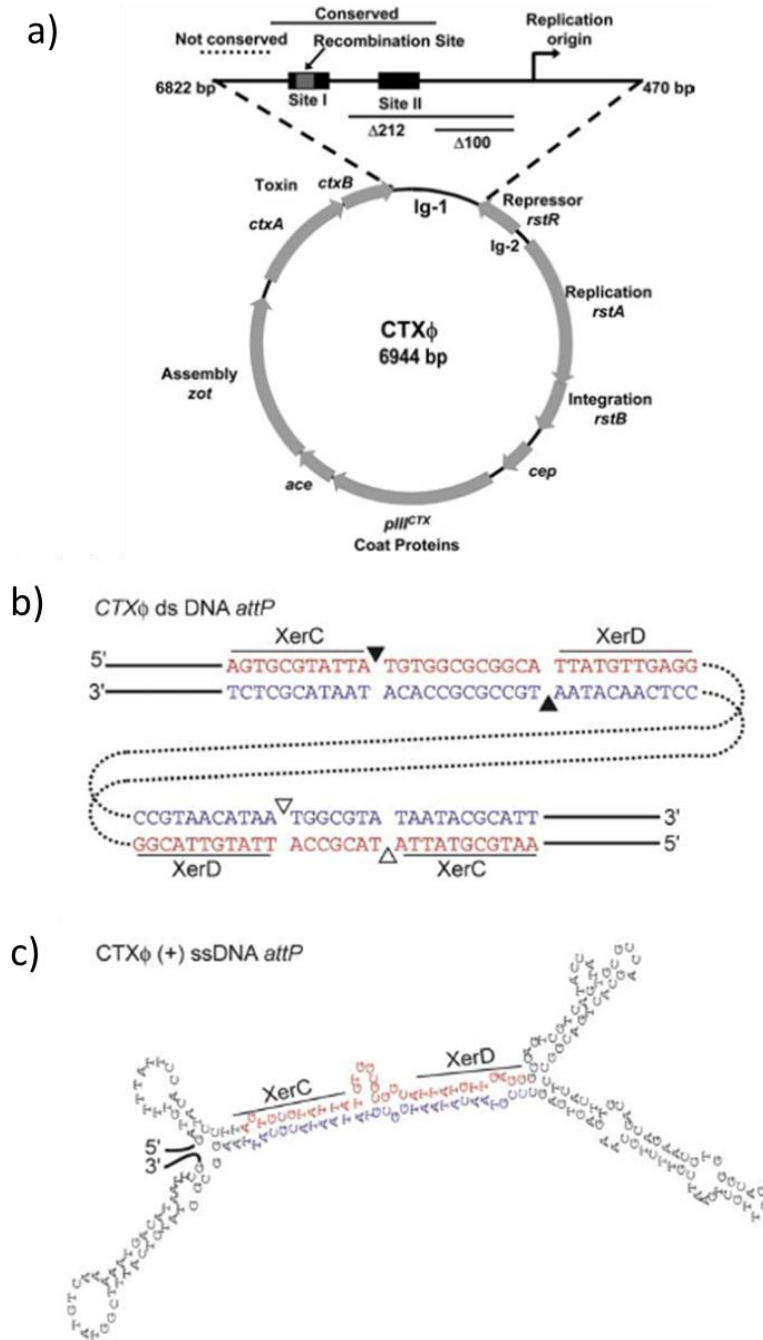


Fig. 17 Représentation schématique du génome et du site *attP* du bactériophage CTXΦ (Adapté de McLeod and Waldor, 2004 ; Val *et al.*, 2005). a) Représentation schématique du génome du bactériophage CTXΦ. Les flèches grises représentent les ORFs. La région intergénique 1 (Ig-1) a été agrandie et contient les deux sites similaires aux sites *dif* (rectangles noirs). **b)** Séquence de la région *attP* de la forme répllicative (double brin) du génome du bactériophage CTXΦ. Deux sites similaires aux sites *dif* sont retrouvés en orientation inverse et séparés par une séquence d'ADN d'environ 90 pb. Les sites de fixation de Xer sont indiqués par un trait noir et les sites de clivages sont représentés par des triangles. **c)** Structuration en tige-boucle de l'ADN simple brin positif du bactériophage CTXΦ.

II.7.3.1) Les IMEXs de type CTXΦ

CTXΦ (« Cholera Toxine phage ») est un phage filamenteux à ADN simple brin infectant la bactérie *V. cholerae* et codant la toxine cholérique. La forme répliquative du génome de ce bactériophage présente deux sites similaires aux sites *dif* (Fig. 17a ; Fig. 17b) en orientation inverse mais aucun des deux sites ne permet l'intégration du bactériophage dans le site *difI* chromosomique. En parallèle, la découverte du mécanisme de recombinaison des cassettes d'intégrons (cf partie II.8) a facilité la compréhension du mécanisme d'intégration du phage CTXΦ. A la suite de ces découvertes, il a été déterminé que la forme intégrative du bactériophage n'était pas un génome double brin, mais plutôt un ADN simple brin positif, replié en tige-boucle permettant la formation d'un site *attP*_{CTX} (Fig. 15e, Fig. 17c) (Val *et al.*, 2005). Une des particularités de ce site est la présence d'une boucle de bases non appariées dans la région centrale.

La réaction d'intégration est catalysée par la recombinase XerC qui permet la formation d'une jonction de Holliday (Fig. 18). L'activité catalytique de XerD n'est pas nécessaire à la recombinaison, mais sa présence est indispensable indiquant un rôle structural important dans la formation du complexe synaptique (Val *et al.*, 2005). Un second échange de brin médié par la recombinase XerD pour résoudre cette jonction de Holliday linéariserait le chromosome et serait létal pour la bactérie, alors qu'un second échange de brin effectué par la protéine XerC entraînerait une résolution de la jonction de Holliday vers les substrats initiaux, empêchant l'intégration du bactériophage CTXΦ. Il a été montré que la protéine Endo III, une protéine de la réparation de l'ADN par excision de base, stabilise cette jonction de Holliday et empêche sa résolution en substrat par XerC (Bischerour *et al.*, 2012). Enfin, comme pour la résolution des dimères de plasmides au niveau du site *cer*, il est proposé que cette jonction de Holliday soit résolue en produits de recombinaison par répliquaison de l'ADN (Val *et al.*, 2005). L'intégration du bactériophage CTXΦ est donc irréversible. En effet, après intégration, la forme double brin du prophage masque de nouveau la forme active du site *attP*, empêchant la réaction d'excision avec le site *dif* (*attR*) (Fig. 18). Cependant, des copies du génome libre du bactériophage CTXΦ peuvent être produites par un mécanisme analogue à la répliquaison en cercle roulant, dans le cas d'intégration multiples du bactériophage (intégration en tandem) (Moyer *et al.*, 2001).

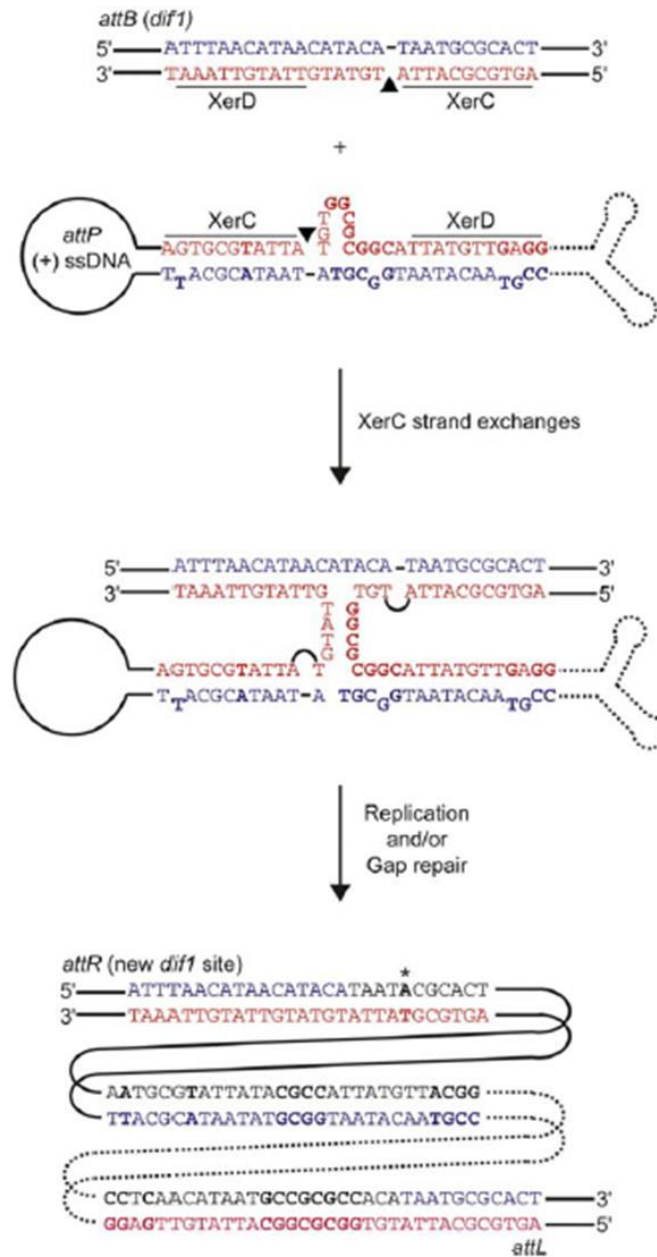


Fig. 18 Représentation schématique de la réaction d'intégration du bactériophage CTXΦ (Val et al., 2005). La réaction de recombinaison a lieu entre le site à ADN double-brin *dif1* (*attB*) de la bactérie *V. cholerae* et l'ADN simple-brin positif, replié en tige-boucle, du bactériophage CTXΦ. Les bases du site *attP* simple-brin qui sont différentes du site *dif1* sont indiquées en gras. Les sites de coupures de XerC sont indiqués par un triangle noir. La réaction de recombinaison s'arrête après un premier échange de brin catalysé par XerC. Il est proposé que la jonction de Holliday est convertie en produit ADN double-brin par des processus cellulaires comme la réparation ou réplication de l'ADN. Au final, l'intégration du bactériophage supprime le site *dif1* et le prophage CTXΦ est encadré par un site *attL* différent et un site *attR* qui diffère d'une paire de base (astérisque) par rapport au site *dif1*. Le site *attR* sert de nouveau site *dif* chromosomique pour la résolution de dimères de chromosomes.

II.7.3.2) Les IMEXs de type VGJΦ

Le bactériophage VGJΦ est un bactériophage filamenteux mais, à la différence du bactériophage CTXΦ, le site *attP*_{VGJ} est retrouvé dans la forme répliquative d'ADN double brin (Fig. 15e). L'intégration de ce bactériophage est médiée par la formation d'une jonction de Holliday dont le premier échange de brin est catalysé par la protéine XerC. Cette jonction de Holliday est probablement résolue en produit de recombinaison par répliquaison de l'ADN. Cependant, l'intégration du bactériophage VGJΦ crée deux sites de recombinaison *attL* et *attR* fonctionnels. L'excision de ce bactériophage est donc possible lorsque la recombinase XerC induit un échange de brin au niveau des sites *attL* et *attR*, formant une jonction de Holliday, résolue par répliquaison de l'ADN en une copie double brin du génome du phage VGJΦ, un chromosome dépourvu du prophage, et un chromosome portant la séquence du prophage (Das *et al.*, 2011).

II.7.3.3) Les IMEXs de type TLCΦ

Le bactériophage TLCΦ (« Toxin-Linked Cryptic phage ») est un bactériophage satellite. Les bactériophages satellites sont des prophages qui ne possèdent pas les gènes codant les composants structuraux, et qui dépendent d'un bactériophage « helper » pour leur survie. Le site *attP*_{TLC} est présent dans la forme répliquative d'ADN double brin de ce virus et présente la particularité d'avoir une région centrale identique à celle du site *difI* et une séquence de fixation de XerD dégénérée (Fig. 15e). Cette séquence dégénérée est un site de fixation de faible affinité de XerD. Cependant, tout comme la résolution de dimères de chromosomes, la jonction de Holliday est formée grâce à un premier échange de brin médié par XerD et résolue par XerC. En revanche, cette activation de XerD se fait indépendamment de la translocase à ADN FtsK (Midonet *et al.*, 2014). A la place, le bactériophage TLCΦ code une protéine XafT (« Xer Activation Factor of TLC »), qui active les fonctions de recombinaison de XerD en interagissant directement avec cette protéine. L'intégration du bactériophage TLCΦ donne lieu à la formation de sites *attL* et *attR* fonctionnels, et ce prophage peut être excisé par formation d'une jonction de Holliday induite par la recombinase XerD activée par XafT, résolue en produit par XerC (Midonet *et al.*, 2019).

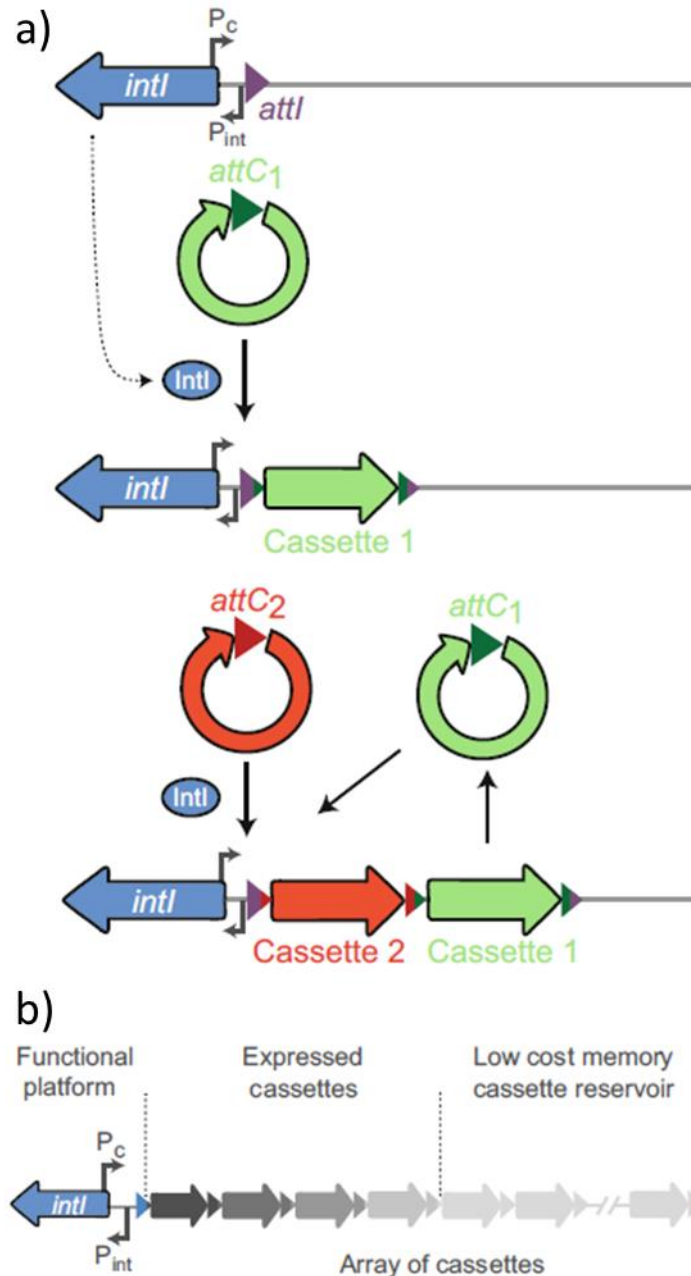


Fig. 19 Organisation d'un intégron (Adapté de Escudero et al., 2018). a) Insertion et excision de cassettes. La plateforme fonctionnelle de l'intégron est composée d'un promoteur P_{int} permettant l'expression du gène *IntI* codant une intégrase (flèche bleue), d'un site *attI* (triangle violet) et d'un promoteur P_c permettant l'expression des cassettes intégrées. Le site *attC* présent sur les cassettes est représenté par un triangle vert ou rouge en fonction de la cassette. Les flèches indiquent l'orientation de l'ORF. L'intégration (*attI/attC*) et l'excision de la cassette (*attC/attC*) sont catalysées par l'intégrase *IntI*. b) Expression des cassettes au sein de l'intégron. Les flèches grises représentent les différentes cassettes de gènes intégrées. Le taux d'expression de chaque gène est représenté par l'intensité de la couleur grise sur chaque flèche. Seules les premières cassettes sont exprimées, et les cassettes non exprimées peuvent être considérées comme constituant une réserve.

II.8) Les intégrons : un mécanisme de recombinaison atypique

Les intégrons ont été découverts à la fin des années 1980 pour leur rôle dans la dissémination de la résistance aux antibiotiques chez les bactéries à Gram négatif (Stokes and Hall, 1989). Au départ, les intégrons étaient retrouvés associés à des éléments transposables et des plasmides conjugatifs permettant leur dispersion et ont été nommés « Mobile Integrons » (MIs). Les MIs sont différenciés en 5 classes en fonction de la séquence de leur intégrase (Mazel, 2006). Les MIs de classe 1 et 3 sont associés avec des transposons dérivés de Tn402. Les MIs de classe 2 sont exclusivement associés avec des transposons dérivés de Tn7. Les MIs de classe 4 et 5 sont retrouvés dans les éléments SXT de *V. cholerae*. Les MIs présentent une petite taille, le plus grand réseau de cassette décrit étant composé de 8 gènes (Naas *et al.*, 2001) et la majorité des cassettes codent pour des fonctions de résistance aux antibiotiques. Des années plus tard, des intégrons de taille plus importante ont été découverts dans les chromosomes bactériens, dont le « superintegron » de la bactérie *V. cholerae* (Mazel *et al.*, 1998). Ces intégrons localisés en dehors d'éléments génétiques mobiles ont été nommés « Sedentary Chromosomal Integrons » (SCIs). A la différence des MIs, ces SCIs sont généralement des intégrons de grandes tailles pouvant posséder plus de 200 cassettes. Ils sont retrouvés dans de nombreux génomes bactériens, puisqu'en 2010, une étude montrait que 17% des génomes séquencés possédait une intégrase d'intégron, suggérant la présence de SCI (Cambray *et al.*, 2010). Les « superintegrons » peuvent stocker de nombreuses cassettes et représenter un pourcentage significatif du chromosome bactérien. Par exemple, le « superintegron » modèle de la bactérie *V. cholerae* possède 175 cassettes et représente 3% du chromosome bactérien. La majorité des cassettes présentes au sein des SCIs n'ont pas de fonction connue.

Les intégrons sont donc des plateformes génétiques qui permettent une évolution rapide des bactéries par l'acquisition, l'excision, le stockage, et le réarrangement de séquences codantes (« Coding DNA Sequence » ou CDS) provenant d'éléments mobiles nommés cassettes (Escudero *et al.*, 2015). Tous les intégrons partagent une structure commune composée d'une plateforme stable et d'un réseau de cassettes variable. La plateforme est composée d'un promoteur P_{int} permettant l'expression du gène *intI* codant une intégrase menant à l'intégration et l'excision de cassettes, d'un site de recombinaison « integron attachment site » *attI* où les cassettes de gènes sont intégrées et d'un promoteur P_c permettant l'expression de ces cassettes (Fig. 19a) (Mazel, 2006). Le réseau de cassettes est variable, allant de 0 à 219 cassettes pour le SCI de la bactérie *Vibrio vulnificus* CMCP6 (Boucher *et al.*, 2006).

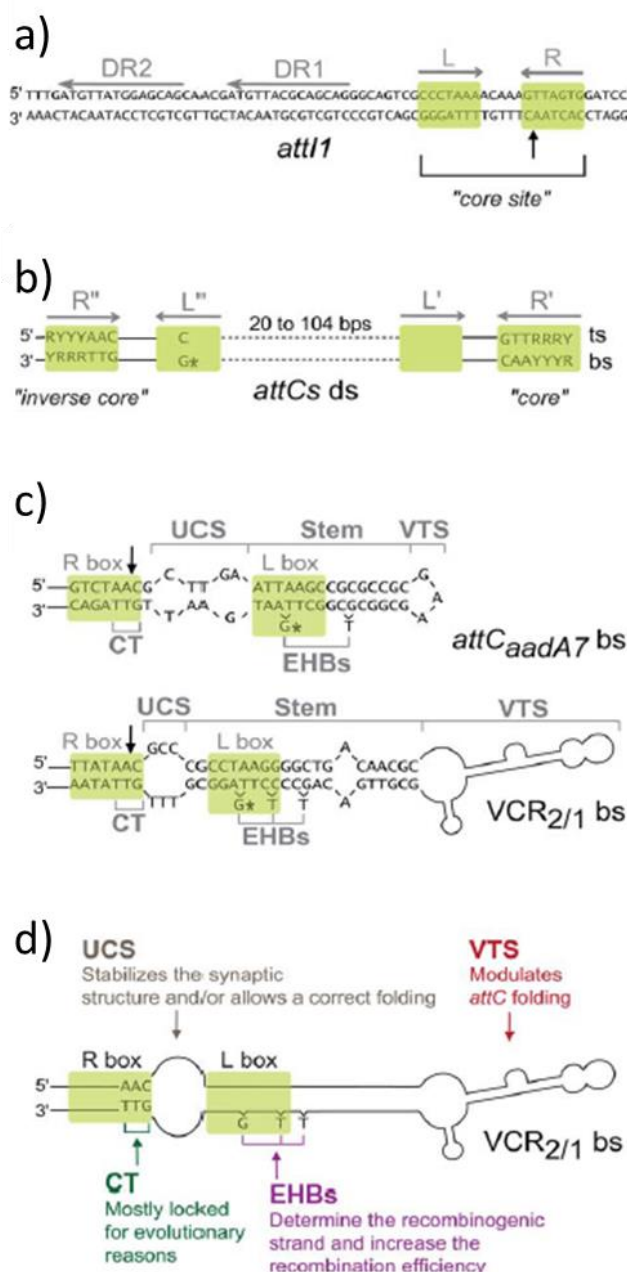


Fig. 20 Sites de recombinaison des intégrons (Escudero *et al.*, 2015). Les sites de l'intégrase *IntI1* sont représentés par des carrés vert. La flèche noire indique le site de clivage. **a)** Séquence du site double brin *attI1*. Les séquences inversées répétées (L et R), et les répétitions en orientation directe (DR1 et DR2) de fixation de *IntI1* sont indiquées par des flèches grises. **b)** Représentation schématique des sites double-brin *attC*. Les séquences inversées répétées (R'', L'', L', R') sont représentées par des flèches grises. Les lignes en pointillés représentent la partie centrale variable. Les nucléotides conservés sont indiqués dans le schéma. Les astérisques montrent le nucléotide G conservé, générant une « ExtraHelical Base » (EHB) dans le brin inférieur du site *attC* courbé. Le brin supérieur est indiqué par « ts » et le brin inférieur est indiqué par « bs ». **c)** Structure secondaire proposée pour les sites *attCaadA7* et *VCR2/1*. Les structures caractéristiques du site *attC* courbé sont indiquées : UCS, « Unpaired Central Segment » ; EHBs, « ExtraHelical Bases » ; « Stem », tige de la tige-boucle ; VTS, « Variable Terminal Structure » et « CT », triplet de nucléotide conservé. **d)** Représentation schématique des caractéristiques structurales du site *VCR2/1*. Le rôle de chaque structure est indiqué.

Les cassettes sont des éléments circulaires non répliatifs (Collis and Hall, 1992a), contenant généralement un CDS sans promoteur et un site de recombinaison « cassette attachment site » *attC* originellement appelé « 59-base element ». Ces CDS deviennent fonctionnels dès leur intégration à l'intérieur du site *attI* où le promoteur P_c permet leur expression (Fig. 19a ; Fig. 19b). L'intégration de différentes cassettes induit un gradient d'expression de ces gènes. En effet, les cassettes étant le plus proche du promoteur sont les plus exprimées et les plus éloignées sont les moins exprimées, jusqu'à ce qu'il n'y ait plus d'expression (Fig. 19b). Une cassette peut être excisée et réintégrée dans le site *attI* par l'intégrase IntI, permettant de nouveau leur expression maximale (Fig. 19a).

II.8.1) Le système de recombinaison des intégrons

II.8.1.1) Le site *attI*

Le site *attI* présent dans l'intégron est le site d'intégration des nouvelles cassettes. Ce site ressemble aux sites classiques des recombinases à tyrosine. En effet, le site minimum *attI* se trouve sous la forme double brin et est composé de deux sites de fixation de l'intégrase en orientation inverse, les sites L et R, formant le site cœur (Fig. 20a). Certains sites *attI* comme dans le système IntII/*attII*, présentent des séquences accessoires permettant la fixation de deux monomères d'intégrase IntII supplémentaires (Fig. 20a) (Gravel *et al.*, 1998). Le rôle de ces séquences accessoires n'est pas complètement connu. Elles ne sont pas indispensables, mais leur présence améliore la réaction de recombinaison au site *attII* d'au moins un facteur 25 lors d'une recombinaison *attII/attC* et d'un facteur 10 lors d'une recombinaison *attII/attII* (Partridge *et al.*, 2000). La recombinaison a lieu au niveau du triplet conservé 5'-GTT-3' situé dans le site R et le site de clivage permettant l'échange de brin a lieu entre le nucléotide A et C au niveau du brin inférieur (Fig. 20a).

II.8.1.2) Le site *attC*

Les sites *attC* sont retrouvés dans les cassettes d'ADN et présentent une structuration différente des sites classiques des recombinases à tyrosine. Sur l'ADN double-brin de la cassette, le site *attC* présente une structuration anormale : deux régions inversées, R''-L'' et L'-R', sont retrouvées et séparées par une région centrale de séquence et de taille variable (Fig. 20b). Ainsi, les sites *attC* présentent peu de similarité de séquence et ont une taille allant de 57 à 141 pb (Stokes *et al.*, 1997). Par comparaison de séquences des sites *attC*, seul le triplet de nucléotides « 5'-GTT-3' » et « 5'-AAC-3' » localisé dans le site R', et R'', respectivement, est conservé. Cette conservation de séquence peut être étendue à un motif inversé répété de 7 pb : « 5'-GTTRRRY-3' » / « 5'-RYYAAC-3' » (R = purine, Y = pyrimidine) (Fig. 20b). Malgré la très faible identité de séquence entre les sites *attC*, ces sites présentent une organisation palindromique permettant la formation d'une structure secondaire par repliement d'un ADN simple-brin en tige-boucle (Fig. 20c). Il a été montré par expérience de retard sur gel que l'intégrase se fixe sur les sites *attC* uniquement sous forme d'ADN simple-brin, avec un fort biais pour le brin inférieur (Francia *et al.*, 1999; Johansson *et al.*, 2004). Par conjugaison, Bouvier *et al.*, ont révélé que le brin inférieur était le substrat préférentiel pour la recombinaison, favorisant une recombinaison d'un facteur 1000 par rapport au brin supérieur (Bouvier *et al.*, 2005). Lors de cette étude, l'importance de la structuration de l'ADN simple-brin a été mise en évidence, validant que le substrat reconnu par l'intégrase était un ADN simple-brin structuré en tige-boucle. Après repliement par hybridation des sites R'-R'' et L'-L'', le triplet de nucléotides conservé 5'-AAC-3' est la région où l'échange de brin a lieu. La séquence du triplet nucléotidique conservé (GTT/AAC) entre les sites *attI* et *attC* n'est pas essentielle à la recombinaison, seule l'identité de séquence du premier nucléotide entre les deux sites de recombinaison l'est (Frumerie *et al.*, 2010).

Une différence majeure par rapport aux sites de recombinaison canoniques des recombinases à tyrosine provient du fait que la forme active du site n'est pas contenue dans la séquence primaire du site *attC* mais plutôt dans sa structure secondaire. En effet, malgré l'absence de similarité de séquence, trois caractéristiques structurales sont communes à tous les sites *attC* (Bouvier *et al.*, 2009). Tout d'abord, après hybridation des sites R'-R'' et L'-L'', 2 à 3 nucléotides situés dans la région R''-L'', nommés « ExtraHelical Bases » (EHBs) (Fig. 20d), ne possèdent pas de bases complémentaires sur la région R'-L' et ressortent de la structure en tige-boucle.

Ces EHBs ont un rôle important pour la recombinaison puisqu'elles déterminent le brin qui sera recombiné (le brin inférieur), permettant ainsi l'intégration de la cassette de gènes dans un sens fonctionnel. En effet, l'intégration d'une cassette de gènes par recombinaison du brin supérieur aboutirait à l'insertion d'un CDS dans le sens opposé au promoteur P_c , ainsi, le brin non codant serait transcrit, empêchant l'expression correcte du gène de la cassette (Bouvier *et al.*, 2009; Nivina *et al.*, 2016). Ces EHBs permettent aussi de stabiliser le complexe nucléoprotéique, grâce à des contacts entre les EHBs et les monomères d'intégrase. Enfin, ces bases induisent un changement conformationnel éloignant la tyrosine catalytique des monomères d'intégrases fixés sur le site L'-L'' (site L), empêchant ainsi un deuxième échange de brin qui linéariserait le produit de recombinaison (MacDonald *et al.*, 2006).

La deuxième caractéristique structurale des sites *attC* est la région « Unpaired Central Segment » (UCS) situé entre les sites L et R. Cette région provient de l'hybridation imparfaite des nucléotides situés entre les sites R'' et L'' avec les nucléotides situés entre les sites R' et L'. Cette région UCS joue un rôle dans la sélection du brin inférieur par stabilisation de la structure formée par ce brin et est essentielle pour obtenir un haut niveau de recombinaison, suggérant un rôle dans l'exclusion du site actif des intégrases et/ou en empêchant la réaction inverse après le premier échange de brin (Bouvier *et al.*, 2009; Nivina *et al.*, 2016).

Enfin, la dernière caractéristique structurale est la « Variable Terminal Structure » (VTS). Cette région correspond à l'extrémité de la tige et de la boucle formée par la structuration du site *attC*. La région VTS joue un rôle dans la sélection du brin inférieur par stabilisation de la structure formée par ce brin (Nivina *et al.*, 2016). Cette structure a aussi un impact sur la recombinaison par modulation du repliement du site *attC*. En effet, la taille et la composition nucléotidique de la région VTS déterminent la capacité du site *attC* à se replier, notamment lorsqu'il se trouve sous forme d'ADN double-brin. Sous cette forme, l'extrusion de structure cruciforme de l'ADN double-brin permet au site *attC* de former une structuration en tige-boucle du brin supérieur et inférieur. Ce phénomène est fortement impacté par des régions VTS de taille importante, possédant une énergie de repliement très défavorable (Loot *et al.*, 2010).

II.8.1.3) L'intégrase IntI

Les intégrases d'intégrons appartiennent à la famille des recombinases à tyrosine et présentent la structure « C-shaped clamp » typique de ces protéines. Cependant, ces intégrases présentent un domaine supplémentaire de 19 AA dans leur domaine catalytique contenant une hélice- α , nommée I2, spécifique aux intégrases d'intégrons et essentielle à l'activité de ces protéines (Messier and Roy, 2001). La structure cristalline du complexe synaptique de l'intégrase du « superintégron » de *V. cholerae* (VchIntIA) avec ces deux sites *attC* (VCR_{bs}) a été obtenue et révèle que l'hélice- α I2 interagit avec les EHBs du site *attC* permettant la stabilisation de cette synapse (MacDonald *et al.*, 2006).

II.8.2) Les réactions de recombinaison des intégrons

Les intégrases d'intégrons IntI sont capables de recombinaison entre un site *attC* et un site *attI*, entre deux sites *attC*, et entre deux sites *attI* (Collis *et al.*, 2001).

II.8.2.1) La recombinaison *attC/attI*

La recombinaison *attC/attI* est la réaction de recombinaison la plus efficace et permet l'intégration de la cassette au niveau du site *attI* plutôt qu'à un autre site *attC* dans le réseau de cassettes, permettant ainsi l'expression du nouveau gène intégré (Collis *et al.*, 1993; Collis *et al.*, 2001). Par exemple, chez le SCI de *V. cholerae*, les cassettes sont préférentiellement intégrées à l'intérieur du site *attIA* malgré la présence de 180 autres sites *attC* (Vit *et al.*, 2021). De par l'efficacité de la réaction *attC/attI*, il semblerait possible que l'intégrase excise les cassettes en première position du réseau de cassettes, aboutissant à la perte des différentes cassettes acquises. Cependant, la nature différente de ces sites rend peu probable le fait qu'un site *attC* sous sa forme simple brin et qu'un site *attI* double brin soient suffisamment proche dans la même molécule afin de permettre l'excision de la première cassette. Ainsi, l'équilibre de la réaction *attC/attI* est déplacé vers une réaction intermoléculaire plutôt qu'intramoléculaire, c'est-à-dire vers l'intégration plutôt que vers l'excision (Escudero *et al.*, 2015). La réaction d'intégration de cassette est atypique puisqu'elle implique un site *attI* double brin et un site *attC* simple brin (Fig. 21a).

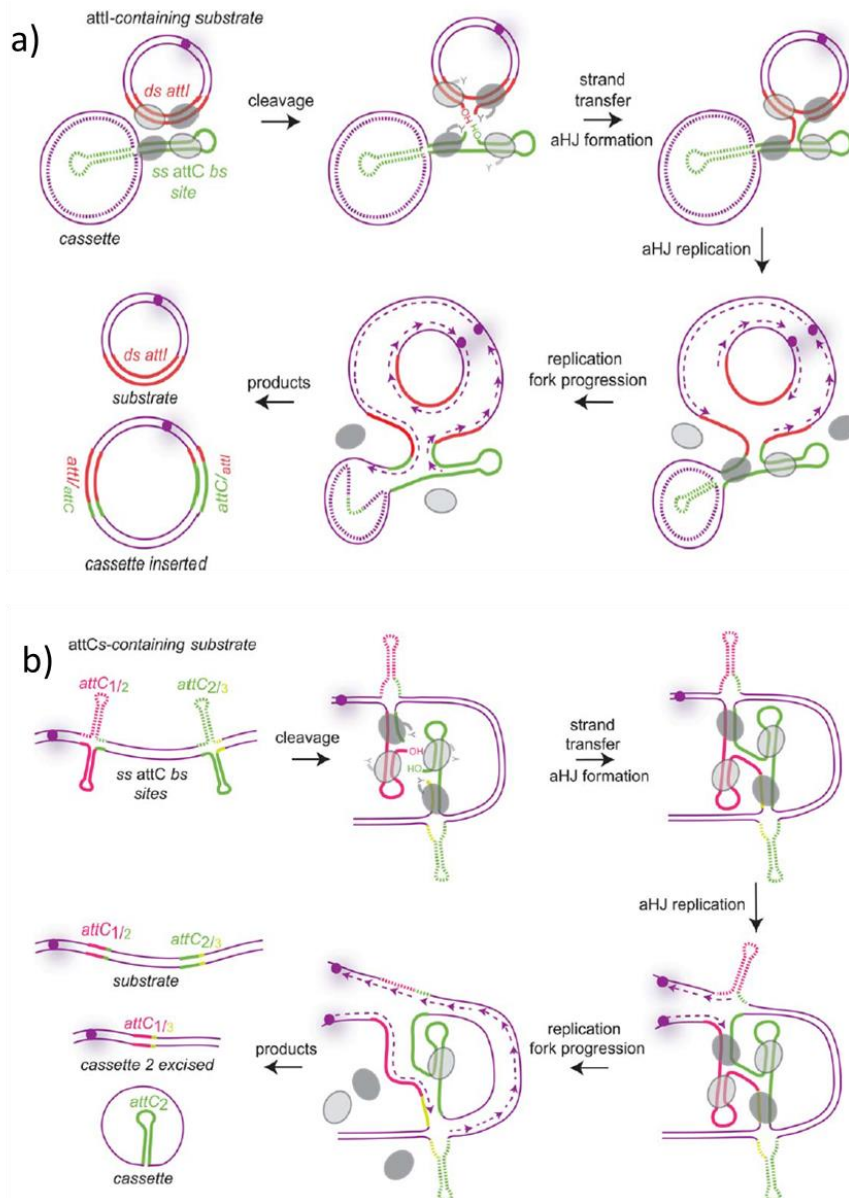


Fig. 21 Représentation des réactions de recombinaison *attC/attI* et *attC/attC* (Adapté de Escudero *et al.*, 2015). Les intégrases sont représentées par des formes ovales gris foncés pour les recombinases actives et gris clairs pour les recombinases inactives. Les cercles violets représentent l'origine de réplication et les nouveaux brins synthétisés (brins continus et discontinus) sont représentés par des pointillés violets. **a)** Intégration des cassettes par recombinaison entre un site *attC* simple-brin inférieur structuré en tige-boucle (vert) et un site *attI* double-brin (rouge). Le brin supérieur du site *attC* est indiqué en pointillé et la nature exacte de la cassette (simple-brin ou double-brin) n'est pas connu. Le complexe synaptique est composé de 4 sous-unités d'intégrases et un brin de chaque site est coupé et échangé afin de former une jonction de Holliday atypique (aHJ). Cette jonction de Holliday est résolue en produit de recombinaison par réplication de l'ADN menant à la formation d'une molécule n'ayant pas intégré la cassette et d'une molécule ayant intégré la cassette. **b)** Excision des cassettes par recombinaison *attC/attC*. La recombinaison a lieu entre deux sites brins inférieurs *attC* structurés en tige-boucle (vert et rose). Les brins supérieurs sont représentés en pointillés. Le complexe synaptique est composé de 4 sous-unités d'intégrases et de deux sites *attC* simple-brin. Les deux intégrases actives réalisent la coupure et l'échange de brin, formant une jonction de Holliday atypique. Il est proposé que cette jonction de Holliday est résolue par réplication de l'ADN. Après réplication, les produits de recombinaison obtenus sont : la molécule originale contenant les deux cassettes, la molécule ne contenant plus la cassette 2 et la cassette 2 excisée.

Cette réaction de recombinaison s'effectue par un échange de brin, permettant la formation d'une jonction de Holliday atypique. Cette jonction de Holliday ne doit pas être résolue par un deuxième échange de brin puisque cela linéariserait le produit de recombinaison. A la place, cette jonction de Holliday est résolue par réplication, aboutissant à l'intégration de la cassette dans un des deux produits de la réplication (Loot *et al.*, 2012).

II.8.2.1) La recombinaison *attC/attC*

La réaction de recombinaison entre deux sites *attC* localisés dans le même réseau de cassettes de gènes aboutit à l'excision d'une cassette sous forme d'ADN double brin (Collis and Hall, 1992a; Collis and Hall, 1992b). A l'intérieur du réseau de cassettes, les sites *attC* sont présents sous la forme d'ADN double brin et doivent être présent sous leur forme repliée soit par extrusion cruciforme de l'ADN, soit sous forme simple brin lors de la réplication de l'ADN, dans le fragment d'Okazaki (Fig. 21b). Une fois que les deux sites *attC* repliés sont recrutés par l'intégrase, un premier échange de brin aboutit à la formation d'une jonction de Holliday atypique. Ensuite, il semblerait que cette jonction de Holliday soit résolue par réplication, menant à l'excision de la cassette de gènes dans l'une des deux molécules formées lors de la réplication.

II.8.2.3) La recombinaison *attI/attI*

La recombinaison entre deux sites *attI* est possible, mais cette réaction est très peu efficace par rapport à une réaction entre un site *attI* et un site *attC* (Collis *et al.*, 2001). Cette recombinaison ne semble pas avoir d'avantage puisque le seul moment où les deux sites *attI* peuvent effectuer une réaction de recombinaison est lors de la réplication des chromosomes où la recombinaison aboutirait à la formation d'un dimère de chromosomes. Par conséquent, cette réaction entre deux substrats double-brin semble être le reliquat de l'activité des ancêtres des intégrases d'intégron. Comme les intégrases d'intégron effectuent leurs réactions de recombinaison avec un seul échange de brin où la jonction de Holliday produite est résolue par réplication, il est possible que ces protéines aient perdu la capacité à réaliser le deuxième échange de brin résolvant la jonction de Holliday.

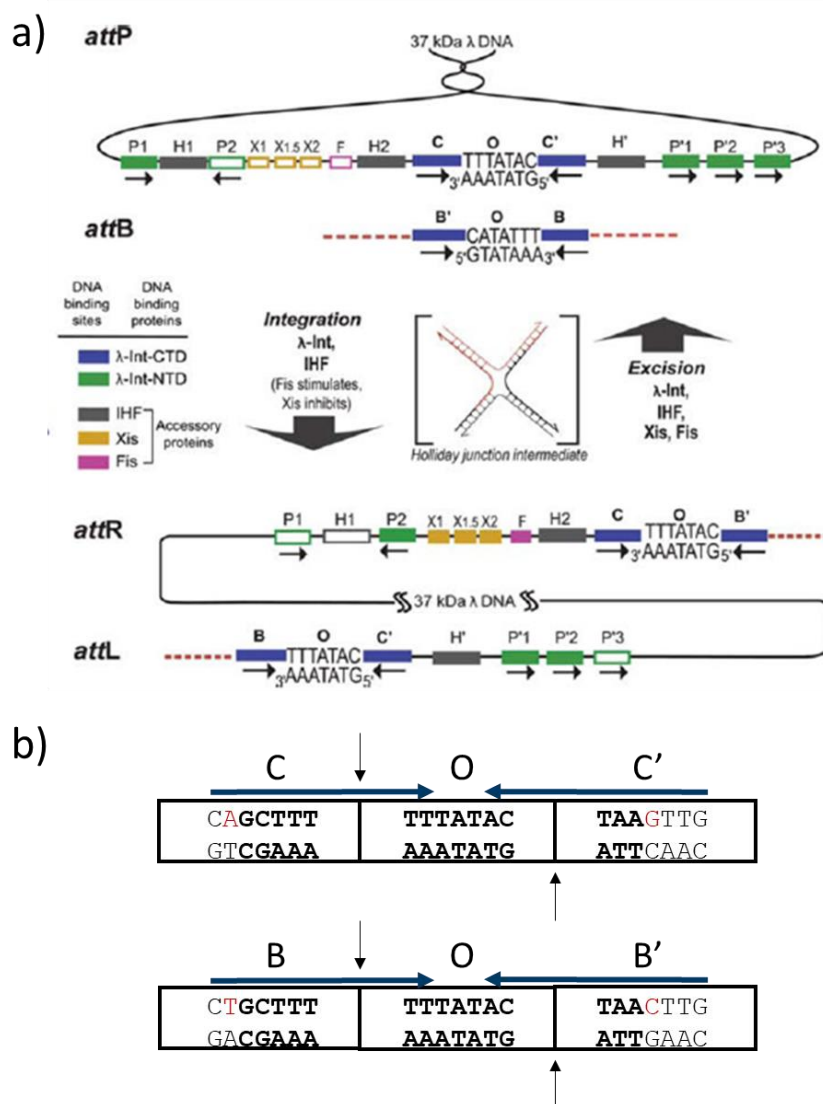


Fig. 22 Voies de recombinaison de l'intégrase λ et structuration des sites (Adapté de Landy, 2015). a) L'intégration s'effectue entre le site *attP* et *attB* et nécessite l'intégrase λ , le facteur d'hôte IHF et est stimulée par la protéine Fis. La recombinaison excisive s'effectue entre les sites *attL* et *attR* afin de régénérer les sites *attB* et *attP* et nécessite en plus l'excisionase Xis. Les deux réactions s'effectuent par la formation et la résolution d'une jonction de Holliday. Les deux réactions utilisent un jeu de sites de fixation protéique différents sur les bras P et P', indiquée par un rectangle coloré pour les sites utilisés, et un rectangle vide pour les sites non utilisés. Les sites bras (sites de forte affinité) de l'intégrase sont représentés par des rectangles vert, les sites cœur (sites de faible affinité) de l'intégrase sont représentés par des rectangles bleu, les sites de l'excisionase par des rectangles jaune, le site de fixation de Fis par un rectangle rose, et les sites de fixation de IHF par des rectangles gris. b) Représentation des sites cœur du site *attP* et *attB*. Les sites cœur B, B', C et C' de fixation de l'intégrase sont représentés par une flèche bleue. La région centrale (« overlap ») est indiquée par la lettre O encadrée par deux flèches noires verticales représentant les deux sites de coupures. Les bases différentes entre *attB* et *attP* sont indiquées en rouge. La région d'identité nucléotidique entre les sites *attB* et *attP*, indiquée en gras, est de 15 pb. Le site *attB* minimum est de 21 pb.

Ainsi, il a été montré que la recombinaison entre deux sites *attI* double-brin pouvait être réalisée soit par une seule coupure de brin, suivie de la résolution de la jonction de Holliday par réplication, soit en suivant la voie de recombinaison classique des recombinases à tyrosine, c'est-à-dire un premier échange de brin formant la jonction de Holliday, suivi d'un deuxième échange de brin permettant sa résolution, suggérant que cette activité de recombinaison est la réaction ancestrale des intégrases d'intégrons (Escudero *et al.*, 2016).

II.9) Le paradigme des recombinases hétérobivalentes : l'intégrase λ

Les intégrases hétérobivalentes sont des recombinases possédant un domaine supplémentaire de liaison à l'ADN permettant à ces protéines de se fixer sur deux séquences d'ADN différentes. Par soucis de clarté, le nom des différentes intégrases hétérobivalentes sera abrégé ^XInt (avec X étant le nom de l'élément génétique mobile dont est issue la protéine). En 1963, Allan Campbell propose un modèle pour l'intégration et l'excision du bactériophage λ dans le chromosome de sa bactérie hôte *E. coli* (Campbell, 1963). Des années plus tard, la purification de l'intégrase λ par Howard Nash (Nash, 1974a), la mise au point d'un système de recombinaison *in vitro* (Nash, 1975), ainsi que la découverte et la purification du facteur d'hôte IHF (« Integration Host Factor ») par Nash et Robertson (Nash and Robertson, 1981) ont permis à ce système de recombinaison d'être le premier à être méticuleusement étudié, et à l'intégrase λ d'être le membre fondateur de la famille des recombinases à tyrosine, originellement appelé « $^\lambda\text{Int family}$ ».

II.9.1) Composants du système de recombinaison

II.9.1.1) Les sites de recombinaison

Le site *attP* est le plus complexe puisqu'il a une taille d'environ 240 pb et possède 14 sites de fixation pour 4 protéines différentes (Fig. 22a) (Landy, 2015). Ce site est composé de deux sites bras, un bras P et un bras P'. Le bras P présente les sites P1, et P2, qui sont des sites de forte affinité de $^\lambda\text{Int}$, les sites de fixation X1, X1.5 et X2 de l'excisionase Xis, un site F de la protéine Fis et les sites H1 et H2 du facteur d'hôte IHF. De son côté, le bras P' est plus simple et présente les sites P'1 P'2 et P'3 de fixation de haute affinité de $^\lambda\text{Int}$, et un site H' pour la protéine IHF. Enfin, le site *attP* partage une région d'identité de séquence avec le site *attB* de 15 pb (Landy and Ross, 1977).

Dans la région cœur du site *attP*, deux séquences inversées imparfaitement répétées de 9 pb, les sites C et C', sont les sites de fixation de faible affinité de λ Int (Fig. 22b). Les sites C et C' encadrent une région de 7 pb, la région centrale « O » (« Overlap »), qui est le lieu de la coupure et de l'échange de brin. La taille de 7 pb pour la région centrale semble être une caractéristique conservée pour les sites cibles des recombinaisons à tyrosine hétérobivalentes. En effet, la région centrale des systèmes CtnDOT, HK022, HP1, L5, P22 est également de 7 pb (Hauser and Scocca, 1992; Kolot and Yagil, 1994; Malanowska *et al.*, 2006; Peña *et al.*, 1996; Smith-Mungo *et al.*, 1994). Pour λ Int, il a été montré que la nature des bases dans la région centrale n'était pas importante pour la recombinaison mais que l'identité de séquence entre les régions centrales des deux sites de recombinaison était un paramètre essentiel à la réaction (Bauer *et al.*, 1985; Weisberg *et al.*, 1983).

De son côté, le site *attB* est simple, d'environ 21 pb (Fig. 22b). Ce site ressemble fortement aux sites *loxP* et *FRT* des recombinaisons à tyrosine simple, puisqu'il ne présente que deux séquences inversées imparfaitement répétées, les sites cœur B et B' (sites de fixation de faible affinité de λ Int), bordant la région centrale « O » (Fig. 22b) (Azaro and Landy, 2002). Les sites *attL* et *attR* sont des sites hybrides, de complexité intermédiaire. Le site *attL* est composé du bras P', avec les sites cœur B et C'. Le site *attR* est composé du bras P avec C et B' comme site cœur (Fig. 22a).

II.9.1.1.1) Les sites de fixation de λ Int

Une des premières découvertes du système λ , fut la capacité de l'intégrase à reconnaître deux séquences d'ADN différentes : une séquence située dans le site bras, et une séquence située dans le site cœur (Ross *et al.*, 1979). Depuis, différentes études ont permis l'identification de 9 sites de fixation de λ Int : 5 sites bras (P1, P2, P'1, P'2, P'3), et 4 sites cœur (B, B', C, C') (Hsu *et al.*, 1980; Ross and Landy, 1982; Ross and Landy, 1983). Les séquences des quatre sites cœur sont toutes différentes, avec le site C' qui présente la séquence consensus « 5'-CAACTTNNT-3' » (N= A, T, G ou C), et les sites B, B' et C qui ont au moins un mésappariement avec cette séquence consensus. L'affinité de λ Int envers ses sites cœur est : C' > C = B > B' (Ross and Landy, 1983). Les séquences des cinq sites bras sont aussi différentes et l'affinité de l'intégrase λ envers ses sites bras est : P1 = P'1 > P2 = P'2 = P'3 (Sarkar *et al.*, 2002).

II.9.1.1.2) Les sites de fixation IHF

A l'intérieur du site *attP*, 3 sites de fixation pour la protéine IHF sont retrouvés : les sites H1 et H2 dans le bras P, et le site H' dans le bras P' (Craig and Nash, 1984). Ces sites partagent une séquence consensus « 5'-WATCAANNNTTR-3' » (W = A/T). La fixation de la protéine IHF sur chacun de ses sites induit une forte courbure estimée à plus de 140° (Thompson and Landy, 1988). Ce résultat a été confirmé par cristallographie sur le site H' où IHF induit une courbure de l'ADN de plus de 160° (Rice *et al.*, 1996). Malgré la similarité de la courbure induite au niveau de ces trois sites, leur rôle dans la recombinaison n'est pas équivalent puisque H2 et H' sont nécessaires pour l'intégration et l'excision, alors que le site H1 est seulement requis pour l'intégration (Bushman *et al.*, 1985; Gardner and Nash, 1986; Thompson *et al.*, 1986).

II.9.1.1.3) Les sites de fixation de l'excisionase Xis et de la protéine Fis

Au départ, une expérience de protection à la néocarzinostatine a permis d'identifier les deux sites de fixation X1 et X2, et une expérience de retard sur gel a permis de montrer la grande coopérativité entre monomère de Xis dans la liaison à ses deux sites. Cette expérience a aussi permis de montrer une forte coopérativité entre l'intégrase λ et son excisionase, puisqu'en présence de Xis, la fixation de l'intégrase λ à son site bras P2, est augmentée d'un facteur 32 (Bushman *et al.*, 1984). Comme pour IHF, la fixation de Xis à ses sites X1 et X2 entraîne une courbure de l'ADN estimée à plus de 140° (Thompson and Landy, 1988). Des années plus tard, il a été montré qu'un troisième monomère de Xis était fixé à l'ADN entre les sites X1 et X2, au niveau d'un site X1.5 (Abbani *et al.*, 2007; Sun *et al.*, 2006). De son côté, la protéine Fis se fixe à un seul site, le site F, présent sur le bras P et chevauchant la séquence du site X2 de Xis (Thompson *et al.*, 1987a).

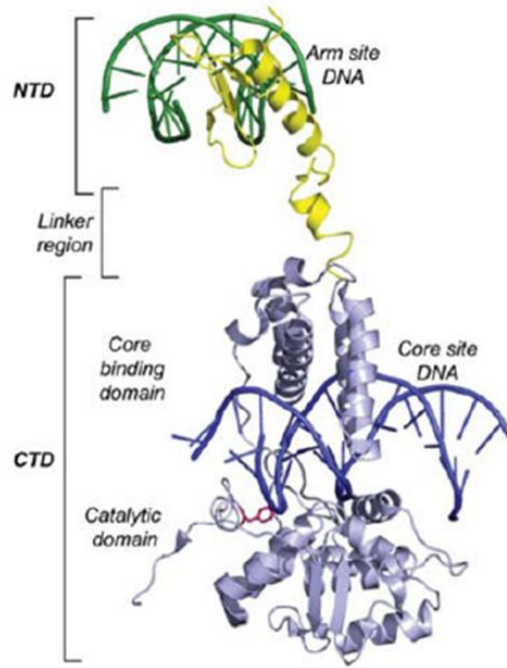


Fig. 23 Structure de l'intégrase λ liée à un site bras et à un site cœur (Landy, 2015). La séquence d'ADN du site bras est représentée en vert, la séquence d'ADN du site cœur est représentée en bleu, le domaine N-ter, se fixant sur le site bras est représenté en jaune et le domaine C-ter se fixant sur le site cœur et possédant l'activité topoisomerase est représenté en gris. La tyrosine nucléophile est représentée en rouge dans le domaine catalytique.

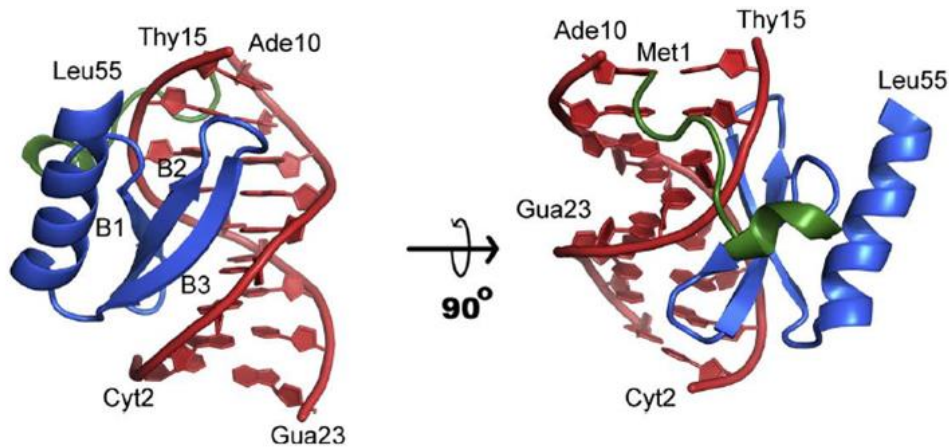


Fig. 24 Structure par RMN du domaine N-ter (1-64) de l'intégrase λ lié à un site bras (Fadeev *et al.*, 2009). L'ADN est représenté en rouge, le domaine N-ter replié en « three-stranded β -sheet » avec une hélice- α C-terminale est représenté en bleu. La partie de la protéine non-ordonnée en absence de fixation d'ADN et ordonnée après fixation à l'ADN est représentée en vert.

II.9.1.2) Les protéines impliquées dans la recombinaison

II.9.1.2.1) L'intégrase λ

L'intégrase λ est une protéine basique de 356 AA, possédant une activité de type topoisomérase I (Hoess *et al.*, 1980; Kikuchi and Nash, 1979) et qui est capable de se fixer de façon hétérobivalente à l'ADN en utilisant deux domaines protéiques différents. Grâce à un clivage à la chymotrypsine, l'intégrase a pu être divisée en deux domaines : un domaine N-ter (1-64) capable de se fixer aux sites bras, et un domaine C-ter (65-356) contenant l'activité topoisomérase et la fixation aux sites cœur (de Vargas *et al.*, 1988). Ce domaine C-ter est fonctionnellement et structurellement similaire aux recombinases Cre, Flp et XerC/D. Des années plus tard, il a été montré que le domaine C-ter pouvait être divisé en deux sous-domaines et que les trois domaines protéiques étaient séparés par 2 « linkers » (Fig. 23). L'intégrase est donc constituée d'un domaine N-ter capable de se fixer aux sites bras (1-64), suivi d'un « linker » (65-74), d'un domaine « core-binding » (CB) (75-159) capable de se fixer aux sites cœur, d'un deuxième « linker » (160-175) et d'un domaine catalytique (176-356) qui contient les résidus conservés RDKHRHY, responsable de l'activité de clivage et de ligature de l'ADN (Biswas *et al.*, 2005; Tirumalai *et al.*, 1997).

II.9.1.2.1.1) Le domaine N-ter de λ Int

La structure du domaine N-ter a été déterminée dans deux études différentes par RMN, en présence ou non de son ligand (Fadeev *et al.*, 2009; Wojciak *et al.*, 2002) (Fig. 24). Ce domaine adopte un repliement de type « three-stranded β -sheet » et reconnaît l'ADN en insérant le motif « three-stranded β -sheet » dans le sillon majeur de l'ADN et en effectuant des contacts avec le sillon mineur *via* la queue N-terminale (Fadeev *et al.*, 2009).

Au départ, le domaine N-terminal était considéré comme un domaine simple, ayant seulement un rôle architectural afin de délivrer le domaine « core-binding » aux sites cœur. En réalité, ce domaine a un rôle beaucoup plus complexe. Le domaine N-ter minimum comprends les acides aminés 1-64, et est capable de se fixer sur un site bras avec la même affinité que l'intégrase entière. En revanche, ce domaine minimal est incapable de se fixer de façon coopérative sur les sites P'1-P'2-P'3, et nécessite 6 acides-aminés supplémentaires présents dans le linker (1-70) (Sarkar *et al.*, 2002; Warren *et al.*, 2005).

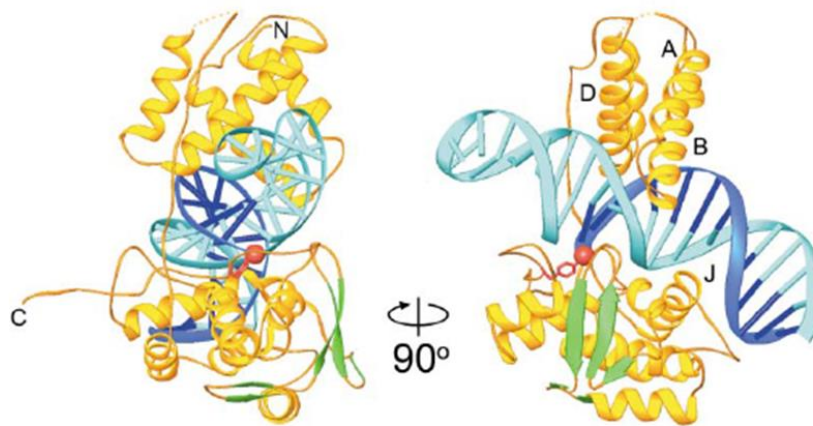


Fig. 25 Structure du domaine C-ter (75-356) de l'intégrase λ liée à un site cœur (Aihara *et al.*, 2003). L'ADN est représenté en bleu et le domaine « core-binding » (situé au-dessus de l'ADN) et catalytique (situé en-dessous de l'ADN) sont représentés en jaune. Le phosphate scissile lié de façon covalente avec la tyrosine est indiqué par une sphère rouge. Le domaine « core-binding » s'insère dans le sillon majeur de l'ADN et le domaine catalytique interagit avec l'ADN *via* le sillon majeur et mineur du côté opposé de l'ADN. Le domaine C-ter forme une structure de type « C-shaped clamp », structure conservée parmi les recombinases à tyrosine.

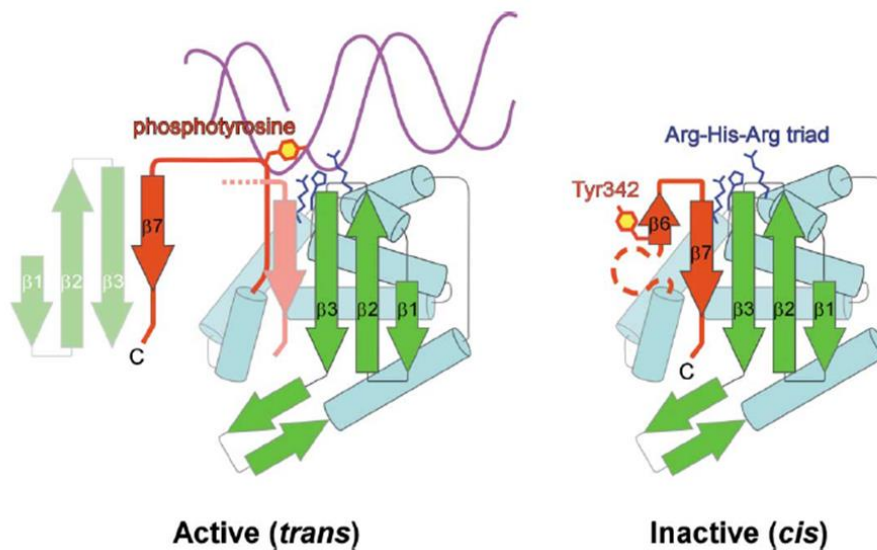


Fig. 26 Changement conformationnel du site actif de l'intégrase λ (Aihara *et al.*, 2003). Dessin illustrant le changement de conformation de l'intégrase lorsqu'elle se fixe à l'ADN et permet le positionnement de la tyrosine nucléophile. L'isomérisation de la forme inactive dans laquelle la tyrosine nucléophile est positionnée loin de la triade catalytique R-H-R, à la conformation active est accompagné par la libération du feuillet $\beta 7$ et son empaquetage en *trans* dans le domaine catalytique d'un protomère voisin.

En plus de la fixation aux sites bras, ce domaine N-ter est capable d'interactions homomérique avec l'intégrase et d'interactions hétéromérique avec l'excisionase Xis *via* son hélice- α située en C-ter (Cho *et al.*, 2002; Swalla *et al.*, 2003; Warren *et al.*, 2003). Le domaine N-ter de l'intégrase fonctionne également comme un modulateur des fonctions de l'intégrase. En effet, la délétion du domaine N-ter augmente la fixation de la protéine aux sites cœur, son activité topoisomérase et la résolution de jonction de Holliday (Sarkar *et al.*, 2001). Lorsque le domaine N-ter est présent en *cis* (intégrase entière), il inhibe l'activité du domaine C-ter, et de façon étonnante, lorsqu'il est présent en *trans* (ajout du domaine N-ter purifié au domaine C-ter purifié), il stimule l'activité du domaine C-ter. Enfin, la fixation du domaine N-ter de l'intégrase entière aux sites bras lève l'inhibition de l'activité du domaine C-ter (Sarkar *et al.*, 2001).

II.9.1.2.1.2) Le domaine C-ter de λ Int

Le domaine C-ter de λ Int (75-356) correspond aux domaines N-ter et C-ter des recombinases Cre, Flp, et XerC/D (Fig. 11). Contrairement à ces protéines, ce domaine est incapable d'effectuer seul une réaction de recombinaison complète, mais possède une activité de topoisomérase, de fixation aux sites cœur et de résolution des jonctions de Holliday (Tirumalai *et al.*, 1997). La structure du domaine C-ter lié à un site cœur a été déterminée (Aihara *et al.*, 2003) (Fig. 25). Le domaine C-ter de λ Int forme une structure de type « C-shaped clamp », avec le domaine « core-binding » interagissant avec l'ADN *via* le sillon majeur et le domaine catalytique interagissant avec l'ADN *via* le sillon majeur et mineur de la face opposée de l'ADN.

Les deux domaines contribuent tous les deux à la reconnaissance du site cœur, mais le domaine « core-binding » confère la majorité de la spécificité de séquence (Kovach *et al.*, 2002; Tirumalai *et al.*, 1998). En comparant la structure du domaine C-ter de λ Int liée à son site cœur (Aihara *et al.*, 2003) avec la structure du domaine catalytique sans ligand (Kwon *et al.*, 1997), une différence majeure est observée au niveau du positionnement de la tyrosine nucléophile. En effet, dans la structure sans ligand, la tyrosine nucléophile est située loin du site actif, stabilisée par l'empaquetage du feuillet β 7 dans la protéine (349-353) (Fig. 26). Dans la structure avec le ligand, les résidus 349-356 d'un protomère s'étendent loin de la protéine et l'empaquetage du feuillet β 7 dans le domaine catalytique d'un protomère adjacent permet le repositionnement de la tyrosine à l'intérieur du site actif (Fig. 26).

Les 7 derniers résidus de ce domaine C-ter ont fait l'objet de différentes études afin de comprendre leur rôle dans la protéine (Kazmierczak *et al.*, 2002; Tekle *et al.*, 2002). La modification de certains résidus ou la délétion des 6 derniers résidus augmente l'activité topoisomérase de l'enzyme, mais la protéine est incapable d'effectuer une recombinaison, suggérant que la queue C-ter de l'enzyme est un élément régulateur de l'activité catalytique, notamment pour la coordination des événements de clivage (Tekle *et al.*, 2002).

Enfin, la spécificité de fixation à la région cœur a pu être étudiée grâce à l'intégrase HK022. ^{HK022}Int et ^λInt sont des intégrases proches, ayant les mêmes 55 premiers acides aminés, partageant 74% d'identité de séquence et reconnaissant des sites bras identiques sur leur site *attP* respectifs. Cependant, aucune des deux intégrases n'est capable d'effectuer une recombinaison avec les sites de l'autre intégrase. La différence majeure entre les sites *attP* et *attB* des deux protéines provient des séquences cœur B/B' et C/C' (Yagil *et al.*, 1989). Dorgai *et al.*, ont identifié 5 changements d'AA « λ-to-HK022 » qui empêchent ^λInt d'effectuer une recombinaison avec ses sites naturels, et lui permettent d'effectuer une recombinaison avec les sites *att* de ^{HK022}Int (Dorgai *et al.*, 1995). Parmi ces modifications, une se trouve dans le domaine « core-binding » (N99D), et les quatre autres se situent dans le domaine catalytique (S282P, G283K, R287K et E319R ; S = sérine, G = glycine).

II.9.1.2.2) IHF

A l'origine, IHF a été identifié comme un facteur présent dans un extrait cellulaire de la bactérie *E. coli*, essentiel à la recombinaison *in vitro* et *in vivo* du système λ, d'où son nom « Integration Host Factor » (Kikuchi and Nash, 1978; Nash and Robertson, 1981). Cette protéine est un hétérodimère constitué d'une sous-unité α de 11,2 kDa et d'une sous-unité β de 10,6 kDa. Ces deux sous-unités partagent environ 30% d'identité entre elles et avec la famille des « DNA binding protein » de type II qui regroupe la plupart des protéines de type « Histone-like », dont la protéine HU (Azaro and Landy, 2002). Cette protéine essentielle à la recombinaison du bactériophage λ s'est rapidement révélée comme étant une protéine importante pour la physiologie de la bactérie *E. coli*. En effet, cette protéine est impliquée dans une grande variété de fonctions cellulaires comme la régulation de la transcription de gènes, l'initiation de la réplication de l'ADN, la transposition... (Friedman, 1988; Johnson *et al.*, 2004; Nash, 1996; Swinger and Rice, 2004).



Fig. 27 Structure de l'hétérodimère IHF fixé sur le site H' du site *attP* de l'intégrase λ (Rice *et al.*, 1996). La sous-unité α est représentée en blanc et la sous-unité β est représentée en rose. La séquence consensus du site est représentée en vert et interagit principalement avec le bras de la sous-unité α et le corps de la sous-unité β . Les deux résidus proline intercalés dans l'ADN et permettant sa courbure sont représentés en jaune.

La structure de cette protéine a pu être obtenue (Rice *et al.*, 1996). IHF, présente un domaine globulaire compact issue de l'entrelacement symétrique d'une sous-unités α et β , à partir duquel s'étendent deux bras « β -ribbon » (Fig. 27). Ces bras s'enroulent autour de l'ADN et interagissent *via* le sillon mineur. Sur la structure, une courbure de l'ADN est observée à deux point de flexion, séparés de 9 pb, produisant une courbure de l'ADN de plus de 160° (Rice *et al.*, 1996).

Le rôle d'IHF dans la recombinaison du bactériophage λ est strictement architectural. En effet, il a rapidement été démontré que la courbure induite par IHF au site H' permettait le rapprochement des sites bras P' et des sites cœur favorisant l'interaction de λ Int avec ces deux sites *via* ces deux domaines protéiques (de Vargas *et al.*, 1989). Depuis, ce rôle strictement architectural a été renforcé par plusieurs études puisqu'un ou plusieurs sites d'IHF ont pu être remplacés par des modules permettant la courbure de l'ADN, soit par de l'ADN intrinsèquement courbé (« A-tract »), soit par des protéines courbant l'ADN comme la protéine HU, ou les protéines eucaryotes HMG1/2 (Giese *et al.*, 1992; Goodman and Nash, 1989; Goodman *et al.*, 1992; Segall *et al.*, 1994). Cependant, malgré le fait que ces différentes constructions puissent compléter l'absence d'IHF, elles n'ont jamais réussi à être aussi efficace que le site sauvage. Cela s'explique par la nécessité d'une précision importante de la courbure, notamment par le fait qu'un décalage d'une paire de base de la courbure induite par IHF réduit fortement la recombinaison (Nunes-Düby *et al.*, 1995b).

II.9.1.2.3) L'excisionase Xis

Xis est une petite protéine basique de 72 AA codée par le bactériophage λ . Cette protéine est essentielle à l'excision, tout en inhibant la réaction d'intégration (Abremski and Gottesman, 1982). Une étude de la séquence de Xis a permis de montrer qu'une excisionase tronquée (1-53) conservait son activité de liaison à l'ADN et la liaison coopérative avec Fis, mais perdait sa liaison coopérative avec l'intégrase, suggérant que la protéine était composée de deux domaine : un domaine N-ter de fixation à l'ADN et un domaine C-ter impliqué dans la coopérativité avec λ Int (Numrych *et al.*, 1992).

La structure de Xis fixée sur son site X2 a été obtenue et montre une protéine de type « winged-helix » qui interagit avec l'ADN afin de former des interactions spécifiques avec le sillon majeur *via* une hélice- α , et des interactions majoritairement non spécifiques avec le sillon mineur *via* le motif « wing » (Fig. 28a) (Sam *et al.*, 2004).

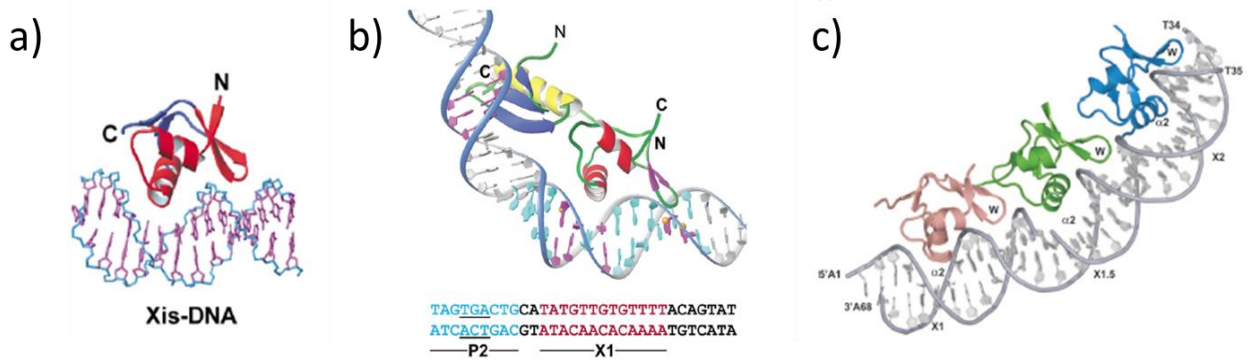


Fig. 28 Complexe de Xis lié à l'ADN (Adapté de Sam *et al.*, 2004 ; Abanni *et al.*, 2007). **a)** Structure de $1-55\text{Xis}^{\text{C28S}}$ lié à son site X2. Le monomère de Xis interagit avec le sillon majeur à l'aide de son hélice- α (interaction spécifique), et avec le sillon mineur de l'ADN à l'aide du motif « wing » (interaction majoritairement non spécifique). **b)** Modèle représentant le complexe ternaire entre le domaine N-ter de l'intégrase λ , Xis et l'ADN. Dans ce modèle, le domaine N-ter de l'intégrase λ interagit dans le sillon majeur avec les bases T-G-A du site P2 (bases soulignées dans la séquence bleue et représentés en violet dans l'ADN). Xis est modélisé sur son site X1, de la même manière que la structure obtenue sur son site X2. La queue C-ter de Xis, désordonnée en solution est localisée proche de l'hélice- α du domaine N-ter de l'intégrase λ et réalise une interaction protéine-protéine avec ce domaine. **c)** Structure de Xis lié à la région X1, X1.5, X2, révélant les bases structurales de leur liaison coopérative. Le monomère rose est lié au site X1, le monomère vert est lié au site X1.5 et le monomère bleu est lié au site X2.

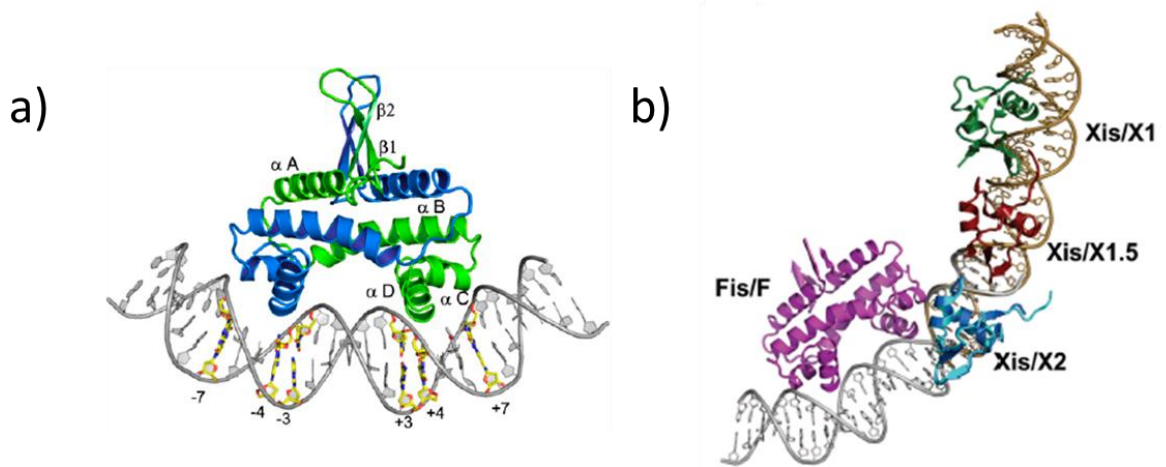


Fig. 29 Structure d'un dimère de Fis fixé à l'ADN (Adapté de Stella *et al.*, 2010 ; Hancock *et al.*, 2019). **a)** Dimère de Fis lié à la séquence F1 (Stella *et al.*, 2010). Un monomère de Fis est représenté en bleu et un monomère de Fis est représenté en vert. **b)** Modèle de la région F-X2-X1.5-X1 du site *attR*. La structure du trimère de Xis fixé à X1 (vert), X1.5 (rouge) et X2 (bleu) (Abanni *et al.*, 2007) a été alignée à la structure de Fis et Xis fixés sur le site F et X2 (Hancock *et al.*, 2019). Le dimère de Fis est représenté en violet. Le recrutement de Xis au site X2 par Fis ne se fait pas au travers d'interaction protéine-protéine mais indirectement au travers de la conformation de l'ADN.

Cette structure de Xis a permis de modéliser le recrutement du domaine N-ter de λ Int au site P2 par le monomère de Xis fixé sur le site X1 (Fig. 28b). En effet, sur ce modèle la queue C-ter de Xis se trouve adjacente à l'hélice- α C-ter du domaine N-ter de λ Int, permettant une interaction protéine-protéine. La structure du complexe coopératif Xis associé à ses sites X1, X1.5 et X2 a été obtenue et montre que les 3 monomères de Xis se fixent dans une orientation directe pour générer un filament micro-nucléoprotéique courbé d'environ 72° (Fig. 28c) (Abbani *et al.*, 2007). Les monomères situés sur les sites X1 et X2 reconnaissent des sites de fixation similaires et participent à une fixation coopérative à l'ADN au travers d'interactions protéine-protéine avec un monomère qui est lié de façon moins spécifique (X1.5). Cette capacité de liaison aspécifique à l'ADN de Xis a été caractérisée par Papagiannis *et al.*, (Papagiannis *et al.*, 2007). Dans cette étude, il est montré que Xis présente une faible activité de liaison spécifique à l'ADN. Ainsi, malgré le fait que cette protéine soit fortement produite dans la cellule par le bactériophage λ , pouvant atteindre jusqu'à 15000 copies par cellule, Xis est relativement inefficace pour mettre en place la réaction d'excision *in vivo* en absence de Fis (Papagiannis *et al.*, 2007).

II.9.1.2.4) Fis

Fis a initialement été identifié comme un facteur favorisant la réaction de recombinaison spécifique de site médiée par les invertases Hin, et Gin, d'où son nom « Factor for Inversion Stimulation » (Johnson *et al.*, 1986; Koch and Kahmann, 1986). Fis est une petite protéine basique de 98 AA, qui se fixe sous forme d'homodimères et courbe l'ADN (Azaro and Landy, 2002). Cette protéine s'est révélée impliquée dans diverses fonctions physiologiques cellulaires comme l'initiation de la réplication de l'ADN, la transposition, la régulation de la transcription de nombreux gènes... (Dorman, 2009; Johnson *et al.*, 2004). La structure de Fis a été obtenue et montre un dimère globulaire composé de 4 hélice- α et d'un domaine de liaison à l'ADN hélice-tour-hélice par monomère (Kostrewa *et al.*, 1991; Yuan *et al.*, 1991). Une des caractéristiques importantes de cette structure est que les deux hélices D sont trop proches pour pouvoir s'insérer dans les sillons majeurs adjacents d'un ADN de type B. Plus récemment, la structure de Fis liée à différents ADN a été obtenue et révèle que Fis sélectionne principalement des sites possédant un sillon mineur intrinsèquement étroit, où la composition en A/T est un facteur important (Fig. 29a) (Stella *et al.*, 2010). Le sillon mineur étroit permet l'insertion des hélices- α D dans les sillons majeurs adjacents et la fixation du dimère de Fis induit une courbure de l'ADN d'environ 65°.

Le rôle de Fis dans la recombinaison du système λ a mis de nombreuses années à être élucidé. Un site de fixation de Fis, le site F chevauchant le site X2 de Xis sur le site *attP* a été découvert et aucun effet de la protéine Fis sur l'excision n'était observé en concentration saturante de Xis et sur l'intégration *in vitro*. Une stimulation de l'excision d'un facteur 20 était observée en concentration limitante de Xis (Thompson *et al.*, 1987a). En revanche, *in vivo*, l'absence de Fis réduit l'excision du bactériophage λ d'un facteur 200 (Ball and Johnson, 1991a), et son intégration d'un facteur 15 (Ball and Johnson, 1991b). L'effet de Fis sur l'intégration a pu être démontré *in vitro*, à concentration sub-optimales de λ Int (Esposito and Gerard, 2003).

Du fait du chevauchement des sites X2 et F, il a été suggéré que les deux protéines ne pouvaient pas se fixer simultanément à l'ADN. Cependant, deux études ont montré que les protéines Fis et Xis se fixaient simultanément et coopérativement aux sites F et X2 (Papagiannis *et al.*, 2007; Sun *et al.*, 2006). Papagiannis *et al.*, ont montré que la protéine Fis *in vitro* sur le site *attR* avait une affinité 100 fois plus grande que Xis et qu'*in vivo* le taux d'excision était diminué d'un facteur 100 en absence de Fis. Cette différence d'affinité les a amenés à proposer qu'*in vivo*, Fis se fixe sur son site F et recrute le monomère Xis sur son site X2, qui recrute ensuite les monomères fixés sur les sites X1.5 et X1. Récemment, un modèle du complexe Fis-Xis-ADN a été proposé (Fig. 29b) (Hancock *et al.*, 2019). Dans cette étude, il a été montré que le recrutement de Xis au site X2 par Fis ne se fait pas par des interactions protéine-protéine, mais de façon indirecte par la conformation de l'ADN : la fixation de Fis rétrécit le sillon mineur de l'ADN, favorisant l'insertion du motif « wing » de Xis, et courbe l'ADN facilitant les contacts Xis-ADN.

II.9.2) La réaction de recombinaison

L'idée d'un complexe nucléoprotéique d'ordre supérieur formé sur le site *attP* est arrivée rapidement après l'observation d'un complexe intégrase-*attP* en microscopie électronique où l'ADN était condensé (Fig. 30a) (Better *et al.*, 1982). Le terme « intasome » a alors été utilisé pour désigner ce complexe intégrase-ADN. Depuis, ce terme est conservé pour décrire les complexes protéine-ADN *att* qui contiennent toutes les protéines nécessaires à la formation de ces structures (Landy, 1989).

La structure de cet intasome est façonnée par trois protéines courbant l'ADN : IHF, Xis et Fis, qui induisent des courbures importantes de l'ADN et par la capacité de λ Int à se fixer à des séquences différentes afin de former des boucles d'ADN par l'utilisation de deux domaines protéique différents. Pour la réaction d'intégration, deux particularités ont été observées : la nécessité d'avoir un site *attP* surenroulé, puisque ce surenroulement facilite l'assemblage de l'intasome sur ce site (Richet *et al.*, 1986), et l'incapacité de l'intégrase à former un complexe stable avec le site *attB* suggérant que ce site *attB* « nu » est capturé par l'intasome préformé sur le site *attP* (Richet *et al.*, 1988).

λ Int semble avoir évolué de façon à effectuer les réactions d'intégration et d'excision de manière mutuellement exclusive. Biologiquement, les réactions d'intégration et d'excision sont l'inverse l'une de l'autre. Cependant, au niveau du mécanisme réactionnel, ces recombinaisons sont distinctes par l'utilisation de protéines et de sites différents. En effet, la protéine Xis essentielle à l'excision inhibe fortement l'intégration (Abremski and Gottesman, 1982). Au niveau de la protéine IHF, les sites H2 et H' sont utilisés pour les deux réactions, mais le site H1 doit être occupé pour l'intégration et vacant pour l'excision (Bushman *et al.*, 1985; Gardner and Nash, 1986; Thompson *et al.*, 1986). Enfin, de son côté, λ Int se fixe sur quatre sites bras pour l'intégration (P1, P'1, P'2 et P'3) et sur trois sites bras pour l'excision (P2, P'1, P'2) (Bauer *et al.*, 1986; Bushman *et al.*, 1985; Hazelbaker *et al.*, 2008; Thompson *et al.*, 1987b). Lors de ces réactions, les pontages entre les sites médiés par λ Int, c'est-à-dire la fixation d'une intégrase à un site bras et à un site cœur, ne sont pas identiques. Tong *et al.*, ont déterminé les différents pontages de l'intégrase dans la réaction d'intégration et d'excision (Tong *et al.*, 2014). Lors de l'intégration, les couples domaine « N-ter de λ Int » fixé à un site bras et « domaine C-ter de λ Int » fixé à un site cœur sont les suivants : P1-B, P'1-C', P'2-C et P'3-B' (Fig. 30b ; Fig. 30c). Lors de l'excision, les couples suivants sont observés : P2-B', P'1-C', P'2-B et l'intégrase fixée au site cœur C ne forme pas de pontage avec les sites bras restant (Fig. 31a) (Tong *et al.*, 2014).

Grâce aux différentes données biochimiques, fonctionnelles, structurales, et aux récents résultats de FRET (« Fluorescence Resonance Energy Transfer ») et de cryo-microscopie électronique du complexe d'excision, il a été possible de créer un modèle précis de chaque voie de recombinaison du bactériophage λ (Laxmikanthan *et al.*, 2016; Seah *et al.*, 2014).

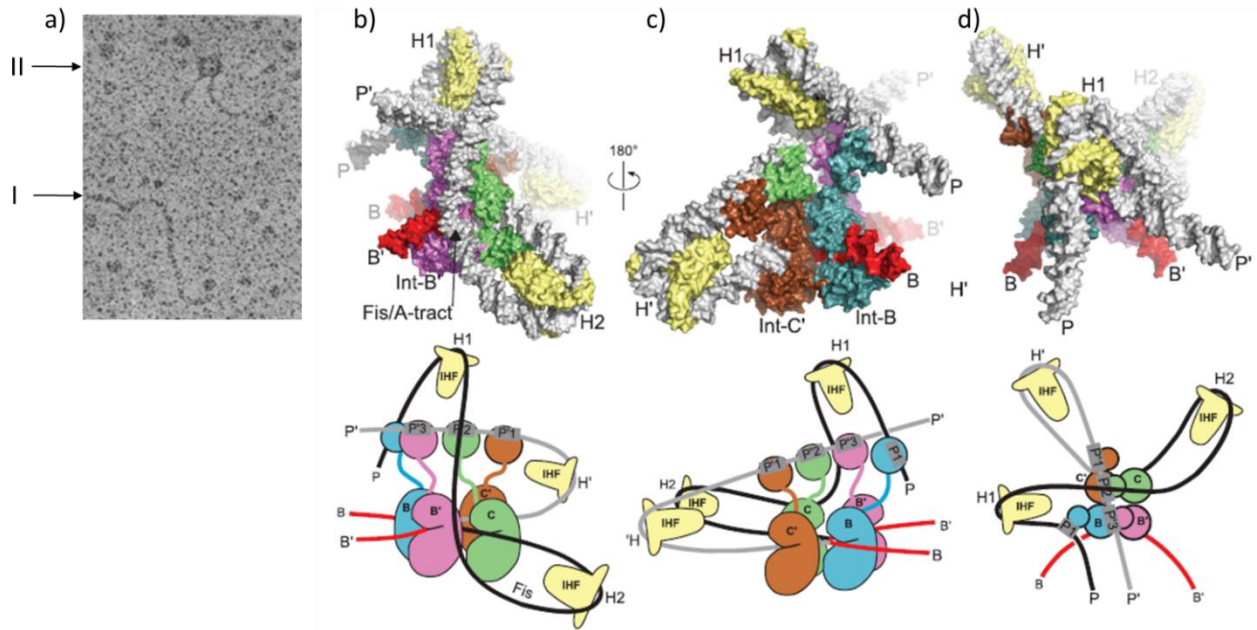


Fig. 30 Formation de l'intasome par le bactériophage λ sur le site *attP* (Adapté de Better *et al.*, 1982 ; Laxmikanthan *et al.*, 2016). **a)** Observation en microscopie électronique de la fixation de l'intégrase λ sur un fragment d'ADN comportant le site *attP*. I, fragment non associé à l'intégrase et II, fragment d'ADN condensé par la fixation de l'intégrase (l'intasome). **b)** Structure de l'intasome sur le côté B'-C. La courbure induite par IHF sur le site H2, combinée à une courbure sur le site Fis et une courbure supplémentaire fournie par le surenroulement négatif dirige le bras P sur le dessus du tétramère λ Int et sur le bras P'. La courbure de la protéine IHF sur le site H1 redirige le bras P vers le bas, le long de la face opposée au tétramère. **c)** Structure de l'intasome sur le côté C'-B. Les pontages P1-B (bleu) et P'1-C' (marron) sont observés ainsi que l'arrangement pratiquement linéaire des domaines N-ter liés aux sites bras P'1, P'2 et P'3. **d)** Vue du haut de l'intasome. Des dessins schématisques du complexe intégratif sont représentés sous les modèles structuraux des parties b, c et d. Le bras P est représenté en noir, et le bras P' en gris. Le centre de la jonction de Holliday n'est pas représenté pour plus de clarté.

II.9.2.1) Le Complexe intégratif

Dans le complexe intégratif (Fig. 30b ; Fig. 30c ; Fig. 30d), l'intasome se forme sur le site *attP* de manière à pouvoir capturer le site *attB*. Sur cet intasome, IHF se fixe sur le site H' induisant une courbure du bras P' (Fig. 30c) ce qui permet aux intégrases fixées par leur domaine N-ter aux sites bras P'1, P'2 de se fixer aux sites cœur C', et C respectivement (Fig. 30b ; Fig. 30c). L'intégrase fixée sur le site bras P'3 est positionnée pour la capture du site cœur B'. Sur le bras P, IHF se fixe aux sites H1 et H2 courbant l'ADN (Fig. 30b), et positionnant le domaine C-ter de l'intégrase fixée au site P1 pour la capture du site cœur B (Laxmikanthan *et al.*, 2016; Seah *et al.*, 2014). Ainsi, le modèle de la réaction d'intégration suggère une interaction transitoire du domaine C-ter de l'intégrase fixée sur le site bras P1, avec les domaines C-ter des intégrases fixées aux sites bras P'1 et P'3 afin de former un complexe tétramérique. Une fois l'intasome formé sur le site *attP*, il peut capturer le site *attB* par un mécanisme encore inconnu. Le modèle suggère que le bras P du complexe intégratif peut s'ouvrir pour tester des séquences *attB* « candidates » et délivrer le domaine C-ter de l'intégrase fixé sur le bras P1 au site cœur B, sur la face opposée de l'ADN (Fig. 30c). Une étude a permis d'avoir plus d'information sur la capture du site *attB* par l'intasome (Tal *et al.*, 2014). Dans cette étude, il est montré que contrairement à l'idée initiale d'un intasome qui cherche activement son site *attB* sur le chromosome, l'ADN phagique reste confiné près de son point d'entrée dans la cellule et profite des mouvements provoqués par la ségrégation du chromosome lors de la réplication pour capturer le site *attB*. Une fois que la fixation au site *attB* est stable, la réaction de recombinaison peut commencer par l'échange du brin supérieur. La structuration de l'intasome dicte la réaction de recombinaison, ainsi, la formation de la jonction de Holliday commence toujours par un échange du brin supérieur (pour l'intégration, et pour l'excision) (Kitts and Nash, 1988; Laxmikanthan *et al.*, 2016; Nunes-Düby *et al.*, 1987). Une fois la recombinaison terminée, les produits *attL* et *attR* sont libérés.

II.9.2.2) Le Complexe d'excision

Dans le complexe d'excision (Fig. 31), IHF courbe le bras P du site *attR* en se fixant sur son site H2. Les sous-unités de Xis se fixent coopérativement aux sites X1, X1.5 et X2, induisant une courbure de l'ADN permettant à l'intégrase fixée au site bras P2 par son domaine N-ter de se fixer au site cœur B' par son domaine C-ter (Fig. 31a). La protéine Fis n'est pas présente, mais sa structure peut être ajustée dans ce modèle (Laxmikanthan *et al.*, 2016).

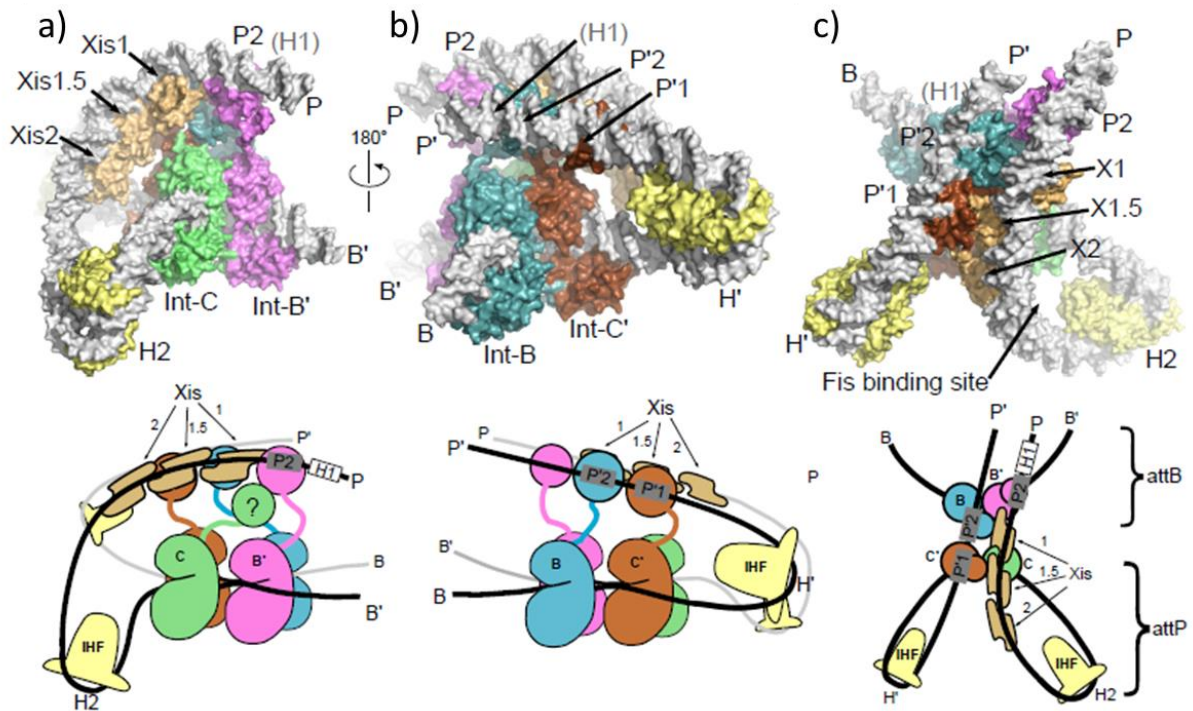


Fig. 31 Structure du complexe excisif du bactériophage λ (Adapté de Laxmikanthan *et al.*, 2016). a) Vue du complexe du côté du site *attR*. La courbure induite par IHF au site H2, combinée à la liaison et la courbure coopérative des trois monomères de Xis et au pontage P2-B' (rose) redirige le bras P vers le haut du tétramère de λ Int. Dans ce modèle structural, le domaine N-ter de l'intégrase liée au cœur C (vert) n'a pas pu être intégré et son emplacement approximatif est représenté dans le dessin schématisé. b) Vue du complexe du côté du site *attL*. La protéine IHF fixée sur le site H' courbe fortement le bras P', permettant aux sous-unités λ Int d'effectuer les pontages P'1-C' (marron) et P'2-B (bleu). c) Vue du haut du complexe excisif. Le bras P et P' sont proches et quasiment parallèles. Une interface importante est formée entre les sous-unités de Xis et les domaines N-ter des sous-unités de λ Int fixées sur le site cœur C' (marron) et B (bleu). Le site de liaison inoccupé mais accessible pour la protéine Fis est indiqué. Dans les vues schématisées, le centre de la jonction de Holliday n'est pas représenté pour plus de clarté. Les accolades *attB* et *attP* indiquent la position des produits d'excision.

Une intégrase se fixe sur le site cœur C avec son domaine C-ter, mais le positionnement de son domaine N-ter est inconnu (Fig. 31a). Sur le site *attL*, IHF se fixe sur le site H' et courbe le bras P' permettant aux intégrases fixées par leur domaine N-ter sur les sites bras P'1 et P'2, de se fixer aux sites cœur C' et B respectivement (Fig. 31b). Dans le modèle, la protéine Xis assure une interface importante protéine-protéine entre le bras P du site *attR* et le bras P' du site *attL*. Les monomères de Xis fixés au site X1 et X1.5 peuvent interagir avec les domaines N-ter des intégrases fixées au site cœur B et C' respectivement, et stabiliser la synapse *attL-attR* (Fig. 31c). Ainsi, Xis serait une protéine multifonctionnelle fondamentale à l'excision, en recrutant l'intégrase à son site P2, en courbant l'ADN permettant la fixation de l'intégrase recrutée au site cœur B', et en effectuant une interface entre les bras P et P', facilitant la synapse *attL-attR* (Laxmikanthan *et al.*, 2016). Une fois la synapse engagée, la réaction de recombinaison peut être réalisée et commencer par l'échange du brin supérieur, puis la résolution de la jonction de Holliday aboutit à la libération des produits *attB* et *attP*.

II.9.2.3) La mise en place de l'intasome : un phénomène conservé parmi les systèmes hétérobivalents

La formation de l'intasome par le système de recombinaison du bactériophage λ a probablement été un processus sélectionné au cours de l'évolution de manière à contrôler la directionnalité (intégration ou excision) de la recombinaison. Selon une perspective évolutive des recombinases spécifiques de site à tyrosine, il est possible que les recombinases simples comme Cre et les recombinases hétérobivalentes comme λ Int aient évolué de manière parallèle à partir d'un ancêtre commun, moins efficace. Il est aussi possible que les recombinases simples non régulées aient précédé les recombinases hétérobivalentes plus complexes. La complexification du système de recombinaison peut se produire par l'acquisition de séquences accessoires, dictant la directionnalité de la réaction. Par exemple, dans le système XerC/D l'ajout des séquences accessoires des sites *cer* et *psi* (ainsi que des protéines accessoires impliquées) crée un filtre topologique imposant aux recombinases la résolution de dimères de plasmides tout en empêchant leur formation. En accord avec cette idée, il a été montré que les éléments accessoires des sites *cer* et *psi* permettent la régulation de la recombinaison de Cre, une recombinase simple non régulée, montrant qu'il est donc possible de construire un système contrôlé par l'acquisition de facteurs accessoires (Gourlay and Colloms, 2004).

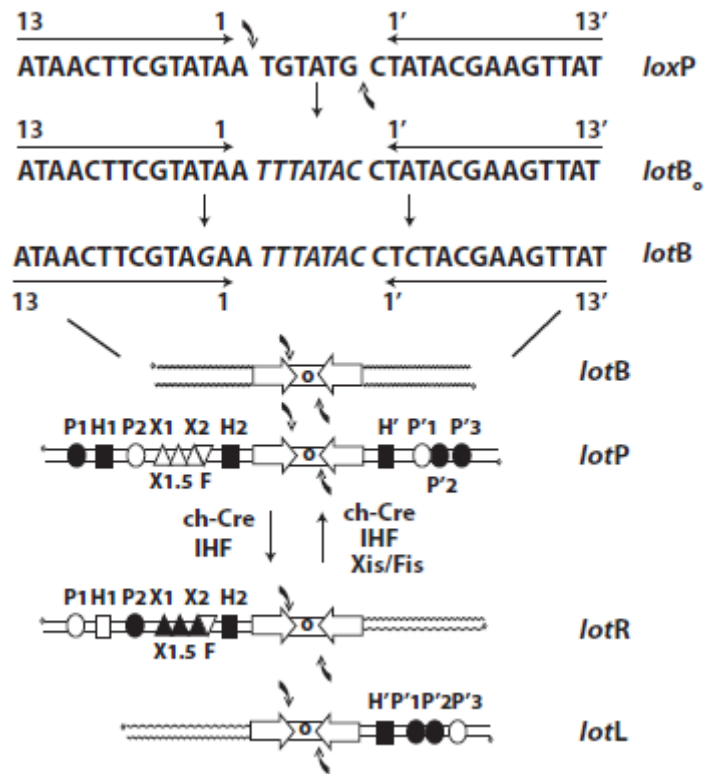


Fig. 32 Description du système de recombinaison chimère λ Int-Cre (Warren *et al.*, 2008). Le site *lotB* provient du site *loxP* où la région centrale a été remplacée par celle des sites de λ Int et où deux substitutions nucléotidiques ont été réalisées afin d'affaiblir l'interaction de Cre avec son site. Le remplacement de la région COC' du site *attP* de λ Int par le site *lotB* a produit le site *lotP*. Les sites bras de fixation du domaine N-ter de λ Int (cercles), de IHF (carrés) de Xis (triangles) et de Fis (triangle inversé) sont présents dans le site *lotP*. Les symboles noirs représentent les sites utilisés lors de la recombinaison *lotB/lotP* et *lotR/lotL*.

Une autre possibilité de complexification du système provient de l'acquisition d'un nouveau domaine protéique. En effet, la différence majeure entre les recombinases simples et les intégrases hétérobivalentes provient de la présence d'un domaine supplémentaire de fixation à l'ADN en N-ter permettant la fixation sur des séquences d'ADN accessoires (les sites bras) (Fig. 10 ; Fig. 11). Ainsi, ce domaine peut être considéré comme un domaine additionnel lié aux besoins de directionnalité et de régulation lors des cycles d'intégration et d'excision des bactériophages. Cette hypothèse a été testée de manière élégante par A. Landy il y a une dizaine d'années (Warren *et al.*, 2008) en construisant un système de recombinaison artificiel constitué d'une chimère λ Int-Cre, obtenue par fusion traductionnelle du domaine N-ter de λ Int à la protéine Cre, et de nouveaux sites de recombinaison, *lotP* et *lotB*, hybrides des sites *loxP* et *lattP* (Fig. 32). Cette protéine chimère devient capable d'effectuer une recombinaison contrôlée entre un site complexe *lotP* et un site simple *lotB* (réaction intégrative), ainsi qu'entre les sites *lotL* et *lotR* (réaction d'excision). Ces réactions présentent les mêmes caractéristiques que celles de λ Int : dépendance au facteur d'hôte IHF et inhibition de la réaction par le facteur directionnel Xis pour la réaction d'intégration et dépendance à IHF, à Xis et stimulation par Fis en concentration limitante de Xis pour l'excision. Par conséquent, l'ajout d'un domaine protéique supplémentaire et des sites bras associés ont permis de finement réguler la directionnalité de recombinaison d'une recombinase simple non régulée.

En plus des besoins de directionnalité de la recombinaison, le système d'intégration et d'excision du bactériophage semble avoir évolué par dépendance aux protéines de l'hôte, permettant de lier la régulation et la directionnalité de la recombinaison à la physiologie de l'hôte bactérien. Ainsi, les protéines de l'hôte IHF et Fis ont un effet sur ces réactions : un excès de protéine IHF inhibe l'excision tandis que Fis l'améliore (Ball and Johnson, 1991a ; Bushman *et al.*, 1985). Au niveau cellulaire, les taux d'IHF augmentent lorsque la bactérie entre en phase stationnaire (Ditto *et al.*, 1994), par conséquent, cette protéine aurait été sélectionnée au cours de l'évolution afin de limiter l'excision du bactériophage et l'entrée dans un cycle lytique à un instant où les ressources intracellulaires sont limitantes. De son côté, la synthèse de la protéine Fis augmente fortement dès que les cellules entrent en croissance exponentielle (Ball *et al.*, 1992), ainsi, l'utilisation de la protéine Fis pourrait favoriser l'excision à une période optimale pour le développement du cycle lytique du phage.

Etude bibliographique

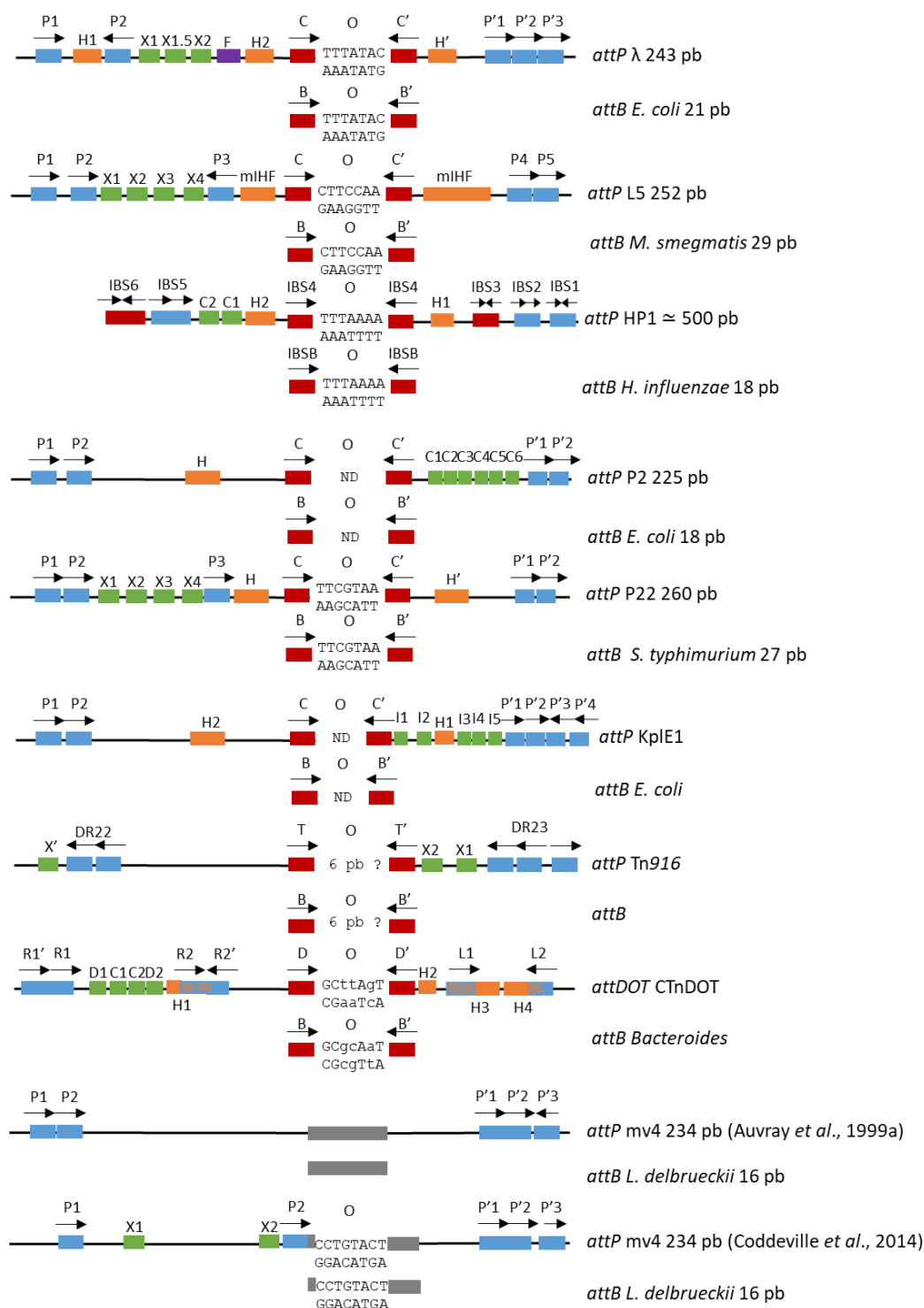


Fig. 33 Représentation des sites de recombinaison des différents systèmes des intégrases hétérobivalentes. Comparaison des sites de recombinaison des systèmes λ , L5, HP1, P2, P22, KplE1, Tn916, CTnDOT et mv4. Les sites bras de fixation des intégrases sont représentés par des rectangles bleu. Les sites cœur de fixation de l'intégrase sont représentés par des rectangles rouge. La flèche noire indique l'orientation du site. Les rectangles orange indiquent les sites de fixation des protéines accessoires ayant un rôle architectural dans la recombinaison (IHF pour λ , HP1, P2, P22 et KplE1, mIHF pour L5, BHFa « *Bacteroides* Host Factor » pour CTnDOT, HU pour Tn916). Dans le cas de CTnDOT, les hachures indiquent le chevauchement d'un site de fixation de IHF, avec un site bras de fixation de l'intégrase. Les rectangles vert indiquent les sites de fixation des excisionases (Xis pour λ , L5, P22, Tn916, mv4, Xis2c et Xis2d pour CTnDOT, Cox pour HP1, P2 et Tor1 pour KplE1). Le rectangle violet indique le site de fixation de la protéine Fis. Les séquences des régions centrales sont représentées, et les lettres en minuscules indiquent les positions où la séquence entre la région centrale des sites *attDOT* et *attB_{Bacteroides}* est différente. Le rectangle gris indique la région d'identité entre les sites *attP* et *attB* du système mv4.

Les contraintes topologiques de la recombinaison de λ Int ont été étudiées et montrent que la recombinaison entre deux sites *attB* et *attP* en orientation directe dans une molécule d'ADN circulaire aboutit à la formation de caténanes alors que la même réaction entre deux sites *attL* et *attR* aboutit majoritairement à la formation de molécules libres lorsque le niveau de surenroulement est faible (Crisona *et al.*, 1999). Ainsi, il semblerait que le complexe d'excision ait évolué au cours du temps de manière à exciser le prophage, en évitant la formation de caténanes qui serait abortif pour le virus. De son côté, le site *attP* semble avoir évolué par la mise en place d'un intasome, un assemblage nucléoprotéique complexe permettant au virus de capturer rapidement et précisément son site cible *attB* parmi les nombreuses séquences possibles dans le chromosome bactérien.

Les intégrases hétérobivalentes ont toutes conservé un domaine supplémentaire en N-ter ainsi que l'utilisation des séquences bras P et P' sur leur site *attP*. Chacun de ces systèmes utilise un facteur d'hôte intercalé entre les sites bras et les sites cœur de fixation de l'intégrase et la directionnalité de la recombinaison est conférée par l'excisionase qui inhibe les réactions d'intégration et est essentielle aux réactions d'excision. Cependant, bien que les sites bras soient conservés parmi ces systèmes, le nombre, l'espacement ainsi que l'orientation des sites de fixation des domaines N-ter de l'intégrase, des protéines de l'hôte et de l'excisionase sont tous différents (Fig. 33). Ainsi, l'assemblage d'un intasome fonctionnel sur le site *attP* semble être un critère essentiel au cycle de vie des bactériophages tempérés utilisant une intégrase hétérobivalente, mais les éléments responsables de la structuration de cet intasome restent caractéristiques de chaque couple phage tempéré/cellule hôte.

II.9.2.3.1) Les systèmes HP1, L5 et KplE1 : une mise en place de l'intasome différent du modèle λ

L'intégrase L5 permet l'intégration du génome du mycobactériophage L5 dans le gène codant un ARN^{Gly} de la bactérie hôte *Mycobacterium smegmatis* (Lee *et al.*, 1991). *In vitro* en présence de λ Int, aucun produit de recombinaison n'est obtenu et l'ajout de la protéine IHF n'a aucun effet sur la réaction (Lee and Hatfull, 1993), suggérant la présence d'un facteur d'hôte spécifique chez *M. smegmatis*.

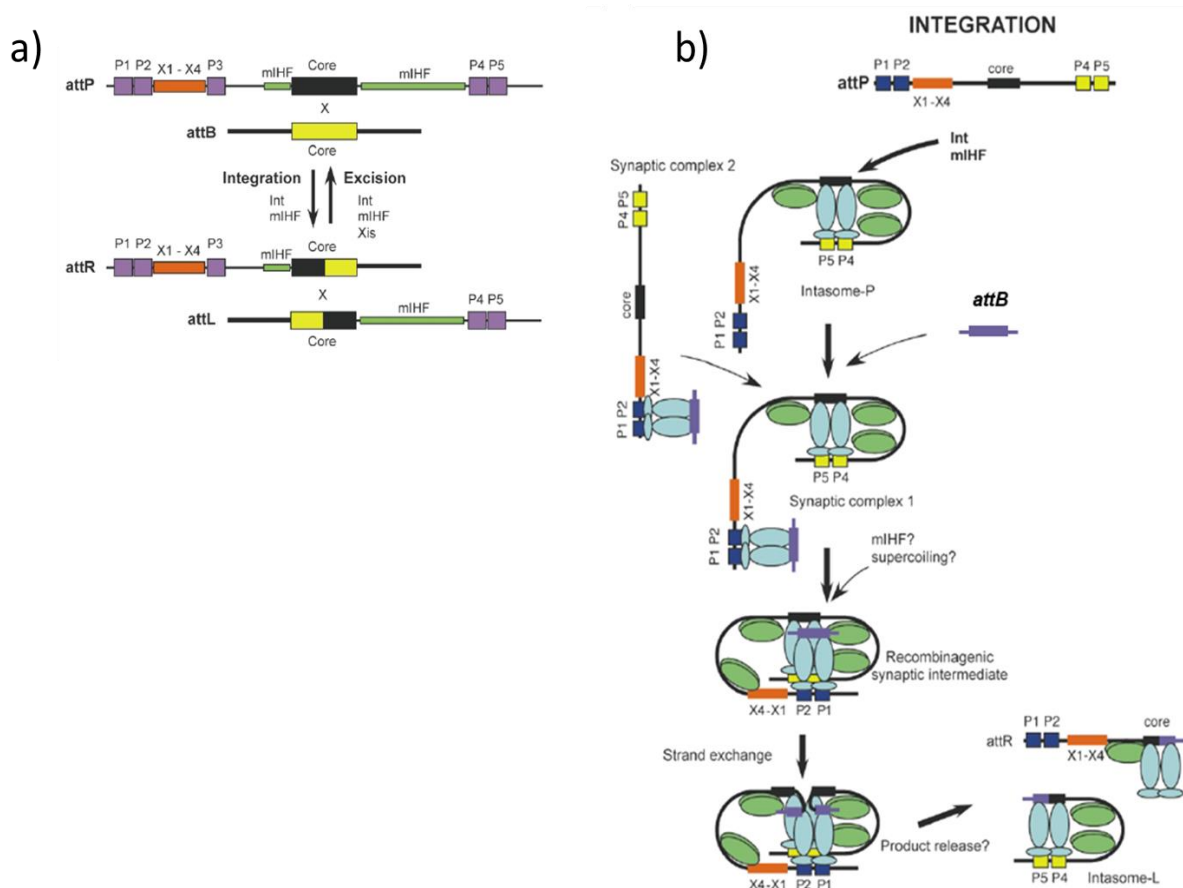


Fig. 34 Voies de recombinaison de l'intégrase L5 (Adapté de Lewis and Hatfull, 2003). **a)** Organisation des sites de recombinaison du système L5. Les sites *attP* et *attB* possèdent les sites de fixation cœur de l'intégrase L5 (rectangle noir et jaune, respectivement). La région cœur est encadrée par des sites bras de fixation de l'intégrase L5 (sites P1 à P5, carrés violet). Les paires de sites P1/P2, P4/P5 sont requises pour la recombinaison intégrative, tandis que P3 est inutile. Deux sites bras supplémentaires sont localisés à droite de P5 (P6-P7), mais ces sites ne sont pas indispensables aux réactions d'intégration et d'excision. Les régions occupées par le facteur d'hôte miHF sont représentées par des rectangles vert. La réaction d'intégration nécessite l'intégrase L5 et la protéine miHF, alors que l'excision nécessite une protéine supplémentaire : l'excisionase Xis. L'excisionase peut se fixer à 4 sites (X1-X4, rectangle orange) au sein du site *attR* et *attP*. **b)** Modèle du complexe intégratif du système L5. L'intégrase L5 (formes bleu clair) et miHF (formes verte) peuvent former un complexe intasome-P capable de capturer le site *attB* par des ponts intermoléculaires avec les intégrases fixées aux sites bras P1/P2 afin de former le complexe synaptique 1. Dans une voie alternative, l'interaction synaptique peut avoir lieu en absence de miHF (complexe synaptique 2), suivi par la liaison de miHF et de l'intégrase L5 aux sites bras P4/P5. Le complexe synaptique 1 doit être converti en complexe recombinogène où l'échange de brin a lieu et cela peut être facilité par le surenroulement de l'ADN, ou bien par la fixation de protéines miHF additionnelles. A la suite de l'échange de brin, les produits *attL* et *attR* sont formés.

Quelques années plus tard, la protéine mIHF, une protéine ne semblant pas appartenir à la grande famille de protéines IHF/HU et se fixant de façon non-spécifique à l'ADN, a été caractérisée et favorise la formation d'un intasome sur le site *attP* révélant une nouvelle fois l'adaptation du bactériophage à son hôte bactérien (Pedulla *et al.*, 1996). Le site *attP* est complexe, de 252 pb et composé de 5 sites bras de fixation et de deux sites cœur de fixation de L^5 Int, encadrant une région centrale de 7 pb (Fig. 34a). Parmi les 5 sites bras, seul les sites P1, P2, P4 et P5 sont nécessaires à l'intégration (Peña *et al.*, 1997). De son côté, le site *attB* est simple, avec une taille de 29 pb et présentant deux sites cœur de fixation de L^5 Int (Peña *et al.*, 1996). Bien que les éléments de la mise en œuvre de l'intasome du bactériophage L5 soient similaire à ceux du paradigme λ Int, le mécanisme utilisé est différent. Dans ce modèle, deux voies sont possibles. Dans une première voie, L^5 Int et la protéine architecturale mIHF peuvent former l'intasome-P par la formation d'un pont intramoléculaire entre les intégrases liées aux sites bras P4-P5 et aux sites cœur du site *attP* (Fig. 34b). La formation de ces ponts intramoléculaire est possible grâce à la courbure de l'ADN induite par la protéine mIHF (Peña *et al.*, 1999). L'intasome-P recrute alors le site *attB* et les intégrases liées aux sites bras P1/P2 formant ainsi le complexe synaptique 1. Dans une voie alternative, en absence de la protéine mIHF, L^5 Int est capable de former le complexe synaptique 2 par fixation des sites bras P1-P2 et des sites cœur du site *attB* (Peña *et al.*, 2000). La liaison de la protéine mIHF et de L^5 Int aux sites bras P4-P5 permet la formation du complexe synaptique 1. Ce complexe semble être un complexe ouvert qui nécessite la fixation de protéines mIHF pour courber l'ADN et permettre la formation d'un tétramère où aura lieu la réaction d'échange de brin. Enfin, la recombinaison permet la production des produits *attL* et *attR*.

De son côté, l'intégrase HP1 permet l'intégration du phage tempéré HP1 dans le gène codant un ARN^{t^{Leu}} de la bactérie hôte *Haemophilus influenzae* (Goodman and Scocca, 1989). En absence de cofacteur, la réaction de recombinaison est peu efficace, mais la recombinaison est fortement stimulée par l'ajout d'un extrait cellulaire d'*H. influenzae* ou bien par la présence de la protéine IHF, une protéine conservée chez cet organisme, évoquant à nouveau l'adaptation du bactériophage à la physiologie de son hôte (Hwang and Scocca, 1990). Le site *attP* est complexe, d'environ 500 pb et contient deux types de sites de fixation de $HP1$ Int : 3 paires de sites de forte affinité (sites bras) (IBS1, IBS2, IBS5) et 3 paires de sites de faible affinité (sites cœur) (IBS3, IBS4, IBS6) (Fig. 35a) (Hakimi and Scocca, 1994). Le site IBS4 est le site où a lieu la réaction d'échange de brin.

Etude bibliographique

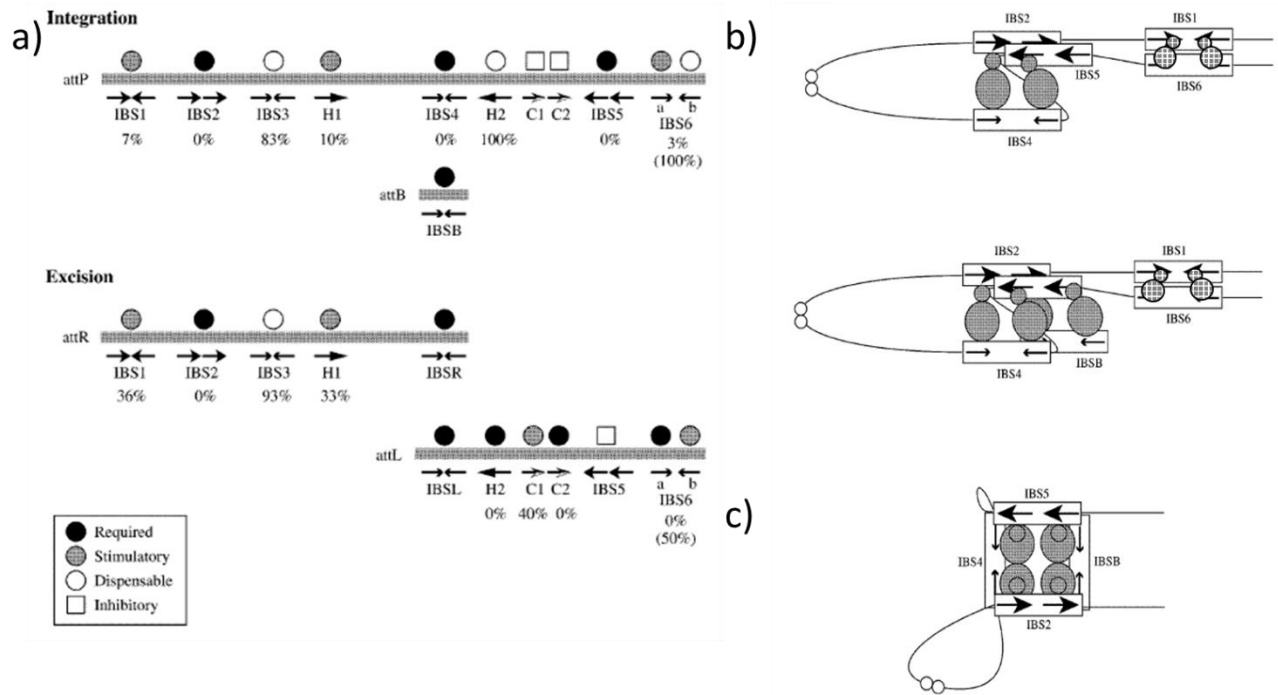


Fig. 35 Système de recombinaison du bactériophage HP1 (Adapté de Esposito *et al.*, 2001). **a)** Représentation graphique des sites de recombinaison de l'intégrase HP1. Les sites de fixation des protéines sont indiqués par des flèches noires et appelés IBS « Intégrase Binding Site » (^{HP1}Int), H (protéine IHF) et C (protéine Cox, « Control Of eXcision »). Les pourcentages indiquent la quantité de recombinaison obtenue pour l'intégration et pour l'excision lorsque le site est muté. Pour le site IBS6, le pourcentage d'activité de la mutation du demi-site IBS6b est représenté entre parenthèse. La couleur des cercles représente la nécessité des sites pour la recombinaison : noir, le site est absolument requis ; gris, le site stimule la recombinaison et blanc, site facultatif pour la recombinaison. Les carrés signifient que l'occupation de ces sites est délétère pour la recombinaison. **b)** Modèle du complexe intégratif de l'intégrase HP1. Les rectangles représentent les sites de fixation de ^{HP1}Int. Les sites cœur sont représentés par des flèches fines et les sites bras par des flèches épaisses. IHF est représenté par deux petits cercles blanc. Le petit cercle gris représente le domaine N-ter de ^{HP1}Int qui se fixe sur les sites bras et le grand cercle gris représente le domaine C-ter de ^{HP1}Int qui se fixe sur les sites cœur. Les monomères d'intégrase supplémentaires fixés sur les demi-sites IBS2 et IBS5 n'ont pas été représentés. **c)** Une vue du haut du complexe synaptique présenté en **b)**.

Une spécificité intrigante de ce site *attP* est donc la présence de deux paires de sites cœur supplémentaires (IBS3 et IBS6) par rapport au système λ où les sites cœur C et C' ne sont présent qu'une seule fois sur le site *attP*. Deux sites de fixation (H1 et H2) de IHF ont été déterminés sur le site *attP* (Hwang and Scocca, 1990), ainsi que deux sites de fixation (C1 et C2) de la protéine Cox, l'excisionase du système (Esposito and Scocca, 1997). De son côté, le site *attB* est simple, avec une taille de 18pb et présentant une seule paire de site cœur (IBSB) (Fig. 35a) (Hauser and Scocca, 1992). En raison de l'architecture inhabituelle du site *attP*, il était présumé que le mécanisme de mise en place de l'intasome serait différent de celui du système λ .

Dans ce modèle, la fixation de IHF au site H1 permet la formation d'un pont intramoléculaire entre l'intégrase fixée sur un demi-site IBS2 par son domaine N-ter et sur un demi-site IBS4 par son domaine C-ter (Fig. 35b). Un deuxième pont intramoléculaire stabilise cet intasome formé sur le site *attP* entre les intégrases fixées au site IBS1 par leur domaine N-ter et au site IBS6 par leur domaine C-ter. Cette boucle permet l'alignement du site IBS5 avec le site IBS4 et permet la fixation d'une intégrase sur un demi-site IBS5 *via* son domaine N-ter, ainsi que sur le deuxième demi-site IBS4 *via* son domaine C-ter. Dans cette structuration, un demi-site de IBS2 et un demi-site de IBS5 sont libres. Grâce à la forte coopérativité de liaison des intégrases, ces sites seront fixés par le domaine N-ter d'une intégrase, laissant deux domaines C-ter libres pour la fixation d'un site de faible affinité tel que *attB*. Le tétramère d'intégrases peut ainsi réaliser les réactions de coupure et d'échange de brins entre les sites cœur IBS4 (*attP*) et IBSB (*attB*) (Fig. 35c) (Esposito *et al.*, 2001).

Enfin, le bactériophage KplE1 (ou CPS-53) est un prophage déficient intégré dans le chromosome de la souche *E. coli* MG1655 dans le gène codant un ARN^t^{Arg} (Rudd, 1999). Un prophage déficient est un virus ayant perdu certaines des fonctions essentielles à la croissance lytique et est donc incapable de libérer des particules infectieuses dans l'environnement. Bien que le prophage KplE1 soit déficient, il est toujours capable d'être excisé *in vivo* (Elantak *et al.*, 2005). Le site *attP* (Fig. 33) est composé de 6 sites bras de fixation et de deux sites cœur de fixation de l'intégrase IntS (Panis *et al.*, 2007). Parmi les 6 sites bras, les mêmes sites (P1, P2, P'1, P'2 et P'4) sont nécessaires à l'intégration et à l'excision contrairement au paradigme λ Int où deux jeux de sites différents sont utilisés (Panis *et al.*, 2010). De plus, seul le site H2 de fixation de la protéine IHF est nécessaire aux réactions bien que la présence de ce cofacteur ne soit pas essentielle à la recombinaison.

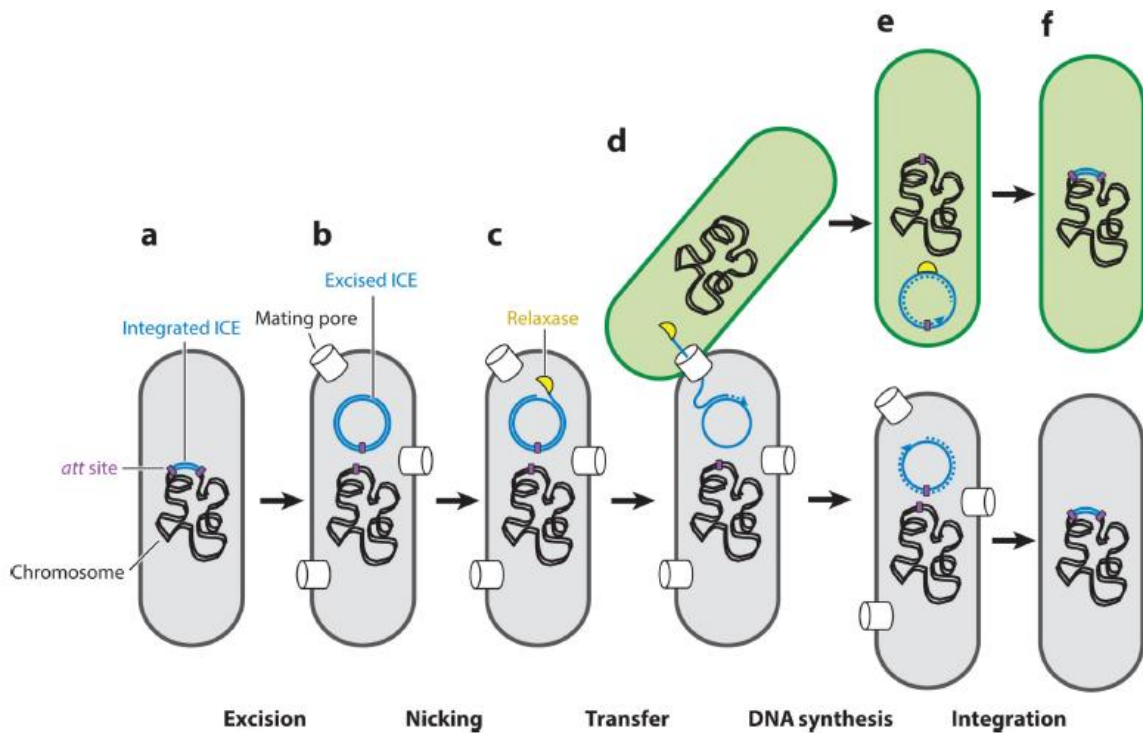


Fig. 36 Cycle de vie des « Integrative and Conjugative Elements (ICEs) (Johnson and Grossman, 2015). La bactérie possédant l'ICE (bactérie donneuse) est représentée en gris et la bactérie recevant l'ICE (bactérie réceptrice) est représentée en vert. L'ADN de l'ICE (bleu), le chromosome bactérien (noir), les sites *att* de recombinaison (violet), la relaxase (forme jaune) et le « mating pore » (forme cylindrique blanche) sont représentés. **a)** L'ICE est présent sous sa forme intégré dans le chromosome de sa bactérie hôte. **b)** Après induction de l'expression des gènes de l'ICE, la réaction de recombinaison est effectuée par l'intégrase et l'excisionase, excisant l'ICE sous une forme d'ADN circulaire double-brin. **c)** La relaxase codée par l'ICE induit une coupure simple-brin de l'ICE et se fixe de manière covalente à l'extrémité 5' de l'ADN coupé permettant la formation de l'ADN de transfert. **d)** Si une cellule réceptrice est disponible, la machinerie de conjugaison permet la transmission de l'ADN de transfert à cette cellule. **e)** Dans la cellule réceptrice, la relaxase permet la ligature des extrémités 5' et 3' de l'ADN afin de former un ADN circulaire simple-brin. L'ADN complémentaire de cet ADN simple-brin est synthétisé afin de former un ADN circulaire double-brin, substrat de la réaction d'intégration dans le chromosome. **f)** L'intégrase permet l'intégration de l'ICE dans le chromosome.

En effet, la présence d'IHF stimule la réaction d'intégration *in vitro* d'un facteur 4-6 et l'absence de cette protéine réduit seulement d'un facteur 20 l'excision du phage KplE1 *in vivo* contrairement au bactériophage λ où cette protéine est essentielle à l'intégration et à l'excision du virus (Panis *et al.*, 2007; Panis *et al.*, 2010). Ainsi, le bactériophage KplE1 et le bactériophage λ semble avoir évolué différemment dans la mise en place de leur intasome chez la même bactérie hôte ce qui renforce l'idée que la formation d'un intasome complexe pour la capture du site *attB* est un processus important mais que sa structuration semble être différente en fonction de chaque couple phage tempéré/cellule hôte étudié.

II.9.2.3.2) La formation de l'intasome : un mécanisme partagé avec les ICEs utilisant une intégrase hétérobivalente

Les « ICEs » ou « Integrative and Conjugative Elements » sont des éléments génétiques mobiles, de taille variable (environ 20 kb à plus de 500 kb) possédant les gènes nécessaires à leur conjugaison, à leur intégration et à leur excision, ainsi que des gènes « cargo » de différentes fonctions. Ces gènes permettent aux cellules l'acquisition de nouveaux phénotypes : résistance à un antibiotique, capacité à métaboliser une nouvelle source de carbone, acquisition de nouveaux gènes de virulence... (Johnson and Grossman, 2015). Comme leur nom l'indique, ces éléments sont retrouvés dans deux états physiologiques : soit intégrés dans le chromosome de leur cellule hôte, soit excisés hors du chromosome où ces éléments peuvent être transmis à une nouvelle cellule par conjugaison. Afin d'effectuer les réactions d'intégration et d'excision, ces éléments codent généralement une recombinase à tyrosine, bien que certains éléments codent une recombinase à sérine ou une transposase de type DDE (acide aspartique-acide aspartique-acide glutamique) (Wozniak and Waldor, 2010). Sous leur forme intégrée (Fig. 36a), lors de conditions spécifiques, les ICEs sont excisés par recombinaison grâce à leur intégrase et leur excisionase aboutissant à la formation d'une molécule d'ADN circulaire double-brin. Ensuite, plusieurs protéines exprimées par l'ICE s'assemblent afin de former le « mating pore » permettant à l'ICE de transférer son ADN (Fig. 36b). Les protéines de l'hôte et de l'ICE reconnaissent l'origine de transfert *oriT* et permettent la formation d'un complexe nucléoprotéique simple-brin appelé ADN de transfert (Fig. 36c). Si une cellule réceptrice est disponible, la machinerie de conjugaison prend en charge l'ADN de transfert afin de permettre sa transmission (Fig. 36d).

L'ADN simple-brin de l'ICE de la cellule réceptrice est répliqué afin de former un ADN circulaire double-brin (Fig. 36e) qui sera le substrat de l'intégration dans le chromosome (Fig. 36f). Si l'ICE est maintenu dans la cellule donneuse, il est proposé qu'il soit réintégré dans le chromosome grâce à un évènement de réintégration médié par l'intégrase (Johnson and Grossman, 2015). La majorité des ICEs utilisent un système de sécrétion de type IV pour transférer leur ADN sous la forme simple-brin. Cependant, les ICEs d'actinomycètes transfèrent leur ADN sous la forme double-brin *via* un système conjugatif utilisant un homologue de la protéine FtsK (Bordeleau *et al.*, 2012). Malgré le grand nombre d'ICEs putatifs trouvés par séquençage génomique, peu de systèmes ont été caractérisés expérimentalement, parmi lesquels Tn916 et CTnDOT (Conjugative Transposon from *Bacteroides thetaiotaomicron* DOT).

Tn916 est un ICE de 18 kb possédant un gène de résistance à la tétracycline et qui a été identifié pour la première fois chez *E. faecalis* (Franke and Clewell, 1981). Cet ICE s'intègre dans de nombreux sites chromosomiques (Gawron-Burke and Clewell, 1982), riches en séquence A/T et est fonctionnel dans de nombreux phyla bactérien (Roberts and Mullany, 2009). L'intégrase ^{Tn916}Int est une intégrase hétérobivalente (Lu and Churchward, 1994) possédant le motif catalytique R-K-H-R-H-Y (Nunes-Düby *et al.*, 1998). La structure RMN en solution du domaine N-ter de ^{Tn916}Int a été déterminée et présente un motif de type « three-stranded β -sheet » similaire à celui de λ Int (Wojciak *et al.*, 1999). Le site de recombinaison *attP* présente une organisation comparable à celle des intégrases hétérobivalentes de phage : 5 sites bras et 2 sites cœur de fixation de ^{Tn916}Int sont retrouvés (Lu and Churchward, 1994) ainsi que 3 sites de fixation de l'excisionase (Fig. 33) (Connolly *et al.*, 2002; Rudy *et al.*, 1997b). De son côté, le site *attB* est simple et riche en A/T.

Bien que la structuration de l'intasome ne soit pas connue, ce système présente plusieurs particularités par rapport aux intégrases hétérobivalentes classiques. Tout d'abord, la région centrale semble être de 6 pb (Taylor and Churchward, 1997), bien qu'il n'est pas exclu qu'elle puisse être de 7 ou 8 pb (Rajeev *et al.*, 2009). La caractéristique de cette région centrale est que les 6 pb sont différentes entre les sites *attB* et *attP*, brisant le dogme de l'identité de séquence dans la région centrale parmi les recombinases à tyrosine. Etant donné l'importante différence au niveau des séquences, il est suggéré que la réaction de recombinaison implique la formation d'une jonction de Holliday qui serait résolue par réplication ou réparation de l'ADN (Wozniak and Waldor, 2010). Une autre différence majeure de Tn916 provient de l'excisionase.

En effet, contrairement aux autres systèmes hétérobivalents, à l'exception du système mv4, la protéine Xis n'inhibe pas la réaction d'intégration dans la cellule réceptrice (Marra *et al.*, 1999). De plus, en fonction de l'organisme hôte, la protéine Xis est soit essentielle, soit stimule la réaction d'excision. Par exemple, chez la bactérie *E. coli*, la protéine Xis stimule la réaction d'excision (Lu and Churchward, 1995; Poyart-Salmeron *et al.*, 1989). Une autre particularité provient de la répartition des sites de fixation de l'excisionase Xis qui sont retrouvés sur les deux bras P et P' indiquant possiblement un rôle particulier de Xis par rapport aux systèmes classiques (Fig. 33). Dans le système Tn916, cette protéine a un rôle double lors de l'excision : sa fixation sur le site X1 et/ou X2 du site *attL* (Fig. 33) stimule la réaction d'excision, notamment en courbant l'ADN et facilitant la fixation de l'intégrase sur ses sites bras et cœur (Connolly *et al.*, 2002) alors que sa fixation sur son site X' du site *attR* inhibe la réaction d'excision (Fig. 33) (Hinerfeld and Churchward, 2001). Ainsi, Xis exerce un rôle régulateur dans l'excision et le cycle biologique de l'ICE Tn916. Enfin, la très faible efficacité de recombinaison *in vitro* (Rudy *et al.*, 1997a) suggérait l'utilisation d'un facteur d'hôte partagé entre les nombreux organismes où Tn916^{Int} est fonctionnelle. Chez *E. coli*, il a été montré que la protéine IHF n'est pas essentielle à l'excision, à la différence de la protéine HU où l'absence de cette protéine engendre une diminution de 90% de la fréquence d'excision (Connolly *et al.*, 2002). HU est une « Nucleoid-Associated Protein » (NAPs) qui semble ubiquitaire parmi les procaryotes et dont la structure protéique est fortement conservée parmi ces organismes (Kamashev *et al.*, 2017) . Ainsi la mise en place de l'intasome de l'ICE Tn916 semble avoir évolué au cours du temps par dépendance envers un cofacteur conservé permettant la dispersion de cet élément génétique chez de nombreux organismes.

CTnDOT est un élément génétique mobile de 65 kb contenant des gènes de résistance à l'érythromycine et à la tétracycline. Au cours de ces dernières années, la prévalence de CTnDOT dans les génomes du genre *Bacteroides* a fortement augmenté, se retrouvant dans plus de 80% des isolats naturel (Wood and Gardner, 2015). A la différence de l'ICE Tn916, CTnDOT présente une sélectivité de site plus restreinte puisqu'il effectue une recombinaison spécifique de site dans environ 7 sites présents sur le chromosome de *Bacteroides* (Bedzyk *et al.*, 1992). L'intégrase IntDOT est responsable de l'intégration ainsi que de l'excision de CTnDOT et présente 5 des 6 résidus caractéristiques des recombinases à tyrosine (R-K-H-R-H-Y) dont la tyrosine catalytique (Cheng *et al.*, 2000). Le premier résidu arginine de ce motif est absent, remplacé par un résidu sérine qui ne semble pas impliqué dans le site catalytique.

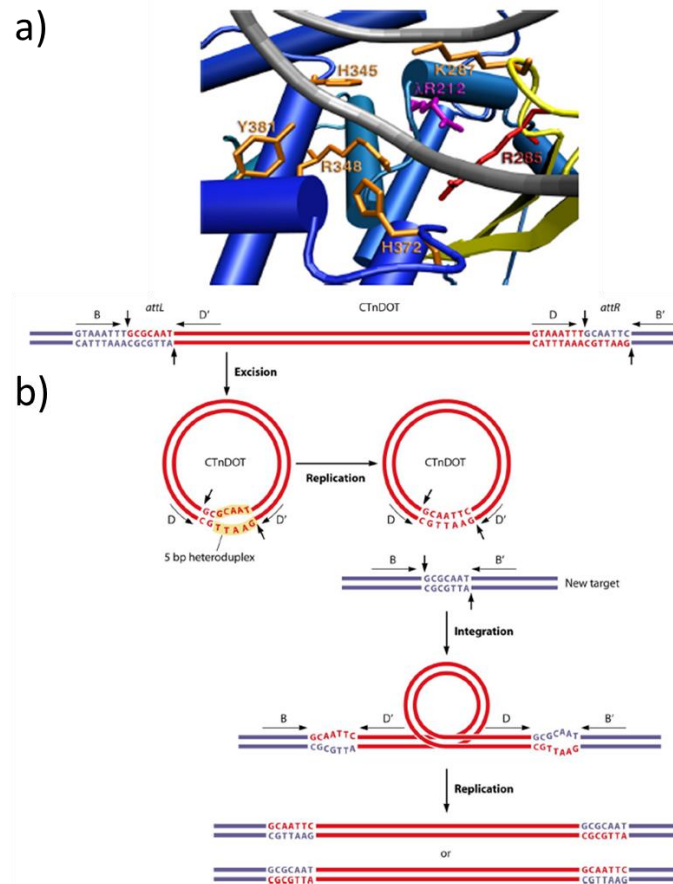


Fig. 37 Modélisation du site actif de l'intégrase IntDOT et représentation des réactions d'intégration et d'excision de l'ICE CTnDOT (Adapté de Rajeev *et al.*, 2009 ; Kim *et al.*, 2010). a) Prédiction de la conformation du site actif de IntDOT et représentation de la fixation du domaine catalytique à un site cœur. L'ADN est représenté en gris et les différentes structures (hélice- α et feuillet- β) du domaine catalytique en bleu. La région de IntDOT qui interagit avec l'ADN et contient les résidus R285 et K287 est indiquée en jaune. La chaîne latérale des résidus du site actif sont indiquées en orange sauf pour le résidu R285 en rouge. Le positionnement classique du résidu R(I) conservé parmi les recombinases à tyrosine, ici le résidu R212 de λ Int, est représenté en violet. Le résidu R212 de λ Int et R285 de IntDOT entre dans le site actif à partir d'une position différente mais peuvent interagir avec la même région de l'ADN. b) Modèle d'intégration et d'excision de CTnDOT. Les sites D et D' sont les sites cœur du site *attDOT* et les sites B et B' sont les sites cœur du site *attB* présents sur le chromosome de la bactérie hôte. L'excision s'effectue entre les sites *attL* et *attR*, par coupure et échange de la région centrale de 7 pb. L'excision aboutit à la formation d'un intermédiaire circulaire, présentant des mésappariements de l'ADN des 5 derniers nucléotides de la région centrale, qui sera résolu après conjugaison et synthèse du brin complémentaire dans la cellule réceptrice. L'intégration de CTnDOT s'effectue par coupure et échange de la région centrale de 7 pb entre le site *attDOT* et le site *attB*. La réaction produit des sites *attL* et *attR* contenant des mésappariements qui seront résolus par réparation ou réplication de l'ADN.

L'absence de ce résidu essentiel pour les recombinases à tyrosine suggère que la structure du site actif de cette intégrase pourrait être unique parmi les recombinases à tyrosine (Kim *et al.*, 2010; Malanowska *et al.*, 2006). En absence de structure, il est proposé qu'un autre résidu arginine situé à une position atypique (R285) soit le résidu impliqué dans la catalyse qui interagirait avec le site cœur d'une manière équivalente au résidu R(I) des recombinases à tyrosine classique (Fig. 37a) (Kim *et al.*, 2010). Comme les intégrases hétérobivalentes, IntDOT est composée de trois domaines : un domaine N-ter capable de se fixer aux sites bras, un domaine « core-binding » ainsi qu'un domaine catalytique capables de se fixer aux sites cœur. Le site *attDOT* est complexe et présente 6 sites bras et 2 sites cœur de fixation de l'intégrase encadrant une région centrale de 7 pb (Fig. 33) (DiChiara *et al.*, 2007; Malanowska *et al.*, 2007; Wood *et al.*, 2010). Bien que la protéine IHF d'*E. coli* améliore les réactions d'intégration et d'excision, il a été montré qu'un cofacteur présent chez l'organisme hôte *Bacteroides* était impliqué dans la recombinaison. Cette protéine BHFa (*Bacteroides* Host Factor A) est un homodimère, membre de la famille des protéines IHF/HU, qui se fixe spécifiquement à 4 régions (H1, H2, H3 et H4) dans le site *attDOT* (Fig. 33) et qui semble favoriser la formation d'un intasome en courbant l'ADN (Ringwald and Gardner, 2015). Classiquement, la réaction d'intégration nécessite l'intégrase IntDOT ainsi que le facteur d'hôte BHFa. En revanche, contrairement aux intégrases hétérobivalentes classiques, la réaction d'excision nécessite en plus 3 protéines codées par CtnDOT : Exc, Xis2c et Xis2d. Exc est une topoisomérase de type Ia dont l'activité catalytique n'est pas nécessaire à l'excision (Sutanto *et al.*, 2002). Cette protéine n'est pas essentielle mais stimule l'excision d'un facteur 3-5. Comme Exc ne semble pas se fixer sur les sites *att*, il est suggéré que cette protéine joue un rôle structural, soit en participant à des interactions protéine-protéine, soit en stabilisant un intermédiaire de la réaction de recombinaison (Keeton and Gardner, 2012). Les protéines Xis2c et Xis2d sont des petites protéines basiques qui agissent comme facteur directionnel de la recombinaison en activant l'excision et inhibant l'intégration (Sutanto *et al.*, 2004). Ces protéines se fixent chacune sur deux régions dans le site *attDOT* (C1, C2 pour Xis2c et D1, D2 pour Xis2d) (Fig. 33) (Keeton *et al.*, 2013).

La particularité du système de recombinaison de CtnDOT provient de la région centrale des sites de recombinaison *attDOT* et *attB* (Fig. 33 ; Fig. 37b). Cette région centrale est composée de 7 nucléotides dont les 2 premiers sont identiques et essentiels à la réaction de recombinaison, ainsi que de 5 nucléotides différents, brisant également le dogme de l'identité de séquence dans la région centrale (Fig. 37b).

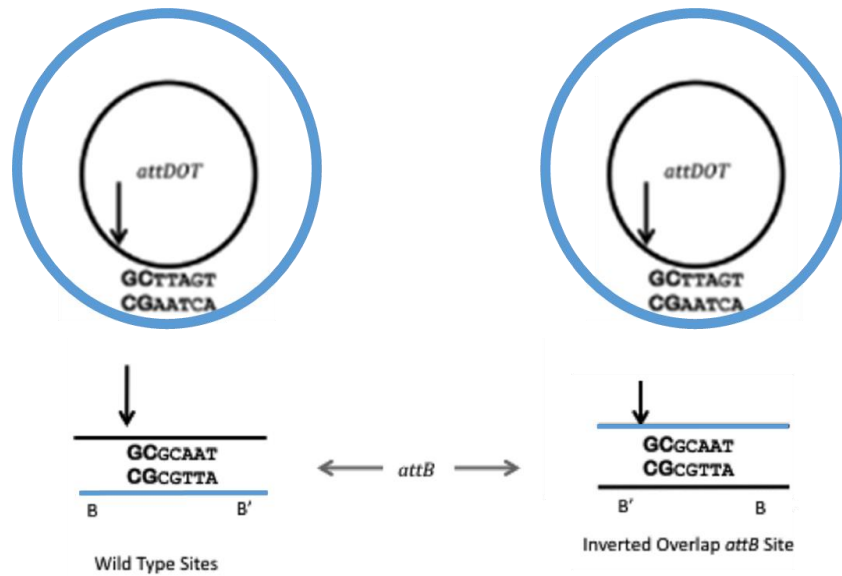


Fig. 38 Schéma représentant l'orientation des sites de recombinaison en fonction de la séquence centrale du site *attB* dans le système CTnDOT (Adapté de Laprise *et al.*, 2010). Les brins supérieurs sont représentés en noir et les brins inférieurs en bleu. La flèche verticale indique le site de clivage initial. Si le substrat *attB* contient les dinucléotides GC sur le brin supérieur en 5' de la région centrale, le brin supérieur sera clivé et échangé en premier alors que si ces dinucléotides se situent sur le brin inférieur en 5' de la région centrale, le brin inférieur sera échangé en premier.

Ainsi, contrairement aux recombinases à tyrosine classiques, IntDOT réalise une réaction de recombinaison en trois étapes : un premier échange de brin dépendant de l'identité de séquence qui nécessite deux paires de bases d'ADN complémentaire au site de clivage, suivi des étapes d'isomérisation, de clivage, d'échange et de ligature de brins indépendantes de l'identité de séquence. La nature de la base des deux premiers nucléotides n'est pas important à la réaction, seule l'identité l'est (Malanowska *et al.*, 2007). Tout comme λ Int, l'ordre d'échange de brins est dicté et commence toujours par un échange du brin supérieur (Malanowska *et al.*, 2007). En revanche, chez λ Int cet ordre défini est conféré par la formation de l'intasome sur les sites bras, aboutissant à une structure où seules les recombinases clivant le brin supérieur sont actives (Kitts and Nash, 1988; Laxmikanthan *et al.*, 2016). Ici, bien que la recombinaison implique la formation d'un intasome, l'ordre d'échange des brins est dicté par l'emplacement de l'identité de séquence dans les 2 régions centrales (Fig. 38). Si les deux premières bases identiques se trouvent sur le brin supérieur, ce brin est échangé en premier, et si elles se trouvent sur le brin inférieur, ce brin sera échangé en premier (Laprise *et al.*, 2010). Il a été montré *in vitro* que la recombinase IntDOT est capable de former une jonction de Holliday et de réaliser une ligature de brins malgré la présence de mésappariements (Malanowska *et al.*, 2007). En revanche, IntDOT est incapable de résoudre une jonction de Holliday synthétique comportant des mésappariements, indiquant l'importance de la formation de l'intasome dans cette étape de la recombinaison (Kim and Gardner, 2011). En accord avec cette idée, il a récemment été montré que la présence des sites bras et du facteur d'hôte BHFa était nécessaire à la résolution de la jonction de Holliday contenant les mésappariements (Ringwald *et al.*, 2017). Ainsi, il semblerait donc que la formation de l'intasome et l'adaptation à son hôte bactérien aient permis à CTnDOT d'évoluer différemment des recombinases hétérobivalentes classiques afin de réaliser une réaction de recombinaison atypique.

En résumé, bien que la structure de l'intasome du système λ soit la seule caractérisée avec précision (Laxmikanthan *et al.*, 2016; Seah *et al.*, 2014; Tong *et al.*, 2014), les différents systèmes de recombinaison utilisant une intégrase hétérobivalente étudiés semblent tous avoir conservé cette propriété. Cette architecture permet à ces systèmes de recombinaison un contrôle de la directionnalité de la réaction (intégration ou excision) et assure le clivage des brins d'ADN uniquement dans le contexte d'un intasome proprement assemblé. Bien que ces systèmes aient évolués différemment, ils présentent tous une dépendance plus ou moins importante à un ou plusieurs facteurs d'hôtes, courbant l'ADN et favorisant la formation de l'intasome.

En effet, les facteurs d'hôtes sont essentiels à la recombinaison des systèmes λ , CTnDOT, L5 et P2 (Nash and Robertson, 1981; Pedulla *et al.*, 1996; Ringwald and Gardner, 2015; Yu and Haggård-Ljungquist, 1993). Enfin, parmi les différents systèmes où un modèle de formation de l'intasome intégratif a été proposé, le mécanisme de mise en place de l'intasome est à chaque fois différent indiquant un nombre important de structuration possible de ce complexe.

II.10) Les recombinases spécifiques de site : de précieux outils d'ingénierie génétique

A la fin des années 1980, le système Cre/*loxP* fut le premier système de recombinaison spécifique de site à être fonctionnel chez une cellule de mammifère (Sauer and Henderson, 1988). Depuis, la forte spécificité ainsi que l'efficacité de réaction des recombinases spécifiques de site ont été détournées dans le but de développer des outils d'ingénierie des génomes, utilisables dans les trois domaines du vivant.

Les intégrases à sérine, des protéines dont la recombinaison est simple (pas de dépendance aux facteurs d'hôte), spécifique et unidirectionnelle, se sont avérées être de précieux outils. Effectivement, il a rapidement été montré que l'intégrase ϕ C31 était fonctionnelle chez les organismes eucaryotes (Groth *et al.*, 2000), bien que son activité ait été optimisée par ajout d'une séquence NLS (« Nuclear Localization Signal ») (Andreas *et al.*, 2002), et par optimisation des codons (Raymond and Soriano, 2007). Cette protéine a permis l'intégration de nombreux transgènes chez différents organismes hétérologues (Fogg *et al.*, 2014; Olorunniji *et al.*, 2016). En revanche, comme ces organismes ne comportent pas de site de recombinaison dans leur chromosome, un « landing pad » (site de recombinaison) doit être intégré au préalable et sera le lieu des différentes intégrations. Des séquences naturellement présentes sur le génome et ressemblant aux sites naturels de l'intégrase (« pseudo-sites ») peuvent aussi être la cible de la recombinaison (Thyagarajan *et al.*, 2001). Les intégrases à sérine ont aussi été utilisées comme outils de biologie synthétique dans le but d'assembler des voies métaboliques *in vitro* par l'utilisation de sites orthogonaux c'est-à-dire des sites possédant plusieurs combinaisons des 2 pb cohésives, et où seuls les sites présentant les 2 pb complémentaires peuvent réaliser la recombinaison. Ainsi, en utilisant différentes combinaisons de sites orthogonaux de l'intégrase ϕ C31, il a été possible d'assembler 5 gènes d'une voie métabolique d'une taille finale de 8 kb dans un plasmide par la technique SIRA (« Serine Integrase Recombinational Assembly ») (Colloms *et al.*, 2014).

Enfin, la capacité de ces protéines à intégrer avec une haute efficacité des ADN de plusieurs dizaines de kb a permis l'intégration d'un vecteur de 133 kb dans le chromosome de *Drosophila melanogaster* (Venken *et al.*, 2006).

Parmi les recombinases à tyrosine, les systèmes simples Cre/*loxP* et Flp/*FRT* ont été les recombinases les plus utilisées, alors que les intégrases hétérobivalentes ont été peu exploitées en raison de leur dépendance aux facteurs d'hôte. Récemment, le système de recombinaison particulier des intégrons a permis le développement de nouveaux outils d'ingénierie. En effet, en 2010 un intégron synthétique a été construit en assemblant l'opéron tryptophane et, grâce à la capacité de l'intégron à exciser et à réintégrer les différentes cassettes *in vivo*, différentes combinaisons de l'opéron ont pu être obtenues dont certaines ont permis d'optimiser la production de tryptophane (Bikard *et al.*, 2010). Plus récemment, un algorithme a été développé et permet la génération de sites *attC* synthétiques pouvant être intégrés dans pratiquement n'importe quel gène codant une protéine (Nivina *et al.*, 2020). Dans cette étude, le potentiel de cet algorithme à générer des sites *attC* fonctionnels a été validé en intégrant un site *attC* à 4 positions différentes dans le gène *lacZ* de la bactérie *E. coli*. L'intégration du site *attC* est généralement peu mutagène pour la protéine, en modifiant de 3 à 5 AA. Ainsi, grâce à la capacité à intégrer différents sites *attC* dans des protéines, il est possible d'envisager des intégrations de sites *attC* dans des protéines multi-domaines et d'effectuer un brassage de ces différents domaines *in vivo* permettant à la protéine d'acquérir de nouvelles fonctions comme la synthèse de composés chimiques d'intérêt (Nivina *et al.*, 2020).

Les recombinases simples Cre et Flp sont les recombinases à tyrosine les plus utilisées dans l'ingénierie génétique chez les organismes procaryotes et eucaryotes, notamment chez les mammifères où Cre est la recombinase spécifique de site la plus utilisée (Olorunniji *et al.*, 2016) et chez la drosophile où l'utilisation de Flp est très répandue (Bischof and Basler, 2008). Naturellement, la protéine Flp fonctionne à une température d'environ 30°C (température de croissance de *S. cerevisiae*) et il a été montré que cette protéine était thermolabile à 37°C, compliquant son utilisation chez les cellules de mammifères (Buchholz *et al.*, 1996). L'évolution dirigée de la séquence protéique de Flp a permis d'obtenir la protéine Flpe, une protéine présentant une plus forte activité à 37°C (Buchholz *et al.*, 1998). Malgré cette optimisation, Flpe était toujours moins efficace que la recombinase Cre dans les cellules de mammifères, ainsi Flpe a été modifiée en Flpo, par optimisation des codons et par ajout d'un signal de localisation nucléaire. Cette nouvelle protéine présente une activité comparable à celle de la recombinase Cre dans les cellules de mammifères (Raymond and Soriano, 2007).

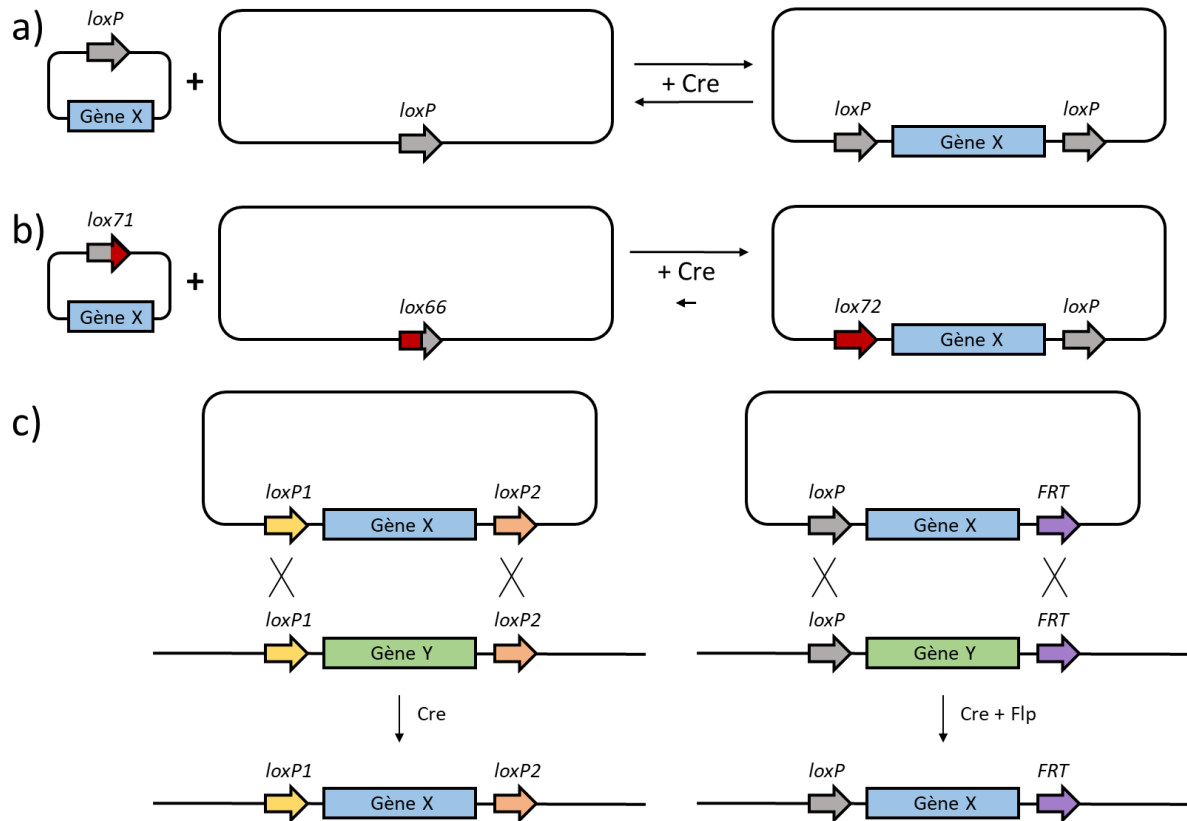


Fig. 39 Intégration, excision et remplacement de cassettes d'ADN par les recombinases Cre et Flp. **a)** Intégration et excision d'une cassette d'ADN par la recombinase Cre. **b)** Stratégie « LE/RE ». La réaction de recombinaison entre deux sites mutés (*lox66* et *lox71*) aboutit à l'intégration de la cassette d'ADN et à la formation d'un site sauvage et d'un site doublement muté, faiblement reconnu par Cre. **c)** Réaction de RMCE et de « dual RMCE ». La réaction de RMCE est représentée à gauche et celle de « dual RMCE » est représentée à droite. Ces réactions permettent le remplacement du gène Y par le gène X grâce à un double événement de recombinaison.

Dans les cellules procaryotes et eucaryotes, les systèmes *Cre/loxP* et *Flp/FRT* sont capables de modifier spécifiquement le génome en réalisant la délétion, l'insertion ou bien l'échange d'ADN (Fig. 39). Un inconvénient de ces systèmes provient du manque de contrôle de la directionnalité des réactions puisqu'après intégration de la cassette d'ADN, la réaction d'excision est possible générant un produit intégré peu stable (Fig. 39a). Rapidement, la stratégie « LE/RE » (« Left Element / Right Element ») a été développée afin d'améliorer le contrôle de la directionnalité de la réaction et donc d'augmenter la stabilité du gène intégré. Dans cette stratégie, les sites de recombinaison contiennent chacun un demi-site sauvage et un demi-site muté (Fig. 39b) et la combinaison de sites LE/RE la plus utilisée est la paire de sites *lox66* (contenant 5 mutations sur le demi-site gauche) et *lox71* (contenant 5 mutations sur le demi-site droit) (Albert *et al.*, 1995). La recombinaison entre ces deux sites permet l'intégration du transgène encadré par un site *loxP* et un site *lox72* (Fig. 39b), un site contenant 5 mutations sur chaque demi-site, faiblement reconnu par Cre et réduisant très fortement l'excision, ce qui améliore la stabilité du produit intégré. Différentes stratégies, nommées RMCE (« Recombinase-Mediated Cassette Exchange ») et « dual RMCE », ont été développées afin de réaliser du remplacement allélique grâce à un double événement de recombinaison au niveau des sites encadrant le fragment (Fig. 39c). Dans la technique RMCE, deux sites incompatibles de la même recombinase sont utilisés (Schlake and Bode, 1994). Ces sites sont incompatibles en raison de l'utilisation de deux régions centrales différentes, ainsi seules les paires de sites avec une région centrale identique peuvent réaliser leur réaction de recombinaison. Dans la technique « dual RMCE », le principe est identique mais repose sur l'utilisation de deux systèmes de recombinaison différents (Fig. 39c) (Lauth *et al.*, 2002). Un inconvénient dans l'utilisation des systèmes *Cre/loxP* et *Flp/FRT* provient de l'obligation d'intégrer un « landing pad » dans le chromosome que l'on souhaite modifier par intégration aléatoire ou par recombinaison homologue avant de pouvoir effectuer les réactions de recombinaison. Si l'on souhaite cibler spécifiquement une séquence naturellement présente dans le chromosome, une évolution dirigée de la recombinase est nécessaire afin d'obtenir des variants capable de reconnaître cette séquence. Par exemple, en utilisant la méthodologie SLiPE (« Substrate-Linked directed Protein Evolution »), Sarkar *et al.*, ont réussi à faire évoluer la recombinase Cre vers une nouvelle protéine nommée Tre, capable d'effectuer une recombinaison dans un site asymétrique retrouvé dans la région LTR (« Long-Terminal Repeat ») du provirus HIV-1, permettant l'excision de ce virus du génome de cellules humaines (Sarkar *et al.*, 2007). En revanche, cette évolution est fastidieuse et dès que l'on souhaite cibler une nouvelle séquence, une nouvelle évolution de la recombinase est obligatoire.

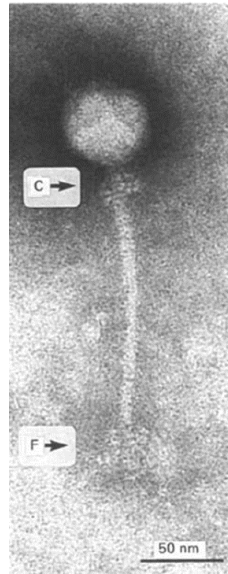


Fig. 40 Observation en microscopie électronique du bactériophage mv4 (Cluzel *et al.*, 1987). La lettre C indique le col, proche de la tête du phage et au-dessus de la queue non contractile. La lettre F indique la structure terminale contenant différentes fibres du bactériophage.

Enfin, bien que les intégrases hétérobivalentes aient été peu utilisées, un système de clonage *in vitro* commercialisé sous le nom de GatewayTM a pu être développé en adaptant le système de recombinaison du bactériophage λ . Ce système exploite la capacité de λ Int à réarranger précisément des fragments d'ADN en utilisant les sites de recombinaison *att*. Par exemple, un fragment d'ADN encadré par deux sites *attB* pourra être remplacé par un autre insert encadré par deux sites *attP* en présence de λ Int et d'IHF en raison des recombinaisons *attB/attP*. L'ajout de la protéine Xis à la réaction permet de restaurer les substrats d'origine grâce aux recombinaisons *attL/attR* (Hartley *et al.*, 2000). Quelques intégrases hétérobivalentes ont aussi pu être utilisées pour intégrer spécifiquement de l'ADN dans un site *attB* présent sur le chromosome bactérien par l'utilisation de plasmides suicides possédant le gène *int*, le site *attP* et un gène de sélection. Par exemple, l'identification de l'intégrase ainsi que du site *attP* du mycobactériophage L5 a permis la construction de vecteur intégratif chez les bactéries *M. smegmatis* et *Mycobacterium tuberculosis* (Lee *et al.*, 1991). Cependant, comme les intégrases hétérobivalentes nécessitent la présence de facteurs d'hôte, et que chaque système *int/attP* est inféodé à son site de recombinaison *attB*, ces systèmes sont difficilement extrapolables à d'autres espèces bactériennes.

III) Le système de recombinaison mv4, un système de recombinaison hétérobivalent atypique

Le bactériophage mv4 est un bactériophage tempéré, isolé à la fin du siècle dernier par induction à la mitomycine C de la souche LT4 de la bactérie *Lactobacillus delbrueckii* ssp. *bulgaricus* (Cluzel *et al.*, 1987; Mata *et al.*, 1986). *L. delbrueckii* est une bactérie lactique appartenant au genre *Lactobacillus* et qui est largement utilisée dans l'industrie agroalimentaire notamment dans la fabrication de yaourt. Les bactériophages issus des bactéries *L. delbrueckii* ssp. *bulgaricus* et *lactis* ont été classés en deux groupes en fonction de la similarité de leur ADN (Lahbib-Mansais *et al.*, 1988; Mata *et al.*, 1986). Le bactériophage mv4 appartient au groupe a, qui est composé de bactériophages lytiques et tempérés (Mata *et al.*, 1986). Son observation par microscopie électronique a permis d'identifier ce virus comme un *Caudovirales* appartenant à la famille des *Siphoviridae*, avec une tête isométrique d'environ 50 nm et une queue non contractile d'environ 180 nm (Fig. 40) (Cluzel *et al.*, 1987). Son génome est de type ADN double brin de 36 kb présentant des extrémités permutées circulaires et redondantes suggérant un mode d'encapsidation de l'ADN phagique par le mécanisme de la « tête pleine » (Lahbib-Mansais *et al.*, 1992).

Différentes fonctions ont été caractérisées : des gènes codant des protéines de structures de la tête (g18 et g34) (Lahbib-Mansais *et al.*, 1992; Vasala *et al.*, 1993), le module de contrôle de la lysogénie (Coddeville *et al.*, 2007), les gènes et sites impliqués dans l'intégration spécifique de site (Auvray *et al.*, 1997, Auvray *et al.*, 1999a; Auvray *et al.*, 1999b; Dupont *et al.*, 1995), ainsi que l'excisionase Xis (Coddeville and Ritzenthaler, 2010). La région impliquée dans la lysogénie est composée de deux opérons transcrits à partir de deux promoteurs en orientation inverse P_{rep} et P_{tec} . Le modulateur lyse-lysogénie *rep-tec* ressemble à celui de λ *cl-cro*, mais présente la particularité de n'avoir qu'un seul site de fixation pour chaque répresseur, indiquant que la régulation lyse-lysogénie pourrait être moins complexe que dans les autres systèmes caractérisés (Coddeville *et al.*, 2007).

Le gène codant l'intégrase mv4 a été identifié (Dupont *et al.*, 1995). Il code une protéine de 427 AA qui présente de l'homologie avec les recombinases à tyrosine (Dupont *et al.*, 1995; Nunes-Düby *et al.*, 1998). Cette protéine permet l'intégration du génome du bactériophage mv4 dans un gène codant un des ARNt^{SER} (Dupont *et al.*, 1995). La majorité des bactériophages tempérés utilisant une intégrase hétérobivalente (59%) intègrent leur génome dans des gènes codant des ARNt (Williams, 2002) comme les bactériophages HP1, L5 et P22 (Goodman and Scocca, 1989; Lee *et al.*, 1991; Smith-Mungo *et al.*, 1994). La plupart de ces phages possèdent l'extrémité 3' du gène dans leur site *attP*, restaurant ainsi un ARNt fonctionnel après leur intégration dans le chromosome. Il est proposé que la conservation des ARNt comme cibles des intégrases hétérobivalente pourrait provenir des caractéristiques structurales de la boucle anti-codon de ces gènes qui présente une séquence inversée-répétée séparée par une séquence de 7 pb (qui peut être utilisée comme région centrale) (Azaro and Landy, 2002). Cela pourrait aussi provenir de la conservation de séquences des ARNt parmi les différentes bactéries, permettant au bactériophage d'avoir une gamme d'hôte plus importante (Reiter *et al.*, 1989; Williams, 2002). A la différence des bactériophages HP1, L5 et P22 appartenant à la classe Ia qui utilisent un site *attB* symétrique dans la boucle anti-codon, le bactériophage mv4 appartient à la classe III, une classe regroupant 25 bactériophages qui ciblent un site *attB* asymétrique dans l'extrémité 3' de la boucle-T (Williams, 2002).

Comme pour toutes les intégrases hétérobivalentes, le site *attP* du phage mv4 est complexe, avec une taille de 234 pb et présentant plusieurs sites bras de fixation de l'intégrase. En 1999, Auvray *et al.*, avaient déterminé une séquence consensus de 9 pb « 5'-TTAACTAGA-3' » avec le motif « 5'-AACTAG-3' » retrouvé 5 fois sur le site *attP*.

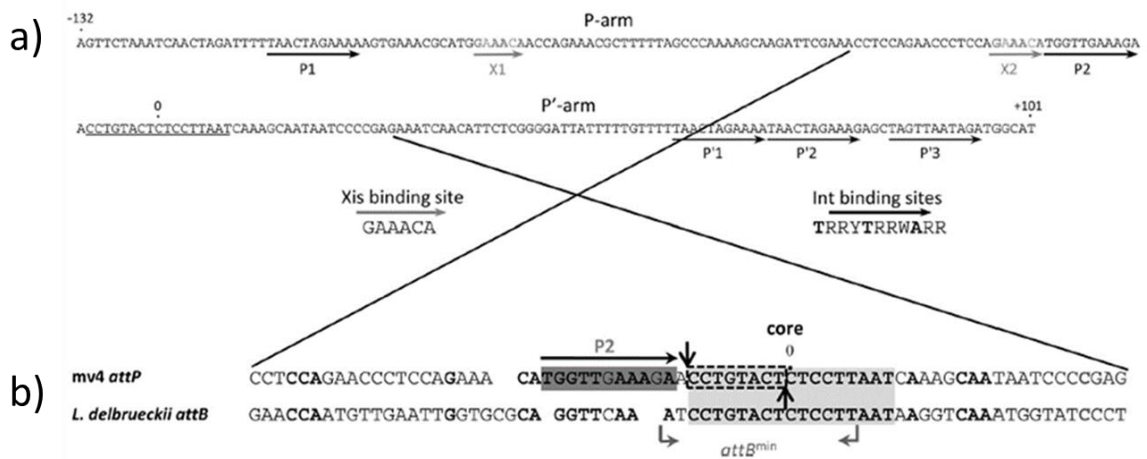


Fig. 41 Représentation des sites de recombinaison *attB* et *attP* du système *mv4* (Adapté de Coddeville *et al.*, 2014). a) Structure et séquence nucléotidique du site *attP*. La région cœur (région d'identité avec *attB*) est soulignée. Le chiffre 0 correspond à la base centrale de la région cœur. Les flèches noires indiquent les 5 sites bras de fixation de *mv4*Int avec la séquence consensus « 5'-TRRYTRRWARR-3' ». Les flèches grises représentent les sites de fixation de Xis avec comme séquence consensus « 5'-GAAACA-3' ». b) Structure et séquence nucléotidique de la région P2-cœur. La séquence cœur de 17 pb (rectangle gris clair), le site bras P2 (rectangle gris foncé) et le site *attB* minimum (flèches grises) sont représentés. Les flèches verticales représentent la région d'échange du brin supérieur et du brin inférieur. La région centrale est indiquée par les pointillés noirs.

Par conséquent, deux sites bras P1 et P2 en orientation directe sont retrouvés sur le bras P, ainsi que trois sites bras P'1, P'2 et P'3 sur le bras P' (Fig. 33). Les sites P'1 et P'2 sont en orientation directe alors que le site P'3 est en orientation inverse (Auvray *et al.*, 1999a). En 2014, Coddeville *et al.*, ont déterminé une nouvelle séquence consensus des sites bras (« 5'-TRRYTRRWARR-3' », W=A/T), déplaçant le site P1 au niveau de l'ancien site P2 et le site P2 à proximité de la région d'échange de brins (Fig. 33 ; Fig. 41a). Ce changement de séquence consensus a aussi déplacé le site P'3 et changé son orientation (Coddeville *et al.*, 2014). Le site P2 se retrouve donc à une position inhabituelle par rapport aux systèmes hétérobivalents décrits (Fig. 33), c'est-à-dire à proximité de la région d'échange de brin, suggérant l'utilisation d'un mécanisme de recombinaison original. Le site *attB* minimum a également été déterminé (Auvray *et al.*, 1999b) : c'est un site de 16 pb, taille la plus courte décrite pour un site *attB* dans la littérature, et décalé de 3 pb de la région cœur (région d'identité de séquence) *attP/attB* (Fig. 41b). Compte tenu de l'asymétrie du site *attB*, aucune séquence inversée répétée n'a pu être caractérisée, ainsi aucun site cœur (B / B' / C / C') de fixation de l'intégrase n'a été déterminé (Auvray *et al.*, 1999a). La région centrale a été déterminée par coupure sur substrat suicide et indique une séquence de 8 pb (Coddeville *et al.*, 2014), une taille inhabituelle puisque les sites de recombinaison des intégrases hétérobivalentes étudiés possèdent plutôt une région canonique de 7 pb (Fig. 33 ; Fig. 41b).

L'excisionase Xis a été caractérisée et correspond à une protéine basique de 56 AA (Coddeville and Ritzenthaler, 2010). Récemment, la structure de la protéine Cox, excisionase du bactériophage P2, a été déterminée et malgré la différence de séquence avec la protéine λ Xis, une structure de type « winged-helix » est conservée (Berntsson *et al.*, 2014). Ainsi, dans le système P2, il est suggéré que les protéines Xis s'assemblent afin de former un filament micro-nucléoprotéique permettant la courbure de l'ADN, comme dans le cas du bactériophage λ . La protéine mv4 Xis contient un domaine putatif de liaison à l'ADN de type Hélice-Tour-Hélice, ainsi cette protéine pourrait fonctionner de façon similaire aux systèmes λ et P2. Deux sites de fixations (X1 et X2) de mv4 Xis ont été déterminés sur le site *attP* (Fig. 33 ; Fig. 41a) et il a été montré que la fixation de cette protéine sur son site X2 induisait une courbure de l'ADN estimée à 90° (Coddeville and Ritzenthaler, 2010). De plus, mv4 Xis est essentielle à la réaction d'excision mais, à la différence des systèmes classiques, cette protéine n'inhibe pas la réaction d'intégration (Coddeville and Ritzenthaler, 2010).

Ainsi, il est proposé que la protéine Xis joue un rôle architectural en courbant l'ADN permettant la formation d'un intasome sur le site *attR*, indispensable à la réaction d'excision, mais que sa fixation sur le site *attP* n'entrave pas la formation d'un intasome essentiel à la réaction d'intégration.

L'intégrase *mv4* (^{mv4}Int) est fonctionnelle dans de nombreux Firmicutes (*L. delbrueckii*, *Lactococcus lactis*, *E. faecalis*, *Streptococcus pneumoniae*) ainsi que chez *E. coli* pour les réactions d'intégration et d'excision (Auvray *et al.*, 1997; Coddeville and Ritzenthaler, 2010; Dupont *et al.*, 1995). *In vitro*, l'intégrase purifiée est fonctionnelle et permet l'obtention d'un produit de recombinaison. De plus, l'ajout des protéines IHF, Fis et H-NS (« Histone-like Nucleoid-Structuring ») n'a pas d'effet visible sur la réaction de recombinaison *in vitro* et seule la protéine HU améliore légèrement la réaction de recombinaison d'un facteur 2-3, suggérant l'indépendance de ce système à un facteur d'hôte (Auvray *et al.*, 1999a).

La réaction de recombinaison de ^{mv4}Int a été étudiée. Cette protéine présente une activité maximale à 42°C, en accord avec la nature thermophile de son hôte bactérien. La réaction de recombinaison est possible entre les sites *attP/attB* et *attP/attL* avec une efficacité similaire, et entre un site *attP/attR* avec une efficacité plus faible d'un facteur 4 (Coddeville and Ritzenthaler, 2010). Pour l'excision, l'efficacité est optimale lorsque l'un des deux sites (*attL* ou *attR*) est surenroulé (Coddeville and Ritzenthaler, 2010). Pour l'intégration, il est essentiel que le site *attP* soit surenroulé (Auvray *et al.*, 1999a). Le rôle de chaque site bras de fixation de ^{mv4}Int dans les réactions d'intégration et d'excision a également été étudié. Pour la réaction d'intégration, le site bras P2 est essentiel et les sites P1 et P'2 ont un rôle important. La mutation des sites P'1 et P'3 n'avait aucun impact négatif sur cette réaction. Pour la réaction d'excision, le site bras de fixation de l'intégrase P'1 joue un rôle important avec le site bras P2 (Fig. 33 ; Fig. 41a) (Coddeville *et al.*, 2014). Ainsi, le site P2 a un rôle important pour les réactions d'intégration et d'excision.

Enfin, l'identification du gène codant ^{mv4}Int et de son site de recombinaison *attP* a conduit à la construction d'un vecteur intégratif (pMC1) contenant ^{mv4}Int et son site *attP* adjacent et permettant l'intégration de ce plasmide dans le gène codant un ARNt^{SER} chez *Lactobacillus plantarum*. La séquence de l'ARNt^{SER} étant très conservée entre *L. delbrueckii* ssp. *bulgaricus* et *L. plantarum*, le site *attB* ciblé est ainsi identique, ce qui explique que le plasmide pMC1 se soit intégré de façon site-spécifique dans ce gène (Dupont *et al.*, 1995). En 1997, Auvray *et al.*, ont déterminé la gamme d'hôte d'intégration du plasmide pMC1.

Ainsi, lorsque les séquences codant un ARNt^{SEr} étaient très conservées (*L. plantarum* et *Lactobacillus casei*), le plasmide était intégré dans cet ARNt. En revanche, si la séquence n'était pas ou peu conservée, le plasmide était intégré dans des sites secondaires (*L. lactis*, *E. faecalis* et *S. pneumoniae*) (Auvray *et al.*, 1997). La capacité du plasmide à être intégré dans des séquences variables suggérait donc une grande flexibilité de l'intégrase vis-à-vis de ces sites de recombinaison. Cette étude montrait notamment l'importance des bases « CCT », et il a été révélé plus tard que ces bases étaient les 3 premiers nucléotides de la région centrale (Coddeville *et al.*, 2014).

En résumé, la réaction de recombinaison médiée par l'intégrase mv4 présente de nombreuses particularités par rapport aux systèmes hétérobivalents classiques. En effet, la réaction de recombinaison est indépendante de la présence de facteurs d'hôte, généralement nécessaire à la courbure de l'ADN et à la formation de l'intasome. Ensuite, les sites de recombinaison *attB* et *attP* présentent une architecture atypique, de par le positionnement du site bras P2 à proximité de la région d'échange de brin, et de par l'absence des sites cœur de fixation de l'intégrase B, B', C et C'. Enfin, le site *attB*, cible de la recombinaison, présente une taille et une structure atypique puisque seulement 2 nucléotides à gauche de la région centrale sont nécessaires à la réaction de recombinaison (Fig. 41b). Ces différentes caractéristiques semblent donc indiquer un mécanisme de recombinaison original parmi les recombinases hétérobivalentes. Plusieurs questions peuvent être posées. En absence des sites cœur B, B', C et C', quel est le rôle du domaine C-ter, et comment l'intégrase est-elle capable de réaliser la réaction de coupure et d'échange de brin au niveau de la région centrale ? Quel est le rôle du site P2, site bras positionné à proximité de la région centrale du site *attP* et essentiel aux réactions d'intégration et d'excision ? En raison de l'asymétrie des sites de recombinaison, l'intégrase est-elle plus permissive à la modification de la région cœur de ses sites par rapport aux intégrases hétérobivalentes classiques ?

L'objectif de ma thèse a donc été de caractériser la région cœur atypique des sites de recombinaison *attB* et *attP* et de redéfinir le système de recombinaison du bactériophage mv4. Le chapitre 1 présentera la redéfinition de la séquence de l'intégrase mv4 ainsi que le développement et l'optimisation des différentes méthodologies utilisées. Le chapitre 2 proposera une méthode basée sur l'utilisation de banques d'ADN présentant des séquences aléatoires afin de caractériser précisément la région cœur des sites de recombinaison. La présence des sites bras et cœur de fixation de mv4Int a pu être validée expérimentalement par retard sur gel et une nouvelle architecture des sites de recombinaison est proposée.

De plus, une analyse NGS des séquences cœur des sites *attB* et *attP* a été réalisée et a permis de classer des milliers de séquences en fonction de leur efficacité de recombinaison. Le chapitre 3 présentera l'étendue de la flexibilité de ^{mv4}Int en fonction des modifications de la région cœur des sites de recombinaison *in vitro* et *in vivo*. Enfin, une discussion générale de la thèse sera suivie d'une conclusion et différentes perspectives de ce travail seront proposées.

**Chapitre 1 : Caractérisation de
l'intégrase mv4 (^{mv4}Int) et mise en place
des outils d'étude**

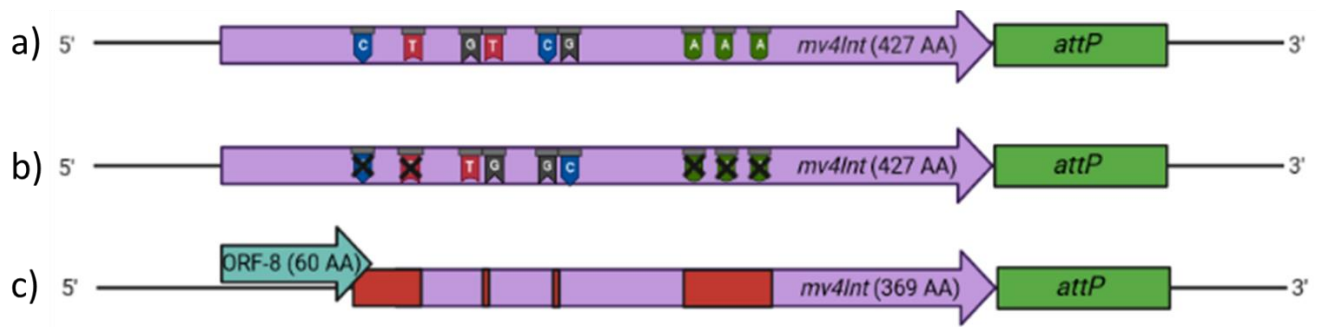


Fig. 42 Représentation schématique de la séquence de l'intégrase mv4 ($mv4Int$) et des conséquences des erreurs de séquençage. a) Séquence de la région $mv4Int/attP$ publiée. Le gène *int* code une protéine de 427 AA. Les bases qui diffèrent : C (bleu), T (rouge), G (noir) et A (vert) sont représentées. Le site *attP* est indiqué en vert et $mv4Int$ en violet. b) Erreurs de séquençage retrouvées dans les plasmides pMC1, pET-Int et sur le génome du bactériophage mv4. Une délétion (croix noire) d'un nucléotide C, d'un nucléotide T et de trois nucléotides A ainsi que l'inversion de la séquence G/T en T/G et C/G en G/C sont retrouvées. c) Impact des mutations sur la séquence de l'intégrase. La couleur rouge indique les régions protéiques différentes par rapport à la séquence publiée.

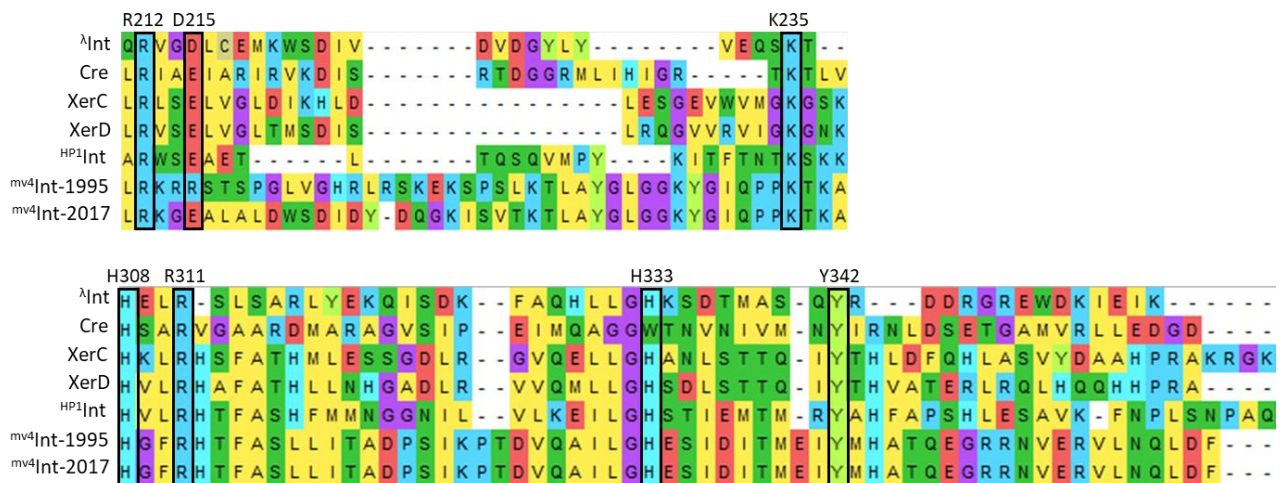


Fig. 43 Alignement des domaines C-ter des recombinases λ Int, Cre, XerC, XerD, $HP1Int$, $mv4Int$ -1995 et $mv4Int$ -2017. Les résidus conservés du domaine catalytique des recombinases à tyrosine sont encadrés en noir. Le résidu lysine (K235) a été ajusté manuellement en s'appuyant sur l'alignement réalisé par Nunes-Düby *et al.*, (Nunes-Düby *et al.*, 1998). Les AA conservés sont numérotés selon la nomenclature λ Int.

Dans ce chapitre nous nous intéresserons à la séquence de l'intégrase mv4 ainsi qu'à la mise en place de la méthodologie d'étude de la réaction de recombinaison. Pour cela, un protocole de purification de l'intégrase a été développé et le stockage de la protéine purifiée a été optimisé. Un test de recombinaison *in vitro* utilisant des fragments fluorescents a été mis au point et optimisé pour l'utilisation d'un extrait cellulaire et d'une protéine purifiée.

I) Redéfinition de la séquence de l'intégrase^{mv4}Int

Afin de continuer l'étude du système de recombinaison du bactériophage mv4, les plasmides pMC1 (Dupont *et al.*, 1995) et pET-Int (Auvray *et al.*, 1999a) ont été récupérés de la collection du laboratoire. En 1995, le gène *int* codant l'intégrase mv4, avait été séquencé au laboratoire (Dupont *et al.*, 1995), identifiant une intégrase à tyrosine de 427 AA (Fig. 42a). Cependant, de nouvelles vérifications de constructions en 2017 indiquaient que le gène *int* du plasmide pMC1 possédait 5 délétions et 4 substitutions nucléotidiques (Fig. 42b) comparé à la séquence publiée. Ces variations ont un fort impact sur la séquence en AA de l'intégrase (Fig. 42c). En effet, la première délétion de la cytosine (Fig. 42b) entraîne l'apparition d'un codon stop prématuré et génère un ORF (ORF-8) codant un peptide de 60 AA. Un nouveau codon d'initiation ATG est retrouvé dans la phase +1 et la délétion suivante d'une thymine (Fig. 42b) a pour conséquence de remettre la séquence de l'ORF sur le même cadre de lecture que celui publié. Les 4 substitutions « GT/TG » et « CG/GC » entraînent un changement de 2 AA et les trois délétions d'adénine dans la seconde moitié de l'ORF entraînent 2 changements successifs puis le retour à la phase de lecture publiée pour le dernier tiers de l'ORF (Fig. 42c). En résumé, les différentes modifications de la séquence nucléotidique du gène *int* donnent une protéine de 369 AA contenant près de 71 AA différents par rapport à la séquence publiée. Les séquences nucléotidiques d'autres plasmides décrits dans la littérature, pA3Int et pET-Int (Auvray *et al.*, 1999a), ainsi que d'un amplicon de la région *int* du phage mv4 ont également montré une séquence correspondant à une intégrase de 369 AA.



Fig. 44 Comparaison des séquences protéiques des deux versions de l'intégrase *mv4Int* et alignement contre d'autres intégrases de *L. delbrueckii* ssp. *bulgaricus*. Les régions différentes de la séquence publiée sont indiquées par les rectangles rouges. MBT8853734, hypothetical protein BTH55_02920 *L. delbrueckii* ssp. *bulgaricus* souche XJ77-7-2 ; MBT9018669, hypothetical protein BTI91_04510 *L. delbrueckii* ssp. *bulgaricus* souche MGC24-1 ; MBT8815572, hypothetical protein BTH97_01860 *L. delbrueckii* ssp. *bulgaricus* souche XJ13-13 ; WP_052933847.1, site-specific integrase *L. delbrueckii* ; AZA16453.1, site-specific integrase *L. delbrueckii* ssp. *bulgaricus* ; WP_119906283.1, site-specific integrase *L. delbrueckii* ; QDH96861.1, site-specific integrase *L. delbrueckii* ssp. *bulgaricus* souche KLD51.1011 ; WP_003622783.1, site-specific integrase *L. delbrueckii* ssp. *bulgaricus*.

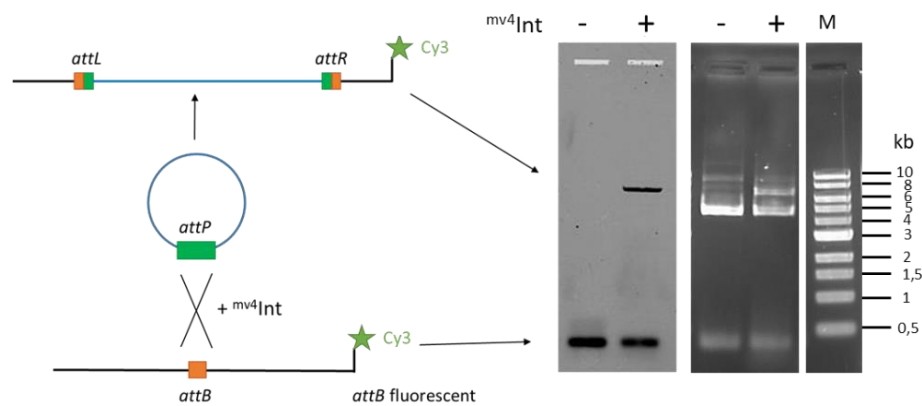


Fig. 45 Principe du test de recombinaison *in vitro*. La présence de *mv4Int* dans la réaction est indiquée au-dessus des pistes. Piste M, échelle de taille 1 kb DNA ladder (New England Biolabs). Gauche, révélation de la fluorescence du gel (cyanine-3) (ChemidocMP, Biorad) ; droite, révélation de la fluorescence de l'ADN après coloration BET et révélation aux UV (ChemidocMP, Biorad).

L'alignement de son domaine C-ter réalisé en 1998 (Nunes-Düby *et al.*, 1998) indiquait que ^{mv4}Int possédait les résidus catalytiques R-K-H-R-H-Y et avait comme particularité de posséder la plus longue distance entre les résidus catalytiques H308 et Y342 (nomenclature ^λInt) (Nunes-Düby *et al.*, 1998). Un nouvel alignement a été réalisé avec les séquences C-ter des recombinaisons ^λInt, Cre, XerC, XerD et ^{HP1}Int (Fig. 43) et indique que, contrairement à la version publiée, le nouveau domaine C-ter comporte bien les 7 résidus conservés de la poche catalytique des recombinaisons à tyrosine : R-D/E-K-H-R-H/W-Y. De plus, l'alignement de cette nouvelle séquence (Fig. 44) montre que ^{mv4}Int est phylogénétiquement très proche d'autres intégrases trouvées dans le génome de différentes souches de *L. bulgaricus*.

II) Caractérisation des protéines impliquées dans la recombinaison

II.1) Mise en place du système de recombinaison *in vitro*

Un test de recombinaison *in vitro* avait été mis en place en 1999 par Auvray *et al.*, (Auvray *et al.*, 1999a). Dans ce test, la réaction nécessitait un fragment linéaire radioactif de 300 pb correspondant à la région chromosomique de *L. bulgaricus* contenant l'ARNt^{SER} de 90 pb (appelé « fragment *attB* » dans la suite du manuscrit), d'un plasmide surenroulé portant le gène *int* suivi de son site de recombinaison *attP* (appelé « plasmide *attP* » dans la suite du manuscrit) et de la protéine ^{mv4}Int (soit purifiée, soit présente dans un extrait cellulaire d'*E. coli*). Ce test de recombinaison *in vitro* a été adapté en remplaçant le fragment radioactif par un « fragment *attB* » identique amplifié par PCR *via* un couple d'amorce dont l'une est marquée à la Cy3 (cf. matériels et méthodes). Le principe de ce test est le suivant (Fig. 45) : après électrophorèse en gel d'agarose du mélange réactionnel, en absence de l'intégrase (piste « - »), une seule bande fluorescente correspondant au fragment linéaire *attB* marqué à la Cy3 est observée, alors qu'en présence de l'intégrase (piste « + »), la réaction de recombinaison aboutit à l'intégration du plasmide dans le fragment linéaire fluorescent, à la formation des sites hybrides *attL* et *attR* et donc à l'apparition d'une bande fluorescente de taille supérieure sur le gel indiquant l'activité de recombinaison.

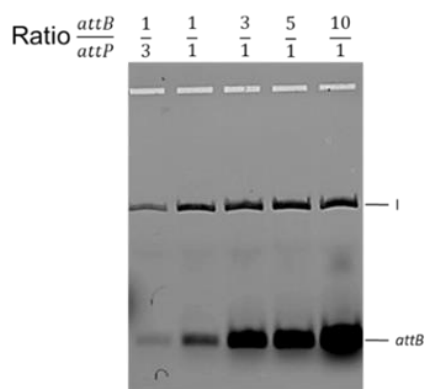


Fig. 46 Etude de la stœchiométrie de la réaction de recombinaison. Les différents ratios $\frac{attB}{attP}$ utilisés dans chaque réaction sont indiqués au-dessus de chaque piste. Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. La réaction contient 40 µg d'extrait cellulaire d'*E. coli* BL21(DE3) ayant surexprimé *mv4Int* et est incubée 1h30 à 42°C.

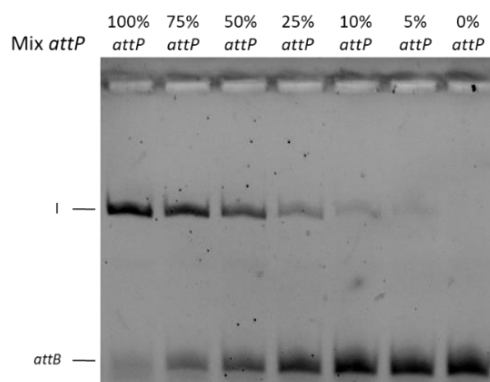


Fig. 47 Etude du seuil minimal de détection du signal de recombinaison *in vitro*. Différents mélanges de plasmides substrats (*attP* et *attP**, *attP* non fonctionnel) ont été utilisés : 300 ng *attP* (100%, pMET311), 225 ng *attP* (75%) et 75 ng *attP** (pMET321), 150 ng *attP* (50%) et 150 ng *attP**, 75 ng *attP* (25%) et 225 ng *attP**, 30 ng *attP* (10%) et 270 ng *attP**, 15 ng *attP* (5%) et 285 ng *attP**, 300 ng *attP** (0%). Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. La réaction contient 7,2 pmol (300 ng) de *mv4Int* supplémentée avec 40 µg d'extrait cellulaire d'*E. coli* préalablement chauffé à 95°C et a été incubée 1h30 à 42°C.

La stœchiométrie de la réaction de recombinaison *in vitro* a été étudiée (Fig. 46). Différents ratios $\frac{attB}{attP}$ ont été utilisés. Le ratio $\frac{1}{3}$ permet l'obtention d'un signal de recombinaison faible et la quasi-totalité des fragments *attB* sont utilisés. A partir du ratio $\frac{1}{1}$ à $\frac{10}{1}$, le signal de recombinaison est intense et saturé puisque l'ajout de fragments *attB* fluorescents n'améliore pas l'intensité du signal de recombinaison. Ainsi, la concentration équimolaire entre les sites *attB* (15 ng) et *attP* (300 ng) permet l'obtention d'un produit de recombinaison intense, tout en limitant l'intensité de fluorescence du site *attB*. Cette condition sera donc conservée par la suite.

Le seuil de détection du système de recombinaison *in vitro* a aussi été analysé (Fig. 47) en mélangeant, dans la réaction, le « plasmide *attP* » avec un plasmide similaire contenant un site *attP* non fonctionnel. Différents mélanges ont été utilisés : 100% *attP*, 75% *attP*, 50% *attP*, 25% *attP*, 10% *attP*, 5% *attP* et 0% *attP*. Plus la quantité de site *attP* fonctionnel est faible, plus le signal de recombinaison obtenu diminue. A une proportion de 5% de plasmide *attP*, le signal de recombinaison est très faible mais reste visible, suggérant que le système de recombinaison *in vitro* que nous avons optimisé est capable de détecter des activités de recombinaison atténuées d'un facteur 20 par rapport au système de recombinaison sauvage.

II.2) Etude de l'importance de l'ORF-8 dans le système de recombinaison

^{mv4}Int/*attP*/*attB*

Bien que rarement décrit dans la littérature, certains systèmes de recombinaison nécessitent plusieurs protéines exprimées par le bactériophage pour catalyser la réaction d'intégration. Par exemple, il a été récemment montré que le gène codant l'intégrase du bactériophage SNJ2, un bactériophage infectant l'Archaea *Natrinema* sp. J7-1, se situe dans un opéron où cette protéine est co-transcrite avec deux autres gènes nécessaires à la recombinaison intégrative (Wang *et al.*, 2018). De plus, bien que hautement improbable, il ne pouvait être exclu que l'intégrase ^{mv4}Int fonctionnelle corresponde en réalité à une protéine plus grande par décalage de phase traductionnelle des 2 ORFs (ORF-8 et ^{mv4}Int) présents dans les plasmides utilisés dans les études précédentes. Dans le but de déterminer le rôle éventuel du peptide codé par l'ORF-8 plusieurs constructions ont été effectuées à partir du plasmide pET-Int (Fig. 48). Ce plasmide permet la surproduction du peptide codé par l'ORF-8 étiqueté en N-ter par une étiquette 6xHis et de l'intégrase mv4 native. A partir de ce plasmide, l'étiquette 6xHis et le gène codant l'ORF-8 ont été supprimées, générant le plasmide pMET332 qui permet la surexpression de l'intégrase mv4 native.

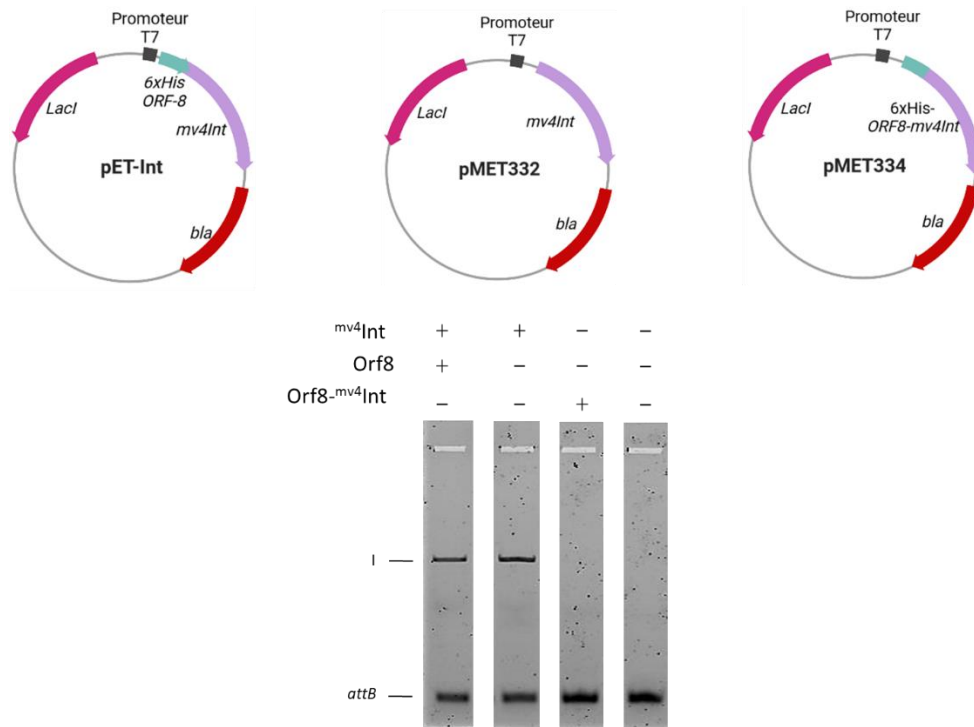


Fig. 48 Etude de l'effet de l'ORF-8 sur la réaction de recombinaison. Les différents plasmides utilisés pour la production de l'extrait cellulaire sont indiqués. L'extrait cellulaire utilisé contient soit *mv4Int* et le peptide 6xHis-ORF8, soit uniquement *mv4Int*, soit la fusion traductionnelle 6xHis-ORF8-*mv4Int*. Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. La réaction contient 40 µg de chaque extrait cellulaire et a été incubée 1h30 à 42°C.

Enfin, la séquence du peptide étiqueté 6xHis-ORF-8 a également été fusionnée avec l'intégrase mv4 afin de mimer un déphasage traductionnel, produisant le plasmide pMET334 (Fig. 48 ; Table 3). Ces trois plasmides ont été transformés dans une souche d'*E. coli* BL21(DE3) et les différentes protéines ont été surexprimées. Un lysat (extrait brut) des différentes souches a été obtenu, et l'activité de recombinaison des différentes constructions a été testée par le test de recombinaison *in vitro* (Fig. 48). Confirmant les résultats précédents (Auvray *et al.*, 1999a), la présence simultanée des protéines 6xHis-ORF-8 et ^{mv4}Int dans un extrait cellulaire permet l'obtention d'un signal de recombinaison intense. Cependant, l'élimination du peptide 6xHis-ORF-8 n'a aucun impact sur l'activité de recombinaison (Fig. 48), signifiant que ce peptide n'est pas important à la réaction et que seule la protéine ^{mv4}Int est responsable de l'activité de recombinaison. Par contre, aucun produit de recombinaison n'est détecté en présence de la protéine chimère 6xHis-ORF-8-^{mv4}Int, suggérant que la fusion traductionnelle est inactive. Chez les intégrases hétérobivalentes classiques, le domaine N-ter est impliqué dans la fixation aux sites bras présents sur le site *attP*, avec une importance des tous premiers AA (Fadeev *et al.*, 2009), il n'est donc pas étonnant que l'ajout de 80 AA en N-ter (pMET334) inactive la recombinaison de l'intégrase, probablement en perturbant la fixation du domaine N-ter aux sites bras.

II.3) Etude de l'effet des modifications K/A et Y/F sur la protéine ^{mv4}Int

La protéine impliquée dans la réaction de recombinaison est donc une intégrase à tyrosine de 369 AA disposant des 7 résidus catalytiques (R-D/E-K-H-R-H/W-Y) (Fig. 43) conservés dans la famille des recombinases à tyrosine (Gibb *et al.*, 2010). Parmi les recombinases modèles, l'importance de la tyrosine nucléophile est généralement démontrée en changeant cet AA par un résidu phénylalanine, comme cela a été montré pour Cre (Gibb *et al.*, 2010), Flp (Pan *et al.*, 1993), ^λInt (Pargellis *et al.*, 1988) et Xer (XerC/D, XerS) (Blakely *et al.*, 1993; Nolivos *et al.*, 2010). De plus, pour certains systèmes, la nécessité de la lysine catalytique (l'acide général de la réaction de coupure de brin) a été vérifiée en changeant la lysine par un résidu alanine chez la recombinase Cre (Lee and Sadowski, 2003), Flp (Chen and Rice, 2003) et XerD (Cao and Hayes, 1999).

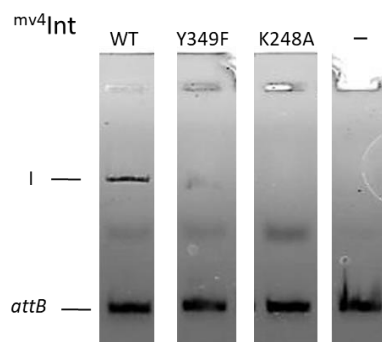


Fig. 49 Etude de l'effet des modifications Y349F et K248A sur l'activité de recombinaison. Les extraits cellulaires contenant les différents variants sont indiqués au-dessus de chaque piste. Pour la réaction de recombinaison, 40 µg de chaque extrait cellulaire a été incubé avec les substrats 1h30 à 42°C. Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. La bande observée dans la piste Y349F est une bande artéfactuelle et ne correspond pas à une activité de recombinaison.

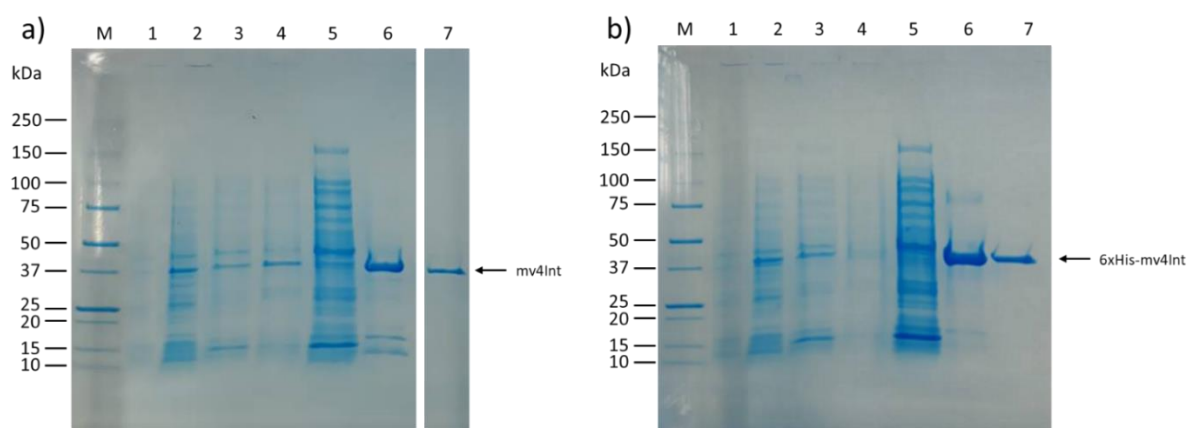


Fig. 50 Surproduction et purification de $mv4Int$ et $6xHis-mv4Int$. **a)** Electrophorèse SDS-PAGE 4-15% d'aliquotes d'extraits protéiques obtenus lors de la purification de $mv4Int$. Pistes : 1, extrait cellulaire *E. coli* BL21(DE3) ; 2, extrait cellulaire induit (IPTG) ; 3, fraction soluble ; 4, fraction insoluble ; 5, fractions non retenues sur la colonne de Ni-NTA ; 6, fractions en sortie de colonne d'héparine ; 7, fractions éluées en sortie de filtration sur gel ; M, marqueur de taille. **b)** Electrophorèse SDS-PAGE 4-15% d'aliquotes d'extraits protéiques obtenus lors de la purification de $6xHis-mv4Int$. L'ordre des pistes est identique à la partie a).

La capacité de recombinaison de deux variants du site actif de ^{mv4}Int a été testée : le mutant Y349F pour confirmer que ^{mv4}Int est une recombinase à tyrosine et le mutant K248A afin de valider expérimentalement que la lysine conservée dans le site actif de ^{mv4}Int décrite en 1998 par Nunes-Düby *et al.*, (Nunes-Düby *et al.*, 1998) correspond bien à la lysine catalytique car chez les recombinases à tyrosine plusieurs résidus lysine se trouvent dans cette région (Fig. 43). Ces deux variants ont été construits par mutagenèse dirigée, permettant de générer les plasmides pMET338 ($^{mv4}Int^{Y349F}$) et pMET339 ($^{mv4}Int^{K248A}$) (Table 3). Ces deux variants ont été surexprimés et produits dans une souche d'*E. coli* BL21(DE3) et les extraits bruts respectifs ont été testés par recombinaison *in vitro* (Fig. 49). Les variants Y/F et K/A ne montre aucune activité de recombinaison, confirmant que les résidus conservés K248 et Y349 sont essentiels à la réaction de recombinaison en jouant respectivement le rôle d'acide général et de nucléophile.

III) Purification de l'intégrase ^{mv4}Int

III.1) Purification de l'intégrase native ^{mv4}Int et de l'intégrase étiquetée 6xHis- ^{mv4}Int

Afin de pouvoir étudier le système de recombinaison ^{mv4}Int , il a été nécessaire de mettre au point un protocole de purification de cette protéine. En 1999, Auvray *et al.*, avaient purifié ^{mv4}Int en utilisant le système pET et la méthode de purification Ni-NTA par surexpression du plasmide pET-Int dans la souche *E. coli* BL21(DE3) (Auvray *et al.*, 1999a), pensant avoir ajouté une étiquette 6xHis à ^{mv4}Int . En réalité, seul le peptide « ORF-8 » est étiqueté et c'est l'affinité naturelle des recombinases à tyrosine pour les colonnes Ni-NTA qui avait permis sa purification de manière suboptimale. Ainsi, de manière à optimiser le protocole de production et purification, nous avons décidé de purifier l'intégrase sous deux formes : une forme native, en utilisant directement le plasmide pET-Int, et une forme étiquetée, obtenue par délétion de l'ORF-8 et fusion traductionnelle de l'étiquette 6xHis et de ^{mv4}Int (plasmide pMET337, Table 3). Des extraits cellulaires surexprimant ces 2 protéines ont été produits puis purifiés par passage successif sur deux colonnes d'affinité : une colonne Ni-NTA puis une colonne d'héparine (Fig. 50). Après passage sur ces colonnes, la fraction éluée n'est pas pure car elle contient en plus de la bande spécifique à ^{mv4}Int d'autres bandes protéiques (pistes 6, Fig. 50a - 50b). Seule une chromatographie supplémentaire sur colonne de filtration sur gel (SUPERDEX 200, 16/60) permet l'obtention de protéines homogènes en coloration bleu de Coomassie (pistes 7, Fig. 50a - 50b).

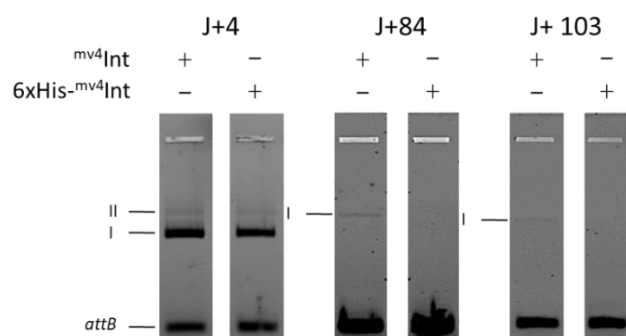


Fig. 51 Evolution de la stabilité des protéines purifiées $mv4Int$ et $6xHis-mv4Int$. La réaction de recombinaison a été effectuée avec 7,2 pmol de $mv4Int$ ou $6xHis-mv4Int$ et a été incubée 1h30 à 42°C. Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. Un deuxième produit de recombinaison (II) est observé sur le gel, avec une taille supérieure à celle du produit I. Comme $mv4Int$ est capable de recombinaison entre un site *attP* et un site *attL* (Coddeville and Ritzenthaler, 2010), il se pourrait que ce produit de recombinaison (II) soit dû à l'intégration d'un plasmide dans le site *attL*. Les temps de stockage des protéines sont indiqués au-dessus des gels.

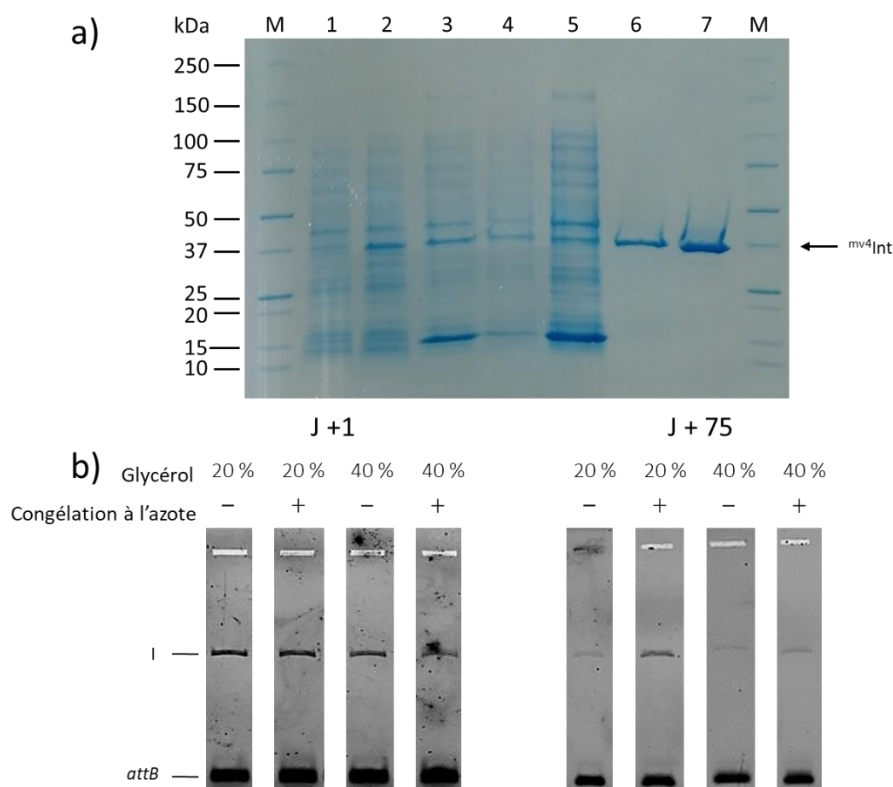


Fig. 52 Purification de $mv4Int$ et étude de la stabilité de la protéine purifiée. a) Electrophorèse SDS-PAGE 4-15% d'aliquotes d'extraits protéiques obtenus lors de la purification de $mv4Int$. Pistes : 1, extrait cellulaire *E. coli* BL21(DE3) ; 2, extrait cellulaire induit (IPTG) ; 3, fraction soluble ; 4, fraction insoluble ; 5, fractions non retenues sur colonne de Ni-NTA ; 6, fractions éluées en sortie de filtration sur gel ; 7, fractions éluées après colonne d'héparine ; M, marqueur de taille. b) Etude de différentes conditions sur la stabilité de $mv4Int$. La réaction de recombinaison a été effectuée avec 7,2 pmol de $mv4Int$ et a été incubée 1h30 à 42°C. Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. Les temps ainsi que les conditions de stockage des protéines sont indiqués au-dessus des gels.

III.2) Stabilité des protéines purifiées et optimisation des conditions de stockage de l'intégrase

Après purification, les protéines mv4 Int et 6xHis- mv4 Int ont été stockées à -80°C (cf. matériels et méthodes). Lors de tests préliminaires, il était suggéré que les protéines purifiées présentaient un problème de stabilité après congélation. Dans le but de contrôler cette stabilité, plusieurs temps de congélation ont été testés (Fig. 51). Ainsi, 4 jours après congélation, il est observé que la protéine étiquetée et la protéine non-étiquetée permettent l'obtention d'un signal de recombinaison intense. Ce résultat confirme que, contrairement aux intégrases hétérobivalentes classiques, l'utilisation de la protéine mv4 Int seule est suffisante pour effectuer une recombinaison *in vitro*, en accord avec les résultats publiés (Auvray *et al.*, 1999a). Cependant, après 84 jours de congélation, la protéine non-étiquetée présente un signal de recombinaison faible et la protéine étiquetée est inactive, cette activité diminuant encore après 100 jours de congélation. Ces résultats indiquent que les conditions de stockage des protéines purifiées ne sont pas optimales et que la protéine native semble plus stable que la protéine étiquetée 6xHis. Par conséquent, les différentes expériences réalisées lors de ce travail ont été effectuées avec la protéine mv4 Int non-étiquetée (forme native).

Dans le but d'améliorer les conditions de stockage de l'intégrase purifiée, deux conditions ont été testées : deux concentrations différentes en cryoprotecteur (glycérol 20% et 40%) avec (ou sans) congélation préalable dans l'azote liquide. Pour cela, une nouvelle purification de l'intégrase native a été effectuée en suivant le protocole optimisé : l'extrait cellulaire est d'abord purifié sur une colonne Ni-NTA, puis par une chromatographie par filtration sur gel (piste 6, Fig. 52a). L'utilisation de ces deux colonnes permet d'obtenir une protéine homogène après coloration au bleu de Coomassie. L'échantillon protéique est enfin concentré à l'aide d'une colonne d'héparine (piste 7, Fig. 52a). Après élution, différentes concentrations de glycérol ont été ajoutées aux échantillons. Les échantillons ont été aliquotés par fraction de 20 μL et congelé ou non dans l'azote liquide avant d'être stockés à -80°C . Deux temps de stockage ont été testés : 1 jour et 75 jours (Fig. 52b). Au premier jour de congélation, les différents échantillons présentent tous une bande de recombinaison intense. Au bout de 75 jours de congélation, les différents échantillons ont perdu en activité de recombinaison. Les conditions 20% glycérol sans congélation à l'azote liquide et 40% glycérol avec et sans congélation à l'azote liquide présentent tous une bande de recombinaison peu intense. Seule la condition 20% de glycérol avec congélation à l'azote montre une bande de recombinaison avec une intensité correcte.

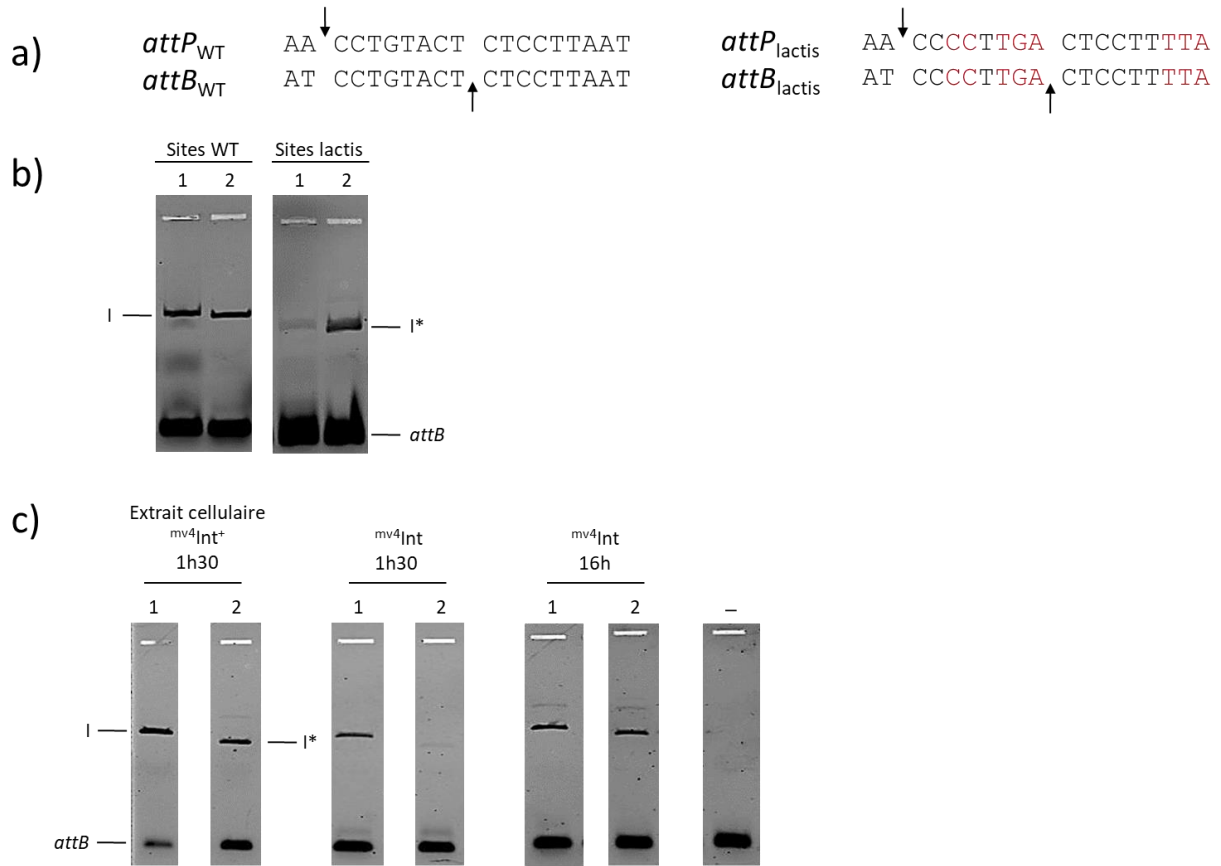


Fig. 53 Optimisation de la recombinaison avec des sites non naturels. a) Représentation de la région cœur du site *attP* et de la séquence *attB* minimum. A gauche : les sites *attP*_{WT} et *attB*_{WT} (Auvray *et al.*, 1999a ; Auvray *et al.*, 1999b). Les flèches indiquent les deux zones de coupures de brin encadrant la région centrale (Coddeville *et al.*, 2014). A droite : les sites modifiés (*attP*_{lactis} et *attB*_{lactis}). Les bases modifiées par rapport à la séquence sauvage sont indiquées en rouge. b) Optimisation du tampon de recombinaison lors d'une recombinaison entre les sites *attP/attB* natifs et les sites *attP/attB* modifiés. La réaction de recombinaison a été effectuée avec 40 µg d'extrait cellulaire ayant surexprimé *mv4Int* et a été incubée 1h30 à 42°C. Le fragment *attB* fluorescent et le produit de recombinaison entre les sites sauvages (I) et les sites modifiés (I*) sont indiqués sur le gel. Les deux bandes de recombinaison ne sont pas à la même taille car le plasmide comportant le site *attP*_{WT} est plus grand de 1 kb par rapport au plasmide possédant le site *attP*_{lactis}. Pistes : 1, tampon « Auvray99 » ; 2, tampon TEND. c) Effet du temps d'incubation sur la recombinaison. La réaction de recombinaison est effectuée dans du tampon TENDP (TEND + 10% PEG8000). Les différentes conditions sont indiquées au-dessus du gel. Pistes : 1, substrats WT ; 2, substrats lactis ; - : témoin sans enzyme. Le fragment *attB* fluorescent et le produit de recombinaison entre les sites sauvages (I) et les sites modifiés (I*) sont indiqués sur le gel. La réaction de recombinaison contient 7,2 pmol de *mv4Int* ou 40 µg d'extrait cellulaire d'*E. coli* surexprimant *mv4Int*.

Ainsi, toutes les conditions testées entraînent à un degrés variable une perte d'activité de recombinaison en fonction de la durée de stockage. La condition 20% de glycérol avec congélation à l'azote est la moins défavorable et est celle qui a été conservée pour les productions ultérieures de l'étude.

IV) Optimisation du système de recombinaison *in vitro* lors de l'utilisation de sites non naturels

Un des objectifs principaux de la thèse a été de tester l'étendue de la flexibilité de ^{mv4}Int par rapport à la modification de ces sites de recombinaison. Ainsi, pour tester la capacité de reprogrammation de la recombinaison vers un site *attB* non naturel, la région cœur du site *attP* natif (*attP*_{WT}) a été adaptée c'est-à-dire qu'elle a été modifiée afin de la rendre identique au site *attB* testé (voir chapitre 3). Dans ce chapitre, les sites non naturels utilisés présentent 8 nucléotides différents dans la région cœur par rapport aux sites *attP*_{WT} et *attB*_{WT} et ont été nommés *attP*_{lactis} et *attB*_{lactis} (Fig. 53a). Au départ, un extrait cellulaire ayant surexprimé ^{mv4}Int était utilisé pour la réaction de recombinaison. Lors de l'utilisation des substrats *attP* et *attB* « lactis », des résultats préliminaires ont indiqué une différence dans l'activité de recombinaison en fonction du tampon utilisé. Ainsi, 17 tampons réactionnels ont été testés dont le tampon utilisé en 1999 (Auvray *et al.*, 1999a) (Table 5). Un nouveau tampon a été sélectionné, le tampon TEND, permettant une recombinaison aussi efficace que le tampon « Auvray99 » pour une recombinaison entre sites sauvages ainsi qu'une réaction 6 fois plus efficace lors de la recombinaison entre sites modifiés (Fig. 53b). Le tampon TEND a donc été conservé pour les différents tests utilisant un extrait cellulaire ayant surexprimée ^{mv4}Int. Une différence majeure entre le tampon TEND et le tampon « Auvray99 » est l'absence de spermidine. Contrairement aux systèmes λ , HP1 et L5 où la réaction de recombinaison est fortement stimulée par l'ajout de spermidine (Goodman and Scocca, 1989; Lee and Hatfull, 1993; Nash, 1975), l'ajout de la polyamine n'a aucun effet sur la recombinaison entre sites sauvages et semble inhiber la recombinaison entre sites « lactis ». Pour la recombinase simple Cre, il a été montré que l'ajout de spermidine favorise la liaison coopérative des dimères de Cre avec les sites *loxP* (Rüfer *et al.*, 2002). Récemment, la structure du complexe intégratif de λ Int a été obtenue et montre des contacts ADN-ADN étroits entre les bras P et P' du site *attP* expliquant le rôle important de la spermidine dans la recombinaison intégrative (Laxmikanthan *et al.*, 2016).

Le tampon TEND présente une concentration plus élevée en sel (150 mM NaCl) par comparaison aux tampons « Auvray99 » (5 mM NaCl, 90 mM KCl) (Auvray *et al.*, 1999a), ^{HP1}Int (50-70 mM NaCl) (Goodman and Scocca, 1989), ^{L5}Int (25 mM NaCl) (Lee and Hatfull, 1993) et ^λInt (70 mM KCl) (Nash and Robertson, 1981). Dans le cas de ^λInt, la nécessité de la polyamine peut être abolie en présence de MgCl₂ ou lorsque la concentration en KCl est augmentée à 100 mM bien que l'ajout de spermidine stimule toujours la recombinaison d'un facteur 3 (Nash and Robertson, 1981). Ainsi, il est possible que la concentration en NaCl du tampon TEND explique l'indépendance de la recombinaison entre sites sauvages par rapport à la spermidine. En revanche, comme le tampon « Auvray99 », l'ajout de 4 mM de spermidine dans le tampon TEND inhibe la recombinaison entre sites « lactis » (données non montrées), suggérant que la polyamine induit une conformation de l'ADN incompatible avec le complexe de recombinaison formé sur les substrats modifiés.

Après purification, ^{mv4}Int a été utilisée dans le but de valider que l'intégrase était capable de réaliser une recombinaison sur des sites non naturels. Le passage d'un extrait brut (environnement complexe) à une protéine purifiée induit une perte importante de l'activité de recombinaison sur les couples « lactis » dans le tampon TEND (données non montrées). Par conséquent, le tampon de recombinaison a de nouveau dû être optimisé pour l'utilisation d'une intégrase purifiée. Dans la cellule, les réactions sont réalisées dans un milieu contenant différentes concentrations de macromolécules (estimée à 300-400 g/L chez *E. coli*) (Zimmerman and Trach, 1991) et ces molécules occupent une partie importante du volume cellulaire (environ 20 à 30%) : c'est ce que l'on appelle le « molecular crowding » (Ellis, 2001). Le « molecular crowding » est un effet physique, non spécifique, décrivant le fait que la concentration intracellulaire de macromolécules est si élevée qu'une proportion du volume est physiquement occupée et donc indisponible pour d'autres molécules (« volume exclu ») (Ellis, 2001). Le « molecular crowding » a donc un effet sur les macromolécules et sur les complexes protéines-ADN (Minton, 2006). De manière à mimer un environnement cellulaire *in vitro*, des « crowding agents » comme le polyéthylène glycol (PEG), le dextran, le Ficoll ou bien des protéines comme l'albumine de sérum bovin sont généralement ajoutés dans le tampon dont le PEG est le « crowding agent » le plus utilisé (Akabayov *et al.*, 2013). Ainsi, nous avons ajouté du PEG8000 (10%) au tampon de recombinaison TEND afin d'obtenir le tampon TENDP qui a été conservé pour les différentes recombinaison utilisant ^{mv4}Int purifiée.

Bien que le tampon ait été optimisé par l'ajout de PEG, l'activité de recombinaison observée pour le couple « lactis » est toujours très faible en 1h30 de réaction avec une protéine purifiée (Fig. 53c), indiquant qu'en plus du « molecular crowding », un facteur supplémentaire présent dans l'extrait cellulaire d'*E. coli* améliore la réaction. En revanche, des signaux équivalents de recombinaison sont obtenus pour les sites natifs et modifiés lorsque le temps d'incubation est augmenté à 16h (« overnight ») (Fig. 53c). Ce résultat indique donc que la protéine ^{mv4}Int est responsable de la réaction de recombinaison sur des sites modifiés, mais qu'un facteur présent dans l'extrait cellulaire de la bactérie *E. coli* stimule la cinétique de cette recombinaison. Ce facteur a été identifié et sera décrit dans le chapitre 3.

En résumé, ^{mv4}Int est bien une intégrase hétérobivalente membre de la famille des recombinases à tyrosine. Sa taille réelle est de 369 AA et cette protéine est responsable de la recombinaison intégrative du bactériophage mv4. Nous avons pu vérifier la particularité de cette enzyme : elle est capable de réaliser la formation d'un intasome et d'effectuer une réaction de recombinaison en absence de facteur d'hôte, une caractéristique rare parmi cette famille de protéine. Nous avons aussi réussi à mettre au point et optimiser les différents protocoles nécessaires à l'étude de cette enzyme : production et purification de l'enzyme, mise en place du système de recombinaison *in vitro* et optimisation des conditions en fonction de la nature de la protéine utilisée (extrait cellulaire ou protéine purifiée). Comme nous contrôlons les différents paramètres de notre système d'étude, nous avons pu commencer à étudier le système de recombinaison particulier du bactériophage mv4, notamment en s'intéressant à la région cœur atypique de ces sites de recombinaison.

Chapitre 2 : Caractérisation des sites de recombinaison du système mv4

Grâce aux différentes méthodologies développées dans le chapitre précédent, nous nous sommes intéressés aux sites de recombinaison du système mv4. Ces sites de recombinaison possèdent plusieurs particularités par rapport aux sites classiques des intégrases hétérobivalentes. En effet, les sites *attB* et *attP* possèdent une architecture atypique, avec des régions cœur dépourvues de site de fixation de mv^4 Int (les sites cœur B, B' sur *attB* et C, C' sur *attP*), une taille inhabituelle de la région centrale de 8 pb (Fig. 41 ; Fig. 55a), une organisation particulière des sites bras de fixation de mv^4 Int sur le site *attP*, avec notamment un site bras (site P2) localisé immédiatement à gauche de la région cœur (Fig. 41 ; Fig. 55a) et un site *attB* asymétrique d'une taille remarquable puisqu'il est le site *attB* le plus court décrit dans la littérature (Auvray *et al.*, 1999b). Le système de recombinaison *in vitro* optimisé lors de cette thèse a été utilisé d'une part pour caractériser les contraintes topologiques liées à la réaction de recombinaison, et d'autre part pour invalider la taille publiée du site *attB*. Une stratégie basée sur l'utilisation de banques aléatoires a ensuite été mise en place pour caractériser plus précisément la région cœur des sites *attB* et *attP* et la nouvelle organisation des régions cœur et des sites bras de fixation de mv^4 Int a été validée par des expériences de retard sur gel. Enfin, la nature et la fréquence des motifs des sites B, B', C et C' permissifs à la recombinaison *in vitro* ont été identifiés par séquençage NGS des produits de recombinaison obtenus après utilisation des banques *attB* Lib6 (séquences B), *attB* Lib8 (séquences B'), *attP* Lib3 (séquences C) et *attP* Lib1 (séquences C') (Fig. 55a).

La topologie de l'ADN joue un rôle important dans de nombreux systèmes de recombinaison (Kanaar and Cozzarelli, 1992), que ce soit chez les résolvases (Krasnow and Cozzarelli, 1983; Reed, 1981b) et les invertases (Johnson and Simon, 1985; Plasterk *et al.*, 1984) pour les recombinases à sérine, ou bien chez le système XerC/D lors de recombinaison sur les sites plasmidiques *cer* et *psi* (Colloms *et al.*, 1996) et les systèmes hétérobivalents pour les recombinases à tyrosine (Goodman and Scocca, 1989; Mizuuchi and Nash, 1976; Panis *et al.*, 2007; Yu and Haggård-Ljungquist, 1993). Chez le modèle λ Int, il a été montré que la réaction d'intégration était dépendante du surenroulement négatif du site *attP* et que cette topologie de l'ADN favorisait la formation de l'intasome sur ce site (Richet *et al.*, 1986).

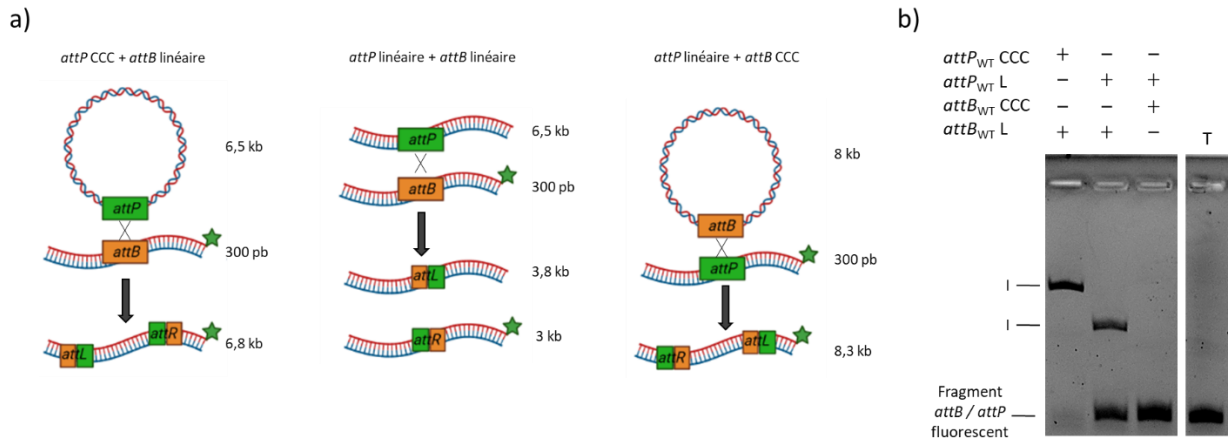


Fig. 54 Effet de la topologie des substrats sur la réaction d'intégration du système mv4. a) Représentation des 3 conditions topologiques testées. b) Résultat de la recombinaison *in vitro* en fonction de la topologie des substrats *attB* et *attP*. La réaction de recombinaison contient 7,2 pmol de *mv4*Int, 40 µg d'extrait cellulaire d'*E. coli* chauffé à 95°C et a été incubée avec les substrats *attP*_{WT} et *attB*_{WT} 1h30 à 42°C. La topologie des différents substrats est indiquée au-dessus des pistes. CCC, forme surenroulée ; L, forme linéaire. Piste T : témoin sans *mv4*Int. Le fragment *attB* ou *attP* fluorescent ainsi que le produit de recombinaison (I) sont indiqués sur le gel. Les expériences ont été réalisées 2 fois.

Le développement du système de recombinaison *in vitro* nous a permis d'étudier l'effet de la topologie des substrats *attB* et *attP* sur la réaction de recombinaison du système mv4. Ainsi, trois conditions ont été testées i) en présence d'un site *attB* linéaire et d'un site *attP* surenroulé, condition optimale pour les intégrases hétérobivalentes, ii) en présence d'un site *attB* surenroulé et d'un site *attP* linéaire et iii) en présence des deux sites linéaires (Fig. 54a). Comme attendu, la réaction de recombinaison la plus efficace est celle contenant un site *attP* surenroulé et un site *attB* linéaire (Fig. 54b). En revanche, contrairement au modèle λ Int (Mizuuchi *et al.*, 1978) et à HP1 Int (Goodman and Scocca, 1989), mv4 Int, comme L5 Int (Peña *et al.*, 1998), est capable de réaliser une recombinaison intégrative entre deux substrats *attB* et *attP* linéaires (Fig. 54b). Pour la recombinase λ Int, il a été montré que la nécessité de surenroulement du site *attP* pouvait être supprimée si la concentration en KCl était abaissée de 70 mM à 25 mM (Pollock and Abremski, 1979). Dans nos conditions, l'absence de surenroulement du site *attP* ne diminue que d'un facteur 2 l'activité de recombinaison, ce qui est en contradiction avec les résultats publiés où aucun produit de recombinaison n'était observé en absence de surenroulement du site *attP* (Auvray *et al.*, 1999a). Les conditions de recombinaison *in vitro* sont différentes entre l'étude de 1999 et notre étude au niveau de la composition du tampon réactionnel, notamment par l'absence de spermidine employée dans le tampon de 1999. Cette molécule est une polyamine capable d'interagir avec l'ADN ce qui peut induire des modifications importantes de la conformation de l'ADN (Allison *et al.*, 1981; Chatteraj *et al.*, 1978). Il est donc possible que la spermidine induise une conformation des ADN linéaires incompatibles avec une réaction de recombinaison. Notre résultat suggère que la nécessité du surenroulement du site *attP* est moins importante pour la mise en place de l'intasome par rapport aux systèmes classiques. Chez λ Int, c'est la combinaison du surenroulement du site *attP* et la présence du facteur d'hôte IHF qui permet la formation d'un intasome fonctionnel (Richet *et al.*, 1986). Le système mv4 ne nécessitant pas de facteur d'hôte (Auvray *et al.*, 1999a), ceci serait cohérent avec une réaction où la topologie de l'ADN *attP* est moins importante pour la mise en place de l'intasome. Dans le système L5, la réaction de recombinaison nécessite le facteur d'hôte mIHF (Pedulla *et al.*, 1996), mais peut être réalisée avec deux sites linéaires, bien que le surenroulement du site *attP* stimule la réaction (Peña *et al.*, 1998). De plus, le surenroulement du site *attB* peut remplacer la nécessité de surenroulement du site *attP* et il a donc été proposé que le surenroulement avait un rôle à une étape plus tardive, après la formation du complexe synaptique (Peña *et al.*, 1998). Contrairement au système L5, le surenroulement du site *attB* exerce un effet inhibiteur sur la recombinaison (Fig. 54b).

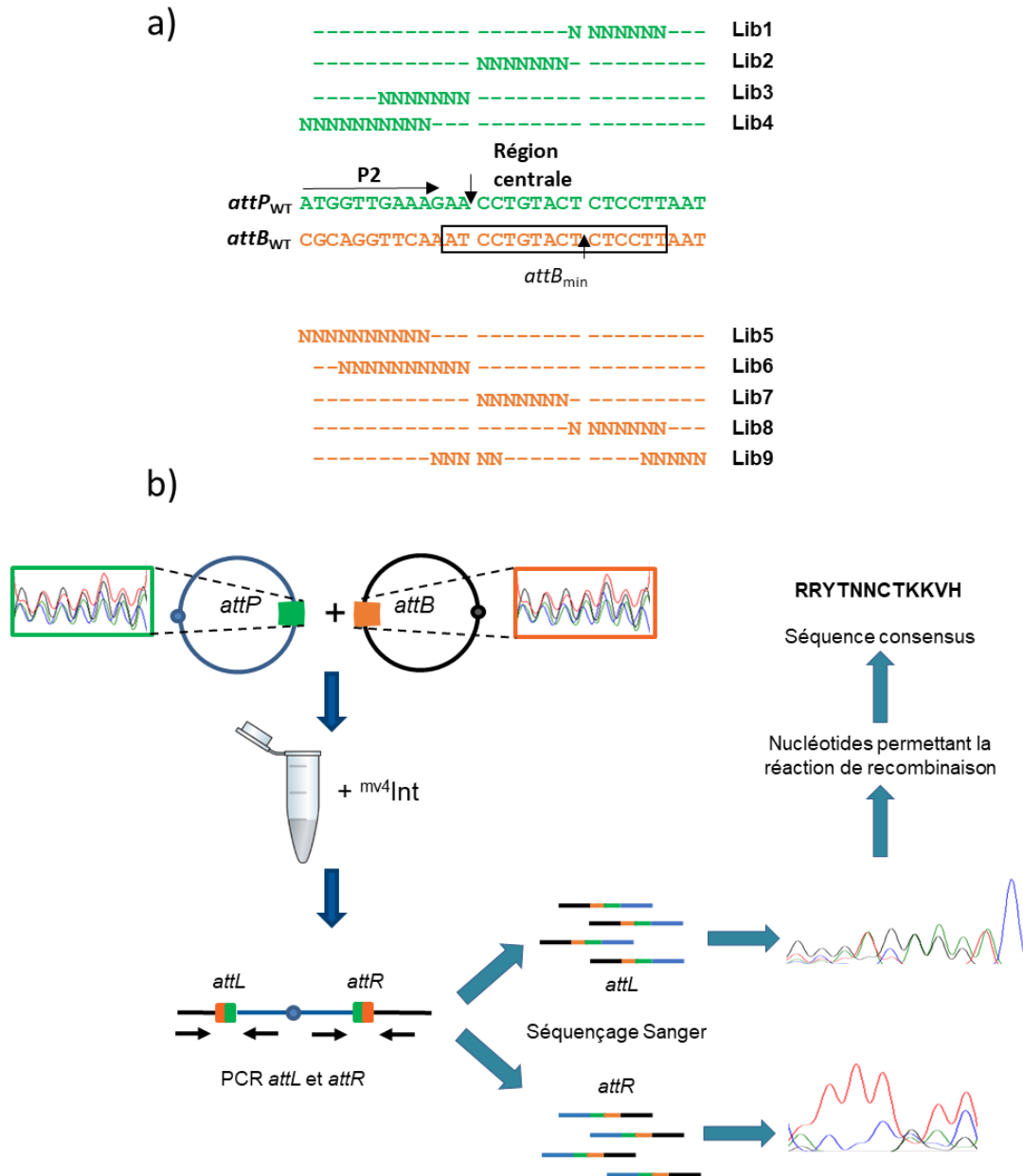


Fig. 55 Principe de l'utilisation de banques aléatoires pour la caractérisation des régions cœur. a) Localisation des banques dans les régions cœur. La région *attB* minimum déterminée en 1999 (Auvray *et al.*, 1999b) est encadrée en noire. Le site bras atypique P2 présent sur le site *attP* est représenté par une flèche horizontale noire. La région centrale est encadrée par les deux sites de coupures (flèches noires verticales) déterminés en 2014 (Coddeville *et al.*, 2014). Les différents motifs aléatoires des banques sont représentés par les lettres « N » et les nucléotides identiques aux séquences natives sont symbolisés par des tirets. b) Stratégie globale de l'utilisation des banques aléatoires. Les chromatogrammes indiquent les différents nucléotides présents à chaque position. Une couleur est associée à chaque nucléotide : bleu, cytosine ; noir, guanine ; rouge, thymine et vert, adénine.

Afin de vérifier si la réaction de recombinaison sur un site *attB* surenroulé était tout de même possible, la réaction de recombinaison a été réalisée entre deux plasmides *in vitro* (cf. partie I.2) et *in vivo* (données non montrées). Dans les deux cas, ^{mv4}Int est capable de réaliser une réaction de recombinaison suggérant que le surenroulement du site *attB* inhibe uniquement la recombinaison lors de l'utilisation d'un site *attP* linéaire. Nous n'avons pas d'hypothèse pour expliquer cette inhibition de la recombinaison.

I) Stratégie expérimentale de l'étude des sites de recombinaison et caractérisation des bornes du site *attB*

I.1) Utilisation de banques aléatoires des sites *attB* et *attP*

En raison de l'originalité de la région cœur des sites *attB* et *attP*, ces régions ont été réanalysées par une approche basée sur l'utilisation de banques d'ADN correspondant soit au site *attB*, soit à la région cœur du site *attP* et contenant chacune des bases aléatoires sur 7 à 10 positions (Fig. 55a). Le principe expérimental est le suivant : des oligonucléotides contenant sur certaines positions les 4 nucléotides possibles ont été synthétisés, amplifiés par PCR et clonés chez *E. coli* dans différents plasmides (Fig. 55b). Chaque banque plasmidique est récupérée, vérifiée par séquençage Sanger, puis utilisée pour différentes expériences de recombinaison *in vitro* avec, soit le site partenaire natif (*attP_{WT}* ou *attB_{WT}*), soit la banque partenaire associée (même région aléatoire entre les 2 sites) (Fig. 55b). Le choix d'utiliser différents plasmides était au départ lié au développement d'une plateforme de recombinaison *in vivo* chez *E. coli* basée sur la recombinaison entre 2 plasmides, un plasmide réplcatif présent à 1 copie par cellule (pCC1Fos) et contenant le site *attB_{WT}* ou ses variants, et un plasmide suicide contenant le site *attP* (plasmide pMET359, Table 3) dépourvu de sa protéine de réplication RepA_{pWV01} et ne se répliquant que dans une souche d'*E. coli* exprimant cette protéine (Table 4). J'ai mis en place cette plateforme et montré qu'elle était fonctionnelle sur un couple de variants *attP/B*, mais comme elle n'a pas été utilisée pour l'étude des sites de recombinaison, elle ne sera pas développée dans ce manuscrit. Après recombinaison, les sites *attL* et *attR* sont amplifiés par PCR et séquencés (Fig. 55b). Parmi les positions contenant les 4 bases, seuls les nucléotides permissifs à la recombinaison seront retrouvés dans ces sites hybrides, ce qui permettra de déterminer les contraintes qui s'exercent sur la nature du nucléotide à chaque position testée.

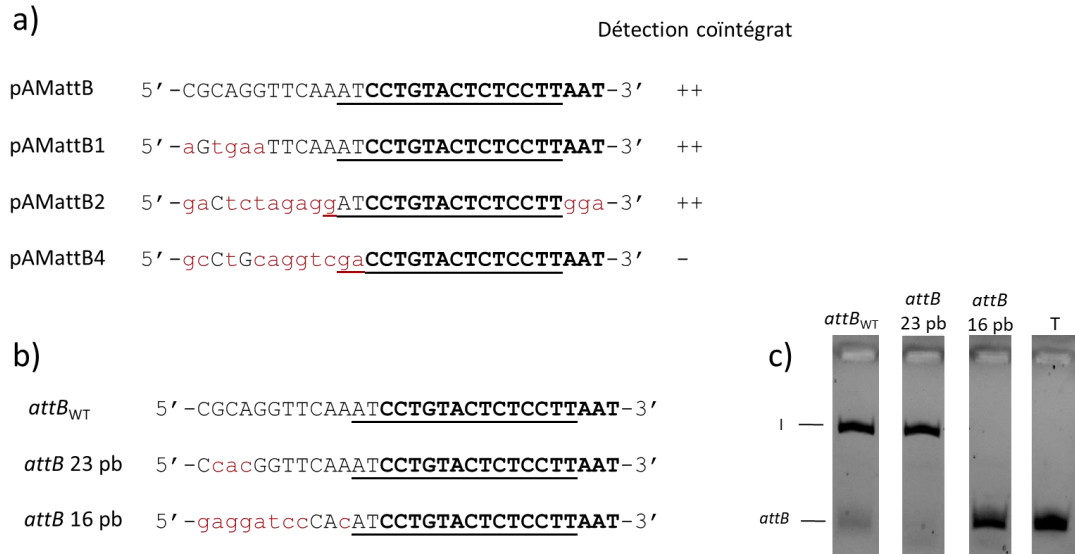


Fig. 56 Effet de la taille du site *attB*_{WT} sur la réaction de recombinaison. a) Caractérisation du site *attB* minimum en 1999 (Auvray *et al.*, 1999b). La séquence *attB* a été successivement tronquée à gauche et à droite. Les bases indiquées en rouge sont les bases qui diffèrent par rapport à la séquence *attB*_{WT}. Les bases soulignées sont les bases du site *attB* minimum et les bases en gras sont les bases de la région cœur (région d'identité entre *attB* et *attP*). b) Séquences utilisées lors de notre étude. c) Effet de la taille du site *attB* sur la réaction de recombinaison. La réaction de recombinaison contient 7,2 pmol de *mv*⁴Int, 40 µg d'extrait cellulaire d'*E. coli* chauffé à 95°C et a été incubée 1h30 à 42°C. La taille des différents sites *attB* est indiquée au-dessus de chaque piste. Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. Piste T : témoin sans *mv*⁴Int.

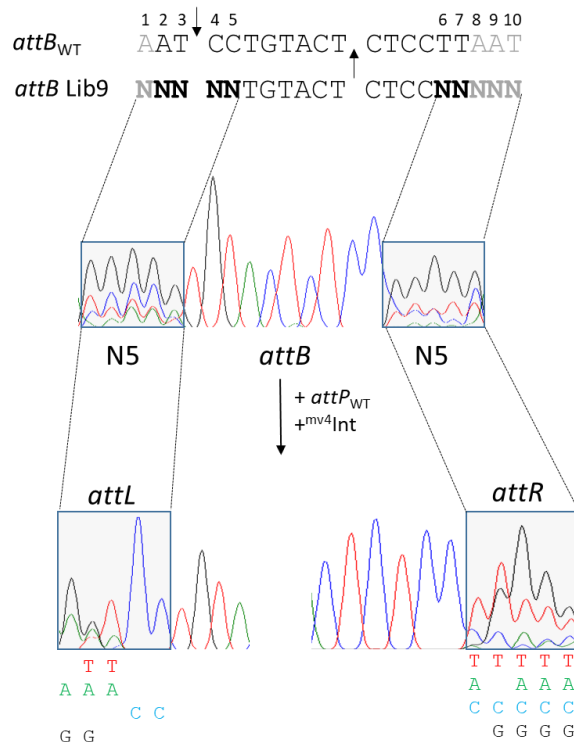


Fig. 57 Principe de la méthodologie d'étude des banques aléatoires : exemple de la banque *attB* Lib9. Résultat du séquençage Sanger des sites *attB*, *attL* et *attR* après recombinaison *attB* Lib9 x *attP*_{WT}. Les bases définies hors du site *attB* minimum (Auvray *et al.*, 1999b) sont représentées en gris et les flèches noires indiquent les deux sites de coupe encadrant la région centrale (Coddeville *et al.*, 2014). Les lettres N indiquent les positions aléatoires de la banque. Une couleur est associée à chaque nucléotide sur le chromatogramme : bleu, cytosine ; noir, guanine ; rouge, thymine et vert, adénine.

I.2) Caractérisation des limites du site *attB* par l'utilisation de banques aléatoires

Dans la littérature, le système mv4 possède le site *attB* le plus court. Ce site *attB* minimum a été déterminé en 1999 par un test de recombinaison *in vivo* entre plasmides répliatifs chez *E. coli* en utilisant une série de sites *attB* tronqués, diminuant petit à petit la taille de ce site (Fig. 56a) (Auvray *et al.*, 1999b). Ainsi, la détection de coïntégrats entre les deux plasmides leur a permis de déterminer un site *attB* minimum de 16 pb décalé de 2 pb par rapport à la région cœur de 17pb identifiée (Fig. 56a). Dans le but de valider la taille particulière de ce site, différentes tailles du fragment *attB*_{WT} ont été amplifiées par PCR (Fig. 56b) pour obtenir des fragments *attB* fluorescents qui ont été testés par recombinaison *in vitro* contre le site *attP*_{WT} (Fig. 56c). Un produit de recombinaison est obtenu avec la région *attB*_{WT} et avec le site *attB* de 23 pb. En revanche, aucun produit de recombinaison n'est obtenu avec le site de 16 pb, indiquant que le site *attB* minimum publié ne semble pas fonctionnel en condition de recombinaison *in vitro* et qu'il nécessite une région plus grande à gauche.

Dans le but de caractériser précisément les bordures du site *attB* minimum, une banque composée de 5 nucléotides aléatoires chevauchant chaque extrémité du site *attB* publié (banque Lib9, Fig. 55a ; Fig. 57) a tout d'abord été construite et testée par recombinaison *in vitro* contre le site *attP* natif (*attP*_{WT}). Après recombinaison, les sites *attL* et *attR* ont été amplifiés par PCR et séquencés (Fig. 57). La banque *attB* initiale contient bien les 4 nucléotides possibles à chaque position, mais un biais important est observé dans la fréquence des bases, avec une sur-représentation en G (ligne noire, Fig. 57), une sous-représentation en A (ligne verte, Fig. 57) et une distribution hétérogène en C (ligne bleue, Fig. 57) et en T (ligne rouge, Fig. 57). Ce biais a été associé à la méthode de synthèse des oligonucléotides et a été corrigé (changement de fournisseur) pour les futures banques décrites. Pour le site *attL*, de fortes contraintes sont observées après recombinaison puisqu'une seule base (cytosine) sur les 4 est autorisée en positions 4 et 5. Ces positions se trouvent dans la région centrale déterminée précédemment par coupure sur substrat suicide (Coddeville *et al.*, 2014) et où les cytosines sont les bases présentes dans les sites *attB* et *attP* natifs. Ce résultat montre que les deux premiers nucléotides de la région centrale doivent être identiques entre les sites *attB* et *attP*, comme observé pour la plupart des systèmes de recombinaison à tyrosine.

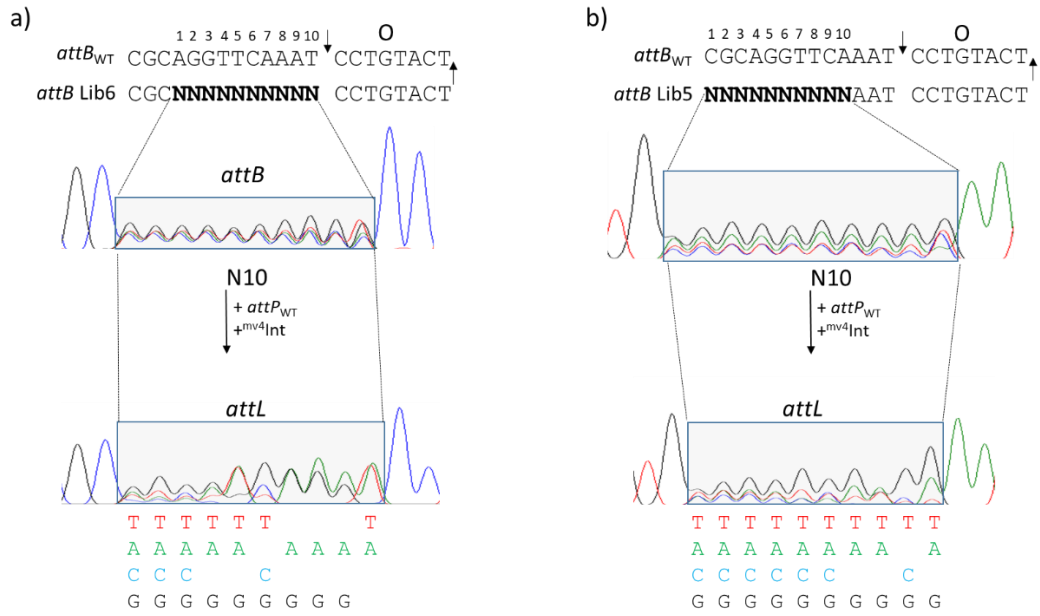


Fig. 58 Détermination de la bordure gauche du site *attB* par recombinaison *in vitro* des banques aléatoires. Le site *attB*_{WT} est représenté, la lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupe encadrant la région centrale (Coddeville *et al.*, 2014). **a)** Résultat de la recombinaison *attB* Lib6 x *attP*_{WT}. **b)** Résultat de la recombinaison *attB* Lib5 x *attP*_{WT}.

Au niveau des positions 2 et 3 (Fig. 57), positions déterminant le début du site *attB* minimum publié (Auvray *et al.*, 1999b), des contraintes plus ou moins importantes sont retrouvées. En effet, au niveau de la position 2, les bases G, A et T sont autorisées, alors qu'à la position 3, seules les nucléotides A et T semblent permissifs à la recombinaison. Enfin, au niveau de la première position, position décrite comme étant hors du site *attB* minimum, une forte contrainte est observée puisque seule une purine est retrouvée. Ces observations sont en accord avec nos résultats de recombinaison *in vitro* montrant que le site *attB* de 16 pb n'est pas fonctionnel et suggèrent un site *attB* plus grand sur l'extrémité gauche. Afin de déterminer précisément la limite gauche du site *attB*, deux autres banques contenant 10 positions aléatoires (Lib5 et Lib6, Fig. 55a) ont été construites et testées par recombinaison *in vitro* contre le site *attP_{WT}* (Fig. 58). Le séquençage du site *attL* issu de la recombinaison *attB* Lib6 x *attP_{WT}* (Fig. 58a) montre que les trois premières positions ne semblent pas présenter de contraintes particulières, suggérant qu'elles sont situées en-dehors du site *attB*. Cette limite gauche a également été validée par la banque *attB* Lib5 qui a confirmé l'absence de contraintes à ces mêmes positions (Fig. 58b). Il semblerait donc que le site *attB* possède 5 bases supplémentaires par rapport à la séquence publiée et corresponde à un site de 21 pb commençant par la séquence « 5'-TTCAAAT-3' » (Fig. 58).

Au niveau de l'extrémité droite du site *attB*, le site *attR* présente des contraintes moins importantes (Fig. 57). En effet, au niveau des positions 8-9-10, les 4 nucléotides sont présents avec un biais proche de celui de la banque *attB* de départ, démontrant que ces bases n'ont pas de rôle dans la recombinaison, bien qu'elles appartiennent à la région cœur (région d'identité entre les sites *attB* et *attP*). Ainsi, ces 3 positions se situent en dehors du site *attB* minimum, comme décrit précédemment (Auvray *et al.*, 1999b). En revanche, au niveau des positions 6 et 7, positions définissant l'extrémité droite du site *attB* minimum, une contrainte est observée à chaque position : absence de G à la position 6 et de A à la position 7, suggérant que le site *attB* se termine par la séquence « 5'-CTCCTT-3' », en accord avec les résultats publiés (Auvray *et al.*, 1999b). En résumé, la nouvelle taille du site *attB_{mv4}* revient à une taille plus classique étant donné que la plupart des sites *attB* des systèmes hétérobivalents caractérisés ont une taille comprise entre 18 pb et 29 pb.

Pour résumé, l'utilisation des banques aléatoires semble une stratégie pertinente pour la caractérisation précise des contraintes s'exerçant en général entre deux sites de recombinaison site-spécifique, notamment sur la nature des nucléotides permissifs à la réaction de recombinaison.

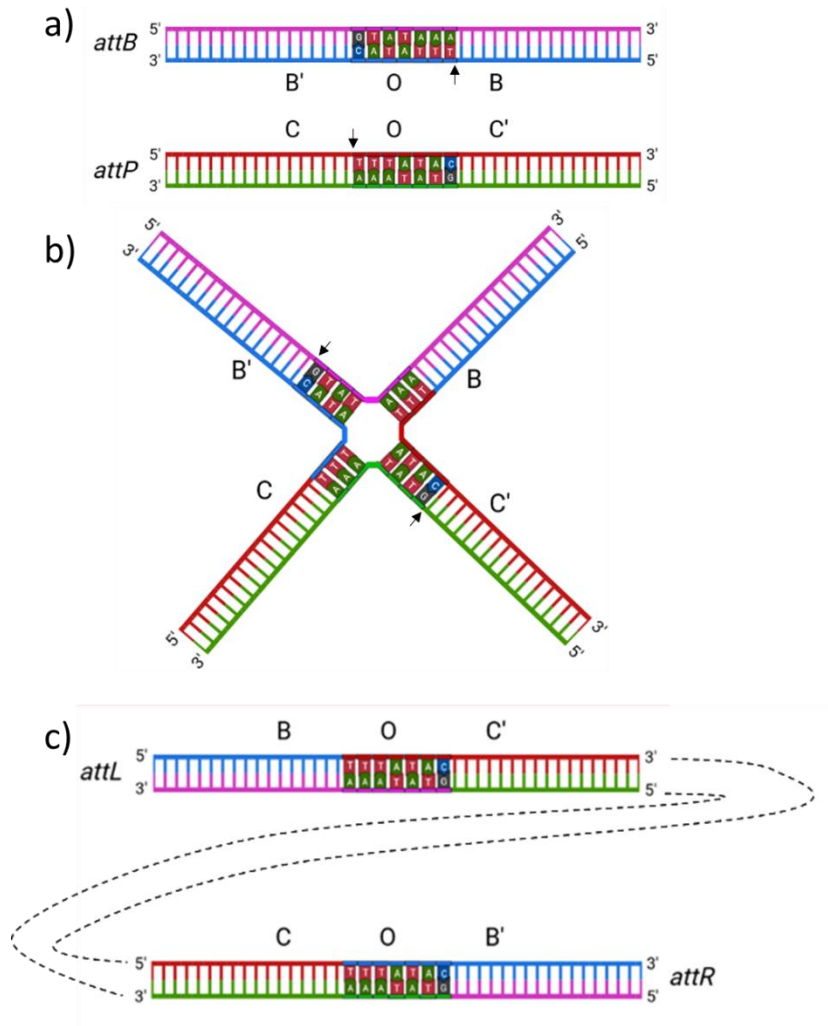


Fig. 59 Représentation de la formation et de la résolution de la jonction de Holliday par λ Int. a) Représentation de la région cœur des sites *attB* et *attP* de λ Int. Les nucléotides de la région centrale sont représentés par les couleurs rouge (T), vert (A), bleu (C), noir (G). Le brin supérieur du site *attB* est représenté en bleu et celui du site *attP* en rouge. Le brin inférieur du site *attB* est représenté en rose, et celui du site *attP* en vert. Les sites *attB* et *attP* sont représentés en orientation antiparallèle. b) Représentation de la formation de la jonction de Holliday par un échange des brins supérieurs. Les 3 premières bases des régions centrales des brins supérieurs (« 5'-TTT-3' ») sont échangées permettant la formation de la jonction de Holliday. L'isomérisation et le deuxième échange de brin n'est pas représenté. c) Résolution de la jonction de Holliday par un échange des brins inférieurs permettant la formation des sites hybrides *attL* et *attR*.

En effet, cette stratégie permet de déterminer i) les positions du site ne présentant aucune contrainte (4 bases possibles à ces positions) et donc pouvant être considérées comme hors du site de recombinaison, ii) les positions présentant des contraintes faibles (3 bases sur 4), modérées (2 bases sur 4) ou fortes (1 seule base possible), et finalement iii) de définir un motif nucléotidique dégénéré permettant la recombinaison. En profitant de la puissance de criblage de cette stratégie expérimentale, nous nous sommes attachés à redéfinir de façon exhaustive la région cœur des sites *attB*_{mv4} et *attP*_{mv4}. Au total, 9 banques ont été conçues : 5 autour du site *attB* et 4 dans la région cœur du site *attP* (Fig. 55a).

II) Etude de la région centrale (région « Overlap »), région d'échange des brins

II.1) Redéfinition de la taille de la région « Overlap »

En 2014, une région centrale de 8 pb avait été déterminée par réaction de coupure sur substrat suicide (Coddeville *et al.*, 2014). La taille de cette région semble atypique, puisque tous les sites de recombinaison des intégrases hétérobivalentes actuellement caractérisés possèdent plutôt une région centrale de seulement 7 pb, comme décrit pour les phages λ (Craig and Nash, 1983), HK022 (Kolot and Yagil, 1994), HP1 (Hauser and Scocca, 1992), L5 (Peña *et al.*, 1996), P22 (Smith-Mungo *et al.*, 1994), ainsi que l'ICE CtnDOT (Malanowska *et al.*, 2006). La connaissance du mécanisme d'échange des brins par les intégrases hétérobivalentes permet de caractériser cette région centrale de manière indirecte sans faire appel à une expérience de coupure sur substrats suicide. En effet, chez les recombinases à tyrosine dont le modèle λ Int, la réaction de recombinaison implique la formation (Fig. 59a ; Fig. 59b) et la résolution d'une jonction de Holliday (Fig. 59b ; Fig. 59c). Ainsi, lors de la réaction, seules les nucléotides présents dans la région d'échange de brins « Overlap » sont échangés.

Nous avons donc profité de cette propriété pour réanalyser la région centrale du système *mv4*Int/*attP*/*attB*. Deux banques comportant un motif de 7 pb aléatoire, dont la première position correspond au nucléotide en 3' de la région centrale publiée ont été construites, les banques *attB* Lib8 et *attP* Lib1 (Fig. 55a). En s'intéressant au devenir de la première position de ces banques, il est théoriquement possible de déterminer si la taille de la région centrale est de 7 ou 8 pb (Fig. 60).

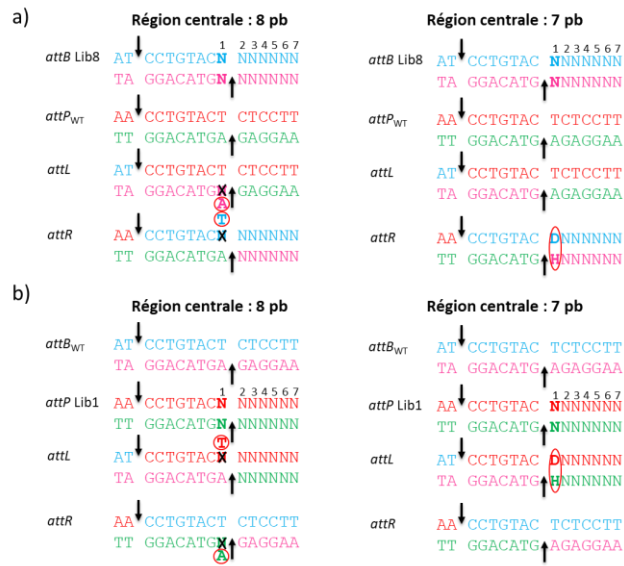


Fig. 60 Représentation du résultat de recombinaison en fonction de la taille de la région centrale. a) Représentation des deux résultats de la recombinaison *attB* Lib8 x *attP*_{WT} en fonction de la taille de la région centrale. La position aléatoire n°1 de la région centrale est représentée en gras et est entourée en rouge dans les sites *attL* et *attR*. La couleur des nucléotides correspond aux couleurs respectives des brins de la figure précédente. Les flèches noires verticales indiquent les deux sites de coupure permettant l'échange des brins. b) Représentation des deux résultats possibles de la recombinaison *attP* Lib1 x *attB*_{WT} en fonction de la taille de la région centrale. La légende est identique à la partie a). Dans les deux cas, si la position n°1 est incluse dans la région d'échange de brin (région de 8 pb), seul le nucléotide complémentaire au site *attP*_{WT} (a) ou *attB*_{WT} (b) sera observé (soit un T sur le brin supérieur et un A sur le brin inférieur). Si cette position est exclue (région centrale de 7 pb), les différents nucléotides permissifs pourront être observés sur un des sites hybrides.

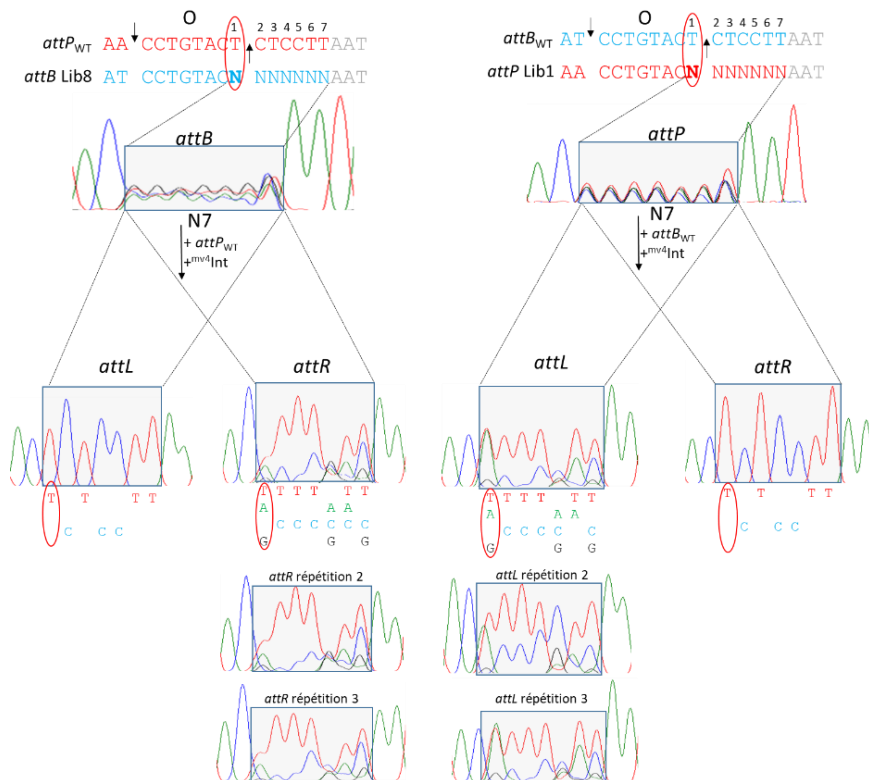


Fig. 61 Détermination de la taille de la région centrale par recombinaison *attP*_{WT} x *attB* Lib8 et *attP* Lib1 x *attB*_{WT}. Les sites *attB*_{WT} et *attP*_{WT} sont représentés, la lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupure encadrant la région centrale. Les répétitions de 3 expériences indépendantes sont montrées.

En effet, si la taille de la région centrale est de 8 pb, comme la recombinaison médiée par les intégrases ne supporte pas les hétéroduplex, on ne pourra donc retrouver sur les deux sites hybrides *attL* et *attR* que la base complémentaire à celle du site natif (*attP*_{WT} pour la recombinaison *attP*_{WT} x *attB* Lib8, et *attB*_{WT} pour la recombinaison *attB*_{WT} x *attP* Lib1), soit un T. En revanche, si la région centrale est de 7 pb, dans la recombinaison *attP*_{WT} x *attB* Lib8 la séquence sauvage sera retrouvée sur le site *attL* et les nucléotides permissifs de la banque *attB* Lib8 seront retrouvés sur le site *attR*, soit un D (A/G/T) puisqu'il existe une contrainte à cette position que l'on analysera dans la partie suivante. Dans la recombinaison *attB*_{WT} x *attP* Lib1, le raisonnement est similaire, la base identique à celle du site *attB* sera retrouvée sur le site *attR* et les nucléotides permissifs de la banque (D) seront retrouvés sur le site *attL* uniquement (Fig. 60).

Les banques *attB* Lib8 et *attP* Lib1 (Fig. 55a) ont été testées par recombinaison *in vitro* contre le site *attP*_{WT} ou *attB*_{WT}, respectivement (Fig. 61). Pour déterminer la taille de la région d'échange de brins, seule la première position de ces banques aléatoires est analysée. Les résultats observés après recombinaison *attP*_{WT} x *attB* Lib8 ou *attP* Lib1 x *attB*_{WT}, c'est-à-dire des sites *attL* et *attR* différents avec un des sites présentant les nucléotides permissifs de la banque au niveau de la première position, sont en accord avec l'hypothèse d'une région d'échange de brins de 7 nucléotides, en contradiction avec les résultats publiés précédemment (Coddeville *et al.*, 2014).

La caractérisation de la région de coupure et d'échange de brins par l'utilisation de banques aléatoires a été observée chez le système des intégrons (Escudero *et al.*, 2016). Dans cette étude, l'utilisation de séquences aléatoires du site *attI* suivi d'un séquençage NGS des événements de recombinaison *in vivo* avaient permis le positionnement précis du site de coupure du brin supérieur lors d'une recombinaison *attI* x *attI*. Classiquement, la région d'échange de brins est déterminée par réaction de coupure sur substrat suicide, or dans les différents systèmes hétérobivalents étudiés, il peut parfois être assez difficile de caractériser les sites de coupure sans ambiguïté, ainsi, deux positions des sites de coupures peuvent parfois être proposées (Hauser and Scocca, 1992). Nous avons montré que l'utilisation de banques aléatoires *in vitro* permet de suivre précisément le devenir de chaque position dans la réaction de recombinaison, et est donc un outil intéressant pour la caractérisation des sites de coupures des recombinases. Grâce à la simplicité et à la rapidité de cette technique, elle pourrait être utilisée pour d'autres systèmes de recombinaison à tyrosine, comme par exemple le système du bactériophage P2 où les sites de coupures de ^{P2}Int n'ont pas encore pu être caractérisés (Sylwan *et al.*, 2010).

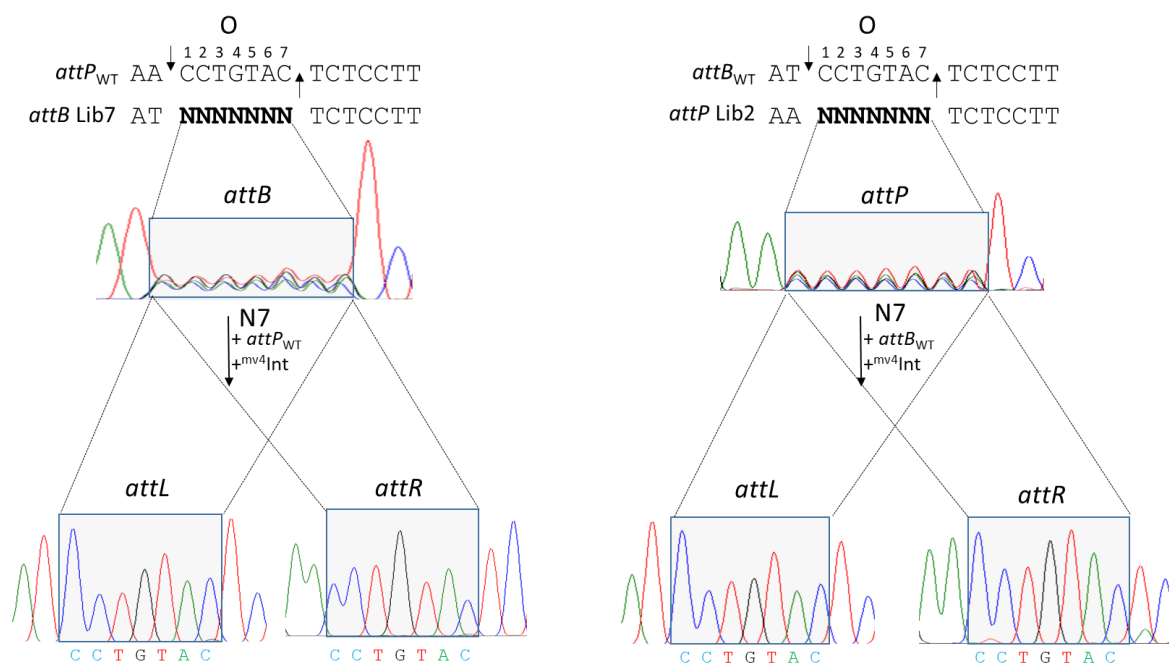


Fig. 62 Résultats de la recombinaison *in vitro* *attB* Lib7 x *attP*_{WT} et *attP* Lib2 x *attB*_{WT}. Les sites *attB*_{WT} et *attP*_{WT} sont représentés, la lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupure encadrant la région centrale.

II.2) Caractérisation des contraintes de la région centrale

La région centrale du système de recombinaison du bactériophage mv4 est donc de 7 pb comme pour tous les systèmes hétérobivalents de bactériophages décrits. Chez le modèle λ Int, il a rapidement été montré que la nature des bases dans la région centrale n'était pas importante pour la recombinaison, mais que l'identité de séquences entre les régions centrales des sites *attB* et *attP* était un paramètre essentiel à la réaction (Bauer *et al.*, 1985; Weisberg *et al.*, 1983). Une caractéristique similaire est observée chez le système de recombinaison du phage HK022 (Kolot *et al.*, 2015) et chez les modèles des recombinases simples Flp (McLeod *et al.*, 1986) et Cre (Hoess *et al.*, 1986), bien qu'il ait été montré que la présence d'un hétéroduplex dans la région centrale des sites *loxP* pouvait parfois être fonctionnelle (Lee and Saito, 1998; Sheren *et al.*, 2007). Ainsi, dans le but de caractériser les contraintes nécessaires à la recombinaison présentes dans la région centrale, deux banques composées d'un motif de 7 pb aléatoire dans la région d'échange de brins ont été construites, les banques *attB* Lib7 et *attP* Lib2 (Fig. 55a), et testées par recombinaison *in vitro* contre les sites *attP* et *attB* natifs, respectivement (Fig. 62). Après recombinaison, parmi les 16000 (4⁷) séquences de départ, seule la séquence correspondant à la région centrale « O » native (« 5'-CCTGTAC-3' ») est retrouvée sur les sites hybrides *attL* et *attR*. Ce résultat montre donc que la recombinaison médiée par $mv4$ Int nécessite une stricte identité entre les régions centrales des sites *attB* et *attP*.

En 1997, Auvray *et al.* avaient montré l'importance des nucléotides « 5'-CCT-3' » dans plusieurs sites *attB* secondaires (Auvray *et al.*, 1997), sans connaître l'existence de la région centrale. Depuis, Coddeville *et al.*, ont démontré que ces nucléotides étaient les 3 premières bases de la région centrale (Coddeville *et al.*, 2014). En étudiant les sites *attB* secondaires de λ Int, Rutkai *et al.* avaient également observé qu'une grande majorité des sites *attB* secondaires de λ Int avaient conservés les 3 premiers nucléotides de la région centrale et il en avait été conclu que ce résultat provenait de l'ordre défini de l'échange des brins (Rutkai *et al.*, 2003). En effet, λ Int commence toujours la formation de la jonction de Holliday par un échange des brins supérieurs c'est-à-dire en échangeant les trois premiers nucléotides de la région centrale (Kitts and Nash, 1988; Nunes-Düby *et al.*, 1987; Nunes-Düby *et al.*, 1995). Ainsi, comme les sites *attB* secondaires du système mv4 ont aussi conservé les 3 premiers nucléotides de la région centrale, il semblerait donc que l'ordre d'échange des brins de $mv4$ Int soit aussi dicté par la formation de l'intasome et que cette réaction commence par un échange des brins supérieurs.

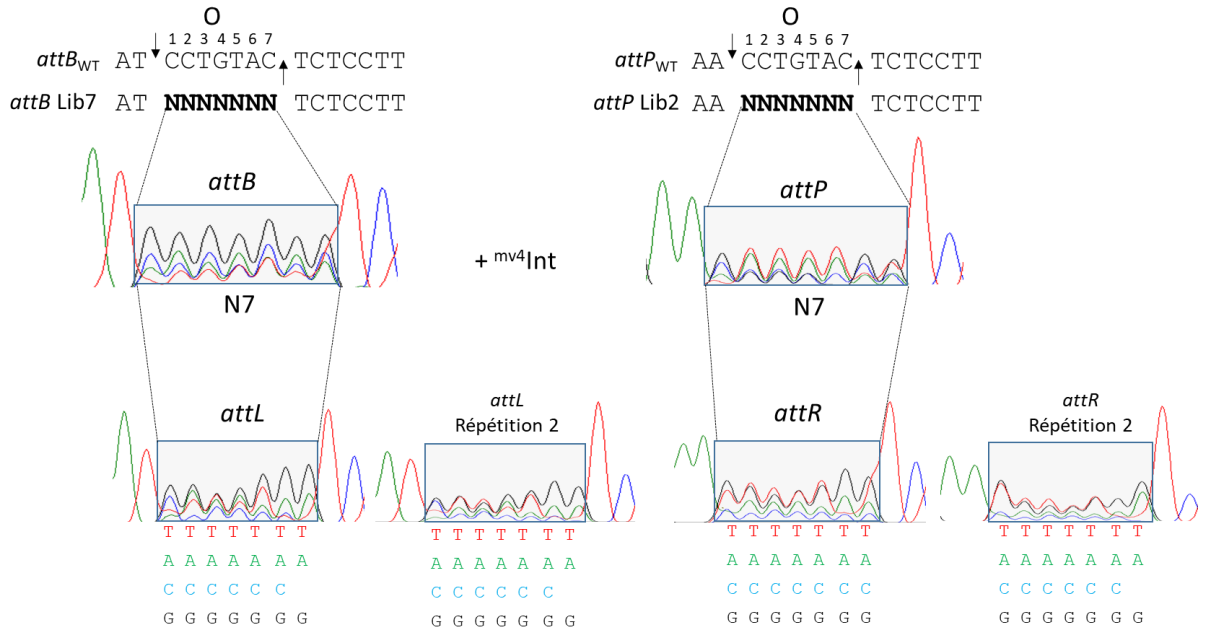


Fig. 63 Résultat de la recombinaison *in vitro* *attB* Lib7 x *attP* Lib2. La lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupure encadrant la région centrale. Les répétitions de 2 expériences indépendantes sont montrées.

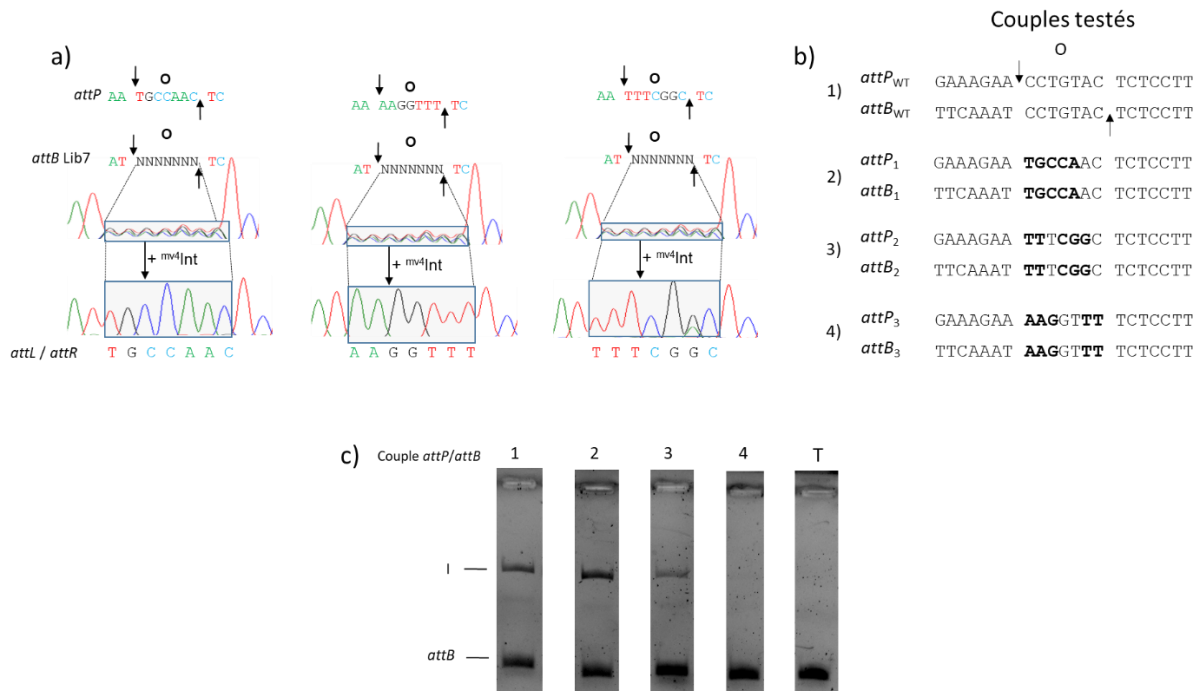


Fig. 64 Etude de 3 régions centrales différentes. **a)** Résultat de la recombinaison *in vitro* entre la banque *attB* Lib7 contre 3 régions centrales différentes du site *attP*. La lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupure encadrant cette région. **b)** Couples de sites *attB/attP* utilisés lors du test de recombinaison fluorescent *in vitro*. Les lettres indiquées en gras sont les nucléotides qui diffèrent par rapport à la séquence sauvage. **c)** Effet de la séquence de la région centrale sur la réaction de recombinaison. La réaction de recombinaison contient 7,2 pmol de *mv4Int*, 40 µg d'extrait cellulaire d'*E. coli* chauffé à 95°C et a été incubée 1h30 à 42°C. Le couple de sites *attB/attP* utilisé est indiqué au-dessus de chaque piste. Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. Piste T, témoin sans *mv4Int*.

De plus, puisque dans les systèmes classiques la nature des bases de la région centrale ne semble pas importante pour la recombinaison, les banques *attB* Lib7 et *attP* Lib2 ont été testées ensembles par recombinaison *in vitro* afin de voir si les contraintes importantes présentes dans la région centrale pouvaient être éliminées grâce à la présence de régions centrales identiques entre les banques (Fig. 63). Après recombinaison, toutes les positions présentent les 4 nucléotides ce qui suggère que de nombreuses séquences sont fonctionnelles, indiquant l'absence de contrainte dans la région centrale, probablement tant que l'identité stricte entre les deux régions centrales testées est conservée. En revanche, en fonction des répétitions et du site hybride observé, certains pics de nucléotides présent en 5' ou en 3' de la région centrale sont absents, ou ambigus. Cela laisse suggérer que certaines contraintes pourraient être présentes aux extrémités de la région centrale, mais qu'elles sont difficilement caractérisables par la méthodologie d'étude utilisée. Le séquençage individuel de chaque site hybride *attL* et *attR* de la population par une approche NGS permettrait de déterminer précisément les contraintes exercées à chaque position dans la région centrale. La présence de contraintes en 5' et en 3' de la région centrale a également été observée chez λ Int, les sites cœur de fixation C et C' chevauchant les 2 premiers nucléotides en 5' et les 2 derniers nucléotides en 3' de la région centrale. Ainsi, il se pourrait que dans le système de recombinaison mv4, les nucléotides présents en 5' et en 3' de la région centrale soient impliqués dans la fixation de $mv4$ Int.

Dans le but de vérifier si la région centrale est modifiable tant que l'identité de séquence est conservée entre les sites de recombinaison, comme c'est le cas pour les recombinases modèles Cre, Flp et λ Int, la banque *attB* Lib7 a été testée par recombinaison *in vitro* contre 3 sites *attP* possédant chacun une région centrale différente (pMET371, pMET372, pMET373, Table 3) (Fig.64a). Pour les trois sites, une seule séquence dans la banque *attB* Lib7 de départ est fonctionnelle : la séquence strictement identique à la région centrale testée, indiquant que la région centrale est modifiable et que le système de recombinaison mv4 se comporte comme les recombinases à tyrosine modèle. Un faible pic de A est observé à la 6^{ème} position lors du test de la région centrale « TTTCGGC », il est donc possible que comme pour Cre, certains hétéroduplex puissent parfois être fonctionnelles. En parallèle, les 3 séquences étudiées ainsi que la séquence sauvage ont été testées par le système fluorescent de recombinaison *in vitro* pour observer l'effet de la séquence de la région centrale sur l'activité de recombinaison (Fig. 64c). Lors de ce test, les sites *attB* et *attP* partagent la même région centrale et possèdent les séquences natives à gauche et à droite de la région centrale (Fig. 64b).

Ainsi, en testant 4 régions centrales différentes, plusieurs effets sont observés après quantification de la fluorescence : aucune différence par rapport à la séquence sauvage (piste 2, Fig. 64c), diminution de l'activité de recombinaison d'un facteur 2 (piste 3, Fig. 64c) et diminution de l'activité d'un facteur 15 (piste 4, Fig. 64c). La séquence « 5'-AAGGTTT-3' » possède un A en première position et un T en dernière position, des nucléotides qui semblent être défavorisés à ces positions dans la région centrale (Fig. 63). Ainsi, l'effet de la séquence de la région centrale sur la réaction de recombinaison était attendu car pour le système Cre/*loxP*, plusieurs études ont montré que la modification de la région centrale de 6 pb pouvait aussi avoir différents effets sur la recombinaison : une légère diminution de l'efficacité de recombinaison par rapport au site *loxP*_{WT} et une forte diminution de l'activité de recombinaison (d'au moins un facteur 20 pour certains variants *lox*) bien que l'identité de séquence entre les régions centrales soit conservée (Lee and Saito, 1998; Sheren *et al.*, 2007). Des résultats similaires ont aussi été obtenus pour le système λ (Nunes-Düby *et al.*, 1997) et Flp/*FRT* (Umlauf and Cox, 1988).

En résumé, la région centrale des sites de recombinaison du système mv4 est une séquence de 7 pb, qui présente les mêmes caractéristiques que celles des modèles Cre, Flp et λ Int, c'est-à-dire qu'une identité stricte est nécessaire à la recombinaison, mais l'identité de séquence ne semble pas l'unique paramètre important pour l'efficacité de la réaction.

III) Caractérisation des contraintes nucléotidiques s'exerçant dans les régions droites des sites *attB* (B') et *attP* (C')

Les banques *attB* Lib8 et *attP* Lib1 décrites dans la partie précédente ont permis de montrer que la taille de la région centrale des sites de recombinaison du système mv4 était de 7 pb. Ainsi, en prenant en compte la taille de la région centrale, ces banques permettent également la caractérisation des contraintes situées à droite de la région d'échange de brins (Fig. 61). Nous avons alors analysé ces contraintes en nous intéressant à l'ensemble des 7 positions aléatoires testées (Fig. 61). Après recombinaison contre les sites *attP* ou *attB* natifs, des contraintes quasi-identiques sont observées pour les deux banques lors de 3 expériences indépendantes. En effet, au niveau de la première position, le C est absent. Ensuite, un triplet de pyrimidines aux positions 2-3-4 semble essentiel à la recombinaison. Au niveau de la position 5, le T est absent, et à la position 7 le A n'est pas observé.

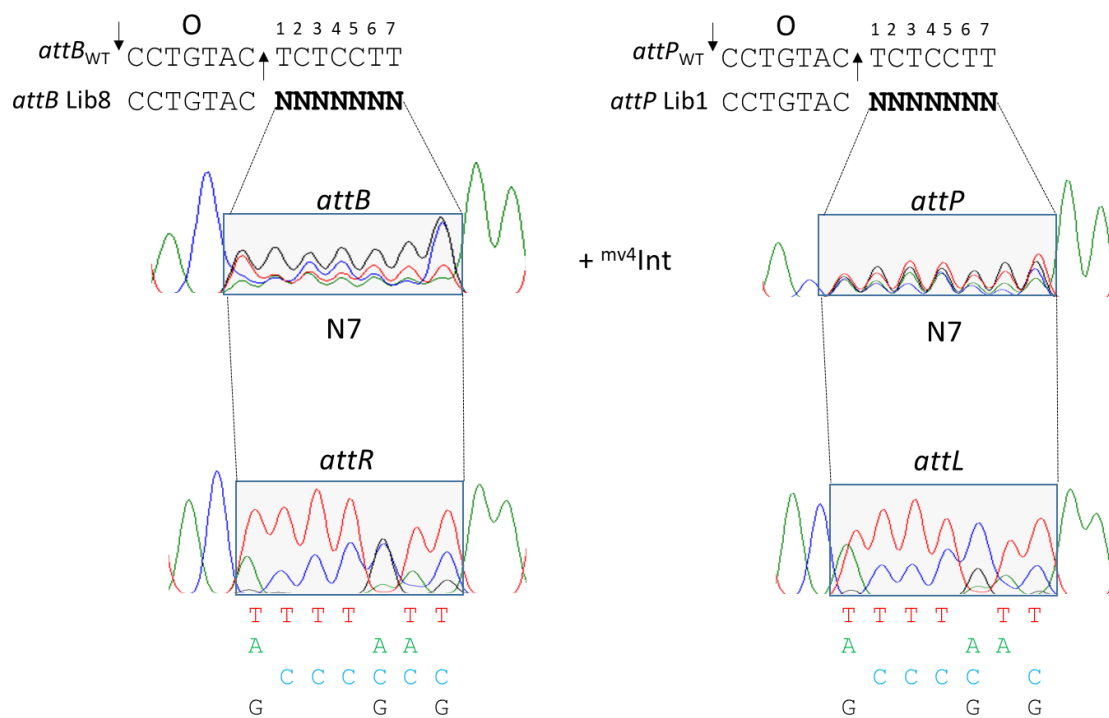


Fig. 65 Résultat de la recombinaison *in vitro* *attB* Lib8 x *attP* Lib1. La lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupure encadrant la région centrale.

La seule différence provient de la position 6 où en plus des nucléotides A et T, un faible pic de C est présent dans la recombinaison *attP*_{WT} x *attB* Lib8 alors que seuls les nucléotides A et T sont retrouvés après recombinaison *attP* Lib1 x *attB*_{WT}. Ces résultats indiquent donc qu'un motif de 7 pb est nécessaire à la recombinaison avec comme motif consensus dégénéré la séquence « 5'-DYYYVHB-3' » (D = A/T/G ; Y = T/C ; V = A/G/C ; H = A/T/C ; B = T/G/C) pour le site *attB* et « 5'-DYYYVWB-3' » (W = A/T) pour le site *attP*. Comme les motifs consensus dégénérés sont très proches, cela suggère que ^{mv4}Int reconnaît un motif particulier à droite de la région centrale, potentiellement un site de fixation dans la région cœur de type B' pour *attB* et C' pour *attP*. Comme les banques *attB* Lib8 et *attP* Lib1 ont permis l'obtention d'un motif dégénéré, cela signifie que différentes séquences sont permissives à la recombinaison et indique qu'une identité stricte entre les sites *attB* et *attP* à droite de la région centrale n'est pas obligatoire pour la réaction. En revanche, il est possible que, comme pour la région centrale, l'identité de séquence entre les sites *attB* et *attP* restaure la fonctionnalité de certains motifs non permissifs. Dans ce cas, une recombinaison entre ces 2 banques (*attB* Lib8 et *attP* Lib 1) devrait donner des sites hybrides *attL* et *attR* présentant les 4 nucléotides possibles à chaque position. Après recombinaison (Fig. 65), les mêmes contraintes que précédemment sont observées, indiquant que ces contraintes sont importantes pour la recombinaison et qu'un motif de type « 5'-DYYYVHB-3' » à droite de la région centrale du site *attB* et « 5'-DYYYVWB-3' » à droite de la région centrale du site *attP* sont nécessaires à la réaction.

IV) Caractérisation des contraintes nucléotidiques s'exerçant dans les régions gauches des sites *attB* (B) et *attP* (C)

IV.1) Caractérisation des contraintes à gauche de la région centrale du site *attB*

La banque *attB* Lib6 décrite dans la partie I.2 a permis de déterminer le début du site *attB* minimum et de montrer que le site *attB* nécessitait la séquence de 7 pb « 5'-TTCAAAT-3' » à gauche de la région centrale au lieu d'une séquence de 2 pb comme publié (Auvray *et al.*, 1999b). Cette banque nous permet également d'analyser les contraintes nucléotidiques essentielles à la recombinaison au niveau des 7 positions situées à gauche de la région d'échange de brins (Fig. 66).

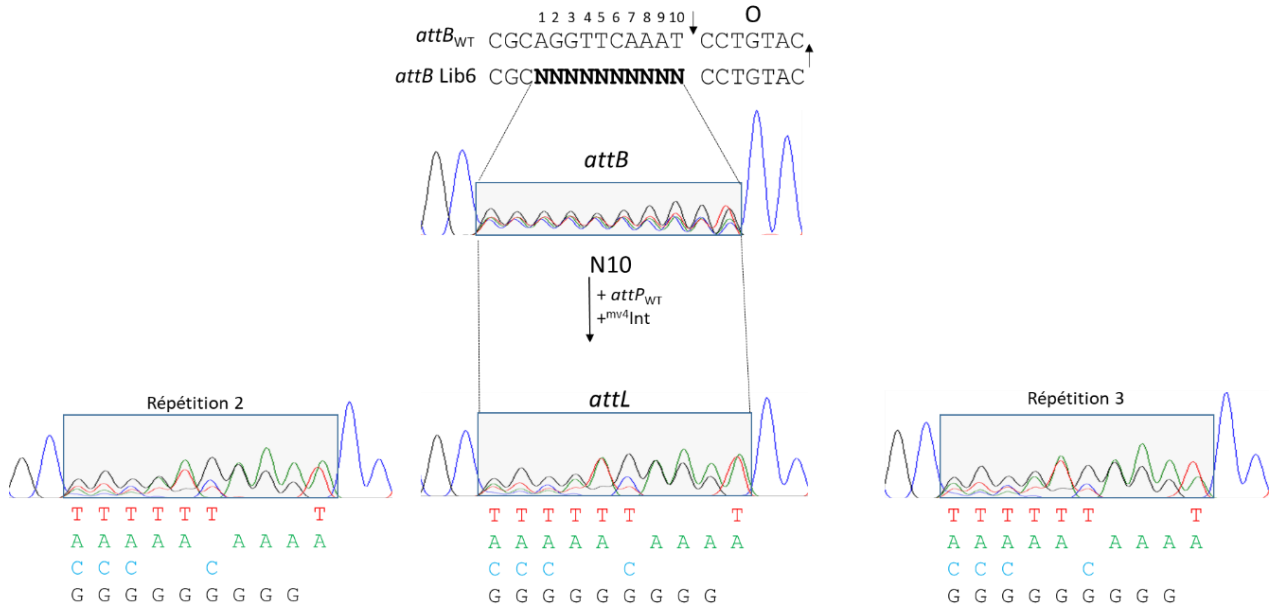


Fig. 66 Résultat de la recombinaison *in vitro* attB Lib6 x attP_{WT}. Cette figure reprend une partie de la figure 58. Le site attB_{WT} est représenté, la lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupe encadrant la région centrale. Les répétitions de 3 expériences indépendantes sont montrées.

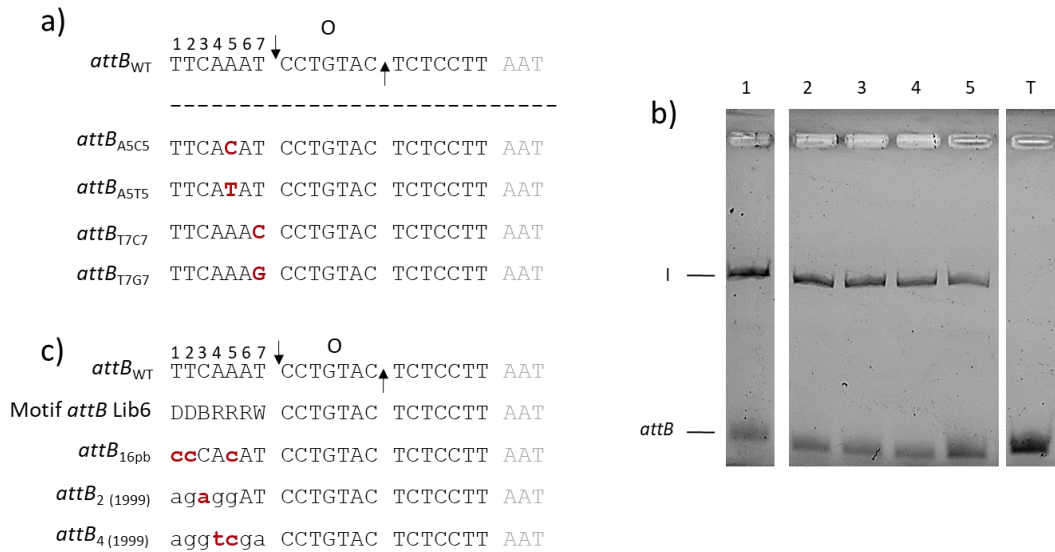


Fig. 67 Etude de différents variants du site attB possédant une ou plusieurs bases non permises dans leurs séquences.

a) Représentation des différents variants construits. Les bases qui ne sont pas retrouvées après recombinaison attP_{WT} x attB Lib6 sont représentées en rouge. Les bases présentes dans la région cœur mais en-dehors du site attB minimum sont représentées en gris. La lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupe encadrant la région. b) Effet de la présence d'un nucléotide non permise dans le site attB sur la réaction de recombinaison. La réaction de recombinaison contient 7,2 pmol de mv4Int, 40 µg d'extrait cellulaire d'*E. coli* chauffé à 95°C et a été incubée avec les substrats attP_{WT} et attB 1h30 à 42°C. Les sites attB sont différents dans chaque piste : 1, attB_{WT}; 2, attB_{A5C5}; 3, attB_{A5T5}; 4, attB_{T7C7}; 5, attB_{T7G7} et T, témoin sans mv4Int. Le fragment attB fluorescent et le produit de recombinaison entre les sites attB et attP (I) sont indiqués sur le gel. c) Comparaison de la séquence consensus dégénérée de 7 pb obtenue par la banque attB Lib6 avec les résultats obtenus en 1999. Les bases en minuscule sont les bases qui diffèrent par rapport au site attB_{WT}. Noir, bases présentes dans le motif « 5'-DDBRRRW-3' » ; rouge, bases absentes du motif « 5'-DDBRRRW-3' ».

Après recombinaison avec le site *attP* natif, un motif de recombinaison de 7 pb est observé. Les positions 4-5-6 présentent une faible contrainte (absence de C ou de A). Ensuite, un triplet de purines aux positions 7-8-9 semble essentiel à la recombinaison. Enfin, à la position 10, seuls les nucléotides A et T sont présents. Ce résultat indique donc la nécessité d'un motif de 7 pb pour la recombinaison de type « 5'-DDBRRRW-3' » (D = A/T/G ; B = T/G/C ; R = G/A ; W = A/T).

Dans le but de vérifier si les contraintes observées sont dues à la présence d'un nucléotide non permissif à une position précise ou bien à un effet de combinaison de bases non permissives avec une recombinaison, 4 variants du site *attB* ont été construits en réalisant des substitutions à des positions où ces bases étaient absentes d'après l'analyse de la banque Lib6 (A₅ → C₅ ; A₅ → T₅ ; T₇ → C₇ ; T₇ → G₇) (Fig. 67a). Ces variants ont été testés par recombinaison *in vitro* (pistes 2 à 5, Fig. 67b) et tous les sites sont fonctionnels, bien que l'activité soit plus faible pour certains. Ce résultat indique donc que la présence d'un nucléotide non-permissif à une seule position n'est pas suffisante pour empêcher la réaction de recombinaison, ce qui suggère que les contraintes observées avec la banque *attB* Lib6 proviennent probablement de l'effet additif de plusieurs bases non permissives.

Le motif dégénéré obtenu pourrait fournir une explication à l'absence de signal de recombinaison avec le site de 16 pb que nous avons testé *in vitro* et pourrait également expliquer comment Auvray *et al.*, ont pu déterminer un site minimum de 16 pb (Fig. 67c). En effet, en comparant le site *attB* de 16 pb, que nous avons testé en recombinaison *in vitro*, à la séquence consensus dégénérée obtenue, 3 nucléotides non-permissifs à la recombinaison sont présents, ce qui semblerait suffisant pour inactiver la recombinaison. Dans l'étude précédente (Auvray *et al.*, 1999b), le site *attB*₂ est le site le plus court permettant une recombinaison *in vivo*. Au niveau de ce site, un seul nucléotide non-permissif est retrouvé (Fig. 67c) et nous avons observé que cela n'était pas suffisant pour inactiver la recombinaison. La perte d'activité du site *attB*₄ avait été attribuée à la modification des nucléotides « AT » par « GA » (positions 6 et 7, Fig. 67c) (Auvray *et al.*, 1999b). Nos résultats indiquent qu'un G à la position 6 et qu'un A à la position 7 sont des nucléotides permissifs pour la recombinaison, par contre, en observant la séquence globale de ce site, 2 pyrimidines sont retrouvées dans le triplet de purines, ce qui pourrait expliquer l'absence de fonctionnalité de ce site.

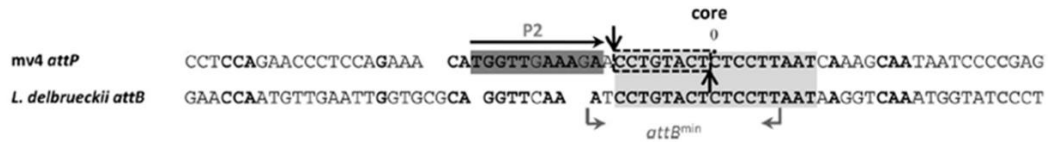


Fig. 68 Organisation du site *attP* du bactériophage mv4. Structure et séquence nucléotidique du site P2 et de la région cœur (adapté de Coddeville *et al.*, 2014). La région cœur de 17 pb est surligné en gris clair, et la séquence du site P2 en gris foncé. Les flèches grises indiquent le site *attB* minimum de 16 pb et les flèches noires correspondent aux sites de coupe encadrant la région centrale.

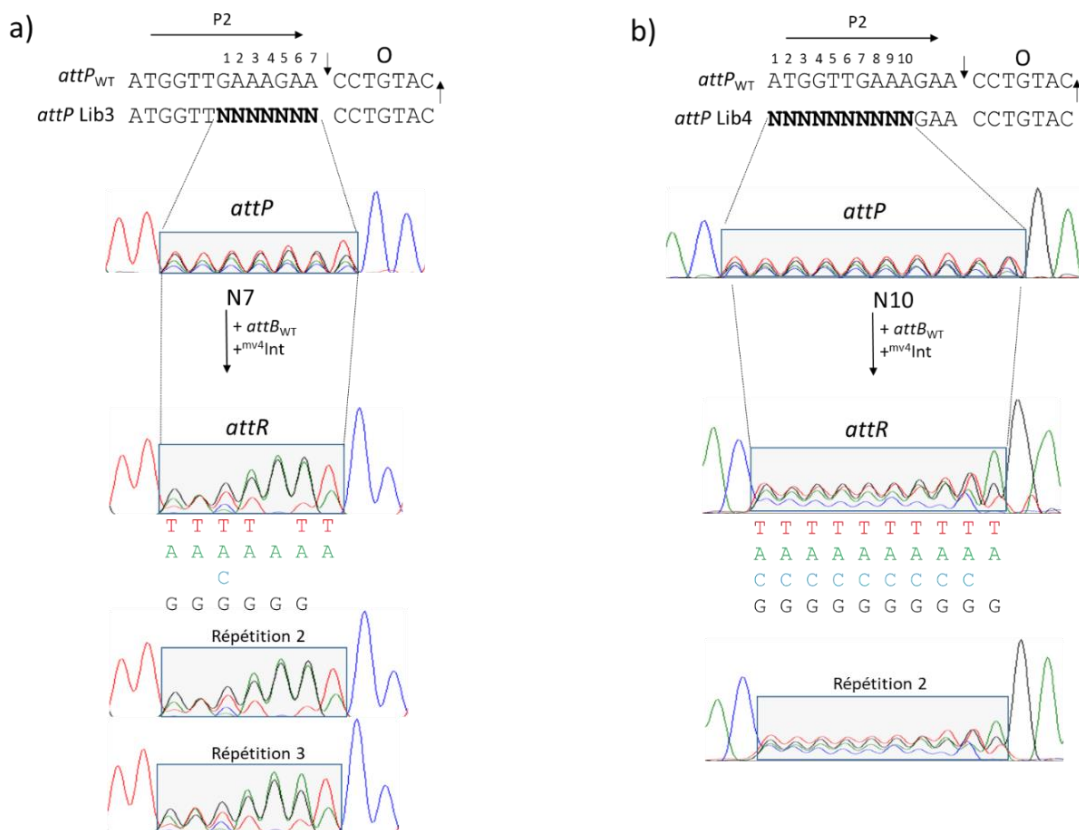


Fig. 69 Caractérisation des contraintes à gauche de la région centrale du site *attP*. Le site *attP*_{WT} est représenté, la lettre O indique la région centrale et les deux flèches indiquent les sites de coupe encadrant la région centrale. Le site bras P2 atypique est représenté par une flèche horizontale. Les lettres N indiquent les positions aléatoires de la banque. Les pics du chromatogramme indiquent les différents nucléotides présents à chaque position. Une couleur est associée à chaque nucléotide sur le chromatogramme : bleu, cytosine ; noir, guanine ; rouge, thymine et vert, adénine. **a)** Résultat de la recombinaison *attP* Lib3 x *attB*_{WT}. **b)** Résultat de la recombinaison *attP* Lib4 x *attB*_{WT}.

IV.2) Caractérisation des contraintes à gauche de la région centrale du site *attP*

Une particularité du site *attP* du bactériophage mv4 provient du positionnement atypique du site bras de fixation de ^{mv4}Int P2, qui se trouve à proximité de la région d'échange de brin (Fig. 68). Dans le but de vérifier la présence du site bras P2, une banque comportant un motif de 7 pb à gauche de la région centrale a été construite (*attP* Lib3, Fig. 55a ; Fig. 69a) et testée par recombinaison *in vitro* contre le site *attB*_{WT}. Après recombinaison, le motif obtenu présente de fortes similarités avec la séquence consensus dégénérée située à gauche de la région centrale du site *attB* (« 5'-DDBRRRW-3' »). En effet, les 2 premières positions ne présentent pas de C et la dernière position ne présente que les nucléotides A et T. De plus, au niveau des positions 4-5-6, il y a une très forte importance des purines, bien qu'un pic de T plus faible soit observé aux positions 4 et 6. Ainsi, un motif de recombinaison de type « 5'-DDNDRDW-3' » semble essentiel à la recombinaison à gauche de la région centrale du site *attP*. La position 7 est une position située hors du site bras P2, mais qui présente une forte contrainte puisque seuls les nucléotides A et T sont permissifs à la recombinaison. Au vu de la similarité avec le motif présent sur le site *attB*, il est fort probable que le site bras P2 soit en réalité un site cœur de fixation de ^{mv4}Int, le site cœur C.

Afin de vérifier l'absence de contraintes au niveau des positions du site bras P2, une nouvelle banque *attP* comportant un motif de 10 nucléotides aléatoires situés dans le site bras P2 a été construite (*attP* Lib4, Fig. 55a ; Fig. 69b). Cette banque a été testée contre le site *attB* natif (Fig. 69b) et, après recombinaison, aucune contrainte n'est observée au niveau des positions 2 à 6, positions appartenant au site P2, indiquant l'absence de site bras à gauche de la région centrale. En revanche, nous aurions dû observer des contraintes (absence de C aux positions 7 et 8) dans le site cœur C (positions 1-2, Fig. 69a ; positions 7-8, Fig. 69b). Malgré les deux répétitions, aucune contrainte n'a été observée à ces positions mais les séquences obtenues ne sont pas très précises puisqu'un bruit de fond est constamment retrouvé sous les pics. Lorsque le bruit de fond est présent dans la région aléatoire, des pics « faux-positifs » peuvent être observés. En regardant les positions 7 et 8 du site *attR* (Fig. 69b), le C est bien moins présent que les 3 autres nucléotides, et il est possible que le signal obtenu soit un « faux-positif ». Ainsi, pour valider que la région importante à gauche de la région centrale du site *attP* est bien une séquence de 7 pb, il faudrait construire une banque de 10 pb, comme la banque *attB* Lib6, ce qui permettrait de montrer que les positions 1 à 3 de cette banque se situent en dehors de la région cœur du site *attP* (Fig. 55a).

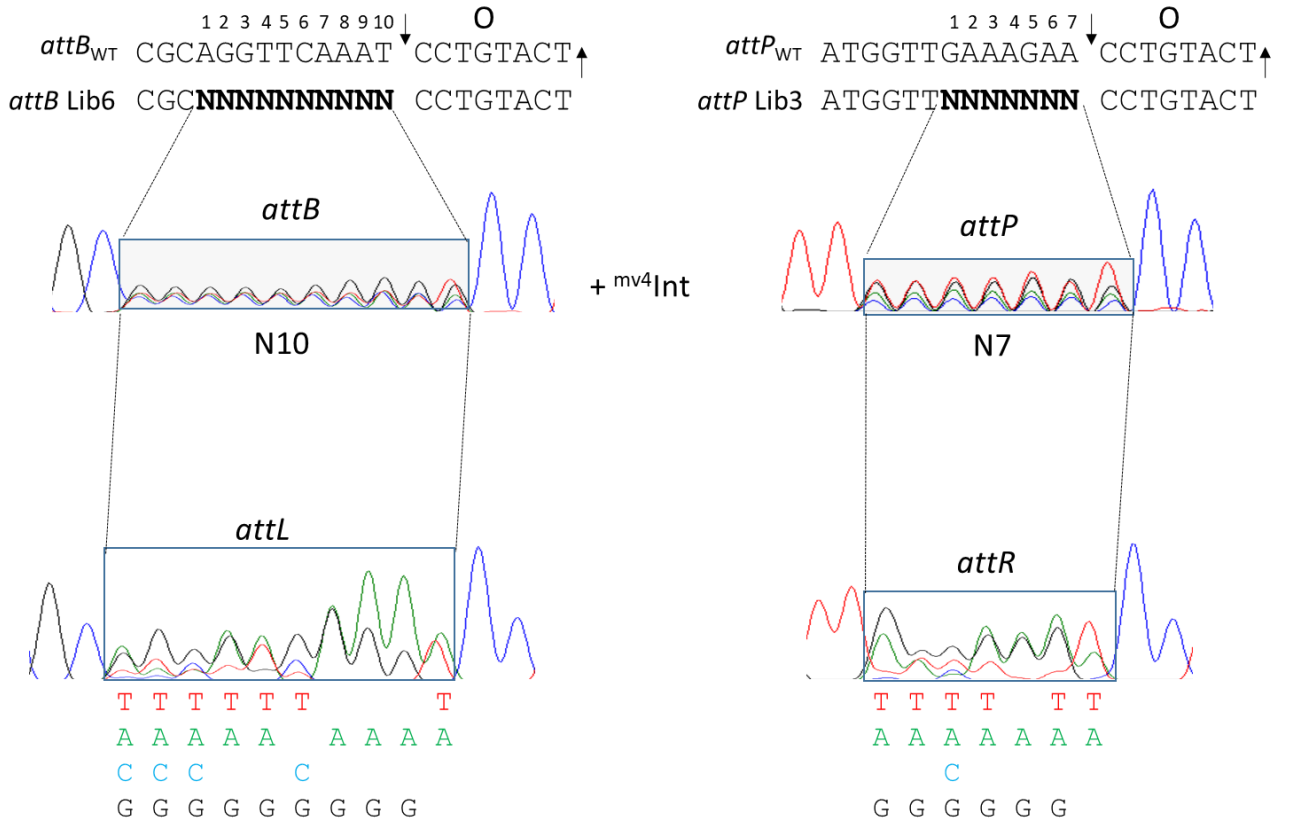


Fig. 70 Résultat de la recombinaison *in vitro* $attB$ Lib6 x $attP$ Lib3. La lettre O indique la région « Overlap » (région centrale) et les deux flèches indiquent les sites de coupure encadrant la région centrale

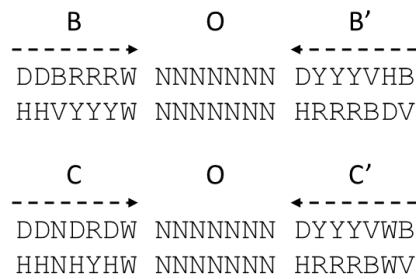


Fig. 71 Séquences consensus dégénérées du site $attB$ et de la région cœur du site $attP$.

En résumé, nos résultats montrent que le site bras atypique P2 n'est pas une séquence de 11 pb, mais plutôt une séquence de 7 pb, qui pourrait être le site cœur C de la région cœur du site *attP*. ^{mv4}Int effectue une recombinaison sur des sites *attB* et *attP* présentant une séquence de 7 pb différente à gauche de la région centrale (« 5'-TTCAAAT-3' » pour *attB* et « 5'-GAAAGAA-3' » pour *attP*, Fig. 68). En revanche, il est possible que, comme pour la région centrale, l'identité de séquence entre les sites *attB* et *attP* restaure la fonctionnalité de certains motifs non permissifs. Dans ce cas, une recombinaison entre les 2 banques *attB* Lib6 et *attP* Lib3 devrait donner des sites hybrides *attL* et *attR* présentant les 4 nucléotides possibles à chaque position. Après recombinaison (Fig. 70), les mêmes motifs de recombinaison ont été obtenus indiquant que seule la présence du motif « 5'-DDBRRRW-3' » sur le site *attB* et « 5'-DDNDRDW-3' » sur le site *attP* semble essentielle à la recombinaison.

En résumé, la comparaison des séquences consensus dégénérées obtenues dans le site *attB* et *attP* (Fig. 71) montre un motif répété-inversé « 5'-DBRRR-3' » encadrant la région centrale de 7 pb sur le site *attB*. L'utilisation de banques aléatoires a donc permis de mettre en évidence les sites cœur de fixation B et B' sur le site *attB* et C et C' sur le site *attP*. Ainsi, le système de recombinaison mv4 présente une architecture plus classique des sites de recombinaison, avec des sites B, B', C et C', encadrant une région centrale dans laquelle la nature des nucléotides ne semble pas importante. En revanche, le système ^{mv4}Int/*attP*/*attB* se distingue des intégrases hétérobivalentes classiques par une forte flexibilité vis-à-vis de ces sites de recombinaison.

V) Etude de la présence des sites cœur de fixation de ^{mv4}Int par retard sur gel

V.1) Mise au point de la technique de retard sur gel

Dans le but de démontrer expérimentalement la présence des sites cœur de fixation de ^{mv4}Int, la méthodologie de retard sur gel a été utilisée. Pour cela, nous nous sommes inspirés du protocole développé par Sarkar *et al.* sur la région COC' de ^λInt (Sarkar *et al.*, 2001). En effet, pendant de nombreuses années, il a été difficile d'étudier la région cœur des sites de recombinaison car la fixation de ^λInt à ses sites était instable. Sarkar *et al.* ont démontré que le domaine N-ter exerce un effet inhibiteur sur le domaine C-ter lorsqu'il n'est pas fixé aux sites bras. En revanche, si de l'ADN possédant les séquences des sites bras P'12 est ajouté à la réaction, cette inhibition est levée et ^λInt est capable de se fixer sur les sites cœur *via* son domaine C-ter (Sarkar *et al.*, 2001).

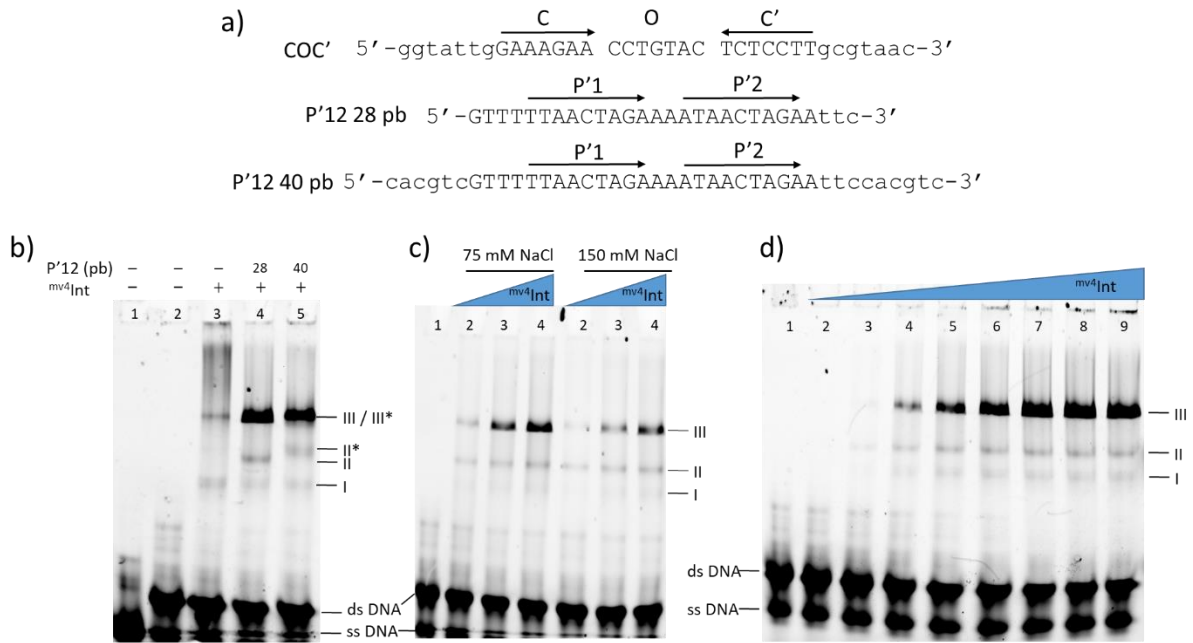


Fig. 72 Mise au point de la méthodologie de retard sur gel. a) Oligonucléotides utilisés. Les bases en majuscules correspondent aux régions présentes dans le site *attP*. Les bases en minuscules sont des nucléotides différents des nucléotides naturellement présent. **b) Effet de la fixation de mv⁴Int aux sites bras P'12 sur la fixation à la région cœur du site attP.** La région cœur du site *attP*, fluorescent à la cyanine-3 (0,87 pmol) a été incubée en présence ou en absence de mv⁴Int (25 pmol) et en présence ou en absence de 4,48 pmol de sites bras P'12 non-marqués. Les réactions ont été analysées par électrophorèse native PAGE 7,5% et la fluorescence a été visualisée sur l'appareil ChemidocMP (Biorad). La présence ou l'absence de mv⁴Int et de la séquence site bras P'12 (28 ou 40 pb) est indiquée au-dessus des pistes. Les différents complexes observés sont indiqués à côté du gel : ssDNA, ADN fluorescent simple-brin ; dsDNA, ADN fluorescent double-brin ; complexe I, monomère de mv⁴Int fixée à la région cœur ; complexe II, monomère de mv⁴Int fixée à la région cœur et à un site bras P'12 de 28 pb (II) ou 40 pb (II*) ; complexe III, dimère de mv⁴Int fixée à la région cœur et aux sites bras P'12 de 28 pb (III) ou 40 pb (III*). **c) Optimisation de la concentration en NaCl.** La réaction contient 0,87 pmol de région cœur du site *attP* marqué à la cyanine-3, 4,48 pmol de sites bras P'12 non marqués, et 0 pmol (1), 3,6 pmol (2), 7,2 pmol (3) ou 10,8 pmol (4) de mv⁴Int. Les deux conditions en NaCl testés sont indiqués au-dessus du gel. La légende est identique à la partie b). **d) Détermination du seuil maximal de détection de la fluorescence par ajout croissant de mv⁴Int.** Les conditions sont identiques aux parties précédentes dans 75 mM de NaCl. La réaction contient 0 pmol (1), 1,2 pmol (2), 3,6 pmol (3), 10,7 pmol (4), 18 pmol (5), 25,1 pmol (6), 35,8 pmol (7), 47,8 pmol (8) ou 60 pmol (9) de mv⁴Int. La légende est identique à la partie b).

Ainsi, pour la mise au point de la technique de retard sur gel, le même principe a été employé (Fig. 72a). Une amorce fluorescente (Cy3) contenant la région cœur COC' native de 21 pb caractérisée grâce aux banques aléatoires a été synthétisée et hybridée en ADN double-brin (Fig. 72a). La migration de l'ADN simple-brin fluorescent (piste 1, Fig. 72b) et de l'ADN double-brin fluorescent (piste 2, Fig. 72b) montre que lors de l'hybridation, tout l'ADN simple brin n'est pas converti en ADN double-brin. Les conditions d'hybridation de l'ADN ont donc été optimisées pour les futures expériences décrites afin d'empêcher la présence d'un mélange ADN simple-brin / ADN double-brin (voir figures suivantes). En présence de mv^4Int mais en absence des sites bras P'12 du site *attP* (piste 3, Fig. 72b), deux complexes sont observés. En considérant les données d'autres systèmes de recombinaison à tyrosine, nos données suggèrent que le complexe I est composé d'un monomère de mv^4Int fixé à un des deux sites cœur (C ou C') par son domaine C-ter et que la liaison d'un deuxième monomère de mv^4Int au deuxième site cœur permet la formation du complexe III. En revanche, un « smear » important est observé dans la piste 3, indiquant que les complexes formés en absence de sites bras sont instables. L'ajout d'un fragment de 28 pb possédant les sites bras P'12, (piste 4, Fig. 72b) stabilise très fortement le complexe III, suggérant que la fixation du domaine N-ter de mv^4Int aux sites bras stabilise la fixation du domaine C-ter du dimère de mv^4Int aux sites cœur. Un complexe supplémentaire, le complexe II, est observé. Il pourrait contenir un monomère de mv^4Int fixé à la région cœur *via* son domaine C-ter et à un site bras P'12 *via* son domaine N-ter. Ainsi, dans la région cœur de 21 pb caractérisée précédemment, 2 molécules de mv^4Int sont capables de se fixer, suggérant la présence des deux sites cœur de fixation C et C' de mv^4Int . Pour confirmer que l'effet de stabilisation observé est dû à la fixation des domaines N-ter de mv^4Int aux sites bras P'12, un fragment de 40 pb, une taille plus importante de 12 pb et incluant la région P'12 a été testé (piste 5 Fig. 72b). Tout comme le fragment de 28 pb, l'ajout de ce fragment permet l'obtention de 3 complexes : le complexe I où un monomère de mv^4Int est fixé à la région cœur par son domaine C-ter, le complexe II* dû à la fixation d'un monomère de mv^4Int à la région cœur et aux sites bras P'12 *via* son domaine N-ter et le complexe III* en raison de la fixation d'un dimère de mv^4Int à la région cœur et aux sites bras P'12. Les complexes II* et III* migrent à une taille supérieure par rapport aux complexes II et III en raison de la présence de 12 pb supplémentaires dans le fragment possédant les sites bras (Fig. 72b). Ainsi, le complexe mv^4Int -région cœur est stabilisé grâce à la fixation spécifique des domaines N-ter de mv^4Int aux sites bras P'12 qui semble induire un réarrangement de la protéine plus favorable à la fixation de la région cœur.

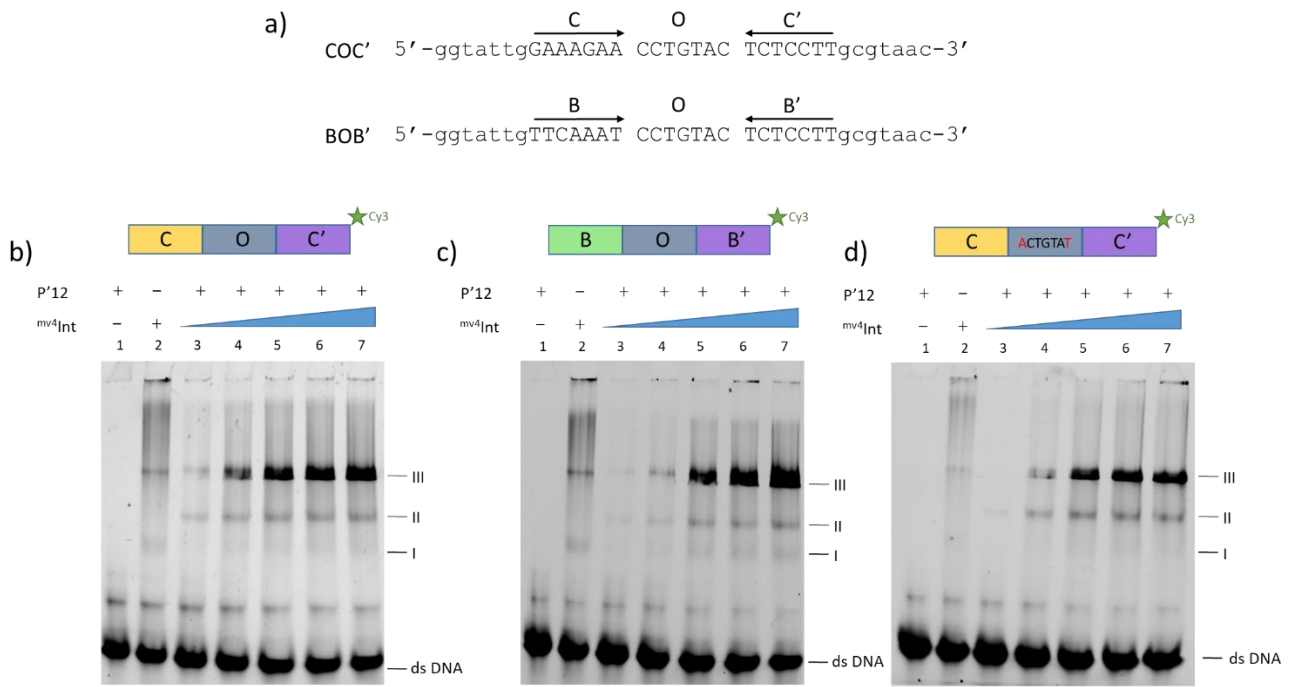


Fig. 73 Caractérisation des sites cœur B, B', C et C' par retard sur gel. **a)** Séquence des fragments fluorescents BOB' et COC' utilisés. **b)** Etude du fragment COC' en présence de *mv*⁴Int et des sites bras P'12. La présence et l'absence de *mv*⁴Int ou des sites bras P'12 est indiquée au-dessus des pistes. Un fragment fluorescent à la cyanine-3 (COC') (0,87 pmol) a été incubé en présence de 4,48 pmol de sites bras P'12 non-marqués. La réaction contient 0 pmol (1), 3,6 pmol (3), 10,7 pmol (4), 18 pmol (5), 25,1 pmol (2 et 6) et 35,8 pmol (7) de *mv*⁴Int. Les trois complexes sont indiqués sur le côté : complexe I, fixation d'un monomère de *mv*⁴Int à la région cœur ; complexe II, fixation d'un monomère de *mv*⁴Int à la région cœur et à un site bras P'12 ; complexe III, fixation d'un dimère de *mv*⁴Int à la région cœur et aux sites bras P'12. **c)** Etude du fragment BOB' en présence de *mv*⁴Int et des sites bras P'12. La légende est identique à la partie **b)**. **d)** Etude de l'effet du changement des premiers et derniers nucléotides de la région centrale par A et T sur la fixation de *mv*⁴Int. La légende est identique à la partie **b)**.

La concentration en NaCl dans le tampon de réaction a aussi été optimisée en testant deux concentrations : 75 mM comme dans le protocole de Sarkar *et al.*, (Sarkar *et al.*, 2001) et 150 mM comme dans le tampon de recombinaison (Fig. 72c). Dans chaque condition testée, le signal de fluorescence du complexe III est le plus intense à 75 mM de NaCl et cette condition sera conservée pour les futures expériences. Enfin, le seuil maximal de fluorescence a été détecté en testant des quantités croissantes de mv^4Int (Fig. 72d). A partir de 35,8 pmol (piste 7, Fig. 72d), le signal de fluorescence est saturé, et cette quantité maximale de mv^4Int a été conservée pour la suite de l'étude.

V.2) Les sites cœur B, B', C et C' sont des sites de fixation de mv^4Int

Dans le but de démontrer expérimentalement la présence des sites cœur de fixation de mv^4Int , deux fragments fluorescents (Cy3) comportant les sites cœur *attB* ou *attP* de 21 pb définis dans les parties précédentes ont été synthétisés (Fig. 73a), hybridés en ADN double-brin à l'aide d'un ADN simple brin complémentaire non marqué, et testés pour leur capacité à fixer mv^4Int (Fig. 73b ; Fig. 73c). En absence de sites bras P'12, mv^4Int forme un complexe instable avec la région cœur de 21 pb du site *attP* (piste 2, Fig. 73b) et avec le site *attB* de 21 pb (piste 2, Fig. 73c). En effet, deux complexes (I et III) sont observés, permettant la fixation d'un monomère de mv^4Int (I) et d'un dimère de mv^4Int (III), mais ces complexes sont peu intenses et un « smear » important est observé. En revanche, l'ajout des sites bras P'12 (pistes 3 à 7, Fig. 73b ; Fig. 73c) stabilise fortement ces complexes puisque, comme précédemment, deux complexes sont observés, celui correspondant à la fixation d'un monomère de mv^4Int à la région cœur des sites *attB* et *attP* ainsi qu'à un site bras P'12 (complexe II) et celui résultant de la fixation d'un dimère de mv^4Int dans la région cœur des sites *attB* et *attP* et au niveau des sites bras P'12 (complexe III). Ainsi, ces résultats suggèrent la présence de deux sites cœur de fixation de mv^4Int , les sites B et B' dans le site *attB* et les sites C et C' dans la région cœur du site *attP*. Comme observé pour λInt , l'absence de fixation du domaine N-ter de mv^4Int aux sites bras P'12 déstabilise fortement le complexe domaine C-ter/région cœur, alors que l'ajout de ces sites bras lève l'inhibition qu'exerce le domaine N-ter et permet la formation de complexes stables avec la région cœur des sites *attB* et *attP*. Ainsi, ces résultats suggèrent l'importance de la formation de l'intasome pour la recombinaison, car en absence de cette structure, l'intégrase semble incapable de former un complexe stable avec la région cœur de ces sites de recombinaison et est donc probablement incapable d'effectuer la réaction de recombinaison.

Chapitre 2

Site cœur	λ Int	HP1Int	L5Int	mv4Int
B	5' -CTGCTTTTT-3'	5' -AAGGATTTA-3'	5' -AGAGCCCTAGT-3'	5' -TTCAAAT-3'
C	5' -CAGCTTTTT-3'	5' -ATGGTATTTA-3'	5' -AGACACCCATT-3'	5' -GAAAGAA-3'
B'	5' -CAAGTTAGT-3'	5' -AGGGATTTTA-3'	5' -GCGTAGCTAGT-3'	5' -AAGGAGA-3'
C'	5' -CAACTTAGT-3'	5' -AGGGATTTTA-3'	5' -GCGTAGCTAGT-3'	5' -AAGGAGA-3'

Fig. 74 Comparaison des séquences B/C et B'/C' pour chaque système de recombinaison des bactériophages λ , L5, HP1 et mv4. Les bases indiquées en rouge sont les bases qui diffèrent entre B et C ou B' et C'. Les séquences B' et C' ont été inversée-complémentée pour pouvoir les comparer avec les séquences B et C.

En revanche, chez une intégrase hétérobivalente d'ICE (CTnDOT), IntDOT est capable de former un complexe stable avec la région cœur en absence des sites bras (Wood *et al.*, 2010), suggérant que l'interaction de IntDOT avec les sites bras est moins importante pour la régulation de l'activité catalytique. Un résultat similaire a été obtenu avec l'intégrase P2 où P2Int est capable de former un complexe stable avec la région cœur du site *attP* en absence des sites bras de fixation (Frumerie *et al.*, 2005). Ainsi, chez ^{mv4}Int, il semblerait que comme pour le modèle ^λInt, les sites bras aient un rôle supplémentaire à leur rôle structural et participent à la régulation de l'activité catalytique. Il serait intéressant de regarder si la délétion du domaine N-ter permet au domaine C-ter de former des complexes stables avec la région cœur et si l'activité de résolution de jonction de Holliday est aussi améliorée.

Comme il est possible de visualiser la fixation de ^{mv4}Int à la région cœur, nous avons profité de cette méthodologie pour tester notre hypothèse émise précédemment. Lors de l'étude de l'effet de la séquence de la région centrale sur l'activité de recombinaison, la séquence « 5'-AAGGTTT-3' », inhibe l'activité de recombinaison d'un facteur 15. Comme dans le système ^λ les sites cœurs chevauchent les deux premiers nucléotides en 5' et en 3' de la région centrale, nous avons proposé que les nucléotides en 5' et en 3' de la région centrale étaient impliqués dans la fixation de ^{mv4}Int, et que la présence d'un A en 5' et d'un T en 3' pouvaient entraver la fixation de ^{mv4}Int sur ses sites cœur. Pour tester cette hypothèse, un fragment fluorescent possédant la région cœur COC', avec le nucléotide A en 5' et T en 3' de la région centrale a été synthétisé (Fig. 73d). On peut observer que ce fragment présente un profil de fixation de ^{mv4}Int similaire à la séquence COC'_{WT} indiquant que la présence d'un A en 5' et d'un T en 3' de la région centrale n'interfère pas avec la fixation de ^{mv4}Int. Comme ces modifications semblent perturber l'activité de recombinaison, il se pourrait que ces bases entravent la recombinaison à une étape ultérieure à la fixation de ^{mv4}Int, potentiellement à l'étape de coupure des brins.

V.3) Caractérisation de la capacité de fixation de ^{mv4}Int sur les sites uniques B, B', C et C'

Les séquences des sites cœur B, B', C et C' des bactériophages ^λ, L5, HP1 et mv4 ont été alignées et l'identité de séquence entre les sites B et C ou B' et C' a été comparé pour chaque système de recombinaison (Fig. 74). Le modèle ^λ se distingue par le fait de n'avoir aucun site cœur identique, mais des séquences entre B et C ou entre B' et C' très proches. Les systèmes HP1 et L5 possèdent des sites cœur B et C proches, et des séquences B' et C' identiques.

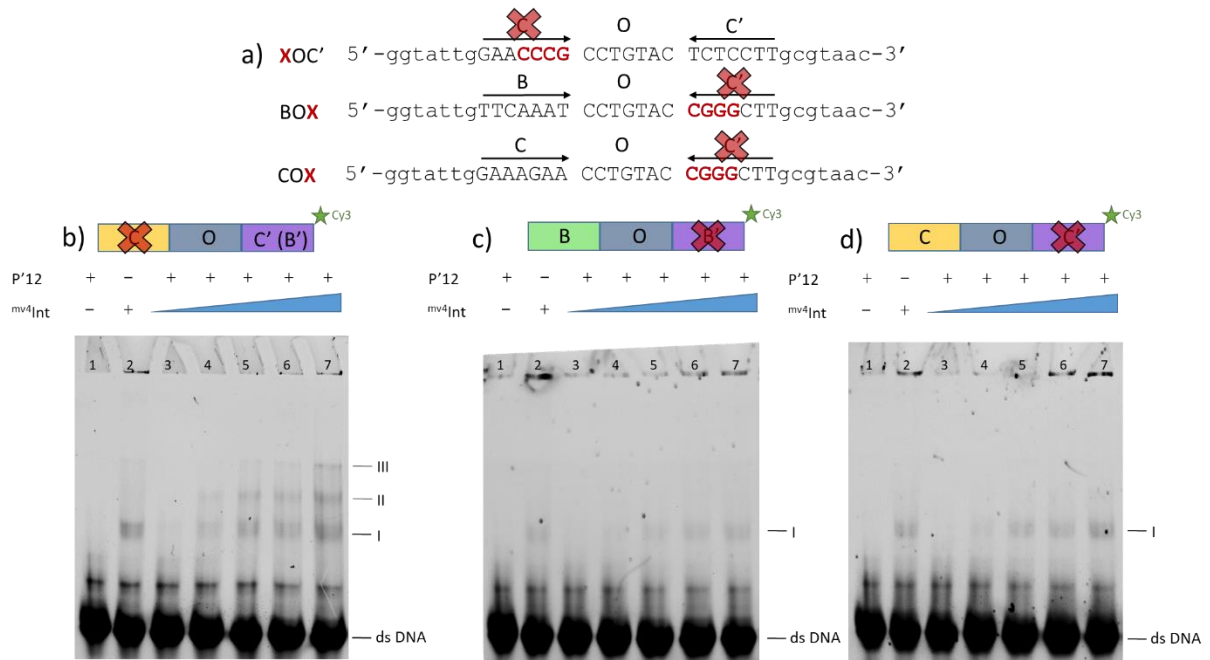


Fig. 75 Affinité des sites cœur B, B', C et C' par retard sur gel. a) Séquence des fragments fluorescents permettant l'étude des sites uniques B, C et B'(C'). b) Site cœur B'(C') en présence de *mv4Int* et des sites bras P'12. La présence et l'absence de *mv4Int* ou des sites bras P'12 est indiquée au-dessus des pistes. Un fragment fluorescent à la cyanine-3 (0,87 pmol) a été incubé en présence de 4,48 pmol de sites bras P'12 non-marqués. La réaction contient 0 pmol (1), 3,6 pmol (3), 10,7 pmol (4), 18 pmol (5), 25,1 pmol (2 et 6) et 35,8 pmol (7) de *mv4Int*. Les trois complexes sont indiqués sur le côté et sont identiques à ceux déterminés précédemment. c) Site cœur B. La légende est identique à la partie b). d) Site cœur C. La légende est identique à la partie b).

Enfin, bien que le système mv4 possède des séquences B' et C' identiques, ce système se distingue par l'utilisation de deux séquences B et C différentes (Fig. 74). Ainsi, $mv4Int$ est capable de reconnaître 3 séquences différentes indiquant une grande flexibilité dans sa capacité à se fixer dans la région cœur de ces sites de recombinaison. Dans le but d'étudier le rôle ainsi que l'affinité de $mv4Int$ envers ces 3 séquences, les séquences B', C et C' ont été inactivées en remplaçant le triplet de purines par un triplet de pyrimidines « CCC », et en remplaçant le triplet de pyrimidines par un triplet de purines « GGG » (Fig. 75a). De cette façon, ces 3 fragments permettent l'étude de la fixation de $mv4Int$ sur le site unique B, C et B'/C' (puisque la séquence de B' est identique à C'). L'affinité de $mv4Int$ sur les sites B' et C' a été étudiée (Fig. 75b) et montre un signal de fluorescence très faible par rapport aux régions cœur comportant deux sites de fixation intacts ce qui indique une coopérativité importante entre les intégrases pour la fixation aux sites cœur. En absence de sites bras, l'ajout de $mv4Int$ permet la formation du complexe I (piste 2, Fig. 75b). En présence des sites bras P'12 (pistes 3 à 7, Fig. 75b), l'augmentation de la quantité de $mv4Int$ fait apparaître les complexes II et III, suggérant la présence de complexes similaires à l'expérience réalisée avec un site natif.

Un profil de fixation de $mv4Int$ complètement différent est observé sur les sites cœur B (Fig. 75c) et C (Fig. 75d). En effet, le signal de fluorescence obtenu est très faible et l'ajout de sites bras de fixation n'a aucun effet sur la fixation de $mv4Int$ aux sites cœur B (pistes 2 et 6, Fig. 75c) et C (pistes 2 et 6, Fig. 75d) suggérant que la liaison aux sites cœur B et C est trop faible pour pouvoir être stabilisée. Ces résultats pourraient suggérer un ordre de fixation des molécules de $mv4Int$ sur la région cœur, $mv4Int$ semble se fixer en premier au site cœur C' puis pourrait recruter un deuxième monomère au niveau du site cœur C de plus faible affinité. Pour la capture du site *attB*, l'ordre de fixation devrait être identique : $mv4Int$ se fixerait en premier sur le site cœur B' et permettrait le recrutement d'un deuxième monomère au site cœur B de plus faible affinité. L'affinité plus élevée de $mv4Int$ envers les sites B' et C' était un résultat attendu au vu de la conservation de séquence (Fig. 74). L'identité de séquence au niveau de la région B' et C' a été sélectionnée au cours de l'évolution et permet de restaurer un gène fonctionnel codant un ARNt^{SER} après intégration, ainsi il semblerait que $mv4Int$ ait évolué en s'adaptant à cette contrainte, afin de se fixer plus fortement sur cette région sur les sites *attB* et *attP*. Comme la plupart des bactériophages ciblant les ARNt possèdent l'extrémité 3' du gène dans leur site *attP*, les sites *attB* et *attP* partagent toujours un site cœur identique. Il serait donc intéressant de vérifier parmi les systèmes HP1 et L5 si comme pour le système mv4, le site conservé B'/C' est celui présentant la plus haute affinité envers l'intégrase.

VI) Caractérisation des sites bras de fixation de ^{mv4}Int par retard sur gel

Les résultats précédents indiquent que le site bras P2, site bras situé à une position atypique sur le site *attP* (Fig. 68), n'est pas un site bras de 11 pb mais est en réalité le site cœur C de 7 pb. Les études de 1999 et de 2014 se différencient au niveau de la séquence consensus des sites bras déterminée : « 5'-TTAACTAGA-3' » (Auvray *et al.*, 1999a) et « 5'-TRRYTRRWARR-3' » (Coddeville *et al.*, 2014), ainsi qu'au niveau de l'architecture et de l'orientation des sites bras de recombinaison (Fig. 76a). Dans le but de caractériser les sites bras et de les positionner précisément sur le site *attP*, la méthodologie des banques aléatoires a été utilisée. Deux banques aléatoires ont été construites dans la région comprenant le site bras P2 (Auvray *et al.*, 1999a) et le site bras P1 (Coddeville *et al.*, 2014), puis testée par recombinaison *in vitro* contre le site *attB_{WT}* (données non montrées). Après recombinaison, aucune contrainte n'a été observée et aucune séquence consensus n'a pu être déterminée. Deux possibilités peuvent expliquer ce résultat : soit ces régions ne sont pas des régions impliquées dans la réaction d'intégration, ainsi aucune contrainte n'est exercée sur les positions modifiées, soit il y a une très forte coopérativité de liaison aux sites bras et l'utilisation de banques aléatoires ne permet pas d'inhiber la fixation de ^{mv4}Int au site modifié. En 2014, la mutation du site P1 (équivalent à P2 défini par Auvray *et al.*, 1999a) diminuait l'activité de recombinaison intégrative d'environ 65% (Coddeville *et al.*, 2014), indiquant que la région ciblée par nos banques aléatoires est une région importante pour la recombinaison. Ainsi, l'hypothèse d'une forte coopérativité de liaison des intégrases aux sites bras est favorisée et est en accord avec l'architecture du site *attP* de 1999 (Fig. 76a), où une paire de site bras P1-P2 est observée (Auvray *et al.*, 1999a).

Comme la méthode de retard sur gel permet d'observer de façon indirecte la fixation aux sites bras P'12 par stabilisation de la fixation de ^{mv4}Int aux sites cœur, cette technique a été utilisée pour vérifier si les sites bras définis par Auvray *et al.* permettent la fixation de ^{mv4}Int. En 1999, deux régions sur le site *attP* présentent des sites bras : une paire de site en orientation directe (P1 et P2) sur le bras P et une région composée de trois sites bras, deux en orientation directe (P'1, P'2) et un en orientation inverse (P'3) sur le bras P' (Fig. 76a) (Auvray *et al.*, 1999a). Dans le but de caractériser la capacité de fixation de ^{mv4}Int à ces sites bras, plusieurs combinaisons de sites ont été utilisés (Fig. 76c). La séquence d'ADN comprenant les sites bras P1 et P2, ainsi que P'1, P'2 et P'3 a été synthétisée et hybridée en ADN double-brin.

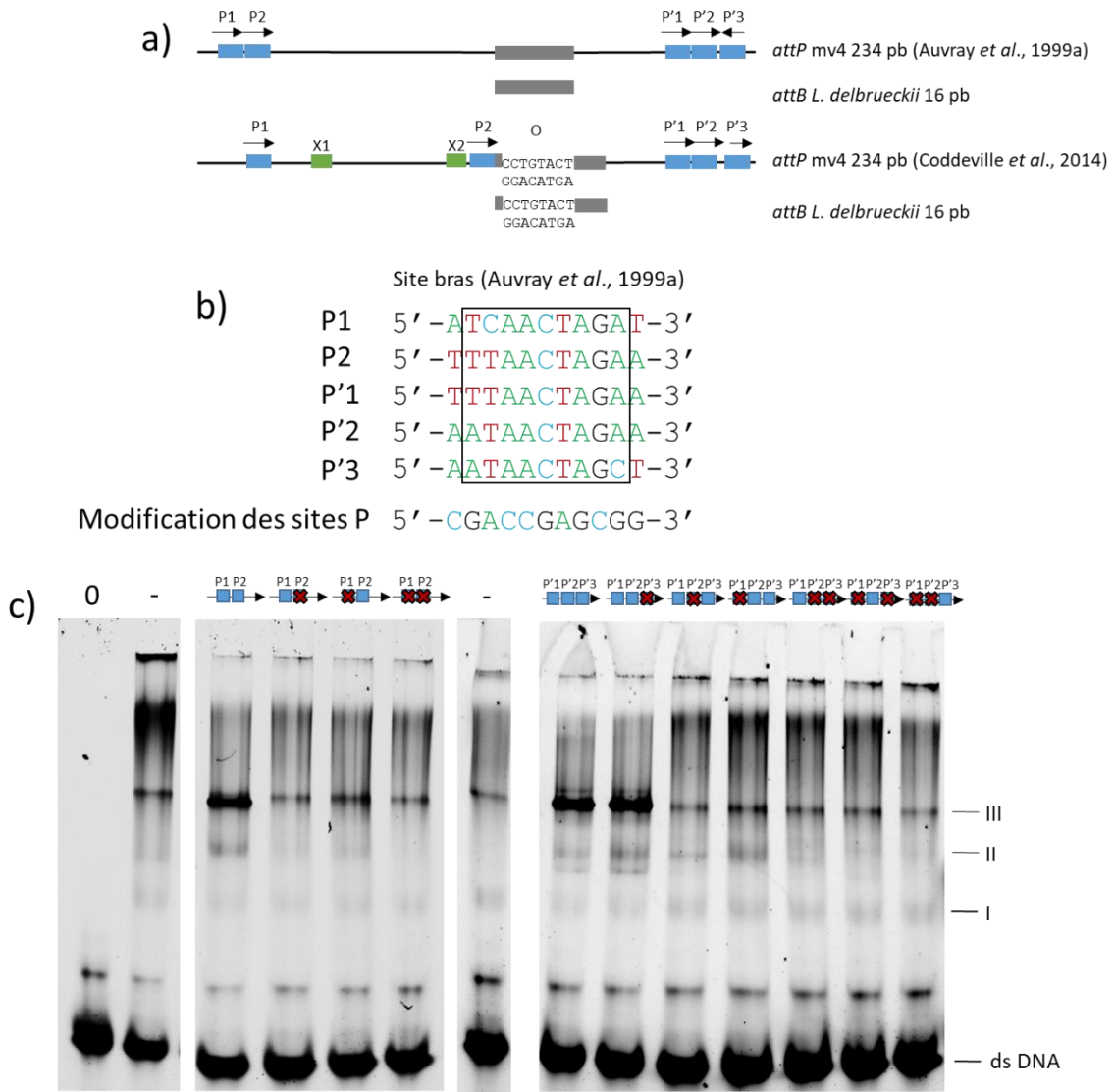


Fig. 76 Effet des différents sites bras sur la stabilisation du complexe $mv4Int-COC'$. a) Comparaison de l'architecture du site $attP_{WT}$ entre l'étude de 1999 et de 2014. Les sites bras de fixation de $mv4Int$ sont représentés en bleu et les flèches noires indiquent l'orientation des sites. Les sites de fixation de l'excisionase $mv4Xis$ sont représentés en vert. La lettre « O » indique la région centrale et les rectangles gris font partie de la région cœur (région d'identité entre $attB$ et $attP$). b) Alignement des séquences des sites bras P1, P2, P'1, P'2 et P'3 décrit en 1999 (Auvray *et al.*, 1999a). La séquence consensus de 9 pb définie par Auvray *et al.*, est encadrée en noir. c) Etude de la stabilisation de la liaison de $mv4Int$ aux sites cœur COC' par l'ajout de différentes combinaisons de sites bras. Un fragment fluorescent à la cyanine-3 (COC') (0,87 pmol) a été incubé en présence de 4,48 pmol de sites bras non-marqués (les différentes combinaisons sont indiquées au-dessus des pistes) et de 25,1 pmol de $mv4Int$. La piste 0 ne contient pas $mv4Int$. L'absence de site bras est indiquée par le signe « - » au-dessus des pistes. Les trois complexes sont indiqués sur le côté et sont identiques à ceux déterminés précédemment.

Les différents sites bras ont été inactivés en remplaçant la région de 11 pb contenant le site bras par une séquence composée de 11 nucléotides où chaque nucléotide est un nucléotide absent dans la séquence consensus (Fig. 76b). Les différentes combinaisons de sites ont ensuite été testées par retard sur gel pour leur capacité à stabiliser la fixation de mv^4Int à la région cœur COC' (Fig. 76c). L'ajout d'un fragment contenant la paire de sites bras P1 et P2 permet la stabilisation de la fixation de mv^4Int aux sites cœur COC' indiquant que mv^4Int est capable de se fixer sur les sites bras P1-P2 avec son domaine N-ter, démontrant expérimentalement pour la première fois la présence de ces sites dans le bras P comme décrit en 1999 (Auvray *et al.*, 1999a). En revanche, dès que le site P1 ou P2 est inactivé, il n'y a plus d'effet de stabilisation de la fixation de mv^4Int aux sites cœur et le signal observé est similaire à celui obtenu en absence de sites bras (piste -, Fig. 76c), suggérant une très forte coopérativité de liaison aux sites bras entre le dimère d'intégrases fixé aux sites cœur puisqu'en présence d'un seul site bras, le complexe dimère-région cœur ne peut être stabilisé. En étudiant le bras droit du site *attP* contenant les sites P'1, P'2 et P'3 (Fig. 76c), seules deux combinaisons stabilisent la fixation de mv^4Int à la région cœur COC' : le fragment composé des sites P'1, P'2 et P'3 ainsi que le fragment portant les sites P'1 et P'2. Comme précédemment, la présence d'un seul site bras ne permet pas la stabilisation du complexe nucléoprotéique. En revanche, l'ajout des deux sites bras P'2 et P'3 ou P'1 et P'3 n'a aucun effet sur la stabilisation de mv^4Int aux sites COC' indiquant que mv^4Int est incapable de se fixer sur le site P'3. Ainsi, il semblerait qu'il n'y ait que deux sites bras de fixation sur le bras P', la paire de sites P'1 et P'2.

En comparant la séquence des différents sites bras (Fig. 76b), peu de variations sont observées entre le site P'3 et les autres sites. Seul un changement est observé à la 10^{ème} position où un C est retrouvé à la place d'un A et il est donc peu probable que cette seule modification explique l'incapacité du site P'3 à fixer mv^4Int . En revanche, en regardant l'orientation des sites bras sur le site *attP*, seul le site P'3 est en orientation inverse (Fig. 76a), ainsi il est possible que cette orientation empêche la liaison coopérative avec le site P'2, expliquant ainsi l'incapacité du site P'3 à fixer mv^4Int . Afin de vérifier si l'orientation des sites bras pouvait jouer un rôle sur la capacité de fixation du domaine N-ter de mv^4Int , plusieurs orientations ont été étudiées : dans la paire de site P'1 et P'2, le site P'2 a été inversé et dans la paire de site P'2 et P'3, le site P'3 a été inversé (Fig. 77a). Ces constructions ont été testées pour leur capacité à stabiliser la fixation de mv^4Int aux sites cœur COC' (Fig. 77b). Dans leur orientation naturelle, les sites P'1 et P'2 permettent la fixation du domaine N-ter de mv^4Int .

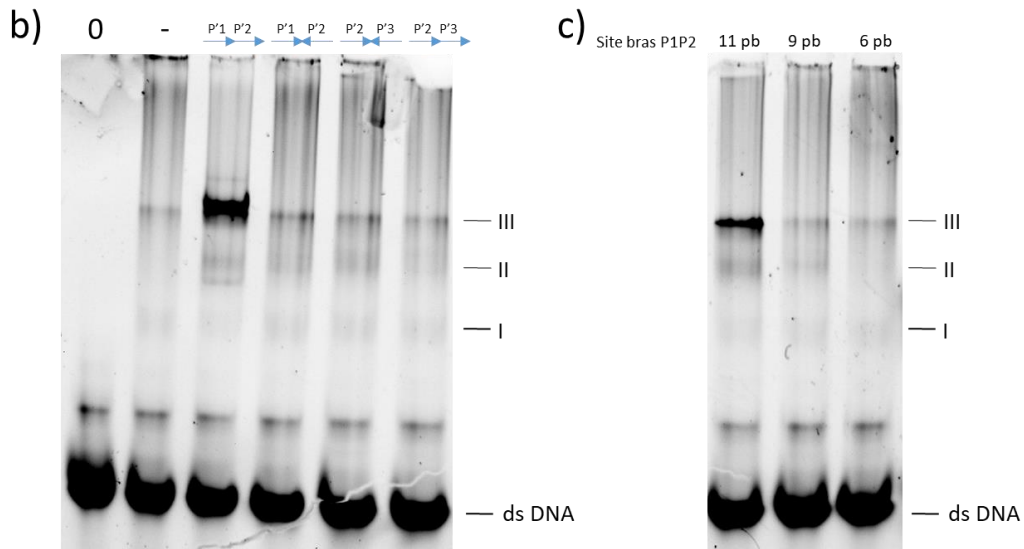
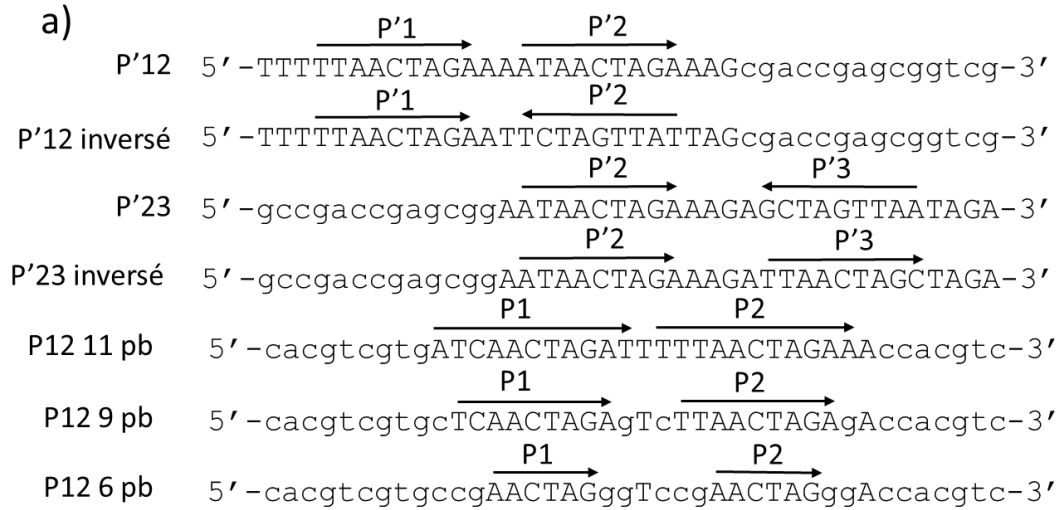


Fig. 77 Etude de l'orientation et de la taille des sites bras. a) Séquence des fragments des sites bras utilisés. Les bases en minuscule sont les nucléotides qui diffèrent par rapport au site *attP_{WT}*. b) Effet de l'orientation des sites bras sur la stabilisation de la liaison de *mv⁴Int* aux sites cœur COC'. Un fragment fluorescent à la cyanine-3 (COC') (0,87 pmol) a été incubé en présence de 4,48 pmol de sites bras non-marqués (les différentes orientations des sites bras sont indiquées au-dessus des pistes) et de 25,1 pmol de *mv⁴Int*. La piste 0 ne contient pas *mv⁴Int*. L'absence de site bras est indiquée par le signe « - » au-dessus des pistes. Les trois complexes sont indiqués sur le côté et sont identiques à ceux déterminés précédemment. c) Effet de la taille des sites bras sur la stabilisation de la liaison de *mv⁴Int* aux sites cœur COC'. La légende est identique à la partie b).

Par contre, l'inversion du site P'2 supprime l'effet stabilisateur suggérant que la liaison coopérative aux sites P'1 et P'2 a été abolie et n'est réalisable qu'en présence de deux sites en orientation directe. Au niveau de la paire de sites bras P'2 et P'3, l'inversion du site P'3 en orientation directe ne stabilise toujours pas la fixation de mv^4Int aux sites cœur, suggérant que la séquence P'3 décrite en 1999 n'est pas un site bras de fixation de mv^4Int . En accord avec ces résultats, la mutation du site P'3 ne semble pas avoir d'effet sur la réaction d'intégration ou d'excision (Coddeville *et al.*, 2014). Deux possibilités peuvent expliquer l'incapacité du fragment contenant les sites P'2 et P'3 en orientation directe à stabiliser la fixation à la région cœur. Tout d'abord, il est possible que la présence d'un A en 10^{ème} position (Fig. 76b) ait un rôle essentiel à la fixation de mv^4Int . Pour vérifier cette hypothèse, il suffirait de modifier les nucléotides A en 10^{ème} position des sites P'1 et P'2 par un C et de vérifier si ce fragment est toujours capable d'être fixé par mv^4Int . En parallèle, sur le fragment contenant les sites P'2 et P'3 en orientation directe, il faudrait modifier le nucléotide C en 10^{ème} position de P'3 par un A, et vérifier la capacité de ce fragment à fixer mv^4Int . Une deuxième explication pourrait provenir de l'espacement entre les sites bras. En effet, en considérant les sites bras comme des sites de 11 pb, dans la paire de sites P1-P2, les sites sont espacés d'un nucléotide. Dans la paire de sites P'1-P'2, les sites sont adjacents tandis que dans la construction où P'2 et P'3 sont en orientation directe, ces sites sont espacés par 3 nucléotides. Ainsi, il est possible que cette distance de 2-3 nucléotides entre les sites bras soit suffisante pour empêcher la liaison coopérative aux sites P'2 et P'3 en orientation directe. Pour valider cette hypothèse, il faudrait donc réduire l'espacement à 0 ou 1 nucléotide entre les sites P'2 et P'3, augmenter la distance entre les sites P'1 et P'2 à 3 nucléotides et tester ces deux fragments pour leur capacité à fixer mv^4Int . Quoi qu'il en soit, la position du site P'3 sur le site *attP* ne permet pas la fixation de mv^4Int et le site *attP* est donc composé de deux paires de sites bras, la paire de sites P1-P2 sur le bras P et P'1-P'2 sur le bras P'.

Nous avons aussi profité de cette stratégie pour tenter de déterminer la taille minimale de la séquence des sites bras de fixation. Pour cela, la paire de site bras P1-P2 a été raccourcie en modifiant certains nucléotides par un nucléotide absent dans la séquence consensus (Fig. 77a). Trois tailles ont été testées : 11 pb (Fig. 77a), 9 pb (taille consensus définie par Auvray *et al.*, 1999a) (Fig. 77a) et 6 pb (Fig. 77a). Après migration, seule la taille de 11 pb permet la stabilisation du complexe nucléoprotéique (Fig. 77c) indiquant que la séquence consensus déterminée en 1999 est en réalité plus grande de 1 pb aux deux extrémités.

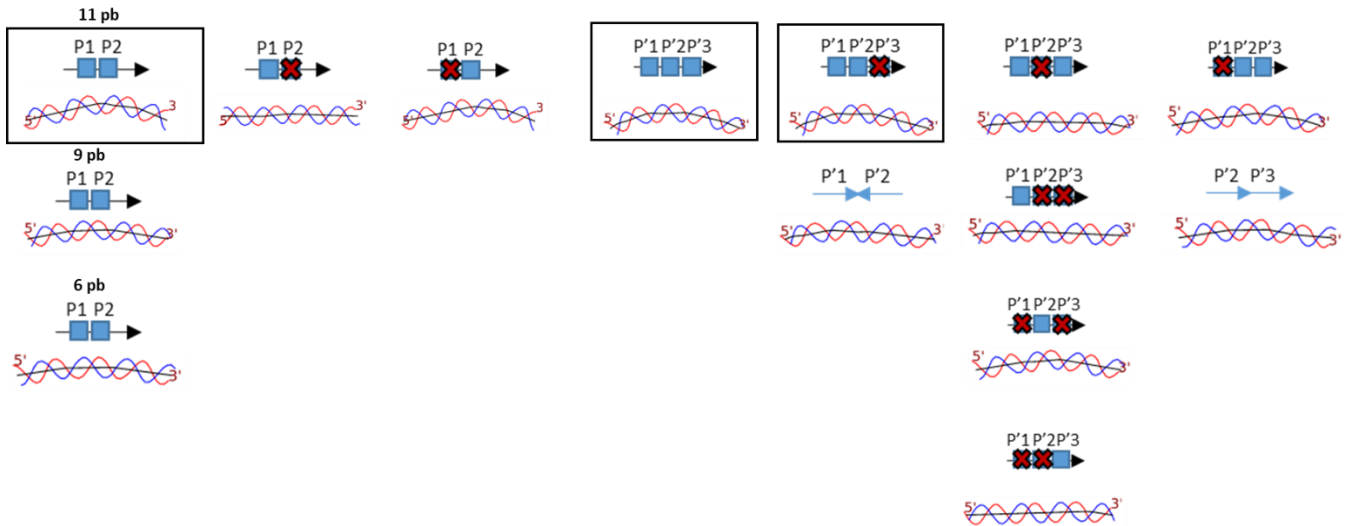


Fig. 78 Structuration théorique des différents fragments utilisés en retard sur gel. La conformation des différentes amorces utilisées dans l'étude a été modélisée grâce au programme « DNA Curvature Analysis » (C. Gohlke, <https://www.lfd.uci.edu/~gohlke/dnacurve/>). Les séquences permettant la fixation du domaine N-ter de $mv4Int$ sont encadrées en noir.

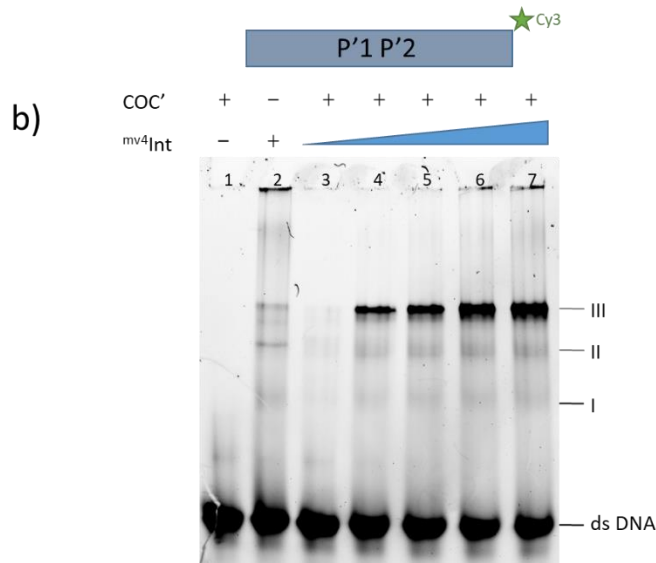
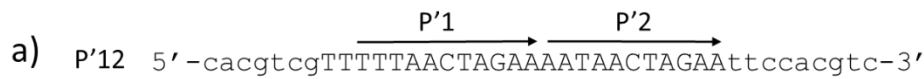


Fig. 79 Etude de la fixation de $mv4Int$ sur la paire de sites bras P'1 et P'2. a) Séquence du fragment fluorescent permettant l'étude de la fixation aux sites bras P'12. b) Etude en retard sur gel du fragment P'12 en présence de $mv4Int$ et des sites cœur COC'. Un fragment fluorescent à la cyanine-3 (P'12) (0,87 pmol) a été incubé en présence de 4,48 pmol de sites cœur COC' non-marqués. La réaction contient 0 pmol (1), 3,6 pmol (3), 10,7 pmol (4), 18 pmol (5), 25,1 pmol (2 et 6) et 35,8 pmol (7) de $mv4Int$. Les trois complexes sont indiqués sur le côté et sont identiques à ceux déterminés précédemment.

La séquence des sites bras (Fig. 76b) montre une proportion importante en nucléotides A et T. Ainsi, du fait de la capacité de courbure intrinsèque des séquences riches en A-T (Nelson *et al.*, 1987), la structuration de l'ADN pourrait avoir un rôle important sur la fixation de mv4 Int aux sites bras. La structure des différents fragments utilisés dans cette étude a été modélisée (Fig. 78) et indique que les fragments permettant la fixation du domaine N-ter de mv4 Int présentent tous une courbure assez importante au milieu du fragment (Fig. 78). La majorité des fragments ne permettant pas la fixation de mv4 Int ne présentent pas ou peu de courbure, à l'exception des fragments P1_{inactivé}-P2, P'1_{inactivé}-P'2-P'3 et P'1_{inactivé}-P'2-P'3_{inactivé}. Il est donc possible que la conformation de l'ADN ait un rôle dans la capacité de fixation du domaine N-ter de mv4 Int mais comme certains fragments non fonctionnels présentent une courbure équivalente, il semble plus probable que la séquence nucléotidique des sites bras ait un rôle majeur dans la fixation du domaine N-ter.

Enfin, dans le but de caractériser la capacité de liaison du domaine N-ter de mv4 Int, une amorce fluorescente comportant la paire de sites bras P'1-P'2 a été synthétisée (Fig. 79a) et testée pour sa capacité à fixer mv4 Int (Fig. 79b). En absence de sites cœur, la fixation de mv4 Int aux sites bras est instable (piste 2, Fig. 79b) et l'ajout de sites cœur stabilise fortement la fixation aux sites bras (pistes 3 à 7, Fig. 79b). Ainsi, mv4 Int semble exercer une inhibition réciproque entre ces deux domaines : en absence de fixation aux sites bras, la fixation aux sites cœur est faible et peu stable et en absence de fixation aux sites cœur, la fixation aux sites bras est aussi faible et peu stable. En revanche, dès que les deux domaines sont fixés, il semblerait y avoir un réarrangement protéique qui stabilise fortement le complexe nucléoprotéique. Une explication à l'inhibition réciproque des domaines N-ter et C-ter peut provenir de la structure de la protéine. En effet, il est possible que les domaines de fixation N-ter et C-ter soient proches et s'obstruent et que la fixation d'un domaine à l'ADN expose et active le second domaine pour la fixation. Chez $^{\lambda}$ Int, le domaine N-ter inhibe la fixation du domaine C-ter mais la présence ou l'absence du domaine C-ter n'a aucun effet sur la fixation du domaine N-ter (Sarkar *et al.*, 2001). A contrario, IntDOT, l'intégrase de l'ICE CTnDOT, est capable de se fixer à ses sites cœur en présence du domaine N-ter, indiquant que ce domaine n'a aucun impact sur le domaine C-ter. En revanche, en absence de fixation du domaine C-ter, IntDOT est incapable de se fixer sur ses sites bras (Wood *et al.*, 2010). Ainsi, bien que les intégrases hétérobivalentes aient conservé deux domaines de fixation de l'ADN différents ainsi qu'une structuration complexe de leur site *attP*, la régulation de chaque domaine semble différente en fonction du système étudié.

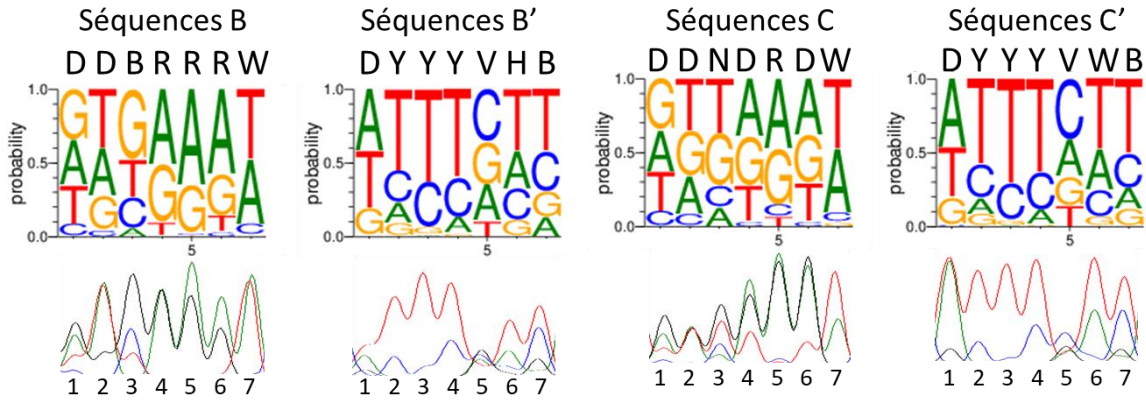


Fig. 80 Analyse des séquences B, B' C et C' après séquençage NGS. La proportion des différents nucléotides à chaque position a été représentée par « *Séquence Logo* ». Le motif de recombinaison obtenu par séquençage Sanger est représenté sous la représentation « *Séquence Logo* ». Sur le chromatogramme, chaque couleur représente un nucléotide : bleu, C ; noir, G ; rouge, T ; vert, A. La proportion de chaque nucléotide à chaque position dans le « *Séquence Logo* » est représentée par sa taille.

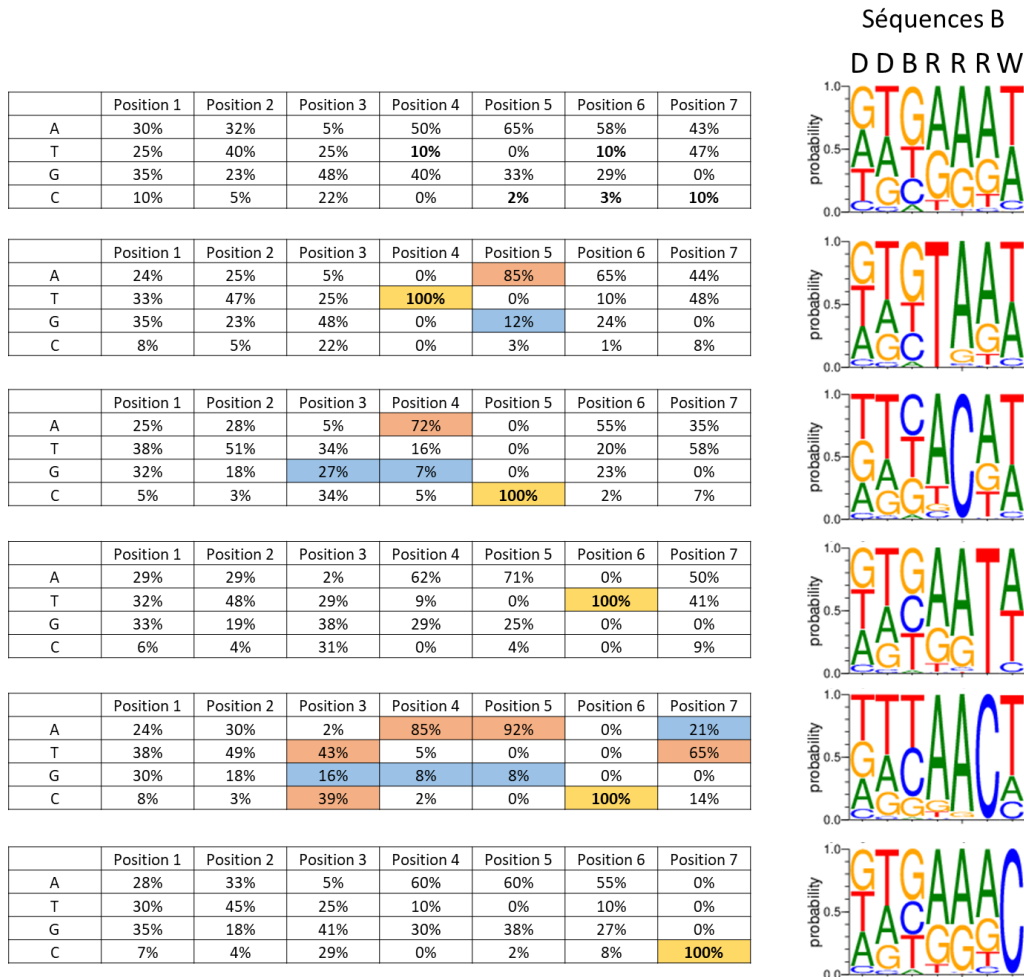


Fig. 81 Effet de la présence d'une pyrimidine dans le triplet de purines et d'un C en position 7 sur le site B. Le premier tableau et « *Séquence Logo* » représente la fréquence des 4 nucléotides en fonction de la position dans la séquence B et proviennent de l'analyse de $1,3 \cdot 10^6$ séquences. Les tableaux et « *Séquence Logo* » suivants décrivent l'effet d'un nucléotide précis à une position sur la fréquence des nucléotides dans les 6 autres positions. La couleur jaune indique la position et le nucléotide fixés pour l'analyse. Les 5 positions fixées sont comparées au motif de recombinaison issue de l'analyse des séquences totales de la banque. La couleur rose indique un effet positif sur la fréquence du nucléotide et la couleur bleue indique un effet négatif sur la fréquence d'un nucléotide.

VII) Caractérisation des motifs B, B', C et C' par séquençage NGS

L'analyse des banques aléatoires par séquençage Sanger a permis l'identification des nucléotides permissifs sur les différents sites B, B', C et C', ainsi que la définition d'une séquence consensus pour les 2 régions cœur. Cependant, cette analyse n'étant pas quantitative, elle ne permet pas de déterminer la fréquence de chaque nucléotide aux différentes positions dégénérées, ni d'identifier d'éventuels motifs nucléotidiques particuliers dans ces sites. Le séquençage NGS (technologie Illumina) des sites signatures de la recombinaison permet d'obtenir des informations plus détaillées sur les contraintes s'exerçant aux sites B, B', C et C'. Les sites *attL* obtenus après recombinaison *attP_{WT}* x *attB* Lib6 (site B) et *attB_{WT}* x *attP* Lib1 (site C'), et les sites *attR* des réactions de recombinaison *attP_{WT}* x *attB* Lib8 (site B') et *attB_{WT}* x *attP* Lib3 (site C) ont donc été séquencées après amplification PCR.

Nous avons tout d'abord commencé l'analyse de ces données NGS en réalisant un « *Séquence Logo* » (Schneider and Stephens, 1990) représentant la proportion de chaque nucléotides aux différentes positions des banques aléatoires (Fig. 80). Par comparaison aux chromatogrammes obtenus par séquençage Sanger, nous pouvons observer que les « *Séquence logo* » générés pour chaque site sont très similaires entre les deux méthodes, à la fois en termes de nature des nucléotides identifiés et de proportionnalité entre signal du chromatogramme et fréquence de chaque base dans le « *Séquence Logo* » (Fig. 80). En revanche, à certaines positions où seulement 2 à 3 pics étaient présents par séquençage Sanger, les 4 nucléotides peuvent parfois être retrouvés par séquençage NGS (par exemple, positions 2, 3 et 6 du site B, Fig. 80). Ceci montre d'une part la plus grande résolution du séquençage NGS et d'autre part que certains nucléotides qui étaient décrits comme non-permissifs dans le motif de recombinaison, peuvent être identifiés, donc impliqués dans un événement de recombinaison. Ce résultat pourrait s'expliquer par l'existence d'un phénomène de compensation entre nucléotides adjacents, un nucléotide non-permissifs à une position précise pouvant être toléré par la présence de nucléotides particuliers à son voisinage. Ne disposant pas des outils de bioinformatique permettant une analyse exhaustive des biais de composition liés à la présence de « sous-motif », tels que les biais en di-, tri- voir tétra-nucléotides, j'ai décidé de réaliser une analyse centrée sur les conséquences de la présence de certains nucléotides non-permissifs sur la composition des différentes positions du site cœur.

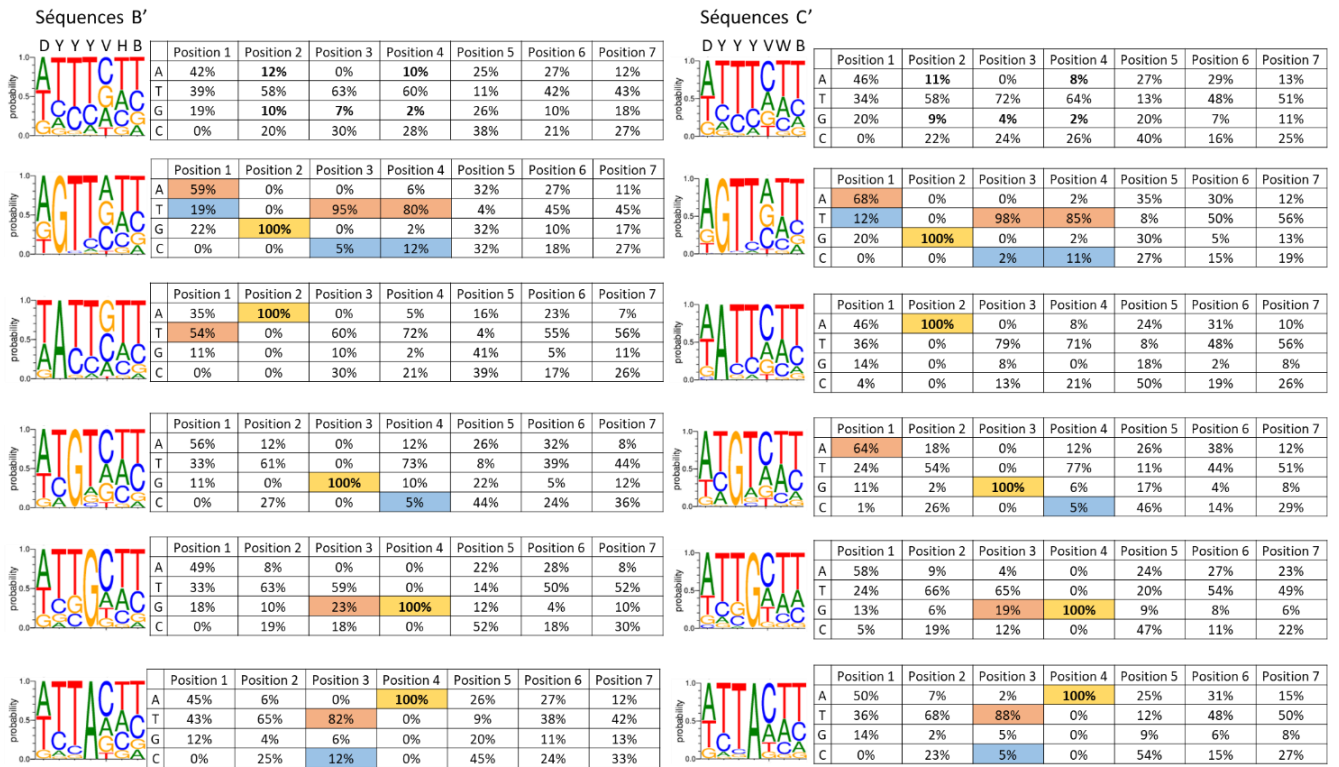


Fig. 82 Effet de la présence d'une purine dans le triplet de pyrimidines sur les sites B' et C'. Le premier tableau et « Séquence Logo » représente la fréquence des 4 nucléotides en fonction de la position dans la séquence B' (analyse de $1,8 \cdot 10^6$ séquences) et dans la séquence C' (analyse de $0,93 \cdot 10^6$ séquences). Les tableaux et « Séquence Logo » suivants décrivent l'effet d'une purine précise à une position sur la fréquence des nucléotides dans les 6 autres positions. La couleur jaune indique la position et le nucléotide fixés pour l'analyse. Les 5 positions fixées sont comparées au motif de recombinaison issue de l'analyse des séquences totales de la banque. La couleur rose indique un effet positif sur la fréquence du nucléotide et la couleur bleue indique un effet négatif sur la fréquence d'un nucléotide.

Etant donné que les contraintes les plus fortes semblent s'exercer aux positions proches de la région centrale (triplet de purines en position 4, 5, 6 et A/T en position 7) pour le site B, des « *Séquence Logo* » ont été générés en fixant un T en position 4, un C en position 5, un T ou un C en position 6, et un C en position 7 (Fig. 81). Tout d'abord, la fréquence des nucléotides au niveau des deux premières positions n'est pas modifiée quel que soit le nucléotide non-permissif fixé, indiquant que ces positions sont trop éloignées pour avoir un effet compensatoire. En présence d'un T en 4^{ème} position, le nucléotide A est favorisé en 5^{ème} position, au détriment d'un G. De façon similaire, la présence d'un C en 5^{ème} position favorise fortement la présence d'un A en 4^{ème} position au détriment d'un G, mais semble aussi diminuer la probabilité d'avoir un G en 3^{ème} position. Au contraire, la présence d'un C en 6^{ème} position change fortement la proportion en nucléotides des positions adjacentes, mais aussi des positions un peu plus éloignées (positions 3 et 4). Enfin, il semblerait que la présence d'un T en position 6, ou d'un C en position 7 n'ait pas d'impact majeur sur la fréquence nucléotidique des 6 autres positions. En résumé, la présence d'un nucléotide non-permissif dans le site cœur B semble modifier la fréquence nucléotidique des positions voisines, mais n'impacte pas les positions les plus éloignées, indiquant que la présence de certains nucléotides adjacents précis permet à l'intégrase de tolérer un nucléotide non-permissif.

Au niveau des sites B' et C', les contraintes les plus importantes semblent s'exercer au niveau du triplet de pyrimidines (aux positions 2, 3 et 4). L'effet de la présence d'une purine dans le triplet de pyrimidines a donc été étudié en réalisant des « *Séquence Logo* » en fixant un G ou un A en position 2, un G en position 3, et un G ou un A en position 4 (Fig. 82). Dans un premier temps, nous pouvons remarquer que les motifs B' et C' réagissent globalement de la même manière à l'ajout d'une purine à une position précise, renforçant les analyses précédentes où une séquence consensus dégénérée très proche étaient obtenus pour les deux sites. De la même manière que le site B, les positions les plus éloignées de la région centrale (positions 5, 6 et 7) sont insensibles à la présence d'un nucléotide non-permissif, suggérant que ces positions sont trop éloignées pour avoir un quelconque effet. La présence d'un G en position 2 favorise la présence d'un A en position 1, au détriment d'un T, ainsi que la présence d'un T aux positions 3 et 4, en défaveur du C. A la position 4, la présence d'un G augmente la probabilité d'avoir le nucléotide G à la position 3, alors que la présence d'un A à cette même position favorise fortement la présence d'un T au détriment d'un C. En résumé, il semblerait donc que la présence d'une purine à une position précise dans le triplet de pyrimidines soit possible grâce à la présence de nucléotides précis au voisinage.

Chapitre 2

Table 1. Synthèse des données de séquençage NGS. L'occurrence théorique moyenne a été calculée en divisant le nombre de séquences totales par le nombre de mots différents séquencés.

Site cœur	Echantillon séquencé	Nombre de séquences totales	Nombre de mots différents	% mots présents (16384 mots théorique)	Occurrence moyenne théorique de chaque mot	Mot avec une occurrence de 1-10	10-100	100-1000	>1000	Occurrence maximale
B	<i>attL</i> Lib6	1,3.10 ⁶	5459	33,30%	238	1862	2105	1152	340	14405
B'	<i>attR</i> Lib8	1,8.10 ⁶	9497	58%	190	4734	2625	1585	553	49310
C	<i>attR</i> Lib3	1,4.10 ⁶	10416	63,60%	134	4013	3423	2724	256	2139
C'	<i>attL</i> Lib1	0,93.10 ⁶	7549	46,10%	123	4115	2095	1112	227	16157

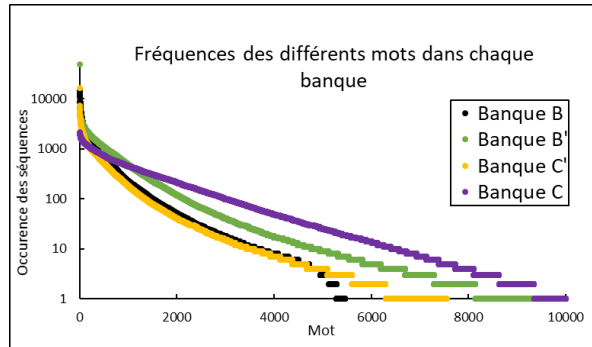


Fig. 83 Représentation de l'occurrence des différents mots dans les banques B, B', C et C'. L'échelle des ordonnées est une échelle logarithmique en base 10. Pour chaque banque, chaque mot est associé à un numéro. Une décroissance exponentielle plus ou moins importante de l'occurrence est observée pour les différentes banques.

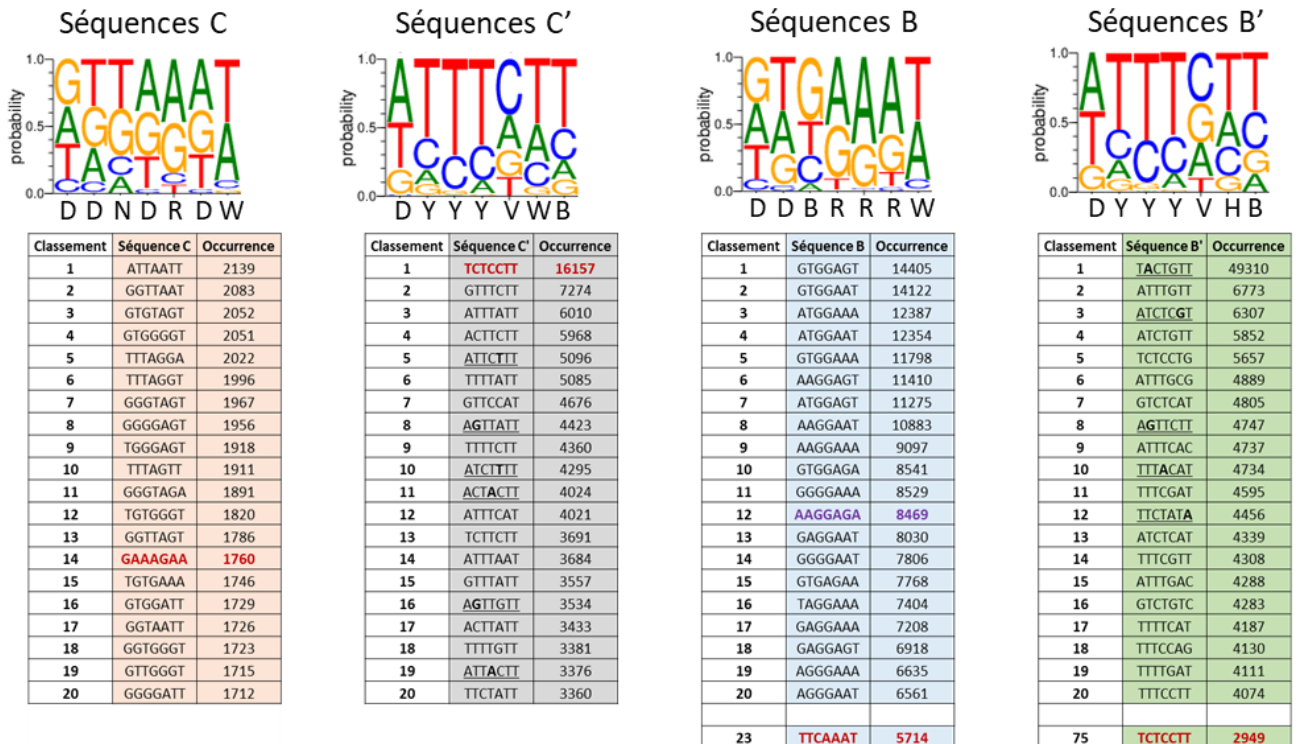


Fig. 84 Représentation des 20 séquences les plus présentes dans les différentes banques après recombinaison. Pour chaque groupe, la séquence native est indiquée en rouge. Les séquences soulignées sont les séquences possédant un nucléotide non-permissif qui est représenté en gras. Dans le groupe B, la séquence violette indique la présence de la séquence B' complétée inversée (« 5-TCTCCTT-3' »).

En conclusion, l'analyse par séquençage NGS confirme la présence des différentes contraintes déterminées précédemment par séquençage Sanger, mais en raison de la plus grande résolution de cette analyse, nos données indiquent que certains nucléotides décrits comme non-permissifs, peuvent en réalité être impliqués dans un événement de recombinaison en présence de « sous-motif » particulier adjacent.

L'analyse de données NGS permettant d'obtenir la séquence individuelle de chaque ADN impliqué dans la réaction de recombinaison, il est donc possible d'identifier et de quantifier chaque motif des sites B, B', C et C' présents dans les sites hybrides *attL* et *attR* (Table 1). Après caractérisation des différents motifs impliqués dans la recombinaison, une forte hétérogénéité est observée dans la représentativité des séquences de chaque banque (Table 1 ; Fig. 83). En effet, parmi le nombre de séquences permissives à la recombinaison, seul un faible pourcentage (2 à 6% séquences) présente une occurrence significativement notable (supérieure à 1000). Par exemple, en fonction des banques, les 100 motifs les plus fréquents représentent 10 à 35% de la totalité des séquences obtenues, indiquant que parmi les milliers de séquences fonctionnelles, seul un faible pourcentage permet une recombinaison à une activité élevée.

Pour chaque site cœur, les 20 séquences les plus présentes ont été extraites et classées selon leur occurrence et les séquences natives de chaque site ont été recherchées (séquences rouges, Fig. 84). Au niveau de la région cœur du site *attP*, la séquence native C' est retrouvée en première position. Dans la partie précédente, il avait été montré que cette séquence était la séquence présentant la plus forte affinité parmi les sites cœur B et C et le fait de la retrouver en première position semble conforter ce résultat. Au niveau du site cœur C, l'occurrence de la séquence native ne se trouve qu'à la 14^{ème} position (Fig. 84), indiquant que d'autres motifs pourrait être plus adaptés lors de la recombinaison *in vitro*. Cependant, il est à remarquer que la valeur maximale de l'occurrence du site le plus représenté n'est que 20% supérieure à celle du site natif, suggérant une plasticité dans la capacité de fixation de l'intégrase sur ce site. Il serait intéressant d'analyser le comportement de ces motifs plus fréquents vis-à-vis de l'affinité de fixation de l'intégrase. Coddeville *et al.*, avaient montré l'importance du site bras atypique P2 pour les réactions d'intégration et d'excision en réalisant des modifications ponctuelles de ce site inactivant la recombinaison (Coddeville *et al.*, 2014). Nous avons démontré que le site P2 est en réalité le site cœur C, et en prenant en compte les modifications réalisées par Coddeville *et al.*, la séquence C obtenue 5'-GACCGTA-3' est retrouvée à une occurrence très faible, de 9, en accord avec l'absence de signal de recombinaison obtenu dans leur étude (Coddeville *et al.*, 2014).

En ce qui concerne le site *attB*, les séquences natives B et B' ne semblent pas les plus optimales en recombinaison *in vitro* puisqu'on les retrouve à la 23^{ème} position pour le site B et à la 75^{ème} position pour le site B' (Fig. 84). Au niveau des séquences B, la séquence complétée inversée B' (séquence violette) est retrouvée à la 12^{ème} position, avec une occurrence 50% supérieure à celle de la séquence native B, suggérant qu'un site *attB* symétrique de type B'OB' serait plus adaptée pour une recombinaison *in vitro* que le substrat natif BOB'.

Au niveau de la banque B', la séquence à la première position présente une fréquence 16 fois plus élevée que la séquence B' native et il est possible qu'un biais expérimental soit responsable de cette occurrence anormalement élevée. Deux types de biais peuvent être envisagés, soit un biais lors de la synthèse et de la construction de la banque B', soit un biais après recombinaison, lors de l'amplification par PCR des sites hybrides. Il faudra donc vérifier l'absence d'artéfact expérimental avant de tester ses capacités à fixer l'intégrase par des expériences de retard sur gel. Un résultat surprenant concerne la fréquence relativement faible du motif B' car ce motif est le plus fréquent lorsqu'il est localisé sur le site *attP* (site C'). Une hypothèse pourrait être que malgré l'identité de séquence, le rôle des séquences B' et C' soit différent lors de la recombinaison. En effet, il se pourrait que pour la mise en place de l'intasome sur le site *attP*, un site précis, de forte affinité, comme le site C' soit requis. Une fois assemblé, il est possible que pour la capture du site *attB*, ^{mv4}Int présente une plus grande flexibilité envers les sites de fixation B et B', d'autres séquences pouvant alors être autant, voire plus, fonctionnelles que les séquences natives. En résumé, parmi les 4 sites cœur natifs, seule la séquence C' semble être la séquence la plus optimale dans le système de recombinaison mv4, ce qui laisse suggérer un rôle essentiel de ce site dans la recombinaison, notamment dans la mise en place de l'intasome sur le site *attP*.

De façon intéressante, parmi les sites B et C, les 20 premières séquences respectent strictement le motif consensus dégénérée obtenu par séquençage Sanger. Au niveau des sites B' et C', entre 5 et 6 séquences parmi les 20 premières présentent un nucléotide non permissif. En revanche, aucune séquence ne présente plus d'un nucléotide non-permissif, suggérant que les séquences les plus fonctionnelles respectent les contraintes déterminées précédemment.

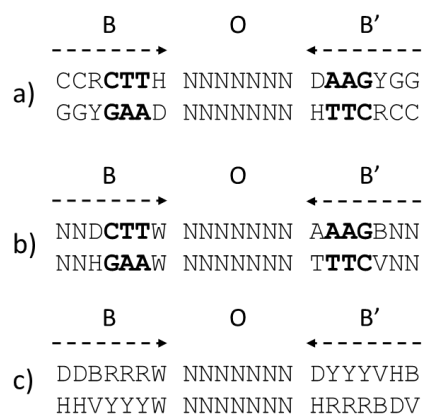


Fig. 85 Séquences consensus des sites *attB* des bactériophages λ , HK022 et mv4. Le triplet de nucléotide conservé « 5'-CTT-3' » est indiqué en gras. **a)** Séquence consensus du site *attB* de λ Int provenant de l'étude de Rutkai *et al.*, (Rutkai *et al.*, 2003). **b)** Séquence consensus du site *attB* ^{HK022}Int provenant de l'étude de Kolot *et al.*, (Kolot *et al.*, 2015). **c)** Séquence consensus du site *attB* de ^{mv4}Int.

Parmi les séquences présentant un nucléotide non-permissif, celles présentant le nucléotide G en 2^{ème} position (séquences 8 pour B', 8 et 16 pour C', Fig. 84) possèdent un A en 1^{ère} position et un T en 3^{ème} et 4^{ème} position, des nucléotides très fortement favorisé lorsqu'un G se situe en 2^{ème} position (Fig. 82), indiquant qu'une séquence possédant le sous-motif 5'-AGTT-3' ne respectant pas la séquence consensus à la 2^{ème} position, peut tout de même être une séquence hautement fonctionnel pour la recombinaison *in vitro*. De leur côté, les séquences présentant le nucléotide A en 4^{ème} position (séquences 10 pour B', 11 et 19 pour C', Fig. 84) possèdent toutes un T en 3^{ème} position, un nucléotide fortement favorisé lorsqu'un A se trouve en 4^{ème} position (Fig. 82), et la séquence possédant le nucléotide A en 2^{ème} position (séquence 1, B', Fig. 84), possèdent un T en première position, un nucléotide favorisé par la présence d'un A en 2^{ème} position.

En résumé, la grande majorité des séquences hautement fonctionnelles respectent les contraintes caractérisées par séquençage Sanger. Cependant, l'analyse plus précise par séquençage NGS a permis de montrer que certains sous-motifs particuliers pouvaient permettre à l'intégrase de tolérer un nucléotide décrit comme non-permissif. Parmi les intégrases hétérobivalentes, peu d'études se sont focalisées sur les modifications du site *attB* tolérées par l'intégrase. En effet, chez le modèle λ Int, alors qu'il a été montré que la région centrale était modifiable tant que l'identité au niveau de cette séquence était conservée entre les sites *attB* et *attP* (Bauer *et al.*, 1985; Weisberg *et al.*, 1983), peu d'information sur la capacité de modification des sites B et B' sont disponibles. Deux études se sont cependant focalisées sur les sites secondaires d'intégration du bactériophage λ , et ont montré que malgré les modifications des sites B et B', une très grande majorité des sites contient la séquence inversée répétée « 5'-CTT-3' » (Rutkai *et al.*, 2003) (Fig. 85a), et que de nombreuses bases du site *attB*_{WT} sont conservées (Tanouchi and Covert, 2017). Cela montre que λ Int est capable de se fixer sur d'autres motifs que les sites B et B' naturels, mais avec une flexibilité réduite.

L'intégrase HK022, une intégrase proche de λ Int a aussi été caractérisée au niveau de ces sites secondaires d'intégration. En effet, chez sa bactérie hôte *E. coli*, cette protéine nécessite le facteur d'hôte IHF or, chez une cellule humaine, cette protéine est fonctionnelle en absence d'IHF (Kolot *et al.*, 2003) suscitant ainsi un intérêt pour la recherche de sites *attB* secondaires dans les cellules humaines (Elias *et al.*, 2020; Kolot *et al.*, 2015).

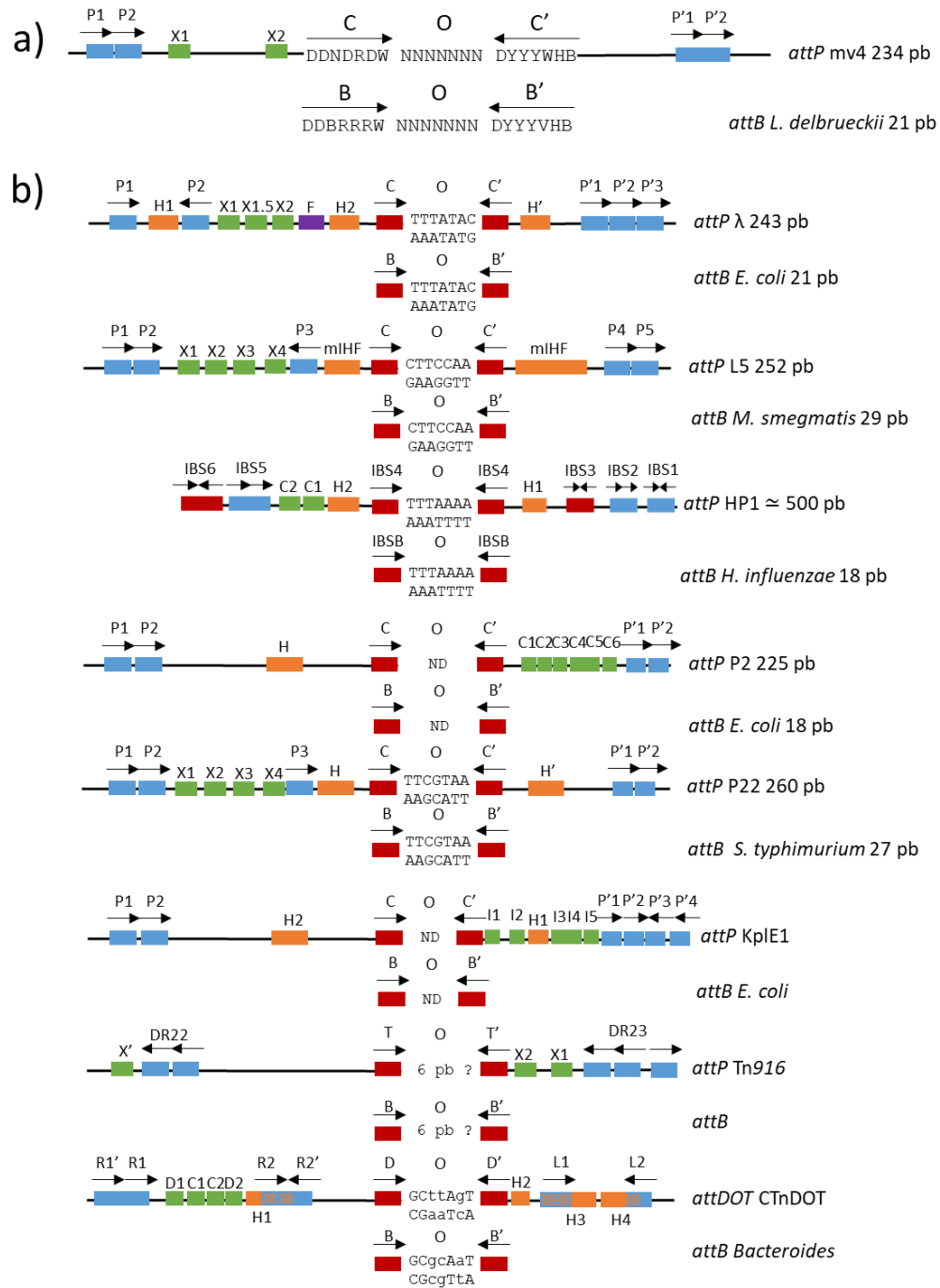


Fig. 86 Représentation des sites *attP* et *attB* du système mv4 et comparaison avec les sites des autres intégrases hétérobivalentes. a) Représentation de la nouvelle architecture des sites *attP*_{mv4} et *attB*_{mv4}. La région centrale est indiquée par la lettre « O ». Les sites bras de fixation sont représentés par des rectangles bleu et la flèche noire indique l'orientation du site. Les rectangles vert indiquent les sites de fixation des excisionases. b) Comparaison des sites de recombinaison des systèmes λ, L5, HP1, P2, P22, KplE1 Tn916, CTnDOT. Les échelles de taille ne sont pas respectées. Les sites cœur de fixation de l'intégrase sont représentés par des rectangles rouge. La flèche noire indique l'orientation du site. Les rectangles orange indiquent les sites de fixation des protéines accessoires ayant un rôle architectural dans la recombinaison (IHF pour λ, HP1, P2, P22 et KplE1, mIHF pour L5, BHFa « *Bacteroides* Host Factor » pour CTnDOT, HU pour Tn916). Dans le cas de CTnDOT, les hachures indiquent le chevauchement d'un site de fixation de IHF, avec un site bras de fixation de l'intégrase. Le rectangle violet indique le site de fixation de la protéine Fis. Les séquences des régions centrales sont représentées, et les lettres en minuscules indiquent les positions où la séquence entre la région centrale des sites *attDOT* et *attB*_{*Bacteroides*} est différente.

Ces études ont testé la capacité de recombinaison de différents sites *attB* secondaires et ont permis la caractérisation d'une séquence consensus du site cible *attB* où la séquence inversée répétée « 5'-CTT-3' » doit aussi être conservée dans les sites cœur de fixation B et B' (Fig. 85b) (Kolot *et al.*, 2015). Ainsi, tout comme λ Int, ^{HK022}Int montre une faible flexibilité au niveau de la reconnaissance de ces sites cœur de fixation. A la différence des intégrases hétérobivalentes classiques où une séquence inversée répétée est nécessaire à la réaction, ^{mv4}Int montre une forte flexibilité au niveau de la reconnaissance de ces 4 sites cœur puisque ces sites sont composés de 3 séquences différentes. De plus, les données NGS montrent que parmi les 20 séquences B et B' les plus représentées, aucune ne présente un nucléotide natif invariable (Fig. 84 ; Fig. 85c) renforçant ainsi l'hypothèse d'une plus grande flexibilité vis-à-vis de la reconnaissance de ces sites de recombinaison que les systèmes classiques.

En conclusion générale, l'utilisation des banques aléatoires combinée à la méthodologie de retard sur gel ont permis la caractérisation précise des sites de recombinaison du système mv4 (Fig. 86a). Les différents résultats obtenus grâce aux banques aléatoires ont permis de définir un motif de recombinaison dans la région cœur des sites *attB* et *attP* et de déterminer les sites cœur de fixation de ^{mv4}Int B, B', C et C' (Fig. 86a). De plus, il a pu être montré que le site bras P2, situé à une position atypique dans l'étude de 2014 est en réalité le site cœur C. Ainsi, l'architecture du site *attP* est semblable à celle des différents systèmes hétérobivalents (Fig. 86b). En effet, le site *attP* présente désormais une paire de site bras en orientation directe sur les bras P et P' (Fig. 86a), comme les systèmes L5, P2 et P22 (Fig. 86b). Une autre caractéristique atypique du système mv4 a pu être clarifiée par l'utilisation des banques aléatoires. En effet, ces banques ont permis de montrer que la région centrale possède une taille de 7 pb, comme tous les systèmes hétérobivalents de bactériophages actuellement décrits. En revanche, l'utilisation des banques aléatoires a permis de distinguer le système ^{mv4}Int/*attP*/*attB* des autres systèmes hétérobivalents par une grande flexibilité vis-à-vis de la reconnaissance de ces sites de recombinaison. L'analyse des banques aléatoires par séquençage NGS renforce l'idée de cette forte flexibilité, et montre que certains nucléotides décrits comme non-permissifs peuvent être fonctionnelle en présence de « sous-motif » particulier. A notre connaissance, la région cœur du système de recombinaison mv4 est la région cœur la mieux caractérisée et il serait intéressant que la méthodologie des banques aléatoires soit utilisée dans d'autres systèmes de recombinaison hétérobivalents afin de déterminer si l'importante flexibilité de ^{mv4}Int vis-à-vis de la région cœur de ces sites de recombinaison est une propriété unique à cette protéine ou bien si elle est partagée avec d'autres intégrases.

Enfin, grâce à l'importante flexibilité de ^{mv4}Int, et au regard des séquences consensus dégénérées des régions cœur, il est envisageable de se demander si la réaction de recombinaison peut être reprogrammée vers un nouveau site cible de recombinaison (*attB*) respectant le motif « 5'-DDBRRRW NNNNNNN DYYYVHB-3' ». Le chapitre 3 a donc eu pour objectif de déterminer l'étendue de la flexibilité de ^{mv4}Int.

**Chapitre 3 : Etude de l'étendue de la
flexibilité de la recombinaison médiée
par ^{mv4}Int**

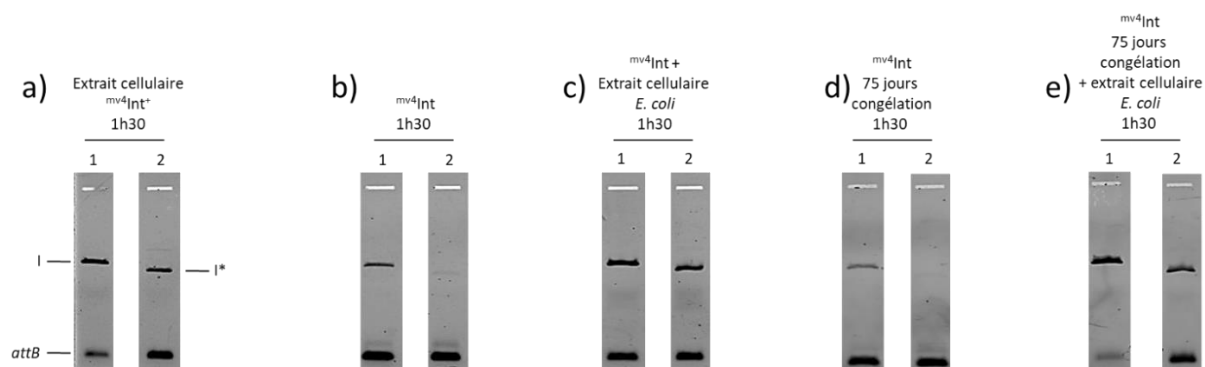


Fig. 87 Effet de l'ajout d'un extrait cellulaire sur la réaction de recombinaison. Les différentes conditions sont indiquées au-dessus du gel. Pistes : 1, recombinaison $attP_{WT} \times attB_{WT}$; 2, recombinaison $attP_{lactis} \times attB_{lactis}$. Le fragment $attB$ fluorescent et les produits de recombinaison entre les sites sauvages (I) ou les sites modifiés (I*) sont indiqués sur le gel. La réaction de recombinaison contient 7,2 pmol de $mv4Int$ ou 40 μg d'extrait cellulaire d'*E. coli*. **a)** Recombinaison à partir d'un extrait cellulaire d'*E. coli* BL21(DE3) après surexpression de $mv4Int$. **b)** Recombinaison avec l'intégrase $mv4Int$ purifiée. **c)** Effet de l'ajout d'un extrait cellulaire (*E. coli* BL21(DE3)) à la réaction de recombinaison **b)**. **d)** Recombinaison de la protéine purifiée $mv4Int$ après stockage à -80°C . **e)** Effet de l'ajout d'un extrait cellulaire (*E. coli* BL21(DE3)) à la réaction de recombinaison **d)**. Les gels des parties **a)** et **b)** sont issus de la figure 53.

Le chapitre 1 a mis en évidence l'importance d'un composé présent dans l'extrait cellulaire d'*E. coli* pour l'obtention d'une recombinaison *in vitro* efficace sur des variants des sites *attB* et *attP*. Ce chapitre relate l'identification de ce facteur accessoire. De plus, comme la région cœur des sites de recombinaison *attB* et *attP* a pu être précisément caractérisée, ce chapitre décrit l'ampleur des modifications nucléotidiques de ces régions cœur tolérée par ^{mv4}Int. Pour cela, l'objectif de ce travail fut de tester la possibilité de reprogrammation *in vitro* et *in vivo* de la réaction de recombinaison vers l'ARNt^{SER} d'autres espèces bactériennes, ainsi que vers d'autres cibles chromosomiques respectant les contraintes de séquences déterminées précédemment.

I) Caractérisation du facteur stimulant la réaction de recombinaison lors de l'utilisation de substrats modifiés

Dans le chapitre 1, une forte différence d'activité de recombinaison sur des substrats modifiés (*attP*_{lactis}, *attB*_{lactis}) était observé lors de l'utilisation d'un extrait cellulaire d'*E. coli* ayant surexprimé ^{mv4}Int (piste 2, Fig. 87a) par rapport à la protéine ^{mv4}Int purifiée (piste 2, Fig. 87b). Cette différence d'intensité, supprimée lorsque la réaction de recombinaison était réalisée pendant 16h, suggère l'existence de composés dans l'extrait d'*E. coli* pouvant avoir un rôle de « cofacteurs facultatifs ». Cette hypothèse a été vérifiée en rajoutant au mélange réactionnel contenant ^{mv4}Int purifiée, un extrait cellulaire soit d'*E. coli* BL21(DE3) (Fig. 87c), soit d'*E. coli* DH5α, ou de *Lactococcus lactis* (données non montrées). Après quantification de l'intensité de fluorescence, ces différents ajouts améliorent tous l'intensité du signal de recombinaison d'un facteur 3 pour les sites natifs, comme montré en 1999 (Auvray *et al.*, 1999a), et d'un facteur 25 à 30 pour une recombinaison entre sites modifiés. Un traitement thermique à 95°C de ces extraits cellulaires ne supprime pas cet effet, contrairement à un traitement à la protéinase K (données non montrées), indiquant que l'élément « stimulateur » correspond à une protéine thermorésistante partagée entre *E. coli* et *L. lactis*. De plus, cet effet stimulateur a également été observé sur des échantillons de la protéine purifiée stockés à -80°C pendant une longue durée (Fig. 87d ; Fig. 87e).

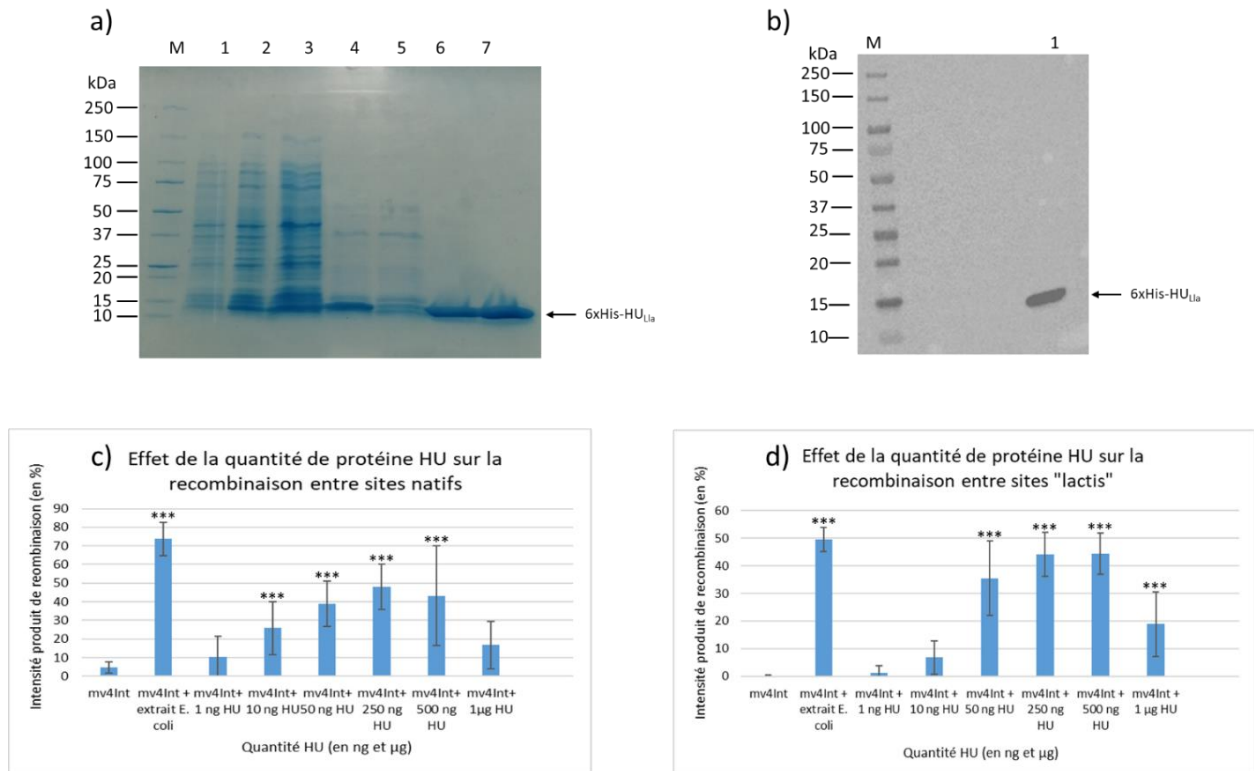


Fig. 88 Purification de la protéine 6xHis-HU_{Lla} et étude de son effet sur la réaction de recombinaison. **a)** Electrophorèse SDS-PAGE 4-15% d'extraits protéiques obtenus lors de la purification de 6xHisHU_{Lla}. Pistes : 1, extrait cellulaire *E. coli* BL21(DE3) non induit ; 2, extrait cellulaire induit (IPTG) ; 3, surnageant avant traitement thermique ; 4, surnageant après traitement thermique 70°C 30 min ; 5, fractions non retenues sur colonne Ni-NTA ; 6, fractions éluées de la colonne Ni-NTA ; 7, fractions éluées de la colonne héparine ; M, marqueur de taille. **b)** Western-blot de la protéine 6xHis-HU_{Lla}. 200 ng de la fraction purifiée 6xHisHU_{Lla} ont été séparé par SDS-PAGE, transféré sur membrane de nitrocellulose et détecté par chemiluminescence après incubation avec des anticorps anti-polyHistidine-Peroxydase (1 :2000). Puits : 1, 200 ng 6xHisHU_{Lla} ; M, marqueur de taille. **c)** Effet de quantités croissantes de 6xHis-HU_{Lla} sur la réaction de recombinaison entre les sites *attP*_{WT} et *attB*_{WT}. L'intensité de fluorescence a été quantifiée avec le logiciel Image Lab (Biorad). Les données présentées sont les valeurs moyennes et les écarts-types calculés à partir de trois mesures indépendantes. Les différences significatives par rapport à la condition *mv4Int* seule sont représentées par le symbole « *** » (Kruskal-Wallis, $p < 0,001$). **d)** Effet de la quantité de 6xHis-HU_{Lla} sur la réaction de recombinaison *attP*_{lactis} et *attB*_{lactis}. Les données présentées sont les valeurs moyennes et les écarts-types de quatre mesures indépendantes. La légende est identique à la partie **c)**.

Comme candidats potentiels, nous avons pensé à une protéine appartenant à la famille des « NAPs » (« Nucleoid-Associated Proteins »), une famille de protéine capable de se fixer et de courber l'ADN (Dillon and Dorman, 2010), car la très grande majorité des facteurs d'hôte des systèmes de recombinaison spécifiques de site appartiennent à cette famille. De plus, les protéines HU, IHF et Fis sont décrites comme plutôt thermorésistantes, comme le facteur recherché. Fis n'étant retrouvé que chez les entérobactéries et IHF chez les Proteobactéries (Kamashev *et al.*, 2017), le candidat le plus probable est la protéine HU, une protéique ubiquitaire où des orthologues sont retrouvés dans tous les phyla bactériens ainsi que chez les Archaea, bien qu'elle ne soit pas abondamment observées chez ces organismes (Johnson *et al.*, 2004). HU est impliquée dans la régulation de la transcription, dans la compaction, la formation de surenroulement négatif et la structuration du chromosome ainsi que dans la réplication, réparation et recombinaison de l'ADN.... (Johnson *et al.*, 2004).

Pour vérifier l'hypothèse de l'implication d'une histone-like dans la réaction de recombinaison, le gène *hllA* codant la protéine HU de *L. lactis* a été cloné en fusion traductionnelle derrière une étiquette 6xHis (pMET369, table 3). Nous avons choisi ce gène car chez *E. coli*, la protéine HU est majoritairement présente sous forme d'hétérodimères (HU- α et HU- β) (Rouvière-Yaniv and Kjeldgaard, 1979), alors que chez *L. lactis* et les firmicutes, HU est codée par un seul gène et est donc présente sous forme d'homodimères, ce qui facilite l'obtention de formes fonctionnelles de cette protéine *in vitro*. La protéine 6xHisHU_{Lla} a été surproduite dans une souche d'*E. coli* BL21(DE3) et, en raison de la thermorésistance de cette protéine, l'extrait cellulaire a été traité à 70°C (piste 4, Fig. 88a) afin d'éliminer les protéines thermolabiles (pistes 3 et 4 ; Fig. 88a). Une fois l'extrait clarifié, la protéine 6xHisHU_{Lla} a été purifiée par passage successif sur deux colonnes d'affinité : une colonne Ni-NTA (piste 6, Fig. 86a) puis sur une colonne d'héparine (piste 7, Fig. 88a) permettant l'obtention d'une protéine homogène après coloration au bleu de Coomassie (Fig. 88a).

L'effet de la protéine HU_{Lla} sur la réaction de recombinaison a été testé. En comparaison avec un extrait cellulaire d'*E. coli* traité à 95°C qui stimule fortement la recombinaison sur les sites *attP_{WT}/attB_{WT}* (Fig. 88c) et sur les sites *attP_{lactis}/attB_{lactis}* (Fig. 88d), après 1h30 de réaction, l'ajout de quantités croissante de la protéine HU_{Lla} stimule de manière dose-dépendante la recombinaison jusqu'à l'obtention d'un effet inhibiteur à 1µg de protéine (Fig. 88c ; Fig. 88d). La protéine HU apparait donc comme un cofacteur facultatif de la réaction catalysée par ^{mv4}Int.

<i>attB</i> (ARNt ^{SER})	B	O	B'		<i>attP</i>	C	O	C'	
<i>L. delbrueckii</i>	TTCAAAT	CCTGTAC	TCTCCTT	AAT	GAAAGAA	CCTGTAC	TCTCCTT	AAT	
<i>L. sakei</i>	TTCAAAT	CCTGTAC	TCTCCTT	TTA	GAAAGAA	CCTGTAC	TCTCCTT	TTA	
<i>L. acidophilus</i>	TTCAAAT	CCTGTAC	ACTCCTT	TTT	GAAAGAA	CCTGTAC	ACTCCTT	TTT	
<i>L. lactis</i>	TTCAAAT	CCCCTTG	ACTCCTT	TTA	GAAAGAA	CCCCTTG	ACTCCTT	TTA	
<i>S. mutans</i>	TTCAAAT	CCCGTCC	TTTCCTT	AAC	GAAAGAA	CCCGTCC	TTTCCTT	AAC	
<i>L. reuteri</i>	TTCAAAT	CCCCTTG	TCTCCTT	AGA	GAAAGAA	CCCCTTG	TCTCCTT	AGA	
<i>L. reuteri 2</i>	TTCAAAT	CCCCTTG	TCTCCAT	AGA	GAAAGAA	CCCCTTG	TCTCCAT	AGA	
<i>Ln. mesenteroides</i>	TTCAAAT	CCCGCTA	GCTCCTT	TAT	GAAAGAA	CCCGCTA	GCTCCTT	TAT	
<i>E. coli</i>	TTCAAAT	CCTCTCT	CCGCCAC	TTT	GAAAGAA	CCTCTCT	CCGCCAC	TTT	
<i>E. faecalis</i>	TTCAAAT	CCCTTAT	CCTCCGT	ACC	GAAAGAA	CCCTTAT	CCTCCGT	ACC	

Fig. 89 Séquences d'ARNt^{SER} utilisées avec les sites *attP* adaptés pour la recombinaison *in vitro*. Les sites cœur B, B', C et C' sont représentés par les flèches horizontales noires. Les deux sites de coupure encadrant la région centrale (O) sont indiqués par les flèches verticales noires. Les nucléotides qui diffèrent de la séquence *attB*_{WT} sont représentés en gras.

L'utilisation de la protéine HU comme cofacteur semble fréquent parmi la famille des β -recombinases, protéines appartenant à la famille des petites recombinases à sérine (Alonso *et al.*, 1995; Petit *et al.*, 1995; Rowland *et al.*, 2002). Parmi les recombinases à sérine, HU est aussi une protéine accessoire pour la recombinaison catalysée par l'invertase Hin (Johnson *et al.*, 1986). Parmi les recombinases à tyrosine, une grande majorité des intégrases hétérobivalentes utilisent la protéine IHF comme cofacteur de recombinaison et chez le modèle λ Int, la protéine HU a pu être utilisée pour remplacer partiellement la protéine IHF (Goodman *et al.*, 1992; Segall *et al.*, 1994). Seul le système hétérobivalent de l'ICE Tn916 utilise la protéine HU comme un cofacteur accessoire chez *E. coli*, améliorant d'un facteur 10 l'excision de l'ICE (Connolly *et al.*, 2002). Enfin, l'ICE CTnDOT utilise la protéine BHFa, une protéine appartenant à la famille IHF/HU, qui pourrait être la protéine HU de *Bacteroides thetaiotaomicron* (Ringwald and Gardner, 2015). En résumé, la protéine HU semble être le facteur accessoire de la recombinaison médiée par mv^4 Int, stimulant d'un facteur 3 la recombinaison entre sites sauvages et d'un facteur 25-30 celle entre sites modifiés.

II) Reprogrammation de la réaction de recombinaison vers d'autres séquences d'ARNt^{SER}

II.1) Reprogrammation de la réaction de recombinaison *in vitro*

En 1997, Auvray *et al.*, ont étudié la capacité d'intégration du système mv^4 dans différentes bactéries. Dans cette étude, il avait été observé que mv^4 Int était capable d'intégrer spécifiquement un plasmide non répliatif contenant mv^4 Int en amont du site *attP*_{WT} (plasmide pMC1), dans le gène codant un ARNt^{SER} des bactéries *Lactobacillus casei* et *Lactobacillus plantarum* en raison de la forte conservation de ce gène entre ces espèces (Auvray *et al.*, 1997). En revanche, dès que les séquences des ARNt^{SER} étaient moins conservées (*L. lactis*, *Enterococcus faecalis*, *Streptococcus pneumoniae*), le plasmide était intégré dans un site secondaire (Auvray *et al.*, 1997). Ainsi, dans le but de voir s'il est possible de reprogrammer la réaction de recombinaison vers des sites cibles autres que la séquence naturelle, nous avons essayé de rediriger *in vitro* la spécificité de cette réaction vers différentes séquences d'ARNt^{SER} bactériennes, des plus conservées aux moins conservées. Pour le test, différents variants du site *attP* avec la séquence de leur région cœur modifiée pour la rendre identique aux différents ARNt^{SER} testés ont été utilisés (Fig. 89).

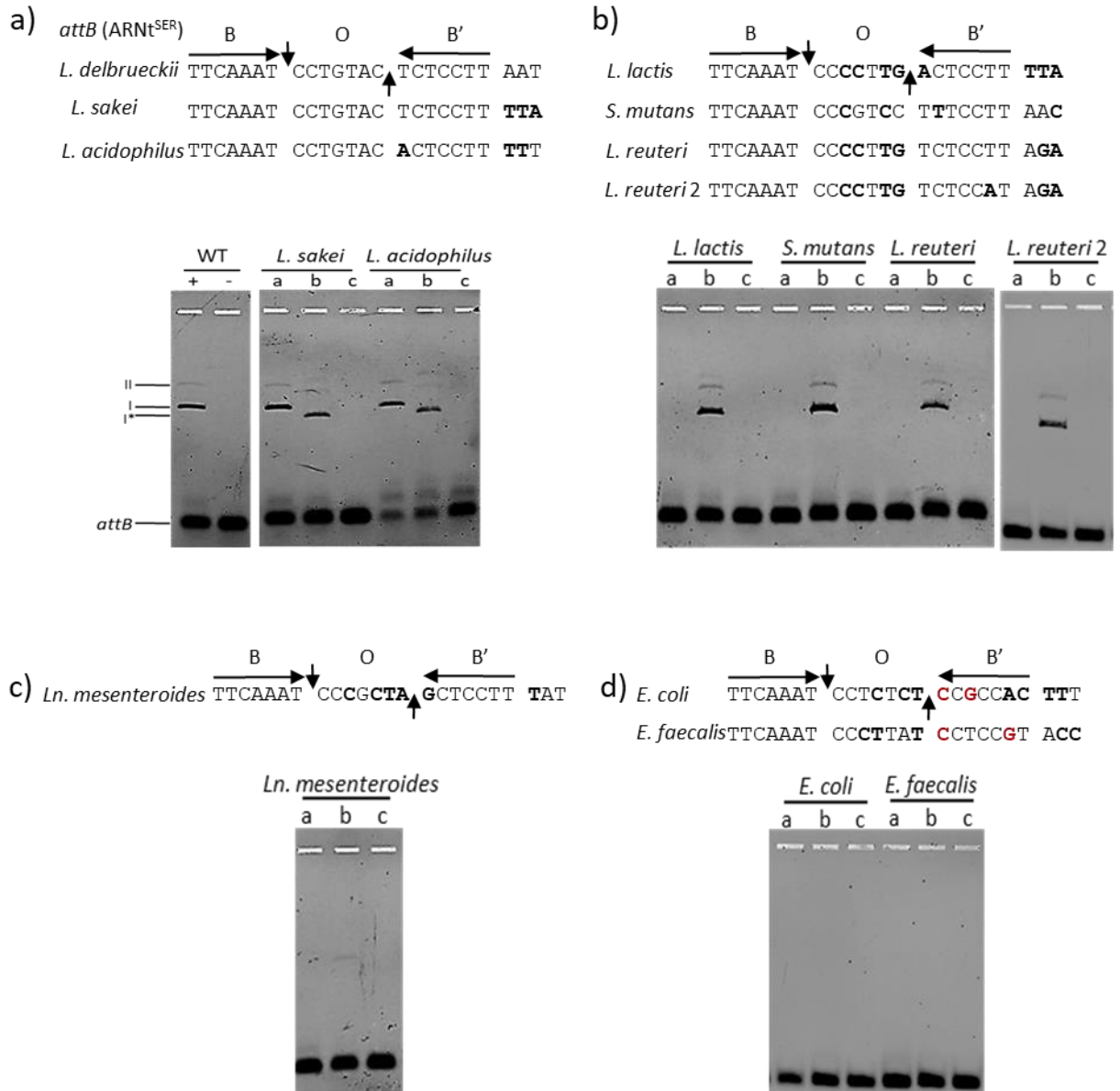


Fig. 90 Recombinaison *in vitro* entre les différents couples *attP_{WT}/attB_X* et *attP_X/attB_X*. La réaction de recombinaison contient 7,2 pmol de *mv⁴Int* et a été incubée 16h à 42°C. Pistes : a, couple *attP_{WT}/attB_X*; b, *attP_X/attB_X* avec *mv⁴Int* ; c, *attP_X/attB_X* sans *mv⁴Int* ; « + » *attP_{WT}/attB_{WT}* avec *mv⁴Int* et « - » couple *attP_{WT}/attB_{WT}* sans *mv⁴Int*. La bactérie d'où provient le site *attB_X* est indiquée au-dessus des puits. Le fragment *attB* fluorescent et le produit de recombinaison avec le site *attP_{WT}* (I) et le site *attP* adapté (I*) sont indiqués sur le gel. Le deuxième produit de recombinaison (II) provient très probablement d'une recombinaison *attP x attL* (II) (Coddeville and Ritzenthaler, 2010). a) Recombinaison possible avec le site *attP_{WT}* et le site *attP* adapté. b) et c) Recombinaison possible uniquement avec le site *attP* adapté. d) Recombinaison impossible, quel que soit le site *attP*.

D'après les résultats des banques du chapitre 2, l'identité de séquence entre les sites *attP* et *attB* n'est essentielle que dans la région centrale. Or, comme la très grande majorité des bactériophages s'intégrant dans des séquences d'ARNt restaurent un gène codant un ARNt fonctionnel (Williams, 2002), nous avons décidé d'adapter la région centrale et le site cœur C' (Fig. 89). En effet, les sites natifs de ^{mv4}Int possèdent des sites cœur B et C différents, mais des sites B' et C' identiques (Fig. 89) ce qui permet au bactériophage de restaurer un ARNt fonctionnel après intégration. L'avantage d'adapter la séquence C' est que l'intégration de nos plasmides ne modifie pas la séquence native de l'ARNt permettant une utilisation de ces constructions aussi bien *in vitro* qu'*in vivo*.

Au total, 8 séquences de gènes codant des ARNt^{SER} ont été utilisées (Fig. 89). Lors des différents clonages, un mutant ponctuel du site *attB* de *Lactobacillus reuteri* (séquence *L. reuteri* 2) a été obtenu et conservé pour les tests de recombinaison *in vitro* (Fig. 89). Ces gènes présentent tous un site cœur B conservé ainsi que plusieurs modifications dans les régions O et B', et ont été testés par recombinaison *in vitro* contre le site *attP*_{WT} et le site *attP* adapté. Après recombinaison, 4 profils différents ont été obtenus (Fig. 90). Le premier profil est composé des ARNt^{SER} de *L. sakei* et *L. acidophilus* pour lesquelles la recombinaison est possible avec les deux versions du site *attP* (pistes a et b, Fig. 90a). L'ARNt^{SER} de *L. sakei* est identique au site *attB*_{WT} (Fig. 90a), seules 3 bases diffèrent dans l'extrémité droite de la région cœur (région d'identité avec *attP*), hors du site *attB* minimum. Par conséquent, l'intégration du site *attP*_{WT} dans ce gène était un résultat attendu. L'ARNt^{SER} de *L. acidophilus* présente un site B' différent (Fig. 90a) mais étant donné qu'un A est toléré à cette position et que la région centrale est identique au site *attP*_{WT}, la recombinaison est possible avec ce site. En accord avec le signal de recombinaison obtenu, après recherche dans le classement des motifs B' du séquençage NGS, le site « 5'-ACTCCTT-3' » se classe à la 204^{ème} position, avec une occurrence élevée, de 1938.

Le deuxième profil est constitué des ARNt^{SER} de *L. lactis*, *Streptococcus mutans*, *L. reuteri* et *L. reuteri* 2 où la recombinaison n'est pas observée avec le site *attP*_{WT} (pistes a, Fig. 90b). En revanche, après adaptation de la région cœur du site *attP* à l'ARNt^{SER} ciblé, la recombinaison est de nouveau possible (pistes b, Fig. 90b). Ces sites cibles présentent de nombreuses modifications dans la région centrale ainsi qu'un changement de nucléotide dans le site B' (Fig. 90b).

L'absence d'identité dans la région centrale de ces ARNt^{SER} avec le site *attP*_{WT} explique l'absence de signal de recombinaison *in vitro* (pistes a, Fig. 90b) et est en accord avec les résultats de Auvray *et al.*, où le plasmide pMC1 n'avait pas pu être intégré dans l'ARNt^{SER} de *L. lactis* mais dans différents sites *attB* secondaires (Auvray *et al.*, 1997). L'adaptation de la région centrale et du site C' du site *attP* à chaque cible restaure la réaction de recombinaison montrant qu'il est possible de reprogrammer la réaction de recombinaison vers d'autres séquences que le site natif (Fig. 90b). En accord avec les signaux de recombinaison obtenu, après recherche dans le classement des motifs B' du séquençage NGS, l'occurrence des séquences B' des ARNt^{SER} de *L. lactis*, *S. mutans* et *L. reuteri* 2 est élevée et de 1938, 4074 et 765 respectivement.

Le troisième groupe est composé de l'ARNt^{SER} de *Leuconostoc mesenteroides*. Cette séquence ne peut être le substrat d'une recombinaison avec le site *attP*_{WT} du fait de l'absence d'identité dans la région centrale (piste a, Fig. 90c) mais permet une recombinaison à une faible efficacité avec le site *attP* adapté (piste b, Fig. 90c). Cette faible activité peut être expliquée de deux manières. Tout d'abord, cette séquence possède un G en première position du site B', un nucléotide toléré par ^{mv4}Int dans le motif de recombinaison B' mais observé en faible proportion par rapport aux nucléotides A et T. Il se pourrait donc qu'un G en première position du site B' soit fonctionnel, mais conduise à une activité de recombinaison faible. En désaccord avec cette idée, la séquence « 5'-GCTCCTT-3' » se classe à la 149^{ème} position, avec une occurrence élevée de 2230, suggérant qu'un autre facteur est responsable de la faible activité. Ainsi, il semble plus probable que la séquence de la région centrale « 5'-CCCGCTA-3' » entraîne une diminution de l'activité de recombinaison car dans le chapitre précédent, il a été montré qu'en fonction de la région centrale utilisé, l'activité de recombinaison pouvait, soit ne pas être impactée, soit diminuer d'un facteur 2 voir d'un facteur 15, malgré la conservation d'identité de séquence entre les régions centrales des sites *attB* et *attP*. Pour vérifier cette hypothèse, il faudrait modifier la région centrale du site *attB* natif par la séquence « 5'-CCCGCTA-3' » et regarder l'effet de cette séquence sur l'activité de recombinaison *in vitro*.

Enfin, le dernier groupe est constitué des ARNt^{SER} d'*E. coli* et *E. faecalis* pour lesquels la réaction de recombinaison n'est observée avec aucun des deux sites *attP* (pistes a et b, Fig. 90d). Ces sites possèdent plusieurs modifications dans la région centrale et dans le site cœur B'. En revanche, contrairement aux séquences précédentes, parmi les modifications, deux nucléotides non-permissifs sont observés dans le motif de recombinaison (bases rouges, Fig. 90d).

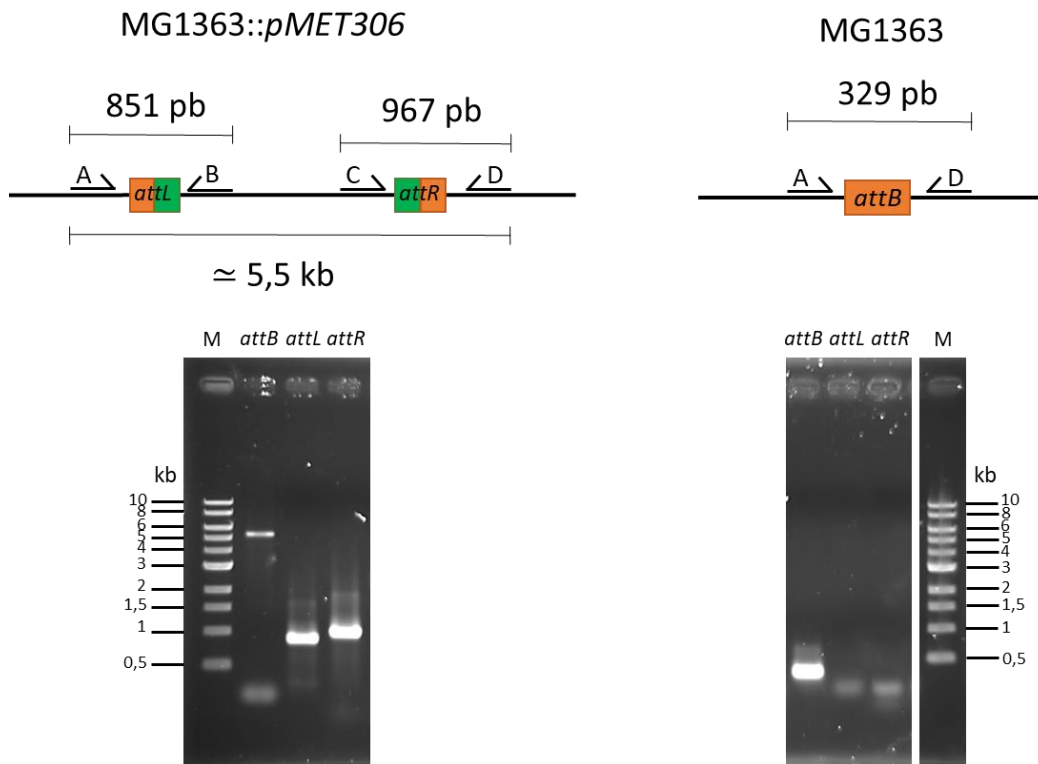


Fig. 91 Amplification par PCR des sites *attB*, *attL* et *attR* d'une colonie *L. lactis* MG1363 ARNt^{SER}::pMET306 et de la souche parentale. Le site amplifié est indiqué au-dessus de chaque piste. M, marqueur de taille 1 kb DNA ladder (New England Biolabs). Une représentation schématique de l'ADN testé est montrée au-dessus des gels.

En effet, un C en première position du site B' (Fig. 90d) est observé pour les deux sites, avec une purine dans le triplet de pyrimidines pour l'ARNt d'*E. coli* et un G à la 6^{ème} position pour l'ARNt d' *E. faecalis*. En accord avec l'absence de signaux de recombinaison, la recherche de ces séquences dans le classement des motifs B' montre que l'occurrence de la séquence de l'ARNt d'*E. coli* est de 0 et que celle de la séquence B' de l'ARNt d'*E. faecalis* est extrêmement faible, de 47. Par conséquent, ces deux séquences ne semblent pas être des sites permettant la recombinaison du fait de la présence de 2 nucléotides non permissifs.

En résumé, les différents résultats obtenus montrent qu'il est possible de reprogrammer la réaction de recombinaison *in vitro* vers de nouvelles séquences cibles *attB* en adaptant la région cœur du site *attP*, si les séquences ciblées respectent le motif consensus des régions cœur « 5'-DDBRRRW NNNNNNN DYYYVHB-3' ».

II.2) Reprogrammation de la réaction de recombinaison *in vivo*

La séquence codant l'ARNt^{SER} de *L. lactis* est un site où la recombinaison peut être reprogrammée *in vitro* par adaptation de la région cœur du site *attP*. Dans le but de vérifier si la reprogrammation de la réaction est également possible *in vivo*, la souche de laboratoire MG1363 de *L. lactis* a été transformée par les plasmides intégratifs pMC1(*attP*_{WT}) (Dupont *et al.*, 1995), ou pMET306 (pMC1-*attP*_{lactis}). Après transformation, aucune colonie résistante à l'érythromycine (Ery^R) n'a pu être obtenue avec le plasmide pMC1. En revanche, après adaptation de la région cœur du site *attP* à l'ARNt^{SER} ciblé (pMET306), une trentaine de colonies Ery^R a été obtenue. L'intégration spécifique dans l'ARNt^{SER} a été vérifiée par PCR en amplifiant l'ARNt^{SER} (*attB*) ainsi que les sites *attL* et *attR* (Fig. 91). Sur 20 colonies testées, toutes ont intégré le plasmide pMET306 dans l'ARNt^{SER} ciblé, indiquant que la réaction de recombinaison spécifique de site a pu être reprogrammée *in vivo* vers un nouveau site *attB*.

III) Reprogrammation de la réaction de recombinaison dans le gène *lacZ* de la bactérie *E. coli*

Le gène codant l'ARNt^{SER} d'*E. coli* n'étant pas un site où la réaction de recombinaison est reprogrammable du fait de la présence de nucléotides non permissifs nous avons décidé de réaliser une analyse *in silico* du génome d' *E. coli* afin d'identifier des sites putatifs de recombinaison appartenant à la séquence « 5'-DDBRRRW NNNNNNN DYYYVHB-3' ».

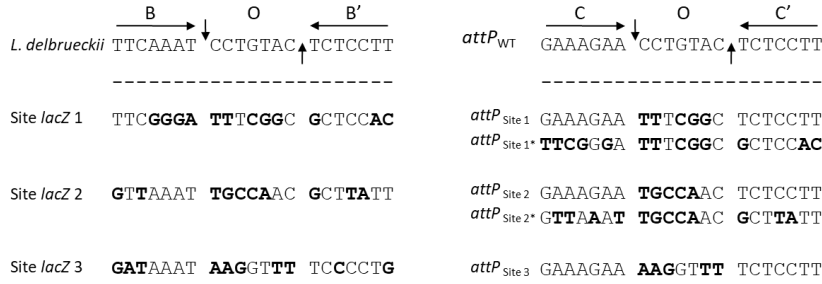


Fig. 92 Représentation des sites *attB* et *attP* artificiels. Les sites cœur B, B', C et C' sont représentés par une flèche noire horizontale et la région centrale est représentée par la lettre « O ». Les nucléotides modifiés par rapport au site *attB_{WT}* et *attP_{WT}* sont indiqués en gras.

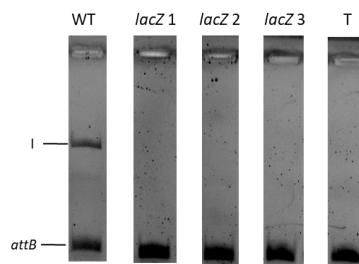


Fig. 93 Recombinaison *in vitro* sur les 3 sites *lacZ*. La réaction de recombinaison contient 7,2 pmol de *mv⁴Int*, 40 µg d'extrait cellulaire d'*E. coli* chauffé à 95°C et a été incubée 1h30 à 42°C. Le couple de sites *attB/attP* utilisé est indiqué au-dessus de chaque piste. Le fragment *attB* fluorescent et le produit de recombinaison (I) sont indiqués sur le gel. Piste T, témoin sans *mv⁴Int*.

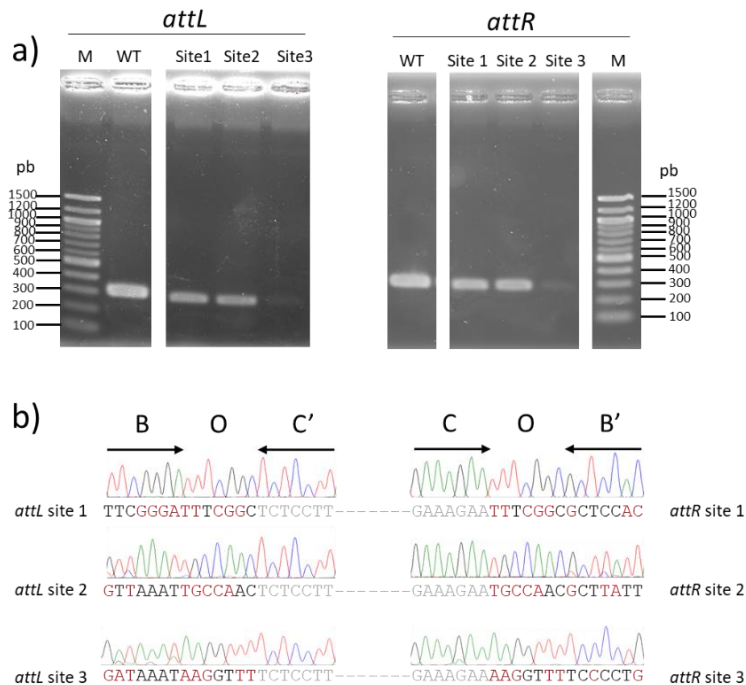


Fig. 94 Résultat de recombinaison entre les sites *lacZ* 1, *lacZ* 2, *lacZ* 3 et les sites *attP_{site1}*, *attP_{site2}* et *attP_{site3}*. **a)** Amplification par PCR des sites *attL* et *attR* après recombinaison. Pistes : WT, recombinaison *attB_{WT}* x *attP_{WT}* ; Site 1, recombinaison *lacZ* 1 x *attP_{site1}* ; Site 2, recombinaison *lacZ* 2 x *attP_{site2}* ; Site 3, recombinaison *lacZ* 3 x *attP_{site3}* et M, marqueur de poids moléculaire 100 pb (New England Biolabs). La réaction de recombinaison contient 7,2 pmol de *mv⁴Int* et 40 µg d'extrait cellulaire chauffé à 95°C d'*E. coli* et a été incubée 1h30 à 42°C. **b)** Séquences des sites *attL* et *attR*. Les bases en gris sont les bases appartenant au site *attP_{WT}* et les bases en rouge sont les bases qui diffèrent par rapport au site *attB_{WT}*.

Après analyse, 7959 sites putatifs ont été identifiés et bien qu'il soit hautement improbable que tous ces sites soient fonctionnels, nous avons décidé de tester la reprogrammation de la réaction de recombinaison vers 3 sites potentiels présent dans le gène *lacZ* (Fig. 92). Ces cibles ont été sélectionnées car elles permettent de cribler facilement les événements de recombinaison soit spécifiques de site, soit hors cible, lors de tests *in vivo*.

La région centrale du site *attP_{WT}* a donc été remplacée par celle du site *lacZ* 1 (*attP_{site1}*, pMET371), *lacZ* 2 (*attP_{site2}*, pMET372) et *lacZ* 3 (*attP_{site3}*, pMET373) (Fig. 92). Dans un premier temps, nous avons choisi de ne pas adapter les sites C et C', et de remplacer uniquement la région centrale car d'après nos résultats obtenus avec les banques aléatoires, l'identité nucléotidique dans les 2 régions centrales devrait suffire pour obtenir une recombinaison spécifique de site. Ces clonages ont été réalisés dans un plasmide à répllication conditionnelle chez *E. coli*. Ces plasmides, dérivés du plasmide pMC1 (Dupont *et al.*, 1995) contiennent le gène *mv⁴int* en amont du site *attP* et ont été transformés dans la souche *E. coli* AG1 (souche *lacZ⁺*, Stratagene). Aucun intégrant n'a pu être obtenu, suggérant soit que la recombinaison n'est pas possible pour ces sites modifiés, soit que l'efficacité de recombinaison est trop faible pour être observée dans le système expérimental mis en place. Afin de tester si la réaction de recombinaison était possible sur ces sites, les 3 couples *attP/attB* ont été testés par recombinaison *in vitro* classique (plasmides portant chacun un site *attP* adapté, contre un fragment PCR marqué portant les sites *attB* correspondants) (Fig. 93). Aucun produit n'a pu être observé suggérant que si une activité de recombinaison est possible, elle se situe sous le seuil de détection du système expérimental utilisé, soit moins de 5% de recombinaison comparé à un couple *attP/B* natif testé dans des conditions similaires (Fig. 47, p 128).

La présence de signaux de recombinaison a alors été recherchée en employant la méthodologie expérimentale utilisée avec les banques aléatoires en amplifiant par PCR les sites *attL* et *attR* (Fig. 94a) puis en les analysant par séquençage Sanger (Fig. 94b). Après recombinaison, les fragments PCR de tailles attendues sont obtenus pour les sites *attL* et *attR* issus de la recombinaison des sites *lacZ* 1 et *lacZ* 2 avec leurs sites *attP* respectifs (*attP_{site1}* et *attP_{site2}*) (Fig. 94a). En revanche, une bande de très faible intensité est obtenue pour les sites *attL* et *attR* issus de la recombinaison du site *lacZ* 3 (Fig. 94a). Ce résultat était attendu car le remplacement de la région centrale du site *attB* natif par la séquence « 5'-AAGGTTT-3' » diminue l'activité de recombinaison de ce site d'un facteur 15 (Fig. 64c, p 160). L'analyse des données de séquences montre que les trois événements de recombinaison observés correspondent bien à des recombinaisons spécifiques de site (Fig. 94b).

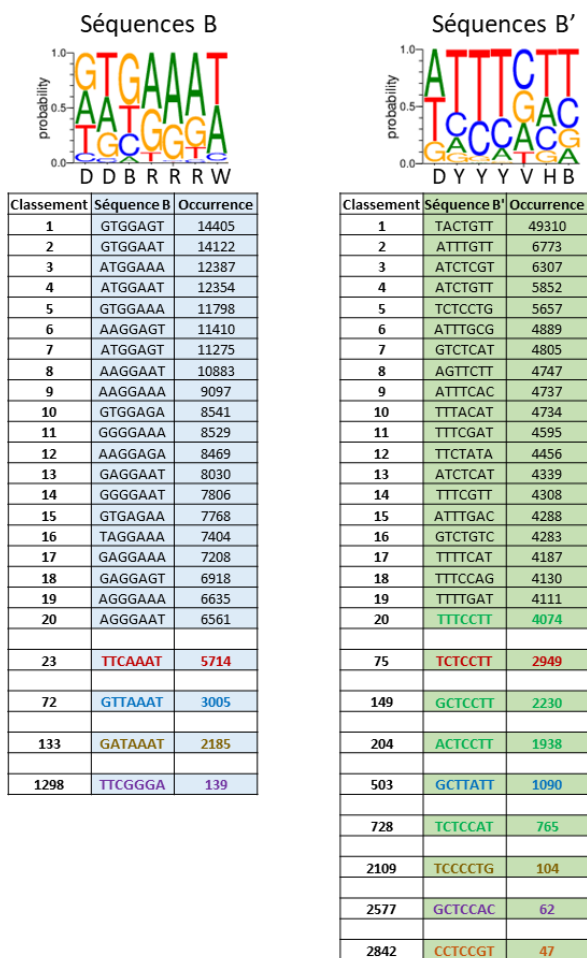


Fig. 95 Classement des 20 premières séquences B et B' après recombinaison ainsi que des variants B et B' utilisés dans le chapitre 3. Pour des raisons de clarté, cette figure reprend une partie de la figure 84. Une couleur est attribuée à chaque séquence : rouge, séquence native ; vert : séquence ARNt fonctionnel ; orange, séquence ARNt non fonctionnel ; violet, site *lacZ* 1 ; bleu, site *lacZ* 2 et dorée, site *lacZ* 3.

Ainsi, il est possible de reprogrammer spécifiquement la réaction de recombinaison vers un site *attB* respectant les contraintes nécessaires à la recombinaison et présentant plus de 50% de différence avec le site *attB*_{WT}. Par contre, bien que cette recombinaison soit possible, l'activité n'est détectable qu'après amplification *in vitro* des sites signatures de la recombinaison.

Dans le but de comprendre la raison de la faible efficacité de recombinaison des sites *lacZ* 1, 2 et 3, les séquences B et B' de chaque site ont été recherchées dans le classement des motifs B et B' obtenus par séquençage NGS (Fig. 95). Après analyse, il est observé que le site *lacZ* 1 possède deux motifs B et B' avec une occurrence extrêmement faible (139 et 62, respectivement) (Fig. 95), que le site *lacZ* 3 possède un motif B présent à une occurrence correcte (2185) mais que le motif B' possède une occurrence très faible (104) (Fig. 95). Ainsi, ces sites de recombinaison semblent peu fonctionnels en raison de la présence d'un ou de deux sites cœur présentant une occurrence extrêmement faible, donc étant des séquences peu fonctionnelles pour la recombinaison. De son côté, le site *lacZ* 2 comporte un site B et B' présent à une occurrence intermédiaire (3005 et 1090, soit à une fréquence 90% et 170% plus faible que les séquences natives respectivement) (Fig. 95). La séquence artificielle *L. reuteri* 2, une séquence permettant une recombinaison à une activité détectable, possède un motif B' retrouvé à une occurrence intermédiaire de 765, c'est-à-dire à une occurrence plus faible que celle du motif B' du site *lacZ* 2 (Fig. 95). En revanche, cette séquence possède un site B natif, retrouvé à une fréquence plus élevée que celle du motif B du site *lacZ* 2 (Fig. 95). Nos données suggèrent donc que pour qu'un site cible soit fonctionnel à un niveau d'activité significatif, il faut que ce site possède au moins un motif cœur présentant une occurrence très élevée. Ainsi, il semblerait que le site *lacZ* 2 soit un site présentant une très faible activité en raison de la présence de deux motifs cœur avec une occurrence intermédiaire. Ces résultats montrent que les séquences choisies dans le gène *lacZ* ne semblent pas être les séquences les plus optimales pour une recombinaison malgré le fait qu'elles respectent la séquence consensus dégénérée déterminée dans le chapitre 2. Ces résultats indiquent donc que pour la reprogrammation de la recombinaison, il semble important de rechercher des séquences cibles comportant des « motifs » B et B' présents à une occurrence élevée, voire très élevée, dans l'expérience de recombinaison réalisée avec les banques aléatoires.

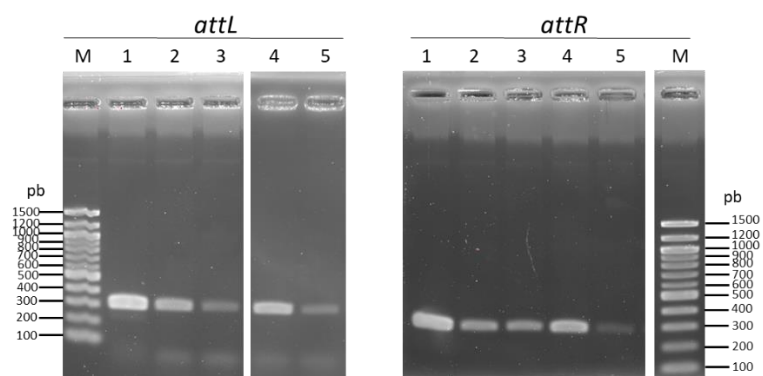


Fig. 96 Effet de l'adaptation de la région centrale et la région cœur COC' sur l'activité de recombinaison. Amplification par PCR des sites *attL* et *attR* après recombinaison. Pistes : 1, recombinaison *attB_{WT}* x *attP_{WT}* ; 2, recombinaison *lacZ* 1 x *attP_{site1}* ; 3, recombinaison *lacZ* 1 x *attP_{site1*}* ; 4, recombinaison *lacZ* 2 x *attP_{site2}* ; 5, recombinaison *lacZ* 2 x *attP_{site2*}* et M, marqueur de poids moléculaire. La réaction de recombinaison contient 7,2 pmol de *mv4*Int et 40 µg d'extrait cellulaire chauffé à 95°C d'*E. coli* et a été incubée 1h30 à 42°C.

Bien que les résultats obtenus *in vitro* dans le chapitre précédent suggèrent que l'identité nucléotidique des sites *attB* et *attP* au niveau des sites B/C et B'/C' des régions cœur n'est pas obligatoire pour obtenir une réaction de recombinaison spécifique de site, il ne peut être exclu que cette identité ait un rôle dans l'efficacité de cette recombinaison.

Dans le but de tester cet effet, la région COC' des sites *attP*_{site1} et *attP*_{site2} a été entièrement adaptée à chaque site ciblé (sites *attP*_{site1*} et *attP*_{site2*} ; Fig. 92). En raison de la très faible activité de recombinaison du site *lacZ* 3, ce site n'a pas été étudié plus en détail. La réaction de recombinaison a été effectuée entre les sites *lacZ* 1 ou *lacZ* 2 et les sites *attP* possédant soit la région centrale adaptée (pistes 2 et 4, Fig. 96), soit la région COC' adaptée (pistes 3 et 5, Fig. 96). La diminution de l'intensité des produits PCR issus de la recombinaison où la région COC' a été adaptée laisse penser que l'adaptation des séquences C et C' n'améliore pas la réaction de recombinaison mais au contraire l'altère (pistes 2 vs pistes 3 ; pistes 4 vs pistes 5, Fig. 96). Ainsi, ces résultats suggèrent que l'identité de séquence entre les sites B/C et B'/C' n'est pas obligatoire et peut même avoir un impact négatif sur l'activité de recombinaison. Il semblerait donc que la protéine ^{mv4}Int se soit adaptée à l'utilisation de ces sites cœur de fixation C et C' pour la mise en place de l'intasome et il n'est donc pas surprenant que la modification de ces séquences puisse avoir un effet négatif sur la recombinaison. Ces résultats ayant été obtenus par PCR en point final, une méthode non quantitative, ils nécessitent d'être validés par PCR en temps réel.

En conclusion, la protéine HU est une protéine accessoire de la recombinaison du système *mv4*, améliorant la recombinaison d'un facteur 2-3 sur les sites natifs et d'un facteur 25-30 sur les sites modifiés. Il est possible de reprogrammer spécifiquement la réaction de recombinaison médiée par ^{mv4}Int vers un site cible choisi *in vitro* et *in vivo*. Pour cela, le site cible choisi doit respecter les contraintes nécessaires à la recombinaison, et posséder des motifs B et B' avec une occurrence élevée, voire très élevée.

Discussions générales, conclusion et perspectives

Dans la littérature, ^{mv4}Int se différencie des intégrases hétérobivalentes en raison de multiples caractéristiques atypiques. En effet, le système mv4 présente une indépendance aux facteurs d'hôtes (Auvray *et al.*, 1999a), ne contient pas de sites cœur de fixation de type B, B', C et C' (Auvray *et al.*, 1999a; Auvray *et al.*, 1999b), possède une architecture atypique des sites bras de fixation (Coddeville *et al.*, 2014), une région centrale de taille inhabituelle (Coddeville *et al.*, 2014) et un site *attB* asymétrique, qui est le plus court décrit dans la littérature (Auvray *et al.*, 1999b).

Les travaux effectués pendant cette thèse ont permis de confirmer et de préciser certains aspects. La séquence protéique de ^{mv4}Int a été corrigée et il a été validé expérimentalement que cette protéine appartient à la grande famille des recombinases à tyrosine. L'indépendance aux facteurs d'hôte lors d'une recombinaison entre sites sauvages a pu être de nouveau démontrée, mais un facteur accessoire ubiquitaire, la protéine HU, stimule fortement la réaction de recombinaison lors de l'utilisation de substrats modifiés.

Une approche par l'utilisation de banques aléatoires a permis la caractérisation de la région cœur inhabituelle des sites *attB* et *attP*. Cette méthodologie est remarquablement adaptée à l'étude des intégrases hétérobivalentes et il serait intéressant de la voir appliquer à d'autres systèmes de recombinaison hétérobivalents, comme par exemple celui du bactériophage P2 d'*E. coli* où la région cœur des sites de recombinaison *attB* et *attP* est complexe, et où la région centrale n'a pas encore pu être identifiée. Dans notre étude, cette approche a permis de suivre le devenir de chaque brin lors de la recombinaison et donc de positionner avec précision les 2 sites de coupures de ^{mv4}Int. Ces résultats montrent que la taille de la région centrale est une séquence canonique de 7pb, comme pour tous les sites de recombinaison des systèmes hétérobivalents de bactériophages décrits jusqu'à présent. Historiquement, les sites de coupures sont déterminés par réaction de coupure sur un substrat suicide, or avec cette technique, il peut parfois s'avérer compliqué de positionner les sites de coupure avec certitude, ainsi, plusieurs positions peuvent parfois être proposés (Hauser and Scocca, 1992). L'utilisation de banques aléatoires semble donc être un nouvel outil permettant la caractérisation précise des extrémités de la région d'échange de brins, comme montré dans le système des intégrons (Escudero *et al.*, 2016) et dans le système mv4.

L'utilisation des banques aléatoires a également permis de rectifier la taille minimale du site *attB*, pour lequel un site de 16 pb avait précédemment été décrit (Auvray *et al.*, 1999b). La méthodologie des banques aléatoires a permis d'une part de montrer avec précision que le site *attB* est en réalité un site plus grand, d'une taille de 21 pb et d'autre part de mettre en évidence la présence de région répétées-inversées permettant de proposer une structuration plus classique (structure de type BOB' pour *attB* et COC' pour *attP*) des régions cœur. L'association de la recombinaison des banques aléatoires avec la méthodologie de retard sur gel a permis de démontrer expérimentalement que ces régions répétées-inversées correspondaient aux sites de fixation cœur de l'intégrase, et d'identifier précisément les sites bras de fixation de ^{mv4}Int.

Enfin, ces travaux ont aussi permis de confirmer les résultats de 1997 qui indiquaient une flexibilité importante de ^{mv4}Int vis-à-vis de ces sites de recombinaison (Auvray *et al.*, 1997). En effet, l'étude des banques aléatoires a permis de caractériser une séquence consensus dégénérée des sites cœur B, B', C et C', suggérant que ^{mv4}Int possède une grande flexibilité dans sa capacité à reconnaître ses sites de recombinaison et que ces sites pourraient être modifiables. Grâce à ces résultats, il a été montré sur différentes séquences d'ARNt^{SER} et sur 3 sites d'intégration putatifs dans le chromosome d'*E. coli*, que la réaction de recombinaison était reprogrammable si la séquence du site *attB* ciblé correspondait à la séquence consensus. L'association de la recombinaison des banques aléatoires au séquençage NGS a permis de classer des milliers de motifs nucléotidiques en fonction de leur efficacité dans la réaction de recombinaison *in vitro*. L'analyse des données de séquençage NGS a permis de montrer i) que la présence d'un nucléotide non-permissif dans le motif de recombinaison pouvait être toléré grâce à la présence de sous-motif particulier, ii) que la grande majorité des séquences les plus fonctionnelles respectaient les contraintes déterminées par séquençage Sanger et iii) que certains motifs respectant strictement la séquence consensus dégénérée pouvait être des motifs peu fonctionnels. Pour finir, les résultats de reprogrammation de la recombinaison suggèrent que pour obtenir une activité de recombinaison significative, il est nécessaire de cibler des séquences cibles contenant des « motifs » B et B' possédant une occurrence élevée voire très élevée.

En perspective, il serait intéressant d'approfondir l'étude du mécanisme de recombinaison de l'intégrase mv4 en raison de la capacité à former un intasome en absence de facteur d'hôte, des protéines généralement essentielles induisant une courbure de l'ADN, facilitant ainsi la fixation des domaines N-ter et C-ter de l'intégrase, mais aussi en raison de la grande flexibilité de ^{mv4}Int à reconnaître la région cœur des sites *attB* et *attP*.

Des analyses de la réaction d'intégration en molécule unique, notamment par TPM (« Tethered Particle Motion ») pourraient fournir des informations au niveau des différentes étapes de la recombinaison, de la fixation de l'intégrase aux sites *attB* et *attP* à la formation des produits de recombinaison. En effet, comme la recombinaison est possible avec un site *attP* linéaire, il est donc envisageable d'étudier le comportement des molécules d'ADN attachées à une bille microscopique et contenant le site *attP* seul, et de visualiser l'effet de l'ajout de ^{mv4}Int. Comme la recombinaison ne nécessite pas de facteur d'hôte, l'intégrase seule est donc théoriquement capable de former un intasome sur le site *attP* et grâce à la méthodologie TPM, il est envisageable de déterminer la vitesse de formation ainsi que la stabilité de ce complexe nucléoprotéique. La formation de courbure de l'ADN associée à la fixation de l'intégrase pourra aussi être caractérisé. De plus, comme la recombinaison intramoléculaire entre deux sites *attP* et *attB* est possible (Kolot *et al.*, 1999; Nash, 1974b), il semble réalisable de déterminer la cinétique de recombinaison des différentes étapes de la recombinaison *attP* x *attB*. Pour cela, une molécule d'ADN attachée à une bille microscopique contenant les sites *attB* et *attP* en orientation directe (excision) ou indirecte (inversion) pourra être utilisée afin de caractériser les vitesses de formation de l'intasome sur le site *attP*, de formation du complexe synaptique *attB/attP* et de formation de la jonction de Holliday.

L'étude de la fixation de ^{mv4}Int sur les sites bras et les sites cœur montre que, contrairement aux autres systèmes caractérisés, la protéine exerce une inhibition réciproque sur ses deux domaines. En effet, en absence de fixation du domaine N-ter, le domaine C-ter est incapable de se fixer de façon stable sur les sites cœur, et en absence de fixation du domaine C-ter, le domaine N-ter est aussi incapable de se fixer stablement sur les sites bras. On pourrait ainsi imaginer que dans la protéine ^{mv4}Int, les domaines N-ter et C-ter soient proches et se gênent mutuellement, mais que la fixation d'un domaine à l'ADN favorise le contact du second domaine avec l'ADN. La détermination de la structure de la protéine en absence et en présence de son ligand pourrait donc être envisagée. De plus, comme le système mv4 présente une indépendance aux facteurs d'hôte, la détermination de la structure du complexe intégratif par microscopie à force atomique ou par cryo-microscopie électronique nous permettrait de caractériser les propriétés permettant à une intégrase de former un intasome en absence de protéines supplémentaires.

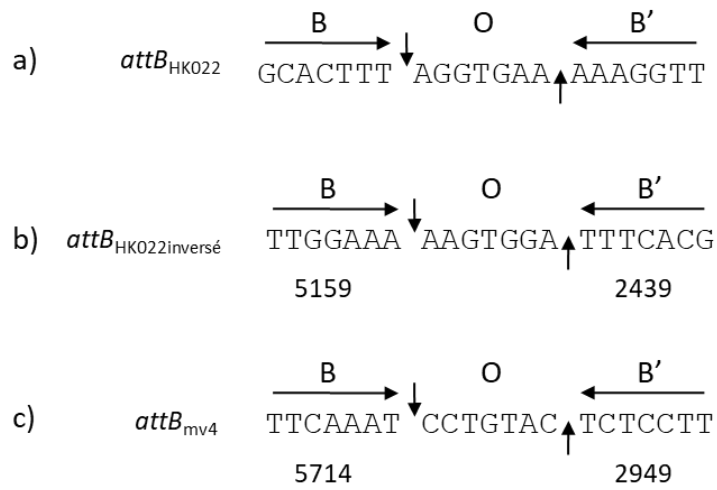


Fig. 97 Représentation des sites *attB* des bactériophages HK022 et mv4. a) Représentation du site *attB*_{HK022}. b) Séquence inversé du site *attB*_{HK022}. L'occurrence des motifs B et B' est indiquée sous la séquence. c) Représentation du site *attB*_{mv4}. L'occurrence des motifs B et B' est indiquée sous la séquence

Pour compléter les informations structurales, des expériences sur les domaines N-ter et C-ter tronqués de mv4 Int pourraient être envisagées. En effet, comme la protéine exerce une inhibition réciproque sur ces domaines, les domaines N-ter et C-ter isolés devraient avoir des interactions stables avec leurs sites de fixation, et le domaine C-ter pourrait avoir une activité catalytique (coupure de brins d'ADN et résolution de jonction de Holliday) plus élevée, comme c'est le cas pour $^{\lambda}$ Int (Sarkar *et al.*, 2001). Ainsi, le domaine N-ter purifié sera testé par retard sur gel en utilisant par exemple la paire de sites bras P'1-P'2, ce qui permettra d'évaluer la capacité de fixation aux sites bras en absence du domaine C-ter. Le domaine C-ter purifié sera aussi testé par retard sur gel en utilisant la région cœur du site *attP* COC' et par des tests de clivage et de résolution de jonction de Holliday. Ces différentes expériences nous fourniront des informations pertinentes sur la protéine mv4 Int ainsi que sur la réaction de recombinaison réalisée par cette protéine.

Au niveau du site *attP*, nos données indiquent la présence de deux paires de sites bras en orientation directe, une sur le bras P et une sur le bras P'. Il serait donc intéressant d'étudier la nécessité de chaque site bras dans la recombinaison dans le but de voir si mv4 Int nécessite la même combinaison de sites bras pour les réactions d'intégration et d'excision, comme c'est le cas pour *intS* du phage Kp1E1 d'*E. coli* (Panis *et al.*, 2010), ou bien nécessite un jeu de sites bras différents, comme c'est le cas pour $^{\lambda}$ Int (Hazelbaker *et al.*, 2008). Pour cela, différentes combinaisons de sites bras inactivés seront construites sur le site *attP* et testées pour leur capacité à réaliser la réaction de recombinaison, *in vivo* et *in vitro*.

En observant la séquence du site *attB*_{HK022} (Fig. 97a), nous nous sommes aperçus que le site *attB*_{HK022inversé} (Fig. 97b) était une séquence respectant le motif « 5'-DDBRRRW NNNNNNN DYYYVHB-3' ». En recherchant l'occurrence des séquences B et B' du site *attB*_{HK022inversé} dans le classement des motifs B et B' obtenus par séquençage NGS, des occurrences très proches de celles des sites B et B' natifs sont obtenues, 5159 pour le site B (contre 5714 pour la séquence native) et 2439 pour le site B' (contre 2949 contre la séquence native) (Fig. 97b ; Fig. 97c). Ainsi, il semble théoriquement possible de reprogrammer la réaction de recombinaison médiée par mv4 Int vers le site *attB* du bactériophage HK022, un phage très proche du modèle λ . Les bactériophages λ et HK022 sont des systèmes de recombinaison évolués, possédant une faible flexibilité vis-à-vis de la reconnaissance de la région cœur de leurs sites de recombinaison et utilisant un facteur d'hôte, la protéine IHF, permettant la formation d'un intasome sur leurs sites *attP* respectifs.

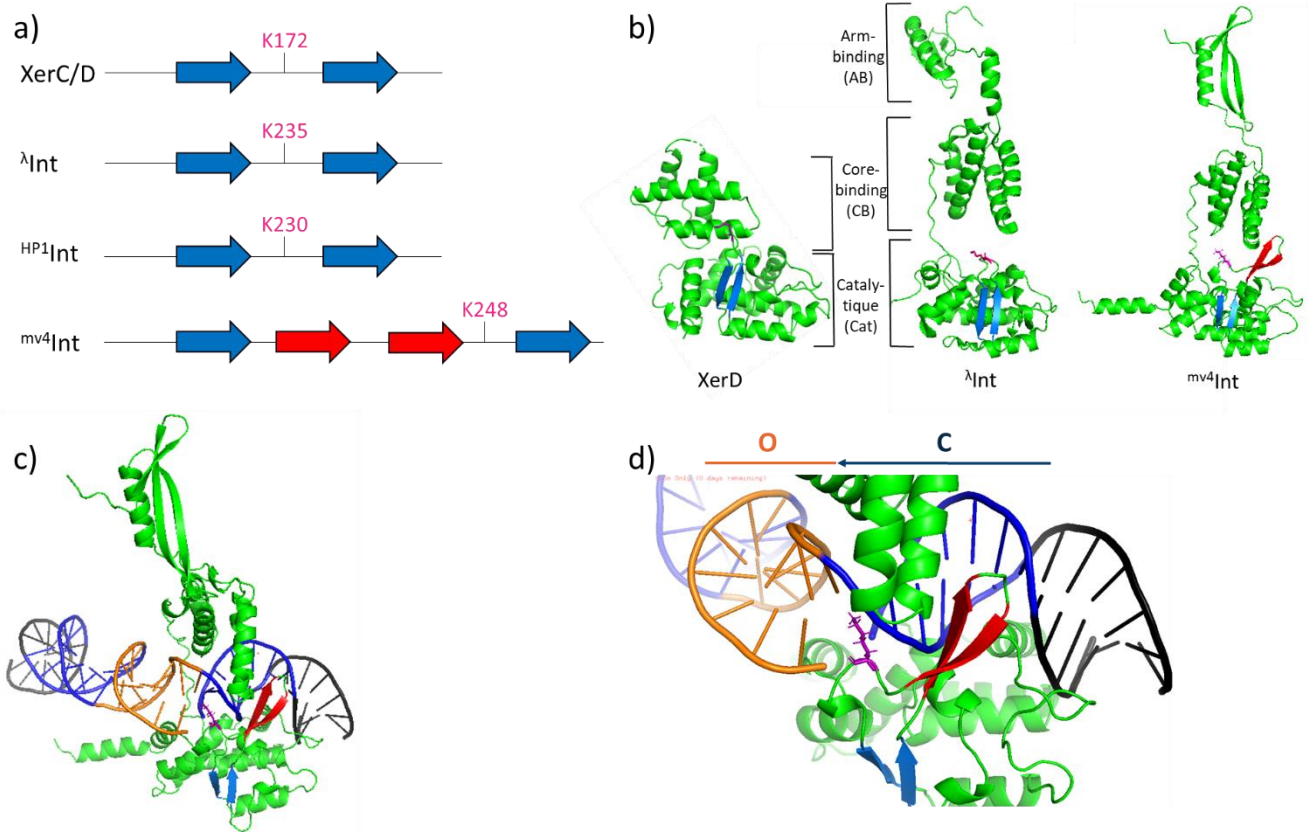


Fig. 98 Particularité structurale de $mv4Int$. **a)** Représentation schématique du positionnement de la lysine dans le domaine catalytique des recombinases à tyrosine. Les feuillettes- β sont représentés par des flèches. Chez les recombinases modèles comme XerC/D, λInt et $HP1Int$, la lysine essentielle à la réaction de recombinaison est positionnée grâce à la présence d'un double feuillet- β antiparallèle (flèches bleues). Chez $mv4Int$, la lysine est aussi positionnée par la présence d'un double feuillet- β antiparallèle (flèches bleues), mais il y a l'insertion d'un double feuillet- β antiparallèle supplémentaire (flèches rouges). **b)** Structures des recombinases XerD, λInt et $mv4Int$. La structure de XerD provient de Subramanya *et al.*, (Subramanya *et al.*, 1997), la structure de λInt provient de Biswas *et al.*, (Biswas *et al.*, 2005) et la structure de $mv4Int$ a été modélisée sur Alphafold (Jumper *et al.*, 2021). Les structures ont été visualisées sur le logiciel pyMOL. La lysine essentielle à la recombinaison est représentée en rose et les feuillettes- β antiparallèles sont représentés en bleu et en rouge. **c)** Modélisation de la fixation d'un monomère de $mv4Int$ aux sites cœur. Cette structure a été obtenue en superposant la structure de $mv4Int$ avec la structure d'un dimère de λInt fixé sur la région cœur COC' (pdb : 1Z19) (Biswas *et al.*, 2005). La séquence des sites cœur C et C' est représentée en bleu, et la séquence de la région centrale « O » est représentée en orange. **d)** Le motif protéique supplémentaire de $mv4Int$ pourrait être impliqué dans la fixation aux sites cœur. Cette figure est un agrandissement de l'image obtenue en partie **c)**. La structure est modélisée sur le site cœur C. Le double feuillet- β antiparallèle supplémentaire (flèches rouges) pourrait interagir avec les sites cœur *via* le sillon majeur de l'ADN.

Au contraire, le système mv4 semble être un système de recombinaison simple, présentant une grande flexibilité dans la capacité à reconnaître la région cœur de ces sites de recombinaison et ne nécessitant pas de facteur d'hôte pour la mise en place d'un intasome. Ainsi, l'intégrase du bactériophage mv4 pourrait être considérée comme un intermédiaire évolutif entre les recombinases simples tels que Cre et Flp qui ne nécessitent pas de protéines accessoires, et les recombinases hétérobivalentes tels que λ Int et HK022 Int, dépendantes des protéines de l'hôte. Si la recombinaison est réalisable au niveau du site $attB_{HK022}$ inversé, il serait intéressant de comparer les structures des protéines λ Int et mv4 Int, dans le but de comprendre les mécanismes impliqués dans l'indépendance aux facteurs d'hôte et dans la plus grande flexibilité de mv4 Int.

La structure de mv4 Int a été modélisée (Jumper *et al.*, 2021) sur alphafold et montre une grande ressemblance avec la protéine λ Int modèle (Fig. 98b). En effet, le domaine N-ter adopte un repliement de type « three-stranded β -sheet », comme le domaine N-ter de Tn916 Int (Wojciak *et al.*, 1999) et λ Int (Fadeev *et al.*, 2009) (Fig. 98b) et au niveau du domaine C-ter, mv4 Int adopte une structure de type « C-shaped clamp », comme l'ensemble des recombinases à tyrosine.

Cependant, quelques différences sont observées par comparaison au modèle λ Int. La première différence se trouve au niveau du « linker » reliant le domaine N-ter et C-ter où une hélice- α présente chez λ Int n'est pas retrouvée chez mv4 Int (Fig. 98b). Chez λ Int, cette hélice- α est impliquée dans la liaison coopérative aux sites bras entre domaines N-ter (Warren *et al.*, 2005). Nos données de retard sur gel indiquent aussi une forte coopérativité entre domaines N-ter pour la fixation aux sites bras et en absence de cette hélice- α dans le « linker », il serait intéressant de comprendre le mécanisme impliqué dans cette liaison coopérative. Un des premiers tests à réaliser serait d'étudier le domaine N-ter purifié, et de tester par retard sur gel sa capacité à se fixer coopérativement aux sites bras P'1-P'2. Au niveau du domaine « core-binding », 4 hélices- α sont observées, comme pour le modèle λ Int (Fig. 98b). Enfin, au niveau du domaine catalytique, 8 hélices- α (contre 7 pour λ Int) et 4 feuillets β (contre 5 pour λ Int) sont observés.

Chez les recombinases à tyrosine modèles comme les recombinases XerC/D, λ Int et HP1 Int, la lysine essentielle à la recombinaison est positionnée grâce à la présence de deux feuillets- β antiparallèles (Fig. 98a-b, flèches bleues). Chez mv4 Int, cette lysine est aussi positionnée grâce à ces deux feuillets- β (Fig. 98a-b, flèches bleues), mais il y a l'insertion de deux feuillets- β antiparallèles supplémentaire (Fig. 98a-b, flèches bleues).

En modélisant la fixation d'un monomère de mv^4Int à la région cœur COC' du site *attP* du bactériophage λ , la structure supplémentaire se retrouve proche de l'ADN et pourrait donc interagir avec les sites cœur C et C' *via* le sillon majeur de l'ADN (Fig. 98c-d, flèches rouges). Au niveau des AA présents dans le motif supplémentaire de mv^4Int , deux résidus prolines sont situés en amont de la lysine permettant d'une part la rigidification de ce motif et d'autre part le positionnement de la lysine dans le site catalytique. De plus, des AA généralement impliqués dans des interactions avec l'ADN sont retrouvés dans ce motif comme les résidus lysine et tyrosine. Ainsi, il est possible que ce domaine additionnel soit impliqué dans la liaison à l'ADN, et pourrait donc avoir un rôle majeur dans la plus grande flexibilité de recombinaison de mv^4Int . Afin de vérifier ce rôle potentiel, des variants de mv^4Int pourront être construits soit par mutagenèse dirigée en remplaçant les résidus possiblement impliqués dans la fixation à l'ADN par des résidus alanine, ou bien en déstructurant les feuillettes- β supplémentaires. Ces variants seront ensuite testés pour leur capacité de recombinaison sur les sites natifs, mais aussi sur différents sites modifiés disponibles dans la collection du laboratoire afin de voir l'impact de la délétion de cette structure sur la flexibilité de recombinaison de mv^4Int . Si cette structure est impliquée dans la flexibilité, il serait aussi intéressant de tester un variant de λInt où le domaine supplémentaire de mv^4Int aura été intégré en amont de la lysine. Ce variant sera ensuite comparé à la protéine native λInt lors de test de recombinaison sur une banque du site B et B' du site *attB* d'*E. coli*.

Un résultat important de ces travaux de thèse est la démonstration de la grande flexibilité de reconnaissance de mv^4Int vis-à-vis de ces sites de recombinaison. Actuellement, la région cœur du système *mv4* est la région la mieux caractérisée parmi les recombinaisons hétérobivalentes. Il serait intéressant d'utiliser la méthodologie des banques aléatoires au niveau de différents systèmes de recombinaison hétérobivalents dans le but d'observer si la plus grande flexibilité de mv^4Int est une caractéristique propre à cette protéine, ou bien partagée avec d'autres systèmes. Ainsi, une étude utilisant des banques aléatoires pourrait être envisagée sur différents groupes de systèmes de recombinaison comme celui des bactériophages λ et HK022, des bactériophages ne s'intégrant pas dans un ARNt, les systèmes des bactériophages HP1, L5 ou P22, des bactériophages s'intégrant dans la boucle anticodon d'un ARNt, le bactériophage $\Phi RV2$ de *M. tuberculosis* (Hendrix *et al.*, 1999) ou ΦCTX de *Pseudomonas aeruginosa* (Hayashi *et al.*, 1993), des bactériophages qui s'intègrent dans la boucle terminale de l'ARNt et celui du bactériophage TPW22 de *L. lactis*, un bactériophage s'intégrant dans l'extrémité 3' de la boucle terminale (Petersen *et al.*, 1999), comme le bactériophage *mv4*.

L'utilisation d'un ou plusieurs systèmes de chaque groupe permettra de révéler la flexibilité de chaque système en fonction de son lieu d'intégration (hors ARNt, boucle anti-codon, boucle terminale et extrémité 3' de la boucle terminale) et il sera possible de déterminer si la flexibilité de ^{mv4}Int est une caractéristique propre à cette protéine, partagée avec d'autres intégrases de la même famille, ou bien généralisable à différents bactériophages s'intégrant dans les ARNt, ou hors des ARNt.

Les résultats obtenus dans le chapitre 3 montre qu'il est possible de cibler spécifiquement une séquence par adaptation de la région cœur du site *attP*. L'association des données obtenus par séquençage Sanger et par séquençage NGS indique que pour reprogrammer la recombinaison vers une séquence cible à une activité significative, il faudrait cibler une séquence respectant les contraintes nucléotidiques pour la recombinaison et que cette séquence contienne des « mots » B et B' présent à une occurrence élevée voire très élevée. De façon arbitraire, en ne choisissant que les 75 premières séquences B et B', et en ajoutant une région centrale de type 7xN, 200 sites potentiels sont retrouvés chez *E. coli* et 300 chez *L. lactis*. Comme ces sites possèdent des séquences B et B' retrouvées à une forte occurrence lors de la recombinaison *in vitro*, il est envisageable de penser que ces différents sites sont fonctionnels à une activité convenable. Il sera donc intéressant de démontrer expérimentalement l'efficacité d'intégration en ciblant quelques sites parmi les 200 à 300 possibles.

En fonction des résultats, il serait alors possible d'envisager le développement d'un outil d'ingénierie des génomes bactériens « à la carte », où l'on pourrait choisir d'intégrer une séquence d'ADN parmi une liste de 200 voire 300 sites. Si un tel outil est mis en place, plusieurs étapes préliminaires seront à réalisées. Pour chaque espèce bactérienne ciblée, il faudra caractériser les différents « pseudo-sites *attB* » permissifs à l'intégration. Pour cela, une banque contenant les 4 nucléotides possibles dans la région centrale de 7 pb sera construite sur un plasmide suicide et transformée dans l'espèce ciblée. En fonction du nombre de clones obtenus, les sites d'intégration seront caractérisés par PCR inverse ou bien par une stratégie haut-débit similaire à la méthodologie TraDIS « Transposon Directed Insertion Sequencing » (Langridge *et al.*, 2009). Après obtention des résultats, il sera intéressant de comparer les sites obtenus *in vivo* par rapport aux sites théoriques caractérisés *in silico* grâce aux données NGS. Après détermination des différents sites d'intégration possibles, les « genomic safe harbors » c'est-à-dire des sites permettant l'expression de l'ADN transgène sans être délétère à l'hôte seront caractérisés.

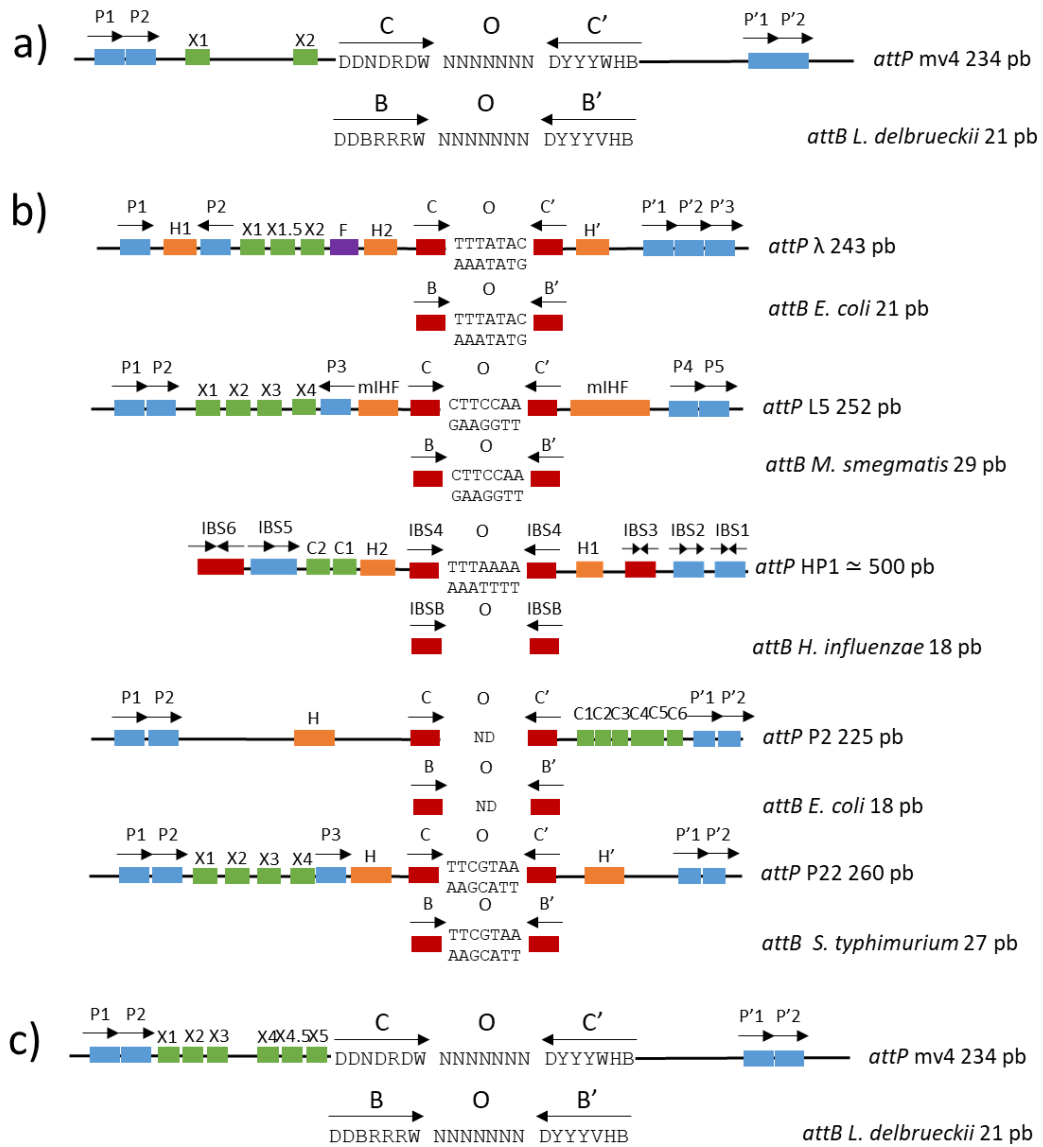


Fig. 99 Comparaison des sites de recombinaison de différentes intégrases hétérobivalentes. Comparaison des sites de recombinaison des systèmes λ , L5, HP1, P2, P22 et mv4, non dessiné à l'échelle. Les sites bras de fixation des intégrases sont représentés par des rectangles bleu. Les sites cœur de fixation de l'intégrase sont représentés par des rectangles rouge. La flèche noire indique l'orientation du site. Les rectangles orange indiquent les sites de fixation des protéines accessoires ayant un rôle architectural dans la recombinaison (IHF pour λ , HP1, P2, P22 et mIHF pour L5). Les rectangles vert indiquent les sites de fixation des excisionases (Xis pour λ , L5, P22, mv4 et Cox pour HP1 et P2). Le rectangle violet indique le site de fixation de la protéine Fis. **a)** Site *attP* mv4 avec les sites Xis déterminés en 2010 (Coddeville and Ritzenthaler, 2010). **b)** Site *attP* de λ Int, λ 5Int, λ HP1Int, λ P2Int et λ P22Int. **c)** Site *attP* mv4 avec les potentiels sites Xis déterminés *in silico*.

Les propriétés de l'outil pourront aussi être étudiées. Pour cela, 5 à 10 sites d'intégration pourraient être sélectionnés et la fréquence d'intégration ainsi que le taux de recombinaison hors-cible seront déterminés. Comme mv^4Int est capable de prendre en charge et d'intégrer le génome du bactériophage mv4 de 36 kb, il semble donc possible d'intégrer de l'ADN allant jusqu'à 36 kb. Afin de déterminer la limite maximale de l'ADN pouvant être pris en charge par l'intégrase, et l'effet de la taille de l'ADN sur l'activité d'intégration, des tailles croissantes d'ADN (5 kb, 20 kb, 100 kb) seront intégrées. Enfin, la gamme d'organisme capable d'utiliser mv^4Int sera identifiée en intégrant de l'ADN dans différents phyla bactérien.

Pour finir, lors de ces travaux de thèse, nous nous sommes uniquement focalisé sur la réaction d'intégration. En perspective, il serait intéressant d'étudier la réaction d'excision, qui présente aussi une caractéristique atypique. En effet, l'excisionase mv^4Xis a été déterminée, et contrairement aux systèmes classiques, cette protéine n'inhibe pas la réaction d'intégration bien qu'elle soit absolument requise pour la réaction d'excision (Coddeville and Ritzenthaler, 2010). Lors de cette étude, 2 sites de fixation de mv^4Xis avec comme séquence consensus « 5'-GAAACA-3' » avaient pu être déterminés (Fig. 99a). Dans les différents systèmes de recombinaison hétérobivalents, les sites de fixation de l'excisionase sont proches permettant aux molécules de Xis de se fixer coopérativement à l'ADN (Fig. 99b). Pour l'excisionase Xis de λInt les structures obtenues montrent que les protéines s'assemblent afin de former un filament micro-nucléoprotéique permettant la courbure de l'ADN (Abbani *et al.*, 2007) et il semblerait que la protéine Cox du bactériophage P2 fonctionne de la même façon (Berntsson *et al.*, 2014). Cette propriété semble être conservée parmi les différents systèmes hétérobivalents puisque les sites de recombinaison caractérisés présentent généralement 2 à 6 sites adjacents de l'excisionase (Fig. 99b). Bien que deux sites de fixation de mv^4Xis aient été déterminés sur le site *attP* (Fig. 99a), ces sites sont éloignés de 50 pb, ce qui rend impossible une fixation coopérative à l'ADN. En observant plus précisément le site *attP*, le motif « 5'-GAAAC-3' » est retrouvé 5 fois sur le bras P (X1 à X5, Fig. 99c) et une séquence « 5'-GAACC-3' » est retrouvée entre le site X4 et X5 (X4.5, Fig. 99c). Ainsi, il se pourrait que le site *attP* contienne en réalité 6 sites de fixation de mv^4Xis , ce qui faciliterait la liaison coopérative à l'ADN ainsi que la formation d'un filament micro-nucléoprotéique. En revanche, à la différence des systèmes classiques, la formation de ce filament n'est pas délétère envers la réaction d'intégration. En perspective, il serait intéressant de vérifier la présence de ces 6 sites de fixation par des études de retard sur gel.

La détermination de la structure cristalline du complexe $^{mv4}Xis/attR$ pourra aussi être envisagée dans le but d'obtenir plus d'information sur le fonctionnement de l'excisionase dans le système de recombinaison $mv4$ et de vérifier si cette protéine fonctionne de la même façon que l'excisionase du système λ . Si notre prédiction des 6 sites de fixation de l'excisionase s'avère correcte, il faudra déterminer leur rôle dans la recombinaison excisive. En raison de l'organisation de ces sites en 2 groupes de 3 sites adjacents, il est possible que chaque triplet de sites ait un rôle différent ou bien que seul un des deux triplets soit nécessaire à la recombinaison excisive. La connaissance des sites nécessaires à l'excision, ainsi que de la structure du complexe excisif pourra permettre d'expliquer comment l'excisionase est essentiel à l'excision, sans être défavorable à la recombinaison intégrative.

Matériels et méthodes

Matériels et méthodes

Table 2. Différentes amorces utilisées pour cette étude.

Amorces	séquence 5'-3'	Description	Tm
LbbulgattB-F	GAATTCCTGCAGCCCAAGC	Amplification site attB sur pbsattB / pMET348	63
Cy3-New-attB-R	GATGTAGATAATTTTGGGCCAAGG	Amplification site attB sur pbsattB / pMET348	62
SDM-attB4-F	GACCTGTACTCTCCTTAATAAGGTCAAATG	Mutagenèse site attB4 sur pbsattB	63-64
SDM-attB4-R	GACCTTATTAAGGAGATACAGGCTTGAAC	Mutagenèse site attB4 sur pbsattB	58-66
SDM-attB5-F	CCTGTACTCTCTGAATAAGGTCAAATGGTATC	Mutagenèse site attB5 sur pbsattB	61-67
SDM-attB5-R	GACCTTATTAAGGAGATACAGGCTTGAAC	Mutagenèse site attB5 sur pbsattB	61-67
SDM-LbsakattB-F	CCTGTACTCTCTTTTAAAGGTCAAATGGTATC	Mutagenèse site attB Lb. sakei sur pbsattB	58-66
SDM-LbsakattB-R	CATTTGACCTTTAAAAGGAGAGTACAGGATTTG	Mutagenèse site attB Lb. sakei sur pbsattB	57-66
SDM-LbacidattB-F	CCTGTACACTCCTTTTAAAGGTCAAATGGTATC	Mutagenèse site attB Lb. acidophilus sur pbsattB	56-67
SDM-LbacidattB-R	CATTTGACCTTTAAAAGGAGTGTACAGGATTTG	Mutagenèse site attB Lb. acidophilus sur pbsattB	55-66
SDM-LbreutattB-F	CCCCTTGCTCCTTAGAAAAGGTCAAATGGTATC	Mutagenèse site attB Lb. reuteri sur pbsattB	51-69
SDM-LbreutattB-R	CTTTTAAAGGAGACAAGGGGATTTGAACCTGCG	Mutagenèse site attB Lb. reuteri sur pbsattB	53-71
SDM-LlactisattB-F	CCCCTTGACTCCTTTTAAAGGTCAAATGGTATCC	Mutagenèse site attB L. lactis sur pbsattB	51-70
SDM-LlactisattB-R	CTTTAAAAGGAGTCAAGGGGATTTGAACCTGCG	Mutagenèse site attB L. lactis sur pbsattB	53-70
SDM-SmutansattB-F	CCCCTCCTTCTTAAACAAGGTCAAATGGTATC	Mutagenèse site attB S. mutans sur pbsattB	51-70
SDM-SmutansattB-R	CTTGTTAAGGAAAGGACGGGATTTGAACCTGCG	Mutagenèse site attB S. mutans sur pbsattB	53-72
SDM-EfaecattB-F	CCCTTATCCTCGTACCAAGGTCAAATGGTATCC	Mutagenèse site attB E. faecalis sur pbsattB	48-70
SDM-EfaecattB-R	CTTGGTACGGAGGATAAGGGATTTGAACCTGCG	Mutagenèse site attB E. faecalis sur pbsattB	53-72
SDM-LnmesentattB-F	CCCGTAGCTCCTTTATAAGGTCAAATGGTATC	Mutagenèse site attB Ln. mesenteroides sur pbsattB	52-69
SDM-LnmesentattB-R	CTTATAAAGGAGTACGGGGATTTGAACCTGCG	Mutagenèse site attB Ln. mesenteroides sur pbsattB	53-71
SDM-EcoliattB-F	CCTCTCTCCGCCACTTTAAGGTCAAATGGTATCC	Mutagenèse site attB E. coli sur pbsattB	52-72
SDM-EcoliattB-R	CTTAAAGTGGCGGAGAGAGGATTTGAACCTGCG	Mutagenèse site attB E. coli sur pbsattB	55-73
SDMgibsonermAM-F	GAGAATATCGTCAACTGTTTACTAAAAATC	Gibson mutation attP pMC1	61
SDMgibsonermAM-R	GTAACAGTTGACGATATTTCTCGATTG	Gibson mutation attP pMC1	62
SDMgibsonattP5-F	AAAGAACCTGTACTCTCCTGAATCAAAGCAATAATC	Mutagenèse site attP5 sur pMC1	62-68
SDMgibsonattP5-R	TTCAGGAGAGTACAGGTTCTTTCAACCATGTTTC	Mutagenèse site attP5 sur pMC1	68-69
SDMgibsonattPsakei-F	AAAGAACCTGTACTCTCCTTTTACAAGCAATAATC	Mutagenèse site attP adapté Lb. sakei sur pMC1	60-67
SDMgibsonattPsakei-R	AAAAGGAGAGTACAGGTTCTTTCAACCATGTTTC	Mutagenèse site attP adapté Lb. sakei sur pMC1	68
SDMgibsonattPacido-F	AAAGAACCTGTACTCTCCTTTTCAAAGCAATAATC	Mutagenèse site attP adapté Lb. acidophilus sur pMC1	57-67
SDMgibsonattPacido-R	AAAAGGAGTGTACAGGTTCTTTCAACCATGTTTC	Mutagenèse site attP adapté Lb. acidophilus sur pMC1	65-69
SDMgibsonattPmutans-F	AAAGAACCCTGCTTCTTCAACCAAGCAATAATCC	Mutagenèse site attP adapté S. mutans sur pMC1	52-72
SDMgibsonattPmutans-R	TTAAGGAAAGGACGGGTTCTTTCAACCATGTTTC	Mutagenèse site attP adapté S. mutans sur pMC1	57-70
SDMgibsonattPreuteri-F	AAAGAACCCTTGCTCTCTTAGACAAAGCAATAATCC	Mutagenèse site attP adapté Lb. reuteri sur pMC1	52-72
SDMgibsonattPreuteri-R	CTAAGGAGACAAGGGTCTTTCAACCATGTTTC	Mutagenèse site attP adapté Lb. reuteri sur pMC1	57-70
SDMgibsonattPlactis-F	AAAGAACCCTGTACTCTTTCACCAAGCAATAATCC	Mutagenèse site attP adapté L. lactis sur pMC1	48-71
SDMgibsonattPlactis-R	AAAAGGAGTCAAGGGTCTTTCAACCATGTTTC	Mutagenèse site attP adapté L. lactis sur pMC1	57-70
SDMgibsonattPfaeca-F	AAAGAACCCTTATCCTCGTACCCAAAGCAATAATCC	Mutagenèse site attP adapté E. faecalis sur pMC1	46-73
SDMgibsonattPfaeca-R	GTACGGAGGATAAGGGTCTTTCAACCATGTTTC	Mutagenèse site attP adapté E. faecalis sur pMC1	57-70
SDMgibsonattPcoli-F	AAAGAACCCTCTCCGCCACTTTCAAAGCAATAATCC	Mutagenèse site attP adapté E. coli sur pMC1	49-73
SDMgibsonattPcoli-R	AAGTGGCGGAGAGAGGTTCTTTCAACCATGTTTC	Mutagenèse site attP adapté E. coli sur pMC1	58-73
SDMgibsonattPmesent-F	AAAGAACCCTAGCTCCTTTATCAAAGCAATAATCC	Mutagenèse site attP adapté Ln. mesenteroides sur pMC1	53-72
SDMgibsonattPmesent-R	TAAAGGAGCTAGCGGTTCTTTCAACCATGTTTC	Mutagenèse site attP adapté Ln. mesenteroides sur pMC1	57-71
pBS-antiattB-F	GCTTGGGCTGCAGGA	Mutagenèse site attB sur pbsattB	60
pBS-antiattB-R	GGAAAGGACATCTAAATCAAATGG	Mutagenèse site attB sur pbsattB	59
pMC1seqprim3-F	TGCGTCCCTGCTCATCACTG	Séquençage intégrase	66
pMC1seqprim2-F	CAAGACGGCCACCGTACAAC	Séquençage intégrase	65
pMC1seqprim1-F	CGGCCGCTAGAACTAGTG	Séquençage intégrase	64
Cseqint3	TCTCCATGGTAATATCGATTGATTC	Séquençage intégrase	60
Cseqint2	GGCATGATGATCTGCTTTC	Séquençage intégrase	61
Cseqint1	TTGCAGGATTACGCTTTGG	Séquençage intégrase	60
Intmv4PCRseq-F	GCTTGGGCTGCAGGA	amplification intégrase sur génome phage mv4	60
Intmv4PCRseq-R	GGAAAGGACATCTAAATCAAATGG	amplification intégrase sur génome phage mv4	60
SDM-Lbreuteri2attB-F	TCCATAGAAAGGTCAAATGGTATC	Mutagenèse site attB Lb. reuteri2 sur pbsattB	60
SDM-Lbreuteri2attB-R	GACAAGGGGATTTGAACCTG	Mutagenèse site attB Lb. reuteri2 sur pbsattB	60
IntmutK248A-F	GCGACTAAGGCAGGGATCCG	Mutagenèse pour remplacer lysine 248 par alanine mv4Int	60-66
IntmutK248A-R	AGGTGGCTGGATCCCG	Mutagenèse pour remplacer lysine 248 par alanine mv4Int	62
IntmutY349F-F	TTTCATGCACGCCACTCAAG	Mutagenèse pour remplacer tyrosine 349 par phenylalanine mv4Int	60-63
IntmutY349F-R	ATCTCCATGGTAATATCGATTGATTC	Mutagenèse pour remplacer tyrosine 349 par phenylalanine mv4Int	60
pET-IntFus-F	ATAAATATGCCAAAGCTAATCTCGCAATCAAAAAATACAC	Amorce pour fusionner ORF8/Intmv4 sur pET-Int	68

I) Amorces, plasmides, souches et conditions de culture

Les différentes amorces, plasmides et souches utilisés dans cette étude sont décrits dans les tables 2, 3 et 4. Les souches de *L. lactis* sont cultivées à 30°C sur milieu M17 (Oxoid) supplémenté avec 0,5% de glucose. L'érythromycine est utilisée à une concentration de 1 µg/mL. Les souches d'*E. coli* sont cultivées à 37°C sur milieu Lysogenic Broth (LB) (Bertani, 1951). Les antibiotiques sont utilisés avec les concentrations suivantes : ampicilline 100 µg/mL, carbénicilline 100 µg/mL, chloramphénicol 12,5 µg/mL, érythromycine 150 µg/mL, kanamycine 50 µg/mL.

II) Manipulation de l'ADN

L'ADN plasmidique est extrait en utilisant le kit QIAprep Spin Miniprep (Qiagen) ou le kit NucleoBond Xtra Midi (Macherey-Nagel). Les réactions PCR (« Polymerase Chain Reaction ») sont effectuées avec la *Onetaq* polymerase (New England Biolabs), la Q5-HF polymerase (New England Biolabs) ou la CloneAmp Hifi polymerase (Takara Bio). La quantité d'ADN matrice utilisé était de 10 à 100 pg lors de l'utilisation d'ADN plasmidique et de 1 à 5 ng lors de l'utilisation d'ADN génomique. La technique « Touchdown » PCR (Don *et al.*, 1991) est parfois utilisée en commençant la réaction à une température d'hybridation de 10°C supérieure au T_m (« melting temperature ») des amorces, et en abaissant cette température de 1°C par cycle pendant 10 cycles. Une fois arrivé à une température équivalente au T_m des amorces, la réaction PCR est effectuée sur 20 cycles. Les produits PCR sont purifiés en utilisant le kit QIAquick PCR purification (Qiagen). La séquence des différentes constructions plasmidiques est vérifiée par séquençage Sanger (Mix2seq, Eurofins). L'ADN génomique (ADNg) est extrait par utilisation du kit DNeasy Blood and Tissue (Qiagen). Les différentes méthodologies sont effectuées en suivant les recommandations du fournisseur. Les clonages sont effectués soit par ligature à bout franc, soit par la méthodologie Gibson (Gibson *et al.*, 2009). Pour la réaction de ligature à bout franc, l'amplicon PCR (insert) est phosphorylé en utilisant l'enzyme T4 PNK (New England Biolabs) selon les recommandations du fournisseur. La réaction de ligature contient 0,03 pmol de vecteur, 0,03 pmol d'insert phosphorylé, du tampon T4 DNA ligase à concentration 1 X et 400 U de T4 DNA ligase (New England Biolabs) et est incubée une nuit à 16°C. Pour le clonage Gibson, le mix NEBuilder Hifi DNA Assembly (New England Biolabs) est utilisé en suivant les recommandations du fournisseur.

Matériels et méthodes

Suite table 2.

pET-IntFus-R	GCAGGATTAGCTTTGGCATATTTATCCACCAATATTC	Amorce pour fusionner ORF8/Intmv4 sur pET-Int	69
oripGh9-F	TAAATTTTACTACTGCAATCGGATGC	Amplification Ori pww01 sans repA sur plasmide pGhost9	60
oripGh9-R	GTATTTTAAATAGCCATGATATAAATACCTTATC	Amplification Ori pww01 sans repA sur plasmide pGhost9	60
pMC1-sans-colE1-F	TGGAACGAAAACCTACGTTAAG	Amplification pMC1 sans Ori ColE1	61
pMC1-sans-colE1-R	TGATTCTGTGGATAACCGTATTAC	Amplification pMC1 sans Ori ColE1	60
attB-Lactis-F	CATCATTCTTCTCTTCAAATTAATATAAATGC	Amplification 300pb génome L. lactis (tRNAser) ou 5500pb si pMC1 intégré	60
attB-Lactis-R	CAGGAGGAAAAGGAGTAAGC	Amplification 300pb génome L. lactis (tRNAser) ou 5500pb si pMC1 intégré	59
pET-Int-deltag-F	CCAAGCGTAATCCTGCAATC	Amplification pEt-Int sans le tag HIS et l'ORF-8	61
pET-Int-deltag-R	CATGGTATATCTCTTCTTAAAGTTAAAC	Amplification pEt-Int sans le tag HIS et l'ORF-8	60
pET-Intdelorf8-R	CATATGGCTGCCGCGC	Amplification pET-Int + tag HIS sans l'ORF-8	63
pCC1FOS-F	GTGGGATCCCCGGGTAC	Amplification pCC1FOS	62
pCC1FOS-R	GTGGGATCCTCTAGAGTCGAC	Amplification pCC1FOS	62
attP-F	GCAGGCGGAATGTTGAAAGAG	Amplification site attP sur pMC1	64
Cy3-attP-R	GGTGTAGTGGTCTACCCGTCAAAC	Amplification site attP sur pMC1	66
attP-N10-N19-R	GTTCTGGAGGTTTCAATCTTG	Amplification vecteur pMET359 pour construction banque attP	61
attP-N10-N19-Vf-F	GTGATAATCGCCTGCCGTTTGAC	Amplification vecteur pMET359 pour construction banque attP	68
anti-attP-N10-N19-F	CAAGATTCGAAACCTCCAGAAC	Amplification fragment attP banque	61
anti-attP-N10-N19-Vf-R	GTCAAACGGGCAGGCGATTATCAC	Amplification fragment attP banque	68
attB-N10-N19-F	GCAGGTCGACTCTAGAGGATCCCAAGCTCGAAATCCG CCGAACCAATG	Amplification fragment attB banque	51
attB-N10-N19-R	GAGCTCGGTACCCGGGATCCCAAGATTCCTACTGGCTA CTTGAATACTTCC	Amplification fragment attB banque	56
attB-gibsonpCC1Fos-F	GCAGGTCGACTCTAGAGGATCCCAAGATTCCTGCAGCCC AAGC	Clonage gibson attB dans pCC1Fos	63-81
attB-gibsonpCC1Fos-R	GAGCTCGGTACCCGGGATCCCAACCCATTTGATTAGAT GTCCTTTC	Clonage gibson attB dans pCC1Fos	61-80
attB Lib9	CGAAATCCGCCGAACCAATGTTGANNNNNNNGCAGGTT CANNNNNNTGACTCTCCNNNNNAAGGTCAAATNNNNNN NNTATAGGAAGTATTTCAAAGTAGCCAGTAGT	Fragment banque attB 2*N5	
SeqbanqueattB-F	TGCCTGCAAGTCTGACTCTAG	Amplification attB / attL (banque)	65
SeqbanqueattL-R	ACGCTAATGCCATCTATTAAGTACG	Amplification attL (banque)	62
SeqbanqueattR-F	GAACAACCCAGAAACGCTTTTATG	Amplification attR (banque)	61
SeqbanqueattB-R	GGCGAATTCGAGCTCGGTAC	Amplification attB / attR (banque)	65
SeqbanqueattP-R	GGTCGACGGTATCGATAAGC	Amplification attP (banque)	63
SDM-pCC1FosCm-F	TCAATGTACCTATAACAGACCGTTGAC	Mutagenèse site attB sur pCC1Fos	65
SDM-pCC1FosCm-R	GAACGGTCTGGTTATAGGTACATTGAG	Mutagenèse site attB sur pCC1Fos	64
SDMpCC1FosattB4-F	GCGCAGGTTCAcgaCCTGTACTCTCCTTAATAAG	Mutagenèse site attB A7->C7 / A8->C8 / T9->A9	62-72
SDMpCC1FosattB4-R	TACAGGtcgTGAACCTGCGCACC	Mutagenèse site attB A7->C7 / A8->C8 / T9->A9	57-71
SDMpCC1FosattB5-F	CTCCTgggaAAGGTCAAATGGTATCCCTATAG	Mutagenèse site attB T23->G23 / A24->G24 / A25->G25 / T26->A26	58-69
SDMpCC1FosattB5-R	CATTTGACCTTtcccAGGAGAGTACAGGATTGAACTGTC	Mutagenèse site attB T23->G23 / A24->G24 / A25->G25 / T26->A26	64-74
SDMpCC1FosattBA1-C1-F	GCGCAGGTTCAcATCCTGTACTCTCCTTAATAAG	Mutagenèse site attB A7->C7	65-72
SDMpCC1FosattBA1-C1-R	TACAGGATgTGAACCTGCGCACC	Mutagenèse site attB A7->C7	60-69
SDMpCC1FosattBA1-T1-F	GCGCAGGTTCAATCCTGTACTCTCCTTAATAAG	Mutagenèse site attB A7->T7	65-69
SDMpCC1FosattBA1-T1-R	TACAGGATaTGAACCTGCGCACC	Mutagenèse site attB A7->T7	60-66
SDMpCC1FosattBT3-C3-F	GCGCAGGTTCAAaCCTGTACTCTCCTTAATAAG	Mutagenèse site attB T9->C9	66-71
SDMpCC1FosattBT3-C3-R	TACAGGgTTTGAACCTGCGCACC	Mutagenèse site attB T9->C9	62-69
SDMpCC1FosattBT3-G3-F	GCGCAGGTTCAAaGCTGTACTCTCCTTAATAAG	Mutagenèse site attB T9->G9	66-71
SDMpCC1FosattBT3-G3-R	TACAGGcTTTGAACCTGCGCACC	Mutagenèse site attB T9->G9	62-69
SDMpCC1FosattBT7-A7-F	CTCCTaAATAAGGTCAAATGGTATCCCTATAG	Mutagenèse site attB T23->A23	59-64
SDMpCC1FosattBT7-A7-R	CATTTGACCTTATTtAGGAGAGTACAGGATTGAACTGTC	Mutagenèse site attB T23->A23	67-70
attBWT-F	ATCCTGTACTCTCCTTAAT	Amorce pour clonage oligo attBWT 16pb	
attBWT-R	ATTAAGGAGAGTACAGGAT	Amorce pour clonage oligo attBWT 16pb	
attB-WTGGTT-F	GGTCAAATCCTGTACTCTCCTTAAT	Amorce pour clonage oligo attBWT 23pb	
attB-WTGGTT-R	ATTAAGGAGAGTACAGGATTGAAACC	Amorce pour clonage oligo attBWT 23pb	
attB Lib7	CGAAATCCGCCGAACCAATGTTGANNNNNNNGCAGGTT CAAATNNNNNNNTCTCCTTAATAAGGTCAAATGNNNNN NNTATAGGAAGTATTTCAAAGTAGCCAGTAGT	Fragment banque attBN7overlap	
attB Lib8	CGAAATCCGCCGAACCAATGTTGANNNNNNNGCAGGTT CAAATCTGTACNNNNNNNAATAAGGTCAAATGNNNNN NNTATAGGAAGTATTTCAAAGTAGCCAGTAGT	Fragment banque attBN7coreright	

III) Production extrait cellulaire

III.1) Surproduction des protéines recombinantes chez *E. coli*

Le plasmide pET-Int (Auvray *et al.*, 1999a), pMET338, pMET339 et pMET369 sont transformés dans des cellules *E. coli* BL21(DE3) ((B F- ompT gal dcm lon hsdSB(rB-mB-) λ (DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λ S)), New England Biolabs) par électropériméabilisation en suivant le protocole de Wu *et al* (Wu *et al.*, 2010). Après étalement et croissance sur boîte, une préculture est réalisée en ensemençant 10 colonies dans du LB à 37°C, overnight. Le lendemain, 40 mL de LB est ensemenché avec 400 μ L de préculture et les cellules sont cultivées à 42°C, 200 rpm jusqu'à obtention d'une DO_{600nm} d'environ 0,6. L'expression de la protéine est induite par ajout de 100 μ M d'IPTG puis incubation sous agitation (200 rpm) à 22°C pendant 3h30. Les cellules sont récoltées par centrifugation (4500 g, 15 min, 4°C), et le culot repris par 1 mL de tampon D (50mM Tris pH 8, 500 mM NaCl, 10% glycérol, 1 mM DTT, 1 mM EDTA). Les cellules sont lysées par sonication (40% d'intensité, 5 cycles de 30 sec 5 cycles de 30 sec de pause). Le lysat est centrifugé (16000g, 15 min, 4°C) et le surnageant stocké à -20°C.

III.2) Extrait cellulaire d'*E. coli* et *L. lactis*

Les cellules d'une culture de nuit (40 mL) sont collectées par centrifugation (4500 g, 15 min, 4°C) et reprises dans 1 mL de tampon D. Les cellules sont lysées par sonication (40% d'intensité, 5 cycles de 30 sec 5 cycles de 30 sec de pause), celles de *L. lactis* ayant été préalablement traitées au lysozyme pendant 30 min (3 mg/mL, Roche). Le lysat est soit centrifugé (16000g, 15 min, 4°C) puis le surnageant stocké à -20°C, soit traité à 95°C 10 min, centrifugé (16000g, 15 min, 4°C) puis le surnageant stocké à -20°C.

IV) Purification des protéines

IV.1) Purification de ^{mv4}Int

Le plasmide pET-Int (Auvray *et al.*, 1999a) est transformé dans des cellules *E. coli* BL21(DE3). Après étalement et croissance sur boîte, une préculture est réalisée en ensemençant 10 colonies dans du LB à 37°C, overnight.

Matériels et méthodes

Suite table 2.

attP Lib2	CAAGATTCGAAACCTCCAGAACCCTCCAGA AACATGGTTG AAAGAANNNNNNNTCTCCTTAATCAAAGCAATAATCCCC GAGAAATCAACATTCTCGGGGATTATTTTTGTTTTAACTAG AAAATAACTAGAAAGAGCTAGTTAATAGATGGCATTAGCG TGATAATCGCCTGCCGTTTGAC	Fragment banque attPN7overlap	
attP Lib1	CAAGATTCGAAACCTCCAGAACCCTCCAGA AACATGGTTG AAAGAACCTGTACNNNNNNAATCAAAGCAATAATCCCC GAGAAATCAACATTCTCGGGGATTATTTTTGTTTTAACTAG AAAATAACTAGAAAGAGCTAGTTAATAGATGGCATTAGCG TGATAATCGCCTGCCGTTTGAC	Fragment banque attPN7coreright	
pMC1delTOPO-F	AAATCGAAACAGCAAAGAATGG	Amplification pMET340 pour ajouter gène bla	60
Bla-F	TGGCACTTTTCGGGAAATG	Amplification gène bla	63
Bla-R	GTGCTACAGAGTCTTGAAGTG	Amplification gène bla	60
attB Lib5	CGAAATCCGCCGAACCAATGTTGAATTGGTGNNNNNNN NNAATCCTGTACTCTCCTTAATAAGGTCAAATGGTATCCCT ATAGGAAGTATTTCAAAGTAGCCAGTAGT	Fragment banque attBminN10left	
attP Lib4	CAAGATTCGAAACCTCCAGAACCCTCCAGA AACNNNNN NNNNGAACCTGTACTCTCCTTAATCAAAGCAATAATCCCC GAGAAATCAACATTCTCGGGGATTATTTTTGTTTTAACTAG AAAATAACTAGAAAGAGCTAGTTAATAGATGGCATTAGCG TGATAATCGCCTGCCGTTTGAC	Fragment banque attPminN10left	
pET-F	TAGCTCGAGGATCCGGCTGCTAAC	amplification pET avec début tagHis	69
pETH-HU-20pbHis-R	GATGATGGCTGCTGCCATGGTATATCTCTTCTTAAAGTTA AACAAAAATATTTTC	amplification pET avec début tagHis	62-74
Hup6xHisgib-F	ATCATCATCACAGCAGCGGCATGGCTAACAAACAAGATCT TATC	amplification protéine Hup de L. lactis MG1363 et ajout de la fin du tagHis en N-ter	59-75
HU-TAA-R	TTATTTAACTGCGTCTTTCAATGCTTTAC	amplification protéine Hup de L. lactis MG1363 et ajout de la fin du tagHis en N-ter	62-74
attB Lib6	CGAAATCCGCCGAACCAATGTTGAATTGGTGCGNNNNN NNNNNCTGTACTCTCCTTAATAAGGTCAAATGGTATCCCT ATAGGAAGTATTTCAAAGTAGCCAGTAGT	Fragment banque attBN10left	
attP Lib3	CAAGATTCGAAACCTCCAGAACCCTCCAGA AACATGGTTN NNNNNCTGTACTCTCCTTAATCAAAGCAATAATCCCCG AGAAATCAACATTCTCGGGGATTATTTTTGTTTTAACTAGA AAATAACTAGAAAGAGCTAGTTAATAGATGGCATTAGCGT GATAATCGCCTGCCGTTTGAC	Fragment banque attPN7left	
attP2N7left	GCAGGCGGAATGTTGAAAGAGTTCTAAATCAACTAGANN NNNNNCTAGAAAAAGTGAACGCATGGAAACAACCCAGA AACGCTTTTTAGCCAAAAGCAAGATTCGAAACCTCCAGA AC	Fragment banque attP2N7left	
attP2N7right	GCAGGCGGAATGTTGAAAGAGTTCTAAATCAACTAGATTTT TAANNNNNNAAGTGAACGCATGGAAACAACCCAGAAA CGCTTTTTAGCCAAAAGCAAGATTCGAAACCTCCAGAAC	Fragment banque attP2N7right	
banquebrasattP-R	CTCTTTCAACATTCCGCTGTC	Amplification vecteur pMET359 pour construction banque attP	64
HupGib-F	CTTTAAGAAGGAGATATACCATGGCTAACAAACAAGATCT TATC	Amplification protéine Hup de L. lactis pour clonage gibson dans pFSKB3X	59-69
HupGibpFSKB3X-R	TGTCACCAAGTCTAGCTTATTTAACTGCGTCTTTCAAT GCTTTAC	Amplification protéine Hup de L. lactis pour clonage gibson dans pFSKB3X	63
pFSKB3X-F	GCTAGCATGACTGGTGGAC	amplif vecteur d'expression protéine pFSKB3X	62
pETRBS-F	GAAATAATTTTGTAACTTTAAGAAAGGAGATATACC	Clonage gibson pour ajouter 6xHisHU dans pFSKB3X	62
Hu-Cter-R	GAATGCAGGAACAACCTGTAGC	Clonage gibson pour ajouter 6xHisHU dans pFSKB3X	61
HU-Cter-F	GCTACAGTTGTTCTGCATTC	Clonage gibson pour ajouter 6xHisHU dans pFSKB3X	61
Site1LacZ-Gib-F	AAAGAATTCGGCTCTCCTTAATCAAAGCAATAATCCC	Mutagenèse gibson overlap attP par overlap site 1 putatif LacZ	61-70
Site2LacZ-Gib-F	AAAGAATGCCAACTCTCCTTAATCAAAGCAATAATCCC	Mutagenèse gibson overlap attP par overlap site 2 putatif LacZ	62-70
Site3LacZ-Gib-F	AAAGAAAAGTTTTCTCCTTAATCAAAGCAATAATCCC	Mutagenèse gibson overlap attP par overlap site 3 putatif LacZ	60-68

La protéine est surproduite par ensemencement de 5 mL de préculture dans 500 mL de milieu LB et induction à l'IPTG selon les conditions précédemment citées. Les cellules sont récoltées par centrifugation (6000g, 15 min, 4°C), lavées une fois dans 10 mL de tampon A (10 mM Tris pH 8, 150 mM NaCl), centrifugées (4500 g, 15 min, 4°C) et le culot repris dans 10 mL de tampon B (50 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole, 10% glycérol) contenant une tablette d'inhibiteur de protéase SIGMAFAST Protease Inhibitor Cocktail Tablets EDTA-Free (Sigma). Les cellules sont lysées par addition de lysozyme (1 mg/mL) puis sonication (40% d'intensité, 10 cycles de 30 sec suivi de 45 sec de pause dans la glace). Le lysat est centrifugé (20 min, 40000 g, 4°C). La protéine est purifiée par succession de chromatographies FPLC (appareil AKTA). Une colonne HisTrap HP (GE Healthcare) est équilibrée en injectant 10 volumes de colonne (10 mL) de tampon B. Les 10 mL de lysat sont injectés, et les protéines non fixées sont éluées par injection de 10 volumes de colonne de tampon B. L'intégrase est éluée avec un gradient (0 à 30%) de tampon C (50 mM Tris pH 8, 500 mM NaCl, 500 mM imidazole, 10% glycérol). La fraction éluée est ensuite injectée sur colonne SUPERDEX 200 (16/60, GE Healthcare) en utilisant le tampon D. La fraction contenant l'intégrase est diluée au ½ avec le tampon E (50 mM Tris pH 8, 10% glycérol, 1 mM DTT, 1mM EDTA). Une colonne HiTrap Heparin HP (GE Healthcare) est équilibrée en injectant 10 volumes de colonne (10 mL) de tampon F (50 mM Tris pH 8, 250 mM NaCl, 20% glycérol, 1 mM DTT, 1 mM EDTA). La fraction contenant l'intégrase est ensuite injectée et les protéines non fixées sont éluées par injection de 10 volumes de colonne de tampon F. L'intégrase est éluée avec un gradient (0 à 100%) de tampon G (50 mM Tris pH 8, 1 M NaCl, 20% glycérol, 1 mM DTT, 1 mM EDTA). La protéine purifiée est aliquotée par 20 µL et congelée dans de l'azote liquide puis stockée à -80°C. La pureté de l'intégrase est évaluée par électrophorèse sur gel d'acrylamide 4-15% (Mini-PROTEAN TGX, Biorad) après coloration au bleu de Coomassie (Instant Blue, Abcam). La concentration protéique est estimée au Nanodrop ND-2000 (Thermofisher) en mesurant la DO_{280nm} .

IV.2) Purification de 6xHis-HU_{Lla}

Le plasmide pMET369 est transformé dans des cellules *E. coli* BL21(DE3). Après étalement et croissance sur boîte, une préculture est réalisée en ensemencant 10 colonies dans du LB à 37°C, overnight. La protéine est surproduite par ensemencement de 2,5 mL de préculture dans 250 mL de milieu LB et induction par 400µM d'IPTG selon les conditions précédemment citées.

Matériels et méthodes

Suite table 2.

Amorces	séquence 5'-3'	Description	Tm
Site1LacZ-Gib-R	TTAAGGAGAGCCGAAATCTTTCAACCATGTTTCTGGAG	Mutagenèse gibson overlap attP par overlap site 1 putatif LacZ	60-70
Site2LacZ-Gib-R	TTAAGGAGAGTTGGCATTCTTTCAACCATGTTTCTGGAG	Mutagenèse gibson overlap attP par overlap site 2 putatif LacZ	60-72
Site3LacZ-Gib-R	TTAAGGAGAAAACCTTTTCTTTCAACCATGTTTCTGGAG	Mutagenèse gibson overlap attP par overlap site 3 putatif LacZ	60-70
AmpGib-F	GAATTATGCAGTGTGCCATAAC	Amorce pour mutagenèse gibson	62
AmpGib-R	GTTATGGCAGCACTGCATAATC	Amorce pour mutagenèse gibson	62
OverlapSite1LacZ-R	AAGGAGAGccgAaaATTGAACTGCGCACC	Mutagenèse gibson overlap attB par overlap site 1 putatif LacZ	60-74
OverlapSite2LacZ-R	AAGGAGAGTggcaATTGAACTGCGCACC	Mutagenèse gibson overlap attB par overlap site 2 putatif LacZ	60-74
OverlapSite3LacZ-R	AAGGAGAaaAccttATTGAACTGCGCACC	Mutagenèse gibson overlap attB par overlap site 3 putatif LacZ	60-71
OverlapSite1LacZ-F	TCAAAATtTcggCTCTCCTTAATAAGGTCAAATGGTATC	Mutagenèse gibson overlap attB par overlap site 1 putatif LacZ	61-69
OverlapSite2LacZ-F	TCAAAATtgccaACTCTCCTTAATAAGGTCAAATGGTATC	Mutagenèse gibson overlap attB par overlap site 2 putatif LacZ	62-69
OverlapSite3LacZ-F	TCAAAATaagGttTCTCCTTAATAAGGTCAAATGGTATC	Mutagenèse gibson overlap attB par overlap site 3 putatif LacZ	60-67
attBsite1LacZ-F	GAAATCCCGAACCTGCGCACCAATTCAACATTG	Mutagenèse attB par site 1 putatif LacZ	65-73
attBsite1LacZ-R	GGCGCTCCACAATAAGGTCAAATGGTATCCCTATAGG	Mutagenèse attB par site 1 putatif LacZ	61-73
attBsite2LacZ-F	GGCAATTTAACCTGCGCACCAATTCAACATTG	Mutagenèse attB par site 2 putatif LacZ	65-72
attBsite2LacZ-R	AAGCTTATTAAATAAGGTCAAATGGTATCCCTATAGG	Mutagenèse attB par site 2 putatif LacZ	61-67
attBsite3LacZ-F	CCTTATTATCCCTGCGCACCAATTCAACATTG	Mutagenèse attB par site 3 putatif LacZ	65-70
attBsite3LacZ-R	TTTTCCCTGAATAAGGTCAAATGGTATCCCTATAGG	Mutagenèse attB par site 3 putatif LacZ	61-70
attPCOC'site1LacZ-F	GGCGCTCCACAATCAAAGCAATAATCCC	Mutagenèse cœur attP par site 1 putatif LacZ	51-70
attPCOC'site1LacZ-R	GAAATCCCGAAACCATGTTTCTGGAGGGTTC	Mutagenèse cœur attP par site 1 putatif LacZ	62-71
attPCOC'site2LacZ-F	AACGCTTATTAAATAAGGTCAAATGGTATCCCTATAGG	Mutagenèse cœur attP par site 2 putatif LacZ	51-62
attPCOC'site2LacZ-R	GGCAATTTAAACAACCATGTTTCTGGAGGGTTC	Mutagenèse cœur attP par site 2 putatif LacZ	62-70
RP	CTCGTATGTTGTGGGAATTGTGAGC	Amorce pCC1Fos pour amplification attB	66
pCC1Fos-Cy3-R	GCGGGCTCTTCGCTATTAC	Amorce pCC1Fos pour amplification attB	65
COC'WT-F	GGTATTGAAAGAACTGTACTCTCTTCGCTAAC	Amorce pour hybridation double brin : fragment COC'WT EMSA	
COC'WT-Cy3-R	GTTACGCAAGGAGAGTACAGTTCTTTCCAATACC	Amorce pour hybridation double brin : fragment COC'WT EMSA	
P'12WT-28bp-F	GTTTTAACTAGAAAATAACTAGAAATTC	Amorce pour hybridation double brin : fragment P'12 28 pb EMSA	
P'12WT-28bp-R	GAATTCTAGTTATTTCTAGTAAAAAC	Amorce pour hybridation double brin : fragment P'12 28 pb EMSA	
P'12WT-40bp-F	CACGTCGTTTTAACTAGAAAATAACTAGAAATCCACGTC	Amorce pour hybridation double brin : fragment P'12 40 pb EMSA	
P'12WT-40bp-R	GACGTGGAATTCTAGTTATTTCTAGTAAAAACGACGTG	Amorce pour hybridation double brin : fragment P'12 40 pb EMSA	
P'12WTdelP'3-F	TTTTAACTAGAAAATAACTAGAAAAGCGACCGAGCGGTGCG	Amorce pour hybridation double brin : fragment P'12 40 pb EMSA	
P'12WTdelP'3-R	CGACCGCTCGTGCTTTCTAGTTATTTCTAGTAAAAA	Amorce pour hybridation double brin : fragment P'12 40 pb EMSA	
P'123WT-F	TTTTAACTAGAAAATAACTAGAAAAGAGCTAGTTAATAGA	Amorce pour hybridation double brin : fragment P'123 40 pb EMSA	
P'123WT-R	TCTATTAAGTCTCTTTCTAGTTATTTCTAGTAAAAA	Amorce pour hybridation double brin : fragment P'123 40 pb EMSA	
P'1WT-F	TTTTAACTAGAACGACCGAGCGGAGCGGCACGCTGTGCG	Amorce pour hybridation double brin : fragment P'1 40 pb EMSA	
P'1WT-R	CGACAGCGTGCCGCTCCGCTCGGTCTAGTTAAAAA	Amorce pour hybridation double brin : fragment P'1 40 pb EMSA	
P'2WT-F	GCCGACCGAGCGGAATAACTAGAAAAGCGGCACGCTGTGCG	Amorce pour hybridation double brin : fragment P'2 40 pb EMSA	
P'2WT-R	CGACAGCGTGCCGCTTTCTAGTTATTCGCTCGGTGCGG	Amorce pour hybridation double brin : fragment P'2 40 pb EMSA	
P'3WT-F	GCCGACCGAGCGGGCGGCACGCTGAGAGCTAGTTAATAGA	Amorce pour hybridation double brin : fragment P'3 40 pb EMSA	
P'3WT-R	TCTATTAAGTCTCTAGCGTCCGCGCTCGGTGCGG	Amorce pour hybridation double brin : fragment P'3 40 pb EMSA	
P'23WT-F	GCCGACCGAGCGGAATAACTAGAAAAGAGCTAGTTAATAGA	Amorce pour hybridation double brin : fragment P'23 40 pb EMSA	
P'23WT-R	TCTATTAAGTCTCTTAGTTATTCGCTCGGTGCGG	Amorce pour hybridation double brin : fragment P'23 40 pb EMSA	

Les cellules sont récoltées par centrifugation (6000g, 15 min, 4°C), lavées une fois dans 10 mL de tampon A, centrifugées (4500 g, 15 min, 4°C) et le culot repris dans 5 mL de tampon H (50 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole) contenant une tablette d'inhibiteur de protéase SIGMAFAST Protease Inhibitor Cocktail Tablets EDTA-Free. Les cellules sont lysées par sonication (40% d'intensité, 6 cycles de 30 sec et 6 cycles de pause de 30 sec dans la glace). Le lysat est centrifugé (16000g, 15 min, 4°C) puis le surnageant chauffé à 70 °C pendant 30 minutes. L'échantillon est centrifugé (16000g 15 min 4°C) puis le surnageant est récupéré et la protéine purifiée par passage successif sur deux colonnes en utilisant le système AKTA Prime Plus FPLC. La colonne HisTrap HP est équilibrée en injectant 10 volumes de colonne (10 mL) de tampon H. Les 5 mL de lysat sont injectés, et les protéines non fixées sont éluées par injection de 10 volumes de colonne de tampon H. La protéine 6xHis-HU_{L1a} est éluée avec un gradient (0 à 70%) de 20 mL de tampon I (50 mM Tris pH 8, 500 mM NaCl, 500 mM imidazole). L'éluution de la protéine est suivie par électrophorèse sur gel d'acrylamide 4-15% (Mini-PROTEAN TGX, Biorad) et par coloration au bleu de Coomassie. Les fractions contenant la protéine 6xHis-HU_{L1a} sont rassemblées et le tampon d'éluution remplacé par le tampon J (50 mM Tris pH 8, 200 mM NaCl) en utilisant une colonne d'ultrafiltration (Amicon Ultra-4, seuil de coupure de 3 kDa, Merck Millipore). Une colonne HiTrap Heparin HP est équilibrée en injectant 10 volumes de colonne de tampon J. L'échantillon contenant la protéine 6xHis-HU_{L1a} est injecté et les protéines non fixées sont éluées par injection de 10 volumes de colonne de tampon J. La protéine 6xHis-HU_{L1a} est éluée avec un gradient (0 à 100%) de 20 mL de tampon K (50 mM Tris pH8, 1 M NaCl) et l'éluution est suivie par électrophorèse sur gel d'acrylamide 4-15% (Mini-PROTEAN TGX, Biorad) après coloration au bleu de Coomassie. Les fractions contenant la protéine HU sont rassemblées et le tampon d'éluution remplacé par le tampon L (50 mM Tris pH 8, 150 mM NaCl) en utilisant une colonne d'ultrafiltration (Amicon Ultra-4, seuil de coupure de 3 kDa, Merck Millipore). Les échantillons sont aliquotés par volume de 20 µL, congelés dans l'azote liquide et stockés à -80°C. La concentration protéique est estimée par la mesure de l'absorbance à 230 nm (Nanodrop ND-2000, Thermofisher), en utilisant la correspondance suivante : une DO_{230nm} de 2,3 est égale à 1 mg/mL de HU (Pellegrini *et al.*, 2000).

Matériels et méthodes

Suite table 2.

Amorces	séquence 5'-3'	Description	Tm
P'13WT-F	TTTTTAAGTAGAACGACCGAGCGGAGAGCTAGTTAATAGA	Amorce pour hybridation double brin : fragment P'13 40 pb EMSA	
P'13WT-R	TCTATTAAGTAGCTCTCCGCTCGGTCTTAGTTAAAAA	Amorce pour hybridation double brin : fragment P'13 40 pb EMSA	
P12WT-F	CACGTCGTGATCAACTAGATTTTTAACTAGAAACCACGTC	Amorce pour hybridation double brin : fragment P12 40 pb EMSA	
P12WT-R	GACGTGGTTCTAGTTAAAAATCTAGTTGATCAGCAGCTG	Amorce pour hybridation double brin : fragment P12 40 pb EMSA	
P1WT-F	CACGTCGTGATCAACTAGATTCGACCGAGCGGACCACGTC	Amorce pour hybridation double brin : fragment P1 40 pb EMSA	
P1WT-R	GACGTGGTCCGCTCGGTGCAATCTAGTTGATCAGCAGCTG	Amorce pour hybridation double brin : fragment P1 40 pb EMSA	
P2WT-F	CACGTCGTGCGACCGAGCGGTTTTAACTAGAAACCACGTC	Amorce pour hybridation double brin : fragment P2 40 pb EMSA	
P2WT-R	GACGTGGTTCTAGTTAAAAACCCTCGGTGCGCAGCAGCTG	Amorce pour hybridation double brin : fragment P2 40 pb EMSA	
Nositebras-F	CACGTCGTGCGACCGAGCGGTGCGGCACGTCGACACGTC	Amorce témoin sans sites bras pour EMSA 40pb	
Nositebras-R	GACGTGGTCAGCGTGCCGCACCGCTCGGTGCGCAGCAGCTG	Amorce témoin sans sites bras pour EMSA 40pb	
P'12WT-40bp-R-Cy3	GACGTGGAATCTAGTTATTTCTAGTTAAAAACGACGCTG	Amorce site bras P'12 fluorescent pour EMSA	
COC'WT-R	GTTACGCAAGGAGAGTACAGGTTCTTTCCAATACC	Amorce site cœur COC' pour fragment COC' froid EMSA	
BOB'WT-F	GGTATTGTTCAAATCCTGTACTCTCTTGCCTAAC	Amorce pour hybridation double brin : fragment BOB'WT EMSA	
BOB'WT-R	GTTACGCAAGGAGAGTACAGGATTTGAACAATACC	Amorce pour hybridation double brin : fragment BOB'WT froid EMSA	
BOB'WT-R-Cy3	GTTACGCAAGGAGAGTACAGGATTTGAACAATACC	Amorce pour hybridation double brin : fragment BOB'WT fluorescent EMSA	
XOC'-F	GGTATTGGAACCCGCTGTACTCTCTTGCCTAAC	Amorce pour hybridation double brin : fragment XOC' : mutation site C pour EMSA	
XOC'-R-Cy3	GTTACGCAAGGAGAGTACAGGCGGGTTCCAATACC	Amorce pour hybridation double brin : fragment XOC' : mutation site C pour EMSA	
COX-F	GGTATTGGAAGAAGCTGTACCGGGCTTGCCTAAC	Amorce pour hybridation double brin : fragment COX : mutation site C' pour EMSA	
COX-R-Cy3	GTTACGCAAGCCCGGTACAGGTTCTTTCCAATACC	Amorce pour hybridation double brin : fragment COX : mutation site C' pour EMSA	
BOX-F	GGTATTGTTCAAATCCTGTACCGGGCTTGCCTAAC	Amorce pour hybridation double brin : fragment BOX : mutation site B' pour EMSA	
BOX-R-Cy3	GTTACGCAAGCCCGGTACAGGATTTGAACAATACC	Amorce pour hybridation double brin : fragment BOX : mutation site B' pour EMSA	
COC'OverlapAT-F	GGTATTGGAAGAACTGTATTCTCTTGCCTAAC	Amorce pour hybridation double brin : fragment COC' avec changement de base dans overlap :C1->A1 / C7->T7	
COC'OverlapAT-R-Cy3	GTTACGCAAGGAGAATACAGTTCTTTCCAATACC	Amorce pour hybridation double brin : fragment COC' avec changement de base dans overlap :C1->A1 / C7->T7	
P12-9pb-F	CACGTCGTGCTCAACTAGAGTCTTAACTAGAGACCACGTC	Amorce pour hybridation double brin : fragment P12 (sequence consensus du site muté de 11 a 9pb)40 pb EMSA	
P12-9pb-R	GACGTGGTCTCTAGTTAAGACTCTAGTTGAGCAGCAGCTG	Amorce pour hybridation double brin : fragment P12 (sequence consensus du site muté de 11 a 9pb)40 pb EMSA	
P12-6pb-F	CACGTCGTGCCGAAGTACGGTCCGAAGTACGGACCACGTC	Amorce pour hybridation double brin : fragment P12 (sequence consensus du site muté de 11 a 6pb)40 pb EMSA	
P12-6pb-R	GACGTGGTCCCTAGTTCCGACCTAGTTCCGACGACGCTG	Amorce pour hybridation double brin : fragment P12 (sequence consensus du site muté de 11 a 6pb)40 pb EMSA	
P'23forward-F	GCCGACCGAGCGGAATAACTAGAAAGATTAAGTAGTAGA	Amorce pour hybridation double brin : fragment P'23 40 pb EMSA avec site P'3 en sens forward et non reverse	
P'23forward-R	TCTAGTAGTTAATCTTTCTAGTTATTCGCTCGGTCCGC	Amorce pour hybridation double brin : fragment P'23 40 pb EMSA avec site P'3 en sens forward et non reverse	
P'12reverse-F	TTTTTAAGTAGAATCTAGTTATTAGCAGCGGACGCTCG	Amorce pour hybridation double brin : fragment P'12 40 pb EMSA avec site P'2 en sens reverse et non forward	
P'12reverse-R	CGACCGCTCGGTGCTAATAACTAGAAATCTAGTTAAAAA	Amorce pour hybridation double brin : fragment P'12 40 pb EMSA avec site P'2 en sens reverse et non forward	

V) Recombinaison *in vitro*

Les différents fragments *attB* fluorescents sont obtenus par amplification *in vitro* par PCR avec le programme suivant : 98°C 5 min ; 30 x (98°C 10 sec, 60°C 15 sec, 72°C 15 sec) ; 72°C 2 min. La réaction (50µL) contient 10 pg ADN matrice, 10 µM des amorces Lbbulg*attB*-F et Cy3-New-*attB*-R (Table 2), 800 µM de dNTP, du tampon Q5 polymérase 1X et 1 U de Q5-HF polymérase. Le fragment *attP* fluorescent est obtenu par PCR avec le programme suivant : 98°C 3 min ; 30 x (98°C 10 sec, 64°C 30 sec, 72°C 10 sec) ; 72°C 2 min. La réaction (50µL) contient 20 pg pMET340, 10 µM des amorces *attP*-F et *attP*-R-Cy3 (Table 2), 800 µM de dNTP, du tampon Q5 polymérase 1X et 1 U de Q5-HF polymérase. Le mélange réactionnel de recombinaison *in vitro* (20 µL) contient 0,08 pmol (15 ng) de fragment linéaire *attB* fluorescent, 0,08 pmol (300 ng) de plasmide comportant le site *attP*, 7,2 pmol (300 ng) de ^{mv4}Int, en présence ou en absence de 40 µg d'extrait cellulaire d'*E. coli* BL21(DE3) (préalablement chauffé à 95°C et centrifugé à 16000g, 15 min, 4°C), dans 25 mM Tris pH 7,5, 1 mM EDTA, 150 mM NaCl, 1 mM DTT, 10% PEG8000 (TENDP 1X). En absence d'extrait cellulaire d'*E. coli*, la réaction est incubée pendant 16h à 42°C, alors qu'en présence d'extrait cellulaire, la réaction est incubée pendant 1h30 à 42°C. La réaction est stoppée par l'addition de 0,1% de SDS, puis les ADN sont séparés par électrophorèse sur gel d'agarose 0,8% (TAE 0,5X). La fluorescence du Cy3 est révélée par excitation à 520-545 nm (ChemidocMP, Biorad) et quantifiée avec le logiciel ImageLab (Biorad).

VI) Construction et analyse des banques aléatoires *attB* et *attP*

VI.1) Construction de la souche NEB5- α RepA⁺

La cassette *glgB-repA-kan-glgB* est amplifiée par PCR à partir de l'ADN génomique de la souche *E. coli* EC1000 (*E. coli* EC101, *repA*⁺) (Leenhouts *et al.*, 1996) avec le programme suivant : 98°C 3 min ; 30 x (98°C 10 sec ; 56°C 30 sec ; 72°C 2 min) ; 72°C 2 min. La réaction (25 µL) contient : 1 ng ADNg, 0,5 µM du couple d'amorce 532-amont-glgb/533-aval-glgb (Table 2), 800µM de dNTP, du tampon Q5 polymérase 1X et 1 U Q5-HF polymérase. L'intégration est réalisée dans la souche NEB5- α ((F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ 80dlacZ Δ M15 Δ (lacZYA-argF)U169, hsdR17(rK-mK⁺), λ -), New England biolad) possédant le plasmide pKD46 par la technique Lambda Red en suivant le protocole de Datsenko and Warner (Datsenko and Wanner, 2000).

Matériels et méthodes

Suite table 2.

Amorces	séquence 5'-3'	Description	Tm
532-amont-glgB	AGTATTACAGCCTTACCTG	Amplification cassette glgB-repA-Kan-glgB	
533-aval-glgB	CAAGGATCAATCAACAACCTT	Amplification cassette glgB-repA-Kan-glgB	
534-5'kan	ATGAGAATATCACCGGAATTG	Amorce pour vérifier intégration de la cassette kanR-RepA au site glgB	
551-aval-glgB	TCCTTGACATGGGCTTCAT	Amorce pour vérifier intégration de la cassette kanR-RepA au site glgB	
535-3'kan	CGCAGAAGGCAATGTCATAC	Amorce pour vérifier intégration de la cassette kanR-RepA au site glgB	
550-amont-glgB	CTTCGTCAACTGGCGTAATC	Amorce pour vérifier intégration de la cassette kanR-RepA au site glgB	

La construction est vérifiée par PCR avec le programme : 98°C 3 min ; 30 x (98°C 10 sec, 60°C 30 sec, 72°C 2 min) ; 72°C 2 min. La réaction (25µL) contient : 2 ng ADNg, 0,5µM des couples d'amorce (532-amont-glgB/533-aval-glgB ; 534-5'kan/551-aval-glgB ; 535-3'kan/550-amont-glgB) (Table 2), 800µM de dNTP, du tampon Q5 polymérase 1X et 1 U de Q5-HF. Les produits PCR sont analysés par électrophorèse sur gel d'agarose 0,8% après coloration au bromure d'éthidium (BET) et la fluorescence est visualisée (ChemidocMP, Biorad).

VI.2) Construction des banques plasmidiques *attB* et *attP*

Les oligonucléotides de 109 pb (*attB*) et 184 pb (*attP*) constituant les différentes banques *attB* et *attP* (Table 2) sont obtenus par synthèse chimique (IDT, Etats-Unis). Ces oligonucléotides sont amplifiés par PCR avec le programme suivant : 98°C 5 min ; 10 x (98°C 10 sec, 63°C 30 sec, 72°C 10sec) ; 72°C 2 min. La réaction (50µL) contient 0,5 pmol (17 ng) de matrice de la banque, 10 µM de chaque couple d'amorce (attB-N10-N19-F/attB-N10-N19-R pour la banque *attB* et anti-attP-N10-N19-F/anti-attP-N10-N19-Vf-R pour la banque *attP*) (Table 2), 800 µM de dNTP, du tampon Q5 polymérase 1X et 1 U de Q5-HF polymérase. Le vecteur de la banque *attB* (pCC1Fos) est amplifié par PCR avec le programme suivant : 98°C 5 min ; 30 x (98°C 15 sec, 62°C 15 sec, 72°C 1 min) ; 72°C 2 min. La réaction (50 µL) contient : 50 pg plasmide pCC1Fos, 0,5 µM du couple d'amorce pCC1Fos-F/pCC1Fos-R (Table 2), et le « mix CloneAmp polymerase » 1 X. Le vecteur de la banque *attP* (pMET359) est amplifié par PCR avec le programme suivant : 98°C 5 min ; 30 x (98°C 10 sec ; 64°C 30 sec ; 72°C 3 min) ; 72°C 2 min. La réaction (50 µL) contient : 100 pg de matrice pMET359, 0,5µM du couple d'amorce attP-N10-N19-vf-F/attP-N10-N19-R (Table 2), 800 µM de dNTP, du tampon Q5 polymérase 1X et 1 U de Q5-HF polymérase. Les banques sont construites par clonage Gibson en utilisant le kit NEBuilder Hifi DNA Assembly (New England Biolabs) et en suivant les recommandations du fournisseur. Les banques *attB* sont transformées dans les cellules commerciales électrocompétentes *E. coli* EPI300 ((F⁻ λ⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 Δ(lac)X74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK rpsL (Str^R) nupG' trfA dhfr, Epicentre)) et propagées avec du chloramphénicol comme agent sélectif. Les banques *attP* sont transformées dans les cellules NEB-5α-*RepA* (Table 4). Ces cellules sont rendues compétentes en suivant le protocole de Wu *et al.*, (Wu *et al.*, 2010) et propagées avec de la carbénicilline comme agent sélectif. Les banques plasmidiques sont extraites en utilisant le kit NucleoBond Xtra Midi.

Matériels et méthodes

Table 3. Différents plasmides utilisés dans cette étude.

Plasmides	Description	Référence / source
pGhost9	Gène de résistance à l'erythromycine, origine de réplication pWV01, repA thermosensible	Maguin <i>et al.</i> , 1996
pBSattB	pBS-KS::attB WT <i>Lb. delbrueckii</i> subsp. <i>bulgaricus</i> résistance à l'ampicilline	Dupont <i>et al.</i> , 1995
pMC1	pRC1::mv4Int, attPWT résistance à l'erythromycine, origine de réplication colE1	Dupont <i>et al.</i> , 1995
pET-Int	pET15b::mv4Int	Auvray <i>et al.</i> , 1999a
pMET301	pMC1 : mutation région cœur attP par clonage gibson (adaptation mutation pMET318) GAAAGAA CCTGTAC TCTCCTGAAT	Cette étude
pMET302	pMC1 : mutation région cœur attP par clonage gibson (adaptation ARNTSer <i>Lb. acidophilus</i>) GAAAGAA CCTGTAC ACTCCTTTT	Cette étude
pMET303	pMC1 : mutation région cœur attP par clonage gibson (adaptation ARNTSer <i>E. coli</i>) GAAAGAA CCTCTCT CCGCCACTTT	Cette étude
pMET304	pMC1 : mutation région cœur attP par clonage gibson (adaptation ARNTSer <i>Lb. mesenteroides</i>) GAAAGAA CCCGCTA GCTCCTTAT	Cette étude
pMET305	pMC1 : mutation région cœur attP par clonage gibson (adaptation ARNTSer <i>E. faecalis</i>) GAAAGAA CCTTAT CCTCCGTACC	Cette étude
pMET306	pMC1 : mutation région cœur attP par clonage gibson (adaptation ARNTSer <i>L. lactis</i>) GAAAGAA CCCCTTG ACTCCTTTA	Cette étude
pMET307	pMC1 : mutation région cœur attP par clonage gibson (adaptation ARNTSer <i>S. mutans</i>) GAAAGAA CCGTCC TTCCTTAAC	Cette étude
pMET308	pMC1 : mutation région cœur attP par clonage gibson (adaptation ARNTSer <i>Lb. reuteri</i>) GAAAGAA CCCCTTG TCTCCTTAGA	Cette étude
pMET309	pMC1 : mutation région cœur attP par clonage gibson (adaptation <i>Lb. reuteri</i> 2) GAAAGAA CCCCTTG TCTCCATAGA	Cette étude
pMET310	pMC1 : mutation région cœur attP par clonage gibson (adaptation ARNTSer <i>Lb. sakei</i>) GAAAGAA CCTGTAC TCTCCTTTA	Cette étude
pMET311	pMC1 IS10::mv4Int	Cette étude
pMET317	pBSattB : mutation attB par clonage gibson (mutation pAMattB4, Auvray <i>et al.</i> , 1999b) TTCAAGA CCTGTAC TCTCCTAAT	Cette étude
pMET318	pBSattB : mutation attB par clonage gibson (mutation pAMattB5, Auvray <i>et al.</i> , 1999b) TTCAAAT CCTGTAC TCTCCTGAAT	Cette étude
pMET320	pBSattB : mutation attB par clonage gibson (mutation ARNTSer <i>Lb. acidophilus</i>) TTCAAAT CCTGTAC ACTCCTTTT	Cette étude
pMET321	pBSattB : mutation attB par clonage gibson (mutation ARNTSer <i>E. coli</i>) TTCAAAT CCTCTCT CCGCCACTTT	Cette étude
pMET322	pBSattB : mutation attB par clonage gibson (mutation ARNTSer <i>E. faecalis</i>) TTCAAAT CCTTAT CCTCCGTACC	Cette étude
pMET323	pBSattB : mutation attB par clonage gibson (mutation ARNTSer <i>L. lactis</i>) TTCAAAT CCCCTTG ACTCCTTTA	Cette étude
pMET324	pBSattB : mutation attB par clonage gibson (mutation ARNTSer <i>S. mutans</i>) TTCAAAT CCGTCC TTCCTTAAC	Cette étude
pMET325	pBSattB : mutation attB par clonage gibson (mutation ARNTSer <i>Ln. mesenteroides</i>) TTCAAAT CCCGCTA GCTCCTTAT	Cette étude
pMET326	pBSattB : mutation attB par clonage gibson (mutation ARNTSer <i>Lb. reuteri</i>) TTCAAAT CCCCTTG TCTCCTTAGA	Cette étude
pMET327	pBSattB : mutation attB par clonage gibson (mutation ARNTSer <i>Lb. sakei</i>) TTCAAAT CCTGTAC TCTCCTTTA	Cette étude
pMET328	pBSattB : mutation attB par clonage gibson (mutation <i>Lb. reuteri</i> 2) TTCAAAT CCCCTTG TCTCCATAGA	Cette étude
pMET332	pET-Int Δ6xHisORF-8	Cette étude
pMET334	pET-Int fusion ORF-8 / mv4Int	Cette étude
pMET337	pET-Int ΔORF-8 : permet la fusion de l'étiquette 6xHis au domaine N-ter de mv4Int	Cette étude
pMET338	pMET332 mutation mv4Int Y349->F	Cette étude
pMET339	pMET332 mutation mv4Int K248->A	Cette étude
pMET340	pMC1 avec Ori colE1 remplacée par Ori pWV01 sans repA	Cette étude
pCC1Fos	pCC1Fos gène de résistance au chloramphénicol, Ori II et Ori V	Epicentre
pMET348	pCC1Fos::attB WT pBSattB	Cette étude
pMET349	pCC1Fos:: 23 pb attBWT GGTTCAAAT CCTGTAC TCTCCTAAT	Cette étude
pMET351	pMET348 : mutation attB par clonage gibson (mutation T23->A23) GGTTCAAAT CCTGTAC TCTCCTAAAT	Cette étude

VI.3) Recombinaison *in vitro* et séquençage des banques aléatoires

Le mélange réactionnel (20 µL) contient 0,08 pmol de plasmide *attB*, 0,08 pmol de plasmide *attP*, 7,2 pmol de ^{mv4}Int, 40 µg d'extrait cellulaire d'*E. coli* BL21(DE3) préalablement chauffé à 95°C dans 25 mM Tris pH 7,5, 1 mM EDTA, 150 mM NaCl, 1 mM DTT, 10% PEG8000 (TENDP 1X). La réaction est incubée 1h30 à 42°C. Les sites *attB*, *attP*, *attL* et *attR* sont amplifiés par PCR avec le programme suivant : 98°C 5 min ; 30 x (98°C 10 sec, 62°C 30 sec, 72°C 10 sec) ; 72°C 2 min. La réaction (50 µL) contient : 1 µL du produit de recombinaison, 0,5µM des couples d'amorce (Seqbanque*attB*-F/seqbanque*attB*-R pour *attB* ; *attP*-F/seqbanque*attP*-R pour *attP* ; seqbanque*attB*-F/seqbanque*attL*-R pour *attL* et seqbanque*attR*-F/seqbanque*attB*-R pour *attR*), 800µM de dNTP, du tampon Q5 polymérase 1X et 1 U de Q5-HF. Les différents produits PCR sont analysés par séquençage Sanger (Mix2seq, Eurofins).

Pour 4 échantillons (*attL* recombinaison banque *attB* Lib6 x *attP*_{WT}, appelé *attL* Lib6 ; *attL* recombinaison *attB*_{WT} x banque *attP* Lib1, appelé *attL* Lib1 ; *attR* recombinaison *attB*_{WT} x banque *attP* Lib3, appelé *attR* Lib3 ; *attR* banque *attB* Lib8 x *attP*_{WT}, appelé *attR* Lib8), 75 ng de chaque produit ont été rassemblés et séquencés par séquençage NGS (Eurofins). Pour des raisons de coût, les 4 échantillons ont été séquencés en un seul séquençage.

VII) Analyse bioinformatique du séquençage NGS

Les deux fichiers fastQ (R1 et R2) ont été appariés et les séquences ont été fusionnées en utilisant le logiciel Genomics Workbench. Les différents programmes utilisés par la suite proviennent du site Galaxy (<https://usegalaxy.org>). Le fichier contenant les séquences fusionnées possède les séquences de 4 échantillons différents. Ces échantillons sont séparés en 4 fichiers en triant les séquences par la taille : *attL* Lib1 (150 pb), *attR* Lib8 (182 pb), *attL* Lib6 (196 pb) et *attR* Lib3 (264 pb) en utilisant le programme Filter FASTA. Afin d'éliminer les séquences provenant d'un autre échantillon, chaque fichier est trié en fonction de sa séquence nucléotidique en utilisant le programme Bowtie2 (Langmead and Salzberg, 2012). Chaque fichier contient un mélange de séquences « forward » et « reverse ». Pour chaque fichier, les séquences « forward » et « reverse » sont séparées en 2 fichiers grâce au programme Filter FASTA. Les séquences « reverse » sont complémentées-inversées en utilisant le programme Seqtk_seq. Ce programme a aussi servi à éliminer les séquences « forward » et « reverse complement » contenant des nucléotides ambigus (hors A/T/G/C).

Matériels et méthodes

Suite table 3.

Plasmides	Description	Référence / source
pMET351	pMET348 : mutation attB par clonage gibson (mutation T23-->A23) GGTTCAAAT CCTGTAC TCTCCTAAAT	Cette étude
pMET352	pMET348 : mutation attB par clonage gibson (mutation A7-->C7 / A8-->G8 / T9-->A9) GGTTCA CGA CCTGTAC TCTCCTTAAT	Cette étude
pMET353	pMET348 : mutation attB par clonage gibson (mutation T9-->C9) GGTTCAAAC CCTGTAC TCTCCTTAAT	Cette étude
pMET354	pMET348 : mutation attB par clonage gibson (mutation T9-->G9) GGTTCAAAG CCTGTAC TCTCCTTAAT	Cette étude
pMET355	pMET348 : mutation attB par clonage gibson (mutation A7-->T7) GGTTCA TAT CCTGTAC TCTCCTTAAT	Cette étude
pMET356	pMET348 : mutation attB par clonage gibson (mutation A7-->C7) GGTTCA CAT CCTGTAC TCTCCTTAAT	Cette étude
pMET357	pCC1Fos:: 16 pb attBWT AT CCTGTAC TCTCCTTAAT	Cette étude
pMET358	pMET348 : mutation attB par clonage gibson (mutation T23-->G23 / A24-->G24 / A25-->G25 / T26--> A26) GGTTCAAAT CCTGTAC TCTCCT GGGA	Cette étude
pMET359	pMET340::bla	Cette étude
pFSKB3X	gène de résistance à la kanamycine, ori colE1, promoteur T7 : pour expression protéine	GTP technology
pMET364	pMET348 : mutation attB par clonage gibson (mutation région centrale par séquence overlap putatif site 1 lacZ) TTCAAAT TTTCGGC TCTCCTT	Cette étude
pMET365	pMET348 : mutation attB par clonage gibson (mutation région centrale par séquence overlap putatif site 3 lacZ) TTCAAAT AAGGTTT TCTCCTT	Cette étude
pMET366	pMET348 : mutation attB par clonage gibson (mutation BOB' par séquence BOB' putatif site 1 lacZ) TT CGGGA TTTCGGC GCTCCAC	Cette étude
pMET367	pMET348 : mutation attB par clonage gibson (mutation BOB' par séquence BOB' putatif site 2 lacZ) GT TAAAT TGCCAAC GCTTAT T	Cette étude
pMET368	pMET348 : mutation attB par clonage gibson (mutation BOB' par séquence BOB' putatif site 3 lacZ) GATA AAAT AAGGTTT TCCCTG	Cette étude
pMET369	pFSKB3X::HU L. lactis (tag 6xHis en N-ter)	Cette étude
pMET370	pMET348 : mutation attB par clonage gibson (mutation région centrale par séquence overlap putatif site 3 lacZ) TTCAAAT TGCCAAC TCTCCTT	Cette étude
pMET371	pMET359 : mutation région cœur attP par clonage gibson (mutation région centrale par séquence overlap putatif site 1 lacZ) GAAAGAA TTTCGGC TCTCCTT	Cette étude
pMET372	pMET359 : mutation région cœur attP par clonage gibson (mutation région centrale par séquence overlap putatif site 2 lacZ) GAAAGAA TGCCAAC TCTCCTT	Cette étude
pMET373	pMET359 : mutation région cœur attP par clonage gibson (mutation région centrale par séquence overlap putatif site 2 lacZ) GAAAGAA AAGGTTT TCTCCTT	Cette étude
pMET374	pMET359 : mutation région cœur attP par clonage gibson (mutation COC' par séquence COC' putatif site 1 lacZ) TTCGGGA TTTCGGC GCTCCAC	Cette étude
pMET375	pMET359 : mutation région cœur attP par clonage gibson (mutation COC' par séquence COC' putatif site 2 lacZ) GT TAAAT TGCCAAC GCTTAT T	Cette étude

Les fichiers « forward » et « reverse-complement » de chaque échantillon sont fusionnés en utilisant le programme Merge.files. Pour les 4 fichiers, le programme FASTA Width formatter est utilisé pour que chaque séquence fasta appartienne à une seule ligne. A partir de cette étape, chaque fichier contient uniquement des séquences à une taille précise (150 pb, 182 pb, 196 pb, 264 pb, en fonction de l'échantillon de départ), dans un sens précis (« forward ») et chaque séquence est écrite sur une seule ligne. Dans chacun de ces 4 fichiers, les séquences de 7 pb d'intérêt sont extraites par le programme Trim sequences (Gordon and Hannon, 2010). Pour chaque fichier, un « *Séquence Logo* » est créé en utilisant le programme Sequence Logo generator for fasta (Crooks *et al.*, 2004) et l'occurrence de chaque séquence est obtenue en utilisant le programme Wordcount. Les fichiers textes issus du programme Wordcount sont analysés sur le logiciel Excel (Microsoft Office).

VIII) Retard sur gel

VIII.1) Amorces

Les différentes amorces utilisées (Eurofins) sont indiquées Table 2. Les amorces contenant le fluorophore Cy3 sont purifiées par HPLC (Eurofins). Les brins complémentaires sont hybridés en ADN double-brin dans 10 mM de Tris pH 7,5, 50 mM NaCl en incubant l'échantillon à 95°C pendant 5 min dans un thermocycleur, et en diminuant la température de 1,5°C/min jusqu'à atteindre 25°C. L'hybridation est vérifiée par électrophorèse en gel d'agarose 3% (QA-Agarose, MP Biomedicals), après coloration au BET et visualisation de la fluorescence (ChemidocMP, Biorad).

VIII.2) Interaction et électrophorèse des complexes ADN-protéines

La réaction (20 µL) est effectuée dans 25 mM Tris pH 8, 75 mM NaCl, 10% glycérol, 0,5 mM DTT, 0,5 mM EDTA, 1 µg polydIdC (Sigma), 0,1 mg/mL BSA, 0,87 pmol d'ADN fluorescent et 4,48 pmol d'ADN non marqué. La réaction est incubée 20 min à température ambiante et les échantillons analysés par électrophorèse sur gel d'acrylamide non dénaturant 7,5% (Mini-PROTEAN TGX, Biorad). La migration est effectuée dans du tampon Tris/Glycine (25 mM Tris pH 8,3, 192 mM Glycine, Biorad), à 4°C, 75 V pendant 2h. La fluorescence du Cy3 est révélée par excitation à 520-545 nm (ChemidocMP, Biorad).

Matériels et méthodes

Table 4. Différentes souches utilisées dans cette étude.

Souches	Description	Référence
<i>E. coli</i> BL21(DE3)	B F–ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacIacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)	New England Biolabs
<i>E. coli</i> EC1000	<i>E. coli</i> EC101, repA +	Leenhouts <i>et al.</i> , 1996
<i>E. coli</i> Neb5α	F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ–	New England Biolabs
<i>E. coli</i> EPI300	F– λ– mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 Δ(lac)X74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK rpsL (StrR) nupG' trfA dhfr	Epicentre
<i>E. coli</i> AG1	endA1 recA1 gyrA96 thi-1 relA1 glnV44 hsdR17(rK- mK+)	Agilent
<i>E. coli</i> Neb5α-RepA	<i>E. coli</i> Neb5α::repA	Cette étude
<i>E. coli</i> BW25113	lacI+rrnBT14 ΔlacZΔWJ16 hsdR514 ΔaraBADAH33 ΔrhaBADLD78 rph-1 Δ(araB–D)567 Δ(rhaD–B)568 ΔlacZ4787(::rrnB-3) hsdR514 rph-1	Datsenko and Wanner, 2000
<i>L. lactis</i> MG1363		Gasson, 1983

Table 5. Différents tampons testés lors des tests de recombinaison *in vitro*.

	Tampon					
	Tris ou Hepes pH 7,5	Spermidine	MgCl ₂ ou EDTA	PEG 8000	DTT	NaCl
Condition 1	25 mM Tris pH 7,5	4 mM	5 mM MgCl ₂	10%	1 mM	150 mM
Condition 2	25 mM Tris pH 7,5	-	5 mM MgCl ₂	10%	1 mM	150 mM
Condition 3	25 mM Tris pH 7,5	-	1 mM EDTA	10%	1 mM	150 mM
Condition 4	25 mM Tris pH 7,5	-	1 mM EDTA	-	1 mM	150 mM
Condition 5	25 mM Tris pH 7,5	-	5 mM MgCl ₂	-	1 mM	150 mM
Condition 6	25 mM Tris pH 7,5	4 mM	1 mM EDTA	10%	1 mM	150 mM
Condition 7	25 mM Tris pH 7,5	4 mM	1 mM EDTA	-	1 mM	150 mM
Condition 8	25 mM Tris pH 7,5	4 mM	5 mM MgCl ₂	-	1 mM	150 mM
Condition 9	25 mM Hepes pH 7,5	4 mM	5 mM MgCl ₂	10%	1 mM	150 mM
Condition 10	25 mM Hepes pH 7,5	-	5 mM MgCl ₂	10%	1 mM	150 mM
Condition 11	25 mM Hepes pH 7,5	-	1 mM EDTA	10%	1 mM	150 mM
Condition 12	25 mM Hepes pH 7,5	-	1 mM EDTA	-	1 mM	150 mM
Condition 13	25 mM Hepes pH 7,5	-	5 mM MgCl ₂	-	1 mM	150 mM
Condition 14	25 mM Hepes pH 7,5	4 mM	1 mM EDTA	10%	1 mM	150 mM
Condition 15	25 mM Hepes pH 7,5	4 mM	1 mM EDTA	-	1 mM	150 mM
Condition 16	25 mM Hepes pH 7,5	4 mM	5 mM MgCl ₂	-	1 mM	150 mM

IX) Recombinaison *in vivo*

IX.1) Recombinaison *in vivo* chez *L. lactis*

Des cellules de la souche *L. lactis* MG1363 sont rendues compétentes et transformées par électropéabilisation (Le Bourgeois *et al.*, 2000) par 1 µg de pMC1 ou 1 µg de pMET306 (Table 3). Les cellules sont étalées sur milieu solide (GM17 agar) supplémenté avec 1 µg/mL d'érythromycine. L'intégration par recombinaison spécifique de site est vérifiée par PCR sur les ADNg des « intégrants » avec les conditions suivantes : 95°C 30 sec ; 30 x (95°C 30 sec, 59°C 60 sec, 68°C 6 min) ; 68°C 5 min. La réaction (50 µL) contient : 800 µM de dNTP, du tampon Onetaq standard à une concentration de 1 X, 5 ng d'ADN génomique, 1,25 U Onetaq et 0,5 µM des couples d'amorce (soit *attBlactis-F/attBlactis-R* ; *attBlactis-F/pMC1sanscolE1-R* ; *pMC1seqprim2-F/attBlactis-R*) (Table 2). Les produits PCR sont analysés par électrophorèse sur gel d'agarose 0,8% après coloration au BET et visualisation de la fluorescence (ChemidocMP, Biorad).

IX.2) Recombinaison *in vivo* chez *E. coli*

Des cellules *E. coli* AG1 ((*endA1 recA1 gyrA96 thi-1 relA1 glnV44 hsdR17(rK- mK+)*, Agilent)) sont transformées par 500 ng de plasmides non répliatif pMET359, pMET371, pMET372 et pMET373 (Table 3) en suivant les recommandations du fournisseur. Après transformation, les cellules ont été incubées 5h à 37°C sous agitation (200 rpm). Les cellules sont étalées sur milieu solide (LB agar) supplémenté avec 100 µg/mL de carbénicilline.

Références bibliographiques

- Abbani, M.A., Papagiannis, C.V., Sam, M.D., Cascio, D., Johnson, R.C., and Clubb, R.T. (2007). Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc Natl Acad Sci USA* *104*, 2109–2114.
- Abremski, K., and Gottesman, S. (1982). Purification of the bacteriophage lambda *xis* gene product required for lambda excisive recombination. *J Biol Chem* *257*, 9658–9662.
- Abremski, K., and Hoess, R. (1984). Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem* *259*, 1509–1514.
- Abremski, K.E., and Hoess, R.H. (1992). Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng* *5*, 87–91.
- Aihara, H., Kwon, H.J., Nunes-Düby, S.E., Landy, A., and Ellenberger, T. (2003). A conformational switch controls the DNA cleavage activity of lambda integrase. *Mol Cell* *12*, 187–198.
- Akabayov, B., Akabayov, S.R., Lee, S.-J., Wagner, G., and Richardson, C.C. (2013). Impact of macromolecular crowding on DNA replication. *Nat Commun* *4*, 1615.
- Albert, H., Dale, E.C., Lee, E., and Ow, D.W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J* *7*, 649–659.
- Alen, C., Sherratt, D.J., and Colloms, S.D. (1997). Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination. *EMBO J* *16*, 5188–5197.
- Allison, S.A., Herr, J.C., and Schurr, J.M. (1981). Structure of viral phi 29 DNA condensed by simple triamines: a light-scattering and electron-microscopy study. *Biopolymers* *20*, 469–488.
- Alonso, J.C., Weise, F., and Rojo, F. (1995). The *Bacillus subtilis* histone-like protein Hbsu is required for DNA resolution and DNA inversion mediated by the beta recombinase of plasmid pSM19035. *J Biol Chem* *270*, 2938–2945.
- Andreas, S., Schwenk, F., Küter-Luks, B., Faust, N., and Kühn, R. (2002). Enhanced efficiency through nuclear localization signal fusion on phage ϕ C31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. *Nucleic Acids Res* *30*, 2299–2306.

- Argos, P., Landy, A., Abremski, K., Egan, J.B., Haggard-Ljungquist, E., Hoess, R.H., Kahn, M.L., Kalionis, B., Narayana, S.V., and Pierson, L.S. (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J* 5, 433–440.
- Aussel, L., Barre, F.-X., Aroyo, M., Stasiak, A., Stasiak, A.Z., and Sherratt, D. (2002). FtsK is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell* 108, 195–205.
- Auvray, F., Coddeville, M., Ritzenthaler, P., and Dupont, L. (1997). Plasmid integration in a wide range of bacteria mediated by the integrase of *Lactobacillus delbrueckii* bacteriophage mv4. *J Bacteriol* 179, 1837–1845.
- Auvray, F., Coddeville, M., Espagno, G., and Ritzenthaler, P. (1999a). Integrative recombination of *Lactobacillus delbrueckii* bacteriophage mv4: functional analysis of the reaction and structure of the *attP* site. *Mol Gen Genet* 262, 355–366.
- Auvray, F., Coddeville, M., Ordonez, R.C., and Ritzenthaler, P. (1999b). Unusual structure of the *attB* site of the site-specific recombination system of *Lactobacillus delbrueckii* bacteriophage mv4. *J Bacteriol* 181, 7385–7389.
- Azaro, M.A., and Landy, A. (2002). λ integrase and the λ Int family. In *Mobile DNA II*, N.L. Craig, R. Craigie, M. Gellert, and A. Lambowitz, eds. (Washington, DC, USA: ASM Press), pp. 118–148.
- Bai, H., Sun, M., Ghosh, P., Hatfull, G.F., Grindley, N.D.F., and Marko, J.F. (2011). Single-molecule analysis reveals the molecular bearing mechanism of DNA strand exchange by a serine recombinase. *Proc Natl Acad Sci USA* 108, 7419–7424.
- Ball, C.A., and Johnson, R.C. (1991a). Efficient excision of phage lambda from the *Escherichia coli* chromosome requires the Fis protein. *J Bacteriol* 173, 4027–4031.
- Ball, C.A., and Johnson, R.C. (1991b). Multiple effects of Fis on integration and the control of lysogeny in phage lambda. *J Bacteriol* 173, 4032–4038.
- Ball, C.A., Osuna, R., Ferguson, K.C., and Johnson, R.C. (1992). Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *J Bacteriol* 174, 8043–8056.

- Bauer, C.E., Gardner, J.F., and Gumport, R.I. (1985). Extent of sequence homology required for bacteriophage lambda site-specific recombination. *J Mol Biol* 181, 187–197.
- Bauer, C.E., Hesse, S.D., Gumport, R.I., and Gardner, J.F. (1986). Mutational analysis of integrase arm-type binding sites of bacteriophage lambda: Integration and excision involve distinct interactions of integrase with arm-type sites. *J Mol Biol* 192, 513–527.
- Bedzyk, L.A., Shoemaker, N.B., Young, K.E., and Salyers, A.A. (1992). Insertion and excision of *Bacteroides* conjugative chromosomal elements. *J Bacteriol* 174, 166–172.
- Berntsson, R.P.-A., Odegrip, R., Sehlén, W., Skaar, K., Svensson, L.M., Massad, T., Högbom, M., Haggård-Ljungquist, E., and Stenmark, P. (2014). Structural insight into DNA binding and oligomerization of the multifunctional Cox protein of bacteriophage P2. *Nucleic Acids Res* 42, 2725–2735.
- Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol* 62, 293–300.
- Better, M., Lu, C., Williams, R.C., and Echols, H. (1982). Site-specific DNA condensation and pairing mediated by the int protein of bacteriophage lambda. *Proc Natl Acad Sci USA* 79, 5837–5841.
- Bigot, S., Saleh, O.A., Cornet, F., Allemand, J.-F., and Barre, F.-X. (2006). Oriented loading of FtsK on KOPS. *Nat Struct Mol Biol* 13, 1026–1028.
- Bikard, D., Julié-Galau, S., Cambray, G., and Mazel, D. (2010). The synthetic integron: an in vivo genetic shuffling device. *Nucleic Acids Res* 38, e153.
- Bischerour, J., Spangenberg, C., and Barre, F.-X. (2012). Holliday junction affinity of the base excision repair factor Endo III contributes to cholera toxin phage integration. *EMBO J* 31, 3757–3767.
- Bischof, J., and Basler, K. (2008). Recombinases and their use in gene activation, gene inactivation, and transgenesis. *Methods Mol Biol* 420, 175–195.
- Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., and Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by λ integrase. *Nature* 435, 1059–1066.

- Blakely, G., Colloms, S., May, G., Burke, M., and Sherratt, D. (1991). *Escherichia coli* XerC recombinase is required for chromosomal segregation at cell division. *New Biol* 3, 789–798.
- Blakely, G., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T., and Sherratt, D.J. (1993). Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell* 75, 351–361.
- Bordeleau, E., Ghinet, M.G., and Burrus, V. (2012). Diversity of integrating conjugative elements in actinobacteria. *Mob Genet Elements* 2, 119–124.
- Boucher, Y., Nesbø, C.L., Joss, M.J., Robinson, A., Mabbutt, B.C., Gillings, M.R., Doolittle, W.F., and Stokes, H.W. (2006). Recovery and evolutionary analysis of complete integron gene cassette arrays from *Vibrio*. *BMC Evol Biol* 6, 3.
- 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.
- 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.
- Bregu, M., Sherratt, D.J., and Colloms, S.D. (2002). Accessory factors determine the order of strand exchange in Xer recombination at *psi*. *EMBO J* 21, 3888–3897.
- Broach, J.R., and Hicks, J.B. (1980). Replication and recombination functions associated with the yeast plasmid, 2 mu circle. *Cell* 21, 501–508.
- Brown, W.R.A., Lee, N.C.O., Xu, Z., and Smith, M.C.M. (2011). Serine recombinases as tools for genome engineering. *Methods* 53, 372–379.
- Bruist, M.F., Glasgow, A.C., Johnson, R.C., and Simon, M.I. (1987). Fis binding to the recombinational enhancer of the Hin DNA inversion system. *Genes Dev* 1, 762–772.
- Buchholz, F., Ringrose, L., Angrand, P.-O., Rossi, F., and Stewart, A.F. (1996). Different thermostabilities of FLP and Cre recombinases: implications for applied site-specific recombination. *Nucleic Acids Res* 24, 4256–4262.

- Buchholz, F., Angrand, P.-O., and Stewart, A.F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* 16, 657–662.
- Bukhari, A.I., and Zipser, D. (1972). Random insertion of Mu-1 DNA within a single gene. *Nat New Biol* 236, 240–243.
- Bushman, W., Yin, S., Thio, L.L., and Landy, A. (1984). Determinants of directionality in lambda site-specific recombination. *Cell* 39, 699–706.
- Bushman, W., Thompson, J.F., Vargas, L., and Landy, A. (1985). Control of directionality in lambda site specific recombination. *Science* 230, 906–911.
- Cambray, G., Guerout, A.M., and Mazel, D. (2010). Integrons. *Annu. Rev. Genet.* 44, 141–166.
- Campbell, A.M. (1963). Episomes. *Adv Genet* 11, 101–145.
- Cao, Y., and Hayes, F. (1999). A newly identified, essential catalytic residue in a critical secondary structure element in the integrase family of site-specific recombinases is conserved in a similar element in eucaryotic type IB topoisomerases. *J Mol Biol* 289, 517–527.
- Carnoy, C., and Roten, C.-A. (2009). The *dif/Xer* recombination systems in proteobacteria. *PLOS ONE* 4, e6531.
- Chattoraj, D.K., Gosule, L.C., and Schellman, A. (1978). DNA condensation with polyamines. II. Electron microscopic studies. *J Mol Biol* 121, 327–337.
- Chen, Y., and Rice, P.A. (2003). The role of the conserved Trp330 in Flp-mediated recombination. Functional and structural analysis. *J Biol Chem* 278, 24800–24807.
- Chen, J.-W., Lee, J., and Jayaram, M. (1992). DNA cleavage in *trans* by the active site tyrosine during Flp recombination: Switching protein partners before exchanging strands. *Cell* 69, 647–658.
- Chen, Y., Narendra, U., Iype, L.E., Cox, M.M., and Rice, P.A. (2000). Crystal structure of a Flp recombinase–Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell* 6, 885–897.

- Cheng, C., Kussie, P., Pavletich, N., and Shuman, S. (1998). Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell* 92, 841–850.
- Cheng, Q., Paszkiet, B.J., Shoemaker, N.B., Gardner, J.F., and Salyers, A.A. (2000). Integration and excision of a *Bacteroides* conjugative transposon, CTnDOT. *J Bacteriol* 182, 4035–4043.
- Cho, E.H., Gumport, R.I., and Gardner, J.F. (2002). Interactions between integrase and excisionase in the phage lambda excisive nucleoprotein complex. *J Bacteriol* 184, 5200–5203.
- Cluzel, P.J., Veaux, M., Rousseau, M., and Accolas, J.P. (1987). Evidence for temperate bacteriophages in two strains of *Lactobacillus bulgaricus*. *J Dairy Res* 54, 397–405.
- Coddeville, M., and Ritzenthaler, P. (2010). Control of directionality in bacteriophage mv4 site-specific recombination: functional analysis of the Xis factor. *J Bacteriol* 192, 624–635.
- Coddeville, M., Auvray, F., Mikkonen, M., and Ritzenthaler, P. (2007). Single independent operator sites are involved in the genetic switch of the *Lactobacillus delbrueckii* bacteriophage mv4. *Virology* 364, 256–268.
- Coddeville, M., Spinella, J.-F., Cassart, P., Girault, G., Daveran-Mingot, M.-L., Le Bourgeois, P., and Ritzenthaler, P. (2014). Bacteriophage mv4 site-specific recombination: the central role of the P2 mv4Int-binding site. *J Virol* 88, 1839–1842.
- Collis, C.M., and Hall, R.M. (1992a). Gene cassettes from the insert region of integrons are excised as covalently closed circles. *Mol Microbiol* 6, 2875–2885.
- Collis, C.M., and Hall, R.M. (1992b). Site-specific deletion and rearrangement of integron insert genes catalyzed by the integron DNA integrase. *J Bacteriol* 174, 1574–1585.
- Collis, C.M., Grammaticopoulos, G., Briton, J., Stokes, H.W., and Hall, R.M. (1993). Site-specific insertion of gene cassettes into integrons. *Mol Microbiol* 9, 41–52.
- 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.

- Colloms, S.D., Sykora, P., Szatmari, G., and Sherratt, D.J. (1990). Recombination at ColE1 *cer* requires the *Escherichia coli xerC* gene product, a member of the lambda integrase family of site-specific recombinases. *J Bacteriol* *172*, 6973–6980.
- Colloms, S.D., McCulloch, R., Grant, K., Neilson, L., and Sherratt, D.J. (1996). Xer-mediated site-specific recombination in vitro. *EMBO J* *15*, 1172–1181.
- Colloms, S.D., Bath, J., and Sherratt, D.J. (1997). Topological selectivity in Xer site-specific recombination. *Cell* *88*, 855–864.
- Colloms, S.D., Alén, C., and Sherratt, D.J. (1998). The ArcA/ArcB two-component regulatory system of *Escherichia coli* is essential for Xer site-specific recombination at *psi*. *Mol Microbiol* *28*, 521–530.
- Colloms, S.D., Merrick, C.A., Olorunniji, F.J., Stark, W.M., Smith, M.C.M., Osbourn, A., Keasling, J.D., and Rosser, S.J. (2014). Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination. *Nucleic Acids Res* *42*, e23–e23.
- Connolly, K.M., Iwahara, M., and Clubb, R.T. (2002). Xis protein binding to the left arm stimulates excision of conjugative transposon Tn916. *J Bacteriol* *184*, 2088–2099.
- Cornet, F., Mortier, I., Patte, J., and Louarn, J.M. (1994). Plasmid pSC101 harbors a recombination site, *psi*, which is able to resolve plasmid multimers and to substitute for the analogous chromosomal *Escherichia coli* site *dif*. *J Bacteriol* *176*, 3188–3195.
- Craig, N.L., and Nash, H.A. (1983). The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell* *35*, 795–803.
- Craig, N.L., and Nash, H.A. (1984). *E. coli* integration host factor binds to specific sites in DNA. *Cell* *39*, 707–716.
- Crisona, N.J., Weinberg, R.L., Peter, B.J., Sumners, D.W., and Cozzarelli, N.R. (1999). The topological mechanism of phage lambda integrase. *J Mol Biol* *289*, 747–775.
- Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res* *14*, 1188–1190.

- Das, B., Bischerour, J., and Barre, F.-X. (2011). VGJ ϕ integration and excision mechanisms contribute to the genetic diversity of *Vibrio cholerae* epidemic strains. *Proc Natl Acad Sci USA* *108*, 2516–2521.
- Das, B., Martínez, E., Midonet, C., and Barre, F.-X. (2013). Integrative mobile elements exploiting Xer recombination. *Trends Microbiol* *21*, 23–30.
- Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* *97*, 6640–6645.
- DiChiara, J.M., Mattis, A.N., and Gardner, J.F. (2007). IntDOT interactions with core- and arm-type sites of the conjugative transposon CTnDOT. *J Bacteriol* *189*, 2692–2701.
- Dillon, S.C., and Dorman, C.J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol* *8*, 185–195.
- Ditto, M.D., Roberts, D., and Weisberg, R.A. (1994). Growth phase variation of integration host factor level in *Escherichia coli*. *J Bacteriol* *176*, 3738–3748.
- Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., and Mattick, J.S. (1991). “Touchdown” PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* *19*, 4008.
- Dorgai, L., Yagil, E., and Weisberg, R.A. (1995). Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases. *J Mol Biol* *252*, 178–188.
- Dorman, C.J. (2009). Nucleoid-associated proteins and bacterial physiology. In *Advances in Applied Microbiology*, (Academic Press), pp. 47–64.
- Duggin, I.G., Dubarry, N., and Bell, S.D. (2011). Replication termination and chromosome dimer resolution in the archaeon *Sulfolobus solfataricus*. *EMBO J* *30*, 145–153.
- Dupont, L., Boizet-Bonhoure, B., Coddeville, M., Auvray, F., and Ritzenthaler, P. (1995). Characterization of genetic elements required for site-specific integration of *Lactobacillus delbrueckii* subsp. *bulgaricus* bacteriophage mv4 and construction of an integration-proficient vector for *Lactobacillus plantarum*. *J Bacteriol* *177*, 586–595.

- Eberl, L., Kristensen, C.S., Givskov, M., Grohmann, E., Gerlitz, M., and Schwab, H. (1994). Analysis of the multimer resolution system encoded by the parCBA operon of broad-host-range plasmid RP4. *Mol Microbiol* *12*, 131–141.
- Elantak, L., Ansaldi, M., Guerlesquin, F., Méjean, V., and Morelli, X. (2005). Structural and genetic analyses reveal a key role in prophage excision for the TorI response regulator inhibitor. *J Biol Chem* *280*, 36802–36808.
- Elias, A., Kassis, H., Elkader, S.A., Gritsenko, N., Nahmad, A., Shir, H., Younis, L., Shannan, A., Aihara, H., Prag, G., et al. (2020). HK022 bacteriophage Integrase mediated RMCE as a potential tool for human gene therapy. *Nucleic Acids Res* *48*, 12804–12816.
- Ellis, R.J. (2001). Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol* *11*, 114–119.
- Escudero, J.A., Loot, C., Nivina, A., and Mazel, D. (2015). The integron: adaptation on demand. *Microbiol Spectr* *3*, MDNA3-0019-2014.
- Escudero, J.A., Loot, C., Parissi, V., Nivina, A., Bouchier, C., and Mazel, D. (2016). Unmasking the ancestral activity of integron integrases reveals a smooth evolutionary transition during functional innovation. *Nat Commun* *7*, 10937.
- Escudero, J.A., Loot, C., and Mazel, D. (2018). Integrons as adaptive devices. In *Molecular Mechanisms of Microbial Evolution*, P.H. Rampelotto, ed. (Cham: Springer International Publishing), pp. 199–239.
- Esposito, D., and Gerard, G.F. (2003). The *Escherichia coli* Fis protein stimulates bacteriophage lambda integrative recombination in vitro. *J Bacteriol* *185*, 3076–3080.
- Esposito, D., and Scocca, J.J. (1997). Purification and characterization of HP1 cox and definition of its role in controlling the direction of site-specific recombination. *J Biol Chem* *272*, 8660–8670.
- Esposito, D., Thrower, J.S., and Scocca, J.J. (2001). Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation. *Nucleic Acids Res* *29*, 3955–3964.

- Fadeev, E.A., Sam, M.D., and Clubb, R.T. (2009). NMR Structure of the Amino-Terminal Domain of the Lambda Integrase Protein in Complex with DNA: Immobilization of a Flexible Tail Facilitates Beta-Sheet Recognition of the Major Groove. *J Mol Biol* 388, 682–690.
- Fogg, P.C.M., Colloms, S., Rosser, S., Stark, M., and Smith, M.C.M. (2014). New applications for phage integrases. *J Mol Biol* 426, 2703–2716.
- Fogg, P.C.M., Younger, E., Fernando, B.D., Khaleel, T., Stark, W.M., and Smith, M.C.M. (2018). Recombination directionality factor gp3 binds ϕ C31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. *Nucleic Acids Res* 46, 1308–1320.
- Francia, M.V., Zabala, J.C., de la Cruz, F., and García Lobo, J.M. (1999). The IntI1 Integron Integrase Preferentially Binds Single-Stranded DNA of the *attC* Site. *J Bacteriol* 181, 6844–6849.
- Franke, A.E., and Clewell, D.B. (1981). Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of “conjugal” transfer in the absence of a conjugative plasmid. *J Bacteriol* 145, 494–502.
- Friedman, D.I. (1988). Integration host factor: a protein for all reasons. *Cell* 55, 545–554.
- Frumerie, C., Sylwan, L., Ahlgren-Berg, A., and Haggård-Ljungquist, E. (2005). Cooperative interactions between bacteriophage P2 integrase and its accessory factors IHF and Cox. *Virology* 332, 284–294.
- Frumerie, C., Ducos-Galand, M., Gopaul, D.N., and Mazel, D. (2010). The relaxed requirements of the integron cleavage site allow predictable changes in integron target specificity. *Nucleic Acids Res* 38, 559–569.
- Gardner, J.F., and Nash, H.A. (1986). Role of *Escherichia coli* IHF protein in lambda site-specific recombination. A mutational analysis of binding sites. *J Mol Biol* 191, 181–189.
- Gawron-Burke, C., and Clewell, D.B. (1982). A transposon in *Streptococcus faecalis* with fertility properties. *Nature* 300, 281–284.
- Gellert, M., and Nash, H. (1987). Communication between segments of DNA during site-specific recombination. *Nature* 325, 401–404.

- Ghosh, K., Lau, C.-K., Gupta, K., and Van Duyne, G.D. (2005a). Preferential synapsis of *loxP* sites drives ordered strand exchange in Cre-*loxP* site-specific recombination. *Nat Chem Biol* 1, 275–282.
- Ghosh, P., Pannunzio, N.R., and Hatfull, G.F. (2005b). Synapsis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites. *J Mol Biol* 349, 331–348.
- Ghosh, P., Wasil, L.R., and Hatfull, G.F. (2006). Control of phage Bxb1 excision by a novel recombination directionality factor. *PLoS Biol* 4, e186.
- Gibb, B., Gupta, K., Ghosh, K., Sharp, R., Chen, J., and Van Duyne, G.D. (2010). Requirements for catalysis in the Cre recombinase active site. *Nucleic Acids Res* 38, 5817–5832.
- Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6, 343–345.
- Giese, K., Cox, J., and Grosschedl, R. (1992). The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* 69, 185–195.
- Goodman, S.D., and Nash, H.A. (1989). Functional replacement of a protein-induced bend in a DNA recombination site. *Nature* 341, 251–254.
- Goodman, S.D., and Scocca, J.J. (1989). Nucleotide sequence and expression of the gene for the site-specific integration protein from bacteriophage HP1 of *Haemophilus influenzae*. *J Bacteriol* 171, 4232–4240.
- Goodman, S.D., Nicholson, S.C., and Nash, H.A. (1992). Deformation of DNA during site-specific recombination of bacteriophage A: Replacement of IHF protein by HU protein or sequence-directed bends. *Proc Natl Acad Sci USA* 89, 11910–11914.
- Gopaul, D.N., Guo, F., and Van Duyne, G.D. (1998). Structure of the Holliday junction intermediate in Cre-*loxP* site-specific recombination. *EMBO J* 17, 4175–4187.
- Gordon, A., and Hannon, G.J. (2010). FASTQ/A short-reads pre-processing tools.
- Gourlay, S.C., and Colloms, S.D. (2004). Control of Cre recombination by regulatory elements from Xer recombination systems. *Mol Microbiol* 52, 53–65.

- Grainge, I., Lesterlin, C., and Sherratt, D.J. (2011). Activation of XerCD- *dif* recombination by the FtsK DNA translocase. *Nucleic Acids Res* 39, 5140–5148.
- 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.
- Grindley, N.D. (2002). The movement of Tn3-like elements: transposition and cointegrate resolution. In *Mobile DNA II*, N.L. Craig, R. Craigie, M. Gellert, and A. Lambowitz, eds. (Washington, DC, USA: ASM Press), pp. 272–302.
- Grindley, N.D.F., Whiteson, K.L., and Rice, P.A. (2006). Mechanisms of site-specific recombination. *Annu Rev Biochem* 75, 567–605.
- Groth, A.C., and Calos, M.P. (2004). Phage integrases: biology and applications. *J Mol Biol* 335, 667–678.
- Groth, A.C., Olivares, E.C., Thyagarajan, B., and Calos, M.P. (2000). A phage integrase directs efficient site-specific integration in human cells. *Proc Natl Acad Sci USA* 97, 5995–6000.
- 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.
- Guo, F., Gopaul, D.N., and Duyne, G.D.V. (1999). Asymmetric DNA bending in the *Cre-loxP* site-specific recombination synapse. *Proc Natl Acad Sci USA* 96, 7143–7148.
- Gupta, K., Sharp, R., Yuan, J.B., Li, H., and Van Duyne, G.D. (2017). Coiled-coil interactions mediate serine integrase directionality. *Nucleic Acids Res* 45, 7339–7353.
- Hakimi, J.M., and Scoocca, J.J. (1994). Binding sites for bacteriophage HP1 integrase on its DNA substrates. *J Biol Chem* 269, 21340–21345.
- Hancock, S.P., Cascio, D., and Johnson, R.C. (2019). Cooperative DNA binding by proteins through DNA shape complementarity. *Nucleic Acids Res* 47, 8874–8887.
- Hartley, J.L., Temple, G.F., and Brasch, M.A. (2000). DNA cloning using in vitro site-specific recombination. *Genome Res* 10, 1788–1795.

- Hatfull, G.F., and Grindley, N.D. (1986). Analysis of gamma delta resolvase mutants in vitro: evidence for an interaction between serine-10 of resolvase and site I of *res*. Proc Natl Acad Sci USA 83, 5429–5433.
- Hauser, M.A., and Scocca, J.J. (1992). Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1. Identification of the points of recombinational strand exchange and the limits of the host attachment site. J Biol Chem 267, 6859–6864.
- Hayashi, T., Matsumoto, H., Ohnishi, M., and Terawaki, Y. (1993). Molecular analysis of a cytotoxin-converting phage, phi CTX, of *Pseudomonas aeruginosa*: structure of the *attP*-*cos*-*ctx* region and integration into the serine tRNA gene. Mol Microbiol 7, 657–667.
- Hazelbaker, D., Azaro, M.A., and Landy, A. (2008). A biotin interference assay highlights two different asymmetric interaction profiles for λ Integrase arm-type binding sites in integrative versus excisive recombination. J Biol Chem 283, 12402–12414.
- Hendrix, R.W., Smith, M.C., Burns, R.N., Ford, M.E., and Hatfull, G.F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. Proc Natl Acad Sci U S A 96, 2192–2197.
- Hickman, A.B., Waninger, S., Scocca, J.J., and Dyda, F. (1997). Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 Integrase at 2.7 Å resolution. Cell 89, 227–237.
- Hinerfeld, D., and Churchward, G. (2001). Xis protein of the conjugative transposon Tn916 plays dual opposing roles in transposon excision. Mol Microbiol 41, 1459–1467.
- Hochman, L., Segev, N., Sternberg, N., and Cohen, G. (1983). Site-specific recombinational circularization of bacteriophage P1 DNA. Virology 131, 11–17.
- Hoess, R.H., Foeller, C., Bidwell, K., and Landy, A. (1980). Site-specific recombination functions of bacteriophage lambda: DNA sequence of regulatory regions and overlapping structural genes for Int and Xis. Proc Natl Acad Sci USA 77, 2482–2486.
- Hoess, R.H., Ziese, M., and Sternberg, N. (1982). P1 site-specific recombination: nucleotide sequence of the recombining sites. Proc Natl Acad Sci USA 79, 3398–3402.

- Hoess, R.H., Wierzbicki, A., and Abremski, K. (1986). The role of the *loxP* spacer region in P1 site-specific recombination. *Nucleic Acids Res* 14, 2287–2300.
- Hsu, P.L., Ross, W., and Landy, A. (1980). The lambda phage *att* site: functional limits and interaction with Int protein. *Nature* 285, 85–91.
- Hwang, E.S., and Scocca, J.J. (1990). Interaction of integration host factor from *Escherichia coli* with the integration region of the *Haemophilus influenzae* bacteriophage HP1. *J Bacteriol* 172, 4852–4860.
- Jayaram, M., Grainge, I., and Tribble, G. (2002). Site-specific recombination by the Flp protein of *Saccharomyces cerevisiae*. In *Mobile DNA II*, N.L. Craig, R. Craigie, M. Gellert, and A. Lambowitz, eds. (Washington, DC, USA: ASM Press), pp. 192–218.
- Jayaram, M., Ma, C.-H., Kachroo, A.H., Rowley, P.A., Guga, P., Fan, H.-F., and Voziyanov, Y. (2015). An overview of tyrosine site-specific recombination: from an Flp perspective. *Microbiol Spectr* 3, 3.4.12.
- Johansson, C., Kamali-Moghaddam, M., and Sundström, L. (2004). Integron integrase binds to bulged hairpin DNA. *Nucleic Acids Res* 32, 4033–4043.
- Johnson, R.C. (2002). Bacterial site-specific DNA inversion systems. In *Mobile DNA II*, N.L. Craig, R. Craigie, M. Gellert, and A. Lambowitz, eds. (Washington, DC, USA: ASM Press), pp. 230–271.
- Johnson, R.C. (2015). Site-specific DNA inversion by serine recombinases. *Microbiol Spectr* 3, 1–36.
- Johnson, C.M., and Grossman, A.D. (2015). Integrative and Conjugative Elements (ICEs): What they do and how they work. *Annu Rev Genet* 49, 577–601.
- Johnson, R.C., and Simon, M.I. (1985). Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. *Cell* 41, 781–791.
- Johnson, R.C., Bruist, M.F., and Simon, M.I. (1986). Host protein requirements for in vitro site-specific DNA inversion. *Cell* 46, 531–539.

Johnson, R.C., Ball, C.A., Pfeffer, D., and Simon, M.I. (1988). Isolation of the gene encoding the *Hin* recombinational enhancer binding protein. *Proc Natl Acad Sci USA* 85, 3484–3488.

Johnson, R.C., Johnson, L.M., Schmidt, J.W., and Gardner, J.F. (2004). Major nucleoid proteins in the structure and function of the *Escherichia coli* chromosome. In *The Bacterial Chromosome*, N.P. Higgins, ed. (Washington, DC, USA: ASM Press), pp. 65–132.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589.

Kamashev, D., Agapova, Y., Rastorguev, S., Talyzina, A.A., Boyko, K.M., Korzhenevskiy, D.A., Vlaskina, A., Vasilov, R., Timofeev, V.I., and Rakitina, T.V. (2017). Comparison of histone-like HU protein DNA-binding properties and HU/IHF protein sequence alignment. *PLOS ONE* 12, e0188037.

Kamp, D., Kahmann, R., Zipser, D., Broker, T.R., and Chow, L.T. (1978). Inversion of the G DNA segment of phage Mu controls phage infectivity. *Nature* 271, 577–580.

Kanaar, R., and Cozzarelli, N.R. (1992). Roles of supercoiled DNA structure in DNA transactions. *Curr Opin Struct Biol* 2, 369–379.

Kazmierczak, R.A., Swalla, B.M., Burgin, A.B., Gumport, R.I., and Gardner, J.F. (2002). Regulation of site-specific recombination by the C-terminus of lambda integrase. *Nucleic Acids Res* 30, 5193–5204.

Keeton, C.M., and Gardner, J.F. (2012). Roles of Exc protein and DNA homology in the CTnDOT excision reaction. *J Bacteriol* 194, 3368–3376.

Keeton, C.M., Hopp, C.M., Yoneji, S., and Gardner, J.F. (2013). Interactions of the excision proteins of CTnDOT in the *attR* intasome. *Plasmid* 70, 190–200.

Khaleel, T., Younger, E., McEwan, A.R., Varghese, A.S., and Smith, M.C.M. (2011). A phage protein that binds ϕ C31 integrase to switch its directionality. *Mol Microbiol* 80, 1450–1463.

Kikuchi, Y., and Nash, H.A. (1978). The bacteriophage lambda *int* gene product. A filter assay for genetic recombination, purification of *int*, and specific binding to DNA. *J Biol Chem* 253, 7149–7157.

- Kikuchi, Y., and Nash, H.A. (1979). Nicking-closing activity associated with bacteriophage lambda *int* gene product. *Proc Natl Acad Sci USA* 76, 3760–3764.
- Kim, S., and Gardner, J.F. (2011). Resolution of Holliday junction recombination intermediates by wild-type and mutant IntDOT proteins. *J Bacteriol* 193, 1351–1358.
- Kim, A.I., Ghosh, P., Aaron, M.A., Bibb, L.A., Jain, S., and Hatfull, G.F. (2003). Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis* *groEL1* gene. *Mol Microbiol* 50, 463–473.
- Kim, S., Swalla, B.M., and Gardner, J.F. (2010). Structure-function analysis of IntDOT. *J Bacteriol* 192, 575–586.
- King, S.R., and Richardson, J.P. (1986). Role of homology and pathway specificity for recombination between plasmids and bacteriophage λ . *Mol Gen Genet* 204, 141–147.
- Kitts, P.A., and Nash, H.A. (1987). Homology-dependent interactions in phage λ site-specific recombination. *Nature* 329, 346–348.
- Kitts, P.A., and Nash, H.A. (1988). Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. *J Mol Biol* 204, 95–107.
- Klemm, P. (1986). Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. *EMBO J* 5, 1389–1393.
- Klippel, A., Mertens, G., Patschinsky, T., and Kahmann, R. (1988). The DNA invertase Gin of phage Mu: formation of a covalent complex with DNA via a phosphoserine at amino acid position 9. *EMBO J* 7, 1229–1237.
- Koch, C., and Kahmann, R. (1986). Purification and properties of the *Escherichia coli* host factor required for inversion of the G segment in bacteriophage Mu. *J Biol Chem* 261, 15673–15678.
- Kolot, M., and Yagil, E. (1994). Position and direction of strand exchange in bacteriophage HK022 integration. *Mol Gen Genet* 245, 623–627.
- Kolot, M., Silberstein, N., and Yagil, E. (1999). Site-specific recombination in mammalian cells expressing the Int recombinase of bacteriophage HK022. *Mol Biol Rep* 26, 207–213.

- Kolot, M., Meroz, A., and Yagil, E. (2003). Site-specific recombination in human cells catalyzed by the wild-type integrase protein of coliphage HK022. *Biotechnol Bioeng* 84, 56–60.
- Kolot, M., Malchin, N., Elias, A., Gritsenko, N., and Yagil, E. (2015). Site promiscuity of coliphage HK022 integrase as a tool for gene therapy. *Gene Ther* 22, 521–527.
- Kono, N., Arakawa, K., and Tomita, M. (2011). Comprehensive prediction of chromosome dimer resolution sites in bacterial genomes. *BMC Genomics* 12, 19.
- Kostrewa, D., Granzin, J., Koch, C., Choe, H.W., Raghunathan, S., Wolf, W., Labahn, J., Kahmann, R., and Saenger, W. (1991). Three-dimensional structure of the *E. coli* DNA-binding protein FIS. *Nature* 349, 178–180.
- Kovach, M.J., Tirumalai, R., and Landy, A. (2002). Site-specific photo-cross-linking between lambda integrase and its DNA recombination target. *J Biol Chem* 277, 14530–14538.
- Krasnow, M.A., and Cozzarelli, N.R. (1983). Site-specific relaxation and recombination by the Tn3 resolvase: recognition of the DNA path between oriented res sites. *Cell* 32, 1313–1324.
- Krogh, B.O., and Shuman, S. (2000). Catalytic mechanism of DNA topoisomerase IB. *Mol Cell* 5, 1035–1041.
- Kuempel, P.L., Henson, J.M., Dircks, L., Tecklenburg, M., and Lim, D.F. (1991). *dif*, a *recA*-independent recombination site in the terminus region of the chromosome of *Escherichia coli*. *New Biol* 3, 799–811.
- Kuhstoss, S., and Rao, R.N. (1991). Analysis of the integration function of the streptomycete bacteriophage ϕ C31. *J Mol Biol* 222, 897–908.
- Kwon, H.J., Tirumalai, R., Landy, A., and Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* 276, 126–131.
- Lahbib-Mansais, Y., Mata, M., and Ritzenthaler, P. (1988). Molecular taxonomy of *Lactobacillus* phages. *Biochimie* 70, 429–435.

- Lahbib-Mansais, Y., Boizet, B., Dupont, L., Mata, M., and Ritzenthaler, P. (1992). Characterization of a temperate bacteriophage of *Lactobacillus delbrueckii* subsp. *bulgaricus* and its interactions with the host cell chromosome. *Microbiology* 138, 1139–1146.
- Landy, A. (1989). Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem* 58, 913–949.
- Landy, A. (2015). The λ Integrase site-specific recombination pathway. *Microbiol Spectr* 3, MDNA3-0051–2014.
- Landy, A., and Ross, W. (1977). Viral integration and excision: structure of the lambda *att* sites. *Science* 197, 1147–1160.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357–359.
- Langridge, G.C., Phan, M.-D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., Charles, I., Maskell, D.J., Peters, S.E., Dougan, G., et al. (2009). Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* 19, 2308–2316.
- Laprise, J., Yoneji, S., and Gardner, J.F. (2010). Homology-dependent interactions determine the order of strand exchange by IntDOT recombinase. *Nucleic Acids Res* 38, 958–969.
- Lauth, M., Spreafico, F., Dethleffsen, K., and Meyer, M. (2002). Stable and efficient cassette exchange under non-selectable conditions by combined use of two site-specific recombinases. *Nucleic Acids Res* 30, e115.
- Laxmikanthan, G., Xu, C., Brilot, A.F., Warren, D., Steele, L., Seah, N., Tong, W., Grigorieff, N., Landy, A., and Van Duyne, G.D. (2016). Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *ELife* 5, e14313.
- Le Bourgeois, P., Langella, P., and Ritzenthaler, P. (2000). Electrotransformation of *Lactococcus lactis*. In *Electrotransformation of Bacteria*, N. Eynard, and J. Teissié, eds. (Berlin, Heidelberg: Springer), pp. 56–65.
- Le Bourgeois, P., Bugarel, M., Campo, N., Daveran-Mingot, M.-L., Labonté, J., Lanfranchi, D., Lautier, T., Pagès, C., and Ritzenthaler, P. (2007). The unconventional Xer recombination machinery of *Streptococci/Lactococci*. *PLoS Genet* 3, e117.

- Lee, G., and Saito, I. (1998). Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene* 216, 55–65.
- Lee, L., and Sadowski, P.D. (2003). Identification of Cre residues involved in synapsis, isomerization, and catalysis. *J Biol Chem* 278, 36905–36915.
- Lee, M.H., and Hatfull, G.F. (1993). Mycobacteriophage L5 integrase-mediated site-specific integration in vitro. *J Bacteriol* 175, 6836–6841.
- Lee, M.H., Pascopella, L., Jacobs, W.R., and Hatfull, G.F. (1991). Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guérin. *Proc Natl Acad Sci USA* 88, 3111–3115.
- Leenhouts, K., Buist, G., Bolhuis, A., ten Berge, A., Kiel, J., Mierau, I., Dabrowska, M., Venema, G., and Kok, J. (1996). A general system for generating unlabelled gene replacements in bacterial chromosomes. *Mol Gen Genet* 253, 217–224.
- Lewis, J.A., and Hatfull, G.F. (2001). Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* 29, 2205–2216.
- Lewis, J.A., and Hatfull, G.F. (2003). Control of directionality in L5 integrase-mediated site-specific recombination. *J Mol Biol* 326, 805–821.
- Li, H., Sharp, R., Rutherford, K., Gupta, K., and Van Duyne, G.D. (2018). Serine integrase *attP* binding and specificity. *J Mol Biol* 430, 4401–4418.
- Lichtenstein, C., and Brenner, S. (1982). Unique insertion site of Tn7 in the *E. coli* chromosome. *Nature* 297, 601–603.
- Loot, C., Bikard, D., Rachlin, A., and Mazel, D. (2010). Cellular pathways controlling integron cassette site folding. *EMBO J* 29, 2623–2634.
- Loot, C., Ducos-Galand, M., Escudero, J.A., Bouvier, M., and Mazel, D. (2012). Replicative resolution of integron cassette insertion. *Nucleic Acids Res* 40, 8361–8370.

- Lovett, S.T., and Kolodner, R.D. (1991). Nucleotide sequence of the *Escherichia coli recJ* chromosomal region and construction of *recJ*-overexpression plasmids. *J Bacteriol* *173*, 353–364.
- Lu, F., and Churchward, G. (1994). Conjugative transposition: Tn916 integrase contains two independent DNA binding domains that recognize different DNA sequences. *EMBO J* *13*, 1541–1548.
- Lu, F., and Churchward, G. (1995). Tn916 target DNA sequences bind the C-terminal domain of integrase protein with different affinities that correlate with transposon insertion frequency. *J Bacteriol* *177*, 1938–1946.
- 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.
- Malanowska, K., Salyers, A.A., and Gardner, J.F. (2006). Characterization of a conjugative transposon integrase, IntDOT. *Mol Microbiol* *60*, 1228–1240.
- Malanowska, K., Yoneji, S., Salyers, A.A., and Gardner, J.F. (2007). CTnDOT integrase performs ordered homology-dependent and homology-independent strand exchanges. *Nucleic Acids Res* *35*, 5861–5873.
- Mandali, S., and Johnson, R.C. (2021). Control of the Serine Integrase Reaction: Roles of the Coiled-Coil and Helix E Regions in DNA Site Synapsis and Recombination. *J Bacteriol* *203*, e0070320.
- Mandali, S., Gupta, K., Dawson, A.R., Van Duyne, G.D., and Johnson, R.C. (2017). Control of recombination directionality by the *Listeria* phage A118 protein Gp44 and the coiled-coil motif of its serine integrase. *J Bacteriol* *199*, e00019-17.
- Marra, D., Pethel, B., Churchward, G.G., and Scott, J.R. (1999). The frequency of conjugative transposition of Tn916 is not determined by the frequency of excision. *J Bacteriol* *181*, 5414–5418.
- Mata, M., Trautwetter, A., Luthaud, G., and Ritzenthaler, P. (1986). Thirteen virulent and temperate bacteriophages of *Lactobacillus bulgaricus* and *Lactobacillus lactis* belong to a single DNA homology group. *Appl Environ Microbiol* *52*, 812–818.

- Mazel, D. (2006). Integrons: agents of bacterial evolution. *Nat Rev Microbiol* 4, 608–620.
- 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.
- McLean, M.M., Chang, Y., Dhar, G., Heiss, J.K., and Johnson, R.C. (2013). Multiple interfaces between a serine recombinase and an enhancer control site-specific DNA inversion. *ELife* 2, e01211.
- McLeod, S.M., and Waldor, M.K. (2004). Characterization of XerC- and XerD-dependent CTX phage integration in *Vibrio cholerae*. *Mol Microbiol* 54, 935–947.
- McLeod, M., Craft, S., and Broach, J.R. (1986). Identification of the crossover site during FLP-mediated recombination in the *Saccharomyces cerevisiae* plasmid 2 microns circle. *Mol Cell Biol* 6, 3357–3367.
- Messier, N., and Roy, P.H. (2001). Integron integrases possess a unique additional domain necessary for activity. *J Bacteriol* 183, 6699–6706.
- Midonet, C., and Barre, F.-X. (2014). Xer site-specific recombination: promoting vertical and horizontal transmission of genetic information. *Microbiol Spectr* 2, 2.6.29.
- Midonet, C., and Barre, F.-X. (2016). How Xer-exploiting mobile elements overcome cellular control. *Proc Natl Acad Sci USA* 113, 8343–8345.
- Midonet, C., Das, B., Paly, E., and Barre, F.-X. (2014). XerD-mediated FtsK-independent integration of TLC ϕ into the *Vibrio cholerae* genome. *Proc Natl Acad Sci USA* 111, 16848–16853.
- Midonet, C., Miele, S., Paly, E., Guerois, R., and Barre, F.-X. (2019). The TLC Φ satellite phage harbors a Xer recombination activation factor. *Proc Natl Acad Sci USA* 116, 18391–18396.
- Minton, A.P. (2006). How can biochemical reactions within cells differ from those in test tubes? *J Cell Sci* 119, 2863–2869.
- Mizuuchi, K., and Nash, H.A. (1976). Restriction assay for integrative recombination of bacteriophage lambda DNA in vitro: requirement for closed circular DNA substrate. *Proc Natl Acad Sci U S A* 73, 3524–3528.

Mizuuchi, K., Gellert, M., and Nash, H.A. (1978). Involvement of supertwisted DNA in integrative recombination of bacteriophage lambda. *J Mol Biol* 121, 375–392.

Mizuuchi, K., Weisberg, R., Enquist, L., Mizuuchi, M., Buraczynska, M., Foeller, C., Hsu, P.L., Ross, W., and Landy, A. (1981). Structure and function of the phage lambda *att* site: size, int-binding sites, and location of the crossover point. *Cold Spring Harb Symp Quant Biol* 45 Pt 1, 429–437.

Moyer, K.E., Kimsey, H.H., and Waldor, M.K. (2001). Evidence for a rolling-circle mechanism of phage DNA synthesis from both replicative and integrated forms of CTX ϕ . *Mol Microbiol* 41, 311–323.

Naas, T., Mikami, Y., Imai, T., Poirel, L., and Nordmann, P. (2001). Characterization of In53, a class 1 plasmid- and composite transposon-located integron of *Escherichia coli* which carries an unusual array of gene cassettes. *J Bacteriol* 183, 235–249.

Nash, H.A. (1974a). Purification of bacteriophage lambda Int protein. *Nature* 247, 543–545.

Nash, H.A. (1974b). Lambda *attB-attP*, a lambda derivative containing both sites involved in integrative recombination. *Virology* 57, 207–216.

Nash, H.A. (1975). Integrative recombination of bacteriophage lambda DNA in vitro. *Proc Natl Acad Sci USA* 72, 1072–1076.

Nash, H.A. (1996). The HU and IHF proteins: accessory factors for complex protein-DNA assemblies. In *Regulation of Gene Expression in Escherichia Coli*, E.C.C. Lin, and A.S. Lynch, eds. (Boston: Springer), pp. 149–179.

Nash, H.A., and Robertson, C.A. (1981). Purification and properties of the *Escherichia coli* protein factor required for lambda integrative recombination. *J Biol Chem* 256, 9246–9253.

Nelson, H.C., Finch, J.T., Luisi, B.F., and Klug, A. (1987). The structure of an oligo(dA).oligo(dT) tract and its biological implications. *Nature* 330, 221–226.

Nivina, A., Escudero, J.A., Vit, C., Mazel, D., and Loot, C. (2016). Efficiency of integron cassette insertion in correct orientation is ensured by the interplay of the three unpaired features of *attC* recombination sites. *Nucleic Acids Res* 44, 7792–7803.

- Nivina, A., Grieb, M.S., Loot, C., Bikard, D., Cury, J., Shehata, L., Bernardes, J., and Mazel, D. (2020). Structure-specific DNA recombination sites: Design, validation, and machine learning-based refinement. *Sci Adv* 6, eaay2922.
- Nolivos, S., Pages, C., Rousseau, P., Le Bourgeois, P., and Cornet, F. (2010). Are two better than one? Analysis of an FtsK/Xer recombination system that uses a single recombinase. *Nucleic Acids Res* 38, 6477–6489.
- Numrych, T.E., Gumport, R.I., and Gardner, J.F. (1992). Characterization of the bacteriophage lambda excisionase (Xis) protein: the C-terminus is required for Xis-integrase cooperativity but not for DNA binding. *EMBO J* 11, 3797–3806.
- Nunes-Düby, S.E., Matsumoto, L., and Landy, A. (1987). Site-specific recombination intermediates trapped with suicide substrates. *Cell* 50, 779–788.
- Nunes-Düby, S.E., Tirumalai, R.S., Dorgai, L., Yagil, E., Weisberg, R.A., and Landy, A. (1994). Lambda integrase cleaves DNA *in cis*. *EMBO J* 13, 4421–4430.
- Nunes-Düby, S.E., Azaro, M.A., and Landy, A. (1995a). Swapping DNA strands and sensing homology without branch migration in λ site-specific recombination. *Curr Biol* 5, 139–148.
- Nunes-Düby, S.E., Smith-Mungo, L.I., and Landy, A. (1995b). Single base-pair precision and structural rigidity in a small IHF-induced DNA loop. *J Mol Biol* 253, 228–242.
- Nunes-Düby, 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.
- Nunes-Düby, 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.
- Olorunniji, F.J., and Stark, W.M. (2010). Catalysis of site-specific recombination by Tn3resolvase. *Biochem Soc Trans* 38, 417–421.
- Olorunniji, F.J., Rosser, S.J., and Stark, W.M. (2016). Site-specific recombinases: molecular machines for the genetic revolution. *Biochem J* 473, 673–684.

- Pan, G., Luetke, K., and Sadowski, P.D. (1993). Mechanism of cleavage and ligation by FLP recombinase: classification of mutations in FLP protein by in vitro complementation analysis. *Mol Cell Biol* 13, 3167–3175.
- Panis, G., Méjean, V., and Ansaldi, M. (2007). Control and regulation of KplE1 prophage site-specific recombination: a new recombination module analyzed. *J Biol Chem* 282, 21798–21809.
- Panis, G., Duverger, Y., Champ, S., and Ansaldi, M. (2010). Protein binding sites involved in the assembly of the KplE1 prophage intasome. *Virology* 404, 41–50.
- Papagiannis, C.V., Sam, M.D., Abbani, M.A., Yoo, D., Cascio, D., Clubb, R.T., and Johnson, R.C. (2007). Fis targets assembly of the xis nucleoprotein filament to promote excisive recombination by phage lambda. *J Mol Biol* 367, 328–343.
- Pargellis, C.A., Nunes-Düby, S.E., de Vargas, L.M., and Landy, A. (1988). Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. *J Biol Chem* 263, 7678–7685.
- Partridge, S.R., Recchia, G.D., Scaramuzzi, C., Collis, C.M., Stokes, H.W., and Hall, R.M. (2000). Definition of the *attI1* site of class 1 integrons. *Microbiology* 146 (Pt 11), 2855–2864.
- Pease, P.J., Levy, O., Cost, G.J., Gore, J., Ptacin, J.L., Sherratt, D., Bustamante, C., and Cozzarelli, N.R. (2005). Sequence-Directed DNA Translocation by Purified FtsK. *Science* 307, 586–590.
- Pedulla, M.L., Lee, M.H., Lever, D.C., and Hatfull, G.F. (1996). A novel host factor for integration of mycobacteriophage L5. *Proc Natl Acad Sci USA* 93, 15411–15416.
- Pellegrini, O., Oberto, J., Pinson, V., Wery, M., and Rouviere-Yaniv, J. (2000). Overproduction and improved strategies to purify the three native forms of nuclease-free HU protein. *Biochimie* 82, 693–704.
- Peña, C.E., Stoner, J.E., and Hatfull, G.F. (1996). Positions of strand exchange in mycobacteriophage L5 integration and characterization of the *attB* site. *J Bacteriol* 178, 5533–5536.

- Peña, C.E., Lee, M.H., Pedulla, M.L., and Hatfull, G.F. (1997). Characterization of the mycobacteriophage L5 attachment site, *attP*. *J Mol Biol* 266, 76–92.
- Peña, C.E., Kahlenberg, J.M., and Hatfull, G.F. (1998). The role of supercoiling in mycobacteriophage L5 integrative recombination. *Nucleic Acids Res* 26, 4012–4018.
- Peña, C.E., Kahlenberg, J.M., and Hatfull, G.F. (1999). Protein-DNA complexes in mycobacteriophage L5 integrative recombination. *J Bacteriol* 181, 454–461.
- Peña, C.E., Kahlenberg, J.M., and Hatfull, G.F. (2000). Assembly and activation of site-specific recombination complexes. *Proc Natl Acad Sci USA* 97, 7760–7765.
- Petersen, A., Josephsen, J., and Johnsen, M.G. (1999). TPW22, a lactococcal temperate phage with a site-specific integrase closely related to *Streptococcus thermophilus* phage integrases. *J Bacteriol* 181, 7034–7042.
- Petit, M.A., Ehrlich, D., and Jannièrè, L. (1995). pAM beta 1 resolvase has an atypical recombination site and requires a histone-like protein HU. *Mol Microbiol* 18, 271–282.
- Plasterk, R.H., Kanaar, R., and van de Putte, P. (1984). A genetic switch in vitro: DNA inversion by Gin protein of phage Mu. *Proc Natl Acad Sci U S A* 81, 2689–2692.
- Pollock, T.J., and Abremski, K. (1979). DNA without supertwists can be an in vitro substrate for site-specific recombination of bacteriophage lambda. *J Mol Biol* 131, 651–654.
- Poulter, R.T.M., and Butler, M.I. (2015). Tyrosine recombinase retrotransposons and transposons. *Microbiol Spectr* 3, MDNA3-0036–2014.
- Poyart-Salmeron, C., Trieu-Cuot, P., Carlier, C., and Courvalin, P. (1989). Molecular characterization of two proteins involved in the excision of the conjugative transposon Tn1545: homologies with other site-specific recombinases. *EMBO J* 8, 2425–2433.
- Radman-Livaja, M., Biswas, T., Ellenberger, T., Landy, A., and Aihara, H. (2006). DNA arms do the legwork to ensure the directionality of λ site-specific recombination. *Curr Opin Struct Biol* 16, 42–50.
- 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.

- Raymond, C.S., and Soriano, P. (2007). High-efficiency FLP and Φ C31 site-specific recombination in mammalian cells. *PLoS One* 2, e162.
- Recchia, G.D., and Sherratt, D.J. (1999). Conservation of xer site-specific recombination genes in bacteria. *Mol Microbiol* 34, 1146–1148.
- Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., and Hol, W.G.J. (1998). Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* 279, 1504–1513.
- Reed, R.R. (1981a). Resolution of cointegrates between transposons gamma delta and Tn3 defines the recombination site. *Proc Natl Acad Sci USA* 78, 3428–3432.
- Reed, R.R. (1981b). Transposon-mediated site-specific recombination: a defined in vitro system. *Cell* 25, 713–719.
- Reed, R.R., and Grindley, N.D.F. (1981). Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein-DNA linkage at the recombination site. *Cell* 25, 721–728.
- Reiter, W.D., Palm, P., and Yeats, S. (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* 17, 1907–1914.
- Rice, P.A. (2015). Serine Resolvases. *Microbiol Spectr* 3, MDNA3-0045–2014.
- Rice, P.A., Yang, S., Mizuuchi, K., and Nash, H.A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell* 87, 1295–1306.
- Richet, E., Abcarian, P., and Nash, H.A. (1986). The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell* 46, 1011–1021.
- Richet, E., Abcarian, P., and Nash, H.A. (1988). Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell* 52, 9–17.
- Ringwald, K., and Gardner, J. (2015). The *Bacteroides thetaiotaomicron* protein *Bacteroides* Host Factor A participates in integration of the integrative conjugative element CTnDOT into the chromosome. *J Bacteriol* 197, 1339–1349.

- Ringwald, K., Yoneji, S., and Gardner, J. (2017). Resolution of mismatched overlap holliday junction intermediates by the tyrosine recombinase IntDOT. *J Bacteriol* *199*, e00873-16.
- Roberts, A.P., and Mullany, P. (2009). A modular master on the move: the Tn916 family of mobile genetic elements. *Trends Microbiol* *17*, 251–258.
- Rojo, F., and Alonso, J.C. (1994). A novel site-specific recombinase encoded by the *Streptococcus pyogenes* plasmid pSM19035. *J Mol Biol* *238*, 159–172.
- Ross, W., and Landy, A. (1982). Bacteriophage lambda int protein recognizes two classes of sequence in the phage *att* site: characterization of arm-type sites. *Proc Natl Acad Sci USA* *79*, 7724–7728.
- Ross, W., and Landy, A. (1983). Patterns of lambda Int recognition in the regions of strand exchange. *Cell* *33*, 261–272.
- Ross, W., Landy, A., Kikuchi, Y., and Nash, H. (1979). Interaction of Int protein with specific sites on lambda *att* DNA. *Cell* *18*, 297–307.
- Rouvière-Yaniv, J., and Kjeldgaard, N.O. (1979). Native *Escherichia coli* HU protein is a heterotypic dimer. *FEBS Lett* *106*, 297–300.
- Rowland, S.-J., Stark, W.M., and Boocock, M.R. (2002). Sin recombinase from *Staphylococcus aureus*: synaptic complex architecture and transposon targeting. *Mol Microbiol* *44*, 607–619.
- Rowland, S.-J., Boocock, M.R., and Stark, W.M. (2006). DNA bending in the Sin recombination synapse: functional replacement of HU by IHF. *Mol Microbiol* *59*, 1730–1743.
- Rowley, P.A., and Smith, M.C.M. (2008). Role of the N-terminal domain of ϕ C31 integrase in *attB-attP* Synapsis. *J Bacteriol* *190*, 6918–6921.
- Rowley, P.A., Smith, M.C.A., Younger, E., and Smith, M.C.M. (2008). A motif in the C-terminal domain of ϕ C31 integrase controls the directionality of recombination. *Nucleic Acids Res* *36*, 3879–3891.
- Rudd, K.E. (1999). Novel intergenic repeats of *Escherichia coli* K-12. *Res Microbiol* *150*, 653–664.

- Rudy, C., Taylor, K.L., Hinerfeld, D., Scott, J.R., and Churchward, G. (1997a). Excision of a conjugative transposon in vitro by the Int and Xis proteins of Tn916. *Nucleic Acids Res* 25, 4061–4066.
- Rudy, C.K., Scott, J.R., and Churchward, G. (1997b). DNA binding by the Xis protein of the conjugative transposon Tn916. *J Bacteriol* 179, 2567–2572.
- Rüfer, A., Neuenschwander, P.F., and Sauer, B. (2002). Analysis of Cre-*loxP* interaction by surface plasmon resonance: influence of spermidine on cooperativity. *Anal Biochem* 308, 90–99.
- Rutherford, K., and Van Duyne, G.D. (2014). The ins and outs of serine integrase site-specific recombination. *Curr Opin Struct Biol* 24, 125–131.
- Rutherford, K., Yuan, P., Perry, K., Sharp, R., and Van Duyne, G.D. (2013). Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res* 41, 8341–8356.
- Rutkai, E., Dorgai, L., Sirot, R., Yagil, E., and Weisberg, R.A. (2003). Analysis of insertion into secondary attachment sites by phage lambda and by int mutants with altered recombination specificity. *J Mol Biol* 329, 983–996.
- Sadowski, P.D. (1995). The Flp Recombinase of the 2- μ m plasmid of *Saccharomyces cerevisiae*. In *Progress in Nucleic Acid Research and Molecular Biology*, (Elsevier), pp. 53–91.
- Sam, M.D., Cascio, D., Johnson, R.C., and Clubb, R.T. (2004). Crystal structure of the excisionase–DNA complex from bacteriophage lambda. *J Mol Biol* 338, 229–240.
- Sarkar, D., Radman-Livaja, M., and Landy, A. (2001). The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions. *EMBO J* 20, 1203–1212.
- Sarkar, D., Azaro, M.A., Aihara, H., Papagiannis, C.V., Tirumalai, R., Nunes-Düby, S.E., Johnson, R.C., Ellenberger, T., and Landy, A. (2002). Differential affinity and cooperativity functions of the amino-terminal 70 residues of lambda integrase. *J Mol Biol* 324, 775–789.

- Sarkar, I., Hauber, I., Hauber, J., and Buchholz, F. (2007). HIV-1 proviral DNA excision using an evolved recombinase. *Science* *316*, 1912–1915.
- Sauer, B., and Henderson, N. (1988). Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci USA* *85*, 5166–5170.
- Schlake, T., and Bode, J. (1994). Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* *33*, 12746–12751.
- Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* *18*, 6097–6100.
- Seah, N.E., Warren, D., Tong, W., Laxmikanthan, G., Van Duyne, G.D., and Landy, A. (2014). Nucleoprotein architectures regulating the directionality of viral integration and excision. *Proc Natl Acad Sci USA* *111*, 12372–12377.
- Segall, A.M., Goodman, S.D., and Nash, H.A. (1994). Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. *EMBO J* *13*, 4536–4548.
- Sheren, J., Langer, S.J., and Leinwand, L.A. (2007). A randomized library approach to identifying functional *lox* site domains for the Cre recombinase. *Nucleic Acids Res* *35*, 5464–5473.
- Sherratt, D.J., and Wigley, D.B. (1998). Conserved themes but novel activities in recombinases and topoisomerases. *Cell* *93*, 149–152.
- Silverman, M., and Simon, M. (1980). Phase variation: Genetic analysis of switching mutants. *Cell* *19*, 845–854.
- Singer, B.S., Gold, L., Gauss, P., and Doherty, D.H. (1982). Determination of the amount of homology required for recombination in bacteriophage T4. *Cell* *31*, 25–33.
- Smith, M.C.M. (2015). Phage-encoded serine integrases and other large serine recombinases. *Microbiol Spectr* *3*, 3.4.06.
- Smith, M.C.M., and Thorpe, H.M. (2002). Diversity in the serine recombinases. *Mol Microbiol* *44*, 299–307.

- Smith-Mungo, L., Chan, I.T., and Landy, A. (1994). Structure of the P22 *att* site. Conservation and divergence in the lambda motif of recombinogenic complexes. *J Biol Chem* 269, 20798–20805.
- Stark, W.M. (2014). The Serine Recombinases. *Microbiol Spectr* 2, MDNA3-0046–2014.
- Stella, S., Cascio, D., and Johnson, R.C. (2010). The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev* 24, 814–826.
- Sternberg, N., Sauer, B., Hoess, R., and Abremski, K. (1986). Bacteriophage P1 *cre* gene and its regulatory region: Evidence for multiple promoters and for regulation by DNA methylation. *J Mol Biol* 187, 197–212.
- Stirling, C.J., Szatmari, G., Stewart, G., Smith, M.C., and Sherratt, D.J. (1988). The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus. *EMBO J* 7, 4389–4395.
- Stokes, H.W., and Hall, R.M. (1989). A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Mol Microbiol* 3, 1669–1683.
- 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.
- Subramanya, H.S., Arciszewska, L.K., Baker, R.A., Bird, L.E., Sherratt, D.J., and Wigley, D.B. (1997). Crystal structure of the site-specific recombinase, XerD. *EMBO J* 16, 5178–5187.
- Summers, D.K., and Sherratt, D.J. (1984). Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* 36, 1097–1103.
- Sun, X., Mierke, D.F., Biswas, T., Lee, S.Y., Landy, A., and Radman-Livaja, M. (2006). Architecture of the 99 bp DNA-six-protein regulatory complex of the lambda *att* site. *Mol Cell* 24, 569–580.
- Sutanto, Y., Shoemaker, N.B., Gardner, J.F., and Salyers, A.A. (2002). Characterization of Exc, a novel protein required for the excision of *Bacteroides* conjugative transposon. *Mol Microbiol* 46, 1239–1246.

- Sutanto, Y., DiChiara, J.M., Shoemaker, N.B., Gardner, J.F., and Salyers, A.A. (2004). Factors required in vitro for excision of the *Bacteroides* conjugative transposon, CTnDOT. *Plasmid* 52, 119–130.
- Swalla, B.M., Cho, E.H., Gumport, R.I., and Gardner, J.F. (2003). The molecular basis of cooperative DNA binding between lambda integrase and excisionase. *Mol Microbiol* 50, 89–99.
- Swinger, K.K., and Rice, P.A. (2004). IHF and HU: flexible architects of bent DNA. *Curr Opin Struct Biol* 14, 28–35.
- Sylwan, L., Frumerie, C., and Haggård-Ljungquist, E. (2010). Identification of bases required for P2 integrase core binding and recombination. *Virology* 404, 240–245.
- Tal, A., Arbel-Goren, R., Costantino, N., Court, D.L., and Stavans, J. (2014). Location of the unique integration site on an *Escherichia coli* chromosome by bacteriophage lambda DNA in vivo. *Proc Natl Acad Sci USA* 111, 7308–7312.
- Taylor, K.L., and Churchward, G. (1997). Specific DNA cleavage mediated by the integrase of conjugative transposon Tn916. *J Bacteriol* 179, 1117–1125.
- Tekle, M., Warren, D.J., Biswas, T., Ellenberger, T., Landy, A., and Nunes-Düby, S.E. (2002). Attenuating functions of the C terminus of lambda integrase. *J Mol Biol* 324, 649–665.
- Thompson, J.F., and Landy, A. (1988). Empirical estimation of protein-induced DNA bending angles: applications to lambda site-specific recombination complexes. *Nucleic Acids Res* 16, 9687–9705.
- Thompson, J.F., Waechter-Brulla, D., Gumport, R.I., Gardner, J.F., Moitoso de Vargas, L., and Landy, A. (1986). Mutations in an integration host factor-binding site: effect on lambda site-specific recombination and regulatory implications. *J Bacteriol* 168, 1343–1351.
- Thompson, J.F., Moitoso de Vargas, L., Koch, C., Kahmann, R., and Landy, A. (1987a). Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. *Cell* 50, 901–908.
- Thompson, J.F., de Vargas, L.M., Skinner, S.E., and Landy, A. (1987b). Protein-protein interactions in a higher-order structure direct lambda site-specific recombination. *J Mol Biol* 195, 481–493.

- Thorpe, H.M., and Smith, M.C.M. (1998). In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci USA* *95*, 5505–5510.
- Thorpe, H.M., Wilson, S.E., and Smith, M.C.M. (2000). Control of directionality in the site-specific recombination system of the *Streptomyces* phage ϕ C31. *Mol Microbiol* *38*, 232–241.
- Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., and Calos, M.P. (2001). Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage ϕ C31 Integrase. *Mol Cell Biol* *21*, 3926–3934.
- Tirumalai, R.S., Healey, E., and Landy, A. (1997). The catalytic domain of lambda site-specific recombinase. *Proc Natl Acad Sci USA* *94*, 6104–6109.
- Tirumalai, R.S., Kwon, H.J., Cardente, E.H., Ellenberger, T., and Landy, A. (1998). Recognition of core-type DNA sites by lambda integrase. *J Mol Biol* *279*, 513–527.
- Tong, W., Warren, D., Seah, N.E., Laxmikanthan, G., Van Duyne, G.D., and Landy, A. (2014). Mapping the λ Integrase bridges in the nucleoprotein Holliday junction intermediates of viral integrative and excisive recombination. *Proc Natl Acad Sci USA* *111*, 12366–12371.
- Trejo, C.S., Rock, R.S., Stark, W.M., Boocock, M.R., and Rice, P.A. (2018). Snapshots of a molecular swivel in action. *Nucleic Acids Res* *46*, 5286–5296.
- Umlauf, S.W., and Cox, M.M. (1988). The functional significance of DNA sequence structure in a site-specific genetic recombination reaction. *EMBO J* *7*, 1845–1852.
- Val, M.-E., Bouvier, M., Campos, J., Sherratt, D., Cornet, F., Mazel, D., and Barre, F.-X. (2005). The single-stranded genome of phage CTX Is the form used for integration into the genome of *Vibrio cholerae*. *Mol Cell* *19*, 559–566.
- Van Duyne, G.D. (2001). A structural view of Cre-*loxP* site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.* *30*, 87–104.
- Van Duyne, G.D. (2002). A structural view of tyrosine recombinase site-specific recombination. In *Mobile DNA II*, N.L. Craig, R. Craigie, M. Gellert, and A. Lambowitz, eds. (Washington, DC, USA: ASM Press), pp. 93–117.

- Van Duyne, G.D. (2015). Cre Recombinase. *Microbiol Spectr* 3, 3.1.10.
- Van Duyne, G.D., and Rutherford, K. (2013). Large serine recombinase domain structure and attachment site binding. *Crit Rev Biochem Mol Biol* 48, 476–491.
- Van Houdt, R., Leplae, R., Lima-Mendez, G., Mergeay, M., and Toussaint, A. (2012). Towards a more accurate annotation of tyrosine-based site-specific recombinases in bacterial genomes. *Mobile DNA* 3, 6.
- de Vargas, L.M., Pargellis, C.A., Hasan, N.M., Bushman, E.W., and Landy, A. (1988). Autonomous DNA binding domains of λ integrase recognize two different sequence families. *Cell* 54, 923–929.
- de Vargas, L.M., Kim, S., and Landy, A. (1989). DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. *Science* 244, 1457–1461.
- Vasala, A., Dupont, L., Baumann, M., Ritzenthaler, P., and Alatosava, T. (1993). Molecular comparison of the structural proteins encoding gene clusters of two related *Lactobacillus delbrueckii* bacteriophages. *J Virol* 67, 3061–3068.
- Venken, K.J.T., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: A BAC transgenic platform for targeted Insertion of large DNA fragments in *D. melanogaster*. *Science* 314, 1747–1751.
- Vit, C., Richard, E., Fournes, F., Whiteway, C., Eyer, X., Lapaillerie, D., Parissi, V., Mazel, D., and Loot, C. (2021). Cassette recruitment in the chromosomal Integron of *Vibrio cholerae*. *Nucleic Acids Res* 49, 5654–5670.
- Voziyanov, Y., Pathania, S., and Jayaram, M. (1999). A general model for site-specific recombination by the integrase family recombinases. *Nucleic Acids Res* 27, 930–941.
- Wang, J., Liu, Y., Liu, Y., Du, K., Xu, S., Wang, Y., Krupovic, M., and Chen, X. (2018). A novel family of tyrosine integrases encoded by the temperate pleolipovirus SNJ2. *Nucleic Acids Res* 46, 2521–2536.
- Warren, D., Sam, M.D., Manley, K., Sarkar, D., Lee, S.Y., Abbani, M., Wojciak, J.M., Clubb, R.T., and Landy, A. (2003). Identification of the lambda integrase surface that interacts with

Xis reveals a residue that is also critical for Int dimer formation. *Proc Natl Acad Sci USA* *100*, 8176–8181.

Warren, D., Lee, S.Y., and Landy, A. (2005). Mutations in the amino-terminal domain of lambda-integrase have differential effects on integrative and excisive recombination. *Mol Microbiol* *55*, 1104–1112.

Warren, D., Laxmikanthan, G., and Landy, A. (2008). A chimeric Cre recombinase with regulated directionality. *Proc Natl Acad Sci USA* *105*, 18278–18283.

Watt, V.M., Ingles, C.J., Urdea, M.S., and Rutter, W.J. (1985). Homology requirements for recombination in *Escherichia coli*. *Proc Natl Acad Sci USA* *82*, 4768–4772.

Weisberg, R.A., Enquist, L.W., Foeller, C., and Landy, A. (1983). Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. *J Mol Biol* *170*, 319–342.

Whiteson, K.L., Chen, Y., Chopra, N., Raymond, A.C., and Rice, P.A. (2007). Identification of a potential general acid/base in the reversible phosphoryl transfer reactions catalyzed by tyrosine recombinases: Flp H305. *Chem Biol* *14*, 121–129.

Williams, K.P. (2002). Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res* *30*, 866–875.

Wojciak, J.M., Connolly, K.M., and Clubb, R.T. (1999). NMR structure of the Tn916 integrase–DNA complex. *Nat Struct Mol Biol* *6*, 366–373.

Wojciak, J.M., Sarkar, D., Landy, A., and Clubb, R.T. (2002). Arm-site binding by lambda - integrase: solution structure and functional characterization of its amino-terminal domain. *Proc Natl Acad Sci USA* *99*, 3434–3439.

Wood, M.M., and Gardner, J.F. (2015). The integration and excision of CTnDOT. *Microbiol Spectr* *3*, MDNA3-0020–2014.

Wood, M.M., DiChiara, J.M., Yoneji, S., and Gardner, J.F. (2010). CTnDOT integrase interactions with attachment site DNA and control of directionality of the recombination reaction. *J Bacteriol* *192*, 3934–3943.

- Wozniak, R.A.F., and Waldor, M.K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat Rev Microbiol* 8, 552–563.
- Wu, N., Matand, K., Kebede, B., Acquah, G., and Williams, S. (2010). Enhancing DNA electrotransformation efficiency in *Escherichia coli* DH10B electrocompetent cells. *Elect J Biotech* 13, 21–22.
- Yagil, E., Dolev, S., Oberto, J., Kislev, N., Ramaiah, N., and Weisberg, R.A. (1989). Determinants of site-specific recombination in the lambdaoid coliphage HK022: An evolutionary change in specificity. *J Mol Biol* 207, 695–717.
- Yu, A., and Haggård-Ljungquist, E. (1993). Characterization of the binding sites of two proteins involved in the bacteriophage P2 site-specific recombination system. *J Bacteriol* 175, 1239–1249.
- Yuan, H.S., Finkel, S.E., Feng, J.-A., Kaczor-Grzeskowiak, M., Johnson, R.C., and Dickerson, R.E. (1991). The molecular structure of wild-type and a mutant Fis protein: Relationship between mutational changes and recombinational enhancer function or DNA binding. *Proc Natl Acad Sci USA* 9558–9562.
- Zimmerman, S.B., and Trach, S.O. (1991). Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J Mol Biol* 222, 599–620.

Annexes

Reprogramming the recombination specificity of a temperate phage integrase

Kevin Debatisse¹, Pierre Lopez¹, Michèle Coddeville², Philippe Rousseau², Muriel Cocaign-Bousquet¹ and Pascal Le Bourgeois^{1,2,*}

¹ TBI, Université de Toulouse, CNRS, INRAE, INSA, 31077 Toulouse, France

² LMGM, Université De Toulouse, UPS, CNRS, 31062 Toulouse, France

* To whom correspondence should be addressed. Tel: +33 561 559 438; Email: pascal.lebourgeois@univ-tlse3.fr

ABSTRACT << The abstract should be a single paragraph, not exceeding 200 words. URLs and references to figures or schemes should NOT be included. However, note that URLs MUST be included in the abstract of manuscripts submitted to the Database and Web Server issues. References should not be included in the abstract>>

INTRODUCTION

Viruses are the most abundant biological entities on Earth, with about $4 \cdot 10^{30}$ viruses in the ocean water (1). Among viruses, bacteriophages are the most abundant organisms with an estimation of more than 10^{30} tailed phages in the biosphere, outnumbering bacteria by a factor of about 10 to 1 (2). Bacteriophages are obligate parasites that can mediate two distinct lifecycles, the lytic and lysogenic cycles. During the infection, most bacteriophages proceed to the establishment of a lytic cycle, where the viruses hijack the host-cell machinery to replicate their DNA, to assemble viral particles and to allow their dissemination in the environment by lysis of their host. However, when the growth conditions are not favorable, some bacteriophages called temperate phages will be able to proceed through a lysogenic cycle resulting in the integration of the viral genome into its host chromosome (3–5). Once integrated, the bacteriophage genome, called prophage is passively replicated along with the bacterial chromosome. Phages and prophages can have significant impact on bacterial genomes through site-specific recombination and homologous recombination (6) and can account for up to 20 % of the bacterial DNA (7). Integration and excision is mediated by a phage-encoded protein called integrase. Integrases belong to the large family of site-specific recombinases and to the serine (S) and tyrosine (Y) recombinases subfamilies (8). Biologically, the integration and excision reactions are the reverse of each other. However, at the recombination mechanism level, these reactions are distinguished by the use of an additional protein for the excision, the phage-encoded excisionase (Xis) or RDF (Recombination Directionality Factors) (9, 10). Bacteriophage integrases catalyse integrative recombination between an

attP site contained in the phage DNA and an *attB* site located in the chromosome of the bacterial host cell, resulting to an integrated prophage flanked by hybrids recombination sites named *attL* and *attR*. An important feature of integrase-mediated recombination is the unidirectionality of the reaction. Indeed, excisive recombination needs in addition to integrase, an RDF protein, allowing the recombination between *attL* and *attR* sites, resulting in the excision of the bacteriophage, restoring the initial *attB* and *attP* sites. Integrase-mediated recombination is conservative as there is no DNA degradation, DNA synthesis, or the need for an energy cofactor (8). Serine integrases were discovered in the 1990's with the *Streptomyces* phage Φ C31 being the first studied (11, 12). Recently a significant advance in the serine integrases field was the structural determination of the C-terminal domain of LI integrase bound to an *attP* half-site (13). Since then, recent publications have focused on understanding the integration and excision mechanism mediated by the serine integrases (14–18). On the other hand, λ Int is the founding member of the Y-recombinase subfamily. The purification of this protein (19) and the development of an *in vitro* assay by Nash (20) have allowed λ Int to be meticulously studied for over 50 years (21–23). λ Int is a heterobivalent protein composed of 2 DNA-binding domains, a N-ter domain that recognizes arm-binding sites (24, 25) and of a C-terminal domain, containing the 7 catalytic conserved residues R-D/E-K-H-R-H/W-Y among the Y-recombinases (26) and that recognizes core-binding sites (25, 27). The *attP* site is complex (~ 240 bp), composed of 14 binding sites for 4 different proteins: 3 excisionase binding-sites, 3 IHF binding sites, 1 Fis binding site and 5 λ Int arm-binding sites and 2 core-binding sites arranged as inverted-repeats (C and C') that surround the 7 bp overlap sequence, which is the region where strand exchange takes place. *attB* is more simple, with a size of 21 bp, composed of 2 inverted repeats (B and B' core-binding sites) surrounding the overlap sequence (23). The recombination reaction involves the formation of a nucleoprotein complex on *attP* called intasome where *attP* supercoiling is required (28) and where the host factor IHF plays a fundamental architectural role (29). Once formed, this intasome can capture its "naked" *attB* site (30). Thanks to the various biochemical, functional and structural data, and to the recent FRET and electron cryo-microscopy results, it was possible to develop a precise model of each bacteriophage λ recombination pathway (31–33). Almost all of what we know about heterobivalent integrases come from studies of λ Int. However, other heterobivalent recombination systems have been studied, such as bacteriophages HP1 (34), KpIE1 (35), L5 (36), P2(37) , P22 (38). Heterobivalent integrases all retained an additional DNA-binding domain in N-ter as well as the use of a complex *attP*. Each of those systems uses a host factor

interposed between the integrase arm-binding and core-binding sites and the directionality of the reaction is controlled by an RDF protein that inhibits integration and is essential for excision. However, although the different protein-binding sites are conserved, the number, spacing and orientation are always different. Thus the assembly of a functional intasome on *attP* seems to be an essential parameter for the temperate phage life cycle using an heterobivalent integrase. On the other hand, it is possible that the formation of the intasome remain characteristic of each temperate phage / host-cell pair. In accordance with this idea, the higher-order structure characterized for HP1 (39), L5 (40–42) and KplE1 (43) are all different from each other and from the λ Int paradigm. Mv4 is a temperate bacteriophage infecting *Lactobacillus delbrueckii*, integrating its DNA in a ^{SER}tRNA gene and encoding a 427 aa protein belonging to the heterobivalent integrase subfamily of Y-recombinase (44, 45). Several studies have shown that ^{m_v4}Int/*attP*/*attB* recombination system is atypical compared to conventional systems, such as the λ recombination system: i) the ^{m_v4}Int-mediated recombination do not require host factors (46), ii) the *attP* and *attB* recombination sites have an atypical architecture, with core-regions lacking ^{m_v4}Int core-binding sites, with a non-canonical 8 bp overlap sequence and an unusual organization of arm-binding sites on *attP* (46, 47), iii) an asymmetric 16 bp *attB*, the shortest describe in the literature (48), iv) an important flexibility of ^{m_v4}Int towards the recognition of *attB* (49) and v) an excisionase protein essential for excision, but that does not inhibit integration (50). These differences suggest an original recombination mechanism among heterobivalent integrase. We were interested in the study of this mechanism, particularly by focusing on the atypical core-region of *attB* and *attP*. We carried out an exhaustive study of this core-region by the use of randomized libraries and EMSA methodology. Those results indicate that i) *attB* contains 5 additional nucleotides on the left, ii) the core-regions have an organization similar to classical systems, with the presence of inverted-repeated pattern (B/B' for *attB* and C/C' for *attP*) used as ^{m_v4}Int core-binding sites, surrounding a canonical 7 bp overlap and iii) the number and location of ^{m_v4}Int arm-binding sites also corresponds to classical systems with two pairs or ^{m_v4}Int arm-binding sites. On the other hand, ^{m_v4}Int differs from classical heterobivalent integrases by a greater tolerance towards the nature of the nucleotides constituting the core-regions, making it possible to define consensus sequences. This significant flexibility allowed us to site-specifically reprogram the recombination reaction to new targets, such as ^{SER}tRNA of other bacterial species, artificial sequences, or a chromosomal gene of *E. coli*, indicating a strong potential of the ^{m_v4}Int/*attP* system as a new tool for bacterial genomes engineering.

MATERIAL AND METHODS

Strains, plasmids, primers and media

The different strains and plasmids that were used in this study are listed in Supplementary Tables S1 and S2. All sequences of primers that were used are available in Supplementary Table S3. NEB5- α *repA*⁺ was built by using the protocol from Datsenko and Warner (51). It was constructed by replacing the *glgB* gene with the *glgB::Kan-repA* gene from *E. coli* EC1000. *E. coli* strains were grown in Lysogenic Broth (LB) at 37°C. Antibiotics were used at the following concentration : ampicillin, 100 µg/mL, carbenicillin, 100 µg/ml, chloramphenicol, 12,5 µg/ml, erythromycin, 150 µg/ml, kanamycin, 50 µg/ml.

DNA procedures

Standard techniques were used for DNA manipulation and cloning. Polymerase chain reaction (PCR) was performed with Q5-HF polymerase (New England Biolabs) or with CloneAmp Hifi polymerase (Takara Bio), according to the manufacturer's instructions. PCR products were purified using the QIAquick PCR purification kit (Qiagen). Plasmids were constructed using Gibson assembly (52) with NEBuilder HIFI DNA Assembly (New England Biolabs) or blunt-end cloning with T4 PNK (New England Biolabs) and T4 DNA ligase (New England Biolabs), according to the manufacturer's instructions. Plasmid DNA was extracted using QIAprep Spin Miniprep kit (Qiagen) or Nucleobond Xtra Midi (Macherey-Nagel) and their sequence was verified by Sanger sequencing (Mix2seq, Eurofins).

Purification of ^{mv4}Int and 6xHis-HU_{L1a}

For ^{mv4}Int purification, pET-Int (46) was transferred to *E. coli* strain BL21 (DE3; New England Biolabs). The resulting strains were grown in LB at 42°C, to an OD₆₀₀ of 0.6. IPTG (0.1 mM) was added to the medium and the culture was incubated at 22°C for 3h. Cells were recovered by centrifugation, resuspended in buffer [50 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole, 10% glycerol, 1 mg/ml lysozyme and SIGMAFAST Protease Inhibitor Cocktail Tablets EDTA-Free (Sigma)], sonicated and the lysate was cleared by centrifugation. ^{mv4}Int was purified on nickel resin (1ml His-trap HP, GE Healthcare) followed by gel filtration (HiLoad 16/60 Superdex 200, GE Healthcare) and heparin (1ml HiTrap Heparin HP, GE Healthcare) columns (Supplementary Fig S1). Purified integrase was stored at -80°C in buffer

containing 50 mM Tris pH 8, 500 mM NaCl, 20% glycerol, 1 mM DTT, 1 mM EDTA. For 6xHis-HU_{Lla} purification, pMET369 was transferred to *E. coli* BL21(DE3). The resulting strains were grown in LB at 37°C, to an OD₆₀₀ of 0.6. IPTG (0.4 mM) was added to the medium and the culture was incubated at 22°C for 3h. Cells were recovered by centrifugation, resuspended in buffer (50 mM Tris pH 8, 500 mM NaCl, 20 mM imidazole and SIGMAFAST Protease Inhibitor Cocktail Tablets EDTA-Free), sonicated and the lysate was cleared by centrifugation. 6xHis-HU_{Lla} was purified on nickel resin followed by heparin column (Supplementary Fig S2). Purified 6xHis-HU_{Lla} was stored at -80°C in buffer containing 50 mM Tris pH 8, 150 mM NaCl.

***In vitro* fluorescent assay**

Reaction mixtures (20 µl) contained 0.08 pmol (300 ng) of supercoiled plasmid carrying the *attP* site, 0.08 pmol (15 ng) of linear fluorescent (Cy3) 300-bp *attB* fragment, 7.2 pmol (300 ng) of ^{mv4}Int and sometimes 40 µg of heated crude-extract from *E. coli* BL21(DE3), in 25 mM Tris pH 7.5, 1 mM EDTA, 150 mM NaCl, 1 mM DTT, 10% PEG8000 (TENDP 1X). The reaction was incubated at 42°C 1h30 (when *E. coli* crude-extract was added) or 16h (when *E. coli* crude-extract was omitted) and it was terminated by adding 0.1% SDS. Samples were analyzed by electrophoresis in 0.8% agarose gels. Fluorescence was revealed by using the ChemidocMP imaging system (Biorad).

Constructing a randomized *attB* library and core-*attP* library

The randomized oligonucleotides (109 bp, *attB* library; 184 bp, core-*attP* library) were synthesized (IDT, USA). PCR was used to create double-stranded DNA using primers attBlibrary-F and attBlibrary-R for *attB* and attPlibrary-F and attPlibrary-R for *attP*. Each PCR product was separately cloned into pCC1Fos for *attB* libraries and pMET359 for *attP* libraries using Gibson assembly (52). Clones were propagated in *E. coli* EPI300 under chloramphenicol selection for *attB* libraries and NEB5-α *repA*⁺ under carbenicillin selection for *attP* libraries.

***In vitro* recombination assay using libraries**

Reaction (20 μ l) contained 0.08 pmol (450 ng) of *attB* plasmid, 0.08 pmol of *attP* plasmid, 7.2 pmol of mv^4Int and 40 μ g of heated crude-extract from *E. coli* BL21(DE3) in TENDP 1X and incubated 1h30 at 42°C. *attB*, *attP*, *attL* and *attR* sites were amplified by PCR using SeqbanqueattB-F/SeqbanqueattB-R primers for *attB*; SeqbanqueattP-F/SeqbanqueattP-R primers for *attP*; SeqbanqueattB-F/SeqbanqueattL-R primers for *attL* and SeqbanqueattR-F/seqbanqueattP-R primers for *attR*. PCR products were purified and analyzed by Sanger sequencing (Mix2seq, Eurofins).

NGS sequencing

Samples (*attL* from recombination *attB* Lib6 x *attP*_{WT}; *attB*_{WT} x *attP* Lib1 and *attR* from recombination *attB*_{WT} x *attP* Lib3; *attB* Lib8 x *attP*_{WT}) used for Sanger sequencing were also used for NGS sequencing (Eurofins). Data were analyzed on the Galaxy website (<https://usegalaxy.org>). *Sequence Logo* were generated using the Sequence Logo generator for fasta (53) and the occurrence of each word was characterized using the Wordcount program.

Electrophoretic Mobility Shift Assay

5' Cy3 end-labeled synthetic oligonucleotides (HPLC purified) were obtained from Eurofins. DNA substrates were prepared by hybridization of complementary labeled and unlabeled oligonucleotides in 10 mM Tris pH 7.5, 50 mM NaCl by incubating the samples 5 min at 95°C in a thermal cycler and reducing the temperature of 1.5°C/min until it reaches 25°C. Binding reactions (20 μ l) were done in buffer containing 25 mM Tris pH 8, 75 mM NaCl, 10% glycerol, 0.5 mM DTT, 0.5 mM EDTA, 1 μ g polydIdC (Sigma), 0.1 mg/ml BSA, and with 0.87 pmol labeled core-type DNA or arm-type DNA and 4.48 pmol unlabeled arm-type DNA or core-type DNA. The protein was mixed and incubated at room temperature for 20 min and samples were loaded onto a native, non-denaturing 7.5% polyacrylamide gel (Mini-PROTEAN TGX, Biorad). The gels were run at 4°C, 75V for 2 h. Fluorescence was revealed by using the ChemidocMP imaging system.

RESULTS

Redefining *mv4*Int sequence

An earlier study showed that *mv4*Int was a 427 AA protein, member of the Y site-specific recombinase family (44). However, new sequencing of the *mv4*Int gene revealed that the pMC1 plasmid (44) harbors 5 deletions and 4 nucleotide substitutions (Fig. S3A) compared to the published sequence. *mv4*Int amino acid sequence is deeply impacted by those variations (Fig. S3A). The various modifications of the nucleotide sequence generate an ORF encoding a 60 AA peptide and a shorter *mv4*Int protein of 369 AA containing 53 different AA compared to the published sequence. The nucleotide sequences of pA3Int (46) and pET-Int (46) as well as an amplicon from the *int* region of prophage mv4 also showed sequences corresponding to an integrase of 369 AA. A new alignment of the catalytic domain of *mv4*Int was realized with the catalytic domains of λ Int, Cre, XerC, XerD and ^{HP1}Int recombinases (Fig. S3B) and indicates that, unlike the published sequence, the new catalytic domain contains the 7 conserved residues R-D/E-K-H-R-H/W-Y of the Y recombinases catalytic pocket (26). We produced and purified the 369 AA *mv4*Int protein (Fig. S1), developed an *in vitro* fluorescent assay (Fig. S3C) and determined that the ORF8 peptide had no effect on the recombination indicating that only the 369 AA *mv4*Int protein is responsible of the recombination activity (Fig. S3D).

Global strategy of the use of randomized libraries

Due to the originality of the core region of *attP* and of the *attB* site, these regions were reanalyzed by an approach based on the use of randomized DNA libraries. Those DNA libraries corresponding either to the *attB* site or to the core region of the *attP* site, contained 7 to 10 randomized positions (Fig. 1A). Oligonucleotides containing at precise positions the 4 possible nucleotides were synthesized, amplified by PCR and cloned in *E. coli* in various plasmids (Fig. 1B). Each plasmid library is recovered, verified by Sanger sequencing, then used for *in vitro* recombination experiments with either the native partner site (*attP*_{WT} or *attB*_{WT}) or the associated partner library (same randomized region between the 2 sites). After recombination, *attL* and *attR* sites are amplified by PCR and sequenced (Fig. 1B). Among the randomized positions, only the nucleotides allowing the recombination will be recovered in these hybrid sites, which will make it possible to determine the constraints exerted on the nature of the nucleotide at each position tested.

Redefining the minimal *attB* site

The *attB*_{min} size was characterized by Auvray et al., and resulted in a 16 bp sequence, the shortest *attB* described in the literature (48). In order to validate the particular size of this site, different sizes of *attB*_{WT} were amplified by PCR (Fig. 1C) to obtain fluorescent *attB* fragments, which were tested by our *in vitro* assay against the *attP*_{WT} site (Fig. 1D). A recombination product is obtained with the *attB*_{WT} region and with the 23 bp *attB* site but not with the 16 bp site indicating that the published site is not functional and that *attB* requires a larger sequence on the left side. To precisely characterize the borders of *attB*, a library composed of 5 randomized positions overlapping each end of the published *attB* site was constructed (*attB* Lib9, Fig. 1A) and tested by *in vitro* recombination against the *attP*_{WT} site. After recombination, *attL* and *attR* sites were amplified by PCR and analyzed by Sanger sequencing (Fig. 2A). On the right side of *attB*, the *attR* site shows few constraints. Indeed, only positions 6 and 7 present a constraint (absence of G and A, respectively), indicating that the *attB* site ends with the sequence 5'-CTCCTT-3', in agreement with the published results (48). On the left side, strong constraints are observed on *attL*. Indeed, after recombination, only one base (C) out of 4 allows recombination at position 4 and 5. These positions are located in the overlap region (47) and where cytosines are the bases present in *attB*_{WT} and *attP*_{WT} indicating that the first two nucleotides in the *attB* and *attP* overlap sequence must be identical, as observed for most Y recombination systems. Constraints are also observed at position 1 (only purines are detected), a position described outside of the minimal *attB* site (48). This result is in agreement with our *in vitro* recombination result (Fig. 1D) and suggests a larger *attB* on the left side. To precisely characterize the beginning of the *attB* site, two other randomized libraries containing 10 randomized positions (Lib 5 and Lib6, Fig. 1A) were constructed and tested *in vitro* against the *attP*_{WT} site (Fig. 2B; Fig. 2C). Sequencing of the *attL* site from *attB* Lib6 x *attP*_{WT} recombination indicates that the first three positions do not show any constraints, suggesting that they are located outside of *attB* (Fig. 2B). This left border was also validated by the sequencing of the *attL* site from *attB* Lib5 X *attP*_{WT} recombination (Fig. 2C), which confirmed the absence of constraints at the same positions. Those results are in agreement and show that the *attB* site contains 5 additional bases on the left side compared to the published sequence and corresponds to a 21 bp site, starting with the sequence 5'-TTCAAAT-3'.

Characterization of the nucleotide constraints allowing recombination on the left side of *attB* and core-*attP*

The *attB* Lib6 library that allowed us to define a 21 bp *attB* also allows us to analyze the nucleotide constraints exerted on the left side of the *attB* overlap sequence (Fig. 2B). After recombination with *attP*_{WT}, a 7 bp degenerated pattern is observed. Positions 4, 5 and 6 show few constraints with an absence of C at positions 4 and 5 and absence of A at position 6. At positions 7, 8 and 9, a purine triplet appears to be essential for recombination. At position 10, only nucleotides A and T seem permissive for recombination. This result indicates the need for a 7 bp degenerated pattern 5'-DDBRRRW-3' on the left side of *attB* for the recombination. Given the localization of this pattern, it could be the B core-binding site of *attB*. An atypical characteristic of *attP*_{mv4} is the presence of an arm-binding site (P2) on the left side of the overlap region (Fig. 1A) (47). To characterize the constraints exerted on the left side of the core region of *attP*, a library containing 7 randomized positions (*attP* Lib3, Fig. 1A) was constructed and tested *in vitro* against the *attB*_{WT} site. After recombination (Fig. 2D), the degenerated pattern observed shows strong similarities with the degenerated pattern obtained on the left side of the *attB* overlap region. Indeed, the first two positions do not present a C and at the last position, only nucleotides A and T seem permissive for recombination. In addition, at positions 4, 5 and 6 there is a strong importance of purines, although a weaker T signal is observed at position 4 and 6. Thus, a 7 bp degenerated sequence 5'-DDNDRDW-3' seems essential for the recombination on the left side of the *attP* overlap region. Due to the fact that position 7, a position outside the P2 arm-type sequence, presents a strong constraints (only A and T seem allowed), and due to the strong similarity with the *attB* degenerated pattern, it is highly possible that the atypical P2 arm-binding site described in an earlier study (47) is actually the C core-binding site of *attP*.

Redefining the size of the *attB* and *attP* overlap region

An earlier study determined an *attB*_{mv4} and *attP*_{mv4} overlap region of 8 bp (47). The size of this region is atypical since all the heterobivalent recombination system currently characterized have an overlap sequence of 7 bp, as described for phages λ (54), HK022 (55), HP1 (56), L5 (57), P22 (38) as well as the ICE CTnDOT (58). Knowing the strand exchange mechanism used by γ recombinases makes it possible to characterize the overlap region indirectly, without using a cleavage experiment on suicide

substrates. Indeed, γ recombinases recombination reaction involves the formation and resolution of a Holliday junction where only the nucleotides present in the overlap region are exchanged (8). We took advantage of this property to reanalyze the overlap region of the $^{mv4}Int/attP/attB$ system. Two libraries containing 7 randomized positions with the first position corresponding to the 3' end of the published overlap sequence were constructed (*attB* Lib8 and *attP* Lib1, Fig. 1A). By following the fate of the first position of these libraries, it is theoretically possible to determine whether the size of the overlap region is 7 bp or 8 bp (Fig. 3A; Fig. 3B). Indeed, on the first hand, if the size of the overlap region is 8 bp, as the integrase-mediated recombination does not support heteroduplex, only the complementary base (T) to that of the native site (*attP*_{WT} for the *attP*_{WT} x *attB* Lib8 recombination and *attB*_{WT} for the *attB*_{WT} x *attP* Lib1 recombination) will be recovered after recombination. On the other hand, if the overlap region is 7 bp, in the *attP*_{WT} x *attB* Lib8 recombination, the wild-type sequence will be found on *attL* and the permissive nucleotides of the *attB* library will be found on *attR* i.e a D (A/G/T) since there is a constraint exerted at this position which will be analyzed later. In the *attB*_{WT} x *attP* Lib1 recombination, the logic is the same, the wild-type sequence will be found on *attR* and the permissive nucleotides of the *attP* library (D) will be found on *attL*. *attB* Lib8 and *attP* Lib1 libraries (Fig. 1A) were tested *in vitro* against *attP*_{WT} or *attB*_{WT}, respectively (Fig. 3C; Fig. 3D). To determine the size of the strand exchange area, only the first position of these randomized libraries will be analyzed in this part. The results observed after recombination i.e different *attL* and *attR* sites with only one of the sites presenting the permissive nucleotides of the libraries at the first position, are in agreement with the hypothesis of a 7 bp overlap, in disagreement with the published results (47).

Characterization of the nucleotide constraints of the *attB* and *attP* overlap region

The overlap sequence of the *mv4* bacteriophage recombination system is a canonical 7 bp sequence. In the paradigm λ Int, it was quickly shown that the nature of the bases in the overlap region was not important for recombination but that the sequence identity between *attB* and *attP* overlap is mandatory (59, 60). A similar feature is observed in the HK022 phage recombination system (61) and in the simple recombinases Flp (62) and Cre (63), although it has been shown that the presence of heteroduplex in the overlap of *loxP* sites can sometimes be functional (64, 65). In order to characterize the constraints exerted in the overlap region, two libraries composed of 7 randomized positions in the overlap region

were constructed (*attB* Lib7 and *attP* Lib2, Fig. 1A) and tested *in vitro* against *attP*_{WT} (Fig. 4A) or *attB*_{WT} (Fig. 4B), respectively. After recombination, among the 16,000 (4⁷) starting sequences, only the sequence corresponding to the native overlap sequence 5'-CCTGTAC-3' is found on *attL* and *attR*. These results indicate that the recombination mediated by *mv4*Int requires a strict identity between *attB* and *attP* overlap sequences. To verify if the strong constraints could be alleviated by the use of identical overlap sequences, the *attB* Lib7 and *attP* Lib2 libraries were tested against each other (Fig. 4C). After recombination, almost all positions show the 4 nucleotides which suggests that many sequences are functional. This result indicates that there is very low nucleotide constraints in the overlap region, probably as long as a strict identity between *attB* and *attP* overlap sequences is maintained. To validate this hypothesis, we tested the *attB* Lib7 library against 3 different *attP* overlap sequence (Fig. 4D). After recombination, only one sequence is functional in the *attB* Lib7 library, the sequence strictly identical to the *attP* overlap tested. This result indicates that the overlap sequence is changeable and that *mv4* recombination system behaves like the model Y recombinases. A weak signal of A at the 6th position is observed during the test of the 5'-TTTCGGC-3' overlap, suggesting that as for Cre, some heteroduplex may sometimes be functional. To obtain information on the effect of the sequence of the overlap region on the recombination activity, the 3 overlap sequences were also tested by the fluorescent *in vitro* assay (Fig. 4E). In this assay, the pair of *attB* and *attP* tested share the same overlap sequence. After quantification of the fluorescence, 3 effects are observed: no difference compared to the native overlap sequence (lane 1, Fig. 5E), a 2-fold decrease in recombination efficiency (lane 2, Fig. 4E) and a 15-fold decrease in recombination efficiency (lane 3, Fig. 4E). Hence, a strict identity is mandatory for *mv4* recombination, but sequence identity does not appear to be the only important parameter for the efficiency of the reaction.

Characterization of the nucleotide constraints allowing recombination on the right side of *attB* and *attP* overlap region.

As the overlap region is a 7 bp sequence, the *attB* Lib8 and *attP* Lib1 libraries used previously allow the characterization of the constraints exerted on the right-side of the overlap region (Fig. 3C; Fig. 3D). In this part, we will analyze the 7 randomized positions. After recombination against the native *attP* or *attB* site, almost identical constraints are observed. Indeed, at the first position, the C is absent. Next, a

pyrimidine triplet at position 2, 3 and 4 appears to be essential for recombination. At position 5, the T is absent, and at position 7, the A is absent. The only difference comes from position 6, where in addition to the A and T nucleotides, a weak signal of C is present for *attP_{WT}* x *attB* Lib 8 recombination. These results indicate that a 7 bp degenerated pattern is necessary for recombination on the right side of the overlap, a 5'-DYYYYVHB-3' pattern for *attB*, and a 5'-DYYYYWHB-3' pattern for *attP*. The libraries reveal the importance of a purine triplet on the left side of the overlap, and of a pyrimidine triplet on the right side of the overlap. This result suggests the presence of an inverted-repeated pattern surrounding the overlap, highlighting the potential core-binding sites, B and B' for *attB* and C and C' for *attP*.

B, B', C and C' core-site are ^{mv4}Int core-binding site

To experimentally demonstrate the presence of ^{mv4}Int core-binding sites, we used EMSA based on the protocol developed by Sarkar et al., on the COC' region of ^λInt recombination system (66). In this study, Sarkar et al., demonstrated that the N-ter domain exerts an inhibitory effect on the C-ter domain when it is not bound to the arm-binding sites. However, if DNA containing the P'12 arm-binding sites was added to the reaction, the inhibition was reversed and ^λInt was able to bind to the core-binding sites. To see if ^{mv4}Int N-ter domain also exerts an inhibitory effect on the C-ter domain, we compared ^{mv4}Int binding to labeled COC', in the presence and absence of unlabeled oligonucleotides containing the P'1-P'2 arm-type binding sites (Fig. 5A). ^{mv4}Int binding to the COC' sequence is quite unstable as a strong background smear is observed (lane 3, Fig. 5A). Addition of a 28 bp oligonucleotide containing the P'1-P'2 sequences greatly stimulates and stabilizes ^{mv4}Int binding to the COC' sequence and leads to the formation of 3 complexes (lane 4, Fig. 5A). Considered with data from other γ recombinase systems, our data suggests that complex I consists of a single ^{mv4}Int monomer bound to one core-binding site, that complex II consists of a single ^{mv4}Int monomer bound to one arm and one core-binding site and that complex III consists of a dimer of ^{mv4}Int bound to arm and core-binding sites. To ensure that the stabilization effect observed is due to the binding of the N-ter domain of ^{mv4}Int to the arm-type binding site, we added a 40 bp oligonucleotides, a larger size oligonucleotide of 12 bp including the P'1-P'2 arm binding sites (lane 5, Fig. 5A). We observed the same stabilization effect on the binding to the COC' sequence, and the formation of two complexes, II* and III* that migrate slower than the complex II and III, in agreement with a larger size of the unlabeled oligonucleotides. Now that we have the optimized

conditions for the binding to the core-sequence, two fluorescent fragments containing the 21 bp *attB* or *attP* core region were synthesized and tested for their ability to bind ^{mv4}Int (Fig. 5B; Fig. 5C). For both DNA, ^{mv4}Int binding is unstable in the absence of arm-binding sites and the addition of this unlabeled DNA results in the formation of the same complexes previously observed, suggesting the presence of 2 core-binding sites, B/B' on *attB* and C/C' on *attP*.

Asymmetric binding of ^{mv4}Int to the core-binding sites

^{mv4}Int is able to recognize two different B and C sequences, and two identical B' and C' sequences. In order to study the role as well as the affinity of ^{mv4}Int toward these 3 sequences, the B, C and C' sites were each inactivated, and the resulting DNA was tested for its ability to bind ^{mv4}Int (Fig. 5D; Fig. 5E; Fig. 5F). This results show a very weak fluorescence signal compared to the core regions containing 2 intact binding sites, indicating a strong cooperativity between ^{mv4}Int molecules for the binding to the core-type sites. ^{mv4}Int binds to the right arm (B'/C', Fig. 5D) with a much higher affinity than to the left arms (B/C, Fig. 5E; Fig. 5F). This result suggest that ^{mv4}Int binds preferentially to the B' and C' core-binding site and then recruits a monomer to the B and C sites. In agreement with this hypothesis, at high ^{mv4}Int concentration, complex III is detected, suggesting that ^{mv4}Int recruits another ^{mv4}Int monomer to the inactivated C binding site (lane 5 to 7, Fig. 5D).

***attP*_{mv4} contains two pairs of direct repeat of ^{mv4}Int arm-binding sites.**

As the EMSA methodology allows us to indirectly observe ^{mv4}Int binding to the arm-binding sites by stabilizing the binding of ^{mv4}Int to the core-type sites, we used this technique to verify if the 5 arm-binding sites described by Auvray et al., (46) allow the binding of ^{mv4}Int. To test this, we used different sets of ^{mv4}Int arm-binding sites (Fig. S4A). Among all the combinations used, only those containing either the P1-P2 pair or P'1-P'2 pair allow the stabilization of ^{mv4}Int binding to COC', therefore the binding of ^{mv4}Int N-ter domain. This results indicates that the P'3 sequence is not an arm-binding site, in disagreement with the published results (46). To understand the binding of the N-ter domain on the arm-type sites, different orientation of the P'1-P'2 pair was tested (i.e direct repeats and inverted repeats, Fig. S4B). The result indicates that ^{mv4}Int binds cooperatively to the arm-binding sites, only if those sites are in

direct repeats (Fig. S4B). We also used EMSA to characterize the size of the arm-binding sequence (Fig. S4B). We found that two direct repeats of 11 bp are necessary for the binding of ^{mv4}Int, suggesting that ^{mv4}Int binding sequence is a 11 bp sequence and not 9 bp as published (46). Finally, the atypical *attP_{mv4}* architecture returns to a more classic structure (Fig. S4C) with 2 pairs of ^{mv4}Int arm-binding sites, one on the P-arm and the other on the P'-arm, with a core regions consisting of 2 ^{mv4}Int core-binding sites surrounding the 7 bp overlap sequence.

^{mv4}Int exerts a reciprocal inhibition between its N-ter and C-ter domains.

In order to characterize ^{mv4}Int binding to its arm-binding sites, a fluorescent fragment containing the P'1-P'2 arm-binding sites was synthesized and tested for its ability to bind ^{mv4}Int (Fig. S4D). Whereas ^{mv4}Int binding to P'1-P'2 is quite unstable (lane 2, Fig. S4D), addition of unlabeled COC' sequence greatly stimulates and stabilizes ^{mv4}Int binding (lanes 3 to 7, Fig. S4D). This result suggests that as the N-ter domain exerts an inhibitory effect on the C-ter domain, the C-ter domain also exerts an inhibitory effect on the N-ter domain.

HU is an accessory protein of ^{mv4}Int-mediated recombination

To study the flexibility of ^{mv4}Int-mediated recombination, we tested the ability to reprogram the recombination to an unnatural *attB* site. To do this, we adapted the core region of *attP_{WT}*, i.e it has been modified to make it identical to the *attB* site tested. Initially, we used a cell extract that has overexpressed ^{mv4}Int for the *in vitro* assay and show that recombination was reprogrammable to a new *attB* site (Fig. 6A). However, when we used the purified protein, a huge decrease in recombination activity was observed for the unnatural *attP* x *attB* recombination (Fig. 6A). This difference in intensity, suppressed when the recombination reaction is elongated to 16 h (data not shown), suggests the existence of accessory factors in *E. coli*. This hypothesis was validated by adding to the reaction mixture containing ^{mv4}Int, a crude extract of either *E. coli* BL21(DE3) (Fig. 6A) or *L. lactis* (data not shown) that improved the recombination activity by a factor 3 for the native sites, as published (46) and by a factor 25 to 30 for recombination between modified sites. Heat treatment at 95°C of the crude extracts does not eliminate this effect, unlike treatment with proteinase K (data not shown) indicating that the accessory

element correspond to a heat-resistant protein shared between *E. coli* and *L. lactis*. As a potential candidate, we thought of HU, a ubiquitous protein where orthologs are found in all bacterial phyla as well as in Archaea, although it is not abundantly observed in these organisms (67). To test this hypothesis, we produced and purified HU from *L. lactis* (Fig. S2) and tested its effect on the recombination (Fig. 6B). In comparison to a heated crude extract that strongly stimulates the *attP*_{lactis} x *attB*_{lactis} recombination, the addition of increasing quantities of HU dose-dependently stimulates the recombination until an inhibitory effect at 1 µg of protein (Fig. 6B). This result show that HU is an accessory protein of the ^{mv4}Int-mediated recombination.

Recombination is reprogrammable if the targets respect the consensus sequence characterized

In order to see if it is possible to reprogram the recombination reaction towards target sites other than the natural *attB* sequence, we tried to redirect *in vitro* the specificity of this reaction towards different bacterial ^{SER}tRNA sequences, from the most to the less conserved genes (Fig. 6C). To do this, we used different variants of *attP* with the overlap and C' sequences modified to make it identical to the overlap and B' sequences of the different tRNA^{SER} tested (adapted *attP*, Fig. 6C). In total, 7 natural tRNA^{SER} sequences and 2 artificial sequences were tested. After recombination, 4 different profiles were obtained. The first profile (Fig. 6D) consists of tRNAs from *L. sakei* and *L. acidophilus* for which the recombination is possible with the native *attP*_{WT} and the adapted *attP*. *attB*_{sakei} is identical to *attB*_{WT} and *attB*_{acidophilus} contains an A in the first position of B', a nucleotide tolerated by ^{mv4}Int, therefore recombination is possible with the native and the modified *attP*. The second profile (Fig. 6E) consists of tRNAs from *L. lactis*, *S. mutans*, *L. reuteri* and artificial sequence 1 where recombination is not observed with the native *attP* (lane a, Fig. 6E). However, after adaptation of the core region of *attP* to the targeted *attB*, the recombination reaction is possible again (lane b, Fig. 6E). These *attB* sites contains modifications in the overlap and in the B' sequences that are all tolerated by ^{mv4}Int. The absence of identity in the overlap sequence explains the absence of recombination signal when tested with the native *attP*. Adaptation of the overlap and C' sequence of *attP* to each target restores the recombination reaction demonstrating that it is possible to reprogram the recombination to sequences other than the native *attB* site. Finally, the third group is composed of *E. faecalis* tRNA and of the artificial sequence 2

for which recombination reaction is not observed with either of the two *attP* sites (Fig. 6F). These sites have several modifications in the overlap and in the B' sequences. However, unlike the previous *attB* targets, among the modifications, two non-permissive nucleotides are found in the recombination sequence. Indeed, a C in first position of B' is observed for both sites, with a purine in the pyrimidine triplet for the artificial sequence 2 and a G at the 6th position of B' for *E. faecalis* tRNA. The results show that it is possible to reprogram *in vitro* the recombination reaction towards new *attB* target sequences by adapting the core region of *attP*. This reprogramming of the recombination seems only feasible if the targeted sequences respect the degenerated consensus sequence 5'-DDBRRRW NNNNNNN DYYYYVHB-3'.

***In vitro* targeting of the *E. coli lacZ* gene**

We performed an *in silico* analysis of the *E. coli* genome in order to identify putative recombination sites belonging to the degenerated consensus sequence. After analysis, 7959 putative sites were identified, and 3 were presents in the *lacZ* gene. We decided to test *in vitro* if the recombination reaction was reprogrammable to those 3 targets. To do this, we replaced the overlap region of *attP* by the overlap sequence of *lacZ* 1, 2 and 3 (Fig. 7A). We chose to keep the native C and C' sequences and to replace only the overlap sequence because according to our results obtained with randomized libraries, the nucleotide identity between the 2 overlaps sequence should be sufficient to obtain a site-specific recombination. To test if the recombination reaction was possible on these 3 sites, we used our fluorescent *in vitro* assay (Fig. 7B). For each *attB/attP* pair, no product was detected suggesting that if a recombination activity is possible, it is below the detection threshold of the experimental system used, i.e less than 5% of recombination compared to a native *attB/attP* pair (data not shown). The presence of recombination signals was then examined by employing the experimental methodology used with the randomized libraries by amplifying the *attL* and *attR* sites by PCR (Fig. 7C) and then analyzing them by Sanger sequencing (Fig. 7D). After recombination, PCR fragments of expected sizes were obtained for *attL* and *attR* from recombination of *lacZ* 1 and *lacZ* 2 *attB/attP* pair. On the other hand, a very low intensity band was detected for *attL* and *attR* resulting from recombination of *lacZ* 3 *attB/attP* pair. This result was expected because the replacement of the overlap of a native *attB* site by the sequence 5'-AAGGTTT-3' induces a 15-fold decrease in recombination activity (Fig. 4E). Analysis of the data by

Sanger sequencing revealed that the three recombination events observed correspond to site-specific recombination (Fig. 7D). Those results indicate that it is possible to specifically reprogram the recombination reaction to an *attB* site respecting the nucleotide constraints and exhibiting more than 50 % of nucleotide difference with *attB*_{WT}. Although this recombination is possible, the recombination activity is only detectable after *in vitro* amplification of the recombination products.

Characterization of the most functional B, B', C and C' sequences

In order to characterize the most functional B, B', C and C' sequences, *attL* from *attP*_{WT} x *attB* Lib6 (B core-site) and *attB*_{WT} x *attP* Lib1 (C' core-site) recombination and *attR* from *attP*_{WT} x *attB* Lib8 (B' core-site) and *attB*_{WT} x *attP* Lib3 (C core-site) recombination were analyzed by NGS sequencing. We first analyzed those data by generating "Sequence Logo" (68) representing the proportion of each nucleotides at the different randomized positions (Fig. 8). By comparison with the chromatograms obtained by Sanger sequencing, the "Sequence Logo" generated for each core-binding site are very similar both in terms of the nature and frequency of the nucleotides observed. In contrast, at some positions where Sanger sequencing detected only 2 to 3 nucleotides, the 4 nucleotides can sometimes be found by NGS sequencing (eg, position 2, 3 and 6 in B, Fig. 8). This indicates on one hand the greater resolution of NGS sequencing and on the other hand that some nucleotides that were described as non-permissive can be identified and therefore involved in a recombination event. After quantification of the occurrence of each sequence, a strong heterogeneity is observed in the representativeness of the sequences in each libraries. Indeed, among the permissive sequences, only to 2 to 6 % has a significant occurrence (> 1000). For example, depending on the libraries, the 100 most frequent sequences represent 10 to 35 % of the totality of the sequences obtained. This indicates that among the thousands of functional sequences, only a small percentage allows recombination at a significant activity. For each core-binding site, the 20 most frequent sequences were extracted and classified according to their occurrence and the native sequences of each core-binding sites were searched (red sequences, Fig. 8). For the C' library, the native sequence is located at the first position. In the previous part, it was shown that this sequence was the sequence exhibiting the strongest affinity among the B and C core-binding sites, in agreement with our NGS data. For the C library, the native sequence is only found at the 14th position indicating that other sequences may be more suitable for *in vitro* recombination.

However, those sequences only differ by 20 % in the occurrence, suggesting flexibility in the ^{mv4}Int recognition of the C core-binding site. For *attB*, the native B and B' sequences do not seem to be the most optimal sequences for *in vitro* recombination since they are found at the 23rd and 75th position, respectively (Fig. 8). For the B' library, the sequence at the first position has a frequency 16 times higher than the native B' sequence. It is possible that an experimental bias is responsible for this abnormally high occurrence, either during the synthesis and construction of the library or during the PCR amplification of the hybrid sites. Interestingly the 3 *lacZ* sites all contained one or two poor recombination substrates, i.e a B or B' sequences below the first 100 most functional sequences. This result can explain why the recombination activity is only detectable after PCR amplification of the recombination product, indicating that despite the fact that the *lacZ* sequences respect the 5'-DDBRRRW NNNNNNN DYYYYVHB-3' pattern, they are not the most optimal sequences for recombination. Sequences carrying highly present B and B' sequences should be tested for reprogramming of the *in vitro* recombination.

DISCUSSION

In the literature, ^{mv4}Int differs from heterobivalent integrases due to multiple atypical characteristics. Indeed, the mv4 system does not need host factors for recombination (46), does not contain core-binding sites (46, 48), has an atypical architecture of ^{mv4}Int arm-binding sites (47), an unusual overlap size (47) and an asymmetric *attB* which is the shortest described in the literature (48). The work described in this paper allowed us to confirm and clarify some aspects. The protein sequence of ^{mv4}Int has been corrected. The atypical independence for host factors was confirmed during native *attB* x *attP* recombination but an ubiquitous factor, HU, strongly stimulates recombination during modified *attB* x *attP* recombination. The use of HU protein as a cofactor is frequent among the family of β -recombinases, proteins belonging to the small serine recombinases family (69–71). HU is also an accessory protein for the recombination catalyzed by Hin invertase (72). Among the tyrosine recombinases, a large majority of heterobivalent recombination system use IHF as a cofactor (34, 38, 73, 74) . For the paradigm λ Int, the host factor IHF was partially replaced by HU (75, 76). Only the ICE Tn916 use HU as an accessory protein in *E. coli* (77) and maybe the ICE CTnDOT as the host factor BHFa could be the HU protein of *Bacteroides thetaiotaomicron* (78). We used randomized libraries to characterize the unusual core region of *attB* and *attP*. This methodology is remarkably suited for the study of heterobivalent recombination systems. Indeed, this methodology allowed us to rectify the minimum size of *attB*, for

which a 16 bp had previously been described (48). The use of randomized libraries allowed on the one hand to show that *attB* is in fact a larger site, with a size of 21 bp and on the other hand to highlight the presence of inverted-repeated regions, making it possible to propose a more traditional structure of the core regions (BOB' on *attB* and COC' on *attP*). The size of *attB_{mv4}* returns to a more classic size because most of the characterized *attB* site from heterobivalent systems have a size between 18 bp (56) and 29 bp (57). The association of randomized libraries with EMSA allowed us to demonstrate experimentally that these inverted-repeated regions corresponded to the core-binding sites. EMSA also allowed us to precisely identify the arm-binding sites and to return to a more classic structure of the *attP* site, with two pairs of direct-repeats, one on the P-arm and the other on the P'-arm, as described for phage L5 (79), P2 (74) and P22 (38). During EMSA experiments, we were able to demonstrate that *mv4*Int exerts a reciprocal inhibition between its N and C-terminal domains. One explanation for the reciprocal inhibition may lie in the structure of *mv4*Int. It is possible that the two DNA-binding domains are close and occluding each other and that DNA binding by one domain would free and expose the second DNA-binding domain. For the model *λ*Int, it was demonstrated that the N-ter domain exerts an inhibitory effect on the C-ter domain, but the C-ter domain does not exert any inhibition on the N-ter domain (66). In the opposite way, IntDOT, the heterobivalent integrase of the ICE CTnDOT, seem to have no inhibitory effect exerted by the N-ter domain, but an inhibitory effect exerted by the C-ter domain (80). Thus, although heterobivalent integrases have retained two different DNA-binding domains as well as a complex structure of their *attP* site, the regulation of each domain appears to be different depending on the system studied. The use of randomized libraries also allowed us to follow the fate of each strand during recombination, and therefore to localize with precision the two *mv4*Int cleavage sites. This results show that the size of the overlap sequence is not an 8 bp sequence as published (47), but a canonical sequence of 7 bp, as described for all recombination sites of bacteriophages heterobivalent system. To our knowledge, it is the first time in the literature that the integrase cleavage sites are determined in this way. Historically, the cleavage sites have been characterized by strand-breakage reaction on a suicide substrate, but with this technique, it can sometimes be complicated to precisely position the cleavage sites, thus, several positions can sometimes be proposed (56). The use of randomized libraries there appears to be a new tool allowing the precise characterization of the borders of the overlap sequence. Due to its simplicity, it could for example be used in the P2 heterobivalent recombination system where the *attB* and *attP* core region are complex, and the overlap sequence has not yet been identified (81).

The randomized libraries allowed us to characterize the nucleotide constraints exerted on the core-binding sites and on the overlap sequence. In the overlap sequence, we found that a strict identity between *attB* and *attP* is mandatory, and that there is very few nucleotide constraints. However, we found that overlap sequence identity does not appear to be the only important parameter for the efficiency of the recombination, as observed for Cre (64, 65), Flp (82) and λ Int (83). We also found that unlike most Y-recombinases, $mv4$ Int exerts a strong flexibility in its ability to recognize the core-binding sites. Indeed, we were able to determine degenerated consensus sequences of the B, B', C and C' core-binding sites. We were also able to reprogram the specificity of the recombination towards new targets, 6 natural tRNA^{SER} sequences, 1 artificial sequence and 3 putative recombination sites in *E. coli lacZ* gene, as long as those sequences strictly respect the *attB* consensus sequence. Among heterobivalent integrases, few studies have focused on modifications of the B and B' sequences tolerated by the integrase. Indeed, in the λ Int model, while it was shown that the overlap region was modifiable as long as the sequence identity between *attB* and *attP* was conserved (59, 60), little information of the nucleotide constraint exerted on the B and B' core-binding sites are available. A study however focused on the secondary integration sites of phage λ (84) and showed that despite the modifications of the B and B' sites, a very large majority of the secondary sites contained the invert-repeated sequence 5'-CTT-3'. This result indicates that λ Int is able to bind sequences other than the native B and B' sequences, but with little flexibility since both sequences must have the 5'-CTT-3' triplet. The same requirement was observed for $HK022$ Int (61, 85). Unlike conventional heterobivalent integrases where an inverted-repeat sequence is necessary for the reaction, $mv4$ Int shows a strong flexibility since this protein is able to recognize 3 different sequences. In addition, the NGS data show that among the 20 most frequent B and B' sequences, there is no invariable native nucleotide, thus reinforcing the hypothesis of a greater flexibility towards its recombination sites. To our knowledge, the $mv4$ Int/*attP*/*attB* core-region is the most characterized recombination system among the heterobivalent recombinases. It would be interesting to use the randomized libraries methodology in other heterobivalent systems in order to observe if the important flexibility of $mv4$ Int is a characteristic of this protein or shared with other heterobivalent integrases. Mv4 belongs to the class III of bacteriophages that integrates their DNA into tRNA genes (86). This class of bacteriophages integrates their DNA in the 3' end of a tRNA gene and therefore uses an asymmetric *attB* site, which does not contain inverted-repeat sequence. It will be interesting to see if this class of bacteriophages also presents a strong flexibility towards the recognition of their

recombination sites. Our data indicates that to reprogram the recombination towards a target sequence with a significant activity, it would be necessary to target a sequence respecting the nucleotide constraints characterized on *attB*, and that this sequence also contains highly represented B and B' sequences. Arbitrarily, by choosing only the first 75 B and B' sequences, and by adding an overlap sequence composed of 7xN, 200 putatives sites are found in *E. coli* and 300 in *L. lactis*. The use of a universal accessory protein, and the hundreds of putative sites found in bacterial genomes suggest the possibility of using $m^{v4}Int$ as a genetic engineering tool, allowing the integration of DNA "à la carte" in all bacterial phyla.

AVAILABILITY

<<Please provide links to all websites where data, software or tools referred to in your article might be hosted, (e.g. github). An example is provided below>>

Genome Maps is an open source collaborative initiative available in the GitHub repository (<https://github.com/compbio-bigdata-viz/genome-maps>)

ACCESSION NUMBERS

<<Please list here all the public repository accession numbers (e.g. PDB, GEO). An example is provided below>>

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data bank under accession number XXXX.

SUPPLEMENTARY DATA

<<Supplementary data should consist of electronic files and should not merely be a link to another web site. All supplementary data MUST be referred to in the main manuscript at an appropriate point in the text. The supplementary data should preferably be saved as one single PDF file, including all text, figures, tables and legends. If this is not possible, a maximum of 10 files is acceptable to make up the supplementary data unit for the article (see below). The maximum size per file should not exceed 1.5 MBytes, and files must be as small as possible so that they can be downloaded quickly.

1. Pick a common cross-platform (PC, Mac, Linux/UNIX, Amiga, etc.) format for your Supplementary data to allow the greatest access.
2. If you cannot convert all your files to PDF (.pdf), please provide the following:
 - text files in MS Word (.doc), HTML (.html) or RTF (.rtf) format.
 - spreadsheet files in MS Excel (.xls) or CSV format. Where possible, combine all tables into a single Excel workbook, saving individual tables on separate clearly labelled worksheets (tabs).

- image files in tif, gif or jpg format. Images should be a maximum size of 640 x 480 pixels (9 x 6.8 inches at 72 pixels per inch).
- sound clips in mp3 format.
- movie clips in mpeg format

Supplementary Data statement:

If applicable, the following statement must be added at the end of the manuscript>>

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog. The quick brown fox jumps over the lazy dog.

FUNDING

<< Details of all funding sources for the work in question, and for the Open Access charge, should be given. An example is shown below.

The following rules should be followed:

- The full official funding agency name should be given, i.e. 'National Institutes of Health', not 'NIH' (full RIN-approved list of UK funding agencies) Grant numbers should be given in brackets as follows: '[grant number xxxx]'
- Multiple grant numbers should be separated by a comma as follows: '[grant numbers xxxx, yyyy]'
- Agencies should be separated by a semi-colon (plus 'and' before the last funding agency)
- Where individuals need to be specified for certain sources of funding the following text should be added after the relevant agency or grant number 'to [author initials]'>>
- The statement should end with 'Funding for open access charge: [insert grant agency name/number]'>>

This work was supported by the National Institutes of Health [AA123456 to A.B., BB123456 to C.D.]; and the Alcohol & Education Research Council [abcde123456]. Funding for open access charge: National Institutes of Health.

CONFLICT OF INTEREST

<<Please disclose any financial interests or connections, direct or indirect, or other situations that might raise the question of bias in the work reported or the conclusions, implications, or opinions stated - including

pertinent commercial or other sources of funding for the individual author(s) or for the associated department(s) or organization(s), personal relationships, or direct academic competition. Further information is available at https://academic.oup.com/journals/pages/authors/authors_fags/conflicts_of_interest>>

REFERENCES

1. Suttle, C.A. (2005) Viruses in the sea. *Nature*, **437**, 356–361.
2. Brüssow, H. and Hendrix, R. (2002) Phage Genomics: Small Is Beautiful. *Cell*, **108**, 13–6.
3. Fogg, P.C.M., Colloms, S., Rosser, S., Stark, M. and Smith, M.C.M. (2014) New applications for phage integrases. *J Mol Biol*, **426**, 2703–2716.
4. Groth, A.C. and Calos, M.P. (2004) Phage Integrases: Biology and Applications. *J Mol Biol*, **335**, 667–678.
5. Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I. and Herskovits, A.A. (2015) A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol*, **13**, 641–650.
6. Menouni, R., Hutinet, G., Petit, M.-A. and Ansaldi, M. (2015) Bacterial genome remodeling through bacteriophage recombination. *FEMS Microbiol Lett*, **362**, 1–10.
7. Casjens, S. (2003) Prophages and bacterial genomics: what have we learned so far? *Mol Microbiol*, **49**, 277–300.
8. Grindley, N.D.F., Whiteson, K.L. and Rice, P.A. (2006) Mechanisms of site-specific recombination. *Annu Rev Biochem*, **75**, 567–605.
9. Abremski, K. and Gottesman, S. (1982) Purification of the bacteriophage lambda *xis* gene product required for lambda excisive recombination. *J Biol Chem*, **257**, 9658–9662.
10. Lewis, J.A. and Hatfull, G.F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res*, **29**, 2205–2216.
11. Kuhstoss, S. and Rao, R.N. (1991) Analysis of the integration function of the streptomycete bacteriophage ϕ C31. *J Mol Biol*, **222**, 897–908.
12. Thorpe, H.M. and Smith, M.C.M. (1998) In vitro site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc Natl Acad Sci USA*, **95**, 5505–5510.
13. Rutherford, K., Yuan, P., Perry, K., Sharp, R. and Van Duyne, G.D. (2013) Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res*, **41**, 8341–8356.
14. Gupta, K., Sharp, R., Yuan, J.B., Li, H. and Van Duyne, G.D. (2017) Coiled-coil interactions mediate serine integrase directionality. *Nucleic Acids Res*, **45**, 7339–7353.
15. Mandali, S., Gupta, K., Dawson, A.R., Van Duyne, G.D. and Johnson, R.C. (2017) Control of recombination directionality by the *Listeria* phage A118 protein Gp44 and the coiled-coil motif of its serine integrase. *J Bacteriol*, **199**, e00019-17.
16. Mandali, S. and Johnson, R.C. (2021) Control of the Serine Integrase Reaction: Roles of the Coiled-Coil and Helix E Regions in DNA Site Synapsis and Recombination. *J Bacteriol*, **203**, e0070320.

17. Fogg,P.C.M., Younger,E., Fernando,B.D., Khaleel,T., Stark,W.M. and Smith,M.C.M. (2018) Recombination directionality factor gp3 binds ϕ C31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. *Nucleic Acids Res*, **46**, 1308–1320.
18. Li,H., Sharp,R., Rutherford,K., Gupta,K. and Van Duyne,G.D. (2018) Serine integrase *attP* binding and specificity. *J Mol Biol*, **430**, 4401–4418.
19. Nash,H.A. (1974) Purification of bacteriophage λ , Int protein. *Nature*, **247**, 543–545.
20. Nash,H.A. (1975) Integrative recombination of bacteriophage lambda DNA in vitro. *Proc Natl Acad Sci USA*, **72**, 1072–1076.
21. Landy,A. (1989) DYNAMIC, STRUCTURAL, AND REGULATORY ASPECTS OF λ SITE-SPECIFIC RECOMBINATION. *Annu Rev Biochem*, **58**, 913–941.
22. Azaro,M.A. and Landy,A. (2002) λ integrase and the λ Int family. In Craig,N.L., Craigie,R., Gellert,M., Lambowitz,A. (eds), *Mobile DNA II*. ASM Press, Washington, DC, USA, pp. 118–148.
23. Landy,A. (2015) The λ Integrase site-specific recombination pathway. *Microbiol Spectr*, **3**, MDNA3-0051–2014.
24. Sarkar,D., Azaro,M.A., Aihara,H., Papagiannis,C.V., Tirumalai,R., Nunes-Düby,S.E., Johnson,R.C., Ellenberger,T. and Landy,A. (2002) Differential Affinity and Cooperativity Functions of the Amino-terminal 70 Residues of λ Integrase. *J Mol Biol*, **324**, 775–789.
25. de Vargas,L.M., Pargellis,C.A., Hasan,N.M., Bushman,E.W. and Landy,A. (1988) Autonomous DNA binding domains of λ integrase recognize two different sequence families. *Cell*, **54**, 923–929.
26. Gibb,B., Gupta,K., Ghosh,K., Sharp,R., Chen,J. and Van Duyne,G.D. (2010) Requirements for catalysis in the Cre recombinase active site. *Nucleic Acids Res*, **38**, 5817–5832.
27. Tirumalai,R.S., Healey,E. and Landy,A. (1997) The catalytic domain of lambda site-specific recombinase. *Proc Natl Acad Sci USA*, **94**, 6104–6109.
28. Richet,E., Abcarian,P. and Nash,H.A. (1986) The interaction of recombination proteins with supercoiled DNA: Defining the role of supercoiling in lambda integrative recombination. *Cell*, **46**, 1011–1021.
29. de Vargas,L.M., Kim,S. and Landy,A. (1989) DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. *Science*, **244**, 1457–1461.
30. Richet,E., Abcarian,P. and Nash,H.A. (1988) Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell*, **52**, 9–17.
31. Laxmikanthan,G., Xu,C., Brilot,A.F., Warren,D., Steele,L., Seah,N., Tong,W., Grigorieff,N., Landy,A. and Van Duyne,G.D. (2016) Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *eLife*, **5**, e14313.
32. Seah,N.E., Warren,D., Tong,W., Laxmikanthan,G., Van Duyne,G.D. and Landy,A. (2014) Nucleoprotein architectures regulating the directionality of viral integration and excision. *Proc Natl Acad Sci USA*, **111**, 12372–12377.
33. Tong,W., Warren,D., Seah,N.E., Laxmikanthan,G., Van Duyne,G.D. and Landy,A. (2014) Mapping the λ Integrase bridges in the nucleoprotein Holliday junction intermediates of viral integrative and excisive recombination. *Proc Natl Acad Sci USA*, **111**, 12366–12371.
34. Goodman,S.D. and Scoocca,J.J. (1989) Nucleotide sequence and expression of the gene for the site-specific integration protein from bacteriophage HP1 of *Haemophilus influenzae*. *J Bacteriol*, **171**, 4232–4240.

35. Rudd, K.E. (1999) Novel intergenic repeats of *Escherichia coli* K-12. *Res Microbiol*, **150**, 653–664.
36. Lee, M.H., Pascopella, L., Jacobs, W.R. and Hatfull, G.F. (1991) Site-specific integration of mycobacteriophage L5: integration-proficient vectors for *Mycobacterium smegmatis*, *Mycobacterium tuberculosis*, and bacille Calmette-Guérin. *Proc Natl Acad Sci USA*, **88**, 3111–3115.
37. Barreiro, V. and Haggård-Ljungquist, E. (1992) Attachment sites for bacteriophage P2 on the *Escherichia coli* chromosome: DNA sequences, localization on the physical map, and detection of a P2-like remnant in *E. coli* K-12 derivatives. *J Bacteriol*, **174**, 4086–4093.
38. Smith-Mungo, L., Chan, I.T. and Landy, A. (1994) Structure of the P22 *att* site. Conservation and divergence in the lambda motif of recombinogenic complexes. *J Biol Chem*, **269**, 20798–20805.
39. Esposito, D., Thrower, J.S. and Scoocca, J.J. (2001) Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation. *Nucleic Acids Res*, **29**, 3955–3964.
40. Peña, C.E., Kahlenberg, J.M. and Hatfull, G.F. (1999) Protein-DNA complexes in mycobacteriophage L5 integrative recombination. *J Bacteriol*, **181**, 454–461.
41. Peña, C.E., Kahlenberg, J.M. and Hatfull, G.F. (2000) Assembly and activation of site-specific recombination complexes. *Proc Natl Acad Sci USA*, **97**, 7760–7765.
42. Lewis, J.A. and Hatfull, G.F. (2003) Control of directionality in L5 integrase-mediated site-specific recombination. *J Mol Biol*, **326**, 805–821.
43. Panis, G., Duverger, Y., Champ, S. and Ansaldi, M. (2010) Protein binding sites involved in the assembly of the KpIE1 prophage intasome. *Virology*, **404**, 41–50.
44. Dupont, L., Boizet-Bonhoure, B., Coddeville, M., Auvray, F. and Ritzenthaler, P. (1995) Characterization of genetic elements required for site-specific integration of *Lactobacillus delbrueckii* subsp. *bulgaricus* bacteriophage mv4 and construction of an integration-proficient vector for *Lactobacillus plantarum*. *J Bacteriol*, **177**, 586–595.
45. Nunes-Düby, 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.
46. Auvray, F., Coddeville, M., Espagno, G. and Ritzenthaler, P. (1999) Integrative recombination of *Lactobacillus delbrueckii* bacteriophage mv4: functional analysis of the reaction and structure of the *attP* site. *Mol Gen Genet*, **262**, 355–366.
47. Coddeville, M., Spinella, J.-F., Cassart, P., Girault, G., Daveran-Mingot, M.-L., Le Bourgeois, P. and Ritzenthaler, P. (2014) Bacteriophage mv4 site-specific recombination: the central role of the P2 mv4Int-binding site. *J Virol*, **88**, 1839–1842.
48. Auvray, F., Coddeville, M., Ordonez, R.C. and Ritzenthaler, P. (1999) Unusual structure of the *attB* site of the site-specific recombination system of *Lactobacillus delbrueckii* bacteriophage mv4. *J Bacteriol*, **181**, 7385–7389.
49. Auvray, F., Coddeville, M., Ritzenthaler, P. and Dupont, L. (1997) Plasmid integration in a wide range of bacteria mediated by the integrase of *Lactobacillus delbrueckii* bacteriophage mv4. *J Bacteriol*, **179**, 1837–1845.
50. Coddeville, M. and Ritzenthaler, P. (2010) Control of directionality in bacteriophage mv4 site-specific recombination: functional analysis of the Xis factor. *J Bacteriol*, **192**, 624–635.
51. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA*, **97**, 6640–6645.

52. Gibson,D.G., Young,L., Chuang,R.-Y., Venter,J.C., Hutchison,C.A. and Smith,H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343–345.
53. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res*, **14**, 1188–1190.
54. Craig,N.L. and Nash,H.A. (1983) The mechanism of phage lambda site-specific recombination: site-specific breakage of DNA by Int topoisomerase. *Cell*, **35**, 795–803.
55. Kolot,M. and Yagil,E. (1994) Position and direction of strand exchange in bacteriophage HK022 integration. *Mol Gen Genet*, **245**, 623–627.
56. Hauser,M.A. and Scocca,J.J. (1992) Site-specific integration of the *Haemophilus influenzae* bacteriophage HP1. Identification of the points of recombinational strand exchange and the limits of the host attachment site. *J Biol Chem*, **267**, 6859–6864.
57. Peña,C.E., Stoner,J.E. and Hatfull,G.F. (1996) Positions of strand exchange in mycobacteriophage L5 integration and characterization of the *attB* site. *J Bacteriol*, **178**, 5533–5536.
58. Malanowska,K., Salyers,A.A. and Gardner,J.F. (2006) Characterization of a conjugative transposon integrase, IntDOT. *Mol Microbiol*, **60**, 1228–1240.
59. Weisberg,R.A., Enquist,L.W., Foeller,C. and Landy,A. (1983) Role for DNA homology in site-specific recombination. The isolation and characterization of a site affinity mutant of coliphage lambda. *J Mol Biol*, **170**, 319–342.
60. Bauer,C.E., Gardner,J.F. and Gumpert,R.I. (1985) Extent of sequence homology required for bacteriophage lambda site-specific recombination. *J Mol Biol*, **181**, 187–197.
61. Kolot,M., Malchin,N., Elias,A., Gritsenko,N. and Yagil,E. (2015) Site promiscuity of coliphage HK022 integrase as a tool for gene therapy. *Gene Ther*, **22**, 521–527.
62. McLeod,M., Craft,S. and Broach,J.R. (1986) Identification of the crossover site during FLP-mediated recombination in the *Saccharomyces cerevisiae* plasmid 2 microns circle. *Mol Cell Biol*, **6**, 3357–3367.
63. Hoess,R.H., Wierzbicki,A. and Abremski,K. (1986) The role of the *loxP* spacer region in P1 site-specific recombination. *Nucleic Acids Res*, **14**, 2287–2300.
64. Lee,G. and Saito,I. (1998) Role of nucleotide sequences of *loxP* spacer region in Cre-mediated recombination. *Gene*, **216**, 55–65.
65. Sheren,J., Langer,S.J. and Leinwand,L.A. (2007) A randomized library approach to identifying functional *lox* site domains for the Cre recombinase. *Nucleic Acids Res*, **35**, 5464–5473.
66. Sarkar,D., Radman-Livaja,M. and Landy,A. (2001) The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions. *EMBO J*, **20**, 1203–1212.
67. Johnson,R.C., Johnson,L.M., Schmidt,J.W. and Gardner,J.F. (2004) Major nucleoid proteins in the structure and function of the *Escherichia coli* chromosome. In Higgins,N.P. (ed), *The bacterial chromosome*. ASM Press, Washington, DC, USA, pp. 65–132.
68. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, **18**, 6097–6100.
69. Alonso,J.C., Weise,F. and Rojo,F. (1995) The *Bacillus subtilis* histone-like protein Hbsu is required for DNA resolution and DNA inversion mediated by the beta recombinase of plasmid pSM19035. *J Biol Chem*, **270**, 2938–2945.

70. Petit, M.A., Ehrlich, D. and Janni re, L. (1995) pAM beta 1 resolvase has an atypical recombination site and requires a histone-like protein HU. *Mol Microbiol*, **18**, 271–282.
71. Rowland, S.-J., Stark, W.M. and Boocock, M.R. (2002) Sin recombinase from *Staphylococcus aureus*: synaptic complex architecture and transposon targeting. *Mol Microbiol*, **44**, 607–619.
72. Johnson, R.C., Bruist, M.F. and Simon, M.I. (1986) Host protein requirements for in vitro site-specific DNA inversion. *Cell*, **46**, 531–539.
73. Panis, G., M jean, V. and Ansaldi, M. (2007) Control and regulation of KpIE1 prophage site-specific recombination: a new recombination module analyzed. *J Biol Chem*, **282**, 21798–21809.
74. Yu, A. and Hagg rd-Ljungquist, E. (1993) Characterization of the binding sites of two proteins involved in the bacteriophage P2 site-specific recombination system. *J Bacteriol*, **175**, 1239–1249.
75. Goodman, S.D., Nicholson, S.C. and Nash, H.A. (1992) Deformation of DNA during site-specific recombination of bacteriophage A: Replacement of IHF protein by HU protein or sequence-directed bends. *Proc Natl Acad Sci USA*, **89**, 11910–11914.
76. Segall, A.M., Goodman, S.D. and Nash, H.A. (1994) Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. *EMBO J*, **13**, 4536–4548.
77. Connolly, K.M., Iwahara, M. and Clubb, R.T. (2002) Xis protein binding to the left arm stimulates excision of conjugative transposon Tn916. *J Bacteriol*, **184**, 2088–2099.
78. Ringwald, K. and Gardner, J. (2015) The *Bacteroides thetaiotaomicron* protein *Bacteroides* Host Factor A participates in integration of the integrative conjugative element CTnDOT into the chromosome. *J Bacteriol*, **197**, 1339–1349.
79. Pe a, C.E., Lee, M.H., Pedulla, M.L. and Hatfull, G.F. (1997) Characterization of the mycobacteriophage L5 attachment site, *attP*. *J Mol Biol*, **266**, 76–92.
80. Wood, M.M., DiChiara, J.M., Yoneji, S. and Gardner, J.F. (2010) CTnDOT integrase interactions with attachment site DNA and control of directionality of the recombination reaction. *J Bacteriol*, **192**, 3934–3943.
81. Sylwan, L., Frumerie, C. and Hagg rd-Ljungquist, E. (2010) Identification of bases required for P2 integrase core binding and recombination. *Virology*, **404**, 240–245.
82. Umlauf, S.W. and Cox, M.M. (1988) The functional significance of DNA sequence structure in a site-specific genetic recombination reaction. *EMBO J*, **7**, 1845–1852.
83. Nunes-D by, 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.
84. Rutkai, E., Dorgai, L., Sirot, R., Yagil, E. and Weisberg, R.A. (2003) Analysis of insertion into secondary attachment sites by phage lambda and by int mutants with altered recombination specificity. *J Mol Biol*, **329**, 983–996.
85. Elias, A., Kassis, H., Elkader, S.A., Gritsenko, N., Nahmad, A., Shir, H., Younis, L., Shannan, A., Aihara, H., Prag, G., *et al.* (2020) HK022 bacteriophage Integrase mediated RMCE as a potential tool for human gene therapy. *Nucleic Acids Res*, **48**, 12804–12816.
86. Williams, K.P. (2002) Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: sublocation preference of integrase subfamilies. *Nucleic Acids Res*, **30**, 866–875.

TABLE AND FIGURES LEGENDS

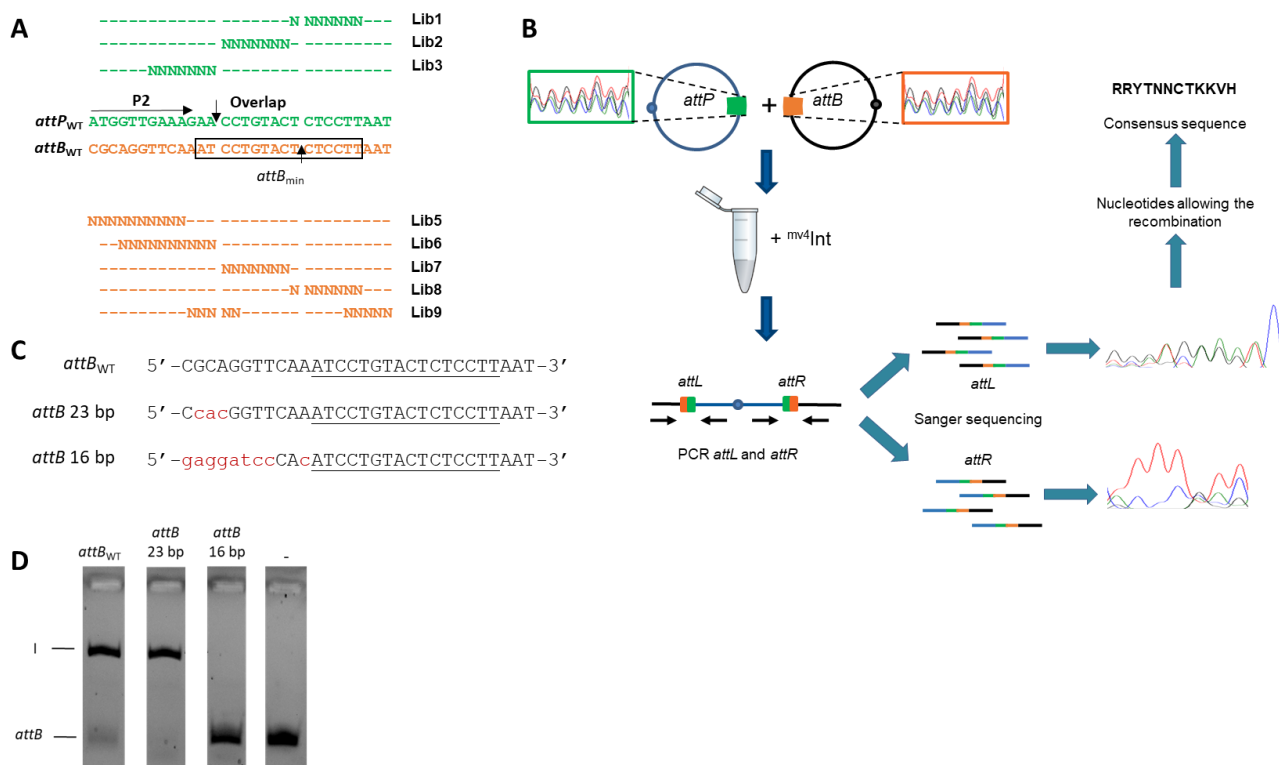


Figure 1. Principle of the use of randomized libraries for the characterization of core regions and *in vitro* characterization of the minimal *attB* size. **(A)** Localization of libraries in the core regions. The minimal *attB* site (1) is in the black box. The atypical P2 arm-type binding site is indicated by a black arrow and the dark vertical arrows indicate the strand exchange loci on top and bottom strands of the 8bp overlap sequence (2). The randomized region of each library is represented by “N” and nucleotides identical to the native sequences are symbolized by dashes. **(B)** Global strategy for the use of randomized libraries. The chromatograms indicate the different nucleotides present at each position. A color is associated with each nucleotide : blue, C; black, G; red, T; green, A. **(C)** Sequences used in this study for the *in vitro* recombination assay. **(D)** Effect of the size of *attB* on the recombination reaction. The recombination reaction contains 7.2 pmol of *mv4*Int, 40 μ g of *E. coli* crude extract heated at 95°C, and the reaction was incubated 1h30 at 42°C. The size of the *attB* used is indicated above each lane. The fluorescent *attB* fragment and the recombination product (I) are indicated on the gel. Lane -, reaction without *mv4*Int.

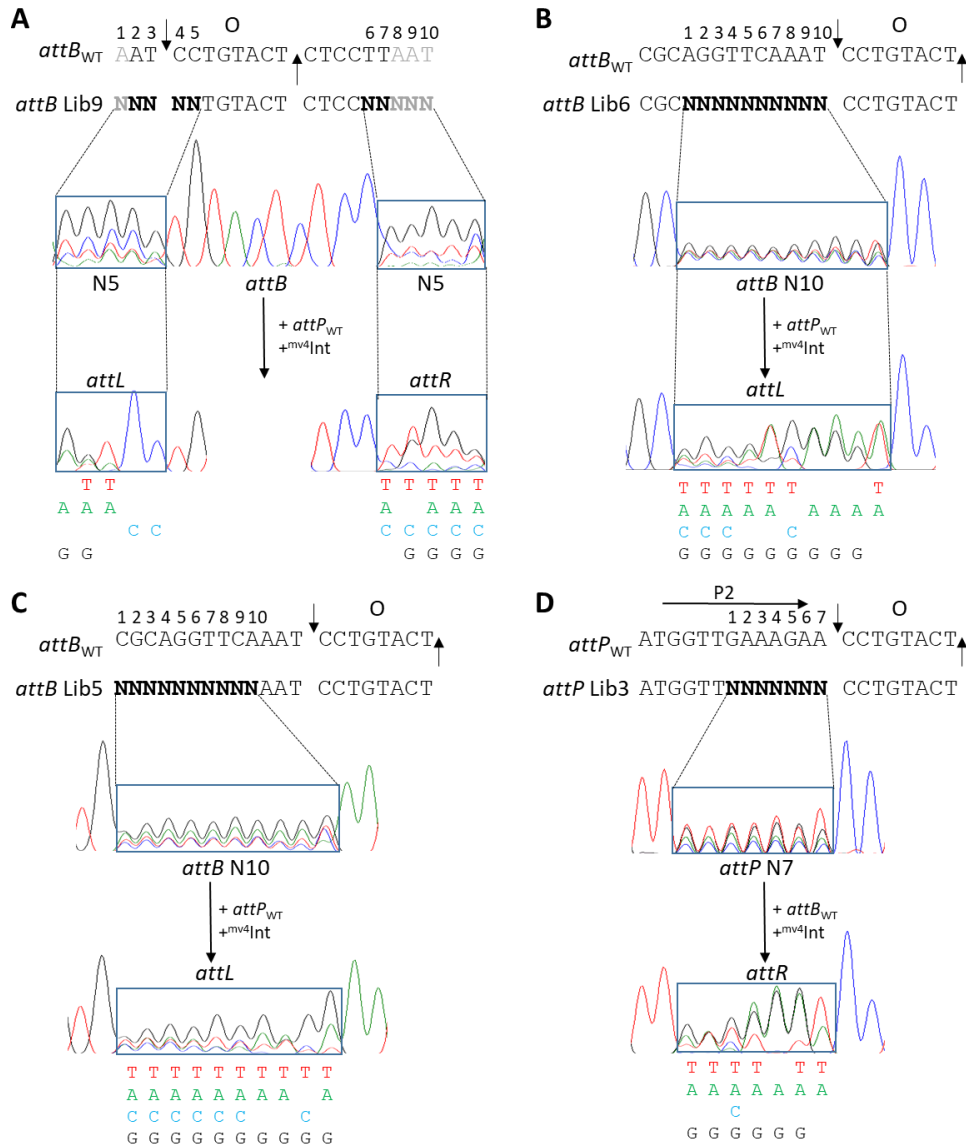


Figure 2. Characterization of the minimal *attB* size and of the constraints on the left side of *attB* and core-*attP*. **(A)** Results of *attB* Lib9 x *attP*_{WT} *in vitro* recombination. The bases defined outside *attB*_{min} (1) are represented in gray. The dark vertical arrows indicate the two cleavage site surrounding the overlap region of 8 bp (2). "O" indicates the overlap region and "N" represents the randomized positions. Each nucleotide is associated with a color on the chromatogram: blue, C; black, G; red, T; green, A. **(B)** Results of *attB* Lib6 x *attP*_{WT} *in vitro* recombination. **(C)** Results of *attB* Lib5 x *attP*_{WT} *in vitro* recombination. **(D)** Results of *attB*_{WT} x *attP* Lib3 *in vitro* recombination. The atypical P2 arm-type binding site is indicated by a black arrow.

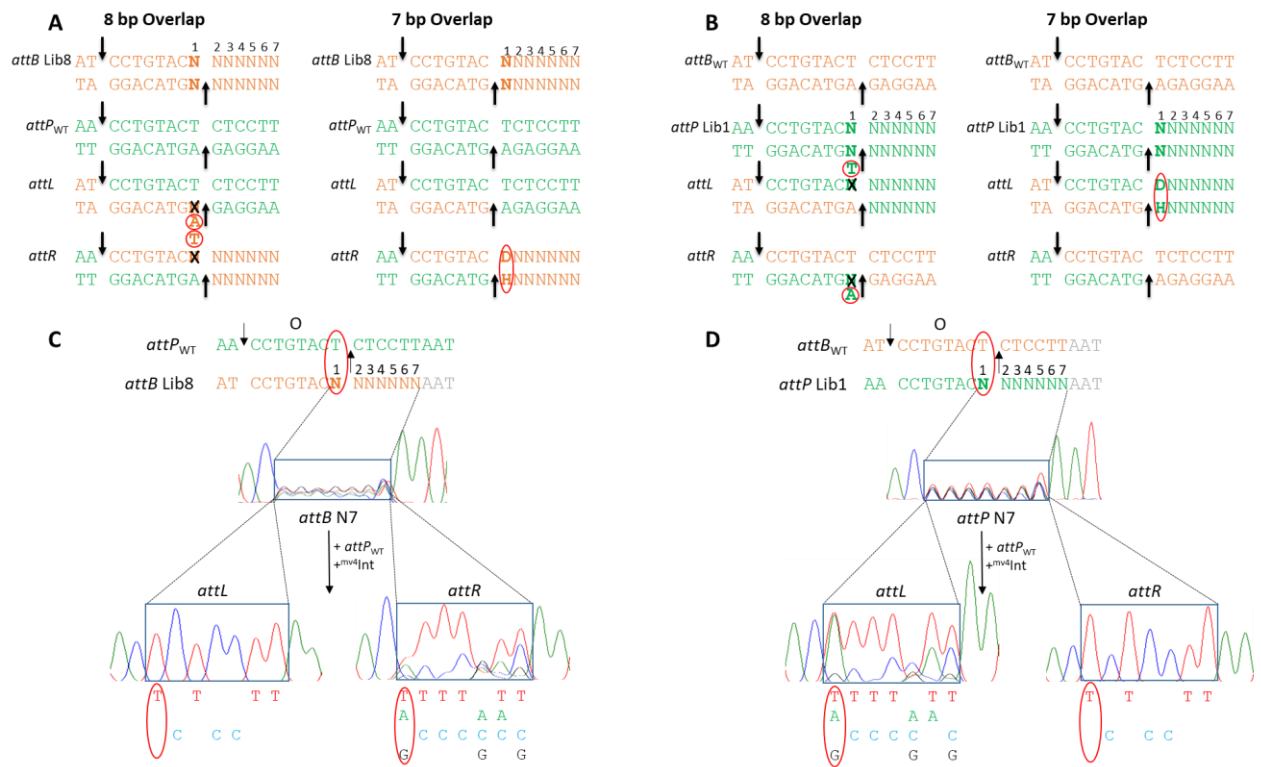


Figure 3. Characterization of the *attB* and *attP* overlap size and of the constraints on the right side of *attB* and core-*attP*. (A) Representation of the two expected results of *attB* Lib8 x *attP* WT recombination depending on the overlap size. The first randomized position is shown in bold and is indicated in red in *attL* and *attR* site. The vertical black arrows indicate the two cleavage sites allowing the strand exchange reaction (2). (B) Representation of the two expected results of *attB* WT x *attP* Lib1 recombination depending on the overlap size. The legend is identical to part (A). In both cases (A and B), if the first position is included in the strand exchange region (8 bp overlap), only the nucleotide complementary to the *attP* WT (A) or *attB* WT (B) site will be observed (i.e a T on the top strand and an A on the bottom strand). If the first position is excluded from the strand exchange region (7 bp overlap), the permissive nucleotides will be observed only in one of the two hybrid sites. (C) Results of *attB* Lib8 x *attP* WT *in vitro* recombination. The dark vertical arrows indicate the two *mv4*Int cleavage sites surrounding the overlap region "O". (D) Results of *attB* WT x *attP* Lib1 *in vitro* recombination.

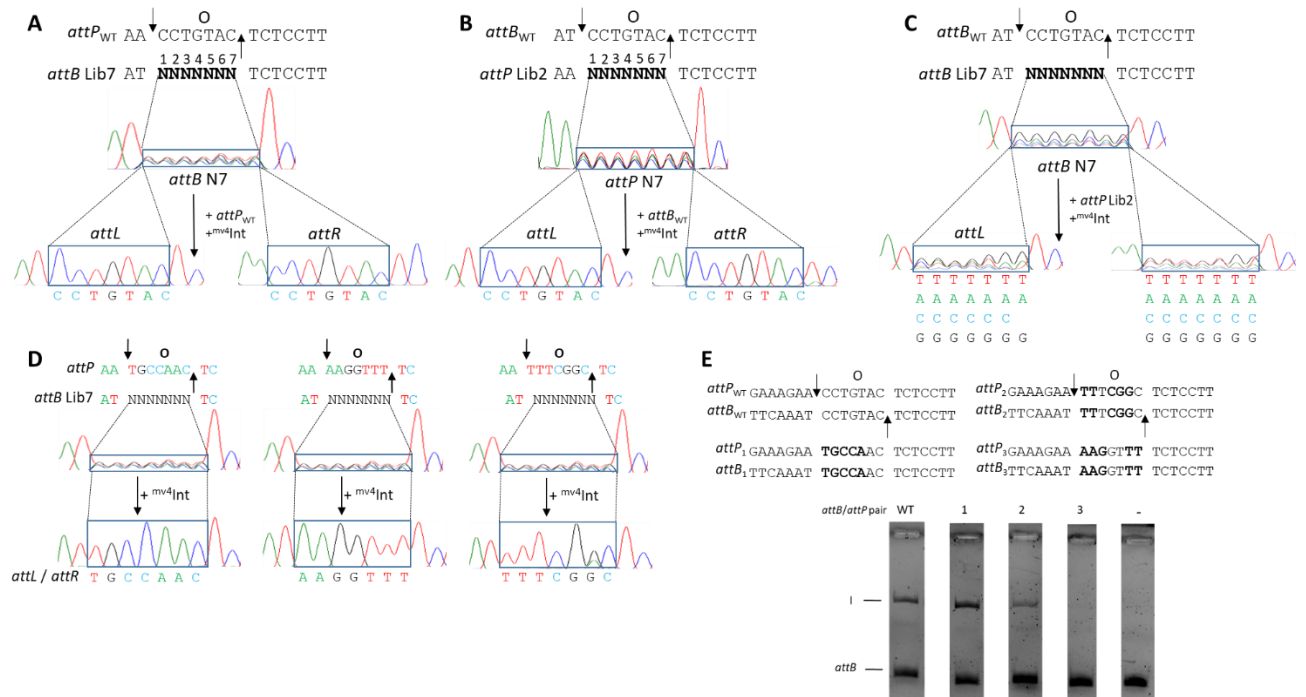


Figure 4. Characterization of the constraints exerted in the overlap region. (A) Results of *attB* Lib7 x *attP*_{WT} *in vitro* recombination. The dark vertical arrows indicate the two *mv*⁴Int cleavage sites surrounding the overlap region “O”. (B) Results of *attB*_{WT} x *attP* Lib2 *in vitro* recombination. (C) Results of *attB* Lib7 x *attP* Lib2 *in vitro* recombination. (D) *In vitro* recombination results of *attB* Lib7 against 3 different *attP* overlap sequences. (E) Effect of the overlap sequence on recombination activity. The 4 different pairs of *attB*/*attP* sharing the same overlap sequence are represented. Lanes WT, *attB*_{WT}/*attP*_{WT} pair; 1, *attB*₁/*attP*₁ pair; 2, *attB*₂/*attP*₂ pair; 3, *attB*₃/*attP*₃ pair, (-), no *mv*⁴Int.

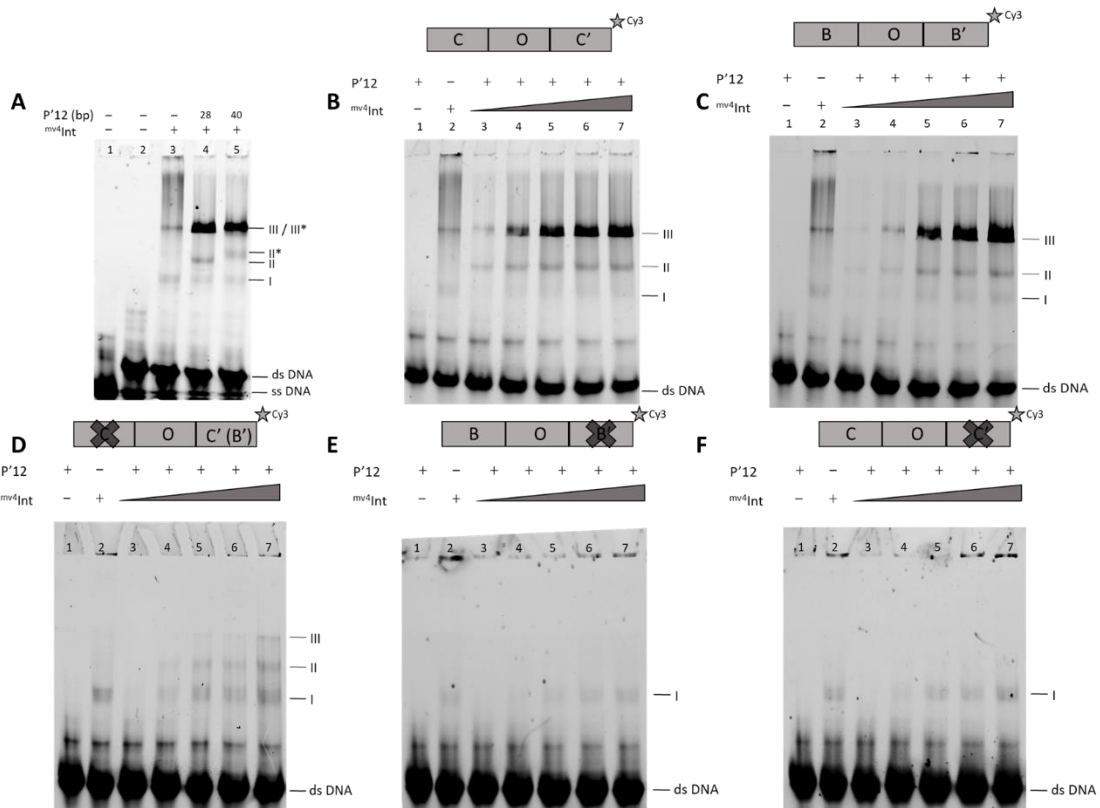


Figure 5. Gel-shift assay. (A) The effect of arm-type oligonucleotide on $mv4Int$ binding to COC'. A fluorescent fragment containing the 21 bp COC' sequence (0.87 pmol) was incubated in the presence or absence of $mv4Int$ (25 pmol) and in the presence or absence of unlabeled DNA containing P'12 arm binding sites (4.48 pmol). Reactions were analyzed by native 7.5% PAGE and fluorescence was visualized on the Chemidoc MP Imaging system (Biorad). The presence or absence of $mv4Int$ and arm-type sites (28 bp or 40 bp) is indicated above the gel. The different complexes are indicated on the gel: ss DNA, single-stranded DNA; ds DNA, double-stranded DNA; I, one monomer of $mv4Int$ bound to the core region; II, one monomer of $mv4Int$ bound to the core and arm region of 28 bp (II) or 40 bp (II*); III, dimer of $mv4Int$ bound to the core and arm region of 28 bp (III) and 40 bp (III*). (B) $mv4Int$ binding to the COC' region of *attP*. Legend is identical to (A). The reaction contains 0 pmol (lane 1), 3.6 pmol (lane 3), 10.7 pmol (lane 4), 18 pmol (lane 5), 25.1 pmol (lanes 2 and 6) and 35.8 pmol (lane 7) of $mv4Int$. (C) $mv4Int$ binding to the BOB' sequence. (D) $mv4Int$ binding to the B'/C' core-binding site. (E) $mv4Int$ binding to the B core-binding site. (F) $mv4Int$ binding to the C core-binding site. Legends are identical to (A) and (B).

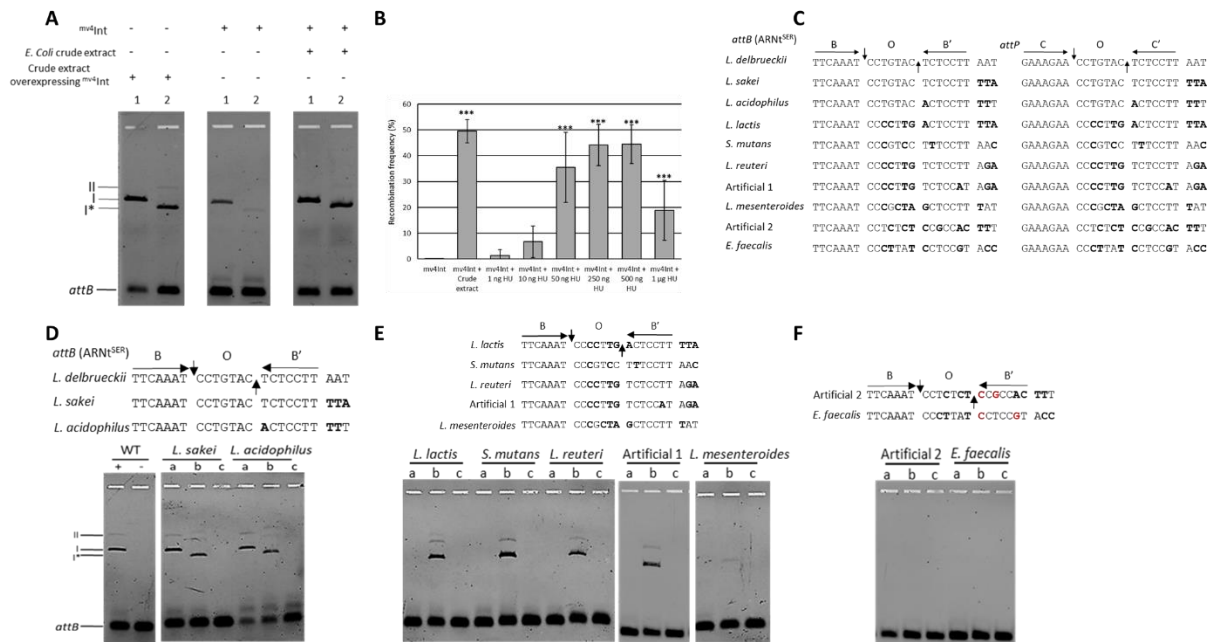


Figure 6. Characterization of an accessory protein and of the extent of *mv4Int* flexibility. (A) Effect of an *E. coli* crude extract on the recombination activity between native sites and modified sites. Lanes : 1, *attP*_{WT} x *attB*_{WT} recombination; 2, *attP*_{lactis} x *attB*_{lactis} recombination. Linear fluorescent *attB* fragment and recombination products between wild-type sites (I) or modified sites (I*) are indicated. The second recombination product (II) is most likely a recombination product from an *attP* x *attL* recombination (3). (B) Effect of the amount of 6xHis-HU_{la} on *attP*_{lactis} x *attB*_{lactis} recombination. Fluorescence intensity was quantified with Image Lab software (Biorad). The data presented are the mean values and standard deviations of four independent measurements. Significant differences from the *mv4Int* condition are represented by the symbol “***” (Kruskal-Wallis p<0.001). (C) tRNA^{SER} sequences and the adapted *attP* core sequence used for the *in vitro* assay. The two *mv4Int* cleavage sites surrounding the overlap region (O) are indicated by the vertical black arrows. Nucleotides that differ from the *attB*_{WT} sequence are shown in bold. (D) Sequences where recombination is possible with *attP*_{WT} and with the adapted *attP*. The recombination reaction contains 7.2 pmol of *mv4Int* and was incubated for 16h at 42°C. Lane: a, *attP*_{WT} x *attB*_x recombination; b, *attP*_x x *attB*_x recombination; c, b without *mv4Int*; “+” *attP*_{WT} x *attB*_{WT} recombination; “-” *attP*_{WT} x *attB*_{WT} recombination without *mv4Int*. The bacteria from which *attB*_x (tRNA^{SER}) comes from is indicated above the gels. (E) Sequences where recombination is only possible with the adapted *attP*. (F) Sequences where recombination is not possible.

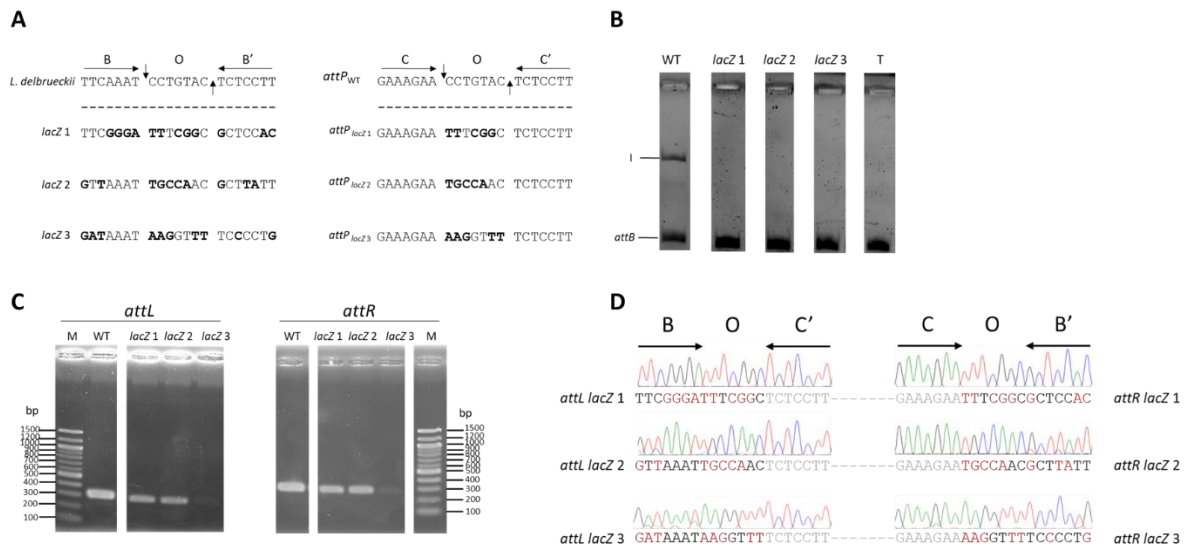


Figure 7. *in vitro* reprogramming of the recombination to 3 putative sites present in *E. coli lacZ* gene. **(A)** Artificial *attB* sites and adapted *attP* used. The nucleotides different from the native *attB_{WT}* are indicated in bold. **(B)** Fluorescent *in vitro* assay against the 3 *lacZ* sites. Recombination reaction contains 7.2 pmol of *mv⁴*Int, 40µg heated *E. coli* crude extract and was incubated 1 h 30 at 42°C. The *attB/attP* pair used is indicated above each lane. Fluorescent linear *attB* and recombination product (I) are indicated on the gel. Lane T, reaction without *mv⁴*Int. **(C)** PCR amplification of *attL* and *attR* for each *attB/attP* pair tested. Lane M, 100 bp DNA ladder (New England Biolabs). **(D)** Sanger sequencing of *attL* and *attR* recombination products. Nucleotides in gray are the nucleotides belonging to *attP_{WT}* and nucleotides in red are the bases that differ from *attB_{WT}*.

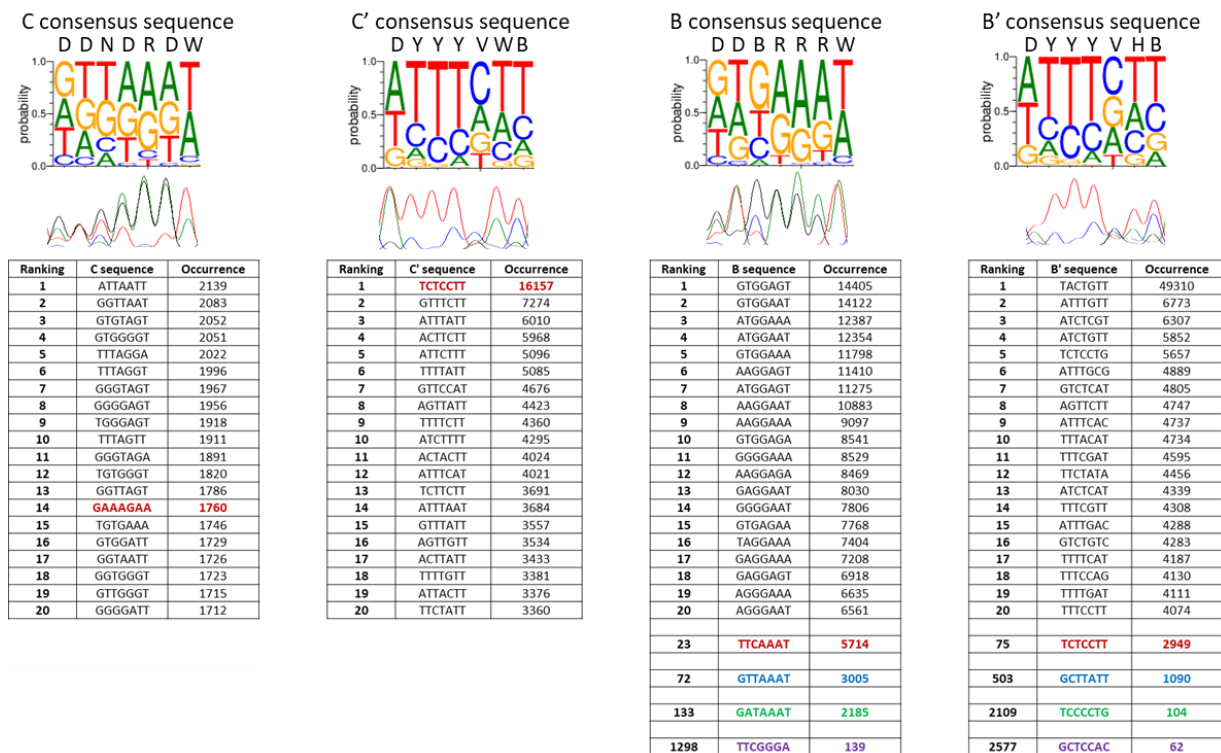


Figure 8. "Sequence Logo" and characterization of the most functional sequences for each core-binding sites. Native sequences are indicated in red. B and B' sequences of *lacZ* 1, 2 and 3 are indicated in purple, blue and green, respectively.

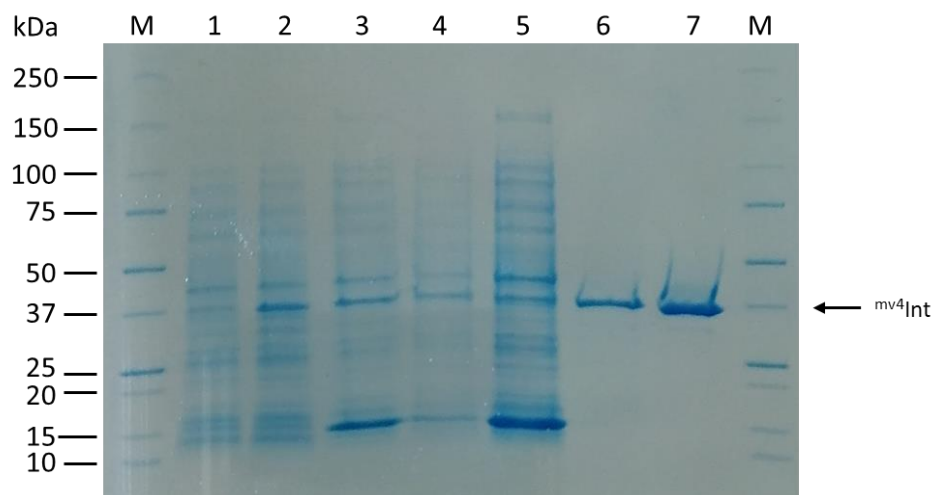


Figure S1. Purification of ^{mv4}Int. Lanes : 1, uninduced *E. coli* BL21(DE3) crude extract; 2, induced crude-extract; 3, soluble fraction; 4, insoluble fraction; 5, Ni-NTA column flow-through; 6, gel filtration eluted fractions; 7, Heparin column eluted fractions; M, protein ladder.

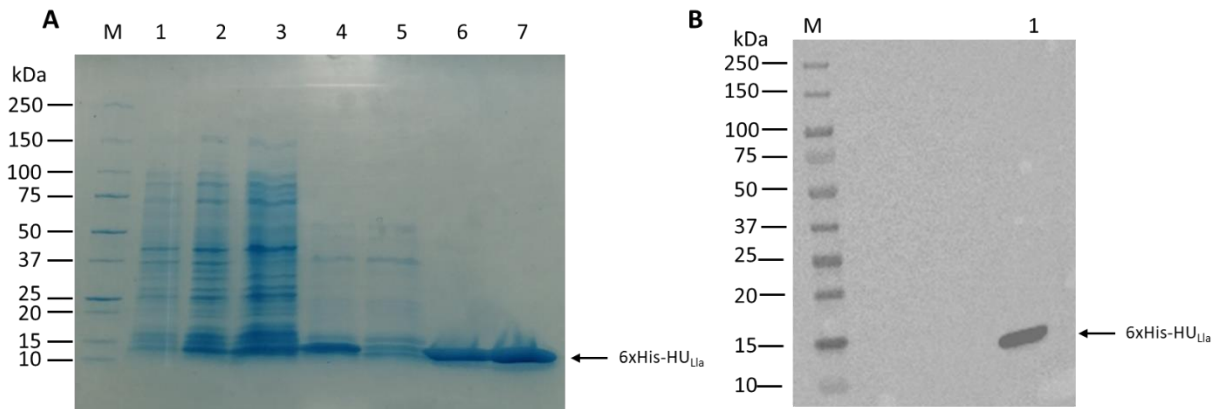


Figure S2. Purification of 6xHis-HU_{Lia}. **(A)** 4-15% SDS-PAGE of protein extracts obtained during the purification of 6xHis-HU_{Lia}. Lanes : 1, uninduced *E. coli* BL21(DE3) crude extract; 2, induced crude-extract; 3, supernatant before heat treatment (70°C, 30 min); 4, supernatant after heat treatment; 5, Ni-NTA column flow-through; 6, Ni-NTA column eluted fractions; 7, Heparin column eluted fractions; M, protein ladder. **(B)** Western-blot of 6xHis-HU_{Lia}. 200 ng of 6xHis-HU_{Lia} were separated by SDS-PAGE, transferred to a nitrocellulose membrane and detected by chemiluminescence after incubation with anti-polyHistidine-Peroxidase antibodies (1: 2000). Lanes : 1, 200 ng 6xHis-HU_{Lia}; M, protein ladder.

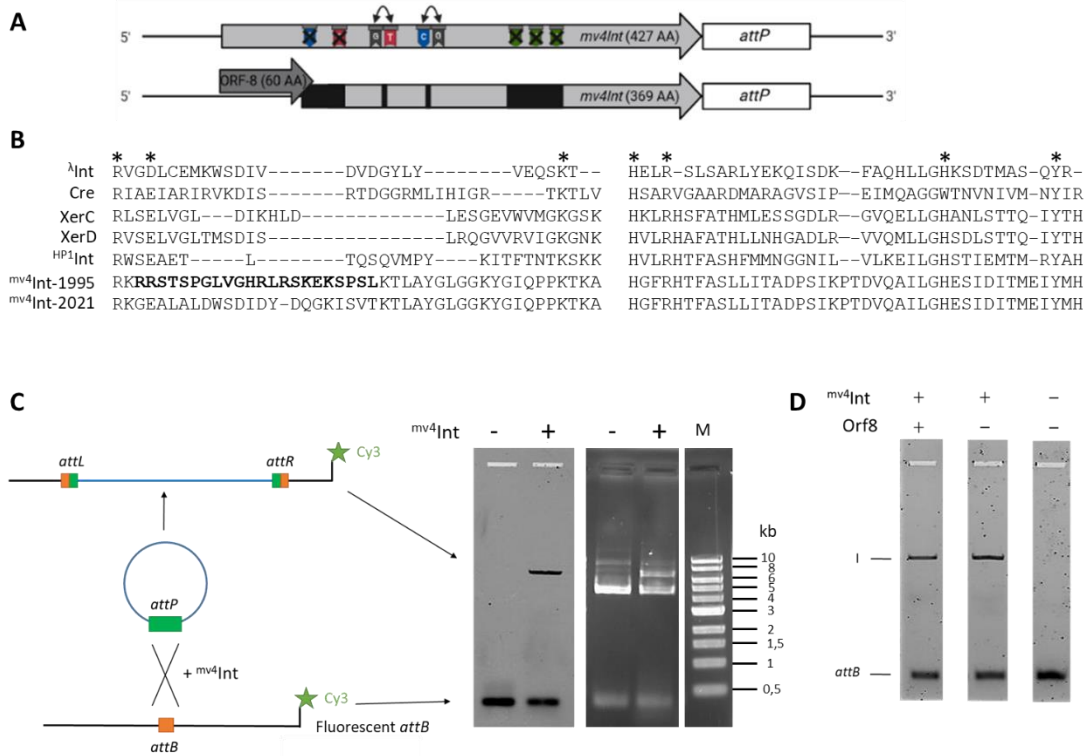


Figure S3. Redefining mv^4 Int sequence and development of an *in vitro* assay. **(A)** Schematic representation of mv^4 Int and consequences of sequencing errors on mv^4 Int sequence. Deletions (black cross) of a C, a T and three A, the inversion of GT/TG and CG/GC (two-way arrow) are found. Black boxes indicates protein regions different from the published sequence. **(B)** Alignment of catalytic domains of λ Int, Cre, XerC, XerD, HP1Int, mv^4 Int-1995 and mv^4 Int-2021. The 7 conserved residues of the tyrosine recombinases catalytic domain are highlighted with an asterisk. The lysine residue (K) was manually adjusted based on the alignment performed by Nunes-Düby et al. (4). **(C)** Principle of the *in vitro* assay. Lane M, 1 kb DNA ladder (New England Biolabs). Left gel, Cy3-labelled DNA ; Right gel, DNA fluorescence after BET staining and UV revelation. **(D)** Effect of ORF8 peptide on recombination. *E. coli* BL21(DE3) crude extract contains either mv^4 Int and the 6xHis-ORF8 peptide or only mv^4 Int. *attB* fluorescent linear fragment and recombination product (I) are indicated on the gel. The reaction contains 40 μ g of each crude extract and was incubated for 1h30 at 42°C.

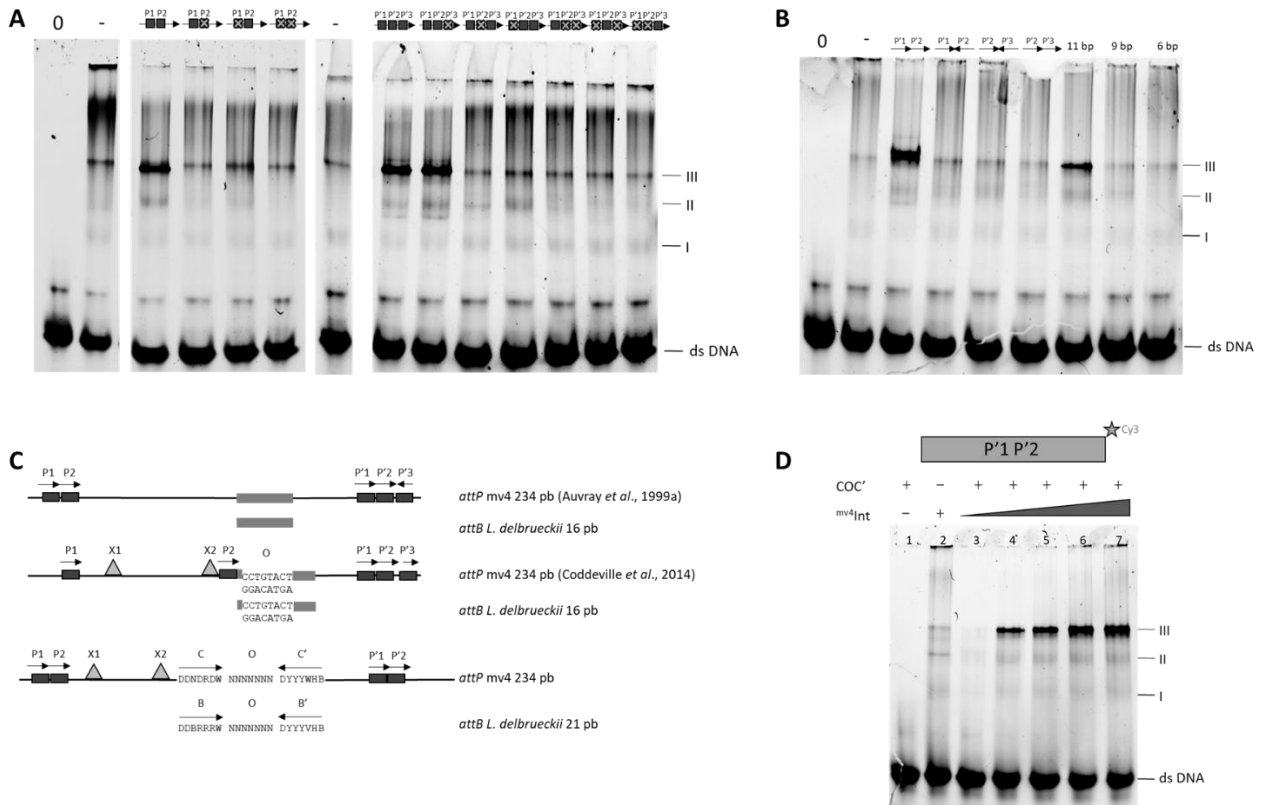


Figure S4. Characterization of the $mv4Int$ arm-binding sites and of the affinity of the N-ter domain to the arm-binding sites. (A) Characterization of the $mv4Int$ arm-binding site by stabilization of $mv4Int$ binding to the COC' core sequence. A fluorescent fragment including the COC' core sequence (0.87 pmol) was incubated in the presence of 4.48 pmol of unlabeled arm-binding sites (the arm-binding region is indicated above the lanes) and 25.1 pmol of $mv4Int$. Lane : 0, no $mv4Int$; -, no arm-binding oligonucleotide. The three complexes are indicated and are identical to those determined previously. **(B)** Effect of the orientation and of the size of the arm-binding sites on the stabilization of $mv4Int$ binding to the COC' core region. The legends is identical to part (A). **(C)** Schematic representation of the *attP* and *attB* sites of the mv4 recombination system. Black boxes indicates arm-binding sites. The orientation of the site is represented by an arrow. Light grey triangles represent the excisionase binding sites (3) and gray boxes represent the core region (region of identity between *attB* and *attP*). **(D)** The effect of core-type oligonucleotide on $mv4Int$ binding to P'1-P'2 arm-binding sites. A fluorescent fragment including the P'12 arm-binding sequence (0.87 pmol) was incubated in the presence of 4.48 pmol of unlabeled core-binding sites (COC'). Reaction contain 0 pmol (1), 3.6 pmol (3), 10.7 pmol (4), 18 pmol (5), 25.1 pmol (2 and 6) and 35.8 pmol of $mv4Int$.

Table S1. Strains used in this study.

Strains	Relevant characteristics	Reference
<i>E. coli</i> BL21(DE3)	B F ⁻ ompT gal dcm lon hsdSB(rB-mB-) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)	New England Biolabs
<i>E. coli</i> EC1000	<i>E. coli</i> EC101, <i>repA</i> ⁺	Leenhouts et al., 1996
<i>E. coli</i> Neb5α	F ⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK-mK+), λ ⁻	New England Biolabs
<i>E. coli</i> EPI300	F ⁻ λ ⁻ mcrA Δ(mrr-hsdRMS-mcrBC) Φ80dlacZΔM15 Δ(lac)X74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK rpsL (StrR) nupG' trfA dhfr	Epicentre
<i>E. coli</i> Neb5α- <i>RepA</i>	<i>E. coli</i> Neb5α:: <i>repA</i>	This study

Table S2. Plasmids used in this study.

Plasmid name	Plasmid description	Plasmid properties and construction
pGhost9		<i>ori</i> pWV01; [Ery ^R] (ref maguin)
pBSattB	pBS:: <i>attB</i> _{mv4}	<i>ori</i> ColE1;[Amp ^R] (Ref dupont)
pMC1	pRC1:: ^{mv4} <i>Int</i> , <i>attP</i> _{mv4}	<i>ori</i> ColE1;[Ery ^R] (Ref dupont)
pET-Int	pET15b:: ^{mv4} <i>Int</i>	<i>ori</i> pBR322; [Amp ^R] (RefAuvray1999)
pCC1Fos		<i>ori</i> F-factor and <i>ori</i> V; [Cm ^R] (Epicentre)
pFSKB3X		<i>ori</i> ColE1;[Kan ^R] (GTP technology)
pMET302	pMC1::core <i>attP</i> tRNA ^{SER} <i>L. acidophilus</i>	
pMET303	pMC1::core <i>attP</i> artificial <i>attB</i> ₁	
pMET304	pMC1::core <i>attP</i> tRNA ^{SER} <i>L. mesenteroides</i>	
pMET305	pMC1::core <i>attP</i> tRNA ^{SER} <i>E. faecalis</i>	
pMET306	pMC1::core <i>attP</i> tRNA ^{SER} <i>L. lactis</i>	
pMET307	pMC1::core <i>attP</i> tRNA ^{SER} <i>S. mutans</i>	
pMET308	pMC1::core <i>attP</i> tRNA ^{SER} <i>L. reuteri</i>	
pMET309	pMC1::core <i>attP</i> artificial sequence <i>L. reuteri</i> 2	
pMET310	pMC1::core <i>attP</i> artificial sequence <i>L. sakei</i>	
pMET311	pMC1 IS10:: ^{mv4} <i>Int</i>	
pMET320	pBSattB::tRNA ^{SER} <i>L. acidophilus</i>	
pMET321	pBSattB::artificial <i>attB</i> ₁	
pMET322	pBSattB::tRNA ^{SER} <i>E. faecalis</i>	

pMET323	pBSattB::tRNA ^{SER} <i>L. lactis</i>	
pMET324	pBSattB::tRNA ^{SER} <i>S. mutans</i>	
pMET325	pBSattB::tRNA ^{SER} <i>L. mesenteroides</i>	
pMET326	pBSattB::tRNA ^{SER} <i>L. reuteri</i>	
pMET327	pBSattB::tRNA ^{SER} <i>L. sakei</i>	
pMET328	pBSattB::artificial sequence <i>L. reuteri</i> 2	
pMET340	pMC1 ::ori pWV01	ori pWV01;[Ery ^R]
pMET348	pCC1Fos::attB _{mv4}	
pMET349	pCC1Fos::attB _{mv4} 23 bp	
pMET357	pCC1Fos::attB _{mv4} 16 bp	
pMET359	pMET340::bla	ori pWV01;[Ery ^R]; [Amp ^R]
pMET364	pMET348::overlap lacZ ₁	
pMET365	pMET348::overlap lacZ ₃	
pMET366	pMET348::attB lacZ ₁	
pMET367	pMET348::attB lacZ ₂	
pMET368	pMET348::attB lacZ ₃	
pMET369	pFSKB3X::6xHis-HU _{Lla}	
pMET370	pMET348::overlap lacZ ₂	
pMET371	pMET359:: overlap lacZ ₁	
pMET372	pMET359:: overlap lacZ ₂	
pMET373	pMET359:: overlap lacZ ₃	
pMET374	pMET359:: core attP lacZ ₁	
pMET375	pMET359:: core attP lacZ ₂	

Table S3. Oligonucleotides used in this study. Sequences are given in 5' → 3' direction.

A. DNA oligonucleotides used for the cloning

Primer name	Sequence
SDM-attB4-F	GACCTGACTCTCCTTAATAAGGTCAAATG
SDM-attB4-R	GACCTTATTAAGGAGAGTACAGGTCTTGAAC
SDM-attB5-F	CCTGACTCTCCTGAATAAGGTCAAATGGTATC
SDM-attB5-R	GACCTTATTAAGGAGAGTACAGGTCTTGAAC
SDM-LbsakattB-F	CCTGACTCTCCTTTTAAAGGTCAAATGGTATC
SDM-LbsakattB-R	CATTTGACCTTTAAAAGGAGAGTACAGGATTTG
SDM-LbacidoattB-F	CCTGTACACTCCTTTTAAAGGTCAAATGGTATC
SDM-LbacidoattB-R	CATTTGACCTTAAAAGGAGTGTACAGGATTTG
SDM-LbreutattB-F	CCCCTGTCTCCTTAGAAAGGTCAAATGGTATC
SDM-LbreutattB-R	CTTTCTAAGGAGACAAGGGGATTTGAACCTGCG
SDM-LlactisattB-F	CCCCTTGACTCCTTTTAAAGGTCAAATGGTATCC
SDM-LlactisattB-R	CTTAAAAGGAGTCAAGGGGATTTGAACCTGCG

SDM-SmutansattB-F	CCCGTCCTTTCCTTAACAAGGTCAAATGGTATC
SDM-SmutansattB-R	CTTGTTAAGGAAAGGACGGGATTTGAACCTGCG
SDM-EfaecattB-F	CCCTTATCCTCCGTACCAAGGTCAAATGGTATCC
SDM-EfaecattB-R	CTTGGTACGGAGGATAAGGGATTTGAACCTGCG
SDM-LnmesentattB-F	CCCGCTAGCTCCTTTATAAGGTCAAATGGTATC
SDM-LnmesentattB-R	CTTATAAAGGAGCTAGCGGGATTTGAACCTGCG
SDM-EcoliattB-F	CCTCTCTCCGCCACTTTAAGGTCAAATGGTATCC
SDM-EcoliattB-R	CTTAAAGTGGCGGAGAGAGGATTTGAACCTGCG
SDMgibsonermAM-F	GAGAATATCGTCAACTGTTTACTAAAAATC
SDMgibsonermAM-R	GTAAACAGTTGACGATATTCTCGATTG
SDMgibsonattP5-F	AAAGAACCTGTACTCTCCTGAATCAAAGCAATAATC
SDMgibsonattP5-R	TTCAGGAGAGTACAGGTTCTTTCAACCATGTTTC
SDMgibsonattPsakei-F	AAAGAACCTGTACTCTCCTTTTACAAAGCAATAATC
SDMgibsonattPsakei-R	AAAAGGAGAGTACAGGTTCTTTCAACCATGTTTC
SDMgibsonattPacido-F	AAAGAACCTGTACTCTCCTTTTCAAAGCAATAATC
SDMgibsonattPacido-R	AAAAGGAGTGTACAGGTTCTTTCAACCATGTTTC
SDMgibsonattPmutans-F	AAAGAACCCGTCCTTTCCTTAACCAAAGCAATAATCCC
SDMgibsonattPmutans-R	TTAAGGAAAGGACGGGTTCTTTCAACCATGTTTC
SDMgibsonattPreuteri-F	AAAGAACCCCTTGCTCCTTAGACAAAGCAATAATCCC
SDMgibsonattPreuteri-R	CTAAGGAGACAAGGGGTTCTTTCAACCATGTTTC
SDMgibsonattPlactis-F	AAAGAACCCCTTGACTCCTTTTACAAAGCAATAATCCC
SDMgibsonattPlactis-R	AAAAGGAGTCAAGGGGTTCTTTCAACCATGTTTC
SDMgibsonattPfaeca-F	AAAGAACCCCTTATCCTCCGTACCCAAAGCAATAATCCC
SDMgibsonattPfaeca-R	GTACGGAGGATAAGGGTCTTTCAACCATGTTTC
SDMgibsonattPcoli-F	AAAGAACCTCTCTCCGCCACTTTCAAAGCAATAATCCC
SDMgibsonattPcoli-R	AAGTGGCGGAGAGAGGTTCTTTCAACCATGTTTC
SDMgibsonattPmesent-F	AAAGAACCCGCTAGCTCCTTTATCAAAGCAATAATCCC
SDMgibsonattPmesent-R	TAAAGGAGCTAGCGGGTCTTTCAACCATGTTTC
pBS-antiattB-F	GCTTGGGCTGCAGGA
pBS-antiattB-R	GGAAAGGACATCTAAATCAAATGG
SDM-Lbreuteri2attB-F	TCCATAGAAAGGTCAAATGGTATC
SDM-Lbreuteri2attB-R	GACAAGGGGATTTGAACCTG
oripGh9-F	TTAAATTTATACTGCAATCGGATGC
oripGh9-R	GTATTTTTAATAGCCATGATATAATTACCTTATC
pMC1-sans-colE1-F	TGGAACGAAAACCTCACGTTAAG
pMC1-sans-colE1-R	TGATTCTGTGGATAACCGTATTAC
pCC1FOS-F	GTGGGATCCCCGGGTAC
pCC1FOS-R	GTGGGATCCTCTAGAGTGCAC
attB-gibsonpCC1Fos-F	GCAGGTCGACTCTAGAGGATCCACGAATTCCTGCAGCCCAAGC
attB-gibsonpCC1Fos-R	GAGCTCGGTACCCGGGGATCCACCCCATTTGATTTAGATGTCCTTTC
attBWT-F	ATCCTGTACTCTCCTTAAT
attBWT-R	ATTAAGGAGAGTACAGGAT
attB-WTGGTT-F	GGTTCAAATCCTGTACTCTCCTTAAT
attB-WTGGTT-R	ATTAAGGAGAGTACAGGATTTGAACC
pMC1delTOPO-F	AAATCGAAACAGCAAAGAATGG

Bla-F	TGGCACTTTTCGGGGAAATG
Bla-R	GTGCTACAGAGTTCTTGAAGTG
pET-F	TAGCTCGAGGATCCGGCTGCTAAC
pETH-HU-20pbHis-R	GATGATGGCTGCTGCCATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTC
Hup6xHisgib-F	ATCATCATCACAGCAGCGGCATGGCTAACAAACAAGATCTTATC
HU-TAA-R	TTATTTAACTGCGTCTTTCAATGCTTTAC
HupGib-F	CTTTAAGAAGGAGATATACCATGGCTAACAAACAAGATCTTATC
HupGibpFSKB3X-R	TGTCCACCAGTCATGCTAGCTTATTTAACTGCGTCTTTCAATGCTTTAC
pFSKB3X-F	GCTAGCATGACTGGTGGAC
pETRBS-F	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATAACC
Hu-Cter-R	GAATGCAGGAACAACCTGTAGC
HU-Cter-F	GCTACAGTTGTTCTGCATTC
Site1LacZ-Gib-F	AAAGAATTTTCGGCTCTCCTTAATCAAAGCAATAATCCC
Site2LacZ-Gib-F	AAAGAATGCCAACTCTCCTTAATCAAAGCAATAATCCC
Site3LacZ-Gib-F	AAAGAAAAGGTTTTCTCCTTAATCAAAGCAATAATCCC
Site1LacZ-Gib-R	TTAAGGAGAGCCGAAATTCCTTCAACCATGTTTCTGGAG
Site2LacZ-Gib-R	TTAAGGAGAGTTGGCATTCTTCAACCATGTTTCTGGAG
Site3LacZ-Gib-R	TTAAGGAGAAAACCTTTTCTTCAACCATGTTTCTGGAG
AmpGib-F	GAATTATGCAGTGCTGCCATAAC
AmpGib-R	GTTATGGCAGCACTGCATAATC
OverlapSite1lacZ-R	AAGGAGAGGccgAaaATTTGAACCTGCGCACC
OverlapSite2lacZ-R	AAGGAGAGTtggcaATTTGAACCTGCGCACC
OverlapSite3lacZ-R	AAGGAGAaaACcttATTTGAACCTGCGCACC
OverlapSite1lacZ-F	TCAAATttTcggCTCTCCTTAATAAAGGTCAAATGGTATC
OverlapSite2LacZ-F	TCAAATtgccaACTCTCCTTAATAAAGGTCAAATGGTATC
OverlapSite3lacZ-F	TCAAATaagGttTCTCCTTAATAAAGGTCAAATGGTATC
attBsite1LacZ-F	GAAATCCCGAACCTGCGCACCAATTCAACATTG
attBsite1LacZ-R	GGCGCTCCACAATAAAGGTCAAATGGTATCCCTATAGG
attBsite2LacZ-F	GGCAATTTAACCTGCGCACCAATTCAACATTG
attBsite2LacZ-R	AACGCTTATTAATAAAGGTCAAATGGTATCCCTATAGG
attBsite3LacZ-F	CCTTATTTATCCCTGCGCACCAATTCAACATTG
attBsite3LacZ-R	TTTTCCCCTGAATAAAGGTCAAATGGTATCCCTATAGG
attPCOC'site1LacZ-F	GGCGCTCCACAATCAAAGCAATAATCCC
attPCOC'site1LacZ-R	GAAATCCCGAAAACCATGTTTCTGGAGGGTTC
attPCOC'site2LacZ-F	AACGCTTATTAATCAAAGCAATAATCCC
attPCOC'site2LacZ-R	GGCAATTTAACAACCATGTTTCTGGAGGGTTC

B. DNA oligonucleotides used for the libraries construction and sequencing

Prim er name	Sequence
Seqb anqu	TGCCTGCAGGTCGACTCTAG

eattB -F	
Seqb anqu eattL -R	ACGCTAATGCCATCTATTAAGTAGC
Seqb anqu eattR -F	GAAACAACCAGAAACGCTTTTTAG
Seqb anqu eattB -R	GGCGAATTCGAGCTCGGTAC
Seqb anqu eattP -R	GGTCGACGGTATCGATAAGC
Seqb anqu eattP -F	GCAGGCGGAATGTTGAAAGAG
attP- N10- N19- R	GTTCTGGAGGTTTCGAATCTTG
attP- N10- N19- Vf-F	GTGATAATCGCCTGCCCGTTTGAC
attPli brary -F	CAAGATTCGAAACCTCCAGAAC
attPli brary -R	GTCAAACGGGCAGGCGATTATCAC
attBli brary -F	GCAGGTCGACTCTAGAGGATCCCACAAGCTTCGAAATCCGCCGAACCAATG
attBli brary -R	GAGCTCGGTACCCGGGGATCCCACGAATTCACTACTGGCTACTTTGAAATACTTCC
attB- gibso npCC 1Fos- F	GCAGGTCGACTCTAGAGGATCCCACGAATTCCTGCAGCCCAAGC
attB- gibso npCC	GAGCTCGGTACCCGGGGATCCCACCCCATTTGATTTAGATGTCCTTTC

1Fos-R	
attB Lib9	CGAAATCCGCCGAACCAATGTTGANNNNNNNCGCAGGTTCAANNNTGTACTCTCCNNNNNAA GGTCAAATNNNNNNNTATAGGAAGTATTTCAAAGTAGCCAGTAGT
attB Lib7	CGAAATCCGCCGAACCAATGTTGANNNNNNNCGCAGGTTCAAATNNNNNNNTCTCCTTAATAAG GTCAAATGNNNNNNNTATAGGAAGTATTTCAAAGTAGCCAGTAGT
attB Lib8	CGAAATCCGCCGAACCAATGTTGANNNNNNNCGCAGGTTCAAATCCTGTACNNNNNNNAATAA GGTCAAATGNNNNNNNTATAGGAAGTATTTCAAAGTAGCCAGTAGT
attP Lib2	CAAGATTCGAAACCTCCAGAACCCTCCAGAAACATGGTTGAAAGAANNNNNNNTCTCCTTAATCA AAGCAATAATCCCCGAGAAATCAACATTCTCGGGGATTATTTTTGTTTTAACTAGAAAATAACTAG AAAGAGCTAGTTAATAGATGGCATTAGCGTGATAATCGCCTGCCCGTTTGAC
attP Lib1	CAAGATTCGAAACCTCCAGAACCCTCCAGAAACATGGTTGAAAGAACCTGTACNNNNNNNAATC AAAGCAATAATCCCCGAGAAATCAACATTCTCGGGGATTATTTTTGTTTTAACTAGAAAATAACTA GAAAGAGCTAGTTAATAGATGGCATTAGCGTGATAATCGCCTGCCCGTTTGAC
attB Lib5	CGAAATCCGCCGAACCAATGTTGAATTGGTGNNNNNNNNNAATCCTGTACTCTCCTTAATAAG GTCAAATGGTATCCCTATAGGAAGTATTTCAAAGTAGCCAGTAGT
attP Lib4	CAAGATTCGAAACCTCCAGAACCCTCCAGAAACNNNNNNNNNGAACCTGTACTCTCCTTAATCA AAGCAATAATCCCCGAGAAATCAACATTCTCGGGGATTATTTTTGTTTTAACTAGAAAATAACTAG AAAGAGCTAGTTAATAGATGGCATTAGCGTGATAATCGCCTGCCCGTTTGAC
attB Lib6	CGAAATCCGCCGAACCAATGTTGAATTGGTGCGNNNNNNNNNCCTGTACTCTCCTTAATAAG GTCAAATGGTATCCCTATAGGAAGTATTTCAAAGTAGCCAGTAGT
attP Lib3	CAAGATTCGAAACCTCCAGAACCCTCCAGAAACATGGTTNNNNNNNNCCTGTACTCTCCTTAATCA AAGCAATAATCCCCGAGAAATCAACATTCTCGGGGATTATTTTTGTTTTAACTAGAAAATAACTAG AAAGAGCTAGTTAATAGATGGCATTAGCGTGATAATCGCCTGCCCGTTTGAC

C. DNA oligonucleotides used for EMSA

Primer name	Sequence
COC'WT-F	GGTATTGGAAAGAACCTGTACTCTCCTTGCGTAAC
COC'WT-Cy3-R	GTTACGCAAGGAGAGTACAGGTTCTTTCCAATACC
P'12WT-28bp-F	GTTTTTAAGTAACTAGAAAATAACTAGAAATTC
P'12WT-28bp-R	GAATTCTAGTTATTTTCTAGTTAAAAAC
P'12WT-40bp-F	CACGTCGTTTTTAAGTAACTAGAAAATAACTAGAAATTCACGTC
P'12WT-40bp-R	GACGTGGAATTCTAGTTATTTTCTAGTTAAAAACGACGTG
P'12WTdelP'3-F	TTTTTAAGTAACTAGAAAATAACTAGAAAGCGACCGAGCGGTGCG
P'12WTdelP'3-R	CGACCGCTCGGTGCGTTTCTAGTTATTTTCTAGTTAAAAA
P'123WT-F	TTTTTAAGTAACTAGAAAATAACTAGAAAGAGCTAGTTAATAGA
P'123WT-R	TCTATTAAGTAACTAGCTCTTTCTAGTTATTTTCTAGTTAAAAA
P'1WT-F	TTTTTAAGTAACTAGAACGACCGAGCGGAGGCGGCACGCTGTCG
P'1WT-R	CGACAGCGTGCCGCCTCCGCTCGGTCTAGTTAAAAA
P'2WT-F	GCCGACCGAGCGGAATAACTAGAAAGGCGGCACGCTGTCG
P'2WT-R	CGACAGCGTGCCGCCTTCTAGTTATTCCGCTCGGTGCGGC
P'3WT-F	GCCGACCGAGCGGGCGGCACGCTGAGAGCTAGTTAATAGA
P'3WT-R	TCTATTAAGTAACTAGCTCTCAGCGTGCCGCCGCTCGGTGCGGC
P'23WT-F	GCCGACCGAGCGGAATAACTAGAAAGAGCTAGTTAATAGA
P'23WT-R	TCTATTAAGTAACTAGCTCTTTCTAGTTATTCCGCTCGGTGCGGC
P'13WT-F	TTTTTAAGTAACTAGAACGACCGAGCGGAGAGCTAGTTAATAGA
P'13WT-R	TCTATTAAGTAACTAGCTCTCCGCTCGGTCTAGTTAAAAA
P12WT-F	CACGTCGTGATCAACTAGATTTTTAAGTAACTAGAAACCACGTC
P12WT-R	GACGTGGTTTTCTAGTTAAAAATCTAGTTGATCACGACGTG
P1WT-F	CACGTCGTGATCAACTAGATTCGACCGAGCGGACCACGTC
P1WT-R	GACGTGGTCCGCTCGGTGCAATCTAGTTGATCACGACGTG
P2WT-F	CACGTCGTGCGACCGAGCGGTTTTAAGTAACTAGAAACCACGTC
P2WT-R	GACGTGGTTTTCTAGTTAAAACCGCTCGGTGCGCACGACGTG
Nositebras-F	CACGTCGTGCGACCGAGCGGTGCGGCACGCTGACCACGTC
Nositebras-R	GACGTGGTCAGCGTGCCGCACCGCTCGGTGCGCACGACGTG
P'12WT-40bp-R-Cy3	GACGTGGAATTCTAGTTATTTTCTAGTTAAAAACGACGTG
COC'WT-R	GTTACGCAAGGAGAGTACAGGTTCTTTCCAATACC
BOB'WT-F	GGTATTGTTCAAATCCTGTACTCTCCTTGCGTAAC
BOB'WT-R	GTTACGCAAGGAGAGTACAGGATTTGAACAATACC
BOB'WT-R-Cy3	GTTACGCAAGGAGAGTACAGGATTTGAACAATACC
XOC'-F	GGTATTGGAACCCGCCTGTACTCTCCTTGCGTAAC
XOC'-R-Cy3	GTTACGCAAGGAGAGTACAGGCGGGTTCCAATACC
COX-F	GGTATTGGAAAGAACCTGTACCGGGCTTGCGTAAC
COX-R-Cy3	GTTACGCAAGCCCGGTACAGGTTCTTTCCAATACC
BOX-F	GGTATTGTTCAAATCCTGTACCGGGCTTGCGTAAC
BOX-R-Cy3	GTTACGCAAGCCCGGTACAGGATTTGAACAATACC
COC'OverlapAT-F	GGTATTGGAAGAAACTGTATTCTCCTTGCGTAAC
COC'OverlapAT-R-Cy3	GTTACGCAAGGAGAATACAGTTTCTTTCCAATACC
P12-9pb-F	CACGTCGTGCTCAACTAGAGTCTTAAGTAACTAGAGACCACGTC

P12-9pb-R	GACGTGGTCTCTAGTTAAGACTCTAGTTGAGCACGACGTG
P12-6pb-F	CACGTCGTGCCGAAGTCCGAACTAGGGACCCACGTC
P12-6pb-R	GACGTGGTCCCTAGTTCGGACCCTAGTTCGGCACGACGTG
P'23forward-F	GCCGACCGAGCGGAATAACTAGAAAGATTAAGTCTAGA
P'23forward-R	TCTAGCTAGTTAATCTTTCTAGTTATTCCGCTCGGTCGGC
P'12reverse-F	TTTTTAACTAGAATTCTAGTTATTAGCGACCGAGCGGTCG
P'12reverse-R	CGACCGCTCGGTCGCTAATAACTAGAATTCTAGTTAAAAA

D. Primers used for the *in vitro* fluorescent assay

Primer name	Sequence
LbulgattB-F	GAATTCCTGCAGCCCAAGC
Cy3-New-attB-R	GATGTAGATAATTTTGGCCAAGG