



Évolution des génomes de bactériophages

Jihane Amarir-Bouhram

► To cite this version:

Jihane Amarir-Bouhram. Évolution des génomes de bactériophages. Sciences agricoles. Université Paris Sud - Paris XI, 2012. Français. NNT : 2012PA112036 . tel-00923139

HAL Id: tel-00923139

<https://theses.hal.science/tel-00923139>

Submitted on 2 Jan 2014

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.

UNIVERSITÉ PARIS-SUD 11

ECOLE DOCTORALE : ABIES

THÈSE DE DOCTORAT

Spécialité : Microbiologie

présentée et soutenue publiquement par

Jihane AMARIR-BOUHRAM

Le 09/03/2012

ÉVOLUTION DES GÉNOMES DE BACTÉRIOPHAGES

Composition du jury : (par ordre alphabétique)

Mireille ANSALDI	Chargée de Recherches (CNRS)
Stéphanie MARSIN	Chargée de Recherches (CEA)
Mauro MODESTI	Directeur de Recherches (CNRS)
Philippe NOIROT	Directeur de Recherches (INRA)
Marie-Agnès PETIT	Directrice de thèse et Directeur de Recherche (INRA)
Suzanne SOMMER	Professeur (Université PARIS-SUD 11)

INRA, UMR 1319, MICALIS
Jouy-en-Josas
France

Remerciements

Je remercie les membres du jury et en particulier Mireille Ansaldi et Stéphanie Marsin d'avoir accepté d'examiner mon travail.

Cela fait trois années que j'ai la chance d'être passagère du magnifique bateau nommé UBLO. Je tenais à remercier chaleureusement toutes les personnes qui ont contribué de près comme de loin à l'aboutissement de ce voyage: en premier lieu, Marie-Agnès qui m'a proposé un sujet de thèse fascinant, m'a formé, m'a fait confiance, m'a épaulé et a également su me remotiver lorsqu'il le fallait. Je remercie également Raphaël Guérois et Anne Lopes avec qui j'ai été heureuse de collaborer.

Qu'aurait été cette traversée sans un équipage formidable: tout d'abord Meriem et David qui m'ont énormément conseillé au début de ce périple. Puis Colin et Marianne qui m'ont aidé à faire avancer ce projet et que je remercie sincèrement d'avoir si gentiment corrigé mon manuscrit. Merci à Geoffrey et Mélodie qui m'ont aidé à faire quelques expériences de plus pendant mon congé maternité. Merci à tous les passagers de UBLO que je n'ai pas nommé, et en particulier à Sandy.

Merci également, à toutes les personnes qui ont pris soin de moi, David, Françoise, Valérie, Aurélie, Célia, et Marianne. Merci pour votre bonne humeur, pour les pauses chocolat, bonbons ou autres qui ont rendues cette croisière inoubliable.

Je remercie mes amis, ma tata Karima et mon tonton Abdelatif, ma belle-famille et tout particulièrement ma famille. J'ai été encouragée et soutenue par mes parents, mes sœurs, mon frère et mon mari pour réaliser ce rêve. Je souhaite avoir donné envie à mes sœurs et à mon frère de faire toujours mieux, car rien n'est impossible, il suffit de s'en donner les moyens.

Je dédie cette thèse à ma maman, à mon papa, à mes sœurs, à mon frère, à ma princesse Yasmine, et enfin à mon mari qui m'a toujours encouragé, supporté et aimé, il est l'amour de ma vie et jamais je ne le remercierai assez. La fin de cette merveilleuse traversée s'est déroulée après quelques mois en mère par la naissance de mon fils Nassim, qui est tout simplement mon rayon de soleil.

L'ancre est maintenant jetée, et c'est avec joie et nostalgie que je regarde à présent le Ublo de ce grand navire s'en aller vers d'autres horizons.

Table des matières

Liste des abréviations	6
Liste des Figures	8
Liste des Tables.....	10
INTRODUCTION	11
Chapitre I. La recombinaison homologue chez les bactéries.	14
I. Rôle et structure de la recombinase RecA.	15
II. La recombinaison homologue chez <i>E. coli</i>	15
1. L'étape pré-synaptique.	16
2. L'étape synaptique.	18
3. L'étape post-synaptique.....	19
III. La recombinaison homologue indépendante de RecBCD.....	19
IV. La recombinaison homologue entre séquences d'ADN non identiques chez <i>E. coli</i>	20
Chapitre II. La recombinaison homologue chez les bactériophages.....	22
I. La recombinase UvsX du bactériophage T4.	22
II. La recombinaison homologue chez le bactériophage Lambda.....	23
1. Rôle et structure de la recombinase Red β	24
2. Les cofacteurs de Red β	25
III. L'appariement d'ADN simple brin par Red β	26
1. L'appariement d'ADN simple brin <i>in vitro</i>	27
2. L'appariement d'ADN simple brin <i>in vivo</i>	27
Chapitre III. Red β et RecT sont-elles capables de faire de l'invasion de brin ?	30
I. L'invasion de brin <i>in vitro</i>	30
II. L'invasion de brin <i>in vivo</i>	31
Chapitre IV : Red β est-elle un cofacteur de RecA?.....	37
I. Rad52 chez les eucaryotes	37
II. Les recombinaisons phagiques de la superfamille Rad52 ayant une activité cofacteur de RecA.....	39
1. Red β est un cofacteur de RecA lorsque Lambda ne peut pas se répliquer	39
2. Le recombineering Red β -dépendant est diminué dans un mutant <i>recA</i>	40
3. Sak et Sak3, deux recombinaisons de bactériophages infectant des bactéries lactiques, sont des cofacteurs de RecA <i>in vitro</i>	41
Chapitre V : Synthèse, formulation de la question poursuivie lors de la thèse	43

MATERIEL ET METHODES	44
I. Souches et plasmides	44
1. Souches.....	44
2. Plasmides.....	44
II. Préparation des cellules électro-compétentes	45
III. Le système expérimental d'appariement d'ADN simple brin <i>in vivo</i>	46
1. Description.	46
2. La quantité d'oligonucléotide.....	46
3. La taille de l'oligonucléotide.....	47
4. Meilleure efficacité de recombinaison avec un oligonucléotide complémentaire au brin tardif.	47
5. Marquages.....	48
1. L'expression des recombinaisons phagiques testées par Western Blot.	48
2. L'expression des recombinaisons phagiques testées par marquage radioactif.....	49
6. Expérience de complémentarité de la sensibilité aux UV de la souche AB1157 <i>recA</i>	50
RESULTATS.....	51
Chapitre I	51
Résumé.....	51
Article	53
Chapitre II	97
Introduction.....	97
I. Activité résiduelle d'appariement d'ADN simple brin chez <i>E. coli</i>	99
II. Les recombinaisons RecA et UvsX se comportent différemment pour l'appariement d'ADN simple brin <i>in vivo</i>	100
III. Le domaine ATPase d'UvsX est essentiel à l'appariement d'ADN simple brin <i>in vivo</i> contrairement à celui de Sak4.....	101
IV. Le domaine C-ter d'UvsX est impliqué dans le mécanisme d'appariement simple brin.....	102
V. L'activité de la recombinaison Sak4 est inactivée par la mutation K123A.....	103
VI. La recombinaison Sak4 du bactériophage HK620 apparie efficacement l'ADN simple brin en présence de ses cofacteurs.	104
VII. Les recombinaisons Red β et Sak4 de HK620 appariant l'ADN simple brin divergent <i>in vivo</i>	105
VIII. Gam stimule l'activité d'appariement d'ADN simple brin de Red β	108

DISCUSSION	110
I. <i>Escherichia coli</i> possède une activité d'appariement d'ADN simple brin en contexte <i>recA</i> ...	110
II. RecA ne fait pas d'appariement d'ADN simple brin <i>in vivo</i>	111
III. UvsX fait de l'appariement simple brin <i>in vivo</i>	112
IV. L'activité d'appariement d'ADN simple brin <i>in vivo</i> de Sak4 est indépendante de son site de fixation de l'ATP.....	112
V. Les recombinaises de bactériophage font-elles toutes de l'appariement d'ADN simple brin lorsque les séquences divergent ?	113
VI. Gam favorise l'appariement simple brin <i>in vivo</i>	116
Conclusion générale	118
BIBLIOGRAPHIE.....	119
Annexe.....	124
Résumé :	134
Summary :	134

Liste des abréviations

aa	acide aminé
AbiK	Abortive infection system K
ACLAME	A CLAssification of Mobile genetic Elements
ADN	Acide désoxyribonucléique
ADNsb	ADN simple brin
ADNdb	ADN double brin
ATP	Adenosine Triphosphate
CDB	Cassure double brin
Chi	Crossover hotspot instigator
C-ter	Carboxy-terminal
<i>E. coli</i>	<i>Escherichia coli</i>
hRad52	human Rad52
Kb	kilobase
<i>L. lactis</i>	<i>Lactococcus lactis</i>
MMR	Methyl-Mismatch Repair
nt	nucleotide
N-ter	N-terminal
pb	paire de base
PCR	Polymerase Chain Reaction

RMP	Recombination Mediator Protein
RPA	Replication protein-A
RH	Recombinaison Homologue
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SSB	Single-strand DNA binding
SF1	Superfamille 1
UV	Ultraviolet

Liste des Figures

Figure 1: Structure primaire de RecA.

Figure 2: Structure cristallographique de RecA associée à l'ADN simple brin.

Figure 3: Mécanisme d'action du complexe RecBCD.

Figure 4: Les principales étapes de la recombinaison homologue chez *E. coli*.

Figure 5: Fréquence de recombinaison homologue lors de la transformation naturelle chez *Bacillus subtilis*, en fonction de la divergence entre les séquences.

Figure 6: Comparaison de la fréquence de recombinaison en fonction du pourcentage de divergence entre deux séquences d'ADN en orientation inverse.

Figure 7: Etape pré-synaptique de la recombinaison homologue chez le bactériophage T4.

Figure 8: L'appariement d'ADN simple brin par Red β .

Figure 9: Fréquences de recombinaison par RecE/RecT (ET) et Red α /Red β ($\alpha\beta$).

Figure 10: Arbre phylogénétique du domaine Pfam Rad52_Rad22.

Figure 11: Effet de la quantité en oligonucléotide sur la fréquence de recombinaison.

Figure 12: Expression des recombinases par Western Blot.

Figure 13: Activité d'appariement d'ADN simple brin *in vivo* de diverses recombinases avec ou sans Flag (étiquette DYKDDDDK en N-ter).

Figure 14: Estimation du niveau de synthèse et de la stabilité des recombinases par marquage des méthionines au ^{35}S .

Figure 15: Estimation du niveau de synthèse et de la stabilité des recombinases Sak4_{PA73} par marquage des méthionines au ^{35}S .

Figure 16: Estimation du niveau de synthèse et de la stabilité des recombinaisons UvsX par marquage des méthionines au ^{35}S .

Figure 17: Sensibilité aux UV par spot test.

Figure 18: Sensibilité aux UV de la souche AB1157 *recA* contenant divers plasmides, ainsi que de la souche sauvage (WT).

Figure 19: Activité d'appariement d'ADN simple brin résiduelle dans la cellule.

Figure 20: Activité d'appariement d'ADN simple brin *in vivo* des recombinaisons de la superfamille Rad51/RecA.

Figure 21: Alignement de la recombinaison RecA d'*E. coli* avec la recombinaison d'UvsX du bactériophage T4 et avec la Sak4 du bactériophage PA73.

Figure 22: Prédiction du résidu de Sak4 interagissant avec l'ADN.

Figure 23: Activité d'appariement d'ADN simple brin *in vivo* de la recombinaison Sak4 du phage HK620 en fonction de la taille de l'oligonucléotide.

Figure 24 : Activité d'appariement d'ADN simple brin *in vivo* par Red β , en fonction du pourcentage de divergence de l'oligonucléotide, en contexte AB1157 *recA*mutS.

Figure 25: Activité d'appariement d'ADN simple brin *in vivo* par Red β , en fonction du pourcentage de divergence de l'oligonucléotide, en contexte AB1157 *recA*.

Figure 26: Activité d'appariement d'ADN simple brin *in vivo* par Sak4_{HK620}, en fonction du pourcentage de divergence de l'oligonucléotide, en contexte AB1157 *recA*mutS.

Figure 27: Activité d'appariement d'ADN simple brin *in vivo*, en fonction de la présence de différents gènes au voisinage de red β , sur le génome de Lambda.

Figure 28: Structure mosaïque des génomes de bactériophages.

Liste des Tables

Table 1 – Souches d'*E. coli*, et plasmides utilisés lors de ce travail.

Table 2 – Séquence des oligonucléotides utilisés dans les tests d'appariement d'ADN simple brin.

Table 3 – Activité d'appariement d'ADN simple brin *in vivo* testé pour pKD46 (Red β en présence de ses cofacteurs) avec différents oligonucléotides.

Table 4 – L'activité d'appariement d'ADN simple brin résiduelle *in vivo* n'est pas diminuée en contexte *recO*.

Table 5 – Activité d'appariement d'ADN simple brin *in vivo* pour les différents membres de la famille Rad51/RecA.

Table 6 - Efficacité de recombinaison avec pKD46 en contexte *recA mutS*, en fonction de la divergence entre séquences appariées.

Table 7 - Efficacité de recombinaison avec pKD46 en contexte *recA* (souche MAC1042), en fonction de la divergence entre séquences appariées.

Table 8 – Activité d'appariement d'ADN simple brin *in vivo* pour pJA15 (Sak4_{HK620}), en contexte *recA mutS* (souche JAC212), en fonction de la divergence des séquences appariées.

Table 9 – Activité d'appariement d'ADN simple brin *in vivo* en présence de divers cofacteurs de Red β .

INTRODUCTION

La recombinaison homologue (RH) permet l'échange d'ADN entre deux molécules et la réparation des dommages à l'ADN. Ce processus biologique permet le maintien de l'intégrité du génome, mais également la création de diversité génétique. Les protéines qui tiennent le rôle principal dans la RH sont parfois appelées synaptases, ici nous les appellerons recombinases. On les retrouve dans les 3 domaines du vivant : RecA chez les bactéries, RadA chez les archées et Rad51 chez les eucaryotes. Ces trois familles de protéines sont homologues entre elles (30% de résidus identiques entre Rad51 de *Saccharomyces cerevisiae* (*S. cerevisiae*) et RecA d'*Escherichia coli* (*E. coli*) sur une longueur de 271 acides aminés). Les gènes codant les recombinases sont en général en copie unique (par haplotype), ils ne sont pas essentiels chez les êtres vivants unicellulaires, mais sont cependant quasi universels (avec quelques exceptions dans les bactéries endosymbiontes obligatoires telles que *Buchnera* et *Blochmania*) (Rocha et al., 2005). Il existe aussi de nombreuses recombinases chez les virus et les bactériophages, elles sont beaucoup plus diverses et n'appartiennent pas toutes à la superfamille Rad51. Ce sont elles qui ont constitué le centre de mon travail de thèse.

Les bactériophages sont les virus infectant les bactéries. En 1915, Frederick Twort décrit pour la première fois des «entités» récupérées de filtrat ayant des effets antimicrobiens. Et c'est seulement en 1917, que Félix d'Herelle à l'Institut Pasteur de Paris nomme ces entités «Bactériophages». On en trouve partout où l'on trouve des bactéries (sol, mer, déserts et eaux d'égouts), ils sont uniquement visibles en microscopie électronique et il est estimé qu'il y a environ 10^{31} bactériophages sur Terre. Il existe deux principaux modes de vie chez les bactériophages. Les bactériophages dits virulents se répliquent à l'intérieur de leurs hôtes bactériens et les détruisent, réalisant alors un cycle lytique, tandis que les bactériophages dits tempérés s'intègrent parfois dans le chromosome de leur hôte (on les appelle alors

prophages), et s'y maintiennent stablement pendant de nombreuses générations, avant de déclencher un cycle lytique suite à la perception d'un signal de stress. Les bactériophages tempérés sont également pourvus d'un système de recombinaison site-spécifique, qui leur permet d'entrer et de sortir du génome de leur hôte. Cet aspect important de la recombinaison chez les bactériophages ne sera pas traité dans cette introduction.

Pendant le cycle de vie lytique du bactériophage, la RH a différents rôles documentés: elle est impliquée dans la réparation du génome (Stahl et al., 1997), la circularisation du génome pour les bactériophages à séquences terminales répétées (Botstein and Matz, 1970; Fenton and Poteete, 1984; Poteete et al., 1988; Weaver and Levine, 1977), et enfin la réplication chez certains bactériophages, comme T4 (Mosig et al., 2001). Dans le cas du bactériophage Lambda, la RH intervient indirectement au cours de la phase d'encapsidation : pendant le cycle lytique le bactériophage se réplique en mode theta puis en mode sigma pour produire de nombreux concatémères de génomes. Au moment de l'encapsidation, ce multi-génome est coupé à des sites spécifiques appelés cos (24 pb) par la terminase du bactériophage. La présence du site cos « à cheval » entre deux génomes crée la nécessité d'avoir au minimum des dimères de chromosomes pour que le génome du bactériophage soit encapsidé. En cas de défaut de production des concatémères, il faut donc que des dimères soient créés par RH à partir des monomères, soit par RecA, soit par la recombinase phagique, Red β , pour que l'encapsidation ait lieu.

Etant donné la redondance entre la fonction de RH codée par l'hôte et celle apportée par le bactériophage, il n'est pas facile de préciser le rôle particulier, ou la raison pour laquelle les bactériophages codent leur propre système de RH. Dans certains cas, comme pour le bactériophage à séquences terminales répétées P22, le gène de RH phagique est essentiel lorsque l'infection a lieu sur une souche déficiente pour *recA* (Botstein and Matz, 1970). Cependant, les souches *recA* étant très rares dans la nature, et étant donné le coût pour la bactérie de perdre le gène *recA*, il paraît

difficile de concevoir que les gènes phagiques soient présents pour parer à l'éventualité d'un défaut de RecA. Il est plus probable que la fonction de RH phagique ait des propriétés distinctes, ou qu'elle modifie à son avantage la fonction de RH de son hôte. Nous reviendrons plus loin sur ces questions.

Dans le premier chapitre de cette introduction, je présenterai la RH chez les bactéries puis dans un deuxième chapitre la RH chez les bactériophages. Je ferai en particulier le point sur l'état des connaissances actuelles sur la recombinaise phagique la plus connue, Red β , qui recombine les séquences d'ADN par un mécanisme d'appariement simple brin. Puis, je synthétiserai les résultats permettant: d'aborder la question de la redondance fonctionnelle entre Red β et RecA : Red β est-elle capable de faire de l'invasion de brin comme RecA ? Et est-elle un cofacteur de RecA ?

Chapitre I. La recombinaison homologue chez les bactéries.

Les premières découvertes des gènes impliqués dans la RH ont été faites avec la bactérie *E. coli*, à partir de mutants hypersensibles aux agents endommageant l'ADN (Clark and Margulies, 1965; Howard-Flanders and Boyce, 1966). En effet, la RH permet la réparation des dommages à l'ADN. Ces dommages sont soit des cassures double brin (CDB), soit des brèches simple brin de l'ADN, soit des fourches de réplication bloquées ou endommagées. Les dommages à l'ADN sont causés par des agents endogènes comme les métabolites cellulaires (par exemple les radicaux libres) ou exogènes comme les radiations ionisantes, les ultraviolets et les agents alkylants. Dans une culture d'*E. coli*, en absence de toute source exogène de dommage à l'ADN, il se produit au moins une CDB par cycle due à un problème lors de la réplication (Kuzminov, 1999). *E. coli* ne peut survivre qu'à seulement deux ou trois CDB dans son chromosome, d'où l'importance de réparer efficacement ces lésions (Kuzminov, 1999).

Chez les bactéries, la RH joue également un rôle important lors du transfert horizontal de gènes : elle permet à de l'ADN entrant dans la bactérie (par transduction généralisée, par conjugaison ou par transformation naturelle) de recombiner avec le chromosome de la bactérie réceptrice, créant de nouvelles combinaisons d'allèles et une plus grande diversité génétique.

La RH chez les bactéries est catalysée par la recombinase RecA. Cette protéine est aussi impliquée dans l'induction du système SOS (Radman, 1975) et de la mutagénèse PolV-dépendante (Dutreix et al., 1989; Sweasy et al., 1990), mais je ne m'intéresserai ici qu'à son rôle dans la RH.

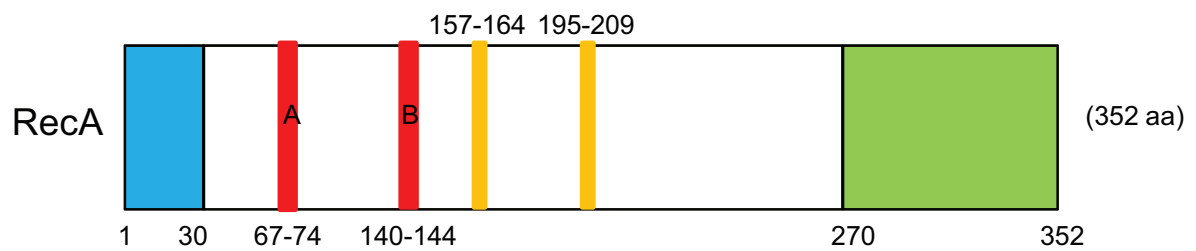


Figure 1: Structure primaire de RecA.

Domaine N-ter: Interaction protéine : protéine (bleu).

Domaine cœur: 2 sites de liaison à l'ATP (motifs Walker A et B) (rouge) et 2 sites de fixation à l'ADN simple brin (orange).

Domaine C-ter: site secondaire de fixation à l'ADN double-brin (vert).

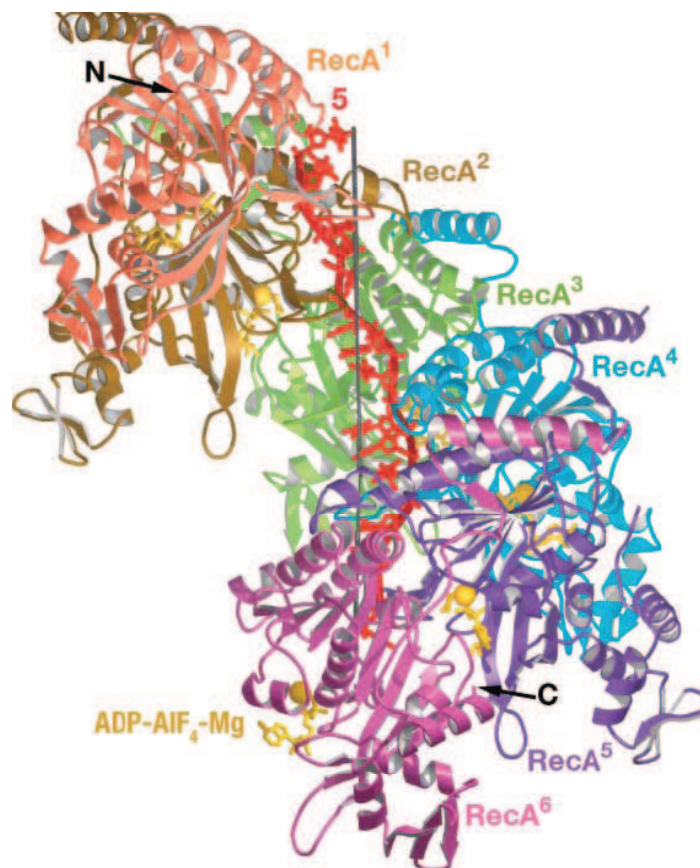


Figure 2: Structure cristallographique de RecA associée à l'ADN simple brin.

Complexe de six protomères de RecA (rose, marron, vert, cyan, violet et magenta) sur un ADN simple brin de 18 nucléotides (rouge).

Tiré de Chen et al., 2008.

I. Rôle et structure de la recombinaise RecA.

Le gène codant la protéine RecA a été identifié pour la première fois à partir de mutants d'*E. coli* hypersensibles aux agents endommageant l'ADN (Clark and Margulies, 1965; Howard-Flanders and Boyce, 1966). Depuis, une analyse bio-informatique de 117 génomes de différentes espèces bactériennes a montré que le gène codant RecA est ubiquitaire et n'est absent que chez *Buchnera* et *Blochmania* (Rocha et al., 2005), qui sont des endosymbiotes obligatoires d'insectes. RecA est une protéine de 38-kDa qui peut être divisée en trois parties : le domaine N-terminal (N-ter) qui contient des sites d'interactions protéine : protéine, la partie centrale qui contient deux sites de liaison à l'ATP (motifs Walker A et B) et deux sites de fixation à l'ADN simple brin, puis la troisième partie, le domaine globulaire Carboxy-terminal (C-ter) qui comporte un site secondaire de fixation à l'ADN double brin (Chen et al., 2008) (Figure 1). Le niveau basal de RecA chez *E. coli* est de 10 000 copies par cellule, et il monte à environ 72 000 copies après induction du système SOS aux ultraviolets (12 J.m⁻²) (Sommer et al., 1998). La structure cristallographique de RecA associée à l'ADN simple brin (Figure 2) montre que RecA se lie à l'ADN simple brin de façon dépendante de l'ATP, en formant un filament nucléoprotéique en hélice qui a 6.2 monomères de RecA par tour d'hélice et environ 3 nucléotides par monomère de RecA (Chen et al., 2008). La présence d'un nucléofilament RecA-ADNsb est nécessaire à l'initiation de la RH.

II. La recombinaison homologue chez *E. coli*.

La recombinaise RecA permet l'appariement et l'échange de matériel génétique entre séquences homologues par un mécanisme appelé l'invasion de brin, et celui-ci est composé de trois phases successives. La première est la phase pré-synaptique, qui aboutit à la formation d'un filament hélicoïdal de RecA sur l'ADN simple brin. Pendant

la deuxième étape, dite synaptique, le nucléofilament de RecA fixé à l'ADN simple brin effectue une recherche d'homologie dans un ADN double brin intact, et permet la réaction d'échange de brin qui va former une structure trois brins ADNsb-RecA-ADNdb appelée « complexe synaptique ». Enfin, l'étape post-synaptique permet la dissociation de RecA du complexe synaptique, puis la résolution de la jonction de Holliday entre les deux molécules d'ADN et la formation d'un nouvel hétéroduplex.

1. L'étape pré-synaptique.

Il existe deux types de substrats pour la recombinaison homologe : (i) une CDB ou une fourche de réplication bloquée et retournée, qui seront pris en charge par RecBCD, et (ii) une brèche simple brin ou une fourche de réplication bloquée du côté de la synthèse du brin tardif, qui seront pris en charge par RecFOR.

a. La voie RecBCD.

Le complexe RecBCD est un hétérotrimère de 330-kDa qui reconnaît les extrémités double brin linéaires. Il comprend deux hélicases de la superfamille 1 (SF1), de polarités opposées et de vitesses différentes: RecB de polarité 3'-5' et RecD de polarité 5'-3'. La protéine RecB possède également une activité nucléase qui dégrade l'ADN double brin de manière asymétrique, l'extrémité 3' étant hydrolysée de manière processive en petit fragments de 10 à 100 nucléotides (nt), alors que l'extrémité 5' est coupée moins fréquemment et en fragments plus longs (quelques kb). La troisième protéine de ce complexe est RecC. Bien que l'analyse de la séquence primaire de RecC ne révèle aucune similarité avec un domaine d'une protéine dont l'activité est connue, sa structure tridimensionnelle révèle un repliement comparable aux hélicases SF1, ce qui suggère qu'une forme ancestrale de cette protéine possédait cette activité (Dillingham and Kowalczykowski, 2008; Singleton et

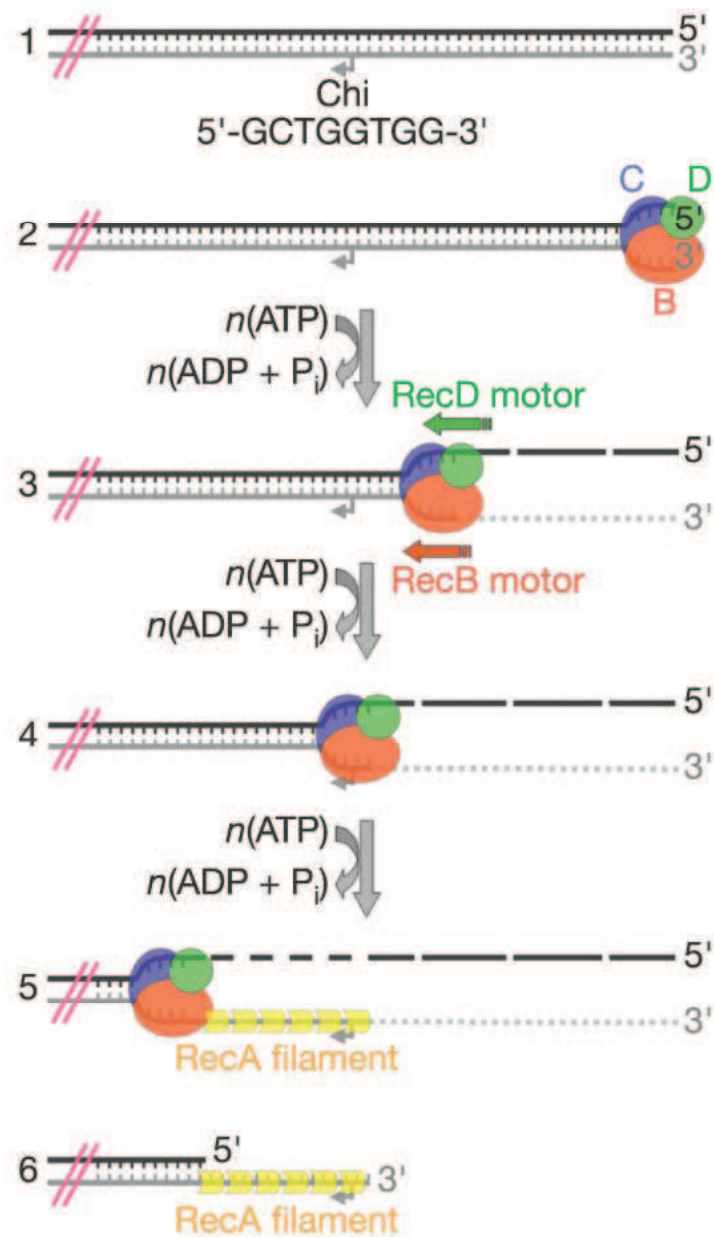


Figure 3: Mécanisme d'action du complexe RecBCD.
Tiré de Singleton et al., 2004.

al., 2004). RecC se lie à l'ADN simple brin et stimule l'activité ATPase et hélicase de RecB. De plus, lorsque RecC reconnaît un site Chi (séquence spécifique de 8 nucléotides), RecD module l'activité de RecB, entraînant un arrêt de l'activité exonucléase 3'-5' au profit d'une meilleure activité exonucléase 5'-3', ce qui permet la formation d'un duplex d'ADN avec une extrémité 3' simple brin sur laquelle RecBCD charge RecA (Figure 3, tiré de Singleton et al., 2004). Le domaine nucléase de RecB interagit directement avec RecA. Le chargement de RecA se fait immédiatement après la formation de cette extrémité 3' simple brin, ce qui empêche la protéine de liaison à l'ADN simple brin SSB (Single-Strand DNA Binding) de venir inhiber la réaction (Dillingham and Kowalczykowski, 2008; Singleton et al., 2004) (Figure 4, tiré de Michel et al., 2007).

b. La voie RecFOR.

Le complexe RecFOR permet la réparation des brèches simple brin de l'ADN. Lorsqu'il y a de l'ADN simple brin dans la cellule celui-ci est pris en charge par la protéine SSB. Celle-ci empêche l'apparition de structure secondaire et protège l'ADN contre la dégradation par des nucléases. Pour former le filament RecA-ADNsb, RecA doit déplacer SSB de l'ADN simple brin, et pour cela il existe des protéines médiatrices de la recombinaison (RMP), telle que RecFOR (Beernink and Morrical, 1999), dont la fonction est de déplacer SSB pour permettre le chargement d'une autre protéine sur l'ADN simple brin. Le complexe RecFOR peut aussi permettre la réparation des CDB dans un mutant *recBC sbcB* avec l'aide de RecJ et de RecQ. Dans ce cas, il est proposé que RecQ et RecJ jouent un rôle analogue à RecBCD : l'hélicase RecQ initie la recombinaison en ouvrant l'ADN et l'exonucléase 5'-3' RecJ dégrade l'ADN simple brin pour former une extrémité 3' qui sera le substrat de RecFOR. Suite à la formation d'une brèche simple brin dans l'ADN, le complexe RecFOR reconnaît la jonction entre la région simple brin et double brin (Figure 4). La

protéine RecF de 40-kDa se fixe à l'extrémité 5' de la jonction. La protéine RecR de 22-kDa interagit avec RecF. Et la protéine RecO de 26-kDa se fixe à RecR. La fixation du complexe RecFOR à la jonction se fait grâce à RecF, alors que la dissociation de SSB et son remplacement par RecA se fait grâce au couple protéique RecOR (Figure 4). Il existe une interaction directe entre SSB et RecO (Bork et al., 2001; Morimatsu and Kowalczykowski, 2003; Persky and Lovett, 2008). De plus, il a été montré que le domaine C-ter de RecA inhibe le déplacement de SSB, et qu'il est probable que RecO et RecR interagissent directement avec RecA en empêchant l'inhibition du domaine C-ter et permettant ainsi le déplacement de SSB par RecA (Eggler et al., 2003).

2. L'étape synaptique.

L'étape synaptique commence par la nucléation du filament, qui correspond à la fixation de cinq à six monomères de RecA sur l'extrémité 3' simple brin. Cette nucléation est suivie de l'extension du filament par addition de RecA dans le sens 5'-3', formant ainsi le filament RecA-ADNsb (Roca and Cox, 1997). Ce filament est alors capable de rechercher de l'homologie dans un duplex d'ADN intact. L'invasion de l'ADN double brin se fait par l'extrémité 3' du filament RecA-ADNsb. Ce processus consiste en l'ouverture de l'ADN double brin et le déplacement du brin homologue par le brin envahissant. Le filament envahissant peut alors s'apparier avec son brin complémentaire, créant un « complexe synaptique ». Une étude en molécule unique a permis d'estimer que la longueur de la région synaptique impliquée dans la réaction d'échange de brin est d'environ 80 pb, correspondant à un complexe synaptique de 80 pb qui se déplace le long de la région homologue et favorise une réaction d'échange de brin. De plus, dans cette étude, les auteurs estiment la vitesse d'appariement de brin à $2,11 \pm 0,05$ nt/s à 22°C et la vitesse d'échange de brin à $2,25 \pm 0,04$ nt/s à 22°C (Van der Heijden et al., 2008). Le deuxième site de fixation d'ADN dans RecA semble lui aussi jouer un rôle déterminant dans la dynamique de l'échange de brin : il permettrait d'une part au filament RecA-ADNsb d'interagir avec

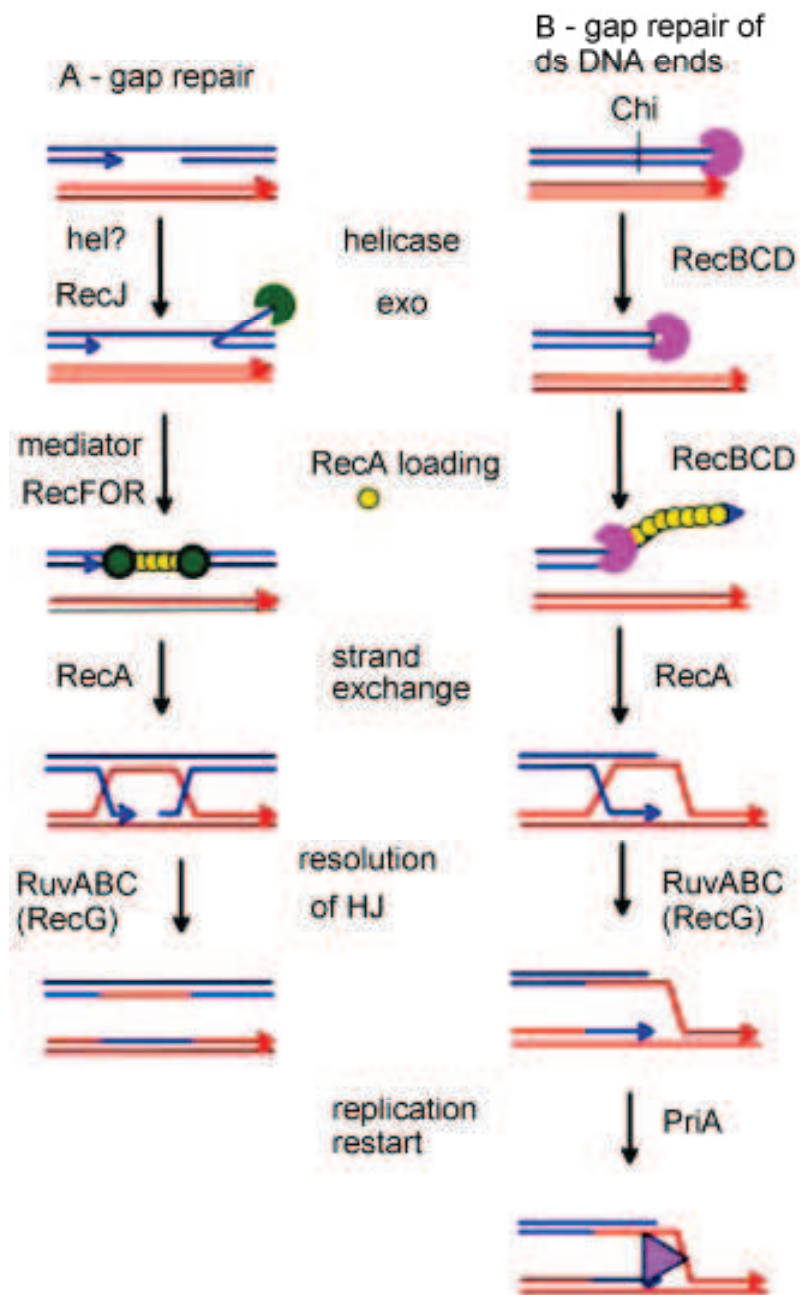


Figure 4: Les principales étapes de la recombinaison homologue chez *E. coli*. A partir d'une brèche d'ADN simple brin (A), ou d'une cassure double brin (B). Tiré de Michel et al., 2007.

le double brin, au cours de la recherche d'homologie, et d'autre part de maintenir le brin homologue à distance de l'hétéroduplex au sein du complexe synaptique (Mazin and Kowalczykowski, 1996; Mazin and Kowalczykowski, 1998). Une analyse par mutagénèse dirigée suggère que ce deuxième site de fixation est situé dans le domaine C-ter de RecA (Kurumizaka et al., 1996).

3. L'étape post-synaptique.

Une fois la réaction d'échange de brin terminée, l'ATP est hydrolysé et RecA se dissocie à partir de l'extrémité 5'. On observe alors la formation d'une structure trois brins ADNdb-ADNsb appelé D-loop pour displacement loop. Cette D-loop est convertie en une structure quatre brins appelée jonction de Holliday (Figure 4), qui est en général résolue par RuvABC. RuvA est une protéine de 22-kDa, dont deux tétramères se fixent autour de la JH. RuvB est une hélicase de 37-kDa qui forme un hexamère autour de la jonction de Holliday. RuvA et RuvB forment un complexe protéique qui catalyse de manière ATP-dépendante une réaction de migration de brin dans le sens 5'-3'. Enfin, l'endonucléase RuvC de 19-kDa assure la résolution du complexe. RuvC forme un dimère autour de la jonction de Holliday et reconnaît préférentiellement un site de coupure (5'-A/TTT^G/C-3'). Elle coupe de manière symétrique les deux brins croisés de la jonction de Holliday qui sont de même polarité. La coupure est suivie d'une ligation qui aboutit à un nouvel hétéroduplex (Figures 4) (West, 1997).

III. La recombinaison homologue indépendante de RecBCD.

Chez *E. coli*, la RH lors de la conjugaison ou de la transduction généralisée est affectée lorsque les gènes *recB* et *recC* sont mutés. Cependant, deux classes de révertants extragéniques capables de recombiner ont été isolés dans les mutants *recBC*. Une de ces deux classes correspond à l'inactivation de *sbcB* (suppresseur de

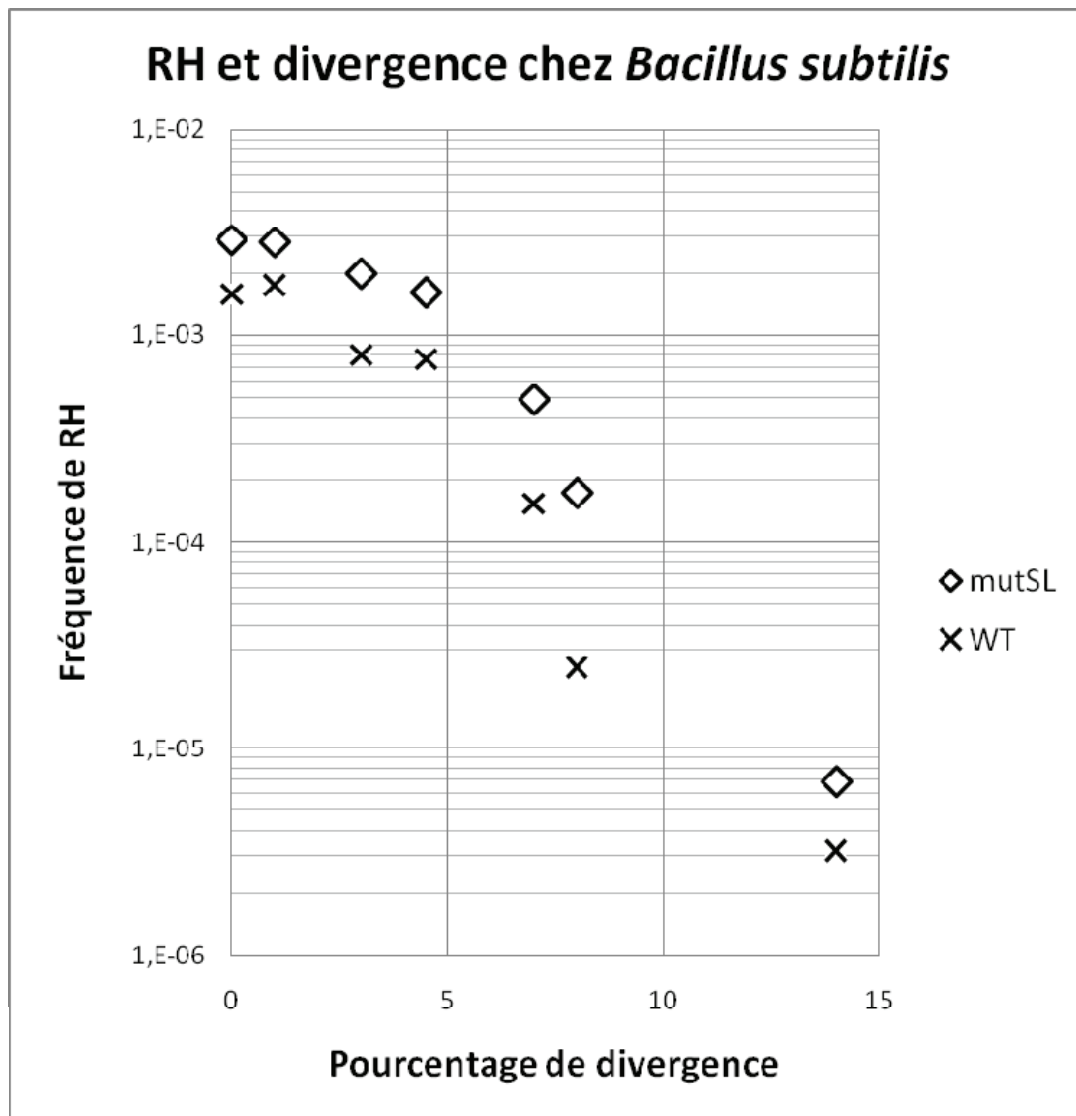


Figure 5: Fréquence de recombinaison homologue lors de la transformation naturelle chez *Bacillus subtilis*, en fonction de la divergence entre les séquences. Adapté de Majewski et Cohan, 1999.

recB recC) dont il a été parlé plus haut : l'absence de l'exonucléase I codée par le gène *sbcB* permet l'activation de la voie RecF (Kushner et al., 1972). L'autre classe correspond aux mutants *recBC sbcA*: la RH se produit alors grâce à l'activation de la recombinase RecT et de l'exonucléase RecE codées par le prophage cryptique Rac (Hall et al., 1993), qui jouent alors le rôle de cofacteurs de la RH (nous reviendrons sur ce point dans le chapitre suivant). RecT est homologue à la recombinase du bactériophage Lambda, Red β , dont je parlerai dans le prochain chapitre. La protéine RecE est une exonucléase 5'-3', ATP indépendante, de fonction analogue à l'exonucléase du bactériophage Lambda Red α . Curieusement, RecE est une protéine de 866 aa alors que Red α n'a que 226 aa, et des études génétiques ont montré que les 300 derniers aa de RecE suffisent pour l'activité nucléasique (Luisi-DeLuca et al., 1988).

IV. La recombinaison homologue entre séquences d'ADN non identiques chez *E. coli*.

RecA peut aussi catalyser l'initiation d'une réaction d'échange de brin entre séquences d'ADN apparentées mais non identiques, cependant celle-ci est généralement avortée par le système de réparation des mésappariements (MMR pour *methyl-directed mismatch repair*, en anglais). Différentes études en contexte muté pour le MMR permettent d'avoir une idée du pourcentage de divergence toléré par RecA. Chez *E. coli*, RecA permet la recombinaison intrachromosomique entre séquences *rhs* (pour recombination hot spot, en anglais) ayant 2 à 3% de divergence (Petit et al., 1991). Chez *Bacillus subtilis*, Majewski et Cohan ont étudié comment l'efficacité de recombinaison lors de la transformation, était affectée par le pourcentage d'identité du fragment PCR avec sa cible, le gène *rpoB*. Une courbe biphasique a été obtenue : la recombinaison est efficace jusque environ 5% de divergence, puis chute, mais continue à se produire jusqu'à 14% de divergence (Figure 5, adapté de Majewski et Cohan, 1999). A des niveaux plus élevés de divergence, de la recombinaison homologue a été observée pendant la conjugaison

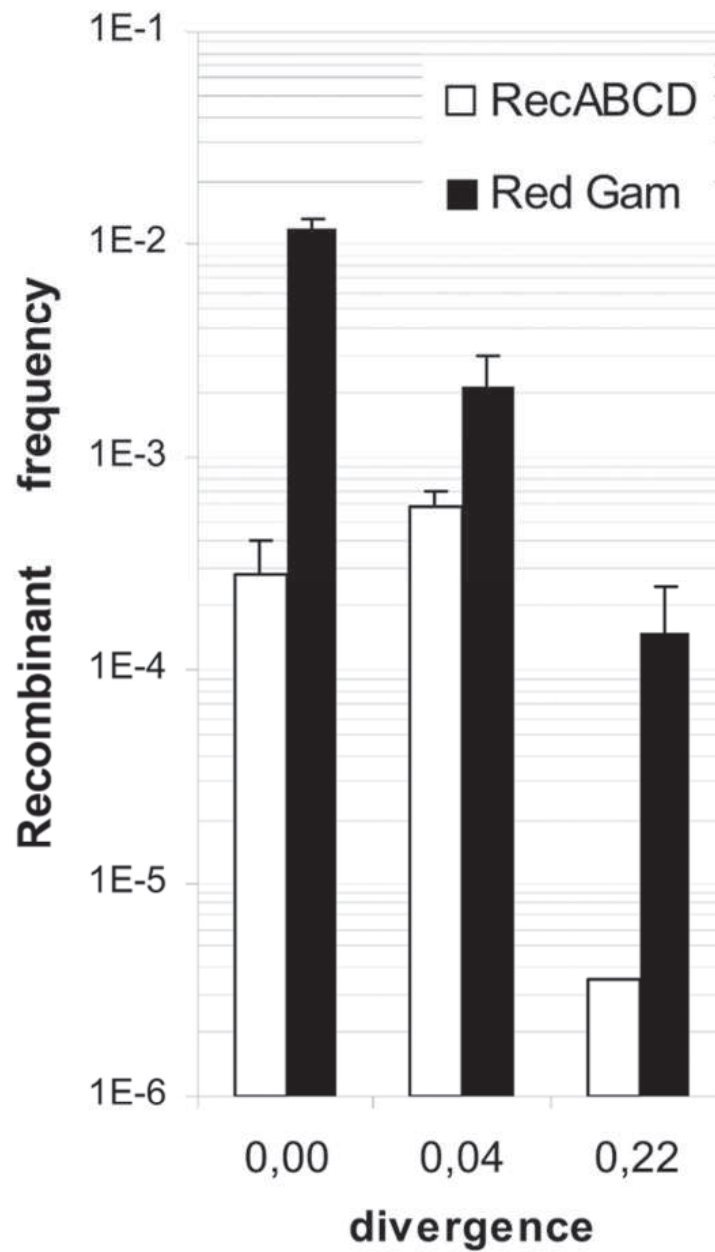


Figure 6: Comparaison de la fréquence de recombinaison en fonction du pourcentage de divergence entre deux séquences d'ADN en orientation inverse. Tiré de Martinsohn et al., 2008.

entre *E. coli* et *Salmonella typhimurium* (le génome de ces deux espèces est en moyenne divergent à 20%), lorsque la bactérie réceptrice est mutée pour les gènes *mutL*, *mutS* ou *mutH*. Cependant, dans la plupart des recombinants obtenus, on ne connaît pas le pourcentage d'identité des séquences des régions où s'est produite la RH (dans trois cas sur 81 où une jonction a pu être identifiée, elle est située dans les séquences des opérons ribosomiques *rrn*, ayant seulement 4 à 5% de divergence entre *E. coli* et *S. typhimurium*) (Matic et al., 1994; Rayssiguier et al., 1989). Une dernière étude a montré que RecA peut permettre à faible fréquence la recombinaison entre deux séquences d'ADN divergées à 22%, dans un modèle de recombinaison intramoléculaire à l'intérieur du génome du bactériophage Lambda (Figure 6, tiré de Martinsohn et al., 2008) (Martinsohn et al., 2008).

Chapitre II. La recombinaison homologue chez les bactériophages.

A la différence des recombinases des autres organismes, qui sont toutes apparentées à RecA, nous avons montré que les recombinases phagiques appartiennent à au moins trois grandes familles de protéines (voir premier chapitre de la partie Résultats), nommées d'après leur représentant le mieux caractérisé: celles de type Rad51/RecA, celles de type Rad52, et celles de type Gp2.5 (Lopes et al., 2010). Le présent chapitre rappellera brièvement ce qui est connu d'UvsX, protéine de type Rad51/RecA codée par le génome du bactériophage T4, et de Red β , protéine de type Rad52 codée par le génome du bactériophage Lambda, avant de présenter dans les deux derniers chapitres de cette introduction les questions en suspens sur les recombinases de type Rad52. Nettement moins de travaux ayant été publiés sur la recombinase de type Gp2.5, celle-ci sera simplement présentée au sein de la première partie des Résultats.

I. La recombinase UvsX du bactériophage T4.

La recombinase UvsX est une protéine de 44-kDa qui possède de l'homologie avec RecA sur toute sa longueur (24% de résidus identiques sur 276 alignés). *In vivo*, un mutant *UvsX* du bactériophage T4 est sensible aux UV, et sa réplication est arrêtée au milieu du cycle (Conkling and Drake, 1984). *In vitro*, UvsX est capable de promouvoir les mêmes réactions d'invasion de brin que RecA d'*E. coli*. Elle en diffère cependant par ses aspects cinétiques : *in vitro*, UvsX est plus rapide que RecA pour la réaction d'échange de brin (Liu and Morrical, 2011). De manière remarquable, le génome de T4 code l'ensemble des protéines pré- et post-synaptiques nécessaires à la RH : Lors de l'étape pré-synaptique, les CDB sont prises en charge par un complexe protéique, composé de Gp46 et Gp47, qui possède une activité exonucléasique de polarité 5'-3' générant une extrémité 3' simple brin, qui sera le

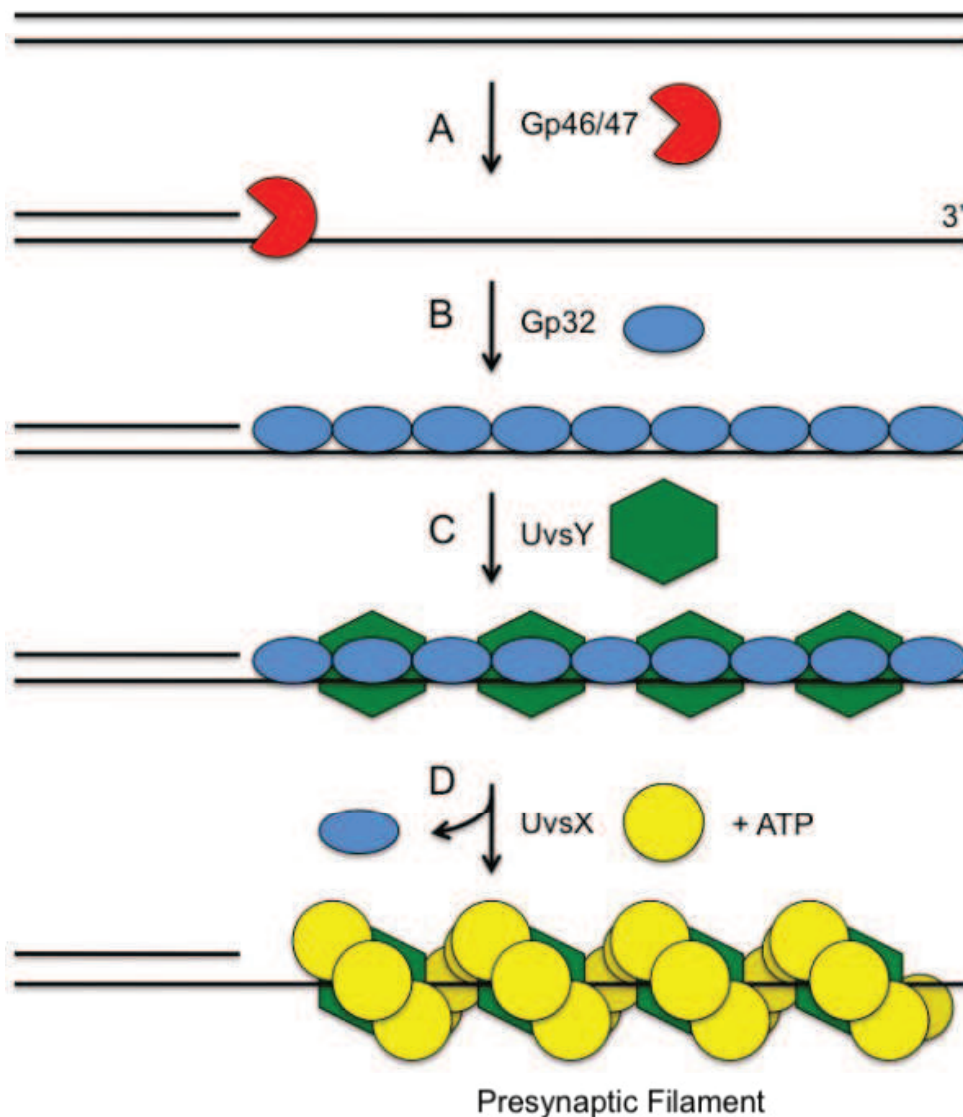


Figure 7: Etape pré-synaptique de la recombinaison homologue chez le bactériophage T4.

Le complexe Gp46/47 dégrade l'extrémité 5' pour créer une extrémité 3' simple brin (A). L'extrémité 3' simple brin est recouverte de protéines Gp32 (B). La protéine médiatrice de la recombinaison UvsY se fixe à l'ADN simple brin couvert de Gp32 (C). Les protéines UvsY permettent le recrutement de la recombinase UvsX, en présence d'ATP, à la place de Gp32. Il y a ainsi formation du filament pré-synaptique.(D)

Tiré de Liu and Morrical, 2011.

substrat d'UvsX (Figure 7, tiré de Liu and Morrical, 2011). Cette extrémité 3' simple brin est recouverte de Gp32, protéine de liaison à l'ADN simple brin, orthologue de la protéine SSB d'*E. coli*. Puis lors de l'étape synaptique, UvsY, qui est une RMP différente de celles décrites chez *E. coli* (voir chapitre I), va déstabiliser le complexe Gp32-ADNsb et faciliter la fixation d'UvsX à l'ADN simple brin (Beernink and Morrical, 1999), pour former le complexe UvsX-UvsY-ADNsb (Figure 7). Ce complexe forme un nucléofilament en forme d'hélice, dans lequel six protomères d'UvsX fixés à un hexamère d'UvsY, se fixent à vingt-quatre nt de l'ADN. Par ailleurs, UvsX peut aussi se lier à l'ADN double brin, mais la présence d'UvsY et d'ATP stabilise sa liaison à l'ADN simple brin. L'activité ATPasique d'UvsX est dépendante de l'ADN simple brin et est différente de celle associée à RecA, car elle permet de former deux produits : l'ADP et l'AMP (RecA ne forme que de l'ADP). Une fois le filament UvsX-UvsY-ADNsb formé, il y a recherche d'homologie dans un ADN double brin intact, appariement du filament à son brin complémentaire et échange de brin pour former la D-loop. Enfin, à l'étape post-synaptique, la protéine UvsW, une hélicase de la superfamille 2 va permettre la migration de brin. Cette migration est suivie par la résolution de la jonction de Holliday grâce à l'endonucléase VII, et aboutit à la formation d'un nouvel hétéroduplex (Liu and Morrical, 2011).

II. La recombinaison homologue chez le bactériophage Lambda.

C'est en 1967 que Brooks et Clark ont mis en évidence que le bactériophage Lambda recombine aussi bien dans une souche d'*E. coli* mutée pour *recA* que dans une souche sauvage. Ils en ont conclu que Lambda avait son propre mécanisme de RH (Brooks and Clark, 1967). Puis d'autres études ont identifié des bactériophages mutés qui recombinaient beaucoup moins dans un mutant *recA* (Franklin, 1967) (Echols et al., 1968; Signer and Weil, 1968). Ces bactériophages étaient mutés dans les gènes *exo* et *bet*. Le bactériophage Lambda possède trois gènes en opéron, *bet*,

exo et *gam*, sous le contrôle du promoteur P_L , qui codent respectivement pour les protéines Red β , Red α et Gam. Ces trois protéines sont impliquées dans la RH, qui est le plus souvent indépendante de RecA lorsque la réplication du bactériophage est active, et se produit par appariement d'ADN simple brin. Lorsque la réplication du bactériophage est inactive, la RH est dépendante de RecA et se produit par invasion de brin (Stahl et al., 1997).

1. Rôle et structure de la recombinase Red β .

Red β est une protéine de 28-kDa, pas encore cristallisée. Cependant, une étude en microscopie électronique de 1999 a montré que Red β existe sous trois formes structurales (Passy et al., 1999). En absence d'ADN elle forme un anneau inactif composé de 11 ou 12 monomères. En présence d'ADN simple brin, elle forme un anneau actif contenant 15 à 18 monomères. Et enfin, lorsque deux ADN simple brin couverts de Red β s'apparient, l'ADN double brin ainsi formé est entouré de Red β formant une hélice gauche. Une étude plus récente d'Erlér et al., en 2009, a montré par microscopie à force atomique les mêmes résultats à faible résolution, en absence d'ADN, mais à forte résolution, ils ont observé que ce qui semblait être un anneau de Red β en absence d'ADN (et détecté également en microscopie à force atomique à faible résolution), formait en réalité une légère hélice droite à forte résolution (Erlér et al., 2009). En présence d'ADN simple brin, ils ont observé des formes hétérogènes, et lorsque deux ADN simple brin couverts de Red β forment un ADN double brin, ils ont également observé une hélice gauche, avec 14 monomères de Red β par tour d'hélice où chaque monomère de Red β se lie à 11 nt (Erlér et al., 2009; Passy et al., 1999).

Plus généralement, les protéines Red β , RecT, ERF (Essential Recombination Function), la recombinase du bactériophage P22, et Rad52, une RMP associée à la recombinase eucaryote Rad51 sont toutes des protéines d'appariement à l'ADN simple brin. Ce sont des analogues fonctionnels qui ont tous la même structure quaternaire en forme d'hélice, mais ils n'ont ni séquences, ni structures secondaires en commun. Une première étude bio-informatique d'Iyer et al, en 2002, avait tenté de regrouper ces diverses protéines par alignement, dans une seule famille, mais sans succès : ils ont pu placer RecT et Red β dans une même famille, ERF et quelques protéines apparentées dans une autre, et Rad52 ainsi que la protéine Sak du bactériophage ϕ 36 infectant *Lactococcus lactis* (*L. lactis*) dans une troisième (Iyer et al., 2002). Cependant, un alignement publié dans une étude plus récente, en 2009, suggèrait que RecT, Red β et Rad52 appartiennent à une même superfamille (Erler et al., 2009).

2. Les cofacteurs de Red β .

Les bactéries possèdent des systèmes de dégradation de l'ADN double brin linéaire, comme par exemple RecBCD chez *E. coli*, RexAB chez *Lactococcus lactis* et *Streptococcus pneumoniae*, et AddAB chez *Bacillus subtilis*. Ces systèmes dégradent inefficacement l'ADN bactérien car celui-ci est protégé par les sites Chi, mais en revanche ils peuvent dégrader l'ADN phagique. Les bactériophages possèdent donc un système de protection pour éviter cette dégradation. Chez Lambda, l'activité anti-RecBCD est possible grâce à la présence de la protéine Gam (Karu et al., 1975; Unger and Clark, 1972). Cette protéine de 16,5-kDa se fixe à RecBCD, et inhibe son activité nucléasique. Ainsi, les extrémités double brin sont disponibles pour être prises en charge par la protéine Red α . C'est une exonucléase de 24-kDa, de polarité 5'-3', dont la structure est un homotrimère en forme toroïdale. Red α est très processive, elle dégrade 1000 bases par seconde, créant ainsi une

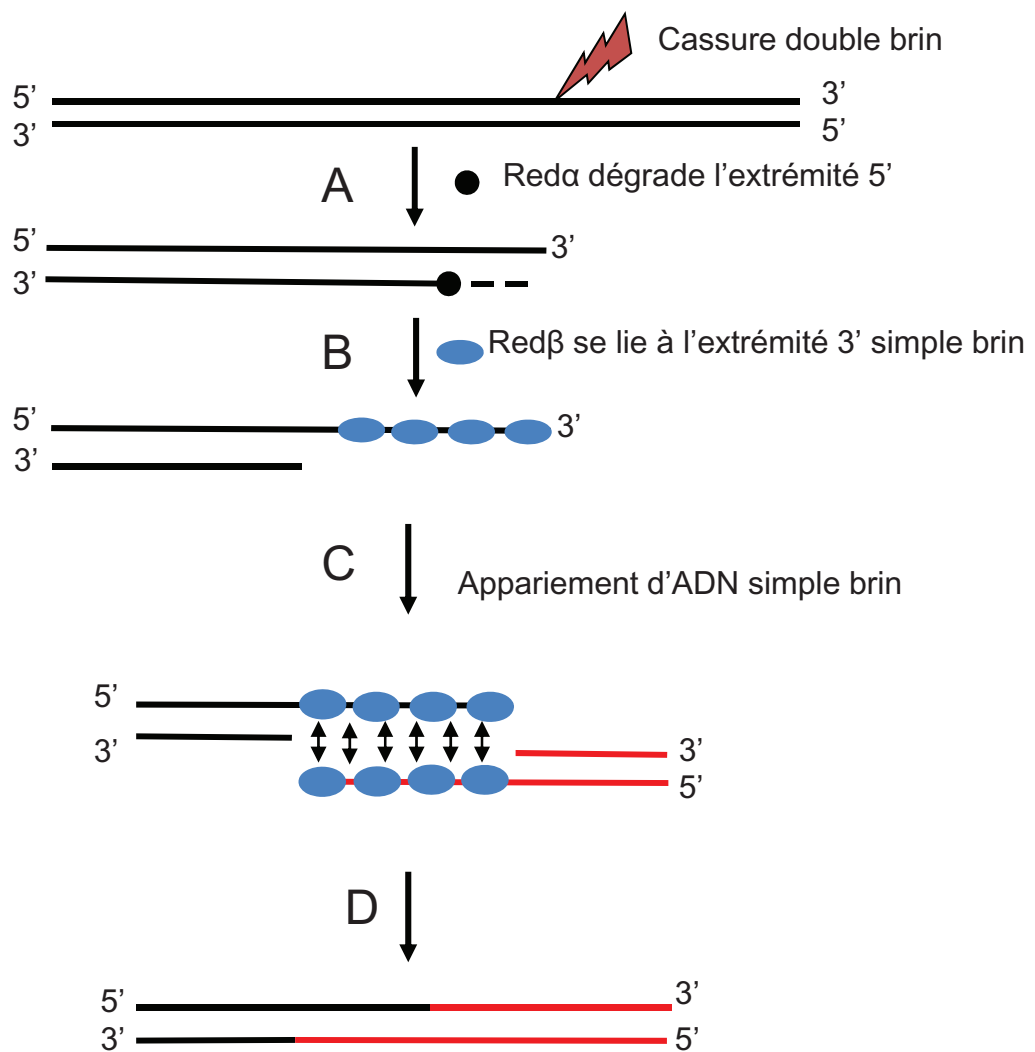


Figure 8: L'appariement d'ADN simple brin par Red β .

Suite à une cassure double-brin dans le chromosome, l'exonucléase Red α dégrade l'extrémité 5' pour créer une extrémité 3' simple brin (A). L'extrémité 3' simple brin est recouverte de recombinaisons Red β (B). Deux molécules d'ADN cassées, recouvertes de Red β , s'apparient pour former un nouvel hétéroduplex intact (D).

extrémité 3' simple brin qui est le substrat de Red β (Figure 8) (Kovall and Matthews, 1997; Subramanian et al., 2003).

Quelques observations suggèrent que Red α interagit physiquement avec Red β : *In vivo*, la RH, fonctionnelle pour les paires Red α /Red β et RecE/RecT, est abolie lorsque les exonucléases sont interchangeées (Muyrers et al., 2000). De plus, biochimiquement, une interaction sur colonne a été observée entre Red α et Red β (Muyrers et al., 2000). Cependant, aucune donnée biochimique plus directe n'existe à notre connaissance, et à l'inverse d'autres données *in vivo* suggèrent que les exonucléases de ces deux recombinaases sont interchangeables: lorsque l'exonucléase Red α est remplacée par RecE dans le bactériophage Lambda, celui-ci est toujours capable de se multiplier (Gillen et al., 1977; Gottesman et al., 1974). De plus, la recombinaison homologue sur le bactériophage Lambda reste efficace lorsque Red α est remplacée par RecE (données personnelles).

III. L'appariement d'ADN simple brin par Red β .

Lorsque deux molécules d'ADN sont cassées en deux points différents, puis prises en charge par Red α qui crée deux extrémités 3' simple brin, Red β s'y fixe dans le sens 3'-5' (sens opposé à la fixation de RecA à l'ADN simple brin qui se fait dans le sens 5'-3') et permet l'appariement de deux ADN simple brin complémentaires de manière ATP indépendante (Figure 8). Puis, les deux molécules d'ADN sont liguées et forment un nouvel hétéroduplex. Red β ne se fixe pas directement à l'ADN double brin, mais lorsque deux ADN simple brin se sont appariés, Red β reste fixé au duplex. Ce complexe ADN_{db}-Red β est plus stable que le complexe ADN_{sb}-Red β . La taille minimale nécessaire pour une réaction stable d'appariement simple brin est de 16 à 20 nt (Erler et al., 2009). De nombreuses études *in vitro* et *in vivo* ont montré que Red β et RecT sont capables de faire de l'appariement d'ADN simple brin. Je

présenterai quelques-unes de ces études afin de faire un point sur l'état des connaissances actuelles.

1. L'appariement d'ADN simple brin *in vitro*.

Des études ont montré que deux molécules d'ADN simple brin complémentaires de 41 kb du bactériophage P22 sont capables de s'apparier *in vitro* en présence de la recombinaise Red β (Kmiec and Holloman, 1981; Muniyappa and Radding, 1986). Le pourcentage de réappariement est défini par le pourcentage d'ADN double brin, créé par l'appariement, résistant à la digestion par la nucléase S1. Selon les études, un pourcentage de réappariement de 44% ou 80% a été observé (Kmiec and Holloman, 1981; Muniyappa and Radding, 1986). Il a également été montré que Red β est capable d'apparier deux ADN simple brin complémentaires du bactériophage Φ X174, molécule de seulement 4,5 kb (au maximum, 60% de l'ADN total est réapparié), également en présence de SSB (au maximum 40% d'appariement à une concentration d'ADN basse, 20% à des concentrations d'ADN plus élevée) (Muniyappa and Radding, 1986).

De même, la protéine RecT est capable de faire de l'appariement d'ADN simple brin. En effet, deux ADN simple brin complémentaires du bactériophage T7 (40 kb) s'apparient avec une efficacité maximale de 65% (Hall et al., 1993).

2. L'appariement d'ADN simple brin *in vivo*

Plusieurs travaux plus récents ont mis en évidence l'activité d'appariement d'ADN simple brin de Red β *in vivo*. Ces expériences consistent à introduire dans une souche d'*E. coli* ayant une mutation ponctuelle dans le gène *galK*, la rendant auxotrophe pour le galactose, un oligonucléotide correspondant au gène *galK*

sauvage. Ainsi, lorsque l'oligonucléotide recombine avec le chromosome, la souche devient prototrophe pour le galactose (Gal⁺). Dans une série de travaux du laboratoire de Donald Court, la souche d'*E. coli* utilisée contient une partie du prophage Lambda, en particulier l'allèle *cl857* du répresseur *cl*, et l'opéron *gam red* dont l'expression est induite par un changement de température. En effet, le répresseur *cl857* est thermosensible, et à une température supérieure à 37°C, la transcription du promoteur *P_L* est déréprimée et l'opéron *gam red* induit. Après induction, les cellules sont préparées pour l'électroporation, puis transformées avec un oligonucléotide codant l'allèle sauvage du gène *galk*. En cas de recombinaison avec cet oligonucléotide, la souche devient donc Gal⁺, permettant d'estimer la fréquence de recombinaison. Les résultats obtenus montrent que la recombinaison est indépendante de RecA (Ellis et al., 2001) (Costantino and Court, 2003) et que Redβ est la seule protéine indispensable (Ellis et al., 2001). Ellis et al. ont montré que la taille de l'oligonucléotide utilisé influence l'efficacité de recombinaison. La fréquence de recombinaison la plus élevée ($2,2 \times 10^5$ recombinants pour 10^8 cellules viables) est obtenue avec un oligonucléotide de 70 nt, alors qu'avec un oligonucléotide dont la taille est comprise entre 40 et 60 nt le taux de recombinants Gal⁺ obtenu est stable et cinq fois plus faible ($4-5 \times 10^4$ recombinants pour 10^8 cellules viables). Et enfin, avec un oligonucléotide de 20 nt le taux de recombinant obtenu est équivalent au taux de mutants Gal⁺ spontanés (Ellis et al., 2001). Une étude plus récente de Sawitzke et al. a mis en évidence qu'avec des oligonucléotides dont la taille est comprise entre 40 et 70 nt, la fréquence de recombinaison augmente (de 5×10^6 à 5×10^7 recombinants pour 10^8 cellules viables). Cette fréquence est diminuée d'un facteur 20 lorsque la taille de l'oligonucléotide est comprise entre 40 et 23 nt et elle est équivalente à la fréquence de mutants Gal⁺ spontanés lorsque l'oligonucléotide comprend 15 à 20 nt (Sawitzke et al., 2011) (Swingle et al., 2010). L'appariement d'ADN simple brin permet aussi bien l'introduction d'une mutation ponctuelle dans le chromosome d'une souche, que la

formation d'une délétion de plusieurs kb (Costantino and Court, 2003; Datta et al., 2008; Ellis et al., 2001).

Si la fonction d'appariement d'ADN simple brin de Red β et RecT est bien établie, les données qui suggèrent qu'elles sont également capables de faire de l'invasion de brin, comme RecA, sont plus contrastées. C'est cet aspect qui sera examiné dans le prochain chapitre.

Chapitre III. Red β et RecT sont elles capables de faire de l'invasion de brin ?

I. L'invasion de brin *in vitro*.

Différentes études biochimiques ont permis de mettre en évidence une activité d'invasion de brin pour Red β et RecT, en utilisant cependant des conditions atypiques par rapport au test effectué avec RecA.

Une première étude de Hall et Kolodner, en 1994, a mis en évidence que RecT et RecE sont capables de former une molécule jointe, de manière ATP indépendante, à partir d'une molécule d'ADN double brin linéaire et d'une molécule d'ADN simple brin circulaire de bactériophage M13mp19. RecE dégrade l'ADN double brin linéaire dans le sens 5'-3', créant une extrémité 3' simple brin. Celle-ci est prise en charge par RecT qui permet son appariement à l'ADN simple brin circulaire, et le brin non complémentaire est déplacé. Les conditions optimales pour former 40 à 50% de molécules jointes sont la présence d'un tétramère de RecT pour 13 nt d'ADN simple brin et d'1,5 tétramères de RecE par extrémité d'ADN double brin linéaire, de 0,3 mM de MgCl₂, d'un pH à 7.0, et d'une température à 37°C. Lorsque l'ADN simple brin circulaire est hétérologue (bactériophage Φ X174), la formation d'une molécule jointe n'est pas observée. De plus, l'appariement et l'échange de brin se fait en absence de RecE, si l'ADN double brin linéaire a une extrémité simple brin 3' ou 5'. Les auteurs concluent que RecT permet l'appariement d'un ADN complémentaire et l'échange de brin à partir d'une extrémité simple brin sans polarité préférentielle (Hall and Kolodner, 1994).

Par la suite, les travaux de Noirot et Kolodner ont montré que RecT pouvait initier l'invasion de brin *in vitro* via la formation de D-loop, en absence ou à très faible concentration (1mM) de MgCl₂, et pour des séquences riches en AT (seulement 16% de G+C). Ils ont ensuite montré que RecT initiait l'invasion de brin à partir d'une

extrémité simple brin sans polarité préférentielle, et que RecT permettait la formation d'une structure trois brins ADNsb-ADNdb (Noirot et al., 2003; Noirot and Kolodner, 1998). De manière intéressante, ils ont observé que la structure trois brins se formait même lorsque l'oligonucléotide était hétérologue. Cependant, il faut noter que l'oligonucléotide de 51 nt utilisé possède une région de 25 nt ayant 76% d'identité avec le substrat double brin.

Une D-loop peut également être formée par Red β à partir d'un plasmide et d'un oligonucléotide de 83 nt riche en AT complémentaire de celui-ci (Rybalchenko et al., 2004). De manière similaire à la réaction catalysée par RecT, la formation de la D-loop dépend de la composition en base de l'oligonucléotide et de la concentration en MgCl₂. En effet, avec l'oligonucléotide composé à 16% de G+C, la formation de D-loop est 12 fois plus efficace qu'avec l'oligonucléotide ayant 37% de G+C. De plus, une concentration supérieure à 2 mM en MgCl₂ inhibe la réaction. De même une réaction d'échange de brin est observée pourvu que l'oligonucléotide soit riche en AT. La réaction est cependant moins rapide et moins complète qu'avec RecA (41% d'échange en 40 minutes pour Red β , contre 75% en 20 minutes avec RecA) (Rybalchenko et al., 2004).

En conclusion, les données biochimiques suggèrent que Red β et RecT ont bien une activité d'invasion de brin, mais faible comparée à RecA, et limitée aux substrats riches en AT.

II. L'invasion de brin *in vivo*

Pour étudier l'invasion de brin *in vivo*, Zhang et al. ont utilisé un système expérimental basé sur la co-transformation chez *E. coli* d'une molécule d'ADN plasmidique circulaire et d'une molécule d'ADN double brin linéaire. L'ADN double

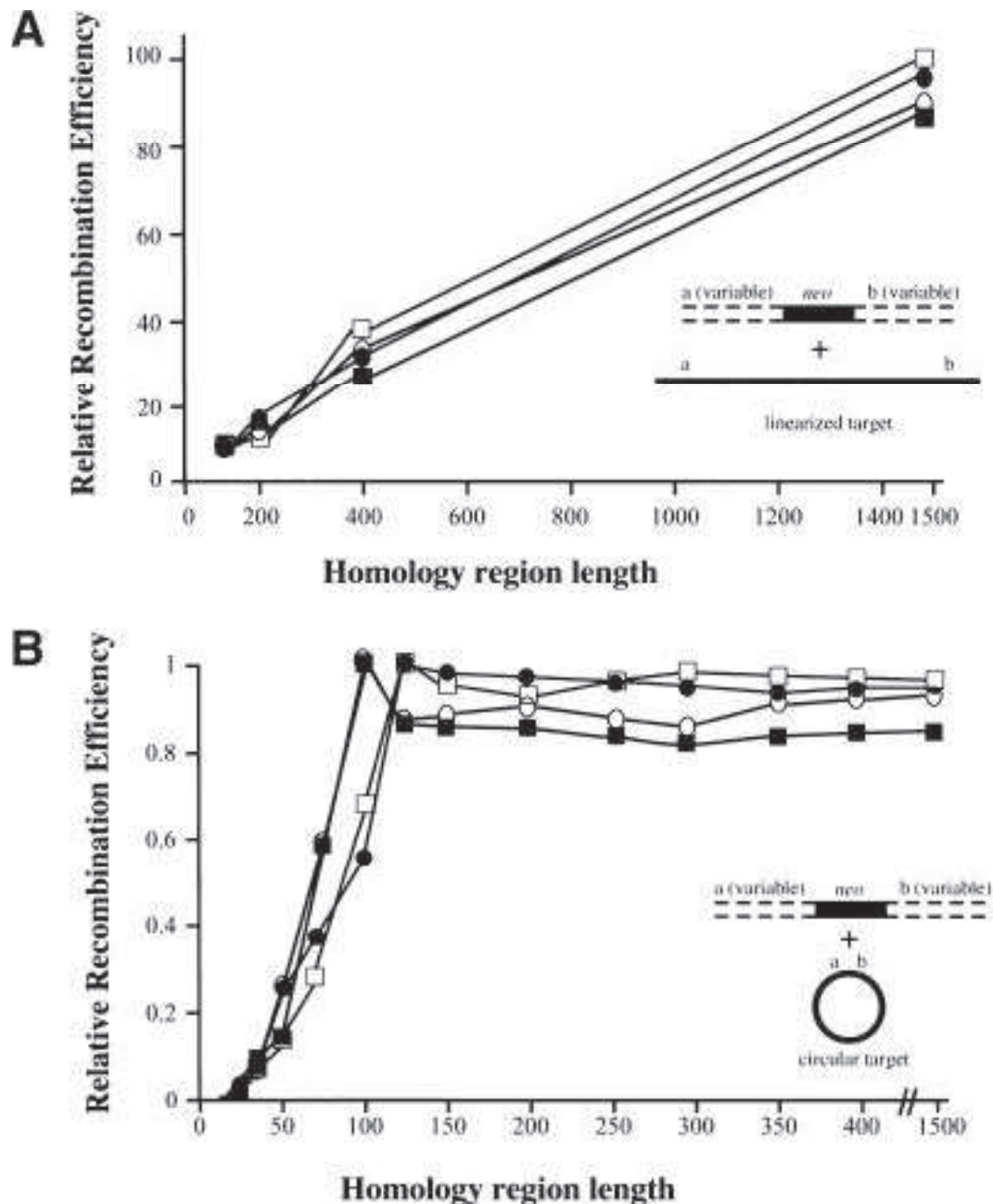


Figure 9: Fréquences de recombinaison par RecE/RecT (ET) et Red α /Red β ($\beta\alpha$). Recombinaison par appariement d'ADN simple brin (A) entre une molécule d'ADN double brin linéaire et différents fragments de PCR, dont la taille des régions homologues (a et b) varie de 20 à 1500 nt. Recombinaison par invasion de brin (B) entre une molécule d'ADN plasmidique et les différents fragments de PCR. (■) RecA⁻ ET; (○) RecA⁺ ET; (●) RecA⁻ $\beta\alpha$; (□) RecA⁺ $\beta\alpha$. Tiré de Muyrers et al., 2000.

brin linéaire possède un gène codant pour la résistance à un antibiotique, entouré de deux régions homologues au plasmide. La recombinaison entre les régions homologues permet l'acquisition du gène de résistance à l'antibiotique par le plasmide, qui est alors facilement sélectionnable. Dans ces conditions, la recombinaison se produit dans un mutant *recBC sbcA*, donc en présence de la recombinase RecT et de l'exonucléase RecE du prophage Rac, et elle est abolie dans le triple mutant *recBC sbcA recE* (Zhang et al., 1998). De manière intéressante, la recombinaison observée est indépendante de RecA (Zhang et al., 1998). Ces travaux montrent donc que RecT peut faire de l'invasion de brin *in vivo*, indépendamment de RecA, cependant, ils ne permettent pas d'apprécier l'importance de cette activité par rapport à l'activité d'appariement d'ADN simple brin.

L'étude de Muyrers et al, en 2000, basée sur un système expérimental similaire, confirme les observations de Zhang et al., et permet de comparer les efficacités d'invasion de brin et d'appariement d'ADN simple brin de RecT et Red β . En effet, en plus de la co-transformation d'un plasmide circulaire et d'un ADN double brin linéaire comme dans le système expérimental de Zhang et al., leurs travaux rapportent des expériences où le plasmide circulaire est lui aussi linéarisé, dans la région homologue aux séquences du fragment linéaire (Figure 9, tiré de Muyrers et al., 2000). Dans ce cas, la recombinaison homologue doit se produire par appariement d'ADN simple brin. L'expérience est faite en faisant varier la longueur d'homologie aux extrémités du fragment linéaire contenant le marqueur antibiotique. L'efficacité de l'appariement d'ADN simple brin augmente linéairement avec la taille de l'homologie entre 20 et 1500 nt. En revanche, dans l'expérience d'invasion de brin, l'efficacité de recombinaison augmente seulement jusqu'à une longueur de 100 nt d'homologie, puis plafonne. Dans ces conditions expérimentales, les auteurs ont donc pu établir que l'invasion de brin, à son efficacité maximale (100-200 nt d'homologie) est 10 fois moins élevée que l'appariement d'ADN simple brin, et que

lorsque la longueur d'homologie augmente, elle devient très minoritaire (80 fois plus d'appariement d'ADN simple brin) (Figure 9) (Muyrers et al., 2000).

Il est donc possible de conclure à ce stade que RecT et Red β font de l'invasion de brin *in vivo*, mais avec une efficacité 10 à 100 fois moindre que l'appariement d'ADN simple brin.

Les recombinaisons Red β et RecT se sont largement popularisées dans les laboratoires de biologie moléculaire depuis l'année 2000, car elles peuvent être utilisées pour remplacer un gène par un autre, pour ajouter ou déléter des gènes (technique du « recombineering », pour recombination-mediated genetic engineering). Trois publications ont rapporté à cette époque une méthode pour déléter un gène à l'intérieur du chromosome d'*E. coli*, basée sur la capacité de Red β à recombiner des fragments ayant une courte homologie avec le chromosome, pourvu qu'elle soit présente à l'extrémité d'une molécule linéaire (Datsenko and Wanner, 2000; Murphy, 1998; Yu et al., 2000). Pour cela, un produit de PCR contenant un gène codant pour une résistance à un antibiotique entouré de régions homologues au gène que l'on veut déléter, est intégré dans le chromosome, en présence d'un plasmide qui code pour Red β , Red α et Gam, sous contrôle du promoteur de force modérée *ParaBAD*, inductible par l'arabinose en présence d'AraC (plasmide pKD46, Datsenko et Wanner). Une variante consiste à produire ces protéines à partir du chromosome, sous contrôle du promoteur fort *P_L*, lui-même réprimé par *ci* (Yu et al.). Ainsi, le gène qui code pour la résistance à l'antibiotique est intégré à la place du gène ciblé. De plus, dans le système de Datsenko et Wanner, les cassettes antibiotiques sont flanquées de sites FRT, ce qui permet leur élimination ultérieure par recombinaison site spécifique (Datsenko and Wanner, 2000). L'étude de Zhang et al, en 1998, met en évidence que RecT est aussi capable d'intégrer un gène dans un épisome, grâce à la présence d'un plasmide qui code pour RecT, RecE et Gam (Zhang et al., 1998). La présence de Gam est nécessaire pour éviter la dégradation du produit de PCR par RecBCD. Cette étude met en

évidence que la région d'homologie doit être d'au moins 42 nt, et que la fréquence de recombinaison augmente avec la taille de la région homologue jusqu'à 60 nt (Zhang et al., 1998).

Dans ce type d'expérience, le fragment PCR ayant à recombinaison avec une molécule circulaire (le chromosome), il était implicitement supposé que la réaction se produisait par un mécanisme d'invasion de brin. Cependant, un travail récent suggère qu'en contexte *recA* la recombinaison a lieu par appariement d'ADN simple brin au niveau de la fourche de réplication du chromosome (Maresca et al., 2010), d'une manière similaire à l'intégration d'oligonucléotides décrite au chapitre précédent. En effet, ce travail montre d'une part que de grandes molécules simple brin (500 nt) peuvent faire de l'appariement simple brin *in vivo*, avec une préférence pour l'appariement avec la matrice du brin tardif, comme pour les oligonucléotides. D'autre part, lorsque les produits PCR ont une extrémité 5' phosphorylée, celle-ci est préférentiellement dégradée par Red α , ce qui conduit *in vivo* à la présence de quelques molécules simple brin de taille intacte, car dégradées uniquement par une extrémité. Enfin, l'argument le plus fort de ce travail vient de l'observation que l'efficacité du recombineering est fortement limitée par la longueur du fragment à introduire : au-delà d'1 kb, cette efficacité chute. Ceci n'est pas attendu si la recombinaison a lieu par invasion de brin. Ce travail conduit donc à reconsidérer les études précédemment citées de Myrner et al. et Zhang et al. : si la molécule linéaire introduite, sensée faire de l'invasion de brin, est convertie entièrement en simple brin, elle pourrait s'introduire dans le plasmide test par appariement simple brin.

Cependant, d'autres travaux suggèrent que Red β fait de l'invasion de brin *in vivo*. En effet, Martinsohn et al., en 2008, ont mis en évidence que Red β permet de la recombinaison entre deux gènes *oxa* insérés en orientation inverse de part et d'autre du promoteur P_L dans Lambda (Martinsohn et al., 2008). Ainsi, lorsque les deux gènes *oxa* recombinent, il y a une inversion du promoteur P_L et un changement phénotypique qui permet de détecter et de compter les recombinants. Cette

recombinaison peut se produire par invasion de brin (en une seule étape), ou par appariement d'ADN simple brin (en plusieurs étapes). Elle a lieu même lorsque les deux séquences homologues ont 22% de divergence. Dans ce cas, il est possible d'analyser le point précis où s'est produit l'échange de brin. Or il a été observé que le point d'échange est symétrique (indiquant une réaction d'échange de brin avec résolution symétrique par RuvABC) lorsque la recombinaison dépend de RecA, ce qui était attendu, mais également, dans une plus faible proportion (20%) lorsque la recombinaison a lieu avec Red β . Ceci suggère que Red β réalise une certaine proportion d'événements d'échange de brin symétriques, difficilement réconciliables avec une succession d'événements d'appariement d'ADN simple brin.

Enfin, deux études rapportent que lors de la RH *in vivo* impliquant une molécule circulaire et une molécule linéaire, en présence de Red β , la molécule circulaire doit se répliquer pour recombiner avec l'ADN linéaire (Maresca et al., 2010; Poteete, 2008). Dans l'étude de Poteete en 2008, il est mis en évidence qu'un bactériophage Lambda linéarisé recombine efficacement avec un plasmide circulaire réplicatif. Par contre si de l'azidothymidine, qui bloque la réplication du plasmide, est ajoutée, la recombinaison est abolie. L'étude de Maresca et al, de 2010, mentionnée ci-dessus, montre également que la recombinaison entre une molécule d'ADN double brin circulaire et un ADN double brin linéaire n'est possible que lorsque la molécule circulaire se réplique. A nouveau, ces deux observations favorisent l'hypothèse d'une recombinaison qui se produit majoritairement par appariement d'ADN simple brin, au niveau d'une fourche de réplication. En effet, dans le cas de recombinaison par invasion de brin, l'étape de réplication du substrat n'est pas nécessaire.

En conclusion, l'ensemble des travaux sur ce sujet conduit aujourd'hui à considérer que Red β fait en priorité de l'appariement d'ADN simple brin *in vivo*, que ce soit entre deux extrémités cassées rendues simple brin et complémentaires, ou au niveau d'une fourche de réplication. Elle semble également capable de produire des

recombinants par invasion de brin, à plus faible efficacité, et dans des conditions biochimiques plus étroites que RecA.

Chapitre IV : Red β est-elle un cofacteur de RecA?

En plus de ses deux activités d'appariement d'ADN simple brin et d'invasion de brin, il semble que Red β soit également capable, dans certaines conditions, d'agir comme cofacteur de RecA. Ceci la rapproche de la protéine Rad52 de la levure, dont la fonction principale consiste à préparer les substrats de recombinaison pour Rad51. Nous commencerons donc ce chapitre par un bref rappel des fonctions connues de Rad52 chez la levure et chez l'homme, avant de présenter les quelques travaux qui suggèrent une activité cofacteur des recombinases phagiques pour RecA.

I. Rad52 chez les eucaryotes

De même qu'un ensemble de travaux a eu pour objectif de comparer Red β ou RecT à RecA, de nombreuses études ont recherché la similarité fonctionnelle des recombinases phagiques avec Rad52. Cela s'explique par l'observation qu'un ensemble de recombinases phagiques (et en particulier la protéine Sak du bactériophage *ul36* infectant *Lactococcus lactis*) a une certaine homologie de séquence avec Rad52 (Iyer et al, 2002).

La protéine Rad52 eucaryote est constituée d'environ 400 résidus, contre seulement 250 pour ses analogues bactériophagiques. L'homologie entre les protéines se situe dans la région N-ter de Rad52. Le C-ter de Rad52 n'est pas structuré, il contient cependant des régions impliquées dans l'interaction avec Rad51 (Singleton et al., 2002), et avec RPA (Replication protein-A) (SSB eucaryote). La structure tridimensionnelle a été obtenue pour le domaine N-ter de la protéine de Rad52 humaine (Kagawa et al., 2002; Singleton et al., 2002). Tout comme Red β , RecT et Sak, Rad52 adopte une structure quaternaire en forme d'anneau en absence d'ADN, constitué de 7 monomères pour la protéine pleine taille (Shinohara and Ogawa,

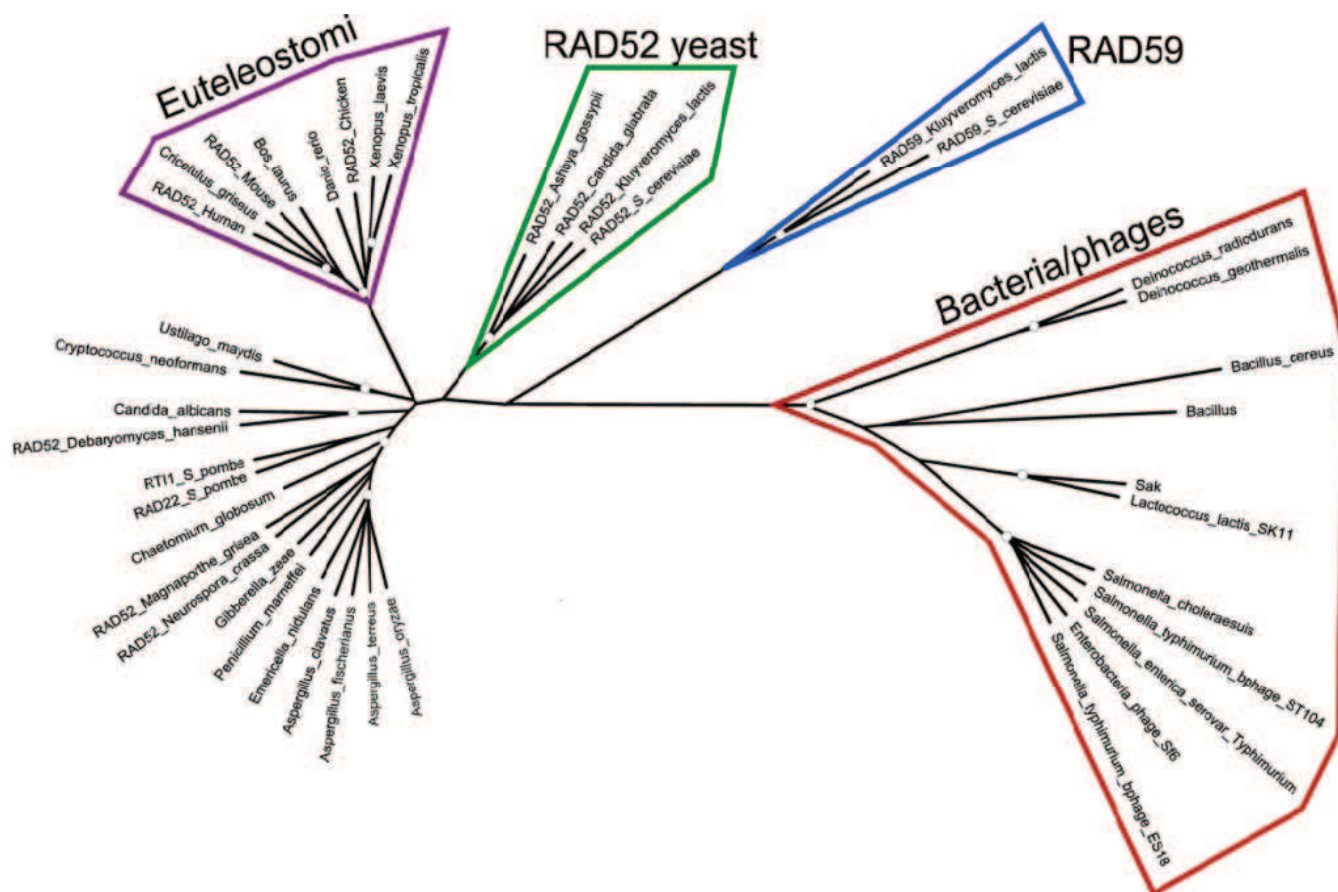


Figure 10: Arbre phylogénétique du domaine Pfam Rad52_Rad22.
Tiré de Ploquin et al., 2008.

1998) (Stasiak et al., 2000), et de 11 pour le domaine N-ter (Kagawa et al., 2002; Singleton et al., 2002).

Chez la levure, Rad52 a deux activités biochimiques principales : d'une part elle permet l'appariement d'ADN simple brin (Mortensen et al., 1996) et d'autre part elle est une RMP de Rad51, qui effectue la RH par invasion de brin. De manière intéressante, les phénotypes des mutants *rad51* et *rad52* ne se recouvrent pas entièrement : le mutant *rad52* est davantage sensible aux radiations ionisantes et aux UV que le mutant *rad51* (Rattray and Symington, 1994). Ceci suggère que Rad52 accomplit réellement deux fonctions distinctes dans la cellule, une liée à Rad51, très probablement sa fonction de médiatrice, et une autre indépendante de Rad51, probablement en lien avec son activité d'appariement d'ADN simple brin. Cependant, on ne sait pas encore comment et quand intervient cette activité d'appariement d'ADN simple brin *in vivo*.

La distribution de Rad52 est sporadique chez les eucaryotes (Figure 10, tiré de Ploquin et al., 2008) : selon les données actuelles, elle est uniquement présente chez les champignons, et chez les vertébrés eutéleostèmes (du Xénope à l'Humain). La protéine Brca2 (« Breast cancer 2 ») est toujours présente chez les espèces dépourvues de Rad52, et possède les mêmes fonctions que Rad52. Il se pourrait donc qu'un remplacement fonctionnel ait eu lieu, la disparition de Rad52 s'accompagnant de l'apparition de Brca2. Cependant, deux études récentes démontrent que ce remplacement n'est pas total, car Brca2 et Rad52 sont toutes deux présentes chez l'Humain. Brca2 joue un rôle de premier plan, étant donné le phénotype cancéreux associé à la perte de fonction. Elle possède une très forte activité médiatrice vis-à-vis de Rad51, étant capable de déloger RPA des substrats simple brin, pour permettre le chargement de Rad51 (Jensen et al., 2010). Cependant, la Brca2 humaine n'a aucune activité d'appariement d'ADN simple brin, contrairement aux Brca2 d'*Ustilago maydis* (charbon du maïs), et de *Cenorhabditis elegans*. Contrairement à Brca2, aucune maladie jusqu'à présent n'est associée à la

perte de fonction de la Rad52 humaine (hRad52). Cette protéine est capable d'apparier de l'ADN simple brin *in vitro*, mais n'a pas d'activité médiatrice de Rad51. Ceci suggère une répartition de fonction entre ces deux protéines. Une fonction de Rad52 a pu être proposée récemment. Des expériences sur cultures de cellules humaines ont montré que l'inactivation par ARN interférence de Rad52 dans des cellules tumorales mutées pour Brca2 supprime leur prolifération. Ce phénotype serait dû à une forte instabilité chromosomique, due à la suppression de l'activité de recombinaison homologue résiduelle de Rad51. Les auteurs en concluent qu'il y a une relation de létalité synthétique entre ces deux gènes, où Brca2 joue le rôle prééminent, tandis que celui de hRad52 est mineur, et se dévoile en contexte *brca2*. Il reste à comprendre si ce schéma s'applique dans toutes les espèces équipées de Brca2 et Rad52. Cela semble déjà ne pas être le cas chez le champignon *U. maydis*, qui possède une Brca2 ayant toutes les activités de Rad52, et pourtant aussi le gène *rad52*.

En résumé, les Rad52 eucaryotes peuvent avoir une ou deux fonctions : la fonction permanente est l'appariement d'ADN simple brin, tandis que la fonction médiatrice peut avoir été perdue (chez l'Humain).

II. Les recombinaisons phagiques de la superfamille Rad52 ayant une activité cofacteur de RecA.

1. Red β est un cofacteur de RecA lorsque Lambda ne peut pas se répliquer

Nous avons vu précédemment que la recombinaison entre deux bactériophages Lambda dépend de la recombinaison phagique, et qu'elle est indépendante de RecA (Brooks and Clark, 1967; Franklin, 1967). Cependant, lorsque la réplication du bactériophage est bloquée, RecA devient nécessaire à la recombinaison. Dans ce cas, on peut se demander si la recombinaison fonctionne avec des fonctions

« purement bactériennes », ou si Red β continue à jouer un rôle, qui serait alors celui d'un cofacteur, au même titre que RecBCD ou RecFOR. Une étude de Stahl et al, en 1997, montre que Red β a ce rôle de cofacteur : cette étude utilise deux bactériophages Lambda parentaux, qui ne peuvent pas se répliquer et dont la transcription des gènes est inhibée (Stahl et al., 1997). Ils ne recombinent pas entre eux en présence de RecA seul. Par contre, en présence de RecA et de Red β , ils recombinent et le taux de recombinants augmente lorsque Gam est aussi présent. Lorsque la même expérience est réalisée dans un mutant *recA*, les auteurs n'observent que très peu de recombinants en présence de Red β et Gam. Donc la présence de RecA, Red β et Gam favorise la formation de recombinants. Une seconde expérience permet de préciser le rôle de Red β dans cette recombinaison. En effet, lorsque l'un des deux bactériophages parentaux est coupé, les bactériophages ne peuvent recombiner qu'en présence de Red β et de RecA. Ceci suggère que Red β stimule l'invasion de brin RecA-dépendante. Enfin, lorsque les deux bactériophages parentaux sont coupés à des sites non-alléliques (situation favorisant la réparation par appariement d'ADN simple brin), les recombinants sont formés en présence de Red β , que RecA soit présente ou non. Ceci confirme que Red β seul ou en présence de RecA est capable de faire de l'appariement d'ADN simple brin (Stahl et al., 1997). En conclusion, lorsque deux bactériophages ne peuvent pas se répliquer, Red β n'est capable de faire de l'invasion de brin qu'en présence de RecA, alors qu'elle est capable de faire de l'appariement d'ADN simple brin seule.

2. Le recombineering Red β -dépendant est diminué dans un mutant *recA*

Dans une des études mettant en évidence l'intérêt des gènes *gam*, *bet* et *exo* pour le recombineering, les auteurs ont observé que l'efficacité de remplacement de gène est diminuée d'un facteur 68 dans une souche mutée pour le gène *recA* ($8,1 \times 10^{-7}$

recombinants par cellule viable dans une souche mutée pour *recA* contre $5,5 \times 10^{-5}$ dans une souche sauvage) (Murphy, 1998). L'efficacité de recombinaison est au contraire augmentée d'un facteur 12 dans une souche mutée pour l'exonucléase RecJ, spécifique de l'ADN simple brin ($7,0 \times 10^{-4}$ recombinants par cellule viable) (Murphy, 1998). Murphy a également construit une souche d'*E. coli*, où les gènes *recBCD* ont été remplacés par les gènes *exo* et *bet* placés sous le contrôle du promoteur *P_{lac}*. Il a observé des niveaux élevés de recombinaison entre le chromosome et un fragment d'ADN linéaire. Là encore, une mutation dans le gène *recJ* stimule cette recombinaison (Murphy, 1998). De plus, la même expérience dans une souche mutée pour *recA* a été réalisée par Poteete et Fenton et ils en concluent aussi que la recombinaison est diminuée d'un facteur 100 dans le mutant *recA* (Poteete and Fenton, 2000). Ces résultats suggèrent donc qu'en contexte RecA⁺, une majorité d'événements de recombineering ferait intervenir RecA, et donc ne se produirait pas par appariement d'ADN simple brin, comme proposé par Maresca et al., dont toute l'étude se déroule en contexte *recA*.

3. Sak et Sak3, deux recombinases de bactériophages infectant des bactéries lactiques, sont des cofacteurs de RecA *in vitro*.

La protéine Sak du bactériophage ϕ 36 infectant *L. lactis* est homologue au N-ter de la protéine RAD52 humaine (Iyer et al., 2002). Ploquin et al. ont purifié cette protéine Sak et ont montré qu'elle avait une affinité préférentielle pour l'ADN simple brin par rapport à l'ADN double brin. Sak permet l'appariement simple brin de deux ADN complémentaires de 400 nt. De plus, en présence de RecA, Sak est capable de faire de l'échange de brin de manière plus efficace que lorsque RecA est seule. Ils ont conclu de cette étude que Sak est capable de faire de l'appariement d'ADN simple brin *in vitro* et qu'elle stimule la recombinaison RecA-dépendante (Ploquin et al., 2008). Une autre étude a montré que seul le domaine N-ter de la protéine Sak est impliqué dans l'appariement d'ADN simple brin, et que le domaine N-ter et C-ter sont

tous les deux nécessaires au mécanisme d'échange de brin stimulé par RecA (Scaltriti et al., 2011). La protéine Sak3 du bactériophage p2 infectant *L. lactis* est une recombinase de la superfamille Rad52 (Lopes et al., 2010). Sak3 a été purifiée, elle a une affinité préférentielle pour l'ADN simple brin par rapport à l'ADN double brin tout comme Sak et Rad52 (Scaltriti et al., 2010). Sak3 est capable d'apparier deux ADN simple brin complémentaires de 63 nt, cependant elle n'est pas capable d'apparier deux ADN complémentaires de 400 nt comme Sak. Seul le domaine N-ter de la protéine est impliqué dans l'appariement d'ADN simple brin. De plus, Sak3 est capable de faire de l'invasion de brin *in vitro* seulement en présence de RecA (Scaltriti et al., 2011).

En résumé, les recombinases Sak et Sak3 sont capables d'apparier de l'ADN simple brin de manière RecA indépendante. De plus, ces recombinases sont des cofacteurs de RecA dans des expériences d'invasion de brin *in vitro*.

Chapitre V : Synthèse, formulation de la question poursuivie lors de la thèse

En conclusion, il ressort de cette analyse bibliographique que les deux recombinaises phagiques les mieux caractérisées ont un comportement assez distinct : UvsX semble fonctionner comme RecA et promouvoir de la recombinaison homologe par invasion de brin, tandis que Red β aurait en priorité une activité d'appariement d'ADN simple brin, mais serait également capable de fonctionner comme cofacteur de RecA dans des réactions d'invasion de brin, et enfin, elle serait capable à faible fréquence de favoriser à elle seule des réactions d'invasion de brin.

L'objectif de mon projet de thèse a été d'étudier le rôle des recombinaises phagiques dans l'évolution des génomes de bactériophages. Pour reprendre la question soulevée en commençant, nous avons formulé l'hypothèse que ces recombinaises pourraient avoir en commun une ou plusieurs facultés les distinguant de RecA, telle qu'une fidélité relâchée lors de la réaction de recherche d'homologie, qui pourrait expliquer en partie la très grande plasticité des génomes de bactériophages.

Une première étape du travail a consisté à identifier de nouvelles recombinaises phagiques par des méthodes bio-informatiques et à confirmer leur activité de recombinaison par la mise au point d'un système expérimental d'appariement d'ADN simple brin *in vivo* (Chapitre I).

Dans un deuxième temps, nous avons étudié plus précisément les nouvelles recombinaises que nous avons détectées, en nous intéressant en particulier aux recombinaises de la superfamille Rad51/RecA. Nous avons comparé la recombinaise Sak4 à UvsX *in vivo*, et testé la capacité de Sak4 à appairer de l'ADN simple brin de séquence divergée. Et enfin, nous avons étudié l'importance des cofacteurs de Red β pour l'appariement d'ADN simple brin (Chapitre II).

Table 1 – Souches d'*E. coli*, et plasmides utilisés lors de ce travail.

n°souche	génotype	source
MAC1022	AB1157 nal ^R strep ^R	S. Delmas, Necker
MAC1024	AB1157 nal ^R strep ^R mutS::spec ^R	S. Delmas, Necker
MAC1031	AB1157 nal ^R strep ^R recA306 srl::tn10	ce travail
MAC1034	AB1157 nal ^R strep ^R recA306 srl tn10 mutS::spec ^R	" "
JAC214	AB1157 nal ^R strep ^R recO:tn5 recA306 srl::tn10	" "
MAC1042	MAC1031 pKD46	" "
MAC1525	MAC1031 pJA21	" "
MAC1528	MAC1031 pJA6	" "
MAC1045	MAC1034 pKD46	" "
MAC1544	MAC1031 pJA12	" "
JAC157	MAC1031 pJA4	" "
JAC160	MAC1022 pJA4	" "
JAC171	MAC10231 pJA13	" "
JAC173	MAC10231 pJA9	" "
JAC 180	MAC1031 pJA7	" "
JAC182	MAC1031 pJA17	" "
JAC186	MAC1031 pJA10	" "
JAC193	MAC1031 pJA5	" "
JAC194	MAC1031 pJA14	" "
JAC195	MAC1031 pJA16	" "
JAC 206	MAC1031 pJA11	" "
JAC 205	MAC1031 pJA8	" "
JAC 210	MAC1031 pJA15	" "
JAC 239	MAC1031 pJA18	" "
JAC 236	MAC1031 pJA19	" "
JAC 237	MAC1031 pJA20	" "
JAC212	MAC1034 pJA15	" "
nom		
plasmide	contenu	source
pKD46	oriSC101 ts Amp ^R araC ParaBA-gam-red β -red α -orf60a	Datsenko et Wanner (2000)
pJA4	oriSC101 ts Amp ^R araC ParaBAD	ce travail
pJA5	pJA4 recA	" "
pJA6	pJA4 uvsX	" "
pJA7	pJA4 uvsX K66A	" "
pJA8	pJA4 uvsX Δ Cter	" "
pJA9	pJA4 uvsX K66A W309L	" "
pJA10	pJA4 uvsX W309L	" "
pJA11	pJA4 uvsX K66A Δ Cter	" "
pJA12	pJA4 sak _{PA73}	" "
pJA13	pJA4 sak _{PA73} K30A	" "
pJA14	pJA4 sak _{PA73} K123A	" "
pJA15	pJA4 sak _{HK620} abc2 hkaK hkaM	" "
pJA16	pJA4 sak _{PA73} K123E	" "
pJA17	pJA4 red β	" "
pJA18	pJA4 red β red α orf60a	" "
pJA19	pJA4 gam red β red α	" "
pJA20	pJA4 gam red β	" "
pJA21	pJA4 gp2.5 _{T7}	" "

MATERIEL ET METHODES

I. Souches et plasmides

1. Souches

Les souches utilisées sont toutes dérivées d'*E. coli* AB1157 *nal^R strep^R recA306 srt::Tn 10* ou *recA306 srt::Tn 10 mutS ::spec^R* (Table 1). Pour exclure la possibilité que les effets observés soient dus à l'activité de médiateur des recombinaisons phagiques, toutes les expériences ont été réalisées dans une souche déficiente pour le gène *recA*, étant donné que Red β peut dans certains cas agir comme cofacteur de RecA (voir Introduction). Les expériences d'appariement d'ADN simple brin identique, à part pour une mutation créant un mésappariement C-C, ont été réalisées dans une souche dont le système MMR est actif, tandis que les expériences d'appariement d'ADN simple brin non identique ont été réalisées dans une souche dont le système MMR est inactif. Ainsi, les mésappariements ne sont pas reconnus et ne sont pas réparés.

2. Plasmides

Les plasmides construits sont tous dérivés de pKD46, lui-même dérivé de pSC101. Les plasmides utilisés sont listés dans la Table 1. Le plasmide pKD46 possède une origine de réplication thermosensible à bas nombre de copie et il porte un gène conférant la résistance à l'ampicilline. pKD46 code pour les gènes *gam*, *bet*, *exo* et *orf60a* du bactériophage Lambda sous le contrôle du promoteur *ParaBAD* dont la transcription est induite par AraC en présence d'arabinose (Datsenko and Wanner, 2000). Un dérivé de pKD46 délété pour la région *gam*, *bet*, *exo* et *orf60a*, pJA4 est utilisé comme vecteur pour cloner les différents gènes à tester. Les recombinaisons à tester sont toutes clonées avec leurs RBS (Ribosome binding site) et certaines ont été mutées par mutagenèse dirigée. Pour cela, une mutation ponctuelle est intégrée dans le gène par PCR. Dans ce but, deux fragments correspondant au début et la fin

du gène sont amplifiés avec deux couples d'oligonucléotides, dont l'un est centré sur la mutation que l'on veut introduire et l'autre est à l'extérieur du gène. Puis, la totalité du gène muté est amplifié, en utilisant comme matrice les deux premiers produits de PCR, et les deux oligonucléotides extérieurs.

Les oligonucléotides ayant servi aux clonages des recombinaisons ont été commandés chez Eurogentec. Toutes les enzymes de restriction proviennent de chez Biolabs et les PCR ont été réalisées avec l'ADN polymérase Phusion « High-Fidelity DNA Polymerase » de chez Finnzymes.

Toutes les souches utilisées sont cultivées à 30°C, en milieu LB supplémenté de 0,025 % de thymine (LBT) et de 50 µg/mL d'ampicilline.

Toutes les constructions ont été vérifiées par séquençage.

II. Préparation des cellules électro-compétentes

Les souches d'*E. coli* électro-compétentes ont été préparées selon le protocole Biorad, avec les modifications suivantes : les souches étant thermosensibles, elles sont cultivées à 30°C, en milieu SLB (5 g/L NaCl, 20 g/L extrait de levure, 35 g/L bacto-tryptone et 1 ml/L NaOH 5N). Les cellules sont cultivées jusqu'en phase exponentielle de croissance (DO₆₀₀ de 0,4), puis 0,2% d'arabinose final est ajouté pour induire les diverses recombinaisons, la croissance est stoppée à une DO₆₀₀ de 0,8 par incubation dans la glace pendant 30 minutes. Ensuite on procède à deux lavages avec de l'eau froide et enfin un dernier au glycérol 10% froid. Les cellules sont resuspendues dans du glycérol 10% froid et congelées rapidement à -80°C. Le niveau de compétence de ces cellules est testé avec 10 ng de plasmide pACYC184. Tous les résultats rapportés ont été obtenus avec des cellules ayant un niveau de compétence compris entre 10⁶ et 10⁷ transformants par microgramme d'ADN.

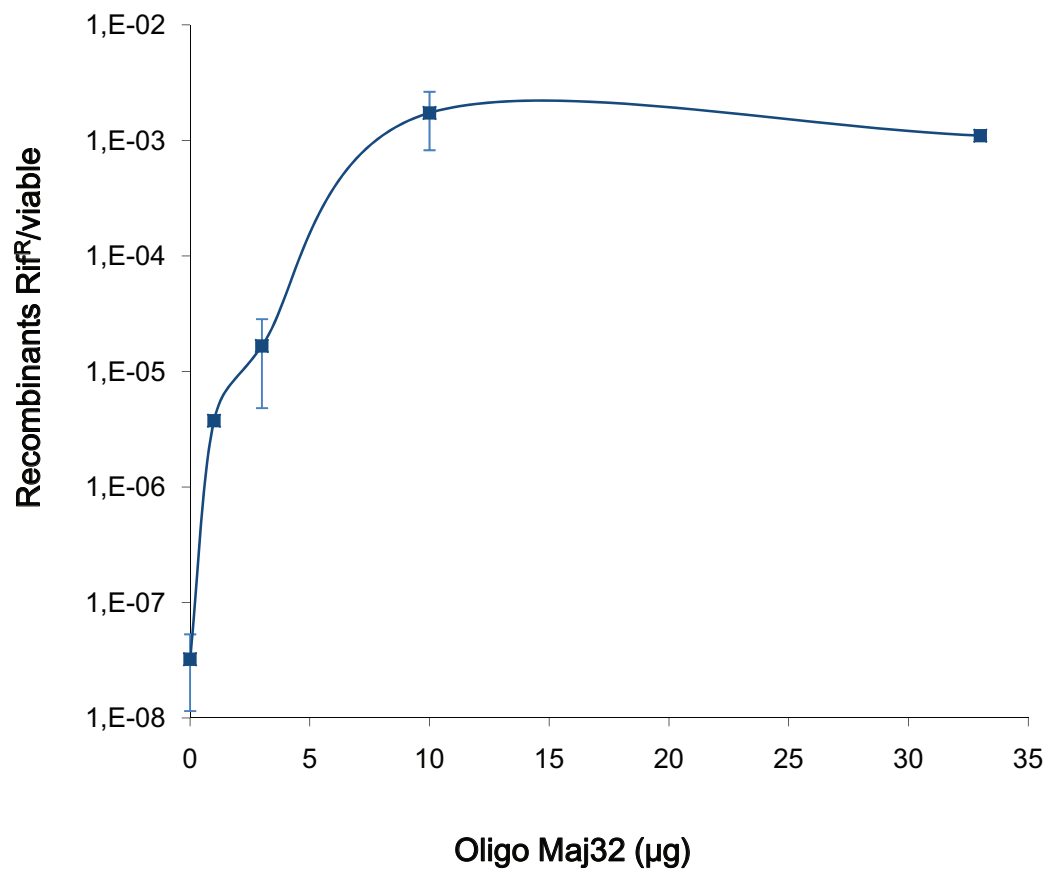


Figure 11: Effet de la quantité en oligonucléotide sur la fréquence de recombinaison.

Test d'appariement d'ADN simple brin *in vivo* par Red β (pKD46), en fonction de la quantité en oligonucléotide Maj32 (51 nt).

III. Le système expérimental d'appariement d'ADN simple brin *in vivo*.

1. Description.

Le système expérimental d'appariement d'ADN simple brin *in vivo* est le même que celui décrit dans le chapitre I de la partie résultat. Brièvement, il est basé sur l'oligonucléotide Maj32 qui comprend 51 nucléotides, centrés sur la position 1576, à partir de l'ATG du gène *rpoB* de la souche MG1655 d'*E. coli* et il est complémentaire au brin tardif. La base 1576 du gène *rpoB* sauvage est un G, alors que l'oligonucléotide contient un C, ainsi le codon CAC du brin codant devient un codon GAC et la protéine RpoB a une mutation H526D qui confère la résistance à la rifampicine (Wolff et al., 2004). De plus, cette mutation produit un mésappariement C-C qui n'est pas reconnu par le MMR (Modrich, 1991). Les souches électro-compétentes, contenant le plasmide qui exprime la recombinaise à tester, sont mélangées avec l'oligonucléotide Maj32, transformées par électroporation, et mises à agiter, pendant une heure, à 30°C, en milieu SOC supplémenté de 0,2% de glucose. Les souches transformées sont ensuite incubées toute la nuit à température ambiante sans agitation et enfin, étalées sur boîte de pétri LB supplémenté de 50 µg/mL de rifampicine et incubées à 37°C.

2. La quantité d'oligonucléotide.

Afin de déterminer la meilleure quantité d'oligonucléotide nécessaire pour les tests d'appariement d'ADN simple brin, des transformations de la souche AB1157 nal^R strep^R *recA306 srl::Tn10* pKD46, ont été réalisées avec différentes quantités de l'oligonucléotide Maj32 (0, 1, 3, 10 et 33 µg) (Figure 11). Les meilleures fréquences de recombinants résistants à la rifampicine (Rif^R) sont obtenues avec 10 ou 33 µg d'oligonucléotide. Cependant, au dessus de 10 µg d'oligonucléotide la fréquence de

Table 2 – Séquence des oligonucléotides utilisés dans les tests d'appariement d'ADN simple brin.

En rouge, le C crée un mésappariement C-C qui n'est pas reconnu par MutS. En vert, les nt échangés pour créer des mutations neutres réparties sur toute la longueur de la séquence, afin d'obtenir divers niveaux de divergence.

Nom oligo	Séquence	Taille (nt)	Divergence (%)
Maj32	CCGAGTGC GGAGATACGACGTTTGT C CGTAATCTCAGACAGCGGGTTGTTTC	51	2
Maj32_41	TGCGGAGATACGACGTTTGT C CGTAATCTCAGACAGCGGGT	41	2
Maj32_61	CTGGGCCGAGTGC GGAGATACGACGTTTGT C CGTAATCTCAGACAGCGGGTTGTTCTGGTC	61	2
Maj32_71	ACCGCCTGGGCCGAGTGC GGAGATACGACGTTTGT C CGTAATCTCAGACAGCGGGTTGTTCTGGTCCATAA	71	1
Maj32_81	GTCAGACCGCCTGGGCCGAGTGC GGAGATACGACGTTTGT C CGTAATCTCAGACAGCGGGTTGTTCTGGTCCATAAACTGA	81	1
Maj33_51_8	CCGAG C GCGGAGATACGACG C TTGT C CGTAATCTCAGAAAGCGGGTTGTTTC	51	8
Maj33_51_14	CC C AGTGC G GA A ATACGACG C TTGT C CGT G ATCTCAGAAAGCGGGTT A TTTC	51	14
Maj33_51_22	CC C AG C GCGGA A AT G CGACG C TTGT C CGT G AT T TCAGAAAG T GGGTT A TTTC	51	22
Maj33_81_9	GT G AGACCGCCTGGGCC C AGTGC G GA A ATACGACGTTTGT C CGTAATCTCAGAAAGCGGGTT A TTCTGGTCCATAAA T TGA	81	9
Maj33_84_13	GGT G AGACCGCCTGG C CC C AGTGC G GA A ATACGACG C TTGT C CGT G ATCTCAGAAAGCGGGTT A TTCTGA T CCATAAA T TGAGA	84	13
Maj63	GAAACCCCTGAAGGTCCGAACATCG C TCTGATCAACTCTCTGTCCGTGTAC	51	2

Table 3 – Activité d'appariement d'ADN simple brin *in vivo* testé pour pKD46 (Redβ en présence de ses cofacteurs) avec différents oligonucléotides.

<i>Souche</i>	<i>Recombinase</i>	<i>Oligo</i>	<i>Ri^R/viable</i>	<i>Nb exp</i>
MAC1042	Redβ	Maj32	1,7 (± 0,9) 10 ⁻³	6
		Maj32_71	4,9 (± 0,7) 10 ⁻³	3
		Maj63	9,2 (± 0,7) 10 ⁻⁶	2

recombinants Rif^R reste stable (Figure 11). Tous les tests d'appariement ont donc été réalisés avec 10 µg d'oligonucléotides.

3. La taille de l'oligonucléotide.

Ellis et al. ont montré qu'un oligonucléotide de 40 à 60 nt s'apparie au gène *galk* à une fréquence stable ($4\text{-}5 \times 10^4$ recombinants Gal⁺ pour 10^8 cellules viables) et que la fréquence la plus élevée est obtenue avec un oligonucléotide de 70 nt ($2,2 \times 10^5$ recombinants Gal⁺ pour 10^8 cellules viables) (Ellis et al., 2001). Nous avons donc réalisé un test d'appariement d'ADN simple brin, en présence du plasmide pKD46 avec l'oligonucléotide Maj32 de 51 nt et avec l'oligonucléotide Maj32_71 de 71nt. (Table 2). Les fréquences de recombinaison obtenues sont respectivement $1,7 \times 10^{-3}$ et $4,9 \times 10^{-3}$ Rif^R par cellule viable (Table 3). Cette différence n'étant pas significative, les tests d'appariement d'ADN simple brin ont été réalisés avec l'oligonucléotide Maj32 de 51 nt.

4. Meilleure efficacité de recombinaison avec un oligonucléotide complémentaire au brin tardif.

Dans une étude d'appariement d'ADN simple brin au gène *galk*, la fréquence d'appariement d'un ADN simple brin complémentaire au brin tardif est 10 fois supérieure par rapport à celle obtenue lorsque l'ADN simple brin est complémentaire du brin précoce (Datta et al., 2008). Nous avons voulu tester si cette observation était reproductible dans notre système expérimental. Pour cela, l'oligonucléotide Maj63 a été comparé à Maj32, l'oligonucléotide de référence : Maj63 possède aussi 51 bases et un mésappariement C-C centré sur la position 1709 du gène *rpoB* (Table 2). La fréquence de recombinants obtenue en présence de pKD46 avec l'oligonucléotide Maj63, qui est complémentaire au brin précoce, est 180 fois inférieure ($9,2 \times 10^{-6}$ Rif^R

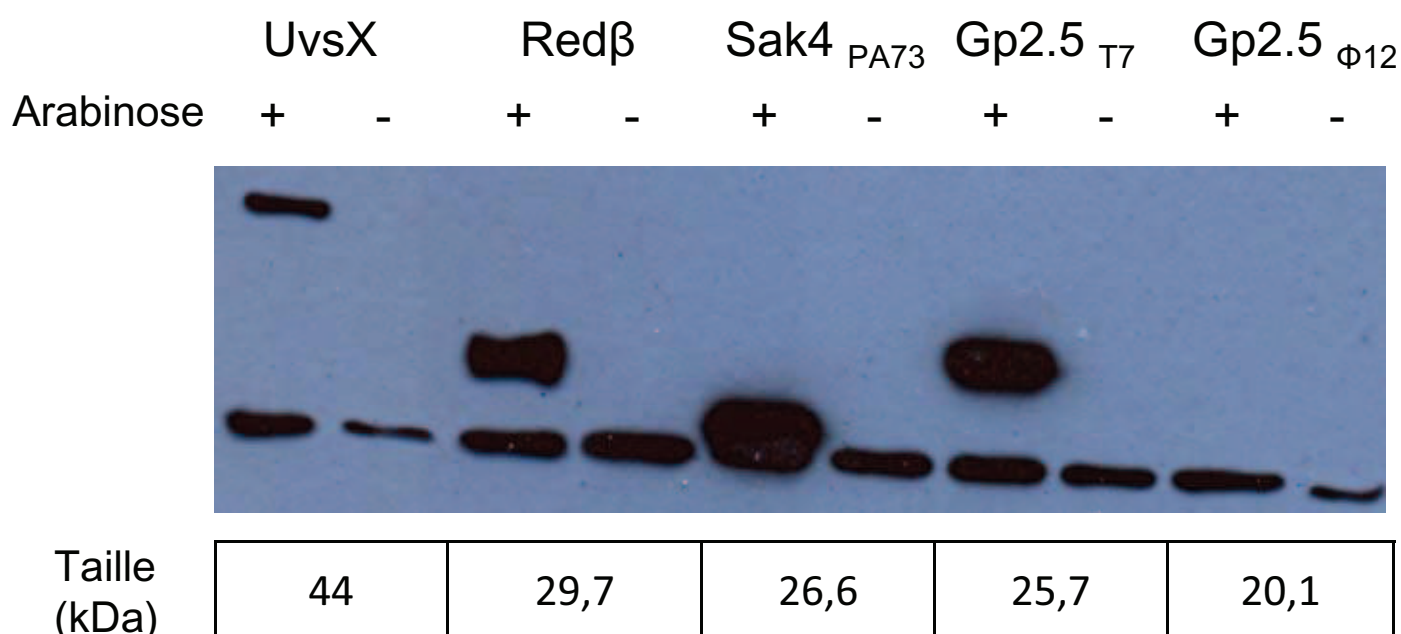


Figure 12: Expression des recombinaises par Western Blot.

Les recombinaises UvsX du bactériophage T4, Red β du bactériophage Lambda, Sak4 du bactériophage PA73, Gp2.5 du bactériophage T7 et Gp2.5 du bactériophage Φ 12 ont été étiquetées avec un Flag (séquence DYKDDDDK) en N-ter, reconnue par un Anticorps primaire Anti-FLAG de souris. Une extraction totale des protéines avant et après induction à l'arabinose a été effectuée, déposée sur gel acrylamide 12%, transférée puis révélées par l'anticorps secondaire couplé à la peroxydase qui reconnaît l'anticorps primaire. Enfin, l'activité peroxydase est révélée et analysée par Western Blot.

par cellule viable) par rapport à la fréquence obtenue avec l'oligonucléotide Maj32, complémentaire au brin tardif ($1,7 \times 10^{-3}$ Rif^R par cellule viable) (Table 3). Nous concluons qu'il existe un fort effet de biais en faveur de l'intégration à la fourche de réplication, du côté de la synthèse du brin tardif. L'effet est même plus prononcé que dans la publication de Datta et al. Ceci pourrait être lié au niveau plus faible d'expression de Red β dans notre système, comparé à celui de Datta et al (basé sur P_L), qui accentuerait le contraste.

Tous les tests d'appariement d'ADN simple brin présentés dans la partie Résultats ont donc été réalisés avec 10 μ g d'oligonucléotide Maj32 ou ses dérivés, complémentaires du brin tardif (Table 2). Le nombre de test d'appariement d'ADN simple brin, la moyenne et les écarts types des fréquences de Rif^R par viable sont présentés dans les tables 3 à 9. Un test de Student unilatéral a été utilisé pour déterminer le caractère significatif des écarts observés avec la souche contrôle contenant le plasmide vecteur qui n'exprime aucune recombinaise (pJA4). Dans le cas des essais réalisés dans le fond génétique *mutS*, les comparaisons ont été effectuées par rapport au taux de mutants spontanés résistants à la rifampicine.

Tous les oligonucléotides ayant servi aux tests d'appariement d'ADN simple brin ont été commandés chez Eurogentec, dessalés, purifiés et resuspendus dans de l'eau à une concentration de 1 mg/ml. Ils sont listés dans la Table 2.

5. Marquages

1. L'expression des recombinaises phagiques testées par Western Blot.

La comparaison des fréquences d'appariement d'ADN simple brin *in vivo* ne peut s'effectuer que si l'on s'assure auparavant que les protéines sont bien produites à

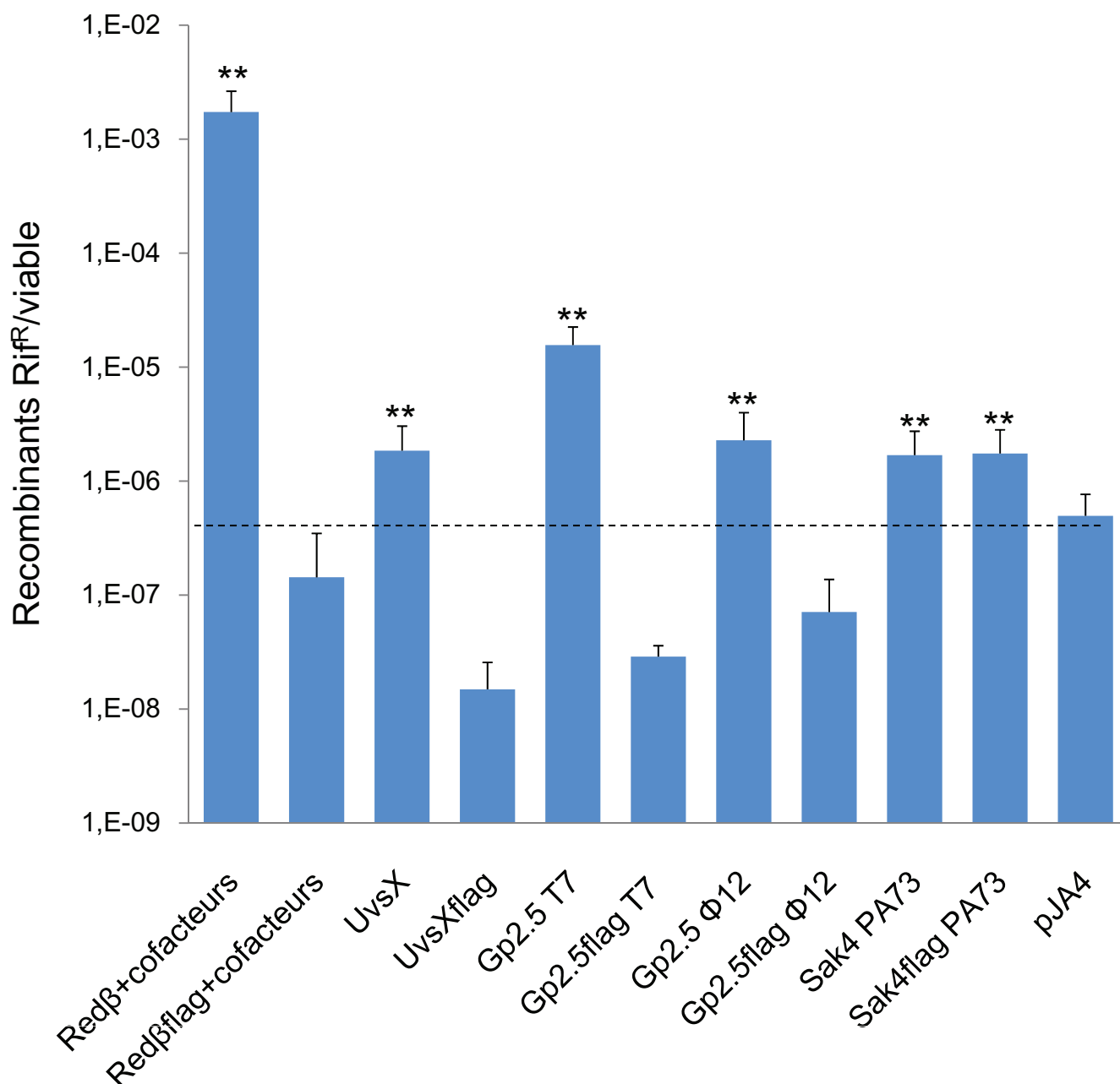


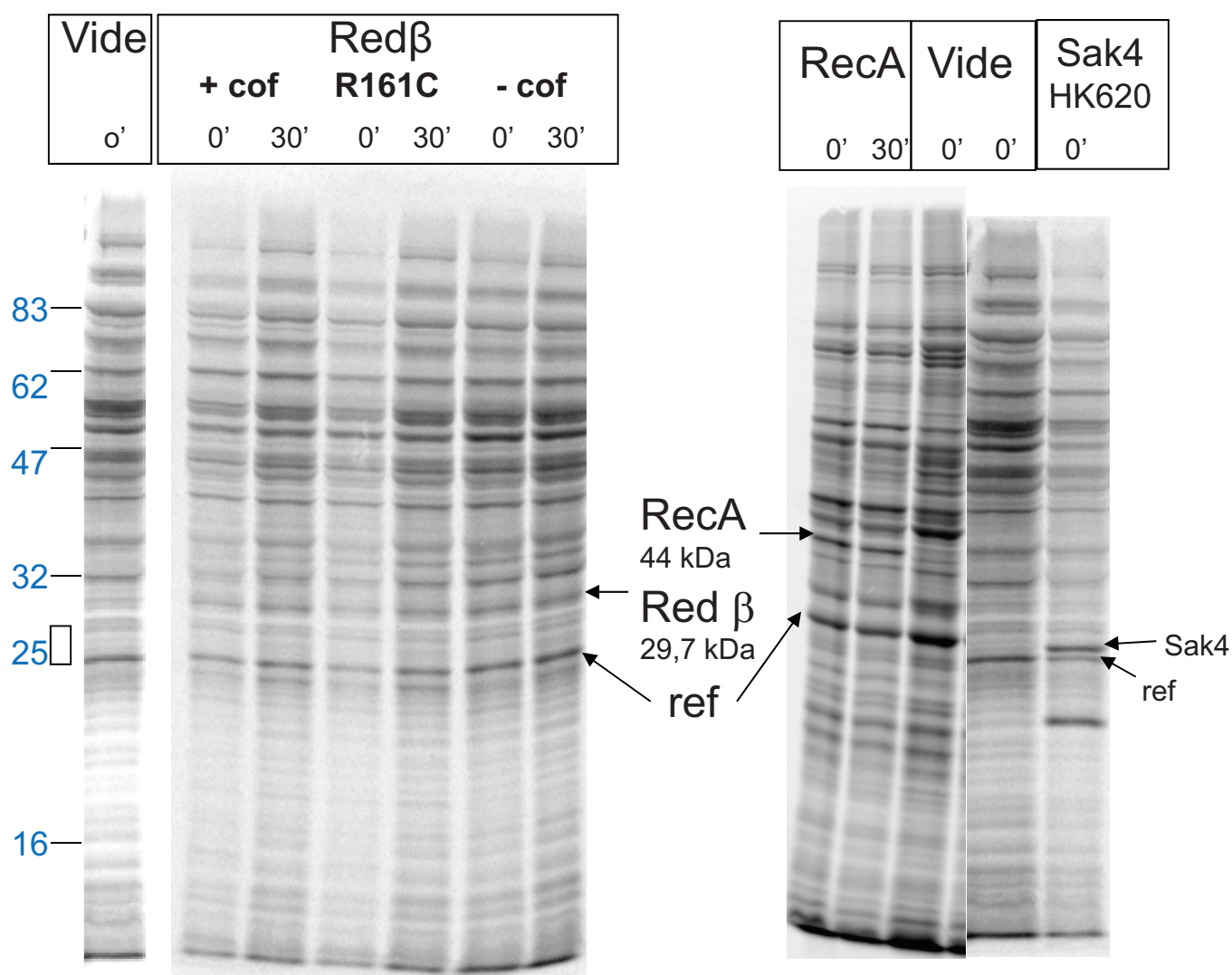
Figure 13: Activité d'appariement d'ADN simple brin *in vivo* de diverses recombinaisons avec ou sans Flag (étiquette DYKDDDDK en N-ter).

La fréquence de recombinants Rif^R est estimée par le ratio de recombinants rifampicine résistant par cellule viable. Chaque valeur correspond à la moyenne d'au moins 3 expériences, réalisées dans la souche AB1157 *recA* qui contient un plasmide dérivé du pSC101 qui exprime la recombinaison à tester après induction à l'arabinose. L'efficacité d'appariement d'ADN simple brin de chaque recombinaison est comparée à l'efficacité obtenue avec le plasmide qui n'exprime aucune recombinaison ('pJA4'), par un test de Student (un double astérisque indique un niveau de significativité à 1%). La ligne en pointillée indique la fréquence de recombinants Rif^R obtenue avec le plasmide vecteur pJA4.

des niveaux équivalents. Pour mesurer ces niveaux, les protéines Red β , Gp2.5 du bactériophage T7, Gp2.5 du bactériophage Φ 12, UvsX du bactériophage T4, et Sak4 du bactériophage PA73 ont été étiquetées avec la séquence DYKDDDDK en N-ter, reconnue par un anticorps Anti-FLAG de souris. Une extraction totale des protéines contenues dans les diverses souches avant et après induction à l'arabinose a été effectuée, déposée sur gel acrylamide 12%, transférée et analysée par Western Blot. Les protéines étiquetées ont ensuite été révélées en utilisant une dilution au 2500^{ème} de l'anticorps Anti-FLAG comme anticorps primaire et une dilution au 20000^{ème} de l'Anticorps secondaire anti-souris couplé à la peroxydase (Les Anticorps ont été commandés chez Sigma). Enfin, l'activité peroxydase est révélée par le kit ECL Direct Nucleic Acid Labelling and Detection System (GE healthcare). Les résultats sont montrés dans la Figure 12. Quatre des 5 protéines étiquetées sont détectées après induction à l'arabinose. Cependant, pour Gp2.5 du bactériophage Φ 12, aucune bande spécifique n'a été détectée. Un test d'appariement d'ADN simple brin *in vivo* réalisé avec ces protéines étiquetées a révélé que l'étiquette inactivait ces recombinaisons, à l'exception de Sak4 (Figure 13).

2. L'expression des recombinaisons phagiques testées par marquage radioactif

Etant donné que nous n'avons pas réussi à détecter la recombinaison Gp2.5 du bactériophage Φ 12 par Western Blot, et que l'introduction d'une étiquette provoque des modifications fonctionnelles de certaines des recombinaisons testées, nous avons effectué des marquages radioactifs au S³⁵ des méthionines de toutes les protéines (la description du protocole est dans l'Article I, les souches sont cultivées en milieu minimum supplémenté d'1 % de LB). Les recombinaisons Red β (produite par pKD46), UvsX, Sak4 du bactériophage PA73 et Gp2.5 du bactériophage Φ 12 sont produites en quantité comparable, alors que Gp2.5 de T7 est produite 5 à 10 fois plus (voir Figure S6 de l'Article 1). Les recombinaisons RecA et Sak4 du bactériophage HK620



	Redβ+ cofacteurs	Redβ R161C + cofacteurs	Redβ - cofacteurs	RecA	Sak4 HK620
Radioactivité (relativement à ref)	1	1,1	1,48	0,88	1,13
Nb méthionines	7	7	7	10	10
Quantité relative de protéine	1	1,1	1,48	0,88	0,67

Figure 14: Estimation du niveau de synthèse et de la stabilité des recombinaisons par marquage des méthionines au ^{35}S . Marquage après induction à l'arabinose, directement après le marquage (0') et 30 minutes après la chasse (30'). « Vide » : plasmide qui n'exprime aucune recombinaison (pJA4). Les quantités de radioactivité sont mesurées relativement à la protéine de référence indiquée « ref ».

ont également été détectées par cette technique, elles sont produites en quantité comparable à Red β et elles sont stables après 30 minutes de chasse (Figure 14). Afin de s'assurer que les pertes d'activité obtenues avec les recombinaases mutées n'étaient pas dues à l'instabilité des protéines produites, les marquages ont également été effectués pour Red β R161C (Figure 14), Sak4_{PA73}K30A, Sak4_{PA73}K123A (Figure 15), UvsX Δ Cter, UvsX K66A, et UvsX W309L (Figure 16). Nous concluons que pour ces différents mutants, les quantités produites sont comparables à un facteur 2 près, et que ces protéines sont stables après 30 minutes de chasse.

6. Expérience de complémentation de la sensibilité aux UV de la souche AB1157 *recA*.

Nous avons voulu déterminer si les différentes recombinaases phagiques testées *in vivo* étaient capables de compléter le défaut de croissance d'une souche mutée pour *recA* après irradiation aux UV. Différentes dilutions des souches exprimant les recombinaases : UvsX du bactériophage T4, Sak4 du bactériophage PA73 et Sak4 du bactériophage HK620 ont été déposées en goutte sur boîte LB, puis exposées à des doses d'UV de 0, 5 et 20 J.m⁻². En contrôle, la souche mutée pour *recA*, la souche sauvage (WT), ainsi que la souche mutée pour *recA* exprimant le gène *recA* à partir du plasmide pJA4 *recA* (pJA5), ont été testées (Figure 17). Les souches exprimant UvsX et Sak4 ne complètent pas la sensibilité aux U.V du mutant *recA*. Cependant, la souche exprimant le gène *recA* à partir du plasmide pJA4 *recA* (pJA5) complète partiellement la résistance aux U.V. Ces résultats sont présentés dans les Figures 17 et 18. La complémentation seulement partielle du plasmide pJA5 pourrait s'expliquer par une production de RecA inférieure à celle de la souche sauvage (non testé).

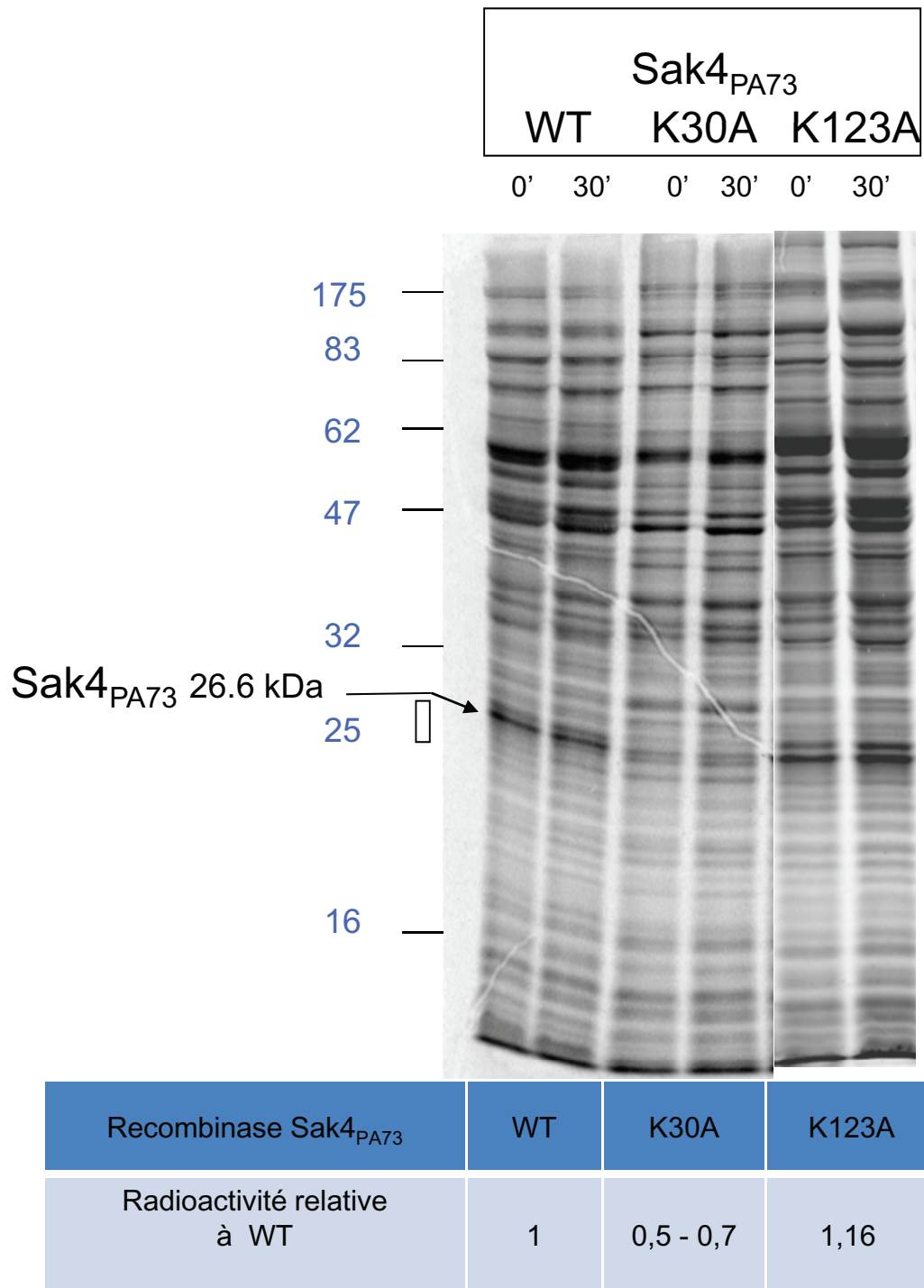


Figure 15: Estimation du niveau de synthèse et de la stabilité des recombinases Sak4_{PA73} par marquage des méthionines au ³⁵S. Marquage après induction à l'arabinose, directement après le marquage (0') et 30 minutes après la chasse (30'). WT pour protéine Sak4 sauvage du bactériophage PA73.

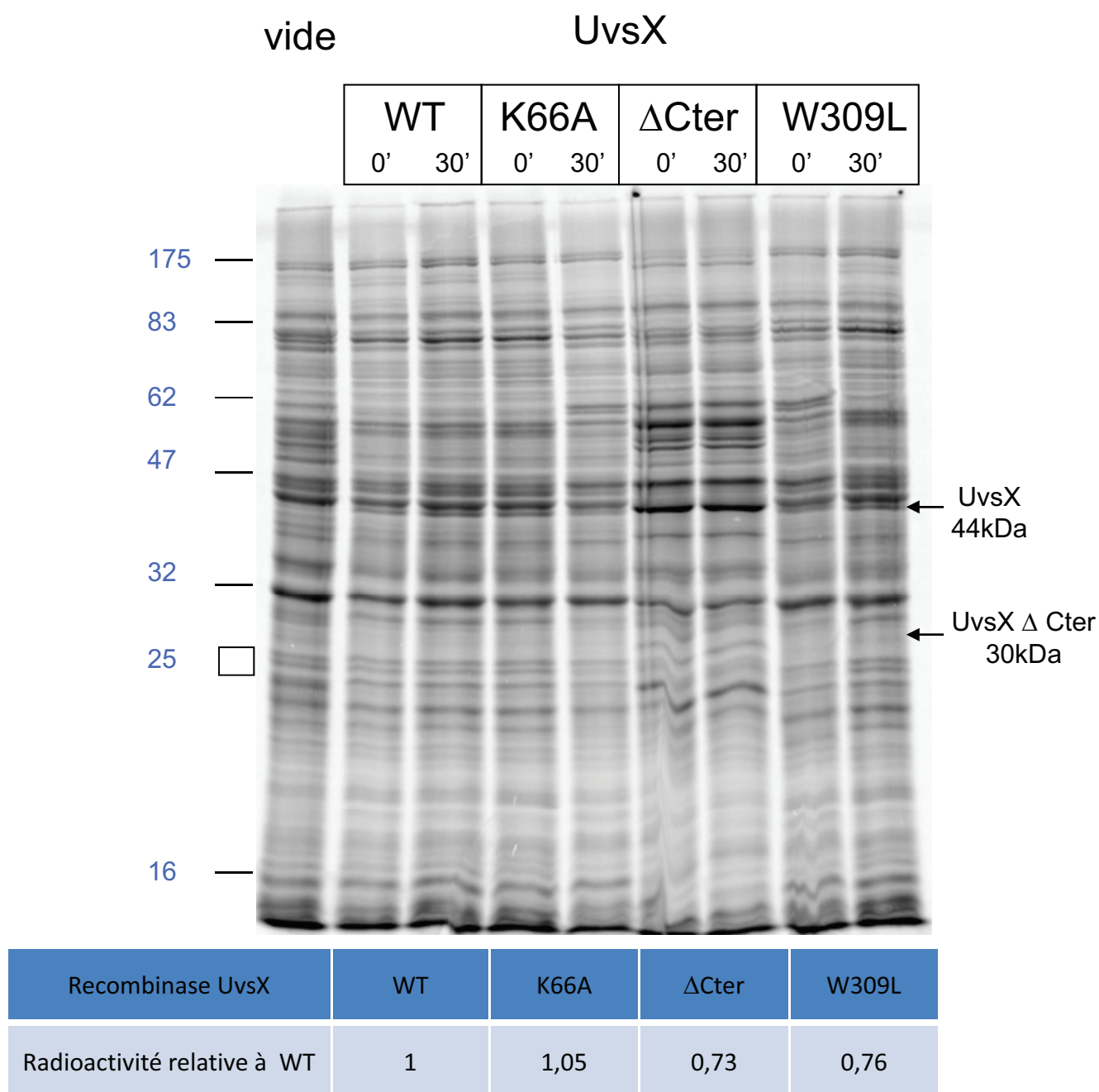


Figure 16: Estimation du niveau de synthèse et de la stabilité des recombinaises UvsX par marquage des méthionines au ^{35}S .

Marquage après induction à l'arabinose, directement après le marquage (0') et 30 minutes après la chasse (30'). « WT » : protéine UvsX sauvage. « Vide » : plasmide qui n'exprime aucune recombinaise (pJA4).

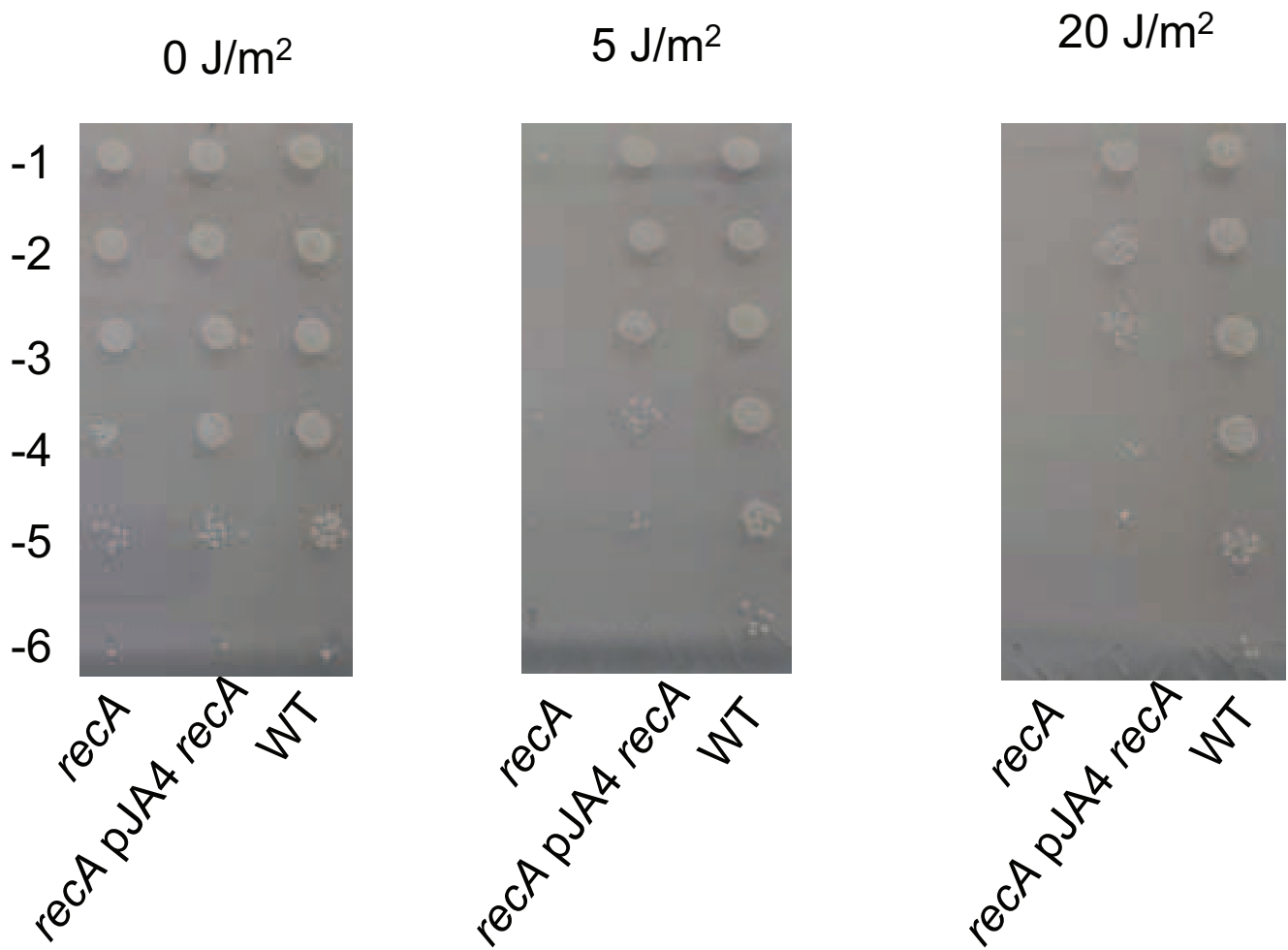


Figure 17: Sensibilité aux UV par spot test.

Sensibilité à des doses d'UV de 0, 5 et 20 J.m⁻² des souches AB1157 *recA*, AB1157 *recA* ayant le plasmide pJA4 *recA* (pJA5) et de la souche WT (AB1157 codant le gène *recA* sauvage).

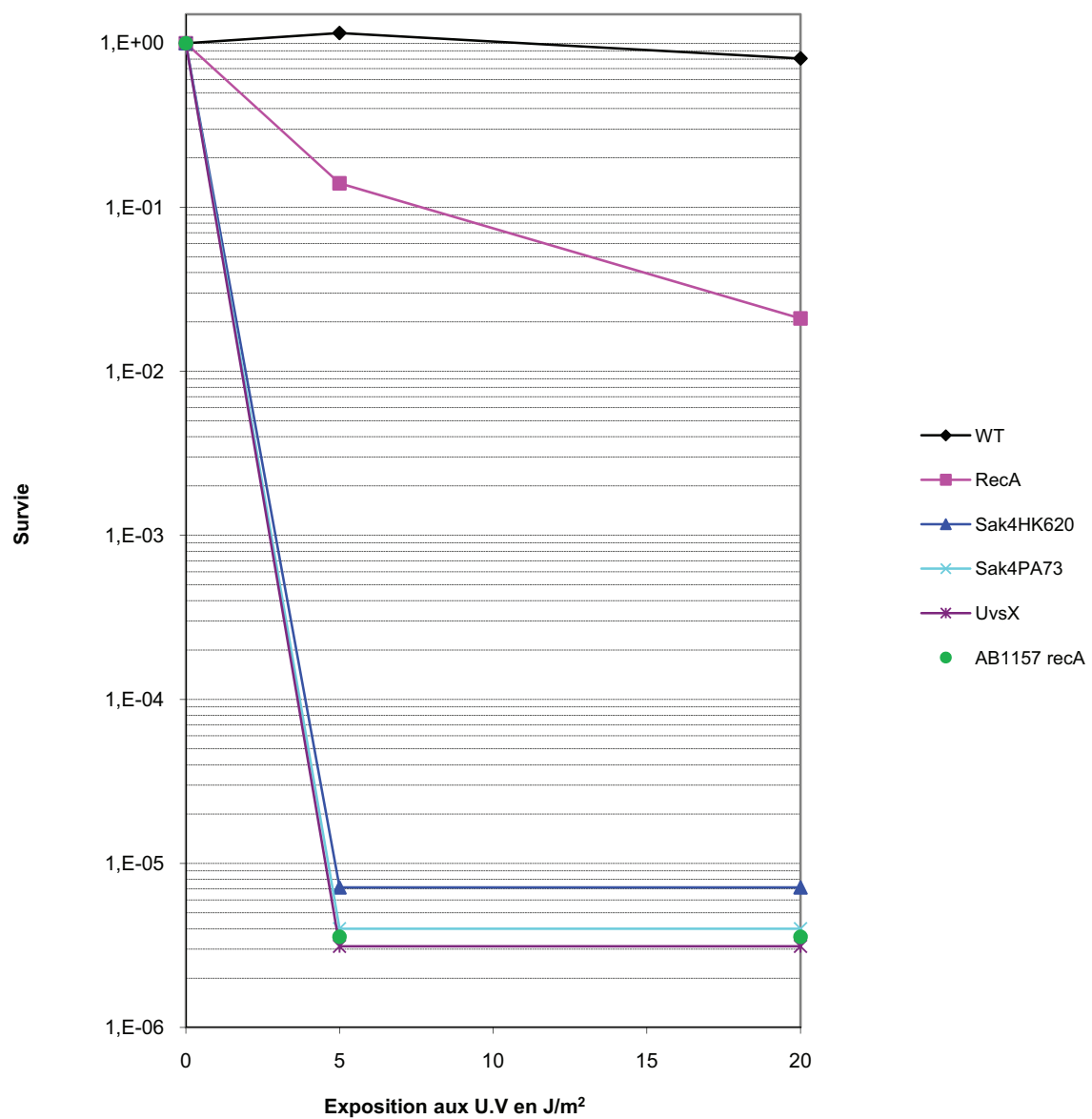


Figure 18: Sensibilité aux UV de la souche AB1157 *recA* contenant divers plasmides, ainsi que de la souche sauvage (WT).

RESULTATS

Chapitre I

Résumé

Les génomes de bactériophages contiennent de nombreux gènes dont la fonction est inconnue et ne peut pas être prédite par des analyses comparatives classiques car les bactériophages sont trop divers. Cette très grande diversité est principalement due au fait que certains des bactériophages séquencés aujourd'hui ont divergé depuis très longtemps, peut-être même selon certains auteurs avant la séparation eucaryotes-procaryotes (Forterre and Prangishvili, 2009). De plus, un très petit nombre de bactériophages ont été bien étudiés, et donc le répertoire de gènes phagiques dont la fonction est connue est très faible au regard de leur diversité. Enfin, les phages évoluent rapidement en raison de leurs taux de mutation (Drake et al., 1998) et de recombinaison (Martinson et al., 2008) élevés. Une étude bio-informatique réalisée par Raphaël Guerois et Anne Lopes (IbITecS, CEA, Gif sur Yvette) a recherché toutes les recombinaisons phagiques présentes dans la base de donnée ACLAME (version 0.3 de mars 2008) (Leplae et al., 2004; Leplae et al., 2010), qui contient 465 génomes de bactériophages entièrement séquencés. Parmi ceux-ci, seuls 22% étaient pourvus d'une recombinaison connue. Cette étude basée sur la détection d'homologie lointaine entre recombinaisons par des comparaisons profil-profil, a permis de conclure à l'existence de trois superfamilles de recombinaisons, l'une apparentée à Rad52, qui comprend les protéines Sak, Red β et Erf (94 génomes sur les 465 analysés possèdent une telle recombinaison), l'autre apparentée à Rad51/RecA, qui comprend les protéines UvsX et Sak4 (71 génomes) et la troisième superfamille contient les recombinaisons apparentées à la protéine Gp2.5 du bactériophage T7 (30 génomes). Au final, il a été possible de détecter une

recombinase dans 42% des génomes de bactériophages présents dans ACLAME (Article 1).

La première partie de ma thèse a consisté à mettre au point un système expérimental d'appariement d'ADN simple brin *in vivo*. Ce système a permis de confirmer que la protéine Gp2.5 du bactériophage Φ 12, un prophage de *Staphylococcus aureus*, et la protéine Sak4 du bactériophage PA73, ont bien une activité de recombinaison (Figure 5 et Table S7 de l'Article 1), validant l'étude bio-informatique. Par ailleurs, nos résultats expérimentaux ont confirmé que les protéines Red β du bactériophage Lambda, Gp2.5 du bactériophage T7 et UvsX du bactériophage T4 étaient capables d'apparier de l'ADN simple brin *in vivo*, et surtout que les deux protéines Sak4 du bactériophage PA73 et Gp2.5 du bactériophage Φ 12 étaient aussi des recombinases (Article 1) (Lopes et al., 2010), alors que jusqu'à présent ces deux dernières protéines n'étaient pas connues comme recombinase.

Article

Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs

Anne Lopes^{1,2}, Jihane Amarir-Bouhram³, Guilhem Faure^{1,2}, Marie-Agnès Petit^{3,*} and Raphaël Guerois^{1,2,*}

¹CEA, iBiTecS, F-91191 Gif sur Yvette, ²CNRS, URA 2096, Gif-sur-Yvette, F-91191 and ³INRA, UMR 1319, Micalis, 78350 Jouy en Josas, France

Received July 27, 2009; Revised February 2, 2010; Accepted February 3, 2010

ABSTRACT

Homologous recombination is a key in contributing to bacteriophages genome repair, circularization and replication. No less than six kinds of recombinase genes have been reported so far in bacteriophage genomes, two (UvsX and Gp2.5) from virulent, and four (Sak, Red β , Erf and Sak4) from temperate phages. Using profile–profile comparisons, structure-based modelling and gene-context analyses, we provide new views on the global landscape of recombinases in 465 bacteriophages. We show that Sak, Red β and Erf belong to a common large superfamily adopting a shortcut Rad52-like fold. Remote homologs of Sak4 are predicted to adopt a shortcut Rad51/RecA fold and are discovered widespread among phage genomes. Unexpectedly, within temperate phages, gene-context analyses also pinpointed the presence of distant Gp2.5 homologs, believed to be restricted to virulent phages. All in all, three major superfamilies of phage recombinases emerged either related to Rad52-like, Rad51-like or Gp2.5-like proteins. For two newly detected recombinases belonging to the Sak4 and Gp2.5 families, we provide experimental evidence of their recombination activity *in vivo*. Temperate versus virulent lifestyle together with the importance of genome mosaicism is discussed in the light of these novel recombinases. Screening for these recombinases in genomes can be performed at <http://biodev.extra.cea.fr/virfam>.

INTRODUCTION

The interest for bacteriophages, viruses infecting bacteria, and for viruses in general, is ever-growing with the discovery of their ubiquity in various natural ecosystems (1–4), their diversity (5–7), and their probable regulatory role in the equilibrium of bacterial populations (8–11). However, phage diversity also represents a challenge when their genomes need to be annotated. At present, each new genome brings in a majority of unknown predicted proteins (5–7). Large-scale analyses with ACLAME (12,13), a database dedicated to the comparative analysis of microbial mobile elements, has confirmed the abundance of orphan proteins in all completely sequenced phage genomes (14). This diversity may appear puzzling when one considers the simplicity of the bacteriophage life cycle, which consists mainly of reproducing its genome and its capsid in a given bacterial host. Phages ought to encode a few basic functions to this end, and one may expect to identify them easily by protein alignments. The fact that the number of unknown phage genes continues to grow suggests either that many phage functions do not derive from a common ancestor or that the current tools for protein alignments are not appropriate. It has been known for a long time that both mutations (15) and recombination rates (16) are much higher in phages than in bacteria, and this may blur the homology signal (17).

For this reason, we decided to apply sensitivity-enhanced techniques to investigate in detail one of these families of phage-encoded proteins. We focused on the phage proteins that are responsible for homologous recombination, which is an essential function in numerous phages: it contributes to genome repair (18), to the circularization of genomes with terminal repetitions (19), and to replication (20). When absent, the function

*To whom correspondence should be addressed. Tel: +33 1 34 65 28 67; Fax: +33 1 34 65 20 65; Email: marie-agnes.petit@jouy.inra.fr
Correspondence may also be addressed to Raphaël Guerois. Tel: +33 1 69 08 67 17; Fax: +33 1 69 08 92 75; Email: raphael.guerois@cea.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2010. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

may be complemented by the host-encoded RecA protein. However, the conservation of this phage-encoded function suggests that phages are better served by their own recombinases, as laboratory experiments may not reflect the variety of growth condition with which the phages and their hosts are faced in nature. Virulent and temperate phages are thought to contain different recombinases. So far, six families of recombinase genes have been reported in bacteriophage genomes, two (UvsX and Gp2.5) from virulent, and four (Sak, Red β , Erf and Sak4) from temperate phages, or from ex-temperate phages that have become virulent after the loss of a lysogeny module (designated collectively hereafter as temperate). Both virulent and temperate phages adopt radically different relationships with their hosts. Virulent phages can only kill them whereas temperate can integrate within their host genome. Different temperate genomes can therefore meet in the same host, and create mosaics by recombination, an event less likely occurring in virulent phages. Recombination processes play therefore a key role in these various lifestyles and the functions of these recombinases need to be further uncovered.

Among virulent phages, UvsX, a RecA-like enzyme present in phage T4, has orthologs in other phages with large genomes and strictly virulent lifestyles (25 orthologs in the ACLAME database version 0.3, March 2008, containing 465 phage genomes). The virulent phage T7 encodes the Gp2.5 recombinase, which shares functional and structural homology with SSB proteins (21), but in contrast to SSB it also provides a recombinase function through its single-strand annealing activity (22,23). It has only seven orthologs in the ACLAME database.

The recombinases identified in temperate phages seem more diverse although they share common properties. Sak, Red β and Erf, which were delineated as three distinct families on the basis of their primary sequences (24), harbour a single-strand annealing activity. All three proteins form similar ring-like quaternary structures visible by electron microscopy (25–27). Red β from phage λ and RecT from prophage *rac* are two well-known members of the Red β -like family that are used for genetic engineering (28,29). Three striking properties distinguish these recombinases from RecA-like enzymes: (i) their ATP-independence, (ii) their high efficiency in pairing short, 40–50 bp long sequences (30) and (iii) their ability to incorporate short single-strand DNA substrates into the bacterial chromosome (31). In addition, the Red β protein was recently reported to be 100-fold more efficient than was RecA for homeologous recombination between 22% divergent sequences (16). All these characteristics may account for the remarkable plasticity and mosaicism of temperate bacteriophage genomes. Interestingly, the Sak protein shares homology with the N-terminal, globular domain of the eukaryotic Rad52 protein (24,26), and a distant homology between Red β and Rad52 has also been reported (32).

It is suspected that other families of recombinases are yet to be discovered. Indeed, most sequenced phage genomes do not encode homologs of any of the aforementioned proteins (78% of phage genomes in ACLAME). A genetic study by the group of Moineau has probably

uncovered one such new family called Sak4 (33). The common characteristic of all proteins called ‘Sak’ is their sensitivity to the plasmid-encoded phage-resistance system AbiK (for abortive infection system K). This plasmid is present in some *Lactococcus lactis* bacterial strains, and confers phage resistance. Sak and Sak2 belong to the same family, and Sak3 was identified as homolog to Erf (33), but Sak4 could not be related to any known recombinase. Up to now, Sak4 has not yet been formally proven to be a recombinase, but the fact that *sak4* bacteriophage mutants are as resistant to AbiK as are *erf* or *sak* recombinase mutants argues strongly for such an activity.

In summary, with the current protein alignment tools, at least six families of phage recombinases have been found or suspected, namely UvsX, Gp2.5, Sak, Red β , Erf and Sak4. The first two systems, UvsX and Gp2.5, seem quite specific to virulent bacteriophages lifestyle, whereas the four others appear to be more broadly distributed. Using in-depth bioinformatic methodologies, coupling profile–profile comparisons, structure-based modelling and gene-context analyses, we provide access to the global landscape of recombinases in bacteriophages: based on the global analysis of 465 bacteriophage genomes archived in the ACLAME database (12,13) we show that the Sak, Red β and Erf families belong to the same Rad52 superfamily. We unveil that Sak4 homologs are widespread and correspond to a shortcut version of Rad51. Coupled with gene-context analyses, these large-scale profile–profile analyses further reveal that Gp2.5-like recombinases are also widespread in bacteriophage genomes and are not limited to relatives of T7. For two newly detected recombinases belonging to the Sak4 and Gp2.5 families, we provide experimental evidence of their recombination activity *in vivo*. We therefore conclude that, previously unsuspected, a majority of bacteriophages encode a recombinase function that can be classified into one of only three superfamilies Rad52, Rad51 or Gp2.5. The discovery of so many recombination systems, which appear to segregate mainly according to the phage temperate or virulent lifestyle, raises fascinating questions as to their DNA repair strategies, their genome diversity and their evolution.

MATERIALS AND METHODS

Bioinformatics procedures

All the 465 genomes used for this study were taken from the ACLAME database version 0.3, release of March 2008 (A CLAssification of Mobile genetic Elements) (12,13). For each protein of the 465 phages, multiple sequence alignments were built using three iterations of PSI-Blast (34) against the nr90 database (non redundant sequence database filtered at 90% identity) and were transformed into HMM profiles using the HHmake algorithm (35) resulting in a total of 28 300 HMM profiles.

The HMM profiles of four recombinases [Sak from ul36 (NP_663647), Red β from 933W (NP_049474), Erf from D3 (NP_061548) and Sak4 from ϕ 31 (AAC48871)] were used as initial queries to screen the 28 300 profiles of the

465 genomes to identify homologs using the profile–profile comparison program HHsearch (v1.6.0), also called HHpred (35). Accordingly, a HHsearch hit was considered as significant when probabilities for being a true positive exceeded 90% with both global and local alignment modes (see Supplementary Figure S1 for justification of the threshold). The same thresholds of 90% were used in the next iterations, in which previously detected profiles were in turn compared to the whole-profile database until convergence is reached. Sak4 identification further required validation because it is composed of an ATPase domain which is a widely represented superfamily. Using Sak4 profile as a query matched not only Sak4 homologs but also other ATPase such as RecA proteins or DnaB helicases. Sak4 proteins are shorter than RecA or DnaB, which contain a DNA binding or a DnaB-like domain in their C-terminal region, respectively. To prevent the selection of any false positive in the assignment of Sak4 homologs, only short sequences (less than 290 residues), excluding the possibility of RecA or DnaB C-terminal domains, were assigned as Sak4.

Models of Sak, Red β and Erf undecamers (11-mer), were generated with Modeller 9v5 (36) using the human Rad52 N-terminal domain as template (PDB:1h2i). Initial alignments between one monomer of each recombinase and the template were obtained from profile–profile comparison (35). The sequence divergence between the three recombinases and the Rad52 template was such that several regions had to be carefully re-aligned in order to provide sequence alignments consistent with the structural constraints imposed by the Rad52-fold and optimizing the scores of Verify3D (37) and Prosa (38) methods assessing the 3D models likelihood (see scores in Supplementary Table S1 and Figure S2). The same procedure was followed for generating Sak4 and Gp2.5 models using RadB from *Thermococcus kodakarensis* (PDB:2cvh) and Gp2.5 from phage T7 (PDB:1je5) as templates, respectively. The multiple sequence alignments were processed with Jalview (39). Structural models were represented using PyMol (40). Conservation analyses were carried out using the Rate4Site algorithm (41). Profile–profile comparisons against the recombinase profiles database can be run and all profiles, models and alignments can be downloaded at <http://biodev.extra.cea.fr/virfam/downloads.html>.

Gene-context analyses were run considering the 20 genes upstream and downstream each of the 133 recombinases identified above. Profile–profile comparisons between the profiles of these genes and the 28 300 profiles database were performed as described above. Related genes were clustered together using full linkage hierarchical agglomerative clustering. This procedure was repeated for all the phage genomes in which no recombinase was detected focusing on the DNA replication modules.

Single-strand annealing assay

In the low-copy number and thermosensitive plasmid pKD46, *red β* of phage λ is cloned under a *ParaB* promoter, and its transcription is induced by AraC in

the presence of arabinose (28). The pKD46 backbone was used to replace the λ fragment encompassing *gam*, *red β* and *red α* by *uvsX* of phage T4, *gp2.5* of phage T7 or phage ϕ 12, or *sak4* of phage PA73, each preceded by its RBS, between the *EcoRI* and *NcoI* sites of pKD46. *EcoRI* and *NcoI* sites were artificially introduced at the 5' and 3' extremities of the recombinase genes, respectively, by PCR. Site-directed mutagenesis was realized by PCR, designing two 40-bp long complementary oligonucleotides centred on the mutation to be added to the gene. Left- and right-arm of the gene were then amplified separately, and both fragments were then combined in a third PCR to amplify the whole gene. All constructions were verified by sequencing. Plasmids, strains and bacteriophages used for the assay are reported in Supplementary Table S2.

Electro-competent cells were prepared essentially as described by the Biorad manufacturer, except that cells were grown at 30°C in SLB (5 g/l NaCl, 20 g/l yeast extract, 35 g/l bacto-tryptone and 1 ml/l NaOH 5N). Ninety minutes prior harvesting, the recombinase was induced by addition of 0.2% arabinose. After 1 h expression of the electroporated cells in SOC medium supplemented with 0.2% glucose, the cultures were further incubated over-night without agitation at room temperature, before plating on LB supplemented with 50 μ g/ml rifampicin, and incubating at 37°C. All competent cells had similar levels of competence of around $2 (\pm 1) \times 10^7$ transformants per microgram, as estimated with the pACYC184 plasmid.

The 51-nt long oligonucleotide Maj32 is centred on the nucleotide 1576, starting from the ATG of the *rpoB* gene of *E. coli* MG1655, and corresponds to the non-coding strand. In this oligonucleotide, a G is changed to a C relative to the wild-type sequence at position 1576, so that the CAC codon on the coding strand becomes GAC, and the RpoB protein has a H526D mutation, which confers resistance to rifampicin (42). Maj32 was ordered from Eurogentec, as desalted and purified, and resuspended at a concentration of 1 mg/ml. Increasing amounts of the oligonucleotide mixture were tested, to determine the saturating concentration above which no more transformants could be obtained with the Red β positive control. Ten micrograms of Maj32 were found to be saturating.

RESULTS

Remote homology detection of recombinases among phage genomes

Three distinct families of recombinases, represented by Red β , Erf, Sak proteins, can be distinguished in bacteriophages, according to the classification scheme of Koonin's group (24) and a fourth one, Sak4, much less well-characterized, has emerged from recent experimental work (33). First, the presence of any of these recombinases among sequenced bacteriophages was assessed based on the 465 genomes archived in the ACLAME database (12,13). Profile–profile comparisons relying on algorithms such as HHsearch (35), are among the most powerful methods to search for homologs among rapidly diverging

sequences. Here, the efficiency of the approach was further enhanced by implementing an iterative and systematic profile–profile strategy. Every gene of every bacteriophage was used as a query to build a multiple sequence alignment so as to create a database of 28 300 profiles. This large ensemble of profiles was screened against each of the four Red β , Erf, Sak and Sak4 recombinase profiles using the HHsearch algorithm (35). This global analysis showed that 126 phages harboured a remote homolog of one of the recombinase superfamilies (Figure 1, thick coloured edges). Among them, 2 Red β -like, 5 Erf-like, 21 Sak-like and 40 Sak4-like were found, which had not been identified in the ACLAME database as recombinases (circles with thick outline in Figure 1). Using each of the 126 profiles as queries, a second iteration of profile–profile comparisons against the 28 300-profile database was sufficient to reach convergence. Seven additional bacteriophages with Sak or Red β recombinase were retrieved. Unexpectedly, this second iteration also

revealed that remote homologs of one recombinase family could match remote homologs of another family (shown as black connections between crowns in Figure 1).

As reported in Figure 1, three distinct recombinase families, namely Red β , Erf and Sak, were connected through remote homology relationships, whereas the Sak4 group of homologs remained isolated. These relationships strongly support the notion in that the three recombinases families belong in fact to a unique superfamily. This result is all the most surprising that none of the direct profile–profile comparisons between the initial Red β , Erf and Sak profiles revealed any significant remote homology signal. Hence, detection of remote homologies between drastically diverged sequences can be considerably enhanced through iterative all-against-all profile–profile comparisons and search for transitive relationships. In Supplementary Table S3, all retrieved recombinases are listed.

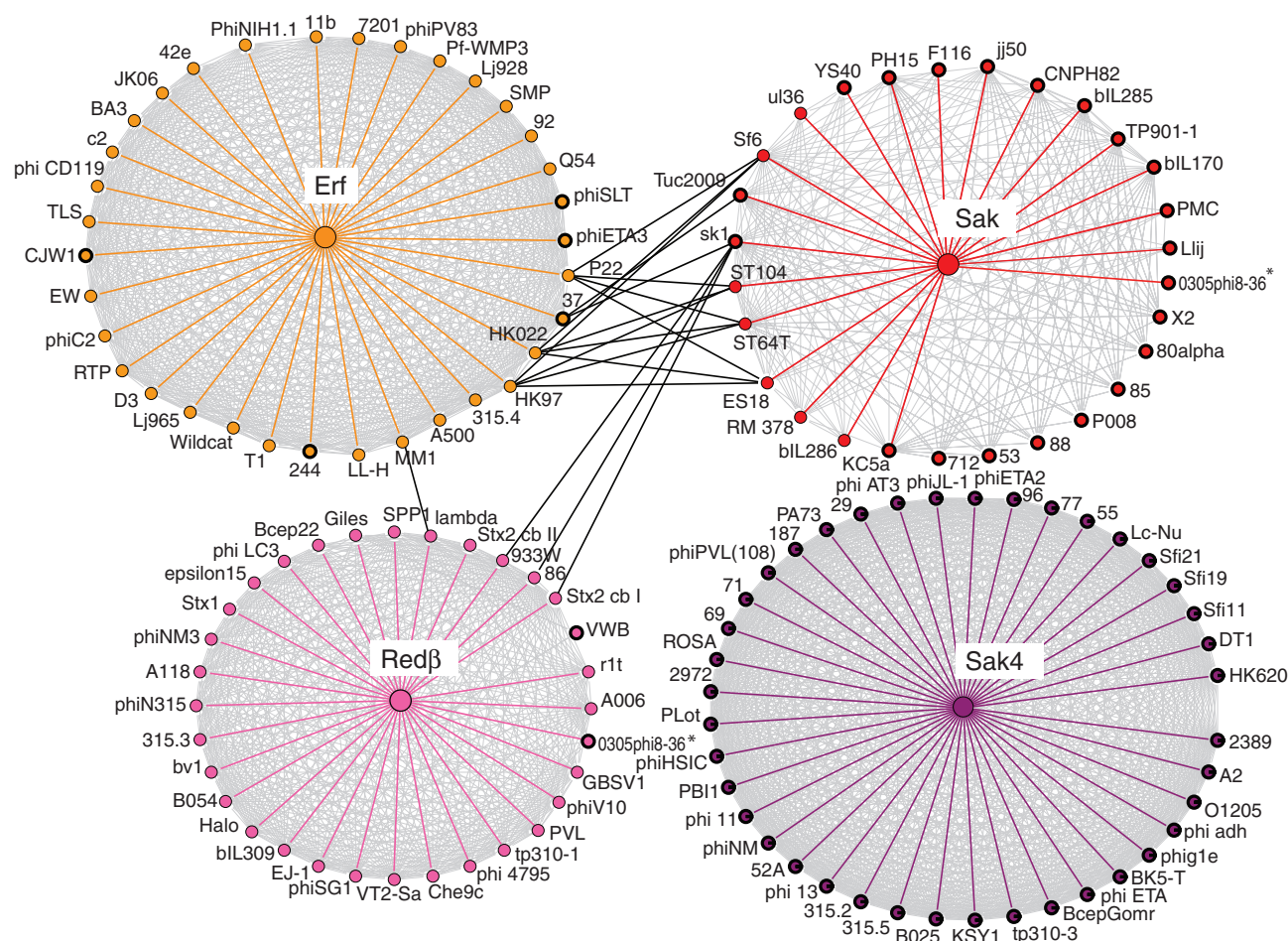


Figure 1. Homology relationships between recombinases revealed by iterative profile–profile comparisons. Network of homology relationships detected between the four recombinase superfamilies Red β /RecT (magenta), Erf (orange), Sak (red) and Sak4 (purple). Nodes represent the bacteriophage genomes for which a homologous recombinase was detected. Edges connect recombinases which were found homologous from the profile–profile comparison using HHsearch. Bold coloured edges correspond to significant hit obtained from the initial screening of the bacteriophage 28 300 profiles database with the four recombinase profiles. A second iteration using the detected recombinase profiles as queries revealed additional relationships among the same superfamily (light grey edges) and, most importantly, between different superfamilies (black edges). Nodes with a black outline pinpoint novel recombinases annotated as unknown function so far. Graph created using the Osprey program (53).

The structures for three families of recombinases are predicted to correspond to mini-Rad52

The remote homology between Sak, Erf and Red β proteins implies they adopt related folds and a structural model would provide critical insights into their common properties. Fortunately, a homology between Sak family and the human Rad52 was initially detected through sequence alignments (24) and further assessed recently through several experimental characterizations (26). The structure of the N-terminal domain of human Rad52 was solved and the domain was found to oligomerize into a ring-like shape bearing 11 subunits (43,44). In solution, full-length Rad52 forms homo-oligomers of seven or more subunits (45). Similarly, electronic microscopy visualization of Sak proteins revealed that they assemble into 11-mer stacked ring structures (26). A structural model of the monomeric Sak protein and of its oligomeric form could be generated using the structure of human Rad52 as a template. With respect to Red β and Erf, a structural model could also be generated based on the above assessed homology. Structural assessment of the models guided the optimization of the alignment (see 'Material and Methods' section) (Figure 2A) and the model obtained for Red β is compared to the Rad52

N-terminus structure in Figure 2B as a ribbon representation.

The major characteristic of the three recombinases with respect to Rad52 core structure is to match a shortcut Rad52 fold. Significant structural deletions are found in the main helix (α 3) and the three strands β 3– β 4– β 5 wrapping around α 3 (dashed regions in the secondary structure cartoon in Figure 2A). Such a deletion in the fold probably accounts for the difficulties encountered so far in detecting homologies among the Rad52-like superfamily. Interestingly, although the fold has been significantly rearranged through evolution, some specific features remained conserved in all these recombinases. Figure 2C maps the sequence conservation index at the surface of the recombinase models as calculated by the Rate4Site algorithm (41). The clusters of red/orange positions in the groove testify to the specific conservation and functional importance of this region. In particular, one position, indicated by a red star in Figure 2A–C, is strictly maintained as a positively charged residue in all three recombinases. Mutation of the corresponding amino acid, K152, in human Rad52 was found to fully abrogate single-strand DNA binding (43). We mutated the corresponding position in Red β (R161) and reached similar

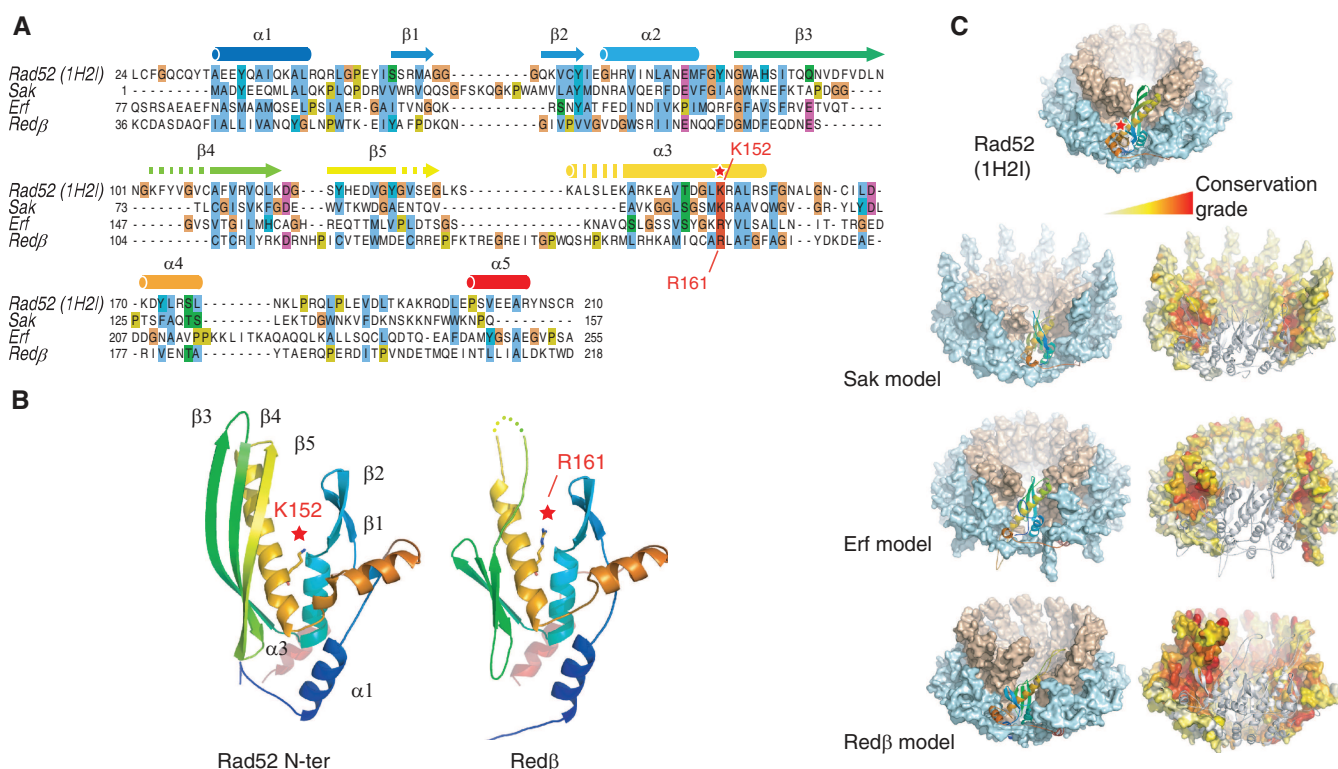


Figure 2. Structural models of the recombinases found evolutionarily related to the Rad52 superfamily. (A) Optimal sequence alignment resulting from the structural modelling procedure between the sequences of human Rad52 corresponding to the PDB (1H2I), Sak (from phage ul36), Erf (D3), Red β (933W). Secondary structures are indicated on top of the alignment from blue to red colour as in (B and C). Truncated secondary structures between Rad52 and the phage recombinases are indicated by dashed secondary structures. The red star indicates the position of K152 in Rad52 whose mutation abrogated single-strand binding. (B) Cartoon representation of the Rad52 crystallized domain and of the Red β structural model with the red star pinpointing the side-chain of K152 and R161, respectively. (C) Surface representation of Rad52 11-mer oligomeric form as observed in the crystal structure 1H2I together with the 11-mer models of the three recombinases Sak, Erf and Red β . On the left, the surface is coloured in either light cyan or orange to help visualizing the external and internal ring, respectively. On the right the surface is coloured with respect to the conservation index calculated using Rate4Site program, conservation grade increasing from white to red.

proteins have lost (or have never acquired) the capacity to promote strand exchange, and maintain only a single-strand annealing activity.

Alternative recombinase systems revealed by combining profile–profile and gene-context analyses

The architecture of the bacteriophage genomes often exhibits a conserved arrangement of gene functions. At a first level, genomes can typically be segmented into modules comprising either head or tail assembly factors, DNA replication machinery or cell lysis factors. Analysis of the gene neighbours for the 133 recombinases detected either as mini-Rad52 (Red β , Erf, Sak) or as mini-Rad51 (Sak4) analysis confirmed that the recombinase was located within a replication module (Supplementary Figure S4, Tables S4 and S5). Furthermore, we found examples of substitutions between the different recombinase families as illustrated in Figures 4A and Supplementary Figure S5. Can the pattern of genes surrounding these recombinases be used further to identify alternative recombinase systems in other phage genomes?

A key step in the gene-context analysis remains the identification of homologous proteins in the vicinity of the recombinases genes. Here again, high sequence divergence among the neighbouring sequences may limit the analysis. We pushed back the efficiency of the homology detection among the neighbours of the recombinases by applying the profile–profile comparisons protocol previously described, focusing on the genes in the DNA replication module. In particular, we looked for orthologs of RecE, a 5′–3′-exonuclease which prepares the single-strand substrate for the recombinase to perform the strand annealing. This protocol revealed 84 genomes containing a *recE*-like homolog, while only 26 could be detected through PSI-Blast analyses (yellow and red striped boxes in Figure 4A). However, among these 84 genomes, only 30 contained both a RecE-like exonuclease and a recombinase (Red β , Erf or Sak4) (shown as pink or dark pink in Figure 4A), suggesting that some other recombinase family may be hidden in the vicinity of the remaining genomes.

Indeed, in 19 genomes, a specific set of homologous but unknown proteins was detected, each of them clustering with a *recE*-like gene. One example of such a protein, found in bacteriophage ϕ 12, is indicated by a white box outlined in black in Figure 4A. Comparing the profiles of these 19 proteins against a database of PDB-derived profiles, revealed an unambiguous homology to the gene protein 2.5 (Gp2.5) of bacteriophage T7. Gp2.5 possesses a strand annealing activity which contributes to genetic recombination during growth of T7 phage (22,23). The structure of Gp2.5 from phage T7 showed that it contains an OB-fold (21) but differs from other SSBs by the presence of the long α 1 helix represented in Figure 4B. The R82C mutation (indicated by a red star) located just downstream of this helix specifically abrogated the single-strand annealing activity suggesting that this extra helix is important for this property (49). In the Gp2.5-like protein of phage ϕ 12, the long helix is also predicted and a conserved basic residue is found at the position

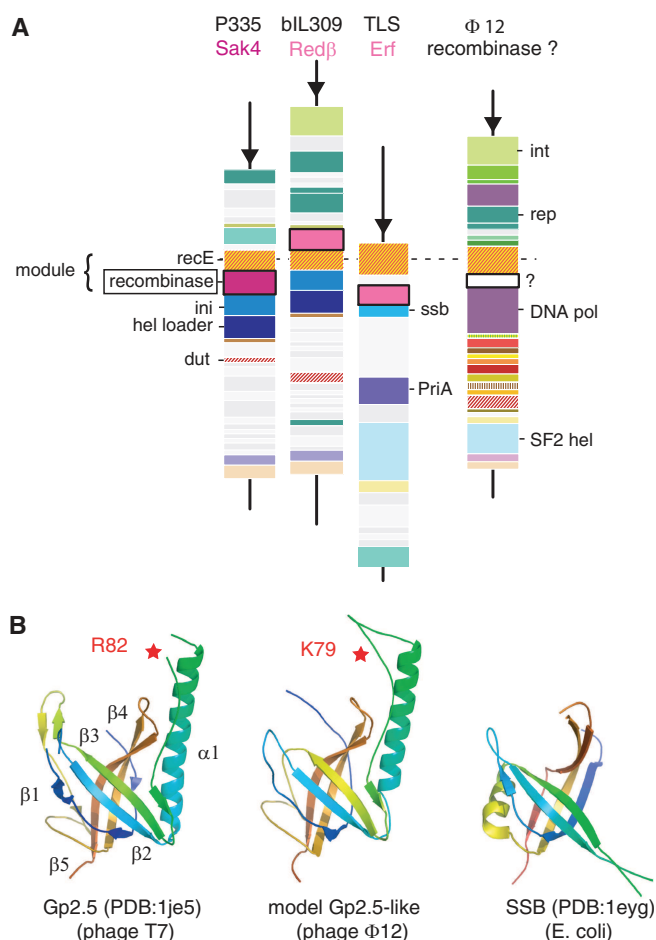


Figure 4. Analysis of the gene context in the DNA replication and recombination module of several bacteriophages. (see also Supplementary Figure S5) (A) Phage names are indicated on top with the type of detected recombinase annotated below. ORFs are represented by ordered boxes with size related to the ORF length. Light and medium grey boxes indicate genes not considered in the analysis. The different predicted recombinases are colored as pink. Other colours, when identical, pinpoint genes sharing a homology relationship (either remote or close). Genes surrounding the homolog of 5′–3′ exonuclease RecE (yellow/red stripes) in P335 *sensu lato* (in fact, phage 4268), bIL309, TLS and ϕ 12 phages. A Rad52 (pink box) or Rad51-like (dark pink box) recombinase is often found in a close neighbourhood of RecE, suggesting that in ϕ 12 (the white gene labelled with a question mark) could itself be a recombinase. Some putative functions have been labelled by a short-term, *ini*, initiator; *hel loader*, DNA-C type helicase loader; *ssb*, single-strand binding; *RecE*, 5′–3′ exonuclease RecE; *int*, phage integrase; *repr*, repressor; *DNA pol*, DNA polymerase (Pol I-type); *dut*, dUTPase; *endo*, endonuclease; *SF2 hel*, superfamily 2 helicase; *PriA*, DNA primase. (B) Structural model of the Gp2.5-like protein identified in phage ϕ 12. From left to right, ribbon representation of (i) the structure of Gp2.5 protein (phage T7) (PDB code: 1je5), (ii) the structural model of Gp2.5 distant homolog from phage ϕ 12 and (iii) the structure of the single-strand binding (SSB) protein of *E. coli* (PDB code: 1eyg). Red stars in Gp2.5 structures pinpoint residues whose mutation abrogated single strand annealing activity [in (49) and in this study].

corresponding to R82 in Gp2.5 of phage T7. Hence, not only the existence of the helix but also the conservation of important positions suggest that protein Phi12P11 (NP_803317) of phage ϕ 12 belongs to the specific Gp2.5 superfamily rather than to the larger SSB family.

As with Rad51- or Rad52-like recombinases, remote homologs of Gp2.5 are not always associated with RecE and the profile–profile comparison revealed three additional Gp2.5. All in all, 30 phages were found to contain a Gp2.5-like recombinase (Supplementary Table S3), underscoring the importance of this class of recombinase, besides the Rad52- and Rad51-like proteins.

Predicted Sak4 and Gp2.5 recombinases have single-strand annealing activities *in vivo*

It has been reported that various phage recombinases of the Rad52-like superfamily are able to promote single-strand annealing *in vivo* in *E. coli* (50). The assay is based on a 70-nt long oligonucleotide complementary to the *gal* gene that reverts cells to a Gal⁺ phenotype. Depending on the recombinase tested, yields of recombinants varied by a factor of 3000 with Red β giving the higher ratio of recombinants (2×10^{-1} Gal⁺ per viable cell), and GP35 of a *Bacillus subtilis* prophage the lower ratio (6.5×10^{-5}) (50). We used a similar assay to determine whether the predicted recombinases of the Sak4 and Gp2.5 families were able to recombine DNA *in vivo*. To do so, we used a 51-nt long single-strand oligonucleotide complementary to the *rpoB* gene of *E. coli*, except for a single point mutation that confers resistance to rifampicin. We then tested whether the presence of one of the predicted recombinases allowed the oligonucleotide integration into the *E. coli* chromosome, upon transformation by electroporation. The *E. coli* recipient strain was chosen to be defective for *recA*, as it is known that RecA is not needed for such a reaction, and proficient for mismatch repair, to minimize the background level of spontaneous rifampicin resistant (Rif^R) clones. The *rpoB* mutation was a C:C mismatch, which escapes from the MutLSHU repair system. We have tested that *in vivo* the five recombinases are produced at similar amounts, except for Gp2.5 from T7, which is produced at 5–10 times the level of the four other ones (Supplementary Figure S6).

With Red β recombinase (of the Rad52 superfamily), a yield of 2×10^{-3} Rif^R per viable cells was obtained at saturating amounts of oligonucleotide (Figure 5, see Supplementary Table S7 for exact numbers). Next, the full size RecA-like enzyme of phage T4, UvsX and the predicted minimal Rad51-like protein of the Sak4 family from phage PA73 (infecting *Pseudomonas aeruginosa*), were compared. Both recombinases produced $1\text{--}2 \times 10^{-6}$ Rif^R transformant per cells, 2000-fold less than Red β . Nonetheless, even this lower frequency was significantly higher than the background level of Rif^R recombinants produced with the control plasmid devoid of recombinase, which was on average at 4.9×10^{-7} per viable cell. Because the Gp2.5 family has various distant homologs, two members of the family were tested, those of phages T7 and ϕ 12 (a prophage of *Staphylococcus aureus*). A yield of 10^{-5} Rif^R clones per viable cells was obtained with the T7 protein, and 10-fold less with the ϕ 12 protein. The higher yield observed with the T7 protein compared to the ϕ 12 protein may be due to its higher expression level. We conclude therefore that the predicted

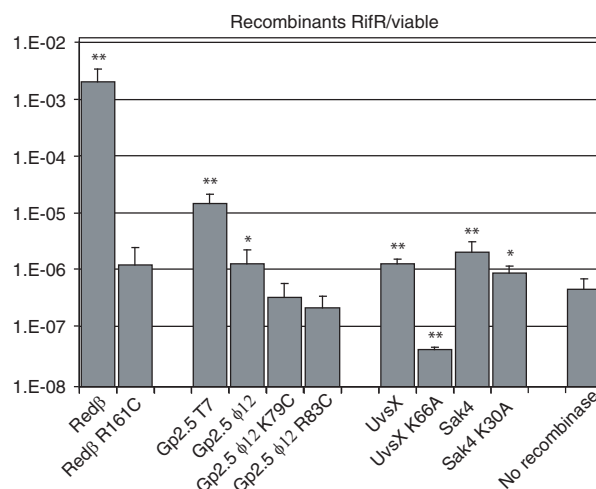


Figure 5. Single-strand annealing activities *in vivo* of the three superfamilies of recombinases. Efficiency of recombination is estimated by the ratio of rifampicin-resistant recombinants (generated by integration of the Maj32 single-strand oligonucleotide into *rpoB*), per viable cells. Each value corresponds to the average of at least three experiments, performed in an *E. coli* AB1157 derivative in which a given recombinase (x-axis) has been induced from a low-copy number pSC101 plasmid derivative. Spontaneous rifampicin-resistant mutants were obtained at a median frequency of 2×10^{-8} . To ascertain the effect of each recombinase, as compared to the empty plasmid vector ('No Recombinase'), a Student test was performed: double asterisks indicate significance at the 1% level, single asterisk indicate significance at the 5% level.

recombinases of phages PA73 and ϕ 12 promote single-strand annealing *in vivo*, albeit with a reduced efficiency with respect to Red β . The reduced efficiency may be related to the heterologous expression of these proteins in *E. coli*, and more complete studies are required to characterize fully these new recombinases.

Alignments of recombinases of the Rad52 family highlighted a conserved arginine/lysine residue across all sub-families (Figure 2, red star). A mutation converting this arginine into a cysteine (R161C) in the Red β gene led to a 1000-fold reduction of *in vivo* annealing efficiency (Figure 5 and Supplementary Table S7), confirming the critical role of this positively charged residue *in vivo*. The structure function analysis of Gp2.5 revealed the R82C substitution as affecting specifically the single strand annealing activity of the protein *in vitro* (49). A substitution affecting the lysine residue located at the equivalent position in the *gp2.5* gene of phage ϕ 12 (K79C) was therefore tested, as well as a neighbouring position, more conserved, R83C (Figure 4B and Supplementary Figure S3). Both protein variants produced a yield of recombinants undistinguishable from the background value. This confirmed the importance of both residues in the ϕ 12 protein, as well as the significance of the low activity detected for the wild-type ϕ 12 protein. Mutants in the Walker A motif of the ATPase of UvsX and Sak4 proteins were done, and interestingly the recombination activity of the Sak4 mutant was not affected at all, in contrast to the UvsX mutant (Figure 5 and Supplementary Table S7). The property of Sak4 is

reminiscent of the behaviour reported for XRCC2, a short version paralog of Rad51 which regulates the length of gene conversion tracts in vertebrate cells, irrespective of its ATPase activity (51).

DISCUSSION

Three major superfamilies of phage recombinases

Our systematic analysis of 465 bacteriophages combining sensitivity-enhanced homology searches, structural modelling and gene contextual analyses revealed many insights into the nature of the homologous recombination systems in bacteriophages (Supplementary Table S3 and Figure S7). First, we were able to draw a unified picture for most recombinases experimentally studied so far. We assessed with high confidence that the Rad52 fold has been substantially rearranged and shortcut throughout bacteriophage evolution, giving rise to a wide variety of recombinases that were previously thought to belong to different families (24). This feature has also been highlighted in two recent works for Sak (26) and Red β (32). In our work, careful structural modelling leads to optimized alignments. As a result, our Red β /Rad52 alignment (Figure 2A) predicts R161 as a hot functional residue in Red β , whereas this residue was not aligned with conserved positions in (32).

Strongly supported by gene-context analyses, we further discovered two unsuspected recombinase systems evolutionarily related either to Rad51 or to Gp2.5 superfamilies. Most members of the Rad51-like family were orthologs of Sak4, which is again a shortcut version of Rad51, and substituted extensively for Rad52 homologs in similar gene contexts. Experimentally, we validated that newly identified members of the Rad51-like and Gp2.5-like superfamilies do act as recombinase *in vivo*. Point mutations abrogating Red β (R161C) and Gp2.5 (R79C, R83C) activity were successfully designed and validate the alignments optimized with the help of the structural templates. The huge majority of genomes have at most one recombinase gene. This almost exclusive presence likely reflects their similar tasks in all these bacteriophages (Supplementary Table S3). The four exceptions (Phages CJW1 and 244 from *Mycobacterium smegmatis*, 0305phi8-36 from *Bacillus cereus*, and YS40 from *Thermus thermophilus*) include cases where a Rad51-like gene is combined with a Rad52-like gene, a situation reminiscent of the eukaryotic kingdom.

Recombinase distribution correlates with genome size

Are these recombinases found in certain classes of bacteriophages or are they randomly distributed? As regards genomes size, all genomes of <20 kb (140 genomes), including single-strand DNA genomes and RNA genomes, were devoid of recombinase. In the 20–80 kb range, 160 genomes out of 286 were identified with a recombinase from either the Rad52-like, Sak4-like or Gp2.5-like superfamily. Rad52-like and Sak4-like are almost exclusively integrated in temperate phages (P -value of $5.5 \cdot 10^{-5}$ and $1.2 \cdot 10^{-4}$, respectively), whereas Gp2.5-like are spread among both temperate and virulent

phages (Supplementary Figure S8B and Table S3). Within the 39 genomes of >80 kb, 24 have a UvsX/RecA-like recombinase, whereas only one have a Rad52-like one. In contrast to Sak4, UvsX/RecA recombinases are significantly associated with virulent phages (P -value of $2 \cdot 10^{-6}$). The presence of a recombinase is therefore strongly correlated with the phage virulence or temperate character (Supplementary Figure S8B and Table S3). Overall, 57% of the genomes >20 kb, have now a recombinase.

Whether large genomes with no detected recombinase (Supplementary Table S6) are really devoid of recombinase function, or encode an undetected one remains an open question. Phage genomes with terminal repeats, if being really orphan of a recombinase gene are not expected to grow in a *recA* host. A way to screen for putative new families of recombinase in the future will consist in testing whether orphan phages with terminal repeats, especially those encoding the *recE*-like accessory gene, grow in *recA* hosts.

Temperate phages and mosaicism

As mentioned in the 'Introduction' section, a major difference between UvsX/RecA and a recombinase such as Red β lies in the fidelity and the stringency of the recombination process. Red β was shown 100-fold more efficient than was RecA for homeologous recombination between 22% divergent sequences (16). We proposed earlier that a manifestation of this reduced stringency in Red β -containing phages may be the existence of genome mosaicism characterized by the presence of dispersed and almost identical segments recently exchanged between divergent phages infecting the same host (16). In contrast, comparative genomics of virulent T4-like genomes (which contain UvsX) gives no evidence for such mosaicism (52). The fidelity of Sak4-like recombinases is not known at present, but our previous study on mosaicism reveals that the nine *Staphylococcus aureus* phages (16) encompassing a Sak4-like recombinase (namely phages 187, 69, 77, 96, ROSA, 71, 55, 29 and 52A) have a high density of mosaics (an average of 10 per comparison), as compared to the genomes encompassing a Rad52-like protein (namely phages 53, 85, 42e, 37, EW, 88, 92, X2; 6.4 mosaics per comparison in average). Accordingly, Sak4 recombinases might behave similarly to Red β towards homeologous recombination. This is unexpected, given the high fidelity of RecA, which belongs to the same superfamily as Sak4.

Hence, two members of the Rad51 superfamily, Sak4 and UvsX/RecA may be associated with very distinct genomic structures related to the fidelity of the recombination process and existence of mosaicism. Comparing the capacity of Sak4 and UvsX proteins to promote homeologous recombination *in vivo* will tell whether these proteins effectively behave differently, and which domain or sequence motifs contribute to the distinct features. Along the same lines, it will be interesting to further probe whether the various Gp2.5, quite evenly distributed among virulent and temperate phages, also play distinct roles with respect to mosaicism.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Raphael Leplae for providing phage classification data and to Johannes Söding for providing the latest version of the HHsearch algorithm. We thank Colin Tinsley and Xavier Veaute for critical reading of the article and the two anonymous referees for their helpful comments and suggestions.

Conflict of interest statement. None declared.

REFERENCES

- Desnues, C., Rodriguez-Brito, B., Rayhawk, S., Kelley, S., Tran, T., Haynes, M., Liu, H., Furlan, M., Wegley, L., Chau, B. *et al.* (2008) Biodiversity and biogeography of phages in modern stromatolites and thrombolites. *Nature*, **452**, 340–343.
- Dinsdale, E.A., Edwards, R.A., Hall, D., Angly, F., Breitbart, M., Brulc, J.M., Furlan, M., Desnues, C., Haynes, M., Li, L. *et al.* (2008) Functional metagenomic profiling of nine biomes. *Nature*, **452**, 629–632.
- Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M., Edwards, R.A., Felts, B., Rayhawk, S. *et al.* (2007) Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. *Appl. Environ. Microbiol.*, **73**, 7059–7066.
- Schoenfeld, T., Patterson, M., Richardson, P.M., Wommack, K.E., Young, M. and Mead, D. (2008) Assembly of viral metagenomes from Yellowstone hot springs. *Appl. Environ. Microbiol.*, **74**, 4164–4174.
- Kwan, T., Liu, J., DuBow, M., Gros, P. and Pelletier, J. (2005) The complete genomes and proteomes of 27 *Staphylococcus aureus* bacteriophages. *Proc. Natl Acad. Sci. USA*, **102**, 5174–5179.
- Kwan, T., Liu, J., Dubow, M., Gros, P. and Pelletier, J. (2006) Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. *J. Bacteriol.*, **188**, 1184–1187.
- Pedulla, M.L., Ford, M.E., Houtz, J.M., Karthikeyan, T., Wadsworth, C., Lewis, J.A., Jacobs-Sera, D., Falbo, J., Gross, J., Pannunzio, N.R. *et al.* (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell*, **113**, 171–182.
- Bossi, L., Fuentes, J.A., Mora, G. and Figueroa-Bossi, N. (2003) Prophage contribution to bacterial population dynamics. *J. Bacteriol.*, **185**, 6467–6471.
- Brown, S.P., Le Chat, L., De Paepe, M. and Taddei, F. (2006) Ecology of microbial invasions: amplification allows virus carriers to invade more rapidly when rare. *Curr. Biol.*, **16**, 2048–2052.
- Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., Kettler, G., Sullivan, M.B., Steen, R., Hess, W.R. *et al.* (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature*, **449**, 83–86.
- Pal, C., Macia, M.D., Oliver, A., Schachar, I. and Buckling, A. (2007) Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*, **450**, 1079–1081.
- Leplae, R., Hebrant, A., Wodak, S.J. and Toussaint, A. (2004) ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic Acids Res.*, **32**, D45–D49.
- Leplae, R., Lima-Mendez, G. and Toussaint, A. (2010) ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res.*, **38**, D57–D61.
- Lima-Mendez, G., Toussaint, A. and Leplae, R. (2007) Analysis of the phage sequence space: the benefit of structured information. *Virology*, **365**, 241–249.
- Drake, J.W., Charlesworth, B., Charlesworth, D. and Crow, J.F. (1998) Rates of spontaneous mutation. *Genetics*, **148**, 1667–1686.
- Martinsohn, J.T., Radman, M. and Petit, M.A. (2008) The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genet.*, **4**, e1000065.
- Hatfull, G.F. (2008) Bacteriophage genomics. *Curr. Opin. Microbiol.*, **11**, 447–453.
- Stahl, M.M., Thomason, L., Poteete, A.R., Tarkowski, T., Kuzminov, A. and Stahl, F.W. (1997) Annealing vs. invasion in phage lambda recombination. *Genetics*, **147**, 961–977.
- Fenton, A.C. and Poteete, A.R. (1984) Genetic analysis of the erf region of the bacteriophage P22 chromosome. *Virology*, **134**, 148–160.
- Mosig, G., Gewin, J., Luder, A., Colowick, N. and Vo, D. (2001) Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *Proc. Natl Acad. Sci. USA*, **98**, 8306–8311.
- Hollis, T., Stattel, J.M., Walther, D.S., Richardson, C.C. and Ellenberger, T. (2001) Structure of the gene 2.5 protein, a single-stranded DNA binding protein encoded by bacteriophage T7. *Proc. Natl Acad. Sci. USA*, **98**, 9557–9562.
- Araki, H. and Ogawa, H. (1981) The participation of T7 DNA-binding protein in T7 genetic recombination. *Virology*, **111**, 509–515.
- Kong, D. and Richardson, C.C. (1996) Single-stranded DNA binding protein and DNA helicase of bacteriophage T7 mediate homologous DNA strand exchange. *EMBO J.*, **15**, 2010–2019.
- Iyer, L.M., Koonin, E.V. and Aravind, L. (2002) Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics*, **3**, 8.
- Passy, S.I., Yu, X., Li, Z., Radding, C.M. and Egelman, E.H. (1999) Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. *Proc. Natl Acad. Sci. USA*, **96**, 4279–4284.
- Ploquin, M., Bransi, A., Paquet, E.R., Stasiak, A.Z., Stasiak, A., Yu, X., Cieslinska, A.M., Egelman, E.H., Moineau, S. and Masson, J.Y. (2008) Functional and structural basis for a bacteriophage homolog of human RAD52. *Curr. Biol.*, **18**, 1142–1146.
- Poteete, A.R., Sauer, R.T. and Hendrix, R.W. (1983) Domain structure and quaternary organization of the bacteriophage P22 Erf protein. *J. Mol. Biol.*, **171**, 401–418.
- 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.
- Muyers, J.P., Zhang, Y., Testa, G. and Stewart, A.F. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res.*, **27**, 1555–1557.
- Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G. and Court, D.L. (2000) An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **97**, 5978–5983.
- Ellis, H.M., Yu, D., DiTizio, T. and Court, D.L. (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl Acad. Sci. USA*, **98**, 6742–6746.
- Erler, A., Wegmann, S., Elie-Caille, C., Bradshaw, C.R., Maresca, M., Seidel, R., Habermann, B., Muller, D.J. and Stewart, A.F. (2009) Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. *J. Mol. Biol.*, **391**, 586–598.
- Bouchard, J.D. and Moineau, S. (2004) Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. *J. Bacteriol.*, **186**, 3649–3652.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Soding, J. (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics*, **21**, 951–960.
- Sali, A. and Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
- Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature*, **356**, 83–85.
- Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins*, **17**, 355–362.

39. Clamp, M., Cuff, J., Searle, S.M. and Barton, G.J. (2004) The Jalview Java alignment editor. *Bioinformatics*, **20**, 426–427.
40. DeLano, W.L. (2002) *The PyMOL Molecular Graphics System*. USA DeLano Scientific, San Carlos, CA.
41. Pupko, T., Bell, R.E., Mayrose, I., Glaser, F. and Ben-Tal, N. (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, **18**(Suppl 1), S71–S77.
42. Wolff, E., Kim, M., Hu, K., Yang, H. and Miller, J.H. (2004) Polymerases leave fingerprints: analysis of the mutational spectrum in *Escherichia coli* rpoB to assess the role of polymerase IV in spontaneous mutation. *J. Bacteriol.*, **186**, 2900–2905.
43. Kagawa, W., Kurumizaka, H., Ishitani, R., Fukai, S., Nureki, O., Shibata, T. and Yokoyama, S. (2002) Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. *Mol. Cell*, **10**, 359–371.
44. Singleton, M.R., Wentzell, L.M., Liu, Y., West, S.C. and Wigley, D.B. (2002) Structure of the single-strand annealing domain of human RAD52 protein. *Proc. Natl Acad. Sci. USA*, **99**, 13492–13497.
45. Stasiak, A.Z., Larquet, E., Stasiak, A., Muller, S., Engel, A., Van Dyck, E., West, S.C. and Egelman, E.H. (2000) The human Rad52 protein exists as a heptameric ring. *Curr. Biol.*, **10**, 337–340.
46. Akiba, T., Ishii, N., Rashid, N., Morikawa, M., Imanaka, T. and Harata, K. (2005) Structure of RadB recombinase from a hyperthermophilic archaeon, *Thermococcus kodakaraensis* KOD1: an implication for the formation of a near-7-fold helical assembly. *Nucleic Acids Res.*, **33**, 3412–3423.
47. Liu, N., Lamerdin, J.E., Tebbs, R.S., Schild, D., Tucker, J.D., Shen, M.R., Brookman, K.W., Siciliano, M.J., Walter, C.A., Fan, W. *et al.* (1998) XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. *Mol. Cell*, **1**, 783–793.
48. Mazin, A.V. and Kowalczykowski, S.C. (1998) The function of the secondary DNA-binding site of RecA protein during DNA strand exchange. *EMBO J.*, **17**, 1161–1168.
49. Rezende, L.F., Willcox, S., Griffith, J.D. and Richardson, C.C. (2003) A single-stranded DNA-binding protein of bacteriophage T7 defective in DNA annealing. *J. Biol. Chem.*, **278**, 29098–29105.
50. Datta, S., Costantino, N., Zhou, X. and Court, D.L. (2008) Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc. Natl Acad. Sci. USA*, **105**, 1626–1631.
51. Nagaraju, G., Hartlerode, A., Kwok, A., Chandramouly, G. and Scully, R. (2009) XRCC2 and XRCC3 regulate the balance between short- and long-tract gene conversion between sister chromatids. *Mol. Cell Biol.*, **29**, 4283–4294.
52. Comeau, A.M., Bertrand, C., Letarov, A., Tetart, F. and Krisch, H.M. (2007) Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology*, **362**, 384–396.
53. Breitkreutz, B.J., Stark, C. and Tyers, M. (2002) Osprey: a network visualization system. *Genome Biol.*, **3**, R22.

Supplementary data

8 Supplementary Figures followed by 7 Supplementary Tables

Supplementary Figures Legends

Figure S1

A complete set of analyses at different probability thresholds for 3 iterations of profile-profile comparisons is presented for the five recombinase families. The precision (red) and recall (blue) values for every recombinase family show that two iterations with a probability threshold of 90 % provide the best compromise. The barplot shown for the members of the Rad52 superfamily represents the number of interconnections detected between Red β , Sak and Erf associated profiles. At 85%, false positives lead to rapid divergence at iteration 3 for ERF and Gp2.5. At 95% several interesting hits would have been missed for Sak. At 90 %, there are 5 ERF false positives that can easily be recognized because they only match the unfolded C-terminal part of Erf with no match within the folded domain. Just below 90% (with a probability score of 89.6 %), 5 Gp2.5 could also be confidently recognized and were added to the set in Table S3. The highest number of interconnections is observed at the 85% threshold and remains rather stable below. When Red β profiles are used as queries only one connexion is found with the other superfamilies at any threshold. In contrast up to 11 connexions to Red β are found from Sak and Erf profiles at a threshold of 85 %.

Figure S2

Model evaluation: distribution of Verify3D scores along the sequence. For each model or Xray structures (indicated by their PDB code), Verify3D scores were plotted relative to the residue positions. Residue numbering corresponds to the PDB index for the Xray structures and to the FASTA file index for the models. We checked that for negative Verify3D scores, the corresponding positions were generally included within loops or disordered regions that cannot be faithfully modelled. The plots were generated using R.

Figure S3

Sequence alignment between Gp2.5-like protein of phage ϕ 12 and the template Gp2.5 from phage T7. Secondary structure elements as observed in the structure of T7 Gp2.5 are presented on top of the sequences and colored with respect to the structure in Figure 4B.

Figure S4

Distribution of the minimal gene separation between a recombinase and a marker of the DNA replication module. The genetic context surrounding every detected recombinase was analyzed using several well annotated protein families as markers of the DNA replication module. These ACLAME families were reliably built in the original ACLAME annotation process and are reported in Table S4 (Leplae et al, NAR 2004;32:D45-9.). Out of the 167 genomes-containing recombinases, only 14 genomes were devoid of any marker (8 %). For the remaining 153 genomes, 87% or 95% contained the predicted recombinase at less than 10 or 20 genes from one marker, respectively. The only notable exception is a member of the Red β superfamily found 51 genes further one of the markers. The corresponding phage has an especially large genome probably resulting from a fusion between two phage genomes (Thomas-JA et al, Virology. 2007 368(2):405-21). It is the only one containing two recombinases (a Sak and a Red β). Interestingly, in this phage, the other Sak-like recombinase is found within the DNA replication module.

Figure S5

Gene architecture in phage ϕ 12 results from a fusion between genomes either related to ETA, PVL, SLT and ϕ 77 or to BPP-1 and APSE-1. The position of the Gp2.5 remote homolog indicated in light pink matches the expected position for a recombinase. Some putative functions have been labelled by a short term, ini, initiator; hel loader, DNA-C type helicase loader; ssb, single strand binding; RecE, 5'-3' exonuclease RecE; int, phage integrase; repr, repressor; DNA pol, DNA polymerase (Pol I-type); dut, dUTPase; endo, endonuclease; SF2 hel, superfamily 2 helicase; PriA, DNA primase; ninB, ninB recombination system functionally related to RecFOR. In this Figure, phages ETA, PVL, SLT and ϕ 77 share a number of homologous genes highlighted by their common colors.

Figure S6

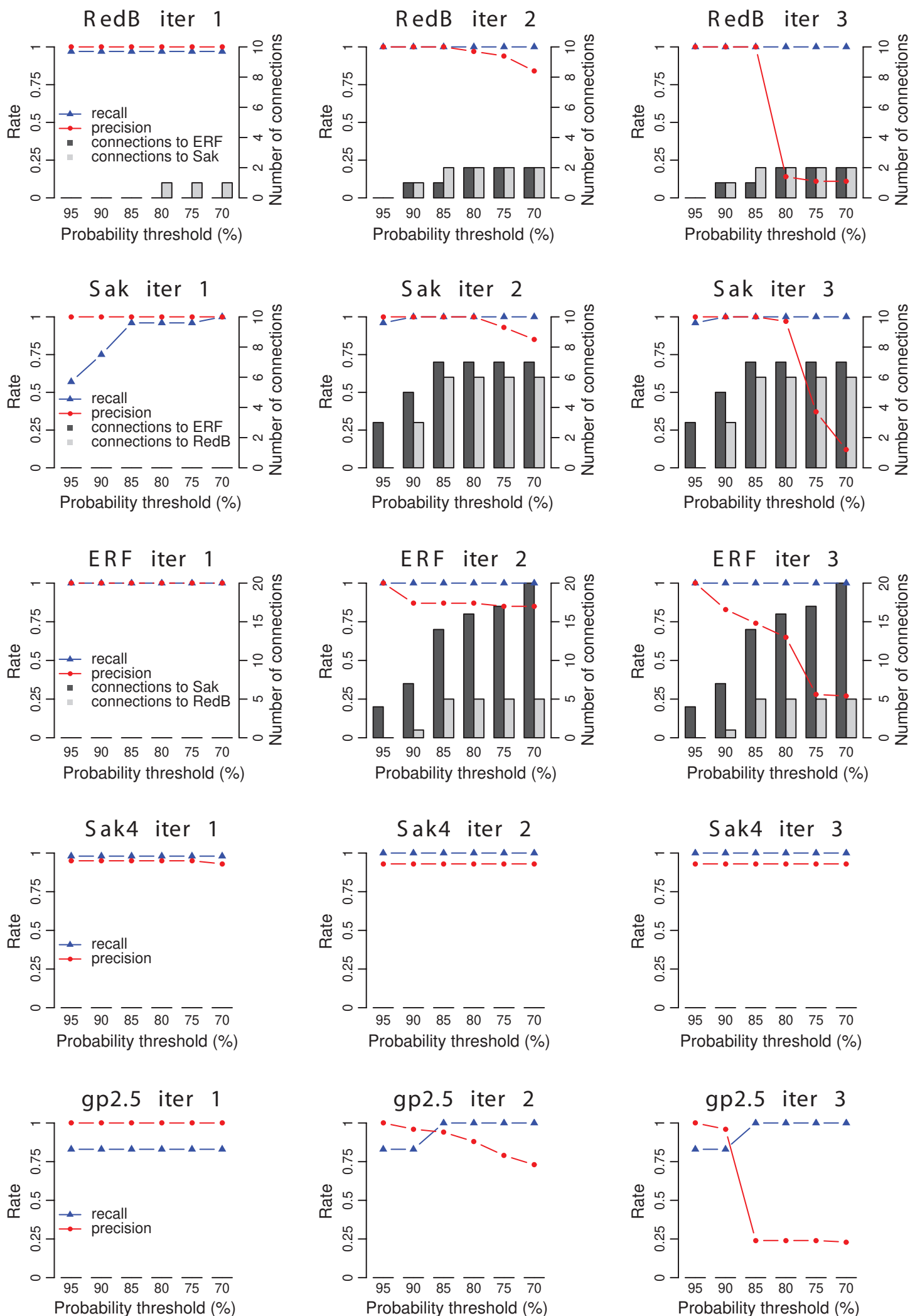
Estimation of the level of synthesis and stability of the various recombinases after arabinose induced expression, by *in vivo* ^{35}S methionine labelling. Cells grown exponentially up to OD 0.5 in M63 minimal medium glucose at 30°C were pelleted and resuspended in M63 supplemented with 0.2% arabinose instead of glucose, to induce recombinase synthesis, during 30 minutes. Induced proteins were pulse-labelled 2 minutes with 20 μCi of ^{35}S methionine per mL (lanes 0'), and chased by the addition of 0.02% cold methionine for 30 more minutes (lanes 30'). Labelled samples were analysed by 12% PAGE, with equal amounts of cells loaded in each lane. Prestained size markers were run next to the radioactive samples, and their positions is reported on the side. Induced proteins are marked with an arrow (Figure S6, upper part). The atypical migration of Sak4 and T7 Gp2.5 proteins, relative to their mass, was confirmed with flagged versions of these proteins (not shown). Estimation of radioactivity amounts was measured on a Storm with the ImageQuant software, and is reported relative to Red β . The complete experiment was repeated twice, and the two estimations are reported (Figure S6, lower part, first lane). To estimate relative protein amounts (FigS6, third lane), the radioactivity signal was divided by the ratio between the number of methionines of each protein, and that of Red β (FigS6, second lane). The stability of all proteins was found to be close to 100% on a 30 minutes period of chase. A parallel experiment with flagged versions of all recombinase proteins was also realised (not shown). It confirmed the similar expression levels of Red β , UvsX and Sak4, and a higher and more fluctuating level for the T7 Gp2.5 protein. However, the flagged Φ 12 Gp2.5 could not be detected with the anti-flag antibodies (we suspect it was masked by an aspecific band).

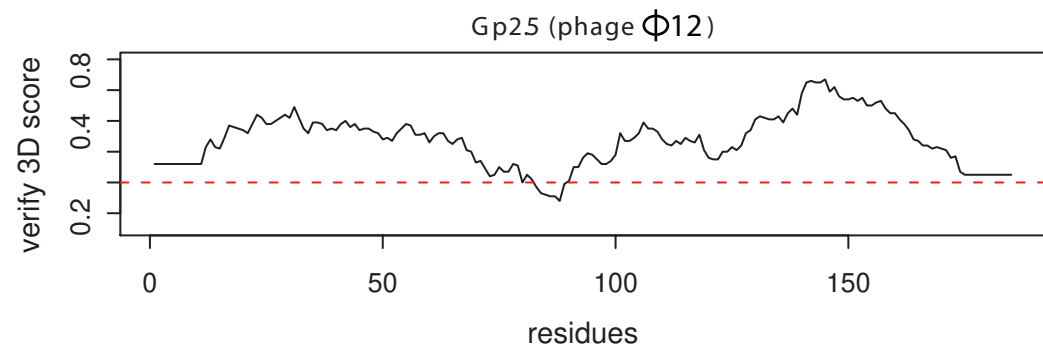
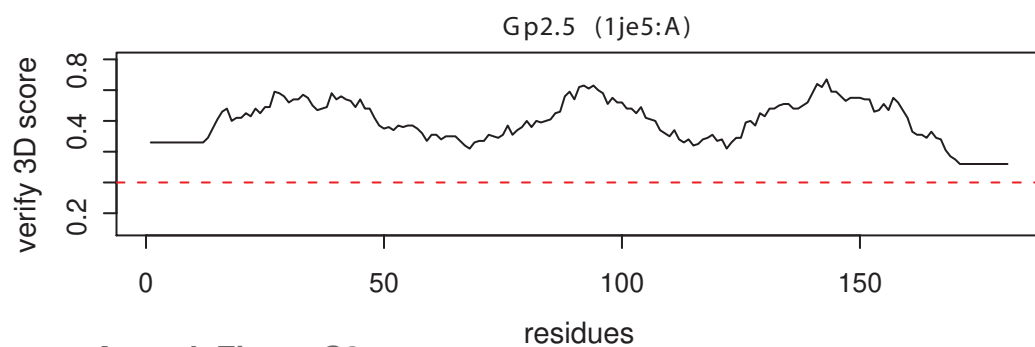
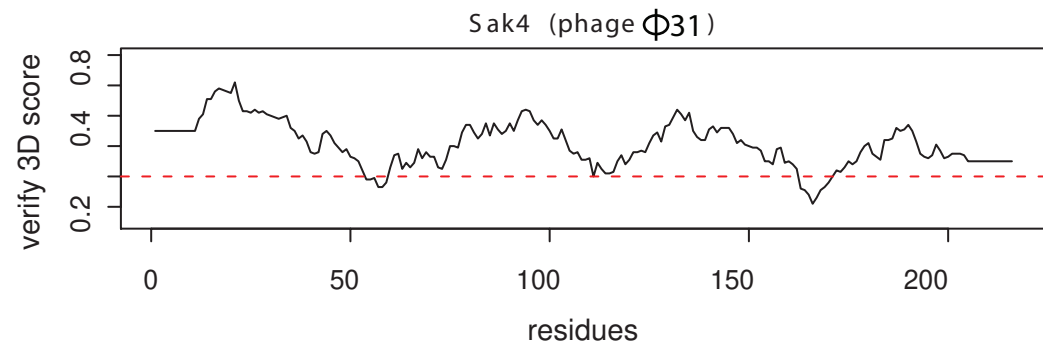
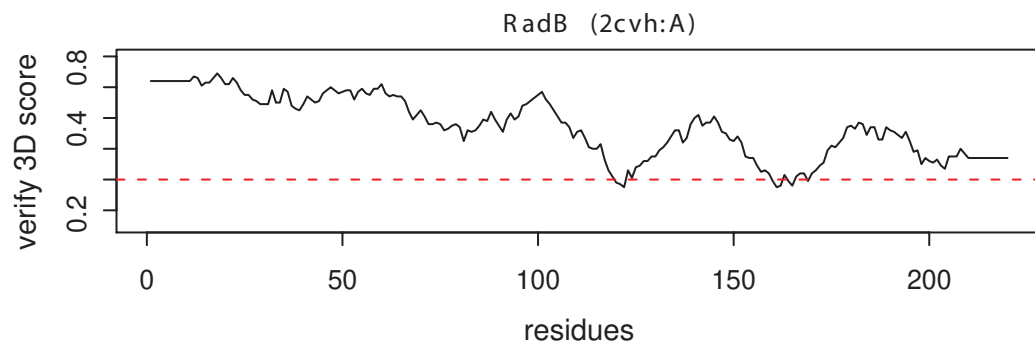
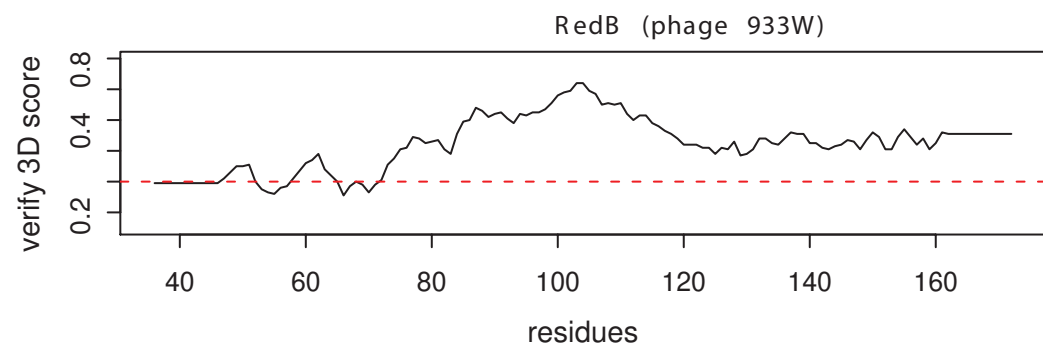
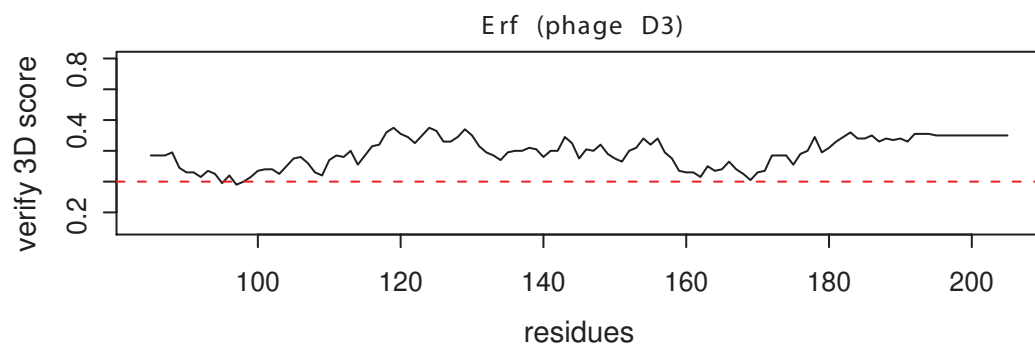
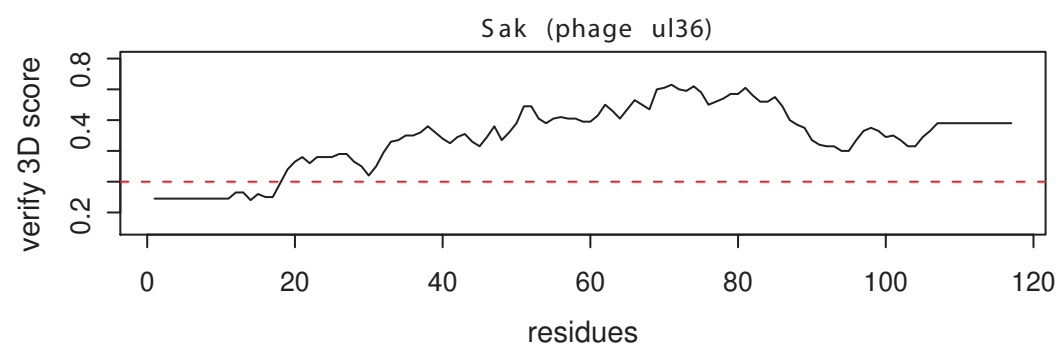
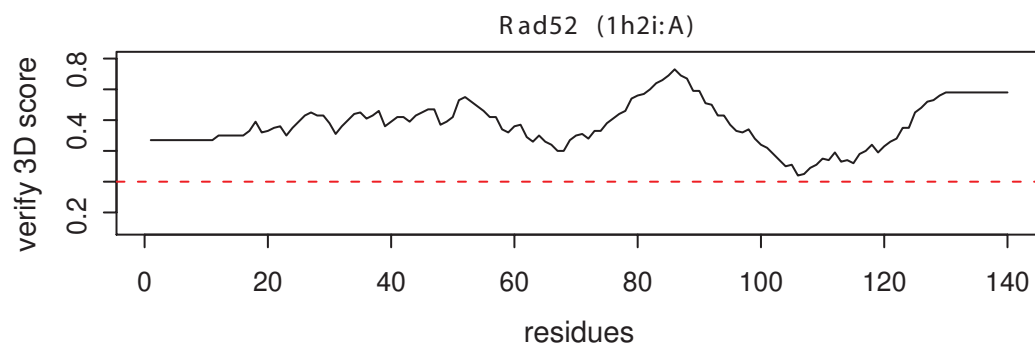
Figure S7

Sequence identity matrices between the various members of the detected recombinase superfamilies. The identity was calculated from the HHsearch alignment used in global mode. A color grade from blue to red reports from highly divergent to highly similar proteins within a given superfamily. Name of the encompassing genome is added along the axes.

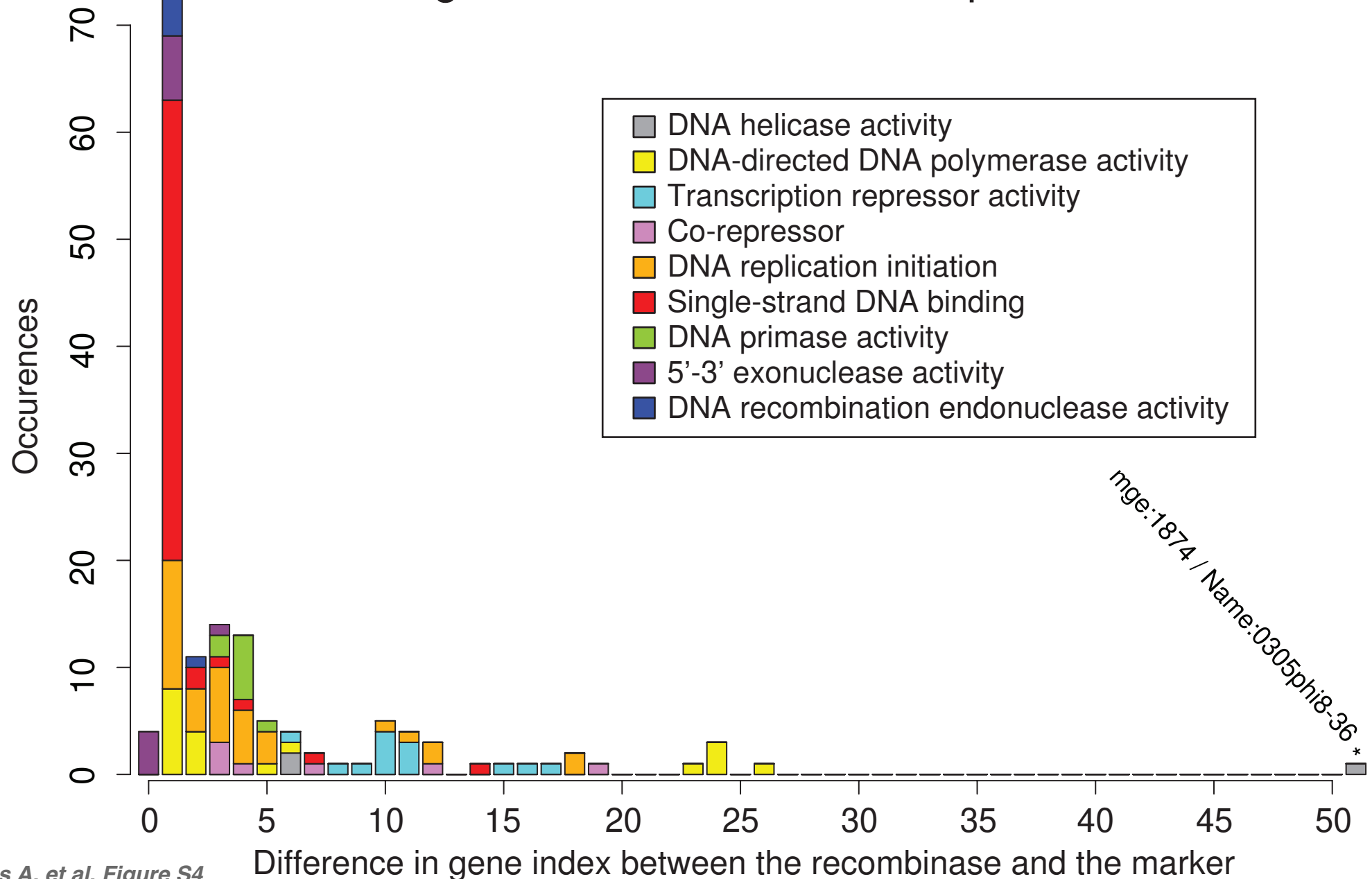
Figure S8

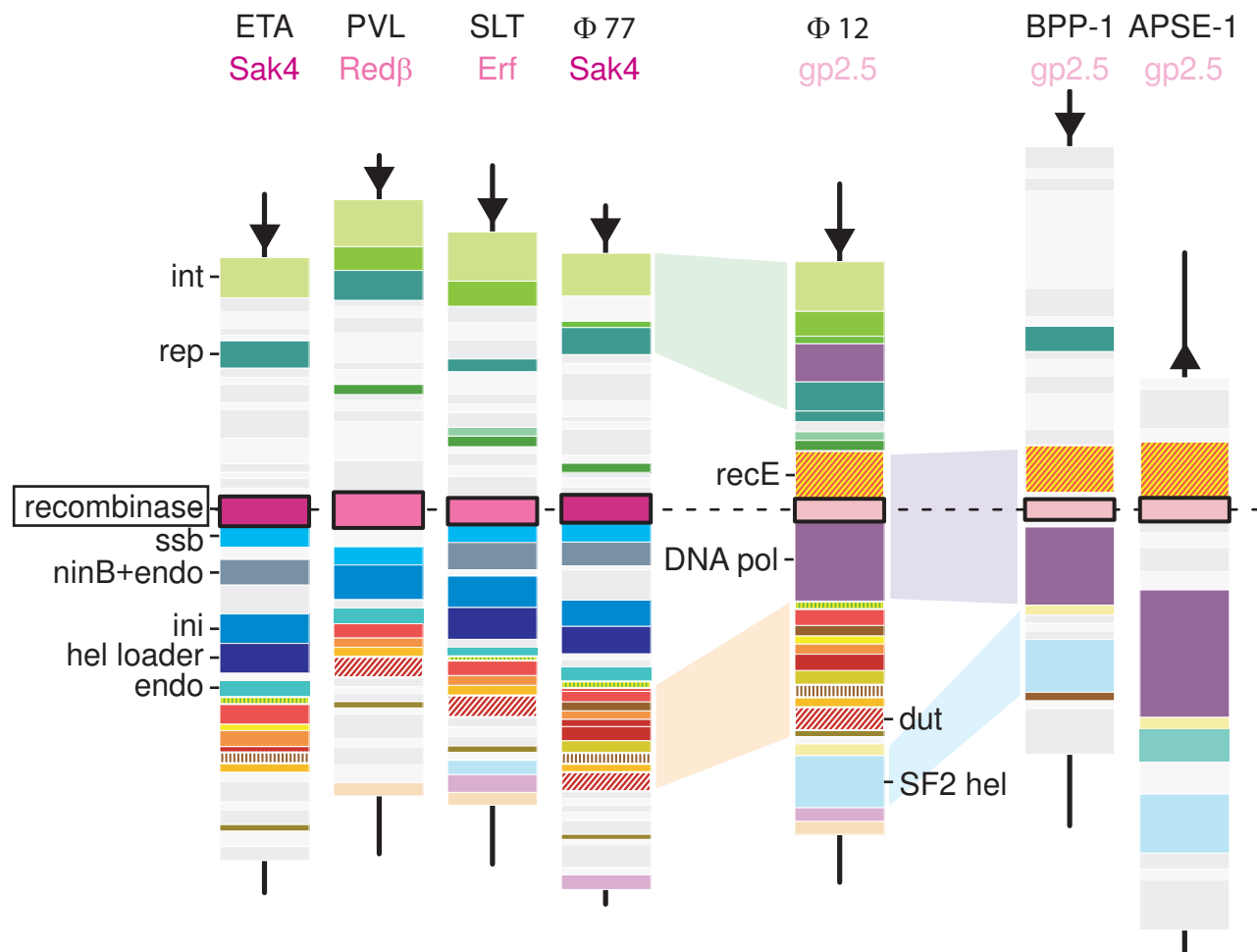
(A) Venn diagram showing the co-occurrence of the five recombinase families among the bacteriophage genomes. Only four genomes contain more than one recombinase, always in association with UvsX. The genome containing three recombinases is discussed in the legend of Figure S5. (B) Distribution of the five recombinase families with respect to the bacteriophage lifestyle. Virulent bacteriophages almost exclusively contain UvsX or Gp2.5 whereas temperate bacteriophages rather contain Red β , Sak, Erf, Sak4 or Gp2.5 recombinases.

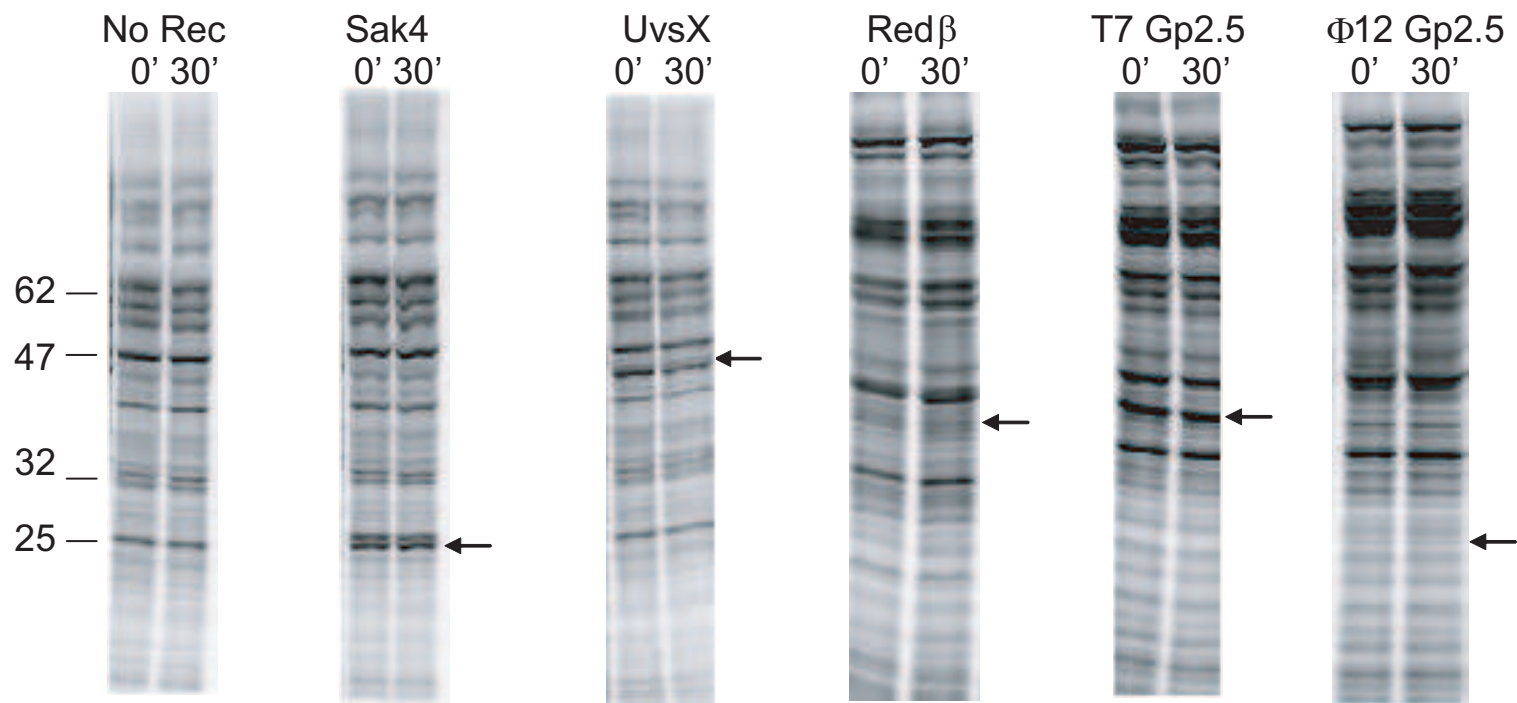




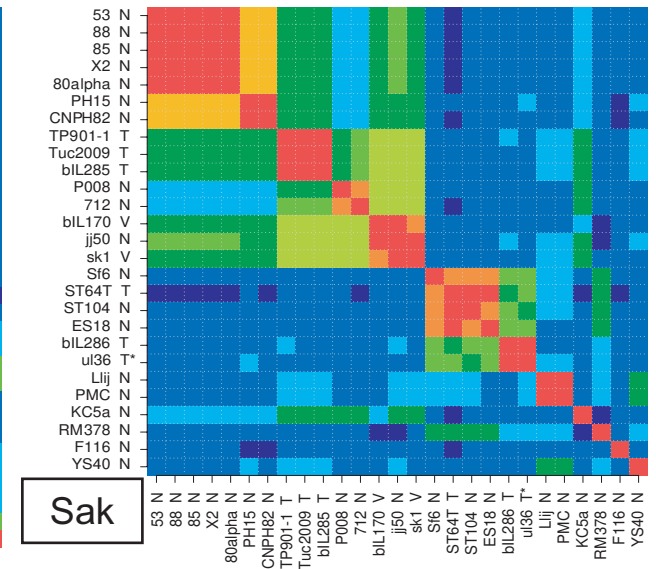
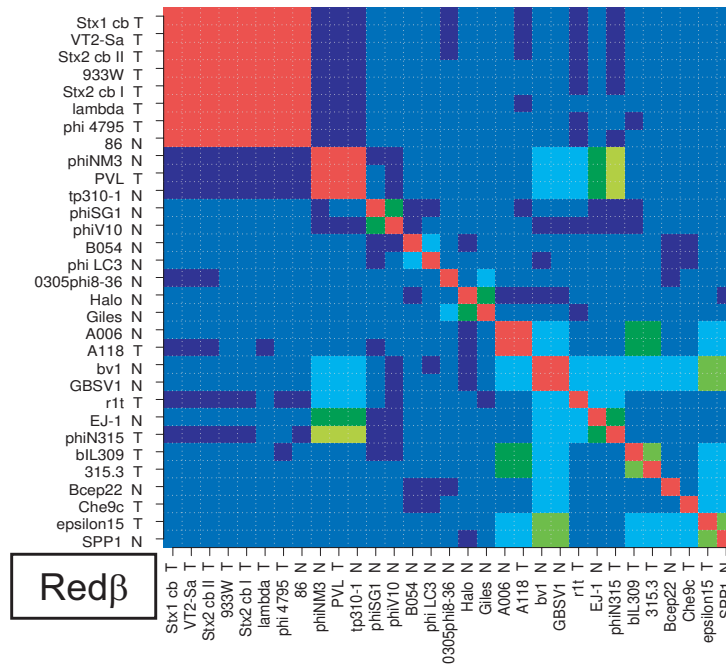
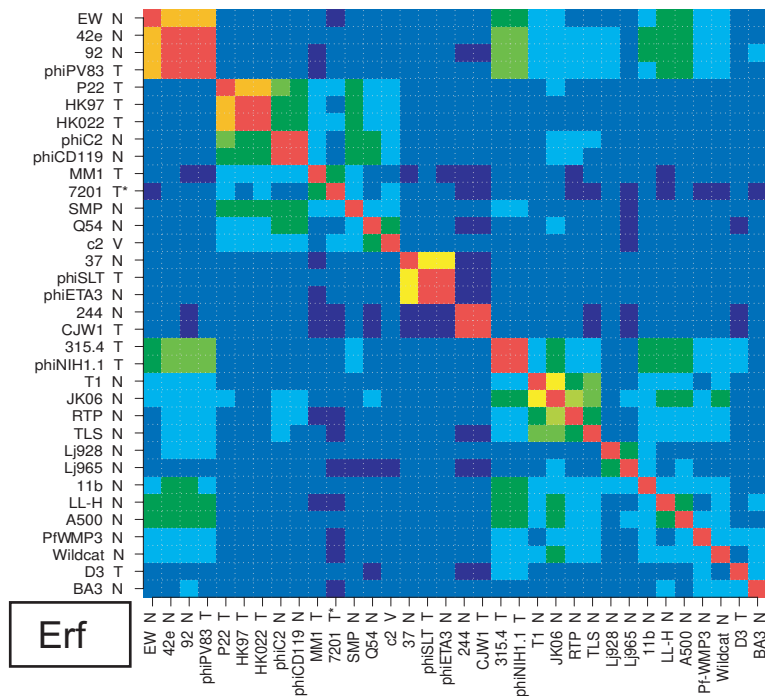
Distribution of the minimal distances between 153 recombinases and one of the gene marker within DNA replication module



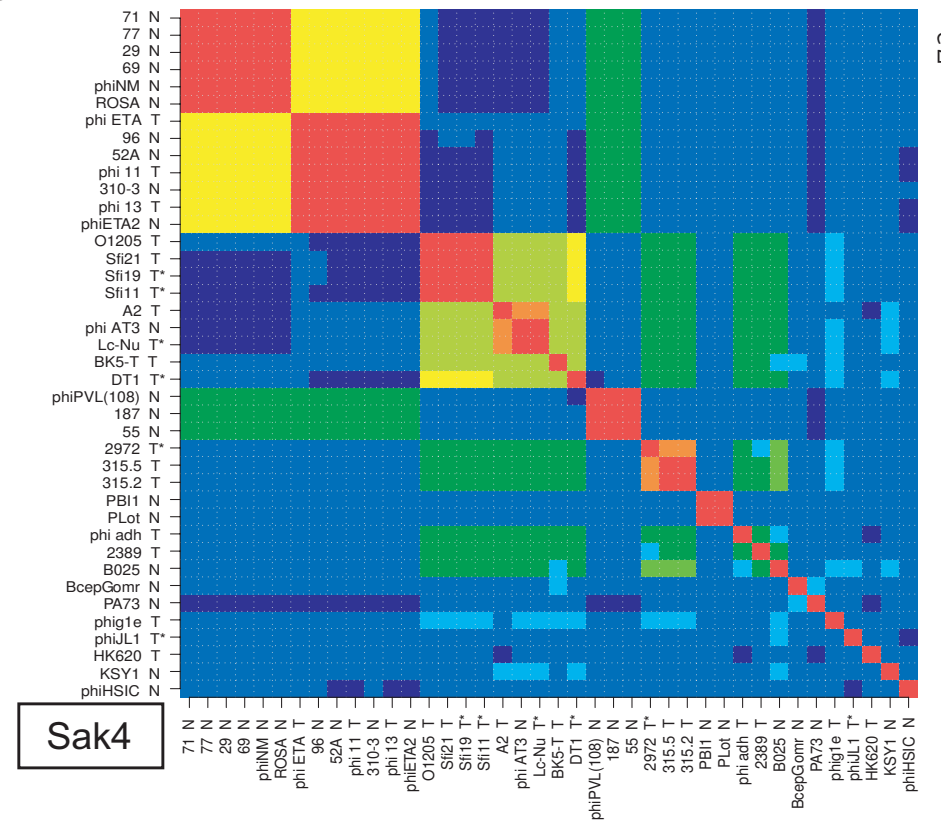
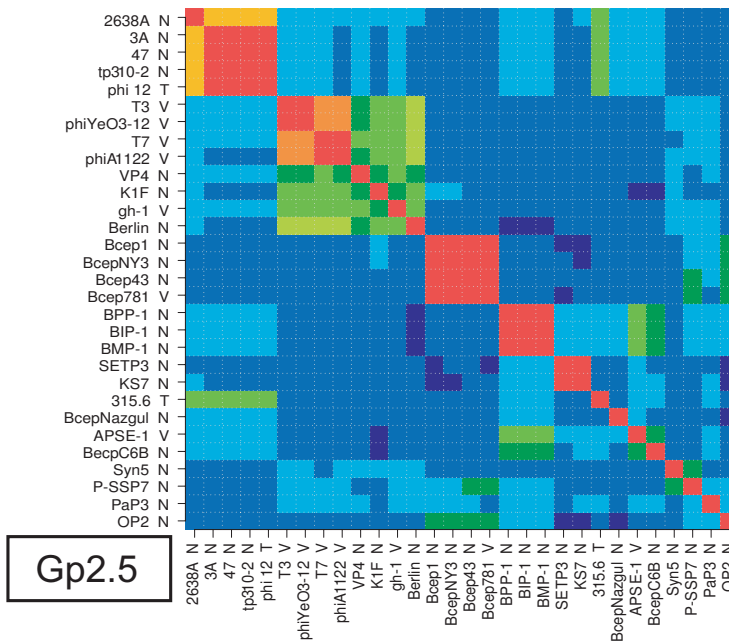
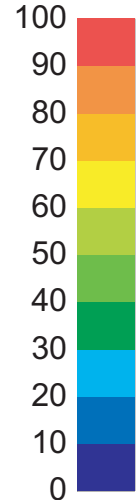




Induced protein	Sak4 (26.6 kDa)	UvsX (44 kDa)	Redβ (29,7 kDa)	T7 Gp2.5 (25.7 kDa)	Φ12 Gp2.5 (20.1 kDa)
Radioactivity, relative to Redβ	3.4 - 4.2	3 - 3.5	1	3.6 - 6.4	0.5
Nb of methionines	16	15	7	5	2
Amount of protein, relative to Redβ	1.5 - 1.8	1.4 - 1.6	1	5 - 9	1.8

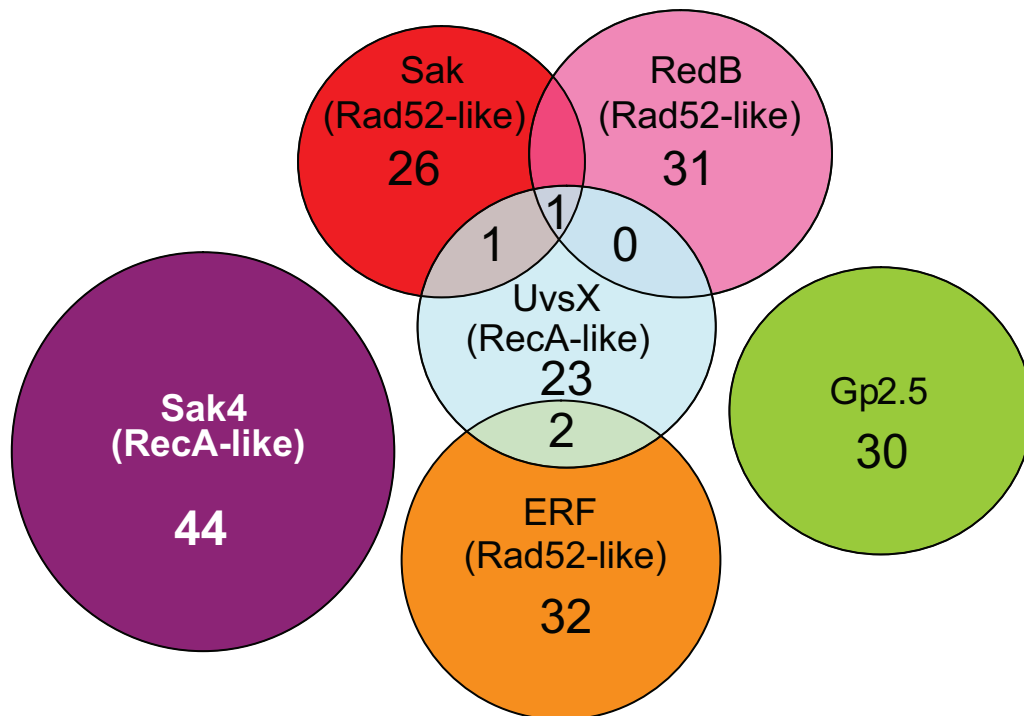


% identity



A

Venn diagram reporting the number of genomes containing one of the recombinases types

**B**

Distribution of the different recombinases with respect to bacteriophage lifestyles



Table S1

Evaluation scores for the models from Verify3D and Prosa.

models	Verify3D scores	Prosa Z-scores
Sak	0.29	-3.87
ERF	0.19	-3.96
Red β	0.24	-3.13
Sak4	0.21	-4.12
gp2.5	0.27	-5.97
Xray		
Rad52 (1h2i:A)	0.37	-5.69
RadB (2cvh:A)	0.34	-7.95
gp2.5 (1je5:A)	0.39	-4.44

Table S2
Plasmids and strains

	Relevant genotype	Source/construction
Plasmids		
pKD46	Para: <i>redβ</i>	(Datsenko and Wanner 2000)
pMAP188	Para: <i>uvrX</i>	<i>uvrX</i> gene of T4 cloned between <i>Eco</i> RI and <i>Nco</i> I sites of pKD46
pMAP189	Para: <i>sak4</i>	<i>sak4</i> gene of PA73 cloned between <i>Eco</i> RI and <i>Nco</i> I sites of pKD46
pMAP190	Para: <i>gp2.5</i>	<i>gp2.5</i> gene of T7 cloned between <i>Eco</i> RI and <i>Nco</i> I sites of pKD46
pMAP191	Para: <i>φ12gp11</i>	<i>gp2.5</i> -like gene of <i>φ12</i> , cloned between <i>Eco</i> RI and <i>Nco</i> I sites of pKD46
pJA157	No recombinase	pKD46 Δ <i>Eco</i> RI - <i>Nco</i> I
pJA151	Para: <i>φ12gp11</i> K79C	Site-directed mutagenesis of pMAP191
pJA153	Para: <i>φ12gp11</i> R83C	Site-directed mutagenesis of pMAP191
pJA148	Para: <i>redβ</i> R161C	Site-directed mutagenesis of pKD46
pJA168	Para: <i>uvrX</i> K66A	Site-directed mutagenesis of pMAP188
pJA169	Para: <i>sak4</i> K30A	Site-directed mutagenesis of pMAP189
<i>E. coli</i> strains		
MAC1031	AB1157 <i>recA</i> 306 <i>srl</i> :Tn10	P1 transduction of the <i>recA</i> 306 allele into a Nal ^R Strep ^R derivative of AB1157
MAC1042	pKD46	transformation of pKD46 into MAC1031
MAC1525	pMAP190	transformation of pMAP190 into MAC1031
MAC1527	pMAP188	transformation of pMAP188 into MAC1031
MAC1535	pMAP191	transformation of pMAP191 into MAC1031
MAC1544	pMAP189	transformation of pMAP189 into MAC1031
JAC157	pJA157	transformation of pJAC157 into MAC1031
JAC148	pJA148	transformation of pJA148 into MAC1031
JAC151	pJA151	transformation of pJA151 into MAC1031
JAC153	pJA153	transformation of pJA153 into MAC1031
JAC171	pJA169	transformation of pJA169 into MAC1031
JAC173	pJA168	transformation of pJA168 into MAC1031
<i>S. aureus</i> strain		
NCTC8325	<i>φ12</i> prophage	Pasteur institute
Bacteriophages		
T4 gp2	mutation in the <i>gp2</i> gene	B. Michel
T7		D'Herelle collection center, Canada
PA73	infects <i>P. aeruginosa</i>	D'Herelle collection center, Canada

Table S3

Detailed information for the genomes in which a recombinase has been detected. *See legend at the bottom*

This table can also be used at <http://biodev.extra.cea.fr/virfam/table.aspx> and the sequences and hmm profiles of every recombinase downloaded.

MGE ^a	Phage Name	nb cds ^b	vir / temp ^c	Host	Type ^d	Gene name	prot:vir ^e	Access Nb ^f	Length ^g	new/known	Presence of RecE
83	ul36 [3]***	61	ex-temperate	<i>Lactococcus lactis</i>	Sak	ORF252	3939	NP_663647	252	known	no
14	Tuc2009	56	temperate	<i>Lactococcus lactis</i>	Sak	orf14	714	NP_108691	207	new	no
1479	712	55	unknown	<i>Lactococcus lactis</i>	Sak	ORF35	93874	YP_764295	184	new	no
18	bIL170	64	virulent	<i>Lactococcus lactis</i>	Sak	e12	902	NP_047161	211	new	no
74	TP901-1	56	temperate	<i>Lactococcus lactis</i>	Sak	Orf11	3588	NP_112674	207	new	no
34	sk1	56	virulent	<i>Lactococcus lactis</i>	Sak	ORF35	1694	NP_044981	211	new	no
1487	jj50	49	unknown	<i>Lactococcus lactis</i>	Sak	ORF34	93978	YP_764348	189	new	no
20	bIL286	61	temperate	<i>Lactococcus lactis</i> ; <i>Lactococcus lactis</i> subsp. <i>lactis</i> Il1403	Sak	Orf14	994	NP_076648	252	known	no
1532	P008	58	unknown	<i>Lactococcus lactis</i> subsp. <i>lactis</i> Il1403	Sak	gp40	94900	YP_762551	184	new	no
19	bIL285	62	temperate	<i>Lactococcus lactis</i> subsp. <i>lactis</i>	Sak	Orf14	932	NP_076586	207	new	no
1502	KC5a	61	unknown	<i>Lactobacillus gasseri</i>	Sak	gp11	105481	YP_529846	267	new	no
1508	88	72	temperate	<i>Staphylococcus aureus</i>	Sak	Orf21	94454	YP_240716	207	new	no
1654	85	78	temperate	<i>Staphylococcus aureus</i>	Sak	Orf23	97178	YP_239777	207	new	no
1549	X2	77	temperate	<i>Staphylococcus aureus</i>	Sak	Orf20	95095	YP_240871	207	new	no
1613	53	79	temperate	<i>Staphylococcus aureus</i>	Sak	Orf24	96396	YP_239694	207	new	no
1858	80alpha	73	unknown	<i>Staphylococcus aureus</i>	Sak	e12	78860	YP_001285330	207	new	no
1679	PH15	68	unknown	<i>Staphylococcus epidermidis</i>	Sak	ph46	105294	YP_950708	208	new	no
1571	CNPH82	65	unknown	<i>Staphylococcus epidermidis</i>	Sak	cn42	107151	YP_950640	208	new	no
164	ST64T	65	temperate	<i>Salmonella typhimurium</i> ; <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i>	Sak	Orf235	9227	NP_720291	235	known	no
1509	ST104	63	temperate	<i>Salmonella typhimurium</i> DT104	Sak	Orf13	100946	YP_006369	235	known	no
1501	ES18	79	temperate	<i>Salmonella typhimurium</i>	Sak	gp49	105738	YP_224187	235	known	no
1636	PMC	104	unknown	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	Sak	gp61	105808	YP_655822	192	new	no
1661	Lij	100	unknown	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	Sak	gp60	102541	YP_655056	192	new	no
1556	Sf6	66	temperate	<i>Shigella flexneri</i>	Sak	gene 28	105392	NP_958204	235	known	no
1540	F116	70	temperate	<i>Pseudomonas aeruginosa</i>	Sak	p19	105567	YP_164283	251	new	no
123	RM 378	146	unknown	<i>Rhodothermus marinus</i>	Sak	p018	5819	NP_835605	164	known	no
1610**	YS40	170	unknown	<i>Thermus thermophilus</i>	Sak	ORF13	102777	YP_874026	191	known	no
1874**	0305phi8-36	246	unknown	<i>Bacillus thuringiensis</i>	Sak	ORF179	79766	YP_001429670	232	new	no
93	D3	95	temperate	<i>Pseudomonas aeruginosa</i>	ERF	Orf52	4386	NP_061548	275	known	no

MGE ^a	Phage Name	nb cds ^b	vir / temp ^c	Host	Type ^d	Gene name	prot:vir: ^e	Access Nb ^f	Length ^g	new/known	Presence of RecE
1602	37	70	temperate	<i>Staphylococcus aureus</i>	ERF	Orf18	96137	YP_240125	212	new	no
102	phiPV83	65	temperate	<i>Staphylococcus aureus</i>	ERF	p17	4673	NP_061607	212	known	no
1643	42e	79	temperate	<i>Staphylococcus aureus</i>	ERF	Orf20	96932	YP_239905	212	known	no
57	phiSLT	61	temperate	<i>Staphylococcus aureus</i>	ERF	p17	2660	NP_075480	216	new	no
1676	92	74	temperate	<i>Staphylococcus aureus</i>	ERF	Orf22	97483	YP_240790	212	known	no
1514	phiETA3	68	unknown	<i>Staphylococcus aureus</i>	ERF	gp16	105933	YP_001004345	216	new	no
1642	EW	77	temperate	<i>Staphylococcus aureus</i>	ERF	Orf021	96893	YP_240203	207	known	no
85	c2	39	virulent	<i>Lactococcus lactis</i>	ERF	p08	4055	NP_043534	179	known	no
1675	Q54	47	unknown	<i>Lactococcus lactis</i>	ERF	gp11	97394	YP_762580	197	known	no
1605	JK06	82	NA	<i>Escherichia coli</i> O157:H7	ERF	JK_79P	103245	YP_277519	130	known	yes
6	HK97	61	temperate	<i>Escherichia coli</i>	ERF	gp40	216	NP_037726	201	known	no
1593	RTP	75	NA	<i>Escherichia coli</i>	ERF	rtp48	104303	YP_398992	216	known	yes
41	HK022	57	temperate	<i>Escherichia coli</i>	ERF	p37	1918	NP_037690	201	known	no
1518	T1	78	virulent	<i>Escherichia coli</i>	ERF	p28	107660	YP_003917	226	known	yes
1872	TLS	87	unknown	<i>Escherichia coli</i> K12	ERF	RecT	79623	YP_001285543	220	known	yes
1646	11b	65	NA	<i>Flavobacterium</i> sp.	ERF	erf	101616	YP_112530	179	known	no
2007	BA3	47	unknown	<i>Thalassomonas loyana</i>	ERF	ERF	108325	YP_001552302	247	known	no
1671	Wildcat	148	unknown	<i>Mycobacterium;Mycobacterium smegmatis</i>	ERF	gp64	98996	YP_655929	211	known	no
1515**	244	142	unknown	<i>Mycobacterium;Mycobacterium smegmatis</i>	ERF	gp70	101739	YP_654825	240	known	yes
150**	CJW1	141	temperate	<i>Mycobacterium;Mycobacterium smegmatis</i>	ERF	gp70	7912	NP_817519	240	known	yes
1886	A500	63	unknown	<i>Listeria monocytogenes</i>	ERF	gp44	80943	YP_001468429	219	known	no
1566	phi CD119	79	unknown	<i>Clostridium difficile</i>	ERF	gp51	102310	YP_529602	202	known	no
1861	phiC2	82	unknown	<i>Clostridium difficile</i>	ERF	p65	79018	YP_001110780	202	known	no
1878	Pf-WMP3	41	unknown	<i>Phormidium foveolarum</i>	ERF	ORF20	80175	YP_001285785	153	known	yes
46	P22	72	temperate	<i>Salmonella typhimurium</i>	ERF	Erf	2071	NP_059596	205	known	no
104	MM1	53	temperate	<i>Streptococcus pneumoniae</i>	ERF	ORF9	4759	NP_150142	235	known	no
61	PhiNIH1.1	55	temperate	<i>Streptococcus pyogenes</i>	ERF	ERF	3012	NP_438125	224	known	no
176	315.4	64	temperate	<i>Streptococcus pyogenes</i> MGAS315	ERF	ERF	9795	NP_795557	224	known	no
105	7201 [6]***	46	ex-temperate	<i>Streptococcus thermophilus</i>	ERF	Orf7	4811	NP_038308	218	known	no
1578	SMP	48	unknown	<i>Streptococcus suis</i>	ERF	gp21	95781	YP_950571	231	known	no
1598	Lj965	46	temperate	<i>Lactobacillus johnsonii</i>	ERF	ERF	106575	NP_958570	278	known	no
1559	Lj928	50	temperate	<i>Lactobacillus johnsonii</i>	ERF	ERF	99520	NP_958522	261	known	no
1873	LL-H	51	unknown	<i>Lactobacillus delbrueckii</i> subsp. <i>lactis</i>	ERF	ORF178B	79697	YP_001285915	178	known	no
66	933W	80	temperate	<i>Escherichia coli</i> O157:H7 EDL933	RedB	Bet prot	3258	NP_049474	261	known	yes
1471	86	81	unknown	<i>Escherichia coli</i>	RedB	gp52	104420	YP_794099	261	known	yes

MGE ^a	Phage Name	nb cds ^b	vir / temp ^c	Host	Type ^d	Gene name	prot:vir: ^e	Access Nb ^f	Length ^g	new/known	Presence of RecE
70	lambda	71	temperate	<i>Escherichia coli</i>	RedB	bet	3455	NP_040617	261	known	yes
16	VT2-Sa	83	temperate	<i>Escherichia coli</i> O157:H7 EDL933; <i>Escherichia coli</i> O157:H7	RedB	bet	778	NP_050513	261	known	yes
1564	phiV10	55	unknown	<i>Escherichia coli</i> O157:H7	RedB	p37	95355	YP_512292	586	known	yes
157	phi 4795	85	temperate	<i>Escherichia coli</i>	RedB	bet	93578	YP_001449250	261	known	yes
59	Stx2 converting bacteriophage I	166	temperate	<i>Escherichia coli</i> O157:H7 EDL933; <i>Escherichia coli</i> O157:H7	RedB	Bet prot	2867	NP_612983	261	known	no
180	Stx2 converting bacteriophage II	170	temperate	<i>Escherichia coli</i> ; <i>Escherichia coli</i> O157:H7	RedB	Bet prot	10221	NP_859351	261	known	no
179	Stx1 converting bacteriophage	167	temperate	<i>Escherichia coli</i> ; <i>Escherichia coli</i> O157:H7	RedB	Bet prot	10055	NP_859185	261	known	no
1891	tp310-1	59	unknown	<i>Staphylococcus aureus</i>	RedB	gp15	81113	YP_001429851	306	known	no
1591	phiNM3	65	unknown	<i>Staphylococcus aureus</i>	RedB	RecT	101269	YP_908809	306	known	no
101	PVL	62	temperate	<i>Staphylococcus aureus</i>	RedB	orf 43	4637	NP_058482	306	known	no
170	phiN315	65	temperate	<i>Staphylococcus aureus</i> subsp. <i>aureus</i> N315	RedB	SA1794	9488	NP_835535	306	known	no
1589	VWB	61	temperate	<i>Streptomyces venezuelae</i>	RedB	p55	98430	NP_958297	214	new	yes
175	315.3	52	temperate	<i>Streptococcus pyogenes</i> MGAS315	RedB	SpyM3_1134	9738	NP_795500	308	known	yes
152	Che9c	84	temperate	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	RedB	gp61	8157	NP_817738	353	known	yes
1611	Halo	65	unknown	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	RedB	gp43	99956	YP_655560	462	known	yes
2004	Giles	79	unknown	<i>Mycobacterium smegmatis</i>	RedB	gp53	108169	YP_001552382	482	known	no
2	A118	72	temperate	<i>Listeria monocytogenes</i>	RedB	orf48	87	NP_463513	271	known	yes
1876	B054	80	unknown	<i>Listeria monocytogenes</i>	RedB	gp50	80096	YP_001468754	261	known	no
1859	A006	62	unknown	<i>Listeria monocytogenes</i>	RedB	gp42	78889	YP_001468882	271	known	yes
125	SPP1	101	NA	<i>Bacillus subtilis</i>	RedB	p076	6027	NP_690727	287	known	no
1874**	0305phi8-36	246	unknown	<i>Bacillus thuringiensis</i>	RedB	p018	79945	YP_001429745	216	new	no
1877	bacteriophage bv1	54	unknown	<i>Bacillus</i> sp. 6k512	RedB	gp01	80151	YP_001425586	289	known	yes
1529	phi LC3	51	temperate	<i>Lactococcus lactis</i> subsp. <i>cremoris</i>	RedB	p10	94760	NP_996684	245	known	no
33	r1t	50	temperate	<i>Lactococcus lactis</i>	RedB	orf10	1617	NP_695038	308	known	no
21	bIL309	56	temperate	<i>Lactococcus lactis</i> ; <i>Lactococcus lactis</i> subsp. <i>lactis</i> II1403	RedB	RecT	1053	NP_076707	268	known	yes
1470	Bcep22	81	NA	<i>Burkholderia cepacia</i>	RedB	Orf13	93644	NP_944238	337	known	no
1567	GBSV1	54	unknown	<i>Geobacillus</i> sp. 6k51	RedB	gp02	95390	YP_764458	289	known	yes
1461	EJ-1	73	temperate	<i>Streptococcus pneumoniae</i>	RedB	p21	102946	NP_945260	323	known	yes
143	epsilon15	51	temperate	<i>Salmonella enterica</i>	RedB	recT	7351	NP_848242	340	known	yes
1463	phiSG1	47	unknown	<i>Sodalis glossinidius</i> str. 'morsitans'	RedB	SGPHI_0025	105534	YP_516203	408	known	yes
25	phi ETA	66	temperate	<i>Staphylococcus aureus</i>	Sak4	ORF17	1212	NP_510911	259	new	no
1588	71	72	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf16	95925	YP_240428	262	new	no
1666	52A	65	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf16	97291	YP_240651	259	new	no
1531	29	75	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf17	94832	YP_240583	262	new	no

MGE ^a	Phage Name	nb cds ^b	vir / temp ^c	Host	Type ^d	Gene name	prot:vir: ^e	Access Nb ^f	Length ^g	new/known	Presence of RecE
1557	187	77	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf13	106649	YP_239540	257	new	no
1607	69	69	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf15	96202	YP_239613	262	new	no
1476	77	69	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf12	100687	NP_958639	262	new	no
1475	55	77	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf17	93771	YP_240507	257	new	no
1612	ROSA	74	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf14	96289	YP_240351	262	new	no
1494	96	79	temperate	<i>Staphylococcus aureus</i>	Sak4	Orf14	94081	YP_240275	259	new	no
1662	phiNM	64	unknown	<i>Staphylococcus aureus</i>	Sak4	gp14	103905	YP_873963	262	new	no
1875	tp310-3	58	unknown	<i>Staphylococcus aureus</i>	Sak4	gp15	80017	YP_001429977	259	new	no
1581	phiPVL(108)	59	unknown	<i>Staphylococcus aureus</i>	Sak4	gp18	98372	YP_918908	257	new	no
1497	phiETA2	69	unknown	<i>Staphylococcus aureus</i>	Sak4	gp16	99758	YP_001004276	259	new	no
165	phi 11	53	temperate	<i>Staphylococcus aureus subsp. aureus</i> NCTC 8325	Sak4	ORF12	9287	NP_803265	259	new	no
167	phi 13	49	temperate	<i>Staphylococcus aureus subsp. aureus</i> NCTC 8325	Sak4	ORF12	9389	NP_803367	259	new	no
108	Sfi19 [1]***	45	ex-temperate	<i>Streptococcus thermophilus</i>	Sak4	p31	4977	NP_049951	233	new	no
1620	2972 [4]***	44	ex-temperate	<i>Streptococcus thermophilus</i>	Sak4	gp35	96465	YP_238517	233	new	no
106	DT1 [1]***	45	ex-temperate	<i>Streptococcus thermophilus</i>	Sak4	p31	4879	NP_049420	233	new	no
107	Sfi11 [1]***	53	ex-temperate	<i>Streptococcus thermophilus</i>	Sak4	gp233	4931	NP_056708	233	new	no
109	Sfi21	50	temperate	<i>Streptococcus thermophilus</i>	Sak4	p35	5026	NP_049998	233	new	no
58	O1205	57	temperate	<i>Streptococcus thermophilus</i>	Sak4	p09	2714	NP_695087	233	new	no
174	315.2	60	temperate	<i>Streptococcus pyogenes</i> MGAS315	Sak4	SpyM3_0962	9683	NP_795445	226	new	no
177	315.5	55	temperate	<i>Streptococcus pyogenes</i> MGAS315	Sak4	SpyM3_1344	9853	NP_795615	226	new	no
32	phig1e	50	temperate	<i>Lactobacillus</i>	Sak4	p57	1604	NP_695186	220	new	no
82	A2	61	temperate	<i>Lactobacillus casei</i>	Sak4	p34	3896	NP_680513	240	new	no
1524	phi AT3	55	temperate	<i>Lactobacillus casei</i>	Sak4	ORF28	100204	YP_025055	235	new	no
1510	phiJL-1 [5]***	46	ex-temperate	<i>Lactobacillus plantarum</i>	Sak4	ORF224	94532	YP_223910	224	new	yes
1473	Lc-Nu [2]***	51	ex-temperate	<i>Lactobacillus rhamnosus</i>	Sak4	ORF29	100855	YP_358787	235	new	no
322	phi adh	63	temperate	<i>Lactobacillus gasseri</i>	Sak4	ORF223	3820	NP_050126	223	new	no
319	BK5-T	63	temperate	<i>Lactococcus lactis</i>	Sak4	p46	4030	NP_116538	234	new	no
1868	KSY1	130	unknown	<i>Lactococcus lactis</i>	Sak4	gp89	79282	YP_001469088	246	new	yes
86	2389	59	temperate	<i>Listeria monocytogenes</i>	Sak4	gp47	4137	NP_511030	230	new	no
1850	B025	65	unknown	<i>Listeria monocytogenes</i>	Sak4	gp57	78366	YP_001468696	230	new	no
1648	PBI1	81	unknown	<i>Mycobacterium;Mycobacterium smegmatis</i>	Sak4	gp56	102247	YP_655252	286	new	yes
1627	PLot	89	unknown	<i>Mycobacterium;Mycobacterium smegmatis</i>	Sak4	gp59	101456	YP_655438	286	new	yes
1629	phiHSIC	47	unknown	<i>Listonella pelagia</i>	Sak4	ORF46	96770	YP_224274	252	new	no
1882	BcepGomr	75	unknown	<i>Burkholderia cepacia</i>	Sak4	gp45	80392	YP_001210265	220	new	yes
5	HK620	58	temperate	<i>Escherichia coli</i>	Sak4	p12	140	NP_112045	226	new	no

MGE ^a	Phage Name	nb cds ^b	vir / temp ^c	Host	Type ^d	Gene name	prot:vir: ^e	Access Nb ^f	Length ^g	new/known	Presence of RecE
1552	PA73	52	unknown	<i>Pseudomonas aeruginosa</i>	Sak4	ORF032	95168	YP_001293439	238	new	yes
117*	Aaphi23	66	temperate	<i>Actinobacillus actinomycetemcomitans</i>	Sak4	p07	5224	NP_852729	297	new	yes
147*	Barnyard	109	unknown	<i>Mycobacterium;Mycobacterium smegmatis</i>	Sak4	gp75	7498	NP_818613	340	new	yes
146*	4268*	49	virulent	<i>Lactococcus lactis</i>	Sak4	p10	7384	NP_839901	306	new	yes
1568*	BCJA1c	58	temperate	<i>Bacillus clarkii</i>	Sak4	gp14	98877	YP_164392	362	new	no
1592	BMP-1	47	temperate	<i>Bordetella bronchiseptica</i>	gp2.5	p39	98495	NP_996602	182	new	yes
1673	BIP-1	48	temperate	<i>Bordetella bronchiseptica</i>	gp2.5	p39	107830	NP_996650	182	new	yes
1537	BPP-1	49	temperate	<i>Bordetella</i>	gp2.5	Bbp40	107422	NP_958709	182	new	yes
1485	3A	67	temperate	<i>Staphylococcus aureus</i>	gp2.5	Orf19	93939	YP_239980	185	new	yes
166	phi 12	49	temperate	<i>Staphylococcus aureus subsp. aureus</i> NCTC 8325	gp2.5	p11	9339	NP_803317	185	new	yes
1506	47	72	temperate	<i>Staphylococcus aureus</i>	gp2.5	Orf22	94380	YP_240049	185	new	yes
1594	2638A	57	temperate	<i>Staphylococcus aureus</i>	gp2.5	Orf18	95983	YP_239830	184	new	yes
1855	tp310-2	67	unknown	<i>Staphylococcus aureus</i>	gp2.5	gp20	78668	YP_001429915	185	new	yes
178	315.6	51	temperate	<i>Streptococcus pyogenes</i> MGAS315	gp2.5	SpyM3_1444	9912	NP_795674	187	new	yes
1604	P-SSP7	53	temperate	<i>Prochlorococcus marinus</i>	gp2.5	gp2.5	100017	YP_214192	189	new	no
1645	BcepC6B	46	temperate	<i>Burkholderia cepacia</i>	gp2.5	gp38	103801	YP_024958	201	new	yes
133	BcepNazgul	73	NA	<i>Burkholderia cepacia</i>	gp2.5	Nazgul70	6391	NP_919004	206	new	yes
1477	Bcep43	65	virulent	<i>Burkholderia cepacia</i>	gp2.5	gp21	101572	NP_958126	235	new	yes
75	Bcep781	66	virulent	<i>Burkholderia cepacia</i>	gp2.5	gp21	3652	NP_705647	235	new	yes
1599	Bcep1	71	virulent	<i>Burkholderia cenocepacia</i>	gp2.5	gp22	106753	NP_944330	236	new	yes
1854	BcepNY3	70	unknown	<i>Burkholderia cenocepacia</i> HI2424	gp2.5	gp20	78580	YP_001294858	236	new	yes
49	T7	60	virulent	<i>Escherichia coli</i>	gp2.5	gp2.5	2173	NP_041970	232	new	no
67	T3	55	virulent	<i>Escherichia coli</i>	gp2.5	p16	3340	NP_523311	232	new	no
1528	K1F	43	virulent	<i>Escherichia coli</i>	gp2.5	gp13	94698	YP_338102	211	new	no
158	PaP3	71	unknown	<i>Pseudomonas aeruginosa</i>	gp2.5	p31	8818	NP_775226	186	new	no
161	gh-1	42	virulent	<i>Pseudomonas putida</i>	gp2.5	p10	8867	NP_813756	233	new	no
1516	Berlin	45	unknown	<i>Yersinia pestis</i>	gp2.5	gp11	94579	YP_918993	235	new	no
184	phiA1122	50	virulent	<i>Yersinia pestis</i>	gp2.5	protein 2.5	10427	NP_848274	232	new	no
31	phiYeO3-12	59	virulent	<i>Yersinia enterocolitica</i> (type O:3)	gp2.5	2.5	1515	NP_052082	232	new	no
1851	SETP3	53	unknown	<i>Salmonella enteritidis</i>	gp2.5	gp11	78413	YP_001110811	208	new	yes
1547	KS7	59	NA	<i>Salmonella enterica subsp. enterica</i> serovar Typhimurium	gp2.5	gp40	95034	YP_224062	126	new	yes
72	APSE-1	54	virulent	<i>Acyrtosiphon pisum</i>	gp2.5	P50	3551	NP_051011	184	new	yes
1856	Syn5	61	unknown	<i>Synechococcus sp. WH 8109</i>	gp2.5	gp21	78707	YP_001285430	192	new	no
1523	VP4	31	virulent	<i>Vibrio</i>	gp2.5	orf24	99698	YP_249584	223	new	no
1493	OP2	62	virulent	<i>Xanthomonas oryzae</i>	gp2.5	ORF17	94030	YP_453634	210	new	yes

MGE ^a	Phage Name	nb cds ^b	vir / temp ^c	Host	Type ^d	Gene name	prot:vir: ^e	Access Nb ^f	Length ^g	new/known	Presence of RecE
139	RB49	274	virulent	<i>Escherichia coli</i>	UvsX	UvsX	6464	NP_891595	356	known	no
2002	JS98	266	unknown	<i>Escherichia coli</i>	UvsX	UvsX	107919	YP_001595166	390	known	no
142	T4	278	virulent	<i>Escherichia coli</i>	UvsX	UvsX	7080	NP_049656	391	known	no
140	RB69	273	virulent	<i>Escherichia coli</i>	UvsX	UvsX	6758	NP_861734	390	known	no
148	Bxz1	225	unknown	<i>Mycobacterium;Mycobacterium smegmatis</i>	UvsX	gp201	7708	NP_818252	362	known	no
151	Che8	112	temperate	<i>Mycobacterium;Mycobacterium smegmatis</i>	UvsX	gp69	8053	NP_817407	293	new	no
1550	Catera	218	unknown	<i>Mycobacterium;Mycobacterium smegmatis</i>	UvsX	gp203	107302	YP_656183	362	known	no
1577	G1	214	virulent	<i>Staphylococcus aureus</i>	UvsX	Orf21	95585	YP_240962	418	known	no
1623	Twort	195	virulent	<i>Staphylococcus aureus</i>	UvsX	Orf19	96661	YP_238613	407	known	no
1655	K	115	virulent	<i>Staphylococcus</i>	UvsX	ORF93	99274	YP_024521	418	known	no
1548	P-SSM4	198	virulent	<i>Prochlorococcus marinus</i>	UvsX	UvsX	104551	YP_214708	336	known	no
1630	P-SSM2	329	virulent	<i>Prochlorococcus marinus</i>	UvsX	UvsX	104784	YP_214417	335	known	no
1542	RB32	270	unknown	<i>Enterobacteriaceae</i>	UvsX	UvsX	103653	YP_802982	393	known	no
1667	RB43	292	virulent	<i>Enterobacteriaceae</i>	UvsX	UvsX	98233	YP_239013	386	known	no
1883	A511	190	unknown	<i>Listeria monocytogenes</i>	UvsX	gp70	80492	YP_001468521	414	known	no
1517	P100	174	virulent	<i>Listeria monocytogenes</i>	UvsX	gp72	63811	82547677	414	known	no
1583	Syn9	229	unknown	<i>Synechococcus sp. WH 8012;Prochlorococcus marinus</i>	UvsX	gp178	103034	YP_717845	336	known	no
1459	S-PM2	236	virulent	<i>Synechococcus sp.;Synechococcus sp. WH 7803</i>	UvsX	UvsX	106936	YP_195169	343	known	no
1474	Aeh1	352	virulent	<i>Aeromonas hydrophila</i>	UvsX	UvsX	106255	NP_943894	411	known	no
1885	phiEF24C	221	unknown	<i>Enterococcus faecalis</i>	UvsX	gp064	80769	YP_001504173	415	known	no
1633	LP65	165	virulent	<i>Lactobacillus plantarum</i>	UvsX	gp053	100740	YP_164688	405	known	no
1530	Ma-LMM01	184	unknown	<i>Microcystis aeruginosa</i>	UvsX	RecA	98710	YP_851022	354	known	no
119	KVP40	381	virulent	<i>Vibrio parahaemolyticus</i>	UvsX	UvsX	5317	NP_899256	366	known	no
1515**	244	142	unknown	<i>Mycobacterium;Mycobacterium smegmatis</i>	UvsX	gp117	101674	YP_654870	320	known	yes
150**	CJW1	141	temperate	<i>Mycobacterium;Mycobacterium smegmatis</i>	UvsX	gp117	7957	NP_817564	320	known	yes
1874**	0305phi8-36	246	unknown	<i>Bacillus thuringiensis</i>	UvsX	gp241	79916	YP_001429755	457	known	no
1610**	YS40	170	unknown	<i>Thermus thermophilus</i>	UvsX	Orf12	102776	YP_874025	339	known	no

^a MGE : genome identifiers in the Aclame database (<http://aclame.ulb.ac.be/>).

^b nb cds : number of coding sequences in the phage genome.

^c Temperate or virulent phage life style as reported in http://aclame.ulb.ac.be/Classification/Phages/life_style.html.

^d Recombinase profile that initially matched this protein profile.

^e Aclame viral protein identifier.

^f NCBI sequence identifier.

^g Length of the recombinase.

*: These Sak4 are unusual by their longer length, which make them resemble RecA. Due to their ambiguous status, they are not considered for the paper statistics.

** : In these genomes, a RecA together with a Rad52-like recombinase was found.

***: Ex-temperate phages which have become virulent (see references below).

&: This phage is named “P335” in the Aclame database or “P335SL” in the phage list from NCBI.

[1] Chopin A., Bolotin A., Sorokin A., Ehrlich SD. and Chopin M. (2001) Analysis of six prophages in *Lactococcus lactis* IL1403: different genetic structure of temperate and virulent phage populations. *Nucleic Acids Res.* **29**, 644-651.

[2] Labrie S. and Moineau S. (2002) Complete genomic sequence of bacteriophage ϕ 36: demonstration of phage heterogeneity within the P335 quasi-species of lactococcal phages. *Virology* **296**, 308-320.

[3] Lévesque C., Duplessis M., Labonté J., Labrie S., Fremaux C., Tremblay D. and Moineau S. (2005) Genomic organization and molecular analysis of virulent bacteriophage 2972 infecting an exopolysaccharide-producing *Streptococcus thermophilus* strain. *Appl. Environ. Microbiol.* **71**, 4057-4068.

[4] Lu Z., Altermann E., Breidt F., Predki P., Fleming HP. and Klaenhammer TR. (2005) Sequence analysis of the *Lactobacillus plantarum* bacteriophage Φ JL-1. *Gene*, **348**, 45-54.

[5] Segev N. and Cohen G. (1981) Control of circularization of bacteriophage PA DNA in *Escherichia coli*. *Virology*, **114**, 333-342.

[6] Tuohimaa A., Riipinen KA., Brandt K. and Alatossava T. (2006) The genome of the virulent phage Lc-Nu of probiotic *Lactobacillus rhamnosus*, and comparative genomics with *Lactobacillus casei* phages. *Arch. Virol.*, **151**, 947-965.

Table S4

Markers used to analyze the genetic context surrounding every recombinase.

To analyze the genetic context around each recombinase, 9 functional markers were selected. These markers correspond to genes that encode for typical functions of the DNA replication module. For every recombinase, the number of genes to any marker was computed. The shortest distance was retained and the distribution of these distances is presented in Figure S5.

FUNCTION	ACLAME FAMILY IDs^a	total number of members^b
DNA replication initiation	6, 18, 36, 242,393	244
Co-Repressor	3, 16, 156, 163, 253	232
Transcription repressor activity	7, 59	132
Single-strand DNA binding	20, 196, 227	86
DNA recombination endonuclease activity	68, 116	52
5'3' exonuclease activity	100, 131	45
DNA primase activity	61	32
DNA helicase activity	95	26
DNA-directed DNA polymerase activity	128	20

^a Each function is represented by a list of annotated genes classified and indexed in different families (according to the ACLAME classification).

^b Total number of proteins that compose the functional family. The families comprising less than 20 proteins were not considered in the analysis.

Table S5

Statistics of the number of recombinases for which the genetic context could be analyzed and for which a marker of the DNA replication module was found at less than 10 or 20 genes.

	Genes separation ≤ 10	Genes separation > 10	Genes separation ≤ 20	Genes separation > 20
nb of recombinase genes (%) of the recombinase database	133	20	147	6
	87	13	96	4

Table S6

Detailed information for the genomes in which no recombinase has been detected.

See legend at the bottom

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir: ^d
320	Mu	55	temperate	<i>Escherichia coli</i>	no	---
95	PF3	9	unknown	<i>Pseudomonas aeruginosa</i>	no	---
77	PhiCPG1	9	unknown	<i>Chlamydomonas reinhardtii</i>	no	---
323	GA-1	35	virulent	<i>Bacillus sp.</i>	no	---
9	Mx8	86	unknown	<i>Myxococcus xanthus</i>	no	---
60	SIFV	73	temperate	<i>Sulfolobus islandicus</i>	yes	2934
114	PBC5	83	temperate	<i>Sinorhizobium meliloti</i>	no	---
11	P27	58	temperate	<i>Escherichia coli</i>	no	---
326	Spiroplasma phage 4	9	unknown	<i>Spiroplasma melliferum</i>	no	---
118	IN93	35	unknown	<i>Thermus aquaticus</i>	no	---
127	phBC6A51	75	temperate	<i>Bacillus cereus</i>	no	---
128	phBC6A52	49	temperate	<i>Bacillus cereus</i>	no	---
122	PY54	67	temperate	<i>Yersinia enterocolitica</i>	no	---
1457	NC1	11	unknown	<i>Escherichia coli</i>	no	---
1458	NC5	11	unknown	<i>Escherichia coli</i>	no	---
1460	Sulfolobus islandicus rod-shaped virus 2	54	temperate	<i>Sulfolobus islandicus</i>	yes	106091
1462	NC3	10	unknown	<i>Escherichia coli</i>	no	---
1464	L17	31	virulent	<i>Salmonella typhimurium</i> LT2	no	---
1465	phiKO2	64	temperate	<i>Klebsiella oxytoca</i>	no	---
1466	C-St	198	temperate	<i>Clostridium botulinum</i> C	no	---
1467	NC51	11	unknown	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i>	no	---
1468	NC6	11	unknown	<i>Escherichia coli</i>	no	---
1469	WA13	10	unknown	<i>Escherichia coli</i>	no	---
1478	Acidianus filamentous virus 1	40	temperate	<i>Acidianus</i>	no	---
1480	B3	59	temperate	<i>Pseudomonas aeruginosa</i>	no	---
1481	PG1	100	NA	<i>Mycobacterium smegmatis</i>	yes	107543
1482	Orion	100	unknown	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	yes	97909
1483	NC19	11	unknown	<i>Escherichia coli</i>	no	---
1484	phi-MhaA1-PHL101	49	unknown	<i>Mannheimia haemolytica</i>	no	---
1486	PT1028	22	temperate	<i>Staphylococcus aureus</i>	no	---
1488	25	242	unknown	<i>Aeromonas salmonicida</i>	no	---

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir: ^d
1489	NC10	11	unknown	<i>Escherichia coli</i>	no	---
1490	W Beta	53	temperate	<i>Bacillus anthracis</i>	no	---
1491	Gamma	53	temperate	<i>Bacillus cereus</i>	no	---
1492	Cherry	51	temperate	<i>Bacillus anthracis str. Vollum</i>	no	---
1495	F108	44	unknown	<i>Pasteurella multocida</i>	no	---
1496	GIL16c	31	temperate	<i>Bacillus thuringiensis</i>	no	---
1498	WA4	11	unknown	<i>Escherichia coli</i>	no	---
1500	phiEL	201	virulent	<i>Pseudomonas aeruginosa</i>	no	---
1503	Fah	50	unknown	<i>Bacillus anthracis</i> ; <i>Bacillus cereus</i>	no	---
1504	T5	162	virulent	<i>Escherichia coli</i>	no	---
1505	Cooper	99	unknown	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	yes	106026
1507	VP5	48	temperate	<i>Vibrio cholerae</i>	no	---
1511	KSF-1phi	12	unknown	<i>Vibrio cholerae</i>	no	---
1512	P1	110	unknown	<i>Escherichia coli</i>	no	---
1519	ID62	10	unknown	<i>Escherichia coli</i>	no	---
1520	BcepB1A	73	virulent	<i>Burkholderia cenocepacia</i>	yes	107717
1521	NC7	11	unknown	<i>Escherichia coli</i>	no	---
1522	D3112	55	temperate	<i>Pseudomonas aeruginosa</i>	no	---
1525	Pf-WMP4	45	unknown	<i>Phormidium</i>	no	---
1526	PR5	31	virulent	<i>Salmonella typhimurium</i> LT2	no	---
1527	mu1/6	52	unknown	<i>Streptomyces aureofaciens</i>	yes	94648
1533	PSP3	42	temperate	<i>Salmonella</i>	no	---
1534	ID12	11	unknown	<i>Escherichia coli</i>	no	---
1535	NC56	11	unknown	<i>Salmonella enterica subsp. enterica serovar Typhimurium</i>	no	---
1536	Sulfolobus turreted icosahedral virus	36	unknown	<i>Sulfolobus solfataricus</i>	no	---
1538	Xp15	84	NA	<i>Xanthomonas campestris pv. pelargonii</i>	yes	94988
1539	NC29	10	unknown	<i>Escherichia coli</i>	no	---
1541	Acidianus two-tailed virus	72	unknown	<i>Acidianus</i>	no	---
1543	NC28	10	unknown	<i>Escherichia coli</i>	no	---
1544	BcepF1	127	unknown	<i>Burkholderia ambifaria</i>	no	---
1545	Chlamydia phage 4	8	unknown	<i>Chlamydophila abortus</i>	no	---
1551	Chlamydia phage 3,	8	unknown	<i>Chlamydophila pecorum</i>	no	---
1553	NC2	11	unknown	<i>Escherichia coli</i>	no	---
1554	PRR1	4	unknown	<i>Pseudomonas</i>	no	---
1555	Vf12	7	unknown	<i>Vibrio parahaemolyticus</i>	no	---
1558	ID34	11	unknown	<i>Escherichia coli</i>	no	---

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir ^d
1560	Sulfolobus spindle-shaped virus 2	34	temperate	<i>Sulfolobus</i> ; <i>Sulfolobus solfataricus</i>	no	---
1561	Felix 01	131	virulent	<i>Salmonella</i>	no	---
1562	NC16	11	unknown	<i>Escherichia coli</i>	no	---
1563	ID1	11	unknown	<i>Escherichia coli</i>	no	---
1565	BcepMu	53	temperate	<i>Burkholderia cenocepacia</i>	no	---
1568	BCJA1c	58	temperate	<i>Bacillus clarkii</i>	no	---
1570	PA11	70	unknown	<i>Pseudomonas aeruginosa</i>	no	---
1572	ID8	11	unknown	<i>Escherichia coli</i>	no	---
1573	ID52	11	unknown	<i>Escherichia coli</i>	no	---
1574	F10	63	unknown	<i>Pseudomonas aeruginosa</i>	no	---
1575	WA3	11	unknown	<i>Escherichia coli</i>	no	---
1576	ID41	11	unknown	<i>Escherichia coli</i>	no	---
1579	ID45	11	unknown	<i>Escherichia coli</i>	no	---
1580	31	247	virulent	<i>Aeromonas salmonicida</i>	no	---
1582	44RR2.8t	252	virulent	<i>Aeromonas salmonicida</i>	no	---
1584	p12J	9	unknown	<i>Ralstonia pickettii</i>	no	---
1585	WA14	11	unknown	<i>Escherichia coli</i>	no	---
1586	N4	72	unknown	<i>Escherichia coli</i> ; <i>Escherichia coli</i> K12	yes	95828
1587	WA2	11	unknown	<i>Escherichia coli</i>	no	---
1590	ID18	11	unknown	<i>Escherichia coli</i>	no	---
1595	BJ1	70	unknown	<i>Halorubrum</i>	no	---
1596	PR3	31	virulent	<i>Salmonella typhimurium</i> LT2	no	---
1597	F8	91	unknown	<i>Pseudomonas aeruginosa</i>	no	---
1600	PR772	31	virulent	<i>Salmonella typhimurium</i> LT2	no	---
1601	phi3396	64	unknown	<i>Streptococcus dysgalactiae</i> subsp. <i>equisimilis</i>	no	---
1603	ID21	10	unknown	<i>Escherichia coli</i>	no	---
1606	ID32	10	unknown	<i>Escherichia coli</i>	no	---
1608	Qyrzula	81	unknown	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	yes	99123
1609	Era103	53	unknown	<i>Erwinia amylovora</i>	no	---
1614	Vf33	7	unknown	<i>Vibrio parahaemolyticus</i>	no	---
1615	Ragged Hills	37	temperate	<i>Sulfolobus</i> sp.	no	---
1616	119X	53	unknown	<i>Pseudomonas aeruginosa</i>	yes	96418
1617	WA11	11	unknown	<i>Escherichia coli</i>	no	---
1618	Pipefish	102	unknown	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	yes	102466
1619	Bcep176	81	NA	<i>Burkholderia cepacia</i>	no	---

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir: ^d
1621	PaP2	58	unknown	<i>Pseudomonas aeruginosa</i>	yes	103410
1622	NC35	10	unknown	<i>Escherichia coli</i>	no	---
1624	VP2	47	temperate	<i>Vibrio cholerae</i>	no	---
1625	NC41	11	unknown	<i>Escherichia coli</i>	no	---
1626	Kamchatka-1	31	temperate	<i>Sulfolobus sp.</i>	no	---
1628	VP882	71	unknown	<i>Vibrio parahaemolyticus</i>	no	---
1632	PA16	0	unknown		no	---
1634	NC11	11	unknown	<i>Escherichia coli</i>	no	---
1635	WA6	11	unknown	<i>Escherichia coli</i>	no	---
1637	phiMFV1	16	temperate	<i>Mycoplasma fermentans</i>	no	---
1638	ID11	11	unknown	<i>Escherichia coli</i>	no	---
1639	phi1026b	83	temperate	<i>Burkholderia pseudomallei</i>	no	---
1640	Bam35c	32	temperate	<i>Bacillus thuringiensis</i>	no	---
1641	phiSM101	53	unknown	<i>Clostridium perfringens</i> SM101	no	---
1644	K1-5	52	unknown	<i>Escherichia coli</i>	no	---
1647	RSM1	15	unknown	<i>Ralstonia solanacearum</i>	no	---
1649	DMS3	52	unknown	<i>Pseudomonas aeruginosa</i>	no	---
1650	NC13	11	unknown	<i>Escherichia coli</i>	no	---
1651	WA10	11	unknown	<i>Escherichia coli</i>	no	---
1652	ID2	11	unknown	<i>Escherichia coli</i>	no	---
1653	OP1	59	virulent	<i>Xanthomonas oryzae</i>	no	---
1656	Che12	98	unknown	<i>Mycobacterium;Mycobacterium tuberculosis</i>	yes	104072
1657	M6	85	unknown	<i>Pseudomonas aeruginosa</i>	no	---
1658	WA5	11	unknown	<i>Escherichia coli</i>	no	---
1659	RSS1	12	unknown	<i>Ralstonia solanacearum</i>	no	---
1660	phiSMA9	7	temperate	<i>Stenotrophomonas maltophilia</i>	no	---
1663	PR4	31	virulent	<i>Salmonella typhimurium</i> LT2	no	---
1665	phi52237	47	temperate	<i>Burkholderia pseudomallei</i>	no	---
1668	SH1	56	unknown	<i>Haloarcula hispanica</i>	no	---
1669	66	27	virulent	<i>Staphylococcus aureus</i>	no	---
1670	WA45	10	unknown	<i>Escherichia coli</i>	no	---
1672	NC37	11	unknown	<i>Escherichia coli</i>	no	---
1674	K1E	62	virulent	<i>Escherichia coli</i>	no	---
1678	ID22	11	unknown	<i>Escherichia coli</i>	no	---
1677	rod-shaped virus 1	45	temperate	<i>Sulfolobus islandicus</i>	yes	103717

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir: ^d
1680	phiAsp2	76	temperate	<i>Actinoplanes</i>	no	---
78	chp1	12	unknown	<i>Chlamydomydia psittaci</i>	no	---
7	K139	44	temperate	<i>Vibrio cholerae</i> O139	no	---
1843	P23-45	117	unknown	<i>Thermus thermophilus</i>	yes	77990
1844	P35	56	unknown	<i>Listeria monocytogenes</i>	no	---
1845	P1201	97	unknown	<i>Corynebacterium glutamicum</i>	no	---
1846	Tweety	109	unknown	<i>Mycobacterium smegmatis</i>	no	---
1847	Min1	77	unknown	<i>Microbacterium nematophilum</i>	no	---
1848	phiE12-2	50	unknown	<i>Burkholderia pseudomallei</i>	no	---
1849	Bethlehem	87	unknown	<i>Mycobacterium smegmatis</i>	yes	78290
1852	Acidianus rod-shaped virus 1	41	unknown	<i>Acidianus</i> sp. Aci26	yes	78429
1853	U2	81	unknown	<i>Mycobacterium smegmatis</i>	yes	78476
1857	phiO18P	45	unknown	<i>Aeromonas media</i>	no	---
1860	LKD16	53	unknown	<i>Pseudomonas aeruginosa</i>	no	---
1862	phiE255	55	unknown	<i>Burkholderia thailandensis</i>	no	---
1863	RSA1	51	unknown	<i>Ralstonia solanacearum</i>	no	---
1865	SAP-2	20	unknown	<i>Staphylococcus aureus</i>	no	---
1866	phiE202	48	unknown	<i>Burkholderia thailandensis</i>	no	---
1867	Phage MP22	51	unknown	<i>Pseudomonas aeruginosa</i> UCBPP-PA14	no	---
1869	Av-1	22	unknown	<i>Actinomyces naeslundii</i>	no	---
1870	P74-26	116	unknown	<i>Thermus thermophilus</i>	yes	79486
1871	cdtl	60	unknown	<i>Escherichia coli</i>	no	---
1879	LKA1	56	unknown	<i>Pseudomonas aeruginosa</i>	no	---
1880	filamentous virus 2	52	unknown	<i>Acidianus pozzuoliensis</i>	no	---
1881	phi644-2	71	unknown	<i>Burkholderia pseudomallei</i>	no	---
1884	PA6	48	unknown	<i>Propionibacterium acnes</i>	yes	80694
1887	L5	0	unknown	<i>Oenococcus oeni</i>	no	---
1888	Phi1	276	unknown	<i>Escherichia coli</i> K12	no	---
1889	Xop411	58	unknown	<i>Xanthomonas oryzae</i>	no	---
1890	P9	53	unknown	<i>Streptococcus equi</i>	no	---
1892	Geobacillus virus E2	71	unknown	<i>Geobacillus</i>	no	---
1893	BFK20	55	unknown	<i>[Brevibacterium] flavum</i>	yes	81257
2000	blL67	37	unknown	<i>Lactococcus</i>	no	---
2001	phiEcoM-GJ1	75	unknown		no	---
329	phi-105	51	temperate	<i>Bacillus subtilis</i>	no	---

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir: ^d
181	VGJphi	13	unknown	<i>Vibrio cholerae</i>	no	---
116	phi29	27	virulent	<i>Bacillus subtilis</i>	no	---
43	lke	10	unknown	<i>Escherichia coli</i>	no	---
37	phiX174	11	unknown	<i>Escherichia coli</i>	no	---
321	G4	11	unknown	<i>Escherichia coli</i>	no	---
163	phiKZ	306	virulent	<i>Pseudomonas aeruginosa</i>	no	---
324	186	46	temperate	<i>Escherichia coli</i>	no	---
30	phiE125	71	temperate	<i>Burkholderia thailandensis</i>	no	---
153	Che9d	111	temperate	<i>Mycobacterium;Mycobacterium smegmatis</i>	no	---
154	Corndog	122	virulent	<i>Mycobacterium;Mycobacterium smegmatis</i>	yes	8298
115	MV-L1	4	unknown	<i>Acholeplasma laidlawii</i>	no	---
316	PhiCh1	98	unknown	<i>Natrialba magadii</i>	no	---
91	P1	11	temperate	<i>Mycoplasma pulmonis</i>	no	---
94	Pf1	14	unknown	<i>Pseudomonas aeruginosa</i>	no	---
162	phi-6	13	unknown	<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i>	no	---
162	phi-6	13	unknown	<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i>	no	---
162	phi-6	13	unknown	<i>Pseudomonas syringae</i> pv. <i>phaseolicola</i>	no	---
159	phi-13	13	unknown	<i>Pseudomonas syringae</i>	no	---
159	phi-13	13	unknown	<i>Pseudomonas syringae</i>	no	---
159	phi-13	13	unknown	<i>Pseudomonas syringae</i>	no	---
53	TM4	89	unknown	<i>Mycobacterium avium;Mycobacterium smegmatis</i>	yes	2551
129	phi-12	15	unknown	<i>Pseudomonas syringae</i>	no	---
129	phi-12	15	unknown	<i>Pseudomonas syringae</i>	no	---
129	phi-12	15	unknown	<i>Pseudomonas syringae</i>	no	---
160	PP7	4	unknown	<i>Pseudomonas aeruginosa</i>	no	---
56	SVTS2	13	unknown	<i>Spiroplasma</i>	no	---
38	P60	80	virulent	<i>Synechococcus</i> sp. <i>WH 7803</i>	no	---
137	MX1	4	unknown	<i>Escherichia coli</i>	no	---
130	phi-8	16	unknown	<i>Pseudomonas syringae</i>	no	---
130	phi-8	16	unknown	<i>Pseudomonas syringae</i>	no	---
130	phi-8	16	unknown	<i>Pseudomonas syringae</i>	no	---
90	MAV1	15	temperate	<i>Mycoplasma arthritis</i>	no	---
113	phiCPAR39	7	unknown	<i>Pseudoalteromonas;Chlamydomphila pneumoniae AR39</i>	no	---
55	SIO1	34	virulent	<i>Roseobacter</i>	no	---
141	SP6	52	virulent	<i>Salmonella typhimurium</i>	no	---

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir: ^d
48	S13	12	unknown	<i>Salmonella</i>	no	---
64	VpV262	67	virulent	<i>Vibrio parahaemolyticus</i>	no	---
121	M11	4	unknown	<i>Escherichia coli</i>	no	---
318	blL310	29	temperate	<i>Lactococcus lactis</i> ; <i>Lactococcus lactis subsp. lactis</i> II1403	no	---
28	phi-C31	53	temperate	<i>Streptomyces</i>	yes	1347
69	blL312	27	temperate	<i>Lactococcus lactis</i> ; <i>Lactococcus lactis subsp. lactis</i> II1403	no	---
24	phi CTX	47	temperate	<i>Pseudomonas aeruginosa</i>	no	---
327	phiK	10	unknown	<i>Escherichia coli</i>	no	---
131	phi-BT1	55	temperate	<i>Streptomyces lividans</i>	yes	6260
76	phage Chp2	8	unknown	<i>Chlamydomonas reinhardtii</i>	no	---
99	1-R8A2B	12	unknown	<i>Spiroplasma citri</i>	no	---
71	L2	14	unknown	<i>Acholeplasma laidlawii</i>	no	---
87	psiM2	32	virulent	<i>Methanothermobacter thermautotrophicus</i>	no	---
98	1-C74	13	unknown	<i>Spiroplasma citri</i>	no	---
73	PM2	22	unknown	<i>Pseudoalteromonas</i>	no	---
62	fs1	15	unknown	<i>Vibrio cholerae</i> O139	no	---
79	HP1	42	temperate	<i>Haemophilus influenzae</i>	no	---
149	Bxz2	86	temperate	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	yes	7827
54	phiMH2K	11	unknown	<i>Bdellovibrio bacteriovorus</i>	no	---
89	L5	85	temperate	<i>Mycobacterium tuberculosis</i>	yes	4278
172	C1	20	virulent	<i>Streptococcus</i> sp. 'group C'	no	---
126	WPhi	44	temperate	<i>Escherichia coli</i>	no	---
117	Aaphi23	66	temperate	<i>Actinobacillus actinomycetemcomitans</i>	yes	5223
92	phiB5	10	unknown	<i>Propionibacterium freudenreichii</i>	no	---
110	Sulfolobus virus 1	32	temperate	<i>Sulfolobus shibatae</i>	no	---
314	phi3626	50	temperate	<i>Clostridium perfringens</i>	no	---
146	P335	49	virulent	<i>Lactococcus lactis</i>	yes	7383
144	fr	4	unknown	<i>Escherichia coli</i>	no	---
35	alpha3	10	unknown	<i>Escherichia coli</i>	no	---
168	phiP68	22	virulent	<i>Staphylococcus aureus</i>	no	---
63	fs-2	9	unknown	<i>Vibrio cholerae</i> O139	no	---
325	N15	60	temperate	<i>Escherichia coli</i>	no	---
315	P2	43	temperate	<i>Escherichia coli</i>	no	---
136	M13	10	unknown	<i>Escherichia coli</i> K12	no	---
27	If1	10	unknown	<i>Escherichia coli</i>	no	---

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir ^d
124	SP	4	unknown	<i>Escherichia coli</i>	no	---
112	Cf1c	9	unknown	<i>Xanthomonas campestris</i>	no	---
52	D29	79	virulent	<i>Mycobacterium tuberculosis</i>	yes	2486
96	ST64B	56	temperate	<i>Salmonella enterica</i> ; <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i>	no	---
13	SPBc2	185	temperate	<i>Bacillus subtilis</i>	no	---
68	VfO3K6	10	unknown	<i>Vibrio parahaemolyticus</i>	no	---
138	NL95	4	unknown	<i>Escherichia coli</i>	no	---
88	psiM100	35	temperate	<i>Methanothermobacter wolfeii</i>	no	---
317	bIL311	22	temperate	<i>Lactococcus lactis</i> ; <i>Lactococcus lactis</i> subsp. <i>lactis</i> II1403	no	---
156	Rosebush	90	virulent	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	yes	8691
12	P4	14	unknown	<i>Escherichia coli</i>	no	---
120	L-413C	40	NA	<i>Yersinia pestis</i>	no	---
132	phiKMV	49	virulent	<i>Pseudomonas aeruginosa</i>	no	---
3	AP205	4	unknown	<i>Acinetobacter</i>	no	---
147	Barnyard	109	unknown	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	yes	7495
145	MS2	4	unknown	<i>Escherichia coli</i>	no	---
169	44AHJD	21	virulent	<i>Staphylococcus aureus</i>	no	---
111	VSK	14	unknown	<i>Xanthomonas campestris</i> ; <i>Vibrio cholerae</i> O139	no	---
173	315.1	56	temperate	<i>Streptococcus pyogenes</i> MGAS315	no	---
134	GA	3	unknown	<i>Escherichia coli</i>	no	---
17	VfO4K68	8	unknown	<i>Vibrio parahaemolyticus</i>	no	---
4	B103	17	virulent	<i>Bacillus subtilis</i>	no	---
42	I2-2	9	unknown	<i>Escherichia coli</i>	no	---
171	SM1	56	temperate	<i>Streptococcus mitis</i>	no	---
155	Omega	237	temperate	<i>Mycobacterium</i> ; <i>Mycobacterium smegmatis</i>	no	---
97	V	53	temperate	<i>Shigella flexneri</i>	no	---
135	KU1	4	unknown	<i>Escherichia coli</i>	no	---
15	VSKK	6	unknown	<i>Vibrio cholerae</i> O139	no	---
182	VHML	57	temperate	<i>Vibrio harveyi</i>	no	---
47	PRD1	31	virulent	<i>Escherichia coli</i>	no	---
183	Xp10	60	virulent	<i>Xanthomonas oryzae</i>	no	---
51	Bxb1	86	temperate	<i>Mycobacterium smegmatis</i>	yes	2392
103	Cp-1	25	virulent	<i>Streptococcus pneumoniae</i>	no	---
328	HP2	37	temperate	<i>Haemophilus influenzae</i>	no	---
2003	phiEcoM-GJ1	75	unknown	<i>Escherichia coli</i>	no	---

MGE ^a	Phage Name	nb cds ^b	vir/temp ^c	Host	Presence of RecE	prot:vir: ^d
2005	SkV1_CR2-3x	13	unknown	<i>Spiroplasma kunkelii</i>	no	---
2006	spindle-shaped virus 4	34	unknown	<i>Sulfolobus islandicus</i>	no	---

^a MGE : genome identifiers in the Aclame database (<http://aclame.ulb.ac.be/>).

^b nb cds : number of coding sequences in the phage genome.

^c Temperate or virulent phage life style as reported in http://aclame.ulb.ac.be/Classification/Phages/life_style.html.

^d Aclame viral protein identifier of the RecE protein.

Table S7Single-strand annealing activities *in vivo* tested for different proteins

<i>Strain</i>	<i>Recombinase</i>	<i>Rif^R/viable cells with oligo Maj32*</i>	<i>Spontaneous Rif^R/viable*</i>
MAC1042	Redβ from λ	2.05 (+/- 1.3) 10 ⁻³	2.1 (+/-1.1) 10 ⁻⁸
JAC148	Redβ R161C	1.35 (+/-1.3) 10 ⁻⁶	6.7 (+/-3.4) 10 ⁻⁸
MAC1525	Gp2.5 from T7	1.56 (+/- 0.7) 10 ⁻⁵	3 (+/- 2.7) 10 ⁻⁸
MAC1535	Gp2.5 from φ12	2.28 (+/- 1.7) 10 ⁻⁶	2 (+/- 1.8) 10 ⁻⁸
JAC151	φ12Gp2.5 K79C	3.7 (+/- 2.5) 10 ⁻⁷	4.4 (+/- 3.9) 10 ⁻⁸
JAC153	φ12Gp2.5 R83C	2.4 (+/- 1.3) 10 ⁻⁷	3 (+/-1.6) 10 ⁻⁸
MAC1527	UvsX from T4	1.42 (+/- 0.3)10 ⁻⁶	2.6 (+/- 2.6) 10 ⁻⁷
JAC173	UvsX K66A	6.05 (+/- 3.5) 10 ⁻⁶	7.9 (+/- 10) 10 ⁻⁸
MAC1544	Sak4 from PA73	1.7 (+/- 1) 10 ⁻⁶	1.9 (+/-2.2) 10 ⁻⁷
JAC171	Sak4 K30A	9.7 (+/- 2.9) 10 ⁻⁷	9.5 (+/- 6.6) 10 ⁻⁸
JAC157	None	4.96 (+/-2.7) 10 ⁻⁷	2.1 (+/-1.6) 10 ⁻⁷

*Average and standard deviations from at least three experiments per strain.

Chapitre II

Introduction

Nous avons montré par une analyse bio-informatique que les recombinaisons de la superfamille Rad51/RecA, celles de type Sak4 sont très majoritairement présentes chez des bactériophages tempérés, contrairement aux recombinaisons de type UvsX, majoritairement présentes chez les phages virulents (Figure S8B et Table S3 de l'Article 1). Les génomes de bactériophages tempérés codant une recombinaison Sak4 sont mosaïques, tout comme ceux codant une recombinaison de la superfamille Rad52 (Martinson et al., 2008) (voir Discussion). Nous nous sommes alors demandés si, comme Red β , Sak4 était capable d'apparier de l'ADN simple brin de séquence divergée. Une telle propriété pourrait être un facteur important dans l'évolution rapide des génomes de bactériophages.

Dans ce chapitre, j'ai choisi de présenter plus précisément les résultats portant sur les recombinaisons de la superfamille Rad51/RecA. Cette famille comporte deux types de recombinaisons : certaines, comme UvsX, ressemblent plus à RecA, tandis que d'autres, comme Sak4, ressemblent plus à des RecA auxquelles il manquerait les domaines N-ter et C-ter ; des « mini-RecA » (Lopes et al., 2010). Alors qu'UvsX est déjà étudiée depuis longtemps (voir Introduction), très peu de choses sont connues sur les protéines de type Sak4. Jusqu'à présent aucune étude n'avait décrit la protéine Sak4 comme étant une recombinaison, même si l'étude génétique de Bouchard et Moineau de 2004 le suggérait (Bouchard and Moineau, 2004). Sak4 est appelée ainsi pour « Sensitivity to AbiK », où Abi signifie Abortive infection. AbiK est un gène porté par un plasmide, présent dans des souches de *L. lactis*, qui les rend résistantes à certains bactériophages en faisant avorter leur cycle lytique. Cependant, plusieurs bactériophages sont insensibles à AbiK lorsqu'ils sont mutés pour les gènes *sak4*, *erf* ou *sak* (Bouchard and Moineau, 2004). Étant donné qu'Erf

et Sak sont des recombinases, les auteurs ont suggéré que Sak4 soit aussi une recombinase. Nous avons montré (Chapitre I des Résultats) que cette protéine est une mini-RecA, elle contient les deux motifs de liaison à l'ATP (Walker A et B) comme RecA, mais il lui manque le domaine N-ter qui contient les domaines d'interaction protéine : protéine et le domaine C-ter qui comporte le site secondaire de fixation à l'ADN double brin (Figure 3 de l'Article I et Figure 1 de l'Introduction) (Lopes et al., 2010).

Nous avons étudié deux protéines Sak4 *in vivo* : la recombinase Sak4 du bactériophage PA73 (collection Félix d'Hérelle) infectant *Pseudomonas aeruginosa* et la Sak4 du bactériophage tempéré HK620 infectant *E. coli*. Nous avons observé : (i) que Sak4 de PA73 a une activité d'appariement d'ADN simple brin *in vivo*, (ii) que cette activité est maintenue dans un mutant inactivant son site de fixation de l'ATP, et (iii) que, tout comme Red β , Sak4 de HK620 apparie de l'ADN simple brin ayant jusqu'à 13% de divergence. Nos travaux ont également montré qu'UvsX était capable de faire de l'appariement d'ADN simple brin *in vivo*, contrairement à RecA. Enfin, nous avons découvert que la protéine Gam du bactériophage Lambda augmente d'un facteur 20 l'activité d'appariement d'ADN simple brin de Red β *in vivo*.

Table 4 – L'activité d'appariement d'ADN simple brin résiduelle *in vivo* n'est pas diminuée en contexte *recO*.

N°Souche	Génotype	N°Plasmide	Recombinase	Rif ^R / viable avec oligo Maj32	Nb exp	Rif ^R /viable sans oligo	Nb exp
JAC157	<i>recA</i>	pJA4	-	$5,0 (\pm 2,7) 10^{-7}$	6	$2,2 (\pm 1,4) 10^{-8}$	4
MAC1031	<i>recA</i>	-	-	$7,8 (\pm 4,4) 10^{-7}$	6	$3,5 (\pm 1,7) 10^{-8}$	5
JAC214	<i>recArecO</i>	-	-	$4,9 (\pm 1,0) 10^{-7}$	3	$1,1 (\pm 0,6) 10^{-7}$	3
JAC193	<i>recA</i>	pJA5	RecA	$1,9 (\pm 0,8) 10^{-7}$	3	$6,3 (\pm 0,7) 10^{-8}$	3
JAC160		pJA4	-	$2,8 (\pm 2,1) 10^{-8}$	3		

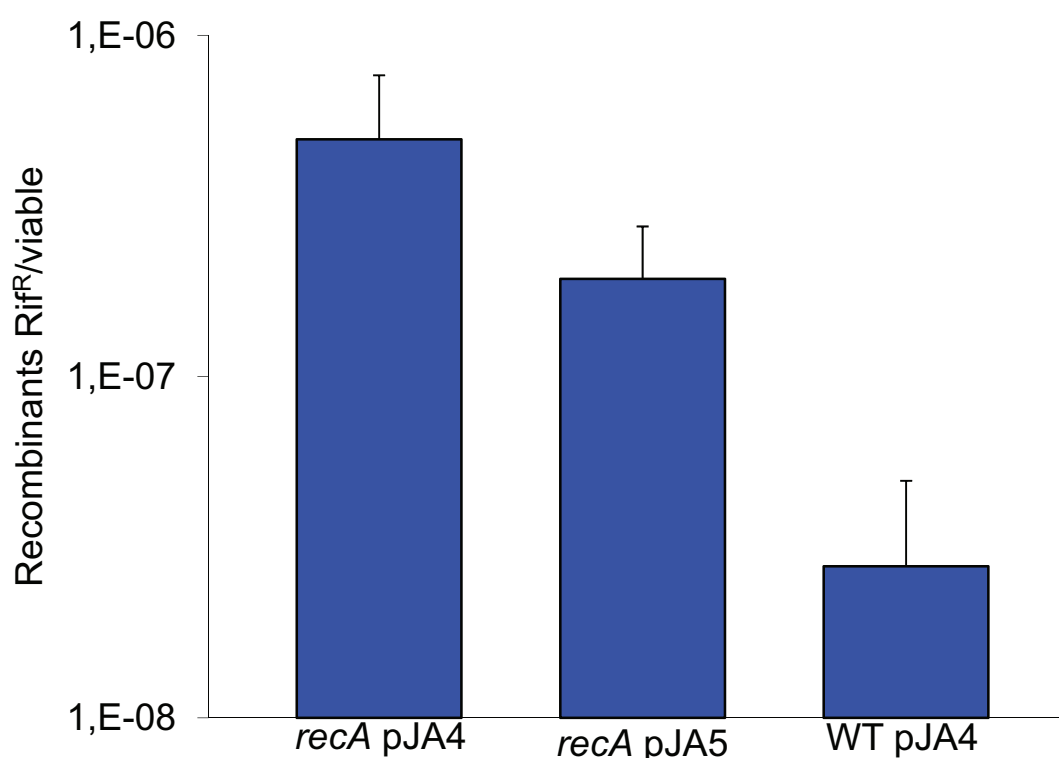


Figure 19 : Activité d'appariement d'ADN simple brin résiduelle dans la cellule.

La fréquence de recombinants Rif^R est estimée par le ratio de recombinants rifampicine résistant par cellule viable. Les expériences ont été réalisées dans la souche AB1157 *recA* en présence du plasmide qui n'exprime aucune recombinaison pJA4 (*recA* pJA4) et en présence du plasmide qui exprime RecA (*recA* pJA5), puis dans la souche sauvage qui exprime RecA (WT pJA4). Ces résultats mettent en évidence que RecA ne fait pas d'appariement d'ADN simple brin *in vivo*. De plus, lorsque RecA est sur le chromosome de la souche WT, elle empêche l'activité cellulaire résiduelle, alors que lorsqu'elle est exprimée à partir de pJA5, elle ne l'empêche que partiellement.

I. Activité résiduelle d'appariement d'ADN simple brin chez *E. coli*.

Comme décrit en détail dans la section Matériel et Méthodes, le test d'appariement d'ADN simple brin *in vivo* que nous avons mis en place consiste à produire diverses recombinaisons à partir d'un plasmide introduit dans la souche AB1157 *recA*, et à tester l'efficacité d'incorporation dans le chromosome, après transformation, de l'oligonucléotide Maj32. Celui-ci confère aux cellules l'ayant incorporé la résistance à la rifampicine. Avant d'effectuer ce test avec différentes recombinaisons, le plasmide vecteur pJA4 qui n'exprime aucune recombinaison a été testé. Une fréquence de recombinaison de $5,0 \times 10^{-7}$ Rif^R par cellule viable a été observée, soit 22 fois plus que la fréquence de mutants spontanés Rif^R ($2,2 \times 10^{-8}$) (Table 4). Il y a donc dans la cellule une activité résiduelle d'appariement d'ADN simple brin. Afin de s'assurer que cette activité d'appariement d'ADN simple brin résiduelle n'est pas causée par la présence du plasmide vecteur dans la cellule, l'expérience a été répétée en absence de celui-ci. Une fréquence également élevée, de $7,8 \times 10^{-7}$ Rif^R par cellule viable a été obtenue (Table 4). La présence du plasmide n'est donc pas la cause de cette activité résiduelle.

Afin de tenter d'identifier une protéine responsable de cette activité d'appariement d'ADN simple brin résiduelle, l'expérience a été réalisée dans une souche déficiente pour le gène *recO*, car il a été décrit que RecO était capable d'apparier de l'ADN simple brin complémentaire (Kantake et al., 2002; Luisi-DeLuca and Kolodner, 1994). La fréquence obtenue est toujours aussi élevée, $4,9 \times 10^{-7}$ Rif^R par cellule viable (Table 4), ce n'est donc pas RecO qui est responsable de cette activité résiduelle. Il existe peut-être une autre protéine dans la cellule capable de faire de l'appariement d'ADN simple brin. Il est aussi possible que l'ADN simple brin s'apparie de manière spontanée.

Dans ce qui suit, les fréquences d'appariement d'ADN simple brin en présence de recombinaisons seront donc comparées à la fréquence obtenue avec le plasmide

Table 5 – Activité d'appariement d'ADN simple brin *in vivo*
pour les différents membres de la famille Rad51/RecA

N° Souche	N° Plasmide	Recombinase	Rif ^R / viable avec oligo Maj32	Nb exp	Facteur relatif à pJA4	Test de student par rapport à pJA4	Rif ^R /viable sans oligo	Nb exp
JAC193	pJA5	RecA	$1,9 (\pm 0,8) 10^{-7}$	3	0,4	p=0,05	$6,3 (\pm 0,7) 10^{-8}$	3
MAC1528	pJA6	UvsX	$1,9 (\pm 1,2) 10^{-6}$	3	4	p=0,01	$8,8 (\pm 8,7) 10^{-8}$	3
JAC180	pJA7	UvsX K66A	$4,5 (\pm 0,5) 10^{-8}$	3	0,09	P=0,01	$2,6 (\pm 2,5) 10^{-8}$	3
JAC205	pJA8	UvsXΔCter	$3,7 (\pm 2,8) 10^{-7}$	6	0,7	p=0,23	$3,5 (\pm 4,0) 10^{-8}$	3
JAC173	pJA9	UvsX K66A W309L	$6,1 (\pm 3,5) 10^{-6}$	4	12	P<0,01	$7,9 (\pm 11) 10^{-8}$	3
JAC186	pJA10	UvsX W309L	$3,6 (\pm 2,9) 10^{-6}$	3	7	p=0,01	$3,0 (\pm 1,6) 10^{-8}$	3
JAC206	pJA11	UvsX K66A ΔCter	$5,4 (\pm 2,2) 10^{-7}$	6	1	p=0,38	$1,2 (\pm 0,6) 10^{-7}$	3
MAC1544	pJA12	Sak4 _{PA73}	$1,7 (\pm 1,1) 10^{-6}$	4	3	p=0,01	$2,2 (\pm 2,5) 10^{-7}$	4
JAC171	pJA13	Sak4 K30A _{PA73}	$9,7 (\pm 2,9) 10^{-7}$	3	2	p=0,02	$9,5 (\pm 6,6) 10^{-8}$	3
JAC194	pJA14	Sak4 K123A _{PA73}	$3,7 (\pm 1,9) 10^{-7}$	3	0,7	p=0,26	$6,3 (\pm 2,3) 10^{-8}$	3
JAC210	pJA15	Sak4 _{HK620}	$1,5 (\pm 0,8) 10^{-5}$	4	30	P<0,01	$3,9 (\pm 1,6) 10^{-8}$	3
JAC157	pJA4	-	$5,0 (\pm 2,7) 10^{-7}$	6	1		$2,2 (\pm 1,4) 10^{-8}$	4

vecteur pJA4 ($5,0 \times 10^{-7}$ Rif^R par cellule viable), même si l'explication de cette fréquence de base d'appariement d'ADN simple brin bactérien n'est pas comprise pour l'instant.

II. Les recombinares RecA et UvsX se comportent différemment pour l'appariement d'ADN simple brin *in vivo*.

L'expérience d'appariement de l'oligonucléotide Maj32 a ensuite été réalisée avec le plasmide exprimant la recombinaire bactérienne RecA (pJA5). Une fréquence de $1,9 \times 10^{-7}$ Rif^R par cellule viable, soit deux fois plus basse que l'activité résiduelle d'appariement d'ADN simple brin a été observée (Table 4). Une expérience similaire, consistant à tester la souche AB1157 codant le gène *recA* sauvage contenant le plasmide vecteur pJA4 aboutit à un résultat encore plus accentué. En effet, la fréquence obtenue est alors de $2,8 \times 10^{-8}$ Rif^R par cellule viable (Table 4), suggérant qu'il y a dans ces conditions une abolition totale de l'appariement d'ADN simple brin résiduel. Ces deux expériences montrent que RecA ne fait pas d'appariement d'ADN simple brin *in vivo*, et que RecA empêche même l'activité cellulaire résiduelle lorsqu'elle est exprimée à partir de ses propres signaux de transcription, sur le chromosome de la souche, et partiellement lorsqu'elle exprimée à partir des signaux de transcription du plasmide pJA5 (Table 4 et Figure 19). La complémentation partielle de RecA lorsqu'elle est codée sur le plasmide a aussi été observée dans l'expérience de complémentation de la sensibilité aux UV (voir le Matériel et Méthodes et la discussion). Ces résultats sont peut-être dus à la présence, dans les souches *recA* contenant pJA5, de moins de molécules de RecA par cellule que dans la souche sauvage.

La recombinaire UvsX du bactériophage T4 est souvent décrite comme étant l'analogue phagique de RecA (voir Introduction générale du manuscrit). Dans le test d'appariement d'ADN simple brin *in vivo*, UvsX est capable d'apparier

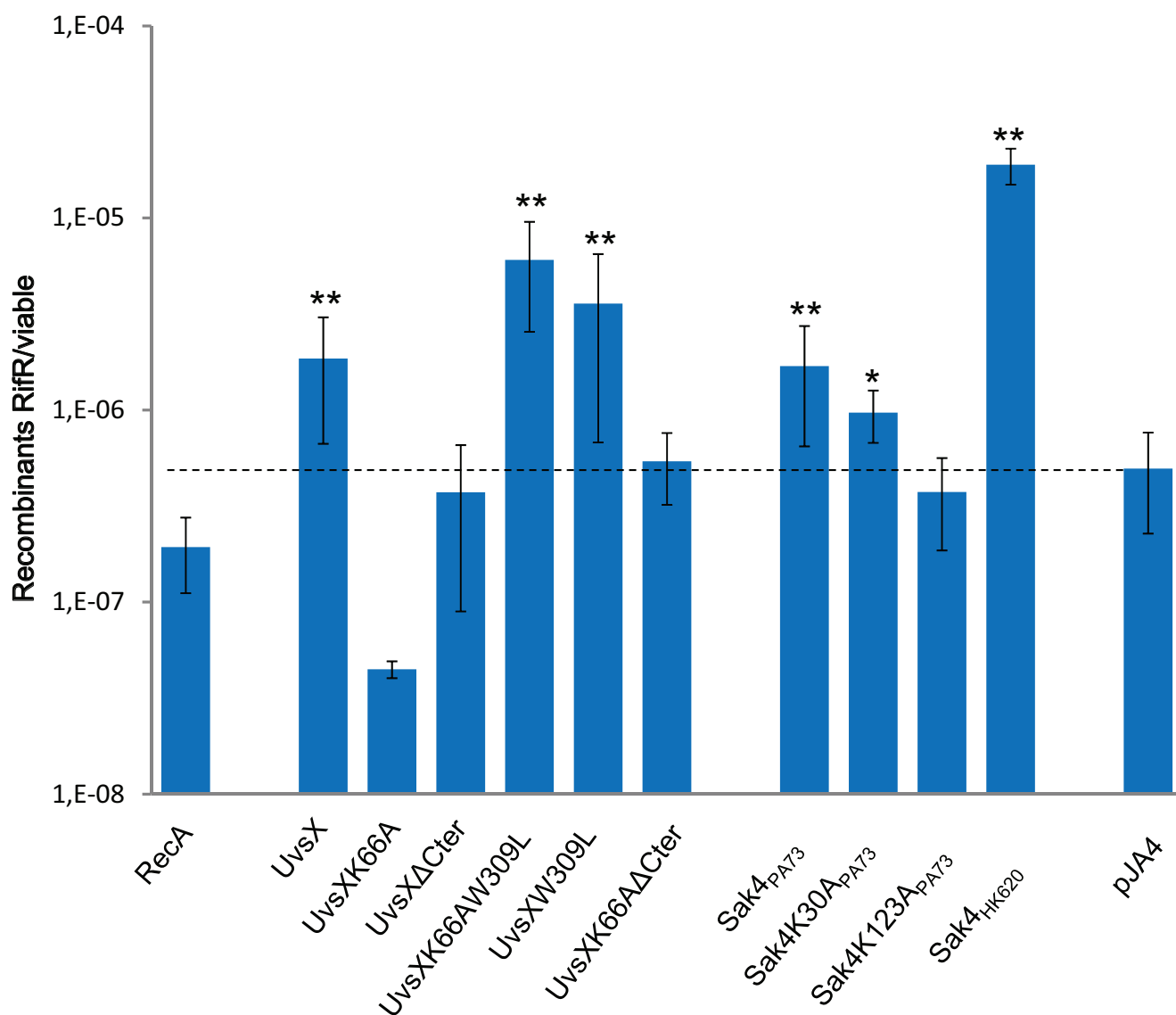


Figure 20: Activité d'appariement d'ADN simple brin *in vivo* des recombinaisons de la superfamille Rad51/RecA.

La fréquence de recombinaisons Rif^R est estimée par le ratio de recombinaisons rifampicine résistant par cellule viable. Les expériences ont été réalisées dans la souche AB1157 *recA* qui contient un plasmide dérivé du pSC101 qui exprime les différentes recombinaisons à tester après induction à l'arabinose. L'efficacité d'appariement d'ADN simple brin de chaque recombinaison est comparée à l'efficacité obtenue avec le plasmide qui n'exprime aucune recombinaison ('pJA4'), par un test de Student (un double astérisque indique un niveau de significativité à 1%, un simple astérisque indique un niveau de significativité à 5%). La ligne en pointillée indique la fréquence de recombinaisons Rif^R obtenue avec le plasmide pJA4.

l'oligonucléotide Maj32 avec une fréquence de $1,9 \times 10^{-6}$ Rif^R par cellule viable (Table 5 et Figure 20), soit 4 fois plus que l'activité basale cellulaire, et de manière significative (test de Student, $p=0,01$) par rapport à la fréquence obtenue avec le plasmide vecteur pJA4 ($5,0 \times 10^{-7}$ Rif^R par cellule viable). Cette fréquence est faible mais permet cependant de conclure qu'UvsX est capable d'apparier de l'ADN simple brin *in vivo*. Contrairement à ce qui était attendu, RecA et UvsX ont donc des propriétés différentes en ce qui concerne l'appariement d'ADN simple brin.

III. Le domaine ATPase d'UvsX est essentiel à l'appariement d'ADN simple brin *in vivo* contrairement à celui de Sak4.

Les recombinaisons UvsX du bactériophage T4 et Sak4 du bactériophage PA73 sont toutes les deux capables d'apparier de l'ADN simple brin *in vivo* à des fréquences respectives de $1,9 \times 10^{-6}$ et $1,7 \times 10^{-6}$ (Table 5 et Figure 20). Ces fréquences, bien que faibles, sont significativement plus élevées comparées à la fréquence obtenue avec le plasmide vecteur pJA4 ($5,0 \times 10^{-7}$ Rif^R par cellule viable), respectivement d'un facteur 4 et 3. Afin de confirmer que cette activité est bien dépendante des protéines UvsX ou Sak4, une mutation ponctuelle capable d'inactiver ces enzymes a été introduite dans leurs gènes respectifs. Le domaine ATPase d'UvsX et de Sak4 a été muté dans le motif Walker A (GXXXXGK^{T/S}) (Figure 21). L'activité de recombinaison d'UvsX K66A est abolie ($4,5 \times 10^{-8}$ Rif^R par cellule viable) (Table 5 et Figure 20). En revanche, Sak4 K30A est toujours capable d'apparier de l'ADN simple brin *in vivo* à une fréquence de $9,7 \times 10^{-7}$ Rif^R par cellule viable (Table 5 et Figure 20). Donc le domaine ATPase d'UvsX est essentiel à l'appariement d'ADN simple brin, contrairement à celui de Sak4. De manière intéressante, le mutant UvsX K66A provoque le même phénomène « d'abolition » de l'appariement d'ADN simple brin résiduel, comme observé pour la souche AB1157 codant le gène *recA* sauvage

(A)	RecA	48	LDIALG---AGGLPMGRIVEIY	GP	ESSGKT	TTLT	QVIAAAQRE--GKTCAFIDAEHALDP	102
			++IAL	GG+	G ++ +	GP	S K+ L +++++ R+ C F D+E + P	
	UvsX	39	MNIALSGEITGGMQSGLLI-LAGPSKSE	ES	SNFGLTMVSSYMRQYPDAVCLFYDSEFGITP			97
		103	IYARKLGVDIDNLLCSQPDTEQ----	ALEICDALARSGAVDVIVVDSVAALTPKAEIEG				158
			Y R +GVD + ++ + + EQ	+ DA+ R V V+ +DS+ L K E E				
		98	AYLRSMGVDPERVIHTFPVQSLEQLRIDMVNQLDAIERGEKV-VVFIDSLGNLASKKETED					156
		159	EIGDSHMG--LAARMMSQAMRKL	AGNLKQSN	TLLIFINQIRMKIGVMFGNPETTGGNAL			216
			+ + + A+ M R + N I IN + MF +T GG					
		157	ALNEKVVSDMTRAKTMKSLFRIVTPYFSTKNIPCI	AINHT-YETQEMFS--KTVMGGGTG				213
		217	KFYASVRLDIRRIGAVKEGENVVGSETRVKVVK	NKIAAPFKQAEFQILY	GEINFGYELV			276
			Y++ + I +K+G ++ G + + V K++ + + + GI+ Y L+					
		214	PMYSADTVFIIGKRQIKDGS	DLQGYQFVLNVEKSRTVKEKSKFFIDVKFDGGIDPYSGLL				273
		277	DLGVKEKLIEKA-GAWYS--YKGEKIGQGKANATAW					309
			D+ ++ + K WY+ + E+ G+ +W					
		274	DMALELGFVVKPKNGWYAREFLDEETGEMIREEKSW					309

(B)	RecA	65	IYGP	ESSGKT	TTL	76
			++G	SGKT	L	
	Sak4 _{PA73}	22	VFGAPG	SGKT	PL	33

Figure 21: Alignement de la recombinaise RecA d'*E. coli* avec la recombinaise d'UvsX du bactériophage T4 (A) et avec la Sak4 du bactériophage PA73 (B). En bleu le site de fixation à l'ATP (motif Walker A), en rouge l'acide aminé muté (Alignement Blastp, NCBI).

contenant le plasmide vecteur pJA4 ($2,8 \times 10^{-8}$ Rif^R par cellule viable) (voir Discussion).

IV. Le domaine C-ter d'UvsX est impliqué dans le mécanisme d'appariement simple brin.

UvsX et Sak4 diffèrent en particulier par la présence du domaine C-ter d'UvsX. Afin de comprendre le rôle de ce domaine dans l'activité d'appariement d'ADN simple brin, le domaine C-ter d'UvsX a été délété à partir du résidu G271. Pour RecA, ce domaine C-ter est supposé intervenir essentiellement dans l'interaction avec l'ADN double brin lors de la recherche d'homologie, ou dans la stabilisation de la D-loop (voir Introduction) Ces propriétés n'étant pas requises pour l'appariement d'ADN simple brin, et Sak4 étant capable d'apparier l'ADN simple brin en l'absence d'un tel domaine, nous nous sommes demandés si la perte de ce domaine dans UvsX affectait son activité d'appariement d'ADN simple brin *in vivo*. Le résultat d'appariement d'ADN simple brin montre qu'UvsX Δ Cter n'est pas capable d'apparier l'oligonucléotide Maj32 ($3,7 \times 10^{-7}$ Rif^R par cellule viable) (Table 5 et Figure 20), indiquant que le domaine C-ter d'UvsX est impliqué dans le mécanisme d'appariement d'ADN simple brin. Il faut noter cependant que ce mutant n'abolit pas, contrairement au mutant ATPase, l'appariement d'ADN simple brin résiduel. Ceci suggère que la protéine UvsX Δ Cter ne peut plus se fixer à l'ADN simple brin, laissant le substrat libre pour la protéine qui permet l'appariement d'ADN simple brin résiduel. Les expériences de marquage (voir Matériel et Méthodes) indiquent que la protéine UvsX Δ Cter est produite à un niveau comparable aux autres constructions. Bien qu'un défaut de repliement soit envisageable, il semble que ce ne soit pas le cas, puisque la protéine serait alors probablement dégradée par les protéases cellulaires. Or, nous avons observé qu'elle est stable pendant au moins 30 minutes après la chasse du marqueur radioactif.

Un autre mutant de la protéine UvsX suggère une interaction forte entre ce domaine C-ter d'UvsX et le reste de la protéine. En effet, lors de la construction de la protéine UvsX K66A, une mutation intragénique a été obtenue spontanément, conduisant à l'introduction de la mutation W309L dans le domaine C-ter. De manière surprenante, le mutant UvsX K66A W309L est capable d'apparier l'oligonucléotide Maj32 avec une fréquence équivalente à la protéine sauvage, de $6,1 \times 10^{-6}$ Rif^R par cellule viable, soit 12 fois plus que la fréquence obtenue avec le plasmide vecteur pJA4 ($5,0 \times 10^{-7}$ Rif^R par cellule viable) (Table 5 et Figure 20). Cette mutation W309L restaure donc l'activité d'UvsX K66A. Le simple mutant codant la protéine UvsX W309L a alors été construit et cette mutation n'affecte pas l'activité d'UvsX ($3,6 \times 10^{-6}$ Rif^R par cellule viable) (Table 5 et Figure 20). Nous nous sommes alors demandé si la délétion complète du domaine C-ter pouvait avoir un effet similaire à la mutation W309L sur la protéine UvsX K66A. La protéine UvsX K66A Δ Cter n'est pas capable d'apparier l'ADN simple brin ($5,4 \times 10^{-7}$ Rif^R par cellule viable) (Table 5 et Figure 20), il n'y a donc pas restauration de l'activité d'UvsX K66A par délétion de son domaine C-ter.

En conclusion, UvsX est capable d'apparier de l'ADN simple brin *in vivo* et son domaine C-ter est impliqué dans ce mécanisme. De plus son domaine ATPase est important pour cette activité, et un révertant capable de restaurer l'activité d'une protéine UvsX mutée dans son domaine ATPase a naturellement été sélectionnée.

V. L'activité de la recombinaise Sak4 est inactivée par la mutation K123A

Afin de prédire les résidus de Sak4 potentiellement impliqués dans l'interaction avec l'ADN, R. Guerois (IbITecS, CEA, Gif sur Yvette) a modélisé la structure en trois dimensions de la recombinaise Sak4 du bactériophage PA73 à partir de celle réalisée sur le filament de RecA-ADNsb. Il a prédit que le résidu K123 devait être particulièrement proche de la molécule d'ADN simple brin et donc être un bon candidat pour un résidu nécessaire à l'interaction de Sak4 avec l'ADN (Figure 22).

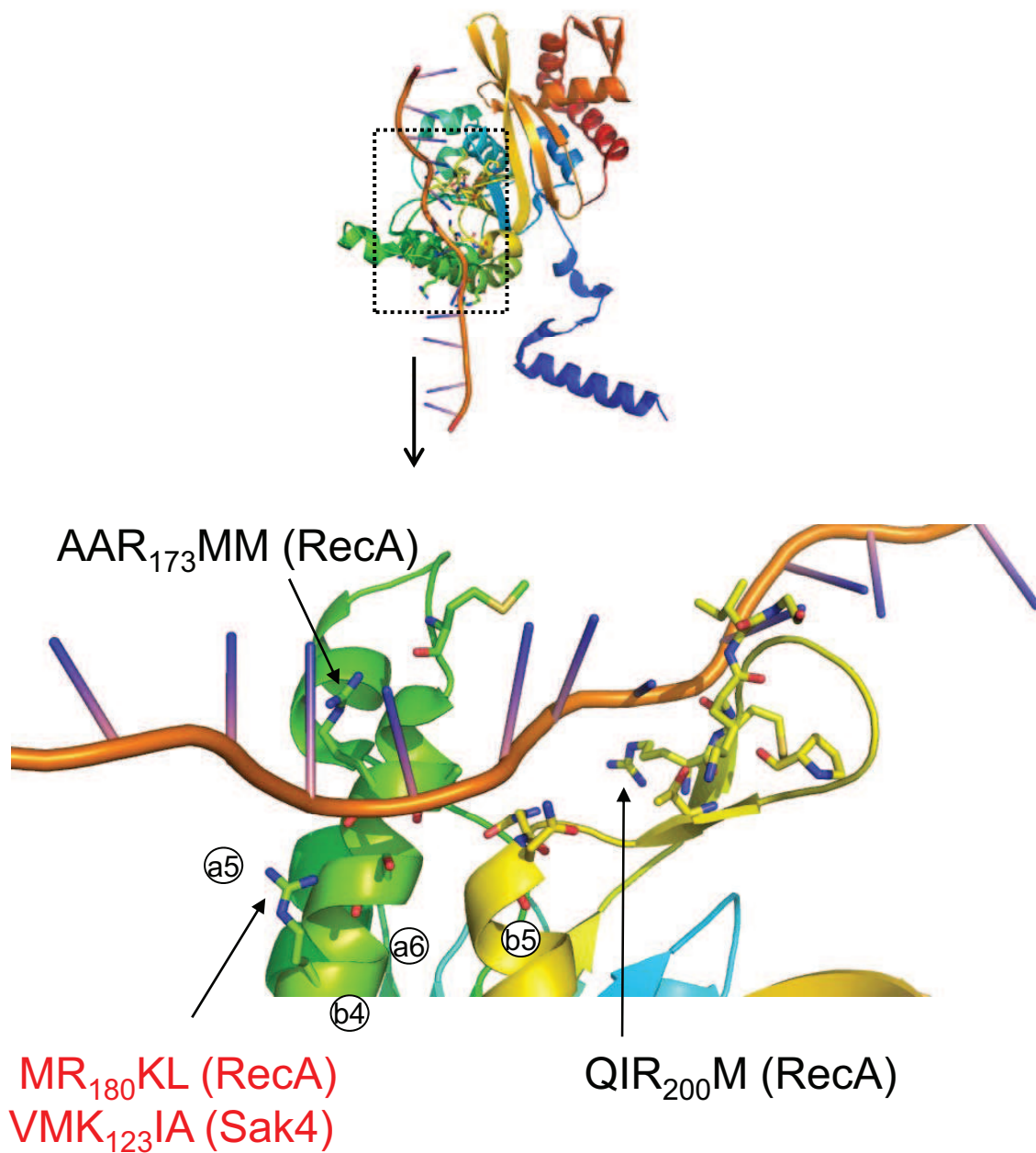


Figure 22: Prédiction du résidu de Sak4 interagissant avec l'ADN. Structure tridimensionnelle de RecA, dans la région d'interaction avec l'ADN simple brin. Les trois résidus les plus proches de l'ADN sont fléchés. Seul R180 de RecA est à une position conservée dans Sak4: il s'agit du résidu K123. Données: R. Guerois.

Nous avons ainsi remplacé ce résidu K123 par un alanine, et testé l'appariement d'ADN simple brin de la nouvelle protéine. De fait, Sak4 K123A n'est plus capable d'apparier de l'ADN simple brin ($3,7 \times 10^{-7}$ Rif^R par cellule viable) (Table 5 et Figure 20). Ceci confirme que Sak4 est capable d'apparier de l'ADN simple brin et montre que le résidu K123 est essentiel à cette activité.

VI. La recombinaise Sak4 du bactériophage HK620 apparie efficacement l'ADN simple brin en présence de ses cofacteurs.

A cette étape du travail, l'activité de la recombinaise Sak4 du bactériophage PA73 étant insuffisante pour permettre les tests d'appariement avec des séquences divergées, nous avons recherché parmi les protéines prédites dans l'Article I (Lopes et al.), une protéine de la famille Sak4 ayant éventuellement une meilleure activité *in vivo*. Pour cela, nous avons sélectionné le bactériophage HK620 infectant *E. coli*, et son gène *sak4* a été cloné dans le plasmide vecteur pJA4, ainsi que ses trois gènes voisins *hkaJ*, *hkaK* et *hkaM*, des cofacteurs potentiels. En effet, une étude de Clark et al, en 2001, comparant l'organisation génomique du bactériophage HK620 à celle du bactériophage P22 qui infecte *Salmonella typhimurium* a montré que le gène *hkaJ* possède 95% d'identité de séquence avec le gène *abc2* de P22. La protéine Abc2 possède une activité inhibitrice de la nucléase RecBCD (Clark et al., 2001). Par contre, aucune homologie avec une protéine de fonction connue n'a été identifiée pour l'instant pour les gènes *hkaK* et *hkaM*. Cependant, nous avons préféré cloner ces 4 gènes, afin de préserver l'organisation génomique autour de Sak4. L'expérience d'appariement de l'oligonucléotide Maj32 a été réalisée et une fréquence de $1,5 \times 10^{-5}$ Rif^R par cellule viable a été obtenue, soit 9 fois plus que pour la Sak4 de PA73 ($1,7 \times 10^{-6}$ Rif^R par cellule viable) (Table 5 et Figure 20). Nous avons ensuite testé l'effet de la longueur de l'oligonucléotide sur l'efficacité de l'appariement. Des oligonucléotides contenant tous la même mutation que celle de

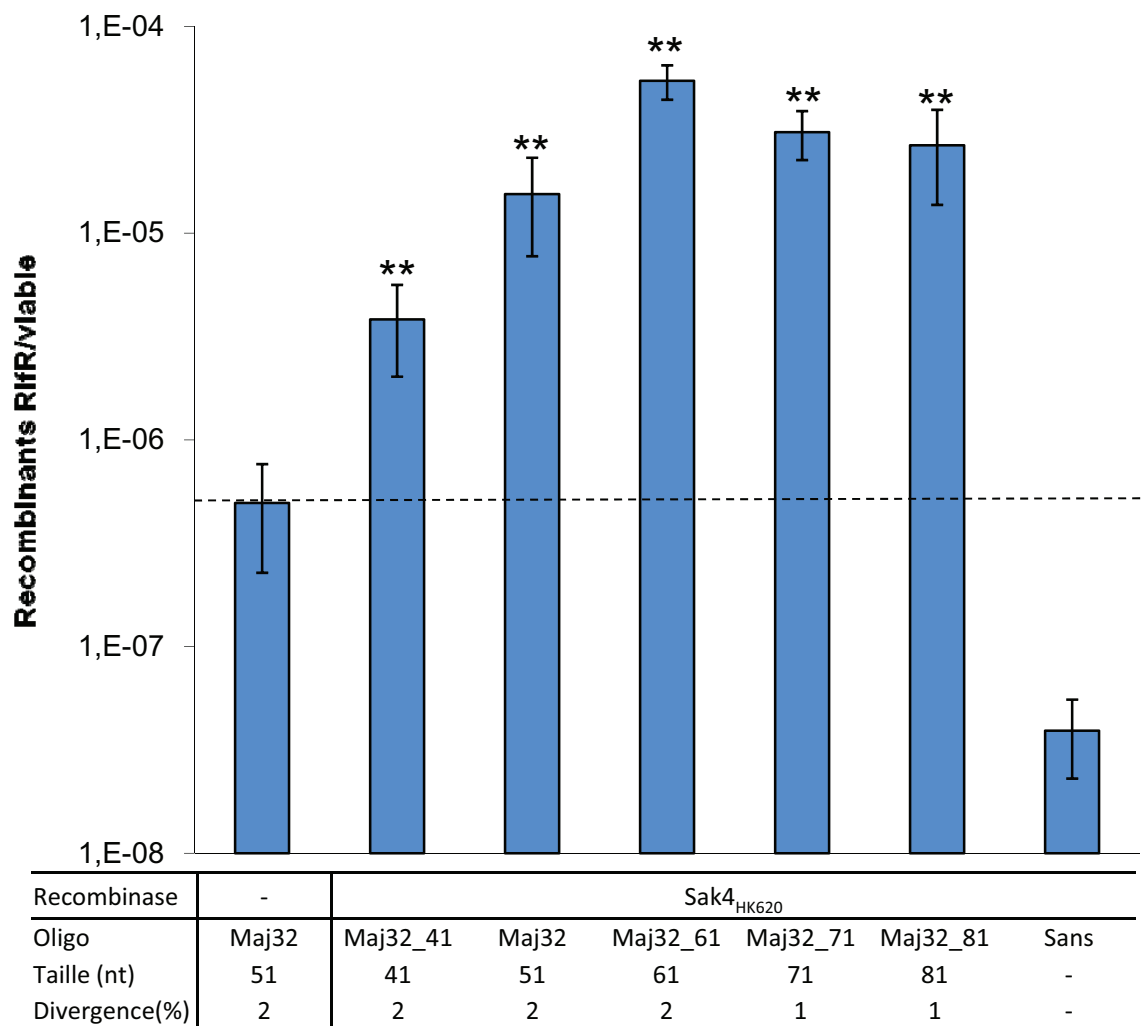


Figure 23: Activité d'appariement d'ADN simple brin *in vivo* de la recombinaise Sak4 du phage HK620 en fonction de la taille de l'oligonucléotide.

La fréquence de recombinants Rif^R est estimée par le ratio de recombinants rifampicine résistant par cellule viable. Chaque valeur correspond à la moyenne d'au moins 3 expériences, réalisées dans la souche AB1157 *recA* pJA15 qui exprime la recombinaise Sak4 après induction à l'arabinose. L'efficacité d'appariement d'ADN simple brin de Sak4 est comparée à l'efficacité obtenue avec le plasmide pJA4 qui n'exprime aucune recombinaise, par un test de Student (un double astérisque indique un niveau de significativité à 1%). La ligne en pointillée indique la fréquence de recombinants Rif^R obtenue avec le plasmide pJA4.

Maj32, mais variant en taille de 41 à 81 nt ont été testés (Table 2). La recombinaise Sak4 du bactériophage HK620 en présence de ses cofacteurs, apparie l'ADN simple brin avec une efficacité aussi élevée pour tous les oligonucléotides d'une taille supérieure ou égale à 51 nt (Figure 23). En revanche, l'efficacité diminue d'un facteur 5 pour l'oligonucléotide de 41 nt.

VII. Les recombinaises Red β et Sak4 de HK620 appariant l'ADN simple brin divergent *in vivo*

Nos résultats mettent en évidence que les recombinaises phagiques Sak4 des bactériophages PA73 et HK620 sont capables d'apparier de l'ADN simple brin de séquence identique à celle du chromosome, à l'exception de la base qui détermine le changement de phénotype. Nous avons ensuite voulu savoir si ces recombinaises phagiques étaient aussi capables de recombiner de l'ADN simple brin de séquence divergée. En effet, dans un système expérimental de recombinaison homologue entre séquences inversées placées sur le génome de Lambda, il a déjà été décrit que Red β était 100 fois plus efficace que RecA pour recombiner de l'ADN divergé à 22% (Martinson et al., 2008). Nous avons recherché si cette capacité de Red β s'observe également dans le test d'appariement simple brin, et si Sak4 a également une telle capacité. Pour cela, le même test d'appariement d'ADN simple brin a été répété avec différents oligonucléotides de séquences divergées, dérivés du Maj32. Ces oligonucléotides de séquences divergées par rapport à *rpoB* sont de longueur de 51 ou 81 nt et possèdent des mutations neutres réparties sur toute leur longueur pour obtenir 8, 14 et 22% de divergence (Table 2). Les expériences ont été réalisées dans un fond génétique *mutS*, afin de supprimer l'édition des intermédiaires appariés par le système de réparation des mésappariements MutLS-dépendant (voir Introduction).

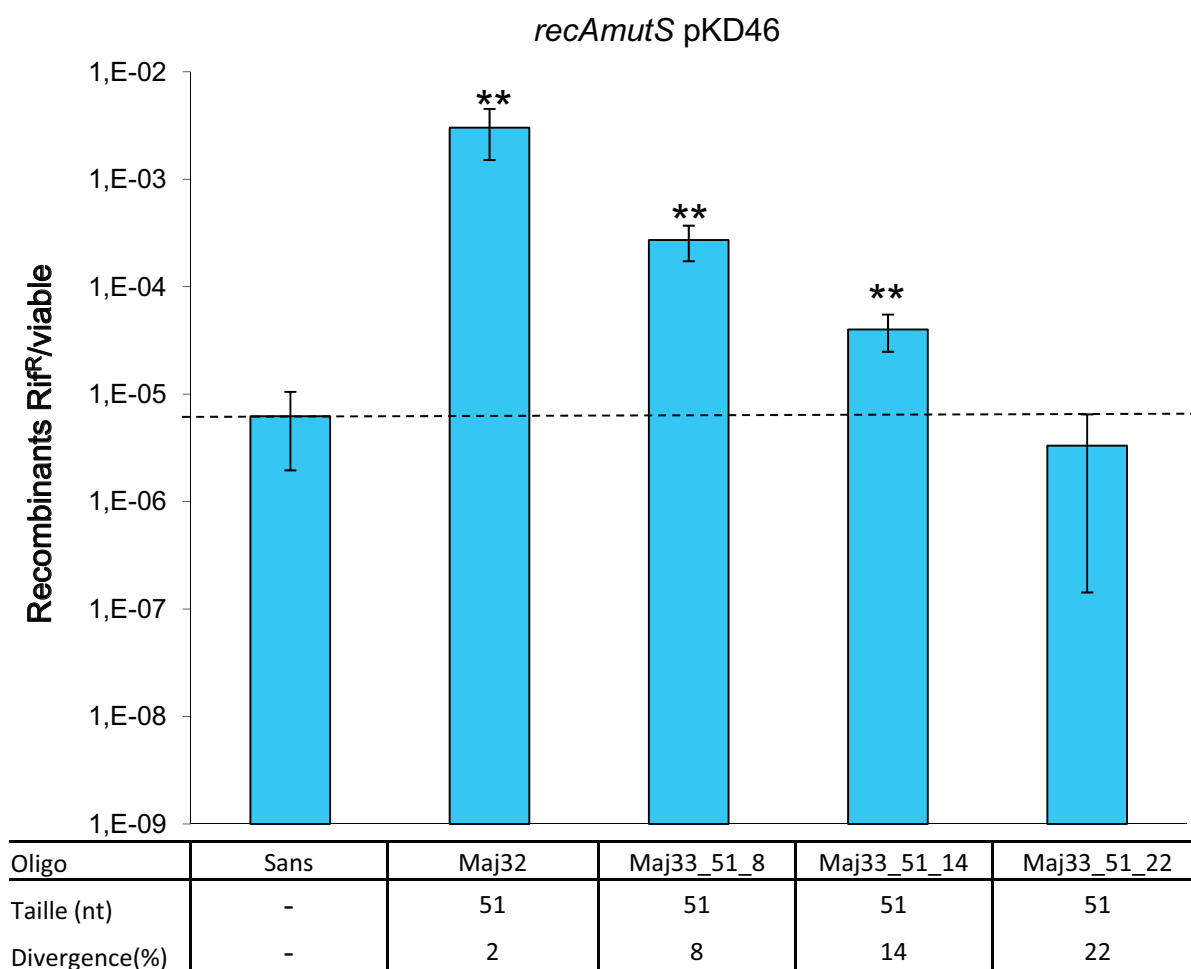


Figure 24 : Activité d'appariement d'ADN simple brin *in vivo* par Red β , en fonction du pourcentage de divergence de l'oligonucléotide, en contexte AB1157 *recAmutS*.

La fréquence de recombinants Rif^R est estimée par le ratio de recombinants rifampicine résistant par cellule viable. Les expériences ont été réalisées en présence de pKD46 qui exprime la recombinaise Red β et ses cofacteurs, après induction à l'arabinose. L'efficacité d'appariement d'ADN simple brin est comparée à l'efficacité obtenue sans oligonucléotide (mutants rifampicine spontanés) par un test de Student (un double astérisque indique un niveau de significativité à 1%). La ligne en pointillé indique le taux de mutants Rif^R spontanés.

Dans un premier temps, ces différents oligonucléotides ont été testés sur l'activité résiduelle d'appariement d'ADN simple brin observée avec la souche AB1157 *recA mutS* contenant le plasmide vecteur pJA4. Comme attendu, un bruit de fond de mutants spontanés Rif^R 120 fois plus élevé ($2,7 \times 10^{-6}$ Rif^R par cellule viable) en contexte *recA mutS* est observé par rapport au contexte *recA mutS+* ($2,2 \times 10^{-8}$ Rif^R par cellule viable). En conséquence, l'activité résiduelle d'appariement d'ADN simple brin n'est pas décelable en ce contexte. Par la suite, en contexte *recA mutS*, les fréquences obtenues sont donc comparées au taux de Rif^R spontanés.

Nous avons ensuite testé le plasmide exprimant Red β ainsi que Red α , Gam et Orf60a (plasmide pKD46), (la présence des cofacteurs permet une meilleure efficacité d'appariement d'ADN simple brin, voir plus bas), en contexte génétique AB1157 *recA mutS*. En présence de l'oligonucléotide Maj32, la fréquence est de $3,0 \times 10^{-3}$ Rif^R par cellule viable (Table 6 et Figure 24), ce qui est comparable à la valeur obtenue en contexte MutS+ ($1,7 \times 10^{-3}$ Rif^R par cellule viable). En effet, que le système MMR soit actif ou non, il n'est pas attendu d'effet sur la fréquence de recombinaison homologue, puisqu'il y a absence de mésappariement reconnaissable par MutS. En présence des oligonucléotides divergents à 8, 14 et 22% les fréquences respectives sont $2,7 \times 10^{-4}$, $4,0 \times 10^{-5}$ et $3,3 \times 10^{-6}$ Rif^R par cellule viable (Table 6 et Figure 24). Il y a donc 10 fois moins de recombinants lorsque l'oligonucléotide est divergé à 8% et 75 fois moins lorsque l'oligonucléotide est divergé à 14%, en comparaison avec la fréquence obtenue avec Maj32. De plus, les fréquences obtenues à 8 et 14% sont significatives comparées au taux de Rif^R spontanés ($6,2 \times 10^{-6}$ Rif^R par cellule viable) (Test de student, $p < 0,01$) (Table 6). Red β permet donc l'appariement d'ADN simple brin divergé jusqu'à 14%, mais son activité sur de l'ADN divergé à 22% n'est pas détectable dans ce système, car le bruit de fond de mutant rifampicine spontané est trop élevé ($6,2 \times 10^{-6}$ Rif^R par cellule viable). La même expérience réalisée en contexte MutS+ n'a pas permis non plus de

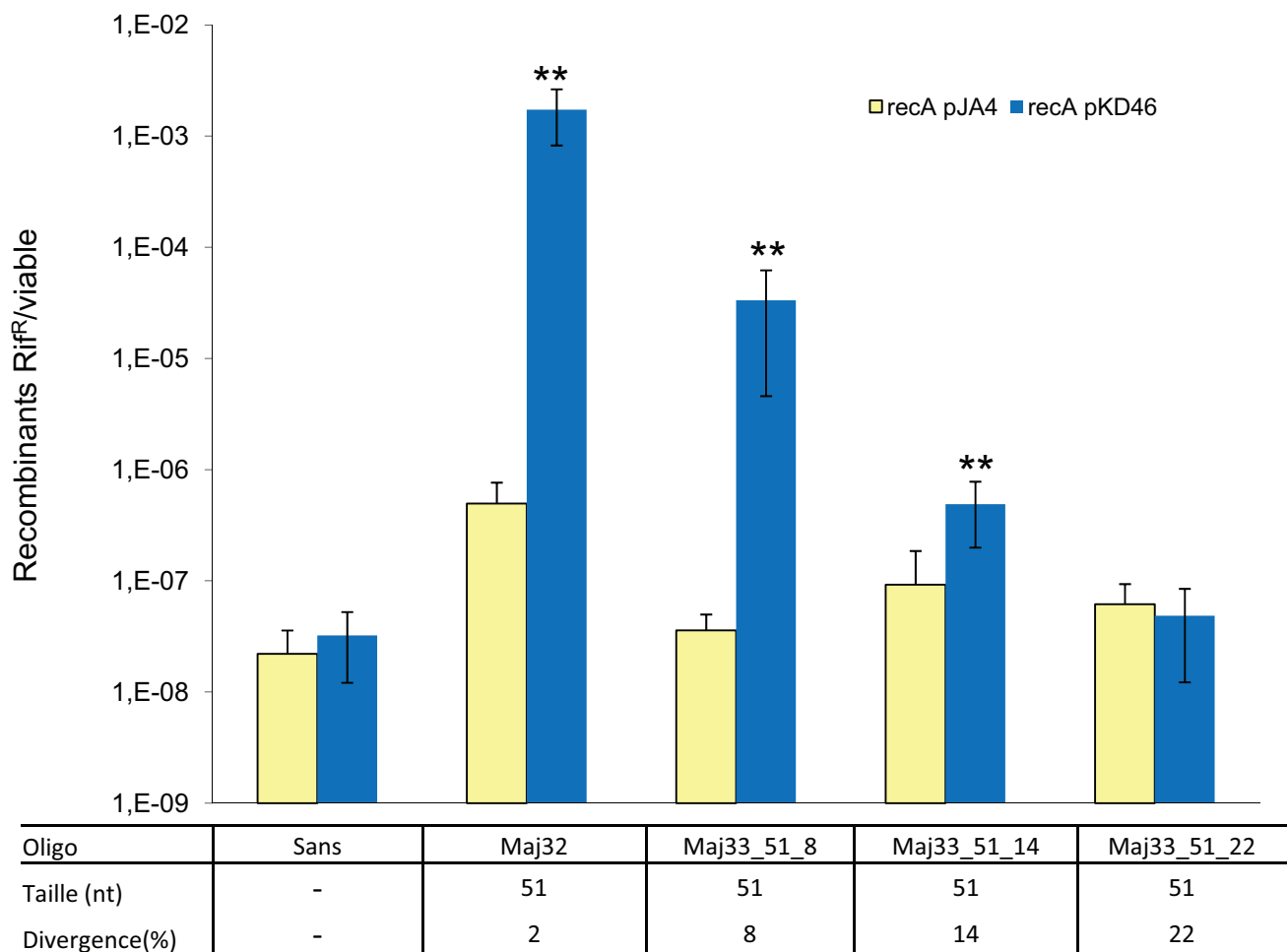


Figure 25: Activité d'appariement d'ADN simple brin *in vivo* par Red β , en fonction du pourcentage de divergence de l'oligonucléotide, en contexte AB1157 *recA*. En bleu, la fréquence de recombinants Rif^R est estimée par le ratio de recombinants rifampicine résistant par cellule viable avec le plasmide pKD46 qui exprime la recombinaise Red β , en présence de ces cofacteurs, après induction à l'arabinose. En jaune, la fréquence de recombinants Rif^R par cellule viable avec le plasmide pJA4 qui n'exprime pas de recombinaise. L'efficacité d'appariement d'ADN simple brin est comparée à l'efficacité obtenue avec pJA4 par un test de Student (un double astérisque indique un niveau de significativité à 1%).

détecter de recombinaison à ce taux de divergence, car l'effet du MMR affecte la détection des recombinants (Table 7 et Figure 25).

Nous avons ensuite testé la recombinaise Sak4 du bactériophage HK620 en présence de ses cofacteurs, en contexte génétique AB1157 *recA mutS*, avec les oligonucléotides cités ci-dessus de 51 nt et de divers taux de divergence (Table 2). Une fréquence de recombinants de $1,4 \times 10^{-4}$ Rif^R par cellule viable (Table 8 et Figure 26) a été obtenue pour l'oligonucléotide Maj32. Lorsque l'oligonucléotide possède 8% et 14% de divergence (Maj33_51_8 et Maj33_51_14), la fréquence de recombinaison est respectivement de $1,3 \times 10^{-5}$ et de $8,9 \times 10^{-6}$ Rif^R par cellule viable (Table 8 et Figure 26). La différence est significative à 5% par rapport au taux de Rif^R spontanés ($8,2 \times 10^{-6}$ Rif^R par cellule viable) à 8% de divergence (Test de student, $p=0,03$), mais pas à 14% (Test de student, $p=0,33$), (Table 8).

Des oligonucléotides de même taux de divergence mais plus longs (81 nt) ont ensuite été testés en contexte génétique AB1157 *recA mutS*. En effet, même si la longueur n'augmente pas la fréquence d'appariement d'un oligonucléotide identique à sa cible pour Sak4 ($1,6 \times 10^{-5}$ Rif^R par cellule viable avec Maj32 et $2,7 \times 10^{-5}$ Rif^R par cellule viable avec Maj32_81) (Figure 23), il est possible qu'en présence de mésappariements, un oligonucléotide plus long favorise un hétéroduplex plus stable et ainsi un meilleur appariement. Avec l'oligonucléotide de 81 nt et 9% de divergence (Maj33_81_9) une fréquence de $1,0 \times 10^{-5}$ Rif^R par cellule viable est obtenue, celle-ci n'est pas significativement différente du bruit de fond. Cependant, à 13% de divergence, une fréquence significativement supérieure au bruit de fond ($2,4 \times 10^{-5}$ Rif^R par cellule viable) a été obtenue (Table 8 et Figure 26). Nous pouvons donc conclure que la recombinaise Sak4 du bactériophage HK620, en présence de ses cofacteurs, est capable d'apparier de l'ADN ayant 13% de divergence, à condition que le système MMR soit inactif et que la longueur de l'oligonucléotide soit de 81 nt.

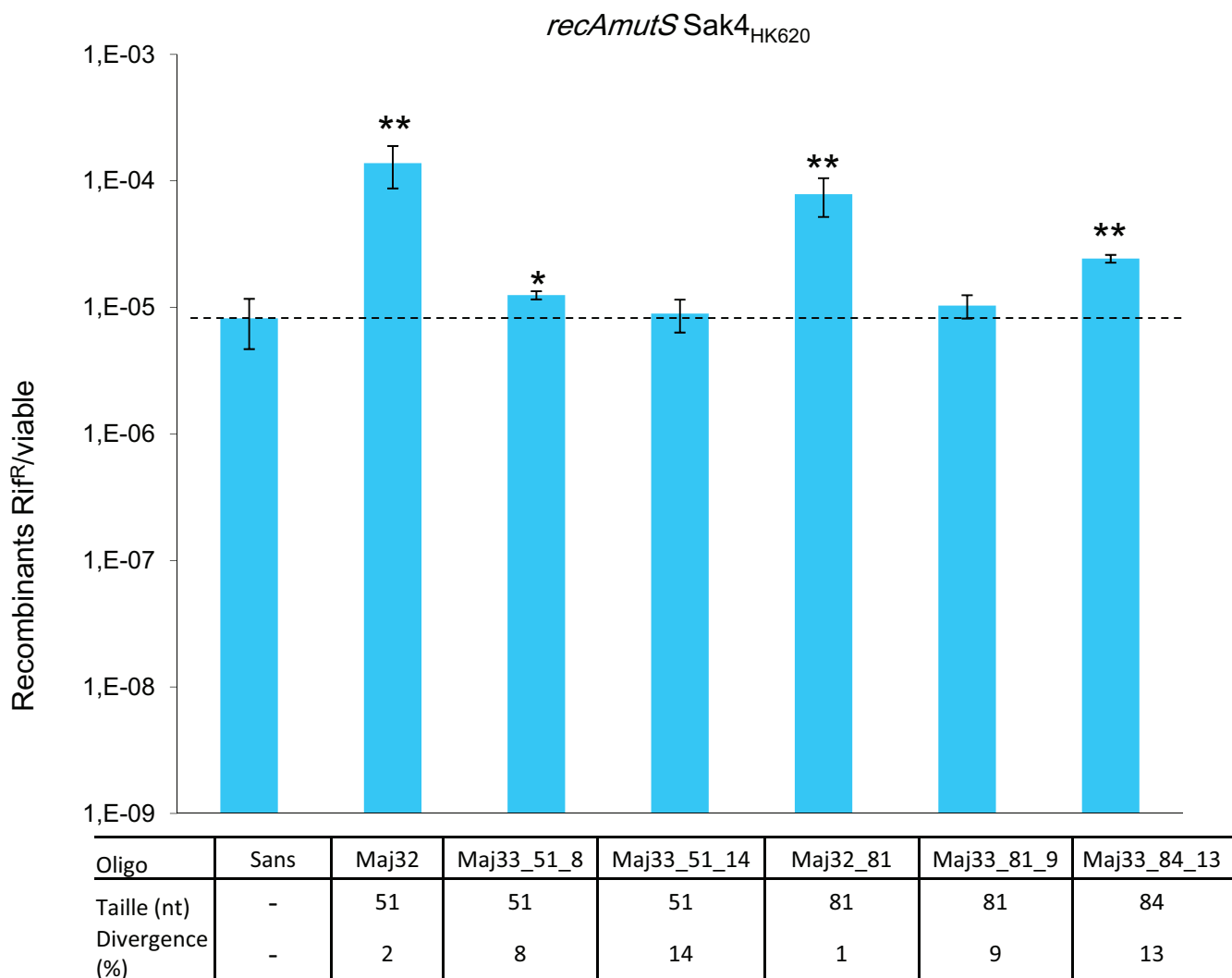


Figure 26: Activité d'appariement d'ADN simple brin *in vivo* par Sak4_{HK620}, en fonction du pourcentage de divergence de l'oligonucléotide, en contexte AB1157 *recAmutS*.

La fréquence de recombinants Rif^R est estimée par le ratio de recombinants rifampicine résistant par cellule viable. Les expériences ont été réalisées en présence de pJA15 qui exprime la recombinaise Sak4_{HK620}, en présence de ses cofacteurs, après induction à l'arabinose. L'efficacité d'appariement d'ADN simple brin est comparée à l'efficacité obtenue sans oligonucléotide (mutants rifampicine spontanés) par un test de Student (un double astérisque indique un niveau de significativité à 1%, un simple astérisque indique un niveau de significativité à 5%). La ligne en pointillé indique le taux de mutants Rif^R spontanés.

VIII. Gam stimule l'activité d'appariement d'ADN simple brin de Red β

Les principaux résultats publiés dans la littérature concernant l'appariement d'ADN simple brin par Red β ont été obtenus avec des constructions exprimant Red α et Gam en plus de la recombinaise. Sur la base des connaissances acquises sur ces deux protéines, il n'était pas supposé qu'elles puissent agir sur un substrat simple brin. Cependant, devant la forte différence d'efficacité d'appariement d'ADN simple brin observée pour Red β (et ses cofacteurs) et toutes les autres recombinaises testées, nous avons voulu tester l'effet des gènes au voisinage de Red β sur le génome de Lambda.

En contexte *recA*, en présence du plasmide pKD46, qui exprime les protéines Gam, Red α et Orf60a en plus de Red β (Figure 27), une fréquence de $1,7 \times 10^{-3}$ Rif^R par cellule viable est observée (Table 9). Cependant, on observe 25 fois moins de recombinants Rif^R ($7,0 \times 10^{-5}$ Rif^R par cellule viable) lorsque le plasmide n'exprime que Red β (Table 9 et Figure 27). Afin de discriminer la ou les protéines impliquées dans cette différence d'activité, les plasmides pJA4 *red β -red α -orf60a* (sans *gam*), pJA4 *gam-red β -red α* (sans *orf60a*) et pJA4 *gam-red β* (sans *exo* et *orf60a*) ont été construits. Avec le plasmide sans *gam*, l'efficacité d'appariement d'ADN simple brin ($6,4 \times 10^{-5}$ Rif^R par cellule viable) est réduite au niveau de celle du plasmide ne contenant que Red β ($7,0 \times 10^{-5}$ Rif^R par cellule viable) (Table 9 et Figure 27). Nous concluons que Gam est nécessaire à la stimulation de l'appariement d'ADN simple brin médié par Red β . Afin de confirmer ce résultat, le plasmide pJA4 *gam-red β* a été testé: une fréquence non significativement différente de celle obtenue avec pKD46 a été observée (Table 9 et Figure 27). Nous concluons donc que Gam semble nécessaire et suffisant pour cette stimulation. De manière intéressante, l'ajout de l'exonuclease Red α sur le plasmide pJA4 *gam-red β* (plasmide pJA4 *gam-red β -red α*), provoque à nouveau une légère diminution d'efficacité de l'appariement d'ADN

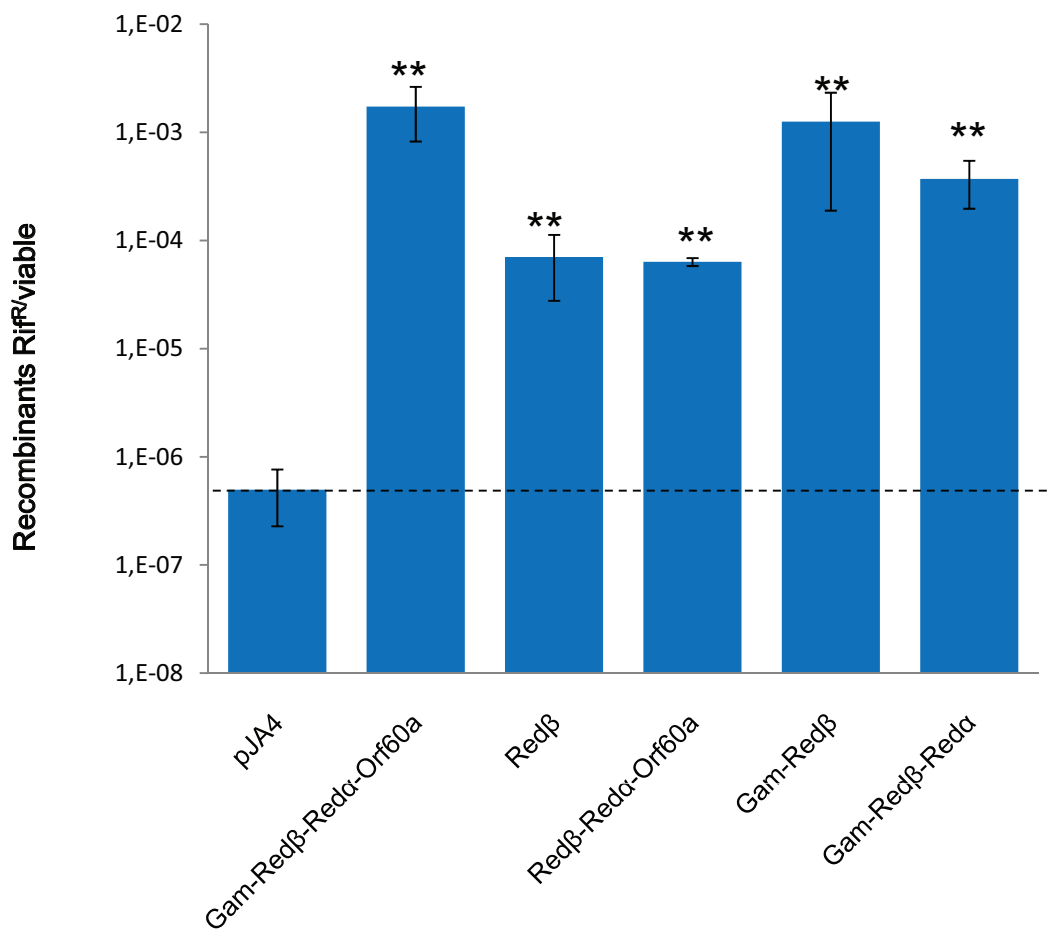


Figure 27: Activité d'appariement d'ADN simple brin *in vivo*, en fonction de la présence de différents gènes au voisinage de redβ, sur le génome de Lambda. L'efficacité d'appariement d'ADN simple brin est comparée à l'efficacité obtenue avec pJA4, par un test de Student (un double astérisque indique un niveau de significativité à 1%).

simple brin de 5 fois (fréquence de $3,7 \times 10^{-4}$ Rif^R par cellule viable) comparé à pKD46 (Table 9 et Figure 27).

En conclusion, les principaux résultats qui se dégagent de ce chapitre, et qui seront discutés dans la dernière partie de cette thèse sont les suivants :

- Le mécanisme d'appariement d'ADN simple brin de Red β est plus efficace en présence de Gam.
- Tout comme Red β , la recombinaise Sak4 du bactériophage HK620 est capable d'apparier deux molécules d'ADN simple brin de séquences divergées de 13%.
- La superfamille de recombinaises Rad51 comporte deux types de recombinaises ayant des propriétés distinctes : le type UvsX fait de l'appariement simple brin de manière ATP-dépendante, contrairement aux recombinaises de type Sak4.

Table 6 - Efficacité de recombinaison avec pKD46 en contexte *recA mutS*, en fonction de la divergence entre séquences appariées.

Oligo	Divergence (%)	Rif ^R / viable [N]	Facteur relatif à sans oligo	Test Student (1)
Sans		6,2 (± 4,3) 10 ⁻⁶ [5]	1	
Maj32	2	3,0 (± 1,5) 10 ⁻³ [3]	480	P<0,01
Maj33_51_8	8	2,7 (± 1,0) 10 ⁻⁴ [3]	40	P<0,01
Maj33_51_14	14	4,0 (± 1,5) 10 ⁻⁵ [3]	6	P<0,01
Maj33_51_22	22	3,3 (± 3,2) 10 ⁻⁶ [3]	0,5	P=0,18

(1) Pour le test de Student, les comparaisons sont faites avec le taux de mutants Rif^R spontanés obtenu dans la souche *recA mutS* sans oligo (ligne 1 du tableau)

Table 7 - Efficacité de recombinaison avec pKD46 en contexte *recA* (souche MAC1042), en fonction de la divergence entre séquences appariées.

Oligo	Divergence (%)	Rif ^R / viable avec pKD46 [N]	Facteur (pKD46/pJA4)	Test Student (1)	Rif ^R / viable avec pJA4 [N]
aucun		3,2 (± 2,0) 10 ⁻⁸ [12]			2,2 (± 1,4) 10 ⁻⁸ [4]
Maj32	2	1,7 (± 0,9) 10 ⁻³ [6]	3400	P<0,01	5,0 (± 2,7) 10 ⁻⁷ [6]
Maj33_51_8	8	3,3 (± 2,9) 10 ⁻⁵ [7]	920	P=0,01	3,6 (± 1,4) 10 ⁻⁸ [5]
Maj33_51_14	14	4,9 (± 2,9) 10 ⁻⁷ [6]	5	P<0,01	9,2 (± 9,3) 10 ⁻⁸ [6]
Maj33_51_22	22	4,8 (± 3,6) 10 ⁻⁸ [3]	0,8	P=0,33	6,2 (± 3,1) 10 ⁻⁸ [3]

(1) Pour le test de Student, les comparaisons sont faites par rapport à la recombinaison résiduelle (pJA4, souche JAC157)

Table 8 – Activité d'appariement d'ADN simple brin *in vivo* pour pJA15 (Sak4_{HK620}), en contexte *recA mutS* (souche JAC212), en fonction de la divergence des séquences appariées.

Oligo	Divergence (%)	Rif ^R / viable	Nb exp	Facteur relatif à sans oligo	Test de Student par rapport sans oligo
Maj32	2	1,4 (± 0,5) 10 ⁻⁴	3	17	P<0,01
Maj33_51_8	8	1,3 (± 0,1) 10 ⁻⁵	3	1,5	p=0,03
Maj33_51_14	14	8,9 (± 2,6) 10 ⁻⁶	6	1	p=0,33
Maj32_81	1	7,9 (± 2,7) 10 ⁻⁵	3	10	P<0,01
Maj33_81_9	9	1,0 (± 0,2) 10 ⁻⁵	7	1,2	p=0,09
Maj33_84_13	13	2,4 (± 0,2) 10 ⁻⁵	3	3	P<0,01
Sans		8,2 (± 3,5) 10 ⁻⁶	9		

Table 9 – Activité d'appariement d'ADN simple brin *in vivo* en présence de divers cofacteurs de Redβ.

Souche	Redβ ± cofacteurs	Rif ^R / viable avec oligo Maj32	Nb exp	Facteur Relatif à MAC1042)	Test de Student par rapport à JAC157	Rif ^R /viable sans oligo	Nb exp
MAC1042	Gam-Redβ-Redα-Orf60a (pKD46)	1,7 (± 0,9) 10 ⁻³	6	1	P<0,01	3,4 (±2,3) 10 ⁻⁸	9
JAC182	Redβ	7,0 (± 4,3) 10 ⁻⁵	6	0,04	P<0,01	1,5 (±0,5) 10 ⁻⁸	3
JAC236	Gam-Redβ-Redα	3,7 (± 1,7) 10 ⁻⁴	5	0,21	P<0,01	3,3 (±0,9) 10 ⁻⁸	3
JAC239	Redβ-Redα-Orf60a	6,4 (± 0,5) 10 ⁻⁵	3	0,04	P<0,01	1,9 (±0,5) 10 ⁻⁸	4
JAC237	Gam-Redβ	1,3 (± 1,1) 10 ⁻³	3	0,75	P<0,01	6,8 (± 5,3) 10 ⁻⁸	3
JAC157	Aucune recombinaison	5,0 (± 2,7) 10 ⁻⁷	6		P<0,01	2,2 (± 1,4) 10 ⁻⁸	4

DISCUSSION

Mon travail de thèse a consisté à caractériser différentes recombinaisons phagiques *in vivo*, en prenant Red β comme référence. J'ai recherché s'il existait des particularités de ces recombinaisons qui les différencieraient de la recombinaison bactérienne RecA, en me concentrant sur l'activité d'appariement simple brin. Je reprendrai dans cette discussion les différentes conclusions auxquelles m'a amené ce travail qui indiquent effectivement certaines particularités des recombinaisons phagiques, et qui pourraient expliquer la plus grande plasticité des génomes de bactériophages, en comparaison aux génomes bactériens.

I. *Escherichia coli* possède une activité d'appariement d'ADN simple brin en contexte *recA*

Dans un premier temps, nous avons observé qu'il y avait dans la cellule une activité résiduelle d'appariement d'ADN simple brin *in vivo*. En effet, lors de la transformation dans une souche *recA* et en absence d'ADN simple brin, le taux de mutation spontanée est de $2,2 \times 10^{-8}$ Rif^R par cellule viable, tandis qu'en présence d'ADN simple brin ce taux est augmenté d'un facteur 22 ($5,0 \times 10^{-7}$ Rif^R par cellule viable). Il y a donc dans la cellule une activité d'appariement d'ADN simple brin à faible fréquence. Cependant, nous n'avons pas pu identifier les facteurs responsables de cette activité. Un candidat potentiel était la protéine RecO, puisque celle-ci réalise de l'appariement simple brin *in vitro*, y compris sur de l'ADN couvert de SSB (Kantake et al., 2002). Or l'appariement simple brin n'est pas diminué dans un mutant *recA recO* (voir Résultats, chapitre II). Nous avons également éliminé l'hypothèse que ce soit une fonction portée par le plasmide qui permette l'appariement simple brin. Il n'est pas à exclure que l'appariement simple brin se produise spontanément à faible fréquence dans un mutant *recA*. Mais il est également possible que la protéine

responsable de cette activité d'appariement d'ADN simple brin résiduelle existe et soit inconnue pour le moment.

De manière intéressante, nous avons observé différentes conditions dans lesquelles cet appariement simple brin était inhibé : En présence de RecA, exprimé à partir de son gène chromosomique, cette activité cellulaire résiduelle est abolie ($2,8 \times 10^{-8}$ Rif^R par cellule viable), contrairement à l'expression de RecA à partir du plasmide pJA5 ($1,9 \times 10^{-7}$ Rif^R par cellule viable). Ceci suggère qu'une telle activité est évitée dans la cellule sauvage (WT). De même, la présence d'UvsX mutée dans son site de fixation de l'ATP, UvsX K66A, exprimée par le plasmide ($4,5 \times 10^{-8}$ Rif^R par cellule viable), inhibe l'appariement d'ADN simple brin cellulaire. Ces résultats suggèrent que RecA et UvsX K66A ont une activité dominante négative sur cette activité : elles peuvent se fixer à l'ADN simple brin et empêcher ainsi l'appariement d'ADN simple brin résiduel. Il serait très intéressant de tester par la suite l'effet d'une mutation du site de fixation de l'ATP dans RecA sur cet appariement simple brin. Il est possible qu'alors, en absence d'activité ATPasique de la protéine, même la copie plasmidique de *recA* ait un effet complètement inhibiteur.

II. RecA ne fait pas d'appariement d'ADN simple brin *in vivo*

Nous avons mis en évidence que la recombinaise bactérienne RecA n'est pas capable d'apparier de l'ADN simple brin *in vivo*. C'est un résultat inattendu, car plusieurs études réalisées *in vitro* avaient montré que RecA était capable d'apparier, de manière ATP-dépendante, deux ADN simple brin de séquences complémentaires (Mazin and Kowalczykowski, 1996; Weinstock et al., 1979). Cependant, à notre connaissance, cette activité de RecA n'a jamais été confirmée *in vivo*. Notre observation pourrait éclairer différentes observations qui ont conduit à proposer que certains intermédiaires de recombinaison homologue sont toxiques dans la cellule (Veaute et al., 2005), et que l'hélicase UvrD était nécessaire pour retirer RecA de ces

intermédiaires. A ce jour, ces intermédiaires toxiques supposés n'ont pas été vraiment caractérisés. Il pourrait s'agir d'une fourche de réplication couverte de RecA sur ses deux brins matrice. Si RecA n'est pas capable d'apparier ces brins, la fourche se retrouve alors effectivement bloquée.

III. UvsX fait de l'appariement simple brin *in vivo*

Etant donné la grande similarité de structure et de fonction entre UvsX et RecA (voir Introduction), il n'était pas attendu qu'elles se comportent différemment dans le test d'appariement simple brin *in vivo*. Cependant, à la différence de RecA, UvsX fait de l'appariement d'ADN simple brin *in vivo*, faiblement ($1,9 \times 10^{-6}$ Rif^R par cellule viable), mais significativement. Nous avons montré de plus que le site de fixation de l'ATP, ainsi que le domaine C-ter d'UvsX sont essentiels à cet appariement.

IV. L'activité d'appariement d'ADN simple brin *in vivo* de Sak4 est indépendante de son site de fixation de l'ATP

Les protéines de type Sak4 forment un sous-groupe dans la superfamille des protéines de type Rad51/RecA, caractérisé par leur petite taille et l'absence de domaine C-ter. Etant donné l'importance de ce domaine pour l'appariement d'ADN simple brin d'UvsX, il n'était pas sûr que les Sak4 aient une telle activité. Nos résultats sont basés sur l'étude de deux Sak4 différentes : celle du bactériophage PA73 qui a été étudiée en l'absence de cofacteurs ($1,7 \times 10^{-6}$ Rif^R par cellule viable), et celle du bactériophage HK620, étudiée en présence de ceux-ci ($1,5 \times 10^{-5}$ Rif^R par cellule viable). Ces deux Sak4 montrent une activité égale ou supérieure à celle d'UvsX ($1,9 \times 10^{-6}$ Rif^R par cellule viable). De plus, l'activité d'appariement d'ADN simple brin *in vivo* de Sak4 est maintenue dans un mutant dont le site de fixation de l'ATP est inactivé. Ceci suggère que le mécanisme d'appariement d'ADN simple brin

diffère entre UvsX et Sak4, et cette question sera abordée plus précisément dans le futur par une approche biochimique.

V. Les recombinares de bactériophage font-elles toutes de l'appariement d'ADN simple brin lorsque les séquences divergent ?

Le centre de mon projet de thèse était de tester la capacité des recombinares phagiques à appairer des séquences d'ADN divergées. Red β étant capable de recombiner de l'ADN divergé (Martinson et al., 2008), nous voulions comprendre si cette propriété était spécifique à Red β ou si elle était généralisable aux autres recombinares phagiques, en particulier celles appartenant aux autres superfamilles.

Pour étudier cet aspect, un ajustement supplémentaire du test d'appariement simple brin a été nécessaire. En effet, le test d'appariement entre séquences divergées est réalisé dans une souche mutée pour *mutS*, afin d'empêcher que le système de réparation des mésappariements élimine les hétéroduplex. Or, dans ces conditions le taux de mutants spontanés résistants à la rifampicine est très élevé (en moyenne $5,7 \times 10^{-6}$ Rif^R par cellule viable). Ceci perturbe l'interprétation des résultats, car les efficacités d'appariement simple brin sont à peine plus élevées que ce bruit de fond. Nous avons donc recherché d'une part à augmenter l'efficacité d'appariement d'ADN simple brin, et d'autre part à diminuer le bruit de fond.

Pour augmenter l'efficacité d'appariement d'ADN simple brin, les recombinares ont été testées en présence de leurs cofacteurs (voir Résultats). Ceci nous a amené à concentrer nos travaux sur deux recombinares : Red β co-exprimée avec Red α , Gam et Orf60a, et Sak4 de HK620 en présence de son cofacteur Abc2 et de deux autres potentiels cofacteurs codés par les gènes *hkaK* et *hkaM*.

Pour diminuer le taux de mutants spontanés Rif^R, nous avons réalisé l'expérience d'appariement d'ADN simple brin dans une souche ayant un système MMR actif, en

présence d'un analogue de l'adénine, le 2-aminopurine (2AP). En effet, le 2AP s'incorpore et se mésapparie aux cytosines pendant la réplication. Pendant le traitement avec du 2AP la souche est déficiente pour le système MMR de manière transitoire (Matic et al., 2003; Pitsikas et al., 2004). Ainsi, nous espérons diminuer le taux de mutant rifampicine spontané. Cependant, nous avons obtenu un taux équivalent à celui obtenu dans une souche *mutS*. Dans ces conditions, nous avons néanmoins pu détecter de l'appariement d'ADN simple brin entre séquences divergées à 14% pour Red β , confirmant les observations rapportés dans la souche *mutS*.

Dans un futur proche, un nouveau système expérimental d'appariement d'ADN simple brin sera testé : il est basé sur l'insertion dans le chromosome d'*E. coli* d'un gène *cat*, conférant la résistance au chloramphenicol et ayant deux mutations ambre. Ensuite, le principe du test est le même que précédemment : les bactéries sont transformées avec un oligonucléotide dont la recombinaison avec le chromosome restaure la séquence sauvage du gène *cat*, rendant les bactéries résistantes au chloramphénicol. Comme le taux de mutants spontanés résistants au chloramphenicol est inférieur à 10^{-8} dans un contexte *mutS*, nous aurons une bien plus grande sensibilité de détection des appariements simple brin.

Par ailleurs, une autre méthode pour améliorer les fréquences de recombinants obtenues serait de tester toutes les recombinases en présence de Gam (voir plus bas).

Dans les conditions actuelles du test d'appariement d'ADN simple brin (test avec un oligonucléotide complémentaire au gène *rpoB*, en contexte *recA mutS*), nous avons déjà pu mettre en évidence que les recombinases Red β et Sak4 appariant des séquences d'ADN simple brin ayant jusqu'à 14% de divergence. Nous pouvons donc conclure que, comme Red β , la recombinase Sak4 du bactériophage HK620 est capable de recombiner de l'ADN de séquence divergée, alors qu'elle appartient à une famille entièrement nouvelle de recombinases de bactériophages tempérés. Une

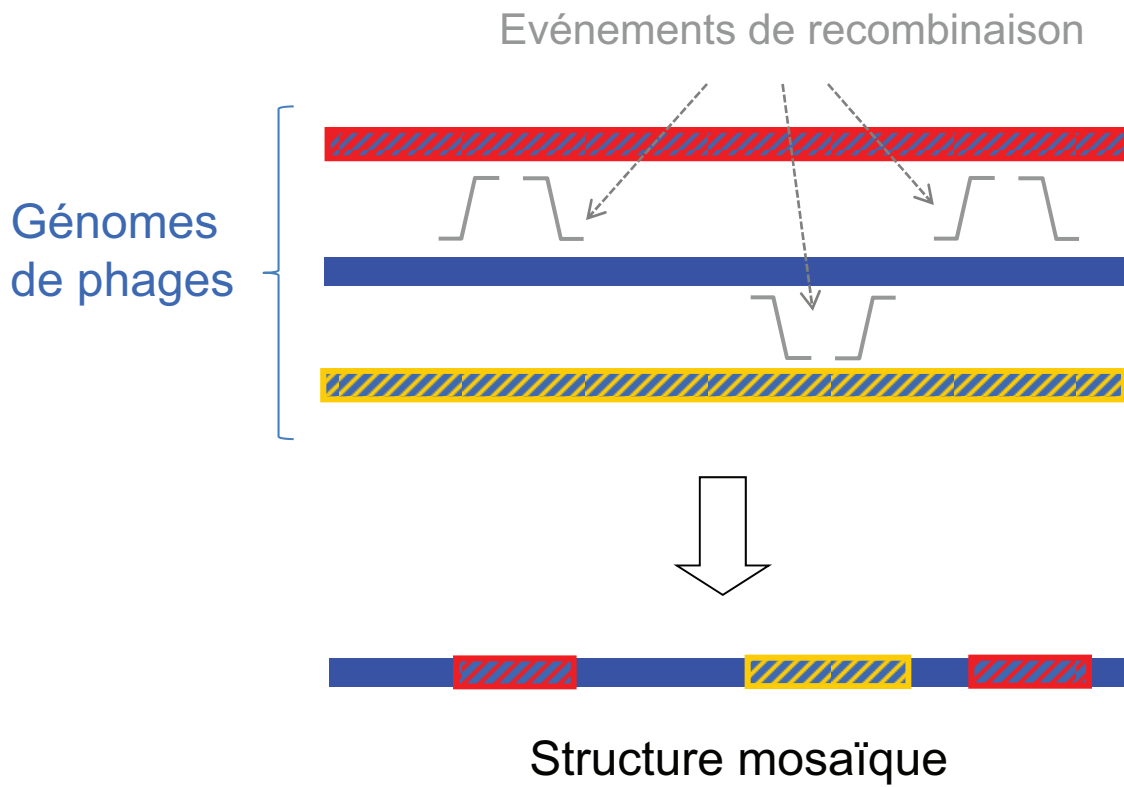


Figure 28: Structure mosaïque des génomes de bactériophages.

Le génome bleu, lorsqu'il est comparé aux génomes rouge et jaune, révèle trois régions de forte similarité avec eux. Ces séquences ont pu être acquises par recombinaison, lors de la rencontre de ces bactériophages au sein d'une même cellule.

telle propriété est peut-être une des voies qui favorise l'évolution rapide des génomes de bactériophages tempérés. En effet, les recombinaisons de type Red β ou Sak4 sont le plus souvent présentes dans des génomes mosaïques de bactériophages tempérés (Figure 28). On désigne par génomes mosaïques deux ou plusieurs génomes qui, comparés entre eux, montrent un ou plusieurs petits blocs (200 pb à quelques kb en général) de très forte identité de séquence (plus de 90%), comme si ces séquences d'ADN avaient été récemment échangées entre des bactériophages infectant le même hôte bactérien (Martinsohn et al., 2008).

Notre hypothèse de travail est que l'apparition de ces mosaïques dans les génomes de bactériophages est liée à la présence d'une recombinaison capable d'apparier des séquences divergées. Quelques génomes de bactériophages virulents de type T4 (codant pour la recombinaison UvsX) ont été séquencés. Jusqu'à présent, les analyses de génomique comparée montrent peu ou pas de traces de mosaïques dans ces génomes ((Comeau et al., 2007) et résultats non publiés de notre laboratoire).

Si notre hypothèse se confirme, nous attendons que les recombinaisons présentes sur des génomes de bactériophages non mosaïques soient davantage fidèles, et ne soient pas capables de recombiner de l'ADN divergent. Dès que le test d'appariement d'ADN simple brin aura acquis une plus grande sensibilité, la recombinaison UvsX sera testée pour son activité d'appariement d'ADN simple brin entre séquences divergées. Puisque nous avons mis en évidence que certains cofacteurs amélioreraient la fréquence d'appariement d'ADN simple brin, UvsX sera également testée en présence de son cofacteur UvsY (voir Introduction).

Une hypothèse alternative pouvant expliquer le degré élevé de mosaïcisme des génomes des bactériophages tempérés est que ces phages ont plus d'opportunité de se retrouver à plusieurs dans une même cellule que les phages lytiques, ce qui augmente la probabilité d'échange de matériel génétique, quel que soit le mécanisme. En ce cas, il n'est pas attendu que les recombinaisons de phages

tempérés se distinguent de celles des phages virulents. Enfin, il n'est pas exclus que la répartition d'UvsX chez les bactériophages virulents de type T4 et de Red β chez les bactériophages tempérés soit simplement due à leur origine phylogénétique (les T4 sont apparentés, donc leurs recombinaisons sont apparentées, et de même pour les lambdoïdes). En conclusion, les analyses de génomique comparée ne font que suggérer des hypothèses de travail, que des expériences moléculaires futures pourront valider ou invalider.

VI. Gam favorise l'appariement simple brin *in vivo*

Nos travaux ont montré que la protéine Gam du bactériophage Lambda permet d'améliorer l'efficacité d'appariement de l'ADN simple brin entre séquences identiques *in vivo* par Red β . En effet, lorsque Red β et Gam sont exprimés ensemble, on observe une augmentation d'un facteur 20 de l'appariement d'ADN simple brin, comparé à Red β seul. Nous avons démontré que Gam est un cofacteur essentiel de Red β pour le mécanisme d'appariement d'ADN simple brin *in vivo*. De manière intéressante, une étude d'Ellis et al, en 2001, basée sur l'appariement d'un oligonucléotide de 70 nt complémentaire au gène *galK* rendant la souche prototrophe pour le galactose avait obtenu un résultat similaire : en présence de Red β seul, les auteurs observaient une fréquence de 8×10^{-4} recombinants par cellule viable. En présence de Red β et Gam, les auteurs obtenaient une fréquence de $6,3 \times 10^{-3}$ recombinants par cellule viable, soit une augmentation d'un facteur 8 (Ellis et al., 2001). Cependant, ces auteurs décrivent l'effet de Gam comme faible et rapportent que Red β seul est nécessaire à l'appariement d'ADN simple brin. Cette observation, rapprochée de nos résultats, tend à suggérer que Gam possède plus qu'une seule fonction inhibitrice : il est connu depuis longtemps que Gam inhibe RecBCD et empêche ainsi la dégradation d'ADN (Karu et al., 1975; Sakaki et al., 1973). Gam inhibe également la nucléase SbcCD qui dégrade les structures d'ADN en épingle à

cheveux, cependant cette activité n'est pas bien caractérisée *in vitro* (Kulkarni and Stahl, 1989). Nous proposons que Gam a en plus un effet d'inhibition de la dégradation de l'ADN simple brin. Cette inhibition pourrait être directement liée à l'interaction de Gam sur RecBCD, car il a été montré que RecBCD peut dégrader l'ADN simple brin (Wang et al., 2000). Dans ce cas, un mutant *recBC* devrait avoir le même effet que la présence de Gam. Alternativement, l'inhibition par Gam pourrait concerner une ou plusieurs des nucléases spécifiques du simple brin : ExoI, ExoVII, ExoX ou RecJ. En ce cas, la mutation d'une ou plusieurs de ces nucléases conduirait au même phénotype que la présence de Gam.

Une étude de Sawitzke et al. suggère que RecJ dégrade l'ADN simple brin lors de l'appariement d'ADN simple brin : dans un mutant *recJ* la fréquence de recombinants Gal⁺ produits par Red β augmente, ainsi que dans un mutant *recQ recJ*. Cependant, cette fréquence n'est pas augmentée dans un mutant *recJ recA* (Sawitzke et al., 2011). Ces expériences sont réalisées en contexte Gam⁺, il faut donc supposer de plus que Gam serait partiellement titrée dans ce système expérimental (on s'attendrait alors à ce que l'inactivation simultanée de Gam et RecJ conduise à une augmentation encore plus forte de la fréquence de recombinaison).

Afin d'approfondir un dernier aspect de cette activité de Gam, il sera intéressant de tester si Gam favorise également l'appariement d'ADN simple brin médié par les autres recombinaisons phagiques, ce qui permettra de supposer que Gam n'a pas besoin d'une interaction spécifique avec Red β pour favoriser l'appariement d'ADN simple brin, et qu'elle agit seule en inhibant l'activité nucléasique de la cellule hôte.

Conclusion générale

En conclusion, nous avons mis en évidence que cinq recombinaises phagiques (UvsX, deux Sak4, deux Gp2.5), en plus de Red β , font de l'appariement d'ADN simple brin *in vivo*. De plus, nous avons montré que la recombinaise bactérienne RecA n'est pas capable d'apparier de l'ADN simple brin *in vivo*. Ceci nous permet de faire l'hypothèse que l'appariement d'ADN simple brin est une propriété générale des recombinaises phagiques qui les distinguent des recombinaises bactériennes, et apportent quelque chose en plus au mode de recombinaison des bactériophages.

Nos travaux ne permettent pas à ce jour une conclusion aussi forte en ce qui concerne la recombinaison entre séquences divergées : nous avons pu montrer que la recombinaise Sak4 de HK620 est capable de recombiner de l'ADN divergent jusque 13%, tout comme Red β . Si cette observation se confirme pour d'autres recombinaises de bactériophages tempérés, et pour des niveaux plus élevés de divergence, nous concluons à une autre marque distinctive des recombinaises phagiques. Si de plus, efficacité d'appariement d'ADN simple brin et facilité à apparier les séquences divergées sont corrélées, ceci suggérera que la moindre fidélité de la recombinaison homologue chez les bactériophages est étroitement associée à leur mécanisme de recombinaison par appariement d'ADN simple brin. Les travaux futurs permettront d'apprécier avec plus de finesse toutes les particularités de la recombinaison homologue chez les bactériophages, et leurs conséquences en termes d'évolution des génomes de bactériophages.

BIBLIOGRAPHIE

- Beernink, H. T., and Morrical, S. W. (1999). RMPs: recombination/replication mediator proteins. *Trends Biochem Sci* 24, 385-389.
- Bork, J. M., Cox, M. M., and Inman, R. B. (2001). The RecOR proteins modulate RecA protein function at 5' ends of single-stranded DNA. *Embo J* 20, 7313-7322.
- Botstein, D., and Matz, M. J. (1970). A recombination function essential to the growth of bacteriophage P22. *J Mol Biol* 54, 417-440.
- Bouchard, J. D., and Moineau, S. (2004). Lactococcal phage genes involved in sensitivity to AbiK and their relation to single-strand annealing proteins. *J Bacteriol* 186, 3649-3652.
- Brooks, K., and Clark, A. J. (1967). Behavior of lambda bacteriophage in a recombination deficient strain of *Escherichia coli*. *J Virol* 1, 283-293.
- Chen, Z., Yang, H., and Pavletich, N. P. (2008). Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453, 489-484.
- Clark, A. J., Inwood, W., Cloutier, T., and Dhillon, T. S. (2001). Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *J Mol Biol* 311, 657-679.
- Clark, A. J., and Margulies, A. D. (1965). Isolation and Characterization of Recombination-Deficient Mutants of *Escherichia coli* K12. *Proc Natl Acad Sci U S A* 53, 451-459.
- Conkling, M. A., and Drake, J. W. (1984). Thermal rescue of UV-irradiated bacteriophage T4 and biphasic mode of action of the WXY system. *Genetics* 107, 525-536.
- Costantino, N., and Court, D. L. (2003). Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* 100, 15748-15753.
- 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 U S A* 97, 6640-6645.
- Datta, S., Costantino, N., Zhou, X., and Court, D. L. (2008). Identification and analysis of recombineering functions from Gram-negative and Gram-positive bacteria and their phages. *Proc Natl Acad Sci U S A* 105, 1626-1631.
- Dillingham, M. S., and Kowalczykowski, S. C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* 72, 642-671, Table of Contents.
- Drake, J. W., Charlesworth, B., Charlesworth, D., and Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics* 148, 1667-1686.
- Duttreix, M., Moreau, P. L., Bailone, A., Galibert, F., Battista, J. R., Walker, G. C., and Devoret, R. (1989). New recA mutations that dissociate the various RecA protein activities in *Escherichia coli* provide evidence for an additional role for RecA protein in UV mutagenesis. *J Bacteriol* 171, 2415-2423.
- Echols, H., Gingery, R., and Moore, L. (1968). Integrative recombination function of bacteriophage lambda: evidence for a site-specific recombination enzyme. *J Mol Biol* 34, 251-260.
- Eggler, A. L., Lusetti, S. L., and Cox, M. M. (2003). The C terminus of the *Escherichia coli* RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. *J Biol Chem* 278, 16389-16396.
- Ellis, H. M., Yu, D., DiTizio, T., and Court, D. L. (2001). High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* 98, 6742-6746.
- Erler, A., Wegmann, S., Elie-Caille, C., Bradshaw, C. R., Maresca, M., Seidel, R., Habermann, B., Muller, D. J., and Stewart, A. F. (2009). Conformational adaptability of Redbeta during DNA annealing and implications for its structural relationship with Rad52. *J Mol Biol* 391, 586-598.
- Fenton, A. C., and Poteete, A. R. (1984). Genetic analysis of the erf region of the bacteriophage P22 chromosome. *Virology* 134, 148-160.
- Forterre, P., and Prangishvili, D. (2009). The origin of viruses. *Res Microbiol* 160, 466-472.
- Franklin, N. C. (1967). Deletions and functions of the center of the phi80 -lambda phage genome. Evidence for a phage function promoting genetic recombination. *Genetics* 57, 301-318.

- Gillen, J. R., Karu, A. E., Nagaishi, H., and Clark, A. J. (1977). Characterization of the deoxyribonuclease determined by lambda reverse as exonuclease VIII of *Escherichia coli*. *J Mol Biol* 113, 27-41.
- Gottesman, M. M., Gottesman, M. E., Gottesman, S., and Gellert, M. (1974). Characterization of bacteriophage lambda reverse as an *Escherichia coli* phage carrying a unique set of host-derived recombination functions. *J Mol Biol* 88, 471-487.
- Hall, S. D., Kane, M. F., and Kolodner, R. D. (1993). Identification and characterization of the *Escherichia coli* RecT protein, a protein encoded by the recE region that promotes renaturation of homologous single-stranded DNA. *J Bacteriol* 175, 277-287.
- Hall, S. D., and Kolodner, R. D. (1994). Homologous pairing and strand exchange promoted by the *Escherichia coli* RecT protein. *Proc Natl Acad Sci U S A* 91, 3205-3209.
- Howard-Flanders, P., and Boyce, R. P. (1966). DNA repair and genetic recombination: studies on mutants of *Escherichia coli* defective in these processes. *Radiat Res, Suppl* 6:156+.
- Iyer, L. M., Koonin, E. V., and Aravind, L. (2002). Classification and evolutionary history of the single-strand annealing proteins, RecT, Redbeta, ERF and RAD52. *BMC Genomics* 3, 8.
- Jensen, R. B., Carreira, A., and Kowalczykowski, S. C. (2010). Purified human BRCA2 stimulates RAD51-mediated recombination. *Nature* 467, 678-683.
- Kagawa, W., Kurumizaka, H., Ishitani, R., Fukai, S., Nureki, O., Shibata, T., and Yokoyama, S. (2002). Crystal structure of the homologous-pairing domain from the human Rad52 recombinase in the undecameric form. *Mol Cell* 10, 359-371.
- Kantake, N., Madiraju, M. V., Sugiyama, T., and Kowalczykowski, S. C. (2002). *Escherichia coli* RecO protein anneals ssDNA complexed with its cognate ssDNA-binding protein: A common step in genetic recombination. *Proc Natl Acad Sci U S A* 99, 15327-15332.
- Karu, A. E., Sakaki, Y., Echols, H., and Linn, S. (1975). The gamma protein specified by bacteriophage gamma. Structure and inhibitory activity for the recBC enzyme of *Escherichia coli*. *J Biol Chem* 250, 7377-7387.
- Kmiec, E., and Holloman, W. K. (1981). Beta protein of bacteriophage lambda promotes renaturation of DNA. *J Biol Chem* 256, 12636-12639.
- Kovall, R., and Matthews, B. W. (1997). Toroidal structure of lambda-exonuclease. *Science* 277, 1824-1827.
- Kulkarni, S. K., and Stahl, F. W. (1989). Interaction between the sbcC gene of *Escherichia coli* and the gam gene of phage lambda. *Genetics* 123, 249-253.
- Kurumizaka, H., Aihara, H., Ikawa, S., Kashima, T., Bazemore, L. R., Kawasaki, K., Sarai, A., Radding, C. M., and Shibata, T. (1996). A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding. *J Biol Chem* 271, 33515-33524.
- Kushner, S. R., Nagaishi, H., and Clark, A. J. (1972). Indirect suppression of recB and recC mutations by exonuclease I deficiency. *Proc Natl Acad Sci U S A* 69, 1366-1370.
- Kuzminov, A. (1999). Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage lambda. *Microbiol Mol Biol Rev* 63, 751-813, table of contents.
- Leplae, R., Hebrant, A., Wodak, S. J., and Toussaint, A. (2004). ACLAME: a CLAssification of Mobile genetic Elements. *Nucleic Acids Res* 32, D45-49.
- Leplae, R., Lima-Mendez, G., and Toussaint, A. (2010). ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* 38, D57-61.
- Liu, J., and Morrical, S. W. (2011). Assembly and dynamics of the bacteriophage T4 homologous recombination machinery. *Virology* 437, 357.
- Lopes, A., Amarir-Bouhram, J., Faure, G., Petit, M. A., and Guerois, R. (2010). Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. *Nucleic Acids Res* 38, 3952-3962.
- Luisi-DeLuca, C., Clark, A. J., and Kolodner, R. D. (1988). Analysis of the recE locus of *Escherichia coli* K-12 by use of polyclonal antibodies to exonuclease VIII. *J Bacteriol* 170, 5797-5805.
- Luisi-DeLuca, C., and Kolodner, R. (1994). Purification and characterization of the *Escherichia coli* RecO protein. Renaturation of complementary single-stranded DNA molecules catalyzed by the RecO protein. *J Mol Biol* 236, 124-138.

- Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y., and Stewart, A. F. (2010). Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol Biol* 11, 54.
- Martinsohn, J. T., Radman, M., and Petit, M. A. (2008). The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism. *PLoS Genet* 4, e1000065.
- Matic, I., Babic, A., and Radman, M. (2003). 2-aminopurine allows interspecies recombination by a reversible inactivation of the Escherichia coli mismatch repair system. *J Bacteriol* 185, 1459-1461.
- Matic, I., Radman, M., and Rayssiguier, C. (1994). Structure of recombinants from conjugational crosses between Escherichia coli donor and mismatch-repair deficient Salmonella typhimurium recipients. *Genetics* 136, 17-26.
- Mazin, A. V., and Kowalczykowski, S. C. (1996). The specificity of the secondary DNA binding site of RecA protein defines its role in DNA strand exchange. *Proc Natl Acad Sci U S A* 93, 10673-10678.
- Mazin, A. V., and Kowalczykowski, S. C. (1998). The function of the secondary DNA-binding site of RecA protein during DNA strand exchange. *Embo J* 17, 1161-1168.
- Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. *Annu Rev Genet* 25, 229-253.
- Morimatsu, K., and Kowalczykowski, S. C. (2003). RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. *Mol Cell* 11, 1337-1347.
- Mortensen, U. H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996). DNA strand annealing is promoted by the yeast Rad52 protein. *Proc Natl Acad Sci U S A* 93, 10729-10734.
- Mosig, G., Gewin, J., Luder, A., Colowick, N., and Vo, D. (2001). Two recombination-dependent DNA replication pathways of bacteriophage T4, and their roles in mutagenesis and horizontal gene transfer. *Proc Natl Acad Sci U S A* 98, 8306-8311.
- Muniyappa, K., and Radding, C. M. (1986). The homologous recombination system of phage lambda. Pairing activities of beta protein. *J Biol Chem* 261, 7472-7478.
- Murphy, K. C. (1998). Use of bacteriophage lambda recombination functions to promote gene replacement in Escherichia coli. *J Bacteriol* 180, 2063-2071.
- Muyrers, J. P., Zhang, Y., Buchholz, F., and Stewart, A. F. (2000). RecE/RecT and Redalpha/Redbeta initiate double-stranded break repair by specifically interacting with their respective partners. *Genes Dev* 14, 1971-1982.
- Noirot, P., Gupta, R. C., Radding, C. M., and Kolodner, R. D. (2003). Hallmarks of homology recognition by RecA-like recombinases are exhibited by the unrelated Escherichia coli RecT protein. *Embo J* 22, 324-334.
- Noirot, P., and Kolodner, R. D. (1998). DNA strand invasion promoted by Escherichia coli RecT protein. *J Biol Chem* 273, 12274-12280.
- Passy, S. I., Yu, X., Li, Z., Radding, C. M., and Egelman, E. H. (1999). Rings and filaments of beta protein from bacteriophage lambda suggest a superfamily of recombination proteins. *Proc Natl Acad Sci U S A* 96, 4279-4284.
- Persky, N. S., and Lovett, S. T. (2008). Mechanisms of recombination: lessons from E. coli. *Crit Rev Biochem Mol Biol* 43, 347-370.
- Petit, M. A., Dimpfl, J., Radman, M., and Echols, H. (1991). Control of large chromosomal duplications in Escherichia coli by the mismatch repair system. *Genetics* 129, 327-332.
- Pitsikas, P., Patapas, J. M., and Cupples, C. G. (2004). Mechanism of 2-aminopurine-stimulated mutagenesis in Escherichia coli. *Mutat Res* 550, 25-32.
- Ploquin, M., Bransi, A., Paquet, E. R., Stasiak, A. Z., Stasiak, A., Yu, X., Cieslinska, A. M., Egelman, E. H., Moineau, S., and Masson, J. Y. (2008). Functional and structural basis for a bacteriophage homolog of human RAD52. *Curr Biol* 18, 1142-1146.
- Poteete, A. R. (2008). Involvement of DNA replication in phage lambda Red-mediated homologous recombination. *Mol Microbiol* 68, 66-74.

- Poteete, A. R., and Fenton, A. C. (2000). Genetic requirements of phage lambda red-mediated gene replacement in *Escherichia coli* K-12. *J Bacteriol* **182**, 2336-2340.
- Poteete, A. R., Fenton, A. C., and Murphy, K. C. (1988). Modulation of *Escherichia coli* RecBCD activity by the bacteriophage lambda Gam and P22 Abc functions. *J Bacteriol* **170**, 2012-2021.
- Radman, M. (1975). SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sci* **5A**, 355-367.
- Rattray, A. J., and Symington, L. S. (1994). Use of a chromosomal inverted repeat to demonstrate that the RAD51 and RAD52 genes of *Saccharomyces cerevisiae* have different roles in mitotic recombination. *Genetics* **138**, 587-595.
- Rayssiguier, C., Thaler, D. S., and Radman, M. (1989). The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* **342**, 396-401.
- Roca, A. I., and Cox, M. M. (1997). RecA protein: structure, function, and role in recombinational DNA repair. *Prog Nucleic Acid Res Mol Biol* **56**, 129-223.
- Rocha, E. P., Cornet, E., and Michel, B. (2005). Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet* **1**, e15.
- Rybalchenko, N., Golub, E. I., Bi, B., and Radding, C. M. (2004). Strand invasion promoted by recombination protein beta of coliphage lambda. *Proc Natl Acad Sci U S A* **101**, 17056-17060.
- Sakaki, Y., Karu, A. E., Linn, S., and Echols, H. (1973). Purification and properties of the gamma-protein specified by bacteriophage lambda: an inhibitor of the host RecBC recombination enzyme. *Proc Natl Acad Sci U S A* **70**, 2215-2219.
- Sawitzke, J. A., Costantino, N., Li, X. T., Thomason, L. C., Bubunencko, M., Court, C., and Court, D. L. (2011). Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. *J Mol Biol* **407**, 45-59.
- Scaltriti, E., Launay, H., Genoio, M. M., Bron, P., Rivetti, C., Grolli, S., Ploquin, M., Campanacci, V., Tegoni, M., Cambillau, C., *et al.* (2011). Lactococcal phage p2 ORF35-Sak3 is an ATPase involved in DNA recombination and AbiK mechanism. *Mol Microbiol* **80**, 102-116.
- Scaltriti, E., Moineau, S., Launay, H., Masson, J. Y., Rivetti, C., Ramoni, R., Campanacci, V., Tegoni, M., and Cambillau, C. (2010). Deciphering the function of lactococcal phage ϕ 36 Sak domains. *J Struct Biol* **170**, 462-469.
- Shinohara, A., and Ogawa, T. (1998). Stimulation by Rad52 of yeast Rad51-mediated recombination. *Nature* **391**, 404-407.
- Signer, E. R., and Weil, J. (1968). Recombination in bacteriophage lambda. I. Mutants deficient in general recombination. *J Mol Biol* **34**, 261-271.
- Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C., and Wigley, D. B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**, 187-193.
- Singleton, M. R., Wentzell, L. M., Liu, Y., West, S. C., and Wigley, D. B. (2002). Structure of the single-strand annealing domain of human RAD52 protein. *Proc Natl Acad Sci U S A* **99**, 13492-13497.
- Sommer, S., Boudsocq, F., Devoret, R., and Bailone, A. (1998). Specific RecA amino acid changes affect RecA-UmuD'C interaction. *Mol Microbiol* **28**, 281-291.
- Stahl, M. M., Thomason, L., Poteete, A. R., Tarkowski, T., Kuzminov, A., and Stahl, F. W. (1997). Annealing vs. invasion in phage lambda recombination. *Genetics* **147**, 961-977.
- Stasiak, A. Z., Larquet, E., Stasiak, A., Muller, S., Engel, A., Van Dyck, E., West, S. C., and Egelman, E. H. (2000). The human Rad52 protein exists as a heptameric ring. *Curr Biol* **10**, 337-340.
- Subramanian, K., Rutvisuttinunt, W., Scott, W., and Myers, R. S. (2003). The enzymatic basis of processivity in lambda exonuclease. *Nucleic Acids Res* **31**, 1585-1596.
- Sweasy, J. B., Witkin, E. M., Sinha, N., and Roegner-Maniscalco, V. (1990). RecA protein of *Escherichia coli* has a third essential role in SOS mutator activity. *J Bacteriol* **172**, 3030-3036.
- Swingle, B., Markel, E., Costantino, N., Bubunencko, M. G., Cartinhour, S., and Court, D. L. (2010). Oligonucleotide recombination in Gram-negative bacteria. *Mol Microbiol* **75**, 138-148.

- Unger, R. C., and Clark, A. J. (1972). Interaction of the recombination pathways of bacteriophage lambda and its host *Escherichia coli* K12: effects on exonuclease V activity. *J Mol Biol* 70, 539-548.
- Van der Heijden, T., Modesti, M., Hage, S., Kanaar, R., Wyman, C., and Dekker, C. (2008). Homologous recombination in real time: DNA strand exchange by RecA. *Mol Cell* 30, 530-538.
- Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M. A. (2005). UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in *Escherichia coli*. *Embo J* 24, 180-189.
- Wang, J., Chen, R., and Julin, D. A. (2000). A single nuclease active site of the *Escherichia coli* RecBCD enzyme catalyzes single-stranded DNA degradation in both directions. *J Biol Chem* 275, 507-513.
- Weaver, S., and Levine, M. (1977). Recombinational circularization of *Salmonella* phage P22 DNA. *Virology* 76, 29-38.
- Weinstock, G. M., McEntee, K., and Lehman, I. R. (1979). ATP-dependent renaturation of DNA catalyzed by the recA protein of *Escherichia coli*. *Proc Natl Acad Sci U S A* 76, 126-130.
- West, S. C. (1997). Processing of recombination intermediates by the RuvABC proteins. *Annu Rev Genet* 31, 213-244.
- Wolff, E., Kim, M., Hu, K., Yang, H., and Miller, J. H. (2004). Polymerases leave fingerprints: analysis of the mutational spectrum in *Escherichia coli* rpoB to assess the role of polymerase IV in spontaneous mutation. *J Bacteriol* 186, 2900-2905.
- Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G., and Court, D. L. (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proc Natl Acad Sci U S A* 97, 5978-5983.
- Zhang, Y., Buchholz, F., Muirers, J. P., and Stewart, A. F. (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. *Nat Genet* 20, 123-128.

Annexe

Low Efficiency of Homology-Facilitated Illegitimate Recombination during Conjugation in *Escherichia coli*

Jihane Amarir-Bouhram^{1,2}, Mélodie Goin^{1,2}, Marie-Agnès Petit^{1,2*}

¹ INRA, UMR 1319, Micalis, Jouy en Josas, France, ² AgroParisTech, UMR 1319, Micalis, Jouy en Josas, France

Abstract

Homology-facilitated illegitimate recombination has been described in three naturally competent bacterial species. It permits integration of small linear DNA molecules into the chromosome by homologous recombination at one end of the linear DNA substrate, and illegitimate recombination at the other end. We report that homology-facilitated illegitimate recombination also occurs in *Escherichia coli* during conjugation with small non-replicative plasmids, but at a low frequency of 3×10^{-10} per recipient cell. The fate of linear DNA in *E. coli* is either RecBCD-dependent degradation, or circularisation by ligation, and integration into the chromosome by single crossing-over. We also report that the observed single crossing-overs are *recA*-dependent, but essentially *recBCD*, and *recFOR* independent. This suggests that other, still unknown, proteins may act as mediator for the loading of RecA on DNA during single crossing-over recombination in *E. coli*.

Citation: Amarir-Bouhram J, Goin M, Petit M-A (2011) Low Efficiency of Homology-Facilitated Illegitimate Recombination during Conjugation in *Escherichia coli*. PLoS ONE 6(12): e28876. doi:10.1371/journal.pone.0028876

Editor: David T. Kirkpatrick, University of Minnesota, United States of America

Received: August 8, 2011; **Accepted:** November 16, 2011; **Published:** December 15, 2011

Copyright: © 2011 Amarir-Bouhram et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Dr. Amarir-Bouhram PhD grant is funded by the French Ministry of Research. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: marie-agnes.petit@jouy.inra.fr

Introduction

Homology-facilitated illegitimate recombination (HFIR) is a hybrid recombination reaction whereby a linear DNA molecule integrates into the bacterial chromosome during natural transformation. Up to now, it has been reported in the three naturally competent bacterial species *Streptococcus pneumoniae* [1], *Acinetobacter baylyi* [2] and *Pseudomonas stutzeri* [3]. In HFIR, one end only of the linear incoming DNA molecule shares homology with the resident chromosome. Homologous recombination at this end is associated with an illegitimate recombination event within the non-homologous region of the linear fragment (see Figure 1A). In the recombination product, a deletion of a size similar to the inserted foreign DNA is generally observed (on average 1 kb insertions and 1 kb deletions). In most cases, a 3 to 10 bp micro-homology is found at the illegitimate recombination junction. Compared to the efficiency of homologous recombination, HFIR is 100 times less frequent in *S. pneumoniae* [1], 10^5 times less frequent in *A. baylyi* [2], and 10^6 times less frequent in *P. stutzeri*. HFIR is always more frequent than strictly illegitimate recombination, and in all cases the reaction is RecA-dependent. Moreover, in *A. baylyi*, HFIR efficiency is 20 fold increased in a *recJ* background [4]. RecJ is a 5' to 3' single-strand-specific DNA exonuclease [5,6], which might degrade the 5' single strand extremity required for the illegitimate recombination event in *A. baylyi*.

We recently performed a systematic analysis of the variable DNA segments found upon multiple alignments of *Escherichia coli* genomes, which revealed that a majority of them are small in size (less than 500 bp), and correspond to insertions coupled with deletions [7], as presented in Figure 1C. This situation could formally originate from HFIR events (Figure 1A and B). However, such recombination events have never been reported in *E. coli*, which is not naturally

competent. We reasoned that other means of horizontal transfer might provide substrates similar to those used in HFIR in naturally competent cells. For instance, during DNA conjugation, single-stranded DNA is transferred, converted into double-stranded DNA, and may remain linear (and therefore a substrate of HFIR), if the DNA transfer is interrupted before completion.

We report the development of a genetic screen to detect HFIR events in *E. coli*, following DNA conjugation with small, non-replicative plasmids. We show that HFIR occurs once in every 3×10^{10} cells involved in conjugation, a frequency a million times lower than homologous recombination, and which might be too low to impact on genome evolution of *E. coli*. We also found evidence that linear DNA has two principal fates in *E. coli*: it is either degraded by RecBCD, or ligated and integrated into the chromosome by single crossing-over (SCO). This led us to enquire into the *E. coli* genetic requirements SCO between a non-replicative circular molecule and the bacterial chromosome. We show that SCO is clearly RecA-dependent, but only slightly dependent on the RecBCD and the RecF pathways. We also observed that SCO is inhibited by RuvABC, and is RecG-independent. Finally, SCO events were inhibited by a factor of 20 by the UvrD helicase. We conclude that the factors mediating the loading of RecA during SCO events, if they exist in *E. coli*, are not RecFOR nor RecBCD, and are unknown at present. Homologous DNA recombination, which has been studied in depth for years in *E. coli*, still has some aspects that remain to be elucidated.

Materials and Methods

Strains

All strains are listed Table 1. Strain β 2163 [8] is a MG1655 derivative hosting RP4-2-Tc::Mu in the *glbB* gene of its

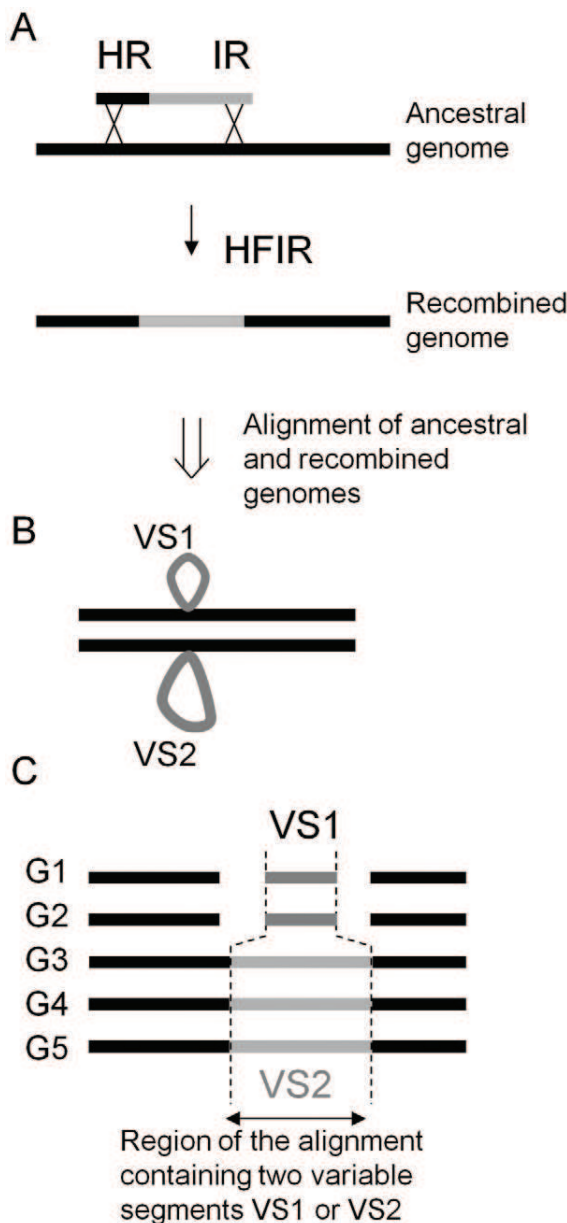


Figure 1. HFIR may account for the occurrence of insertions coupled with deletions, observed upon multiple genomes alignments. **A.** During Homology-Assisted Illegitimate Recombination (HFIR), a linear fragment sharing homology with the chromosome (black part) enters into the recipient chromosome by an atypical double crossing-over event, with homologous recombination (HR) at the left end of the molecule, and illegitimate recombination (IR) at the other end. **B.** Alignment of the resulting recombinant with the ancestral genome will produce two small “variable regions” flanked by the backbone sequences, common to both strains. **C.** Diversification of strains produced by HFIR, leads to the detection of insertions (variable segment VS2 in genomes G3–G5) coupled with deletions (variable segment VS1 in genomes G1–G2) upon multiple genome alignments of various *E. coli* strains (here 5 strains). doi:10.1371/journal.pone.0028876.g001

chromosome [9]. RP4-2-Tc::Mu contains the transfer functions of the conjugative plasmid RP4, together with a kanamycin resistance gene, and a Mu-prophage inserted in its tetracycline resistance gene. In addition, this strain contains the *pir* gene and

erythromycin resistance gene integrated at the *dap* locus, so that it requires 0.3 mM diaminopimelate for growth. An *EcoK* restriction and modification mutant derivative of strain β 2163 was constructed (strain MAC1306), in which the two genes *hsdR* *hsdM* were replaced by a phleomycin resistance (*phleo^R*) gene. Like strain β 2163, MAC1306 contains the Mu prophage [9]. We verified that Mu did not interfere with the recombination assay (not shown). The single mutant Δ *hsdR*::FRT derivative of β 2163 was also constructed (strain MAC1308), to use as a donor strain in conjugation experiments when the plasmid had to be protected from *EcoK* restriction in the recipient strain.

The MAC1306 or MAC1308 strains containing the various plasmids described below were used as a donor strain in the conjugation experiments. Recipient strains were either MG1655 (wild type strain) or MG1655 mutant derivatives listed in Table 1, containing plasmid pISceI. Mutant phenotypes were verified by UV sensitivity tests (for *recA*, *recB*, *recF*, *recB* *recF*, *recO*, *recR*, *recG*, *recJ* and *ruvA* mutants), T4gp2 sensitivity (*recD* allele), mutator phenotype (*uvrD*), and sensitivity to *EcoK* restriction of phage Lambda grown on non-modifying strains (*hsdR* and *hsdM* phenotypes).

Plasmid constructions

To place DNA in a situation in which HFIR may occur, a system allowing plasmid delivery with high efficiency, and maintaining this plasmid linear in the recipient cell for as long as possible was designed. To reach high efficiency of plasmid entry, we set up a genetic system based on conjugation in *E. coli*, taking advantage of the mobilizable suicide plasmid pSW23T [8]. This plasmid carries the *oriT* transfer origin of RP4, and relies for conjugation on the RP4 conjugation genes provided in *trans* (RP4-2-Tc::Mu is chromosomally integrated in the donor strain). It also contains the vegetative replication origin of plasmid R6K, which relies on the *Pir* replication protein provided in *trans* in the donor strain. This system is conceived such that, once mobilized and introduced in a recipient strain, pSW23T no longer replicates and can only survive by recombining with the recipient chromosome.

Plasmid pSW23T was engineered further to maintain it in a linear form in the recipient strain, and thus facilitate HFIR detection. For this, a “cut cassette” composed of two *EcoK* restriction sites flanking an *I-sceI* restriction site was cloned into pSW23T, giving pJA1 (see Figure 2). *EcoK* is a type I restriction enzyme which loads DNA on the AAC(N6)GTGC sequence, then tracks along the DNA from its loading site, by pulling it from both sides [10]. Cutting occurs at random positions, upon collision with a second *EcoK*/DNA complex. Therefore, a minimum of two *EcoK* restriction sites are needed on a plasmid to provoke cutting. An internal segment of the *yegN* gene of *Bacillus subtilis*, which contains two *EcoK* sites 450 bp apart was PCR amplified, using primers j1 and j2 (Table 2), containing at their 5′ extremities a Chi site oriented such that it protects pJA2 from RecBCD degradation, once linearised in the cut cassette. A linker composed of the *I-sceI* cutting site (TAGGGA-TAACAGGGTAAT) and compatible ends (oligonucleotides j7 and j8, Table 2) was then ligated into the *yegN* PCR fragment, between the *Clal* and *BclI* sites. Finally, this synthetic cassette was cloned at the *SmaI* site of pSW23T, giving plasmid pJA1. The presence of all critical elements of this cassette was verified by DNA sequencing. This cassette allows the production, in the recipient strain, of a linear plasmid flanked by two Chi sites, properly oriented to protect linear DNA from RecBCD degradation.

In a further step, to provide the conjugative plasmid with homology to the recipient chromosome, a 1 kb segment corresponding to the 3′ terminal part of the *lacZ* ORF was amplified with primers j3 and j4, and cloned into pJA1 between the *Bam*HI and *Sac*II sites, giving pJA2 (Figure 2). Linearization of pJA2 at the cut

Table 1. Strains and plasmids used in this study.

Strains		
Name	Relevant genotype (all strains are MG1655 derivatives)	Source, construction
β2163	RP4-2-Tc::Mu Δ dapA:(Erm-pir)	[8]
MAC1306	RP4-2-Tc::Mu Δ dapA:(Erm-pir) Δ hsdR-hsdM::phleo ^R	This work
MAC 1308	RP4-2-Tc::Mu Δ dapA:(Erm-pir) Δ hsdR::FRT	This work
JAC7	plscel (Amp ^R)	This work
JAC21	recA306 srl::Tn10	This work
MAC1348	recA306 srl::Tn10 plscel (Amp ^R)	This work
MAC1394	recB268::Tn10	This work
MAC1397	recB268::Tn10 plscel (Amp ^R)	This work
MAC1470	recF400::Tn5	This work
MAC1473	recF400::Tn5 plscel (Amp ^R)	This work
MAC1497	ruvA60::Tn10	This work
MAC1357	ruvA60::Tn10 plscel (Amp ^R)	This work
MAC1476	recG::Tet6200	This work
MAC1479	recG::Tet6200 plscel (Amp ^R)	This work
MAC1354	recD::Tn10	This work
MAC1352	recD::Tn10 plscel (Amp ^R)	This work
MAC1500	recJ::PhleoR	This work
MAC1350	recJ::PhleoR plscel (Amp ^R)	This work
MAC1503	recO::Tn5	This work
MAC1507	recR::Tn5	This work
MAC1662	uvrD::phleoR	This work
MAC1511	recF400::Tn5 recB268::Tn10	This work
Plasmids		
pSW23T	oriV R6K γ , oriT RP4, Cm ^R	[8]
plscel	pUC, <i>cl</i> 857, pL: <i>l-scel</i>	A. Lindner and M. Elez (Inserm U1001, Paris)
pJA1	pSW23T, cut cassette, Cm ^R	This work
pJA2	pJA1, <i>lacZ end</i>	This work
pJA3	pJA2, <i>lacZ beg</i>	This work

doi:10.1371/journal.pone.0028876.t001

cassette produces a molecule with 1 kb of homology to the *lacZ* chromosomal gene at one end (except for a few hundreds of terminal bp corresponding to the distance between the cutting site and the region of *lacZ* homology), whereas the other end has no homology with the recipient chromosome.

As a control for the efficiency of linearization, a pJA3 plasmid was derived from pJA2, containing the 5' end of the *lacZ* ORF. This segment of *lacZ* was amplified with primers j5 and j6, and cloned between the *Pst*I and *Sal*I sites of pJA2. Linearization of pJA3 produces a molecule with 1 kb *lacZ* sequences at both ends, which are correctly oriented to integrate the plasmid by double crossing-over recombination in the *lacZ* gene in the *E. coli* chromosome, and produce a Lac⁻Cm^R ex-conjugant (Figure 2). If on the contrary pJA3 integrates into the chromosome as a circular form, by a single-crossing over recombination in either of the two *lacZ* segments present on pJA3, an intact *lacZ* ORF is restored, and a Lac⁺Cm^R phenotype is expected.

Conjugations

Conjugations were performed between the MAC1306 or MAC1308 donor strains containing plasmid pJA1, pJA2 or pJA3, and various derivatives of the wild type strain MG1655 as recipient.

The *Eco*K restriction system is active in MG1655. Plasmid plscel is a pUC derivative encoding the *cl* 857 thermosensitive repressor of bacteriophage Lambda, and the gene encoding the *Isc*I endonuclease cloned downstream of the Lambda pL promoter. Transcription of the endonuclease is induced by a shift from 37°C to 41°C for 1h30 prior to conjugation. Conjugations were done on filters at 41°C, unless otherwise stated, for two hours, starting from exponentially growing cells of the donor and recipient strains (OD₆₀₀ of 0.3), and using a ratio of 2.5 recipient per donor cell. After conjugation, filters were taken, cells resuspended by vortexing in LB, and various dilutions of the mixtures were then plated at 37°C on LB rich medium devoid of diamminopimelate, to counterselect the donor cells, and to count recipient cells. Cm^R recombinants, both Lac⁺ and Lac⁻, were detected on LB plates supplemented with 20 µg/ml of chloramphenicol, 0.5 mM isopropyl-b-D-1-thiogalactopyranoside (IPTG) and 40 µg/ml of 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal).

Results

Experimental set-up to test HFIR

To detect HFIR during conjugation in *E. coli*, we used plasmid pJA2, and its two controls pJA1 and pJA3, presented in Figure 2.

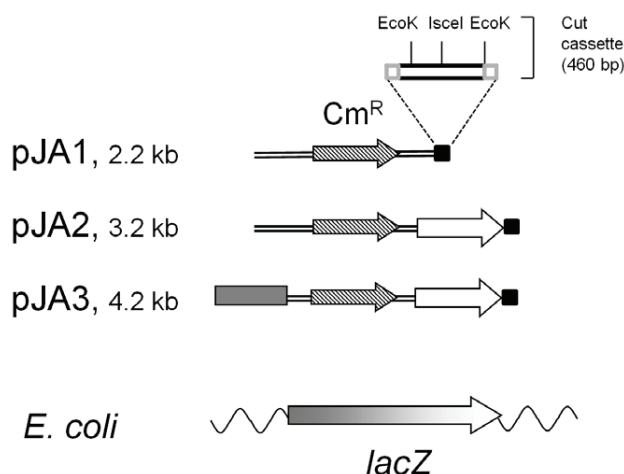


Figure 2. Maps of plasmids pJA1, pJA2 and pJA3 after linearization in the cut cassette. The 1 kb 5' part of *lacZ* is shown with a grey box, and the 1 kb 3' part of *lacZ* (*lacZ*end) by a white arrow. The *cat* gene encoding resistance to chloramphenicol (Cm^R) is shown as a stripped arrow. The cut cassette is indicated by a small black box, linearization of pJA plasmids *in vivo* might happen either at the *IseI* recognition sequence, or between the two *EcoK* sites. The cut cassette map is shown in details, with two grey boxes indicating the Chi sites. Below the three plasmids, the *E. coli* chromosome with which plasmids might recombine is shown (wavy line) in the *lacZ* region (grey to white arrow).

doi:10.1371/journal.pone.0028876.g002

All plasmids encode a gene conferring resistance to chloramphenicol (Cm^R), carry the *oriT* transfer origin of RP4, and the vegetative replication origin of plasmid R6K. Once mobilized and introduced in a recipient strain, such plasmids no longer replicate (nor conjugate), and can only survive by recombining with the recipient chromosome. To maintain the plasmid in a linear form in the recipient strain, the three plasmids contain a 'cut cassette', composed of sites cut by the two endonucleases *EcoK* and *IseI*, framed by two Chi sites. Plasmid pJA2 has 1 kb of homology at the 3' end of *lacZ*, so that the linear pJA2 has homology to the chromosome at one end, and no homology at the other. With such a set up, HFIR is expected to produce $\text{Cm}^R \text{Lac}^-$ exconjugants as

represented Figure 3 (central part), which can easily be distinguished from the $\text{Cm}^R \text{Lac}^+$ exconjugants formed by single crossing-over due to possible residual circular molecules (Figure 3, left part). Plasmid pJA3 is a control plasmid derived from pJA2 (see Figure 2), which permits measuring the efficiency of linearization. It contains a second 1 kb region of homology with *lacZ* (the 5' end of the *lacZ* ORF). Finally, pJA1 has no homology to the chromosome.

Estimation of the efficiency of linearization of the incoming plasmids after conjugation

All conjugative plasmids used in this study enter as single strand, linear DNA in the recipient cell. However, the plasmid recircularizes once entry is completed, and converts into double strand (ds) DNA. This conversion to ds DNA is probably rapid, as recombination of pJA2 into the chromosome is sensitive to *EcoK* and *IseI* ds endonucleases (see below and Figure 3). In order to maintain the incoming ds DNA molecules linear, we made use of the resident restriction enzyme *EcoK*, and added an additional *IseI* endonuclease provided by plasmid pIsceI. The efficiency of cutting by each system was estimated using pJA3. If all pJA3 molecules are kept linear, 100% of Lac^- recombinant clones are expected, because all Cm^R exconjugants will result from double crossing-over events (DCO) involving the two *lacZ* fragments present on the plasmid. If some of the molecules remain circular on the contrary, they will generate recombinants by single crossing-overs (SCO), which will reconstitute the full length *lacZ* ORF, and be Lac^+ . Finally, when no restriction is provided (*i.e.*, the donor strain is Modification⁺ for *EcoK*, and the recipient does not have plasmid pIsceI), the proportion of Lac^- clones will indicate how many DCO are generated starting from a circular substrate.

Conjugations with MAC1306 (*EcoK* Modification⁻) or MAC1308 (*EcoK* Modification⁺) donor cells containing pJA3 and recipient MG1655 wild-type cells, expressing or not expressing the *IseI* site specific nuclease, were carried out on filters for 2 hours. Cells were then resuspended and appropriate dilutions were plated either on LB medium lacking DAP, to counterselect donor cells and count recipient cells only, or on the same plates supplemented with Xgal, IPTG and chloramphenicol, to count recombinant recipients and distinguish Lac^+ from Lac^- . Results are reported in Figure 4, left panel. We found that the background proportion of Lac^- clones obtained with pJA3 when no restriction is provided was 61%. Providing *EcoK* cutting or *IseI* cutting increased the level of Lac^- clones to 86% and 91%, respectively. Finally, when both endonucleases were active, Lac^- clones were 98.7% of all exconjugants. We conclude that both cutting systems are efficient, even though 1.3% of exconjugants remain Lac^+ when both systems are acting together, suggesting that a maximum of 1.3% of the plasmid molecules have remained circular (we suppose here that circular and linear molecules recombine with similar efficiencies).

Recombination with linear pJA2 is decreased compared to linear pJA3

Next, conjugations with MAC1306 or MAC1308 donor cells containing pJA2 and recipient MG1655 wild-type cells, expressing or not the *IseI* site-specific endonuclease, were carried out. Results are reported in Figure 4, middle panel. Efficiency of recombination with pJA2 decreased 10 fold when one endonuclease was active on the substrate, and 50 fold when both were active. The decrease in recombination yield for the linear molecules, compared to pJA3, is most likely due to the lack of homology at one end of the linear DNA. Despite the use of two

Table 2. Sequence of the oligonucleotides used in this study.

Name	Sequence
j1	AACTGCAGCTGGTGGAAACACAGTGGTTCTGAAACC
j2	AATACCCGGGCTGGTGGCAGGTCCATATAAACATCT
j3	ATACCTCCGGGGCAACTCTGGCTCACAGTAC
j4	ATAGGATCCTATTTTGGACACAGACCA
j5	TTACTGCAGATGACCATGATTACGGATTC
j6	TTAGTCGACAAACGACATCGCAGGCTTC
j7	CGATAGGGATAACAGGGTAAT
j8	GATCATTACCTGTTATCCCTAT
j16	CCCAACCGCTGGCACAACA
j17	GATTGAAATGGTCTGCTGC
j18	AATGCTCAAAATGTTCTTACGA
j19	GCAACTCTGGCTCACAGTAC

doi:10.1371/journal.pone.0028876.t002

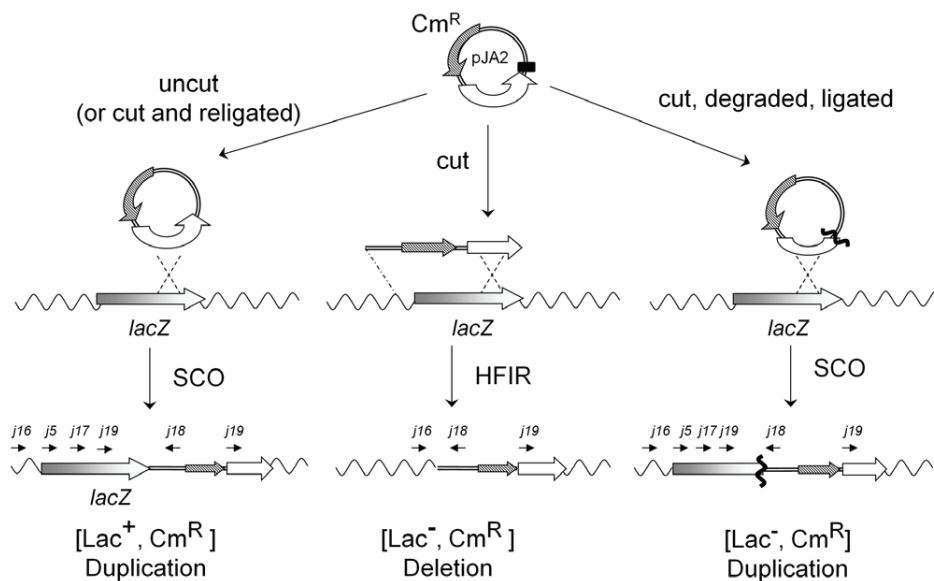


Figure 3. Inferred chromosomal recombination products obtained with plasmid pJA2. **Left part:** Single crossing-over (SCO) between circular pJA2 and the chromosome, as produced if pJA2 remains uncut (or if pJA2 is cut and resealed). The resulting exconjugants are $\text{Lac}^+ \text{Cm}^R$. **Central part:** HFIR between linear pJA2 and the chromosome. Exconjugants are $\text{Lac}^- \text{Cm}^R$, and therefore distinguishable from the previous SCO events. **Right part:** If pJA2 is cut, degraded up into the *lacZ* sequence, and recircularized, a SCO between a circular, deleted (wavy bold line) pJA2 molecule and the chromosome will give rise to $\text{Lac}^- \text{Cm}^R$ exconjugants. White arrow, 3' part of the *lacZ* gene. Grey box, remaining part of the *lacZ* gene. Wavy lines, other sequences of the *E. coli* chromosome. Straight line, pJA2 plasmid sequences. Black box, "cut cassette". Primers j18, j19, j17, j5, j16 complementary to the plasmid pJA2 and the chromosome, are shown above each chromosomal recombination product.
doi:10.1371/journal.pone.0028876.g003

endonuclease systems, *EcoK* and *I-sceI*, to linearise the incoming plasmids, a background ratio of 4×10^{-6} $\text{Lac}^+ \text{Cm}^R$ per recipient cells was obtained, which represents 2% of the recombinants obtained when the incoming plasmid is not cut. This is a ratio similar to that observed for pJA3 (1.3%), and corresponds to the inescapable background production of Lac^+ clones resulting from single crossing-overs. We nevertheless detected 4×10^{-9} $\text{Lac}^- \text{Cm}^R$ clones in the "double cut" context. No $\text{Lac}^- \text{Cm}^R$ clones were scored when either one of the cutting system was used alone. We cannot rule out that this absence of Lac^- is due to the technical difficulty of screening such clones among a wealth of Lac^+ colonies. Alternatively, it may be that the sticky ends produced by each single restriction system reseat more easily when the two restriction enzymes are not combined.

A control experiment with plasmid pJA1 (Figure 4, right panel) sharing no homology with the *E. coli* chromosome showed a marked, 1000 decrease in the yield of $\text{Cm}^R \text{Lac}^+$ clones compared

to pJA2 or pJA3 when no cleavage was applied, as expected. These most likely were produced by illegitimate recombination. Interestingly, their yield was also reduced 10 to 100 fold upon cleavage. These clones were not analysed further.

Detection of HFIR events with pJA2

We next proceeded to analyse the few $\text{Lac}^- \text{Cm}^R$ recipients produced with pJA2 when both cutting systems were active, as these were expected to correspond to HFIR events (see first section of the results). $\text{Lac}^- \text{Cm}^R$ recipients were found at a frequency of $5.8 (+/-4.0) \times 10^{-9}$ recombinants per recipient cells (60 Lac^- clones recovered over a total of 31 conjugation experiments, 0 to 7 Lac^- clones per experiment). All 60 Lac^- clones were analysed by PCR to test whether the integration of pJA2 into the chromosome had generated a deletion of part of *lacZ* as predicted for HFIR (Figure 3, middle). In most cases, unexpectedly, a duplication of part of *lacZ*, rather than a deletion, was observed (Figure 3, right

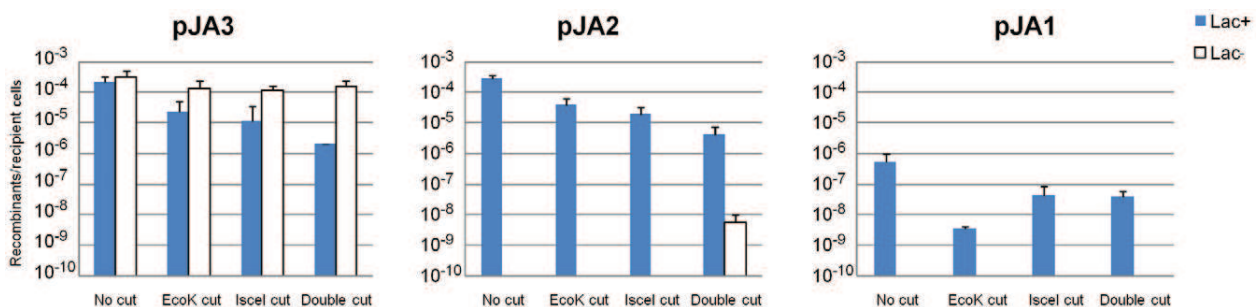


Figure 4. Efficiency of post-conjugative recombination with plasmids pJA3, pJA2 and pJA1. The Lac^+ (blue bars) and Lac^- (white bars, when present) Cm^R ex-conjugants reported to the viable recipient cells, are scored separately.
doi:10.1371/journal.pone.0028876.g004

part). The PCR test diagnostic for duplication use oligonucleotides j19 and j18, with j19 being placed at the limit of the *lacZ* end segment cloned into pJA2, and j18 facing j19, next to the Cm^R sequence (see Figure 3 lower panel for the positions of the oligonucleotides). In case of an HFIR event, these two oligonucleotides do not face each other (Fig. 3, middle) and no product is expected. In 56 of the 60 clones, a PCR product was found, whose size varied between 500 and 2 kb. Sequencing of 14 such clones confirmed the presence of a duplication of part of the *lacZ*-end homologous locus, with a size between 109 to 950 bp. We concluded that this category of Lac^- recombinants resulted from a single crossing-over between a pJA2 circular molecule lacking some of the *lacZ*-end fragment and the chromosome (see Figure 3, right part). Loss of some *lacZ* sequences in pJA2 could result from pre-existing plasmid deletion, or linearization by EcoK or IseI followed by partial, RecBCD dependent degradation of *lacZ*-end and re-circularisation. The results below, showing that a *recD* mutation decreased the yield of Lac^- clones, strongly suggests that the second alternative is more likely.

The remaining 4 of the 60 $\text{Cm}^R \text{Lac}^-$ clones appeared to be the result of HFIR. To detect the new junction expected, a set of primers upstream from the *lacZ* 5' locus (j17, j5 and j16, all 1 kb apart from each other, Figure 3, lower part) were tested in combination with j18, the pJA2-specific primer oriented towards the new junction. PCR products of different lengths were detected in all clones (one with the oligonucleotides combination j17-j18, three with j5-j18). Table 3 summarizes the deletion/integration length and the presence of homology at the junction for these clones. For 3 of the 4 clones, no homology was found at the new junction. We conclude that HFIR is detectable during conjugation in *E. coli*, albeit at a very low frequency. Based on the fact that 27 conjugations among 31 did not produce any HFIR clone, and using the first term of the Poisson distribution, we can estimate a frequency of HFIR of 3.5×10^{-10} , which is 10^6 fold below the frequency of homologous recombination events.

SCO are nearly independent of RecBCD and RecFOR, and inhibited by RuvA

To better understand recombination in our assay, experiments were performed in various mutants. RecA is the central actor of all homologous recombination reactions, while RecBCD and RecFOR are mediator proteins facilitating access of RecA to single-stranded DNA. The RecBCD complex is specialised in promoting RecA loading at the extremity of linear ds DNA substrates, whereas RecFOR act on single-stranded gaps. The RuvABC complex resolves Holliday junctions into recombination products at the last step of recombination, and RecG is involved in a

redundant pathway of resolution (for a review, see [11]). RecJ is a 5' to 3' single-strand exonuclease acting in the RecF pathway, and has been reported to prevent HFIR in *S. pneumoniae* [4], and UvrD displaces toxic RecA filaments [12].

Before testing recombination in strains permitting the linearization of pJA2, where several events are taking place (SCO and HFIR), we first examined the genetics of SCO, using donor and recipient strains that do not lead to pJA2 linearization. Efficiencies of recombination, measured as $\text{Cm}^R/\text{Recipient cells}$, are reported in Table 4. Recombination was found to be *recA* dependent, but *recB*, *recD* and *recF* independent. Mutants for *recD*, *recO* and *recR* were not affected either. The double *recB recF* mutant yielded to a 2 fold reduction in recombination efficiency. Finally, recombination was increased by a factor of 17 in the *ruvA* mutant, and 22 in a *uvrD* mutant.

Intrigued by the limited effects of the *recB* and *recF* mutations on SCO during conjugation, we asked whether the efficiency of SCO during electroporation was also unaffected by these mutations. The same bacteria were prepared for electroporation, and transformed with pJA2. As a control for transformation efficiency, 1 ng of the replicative plasmid pACYC184 was used. Recombination yields were measured relative to the transformation efficiency. In the MG1655 wild type strain, a yield of 1.4×10^{-4} Cm^R recombinants per transformed cells was obtained. In the *recB* and *recF* derivatives, yields of 4.8 and 4.3×10^{-5} respectively were found, meaning once again that SCO recombination is barely affected (3 fold) by defects in the *recF* or *recB* pathways.

We finally measured recombination efficiencies in mutants with linear pJA2, i. e. under conditions of cleavage by both EcoK and IseI. Both $\text{Lac}^+ \text{Cm}^R$ (SCO products) and $\text{Lac}^- \text{Cm}^R$ (essentially pJA2 trimming, followed by SCO) are reported in Table 5. Compared to circular pJA2, similar effects of the *recA*, *ruvA* and *recJ* mutations were observed. A slight change was found for the *recB*, *recF* and *recG* mutations, as a small but significant 4 to 5 fold reduction in the yield of recombinants was observed, compared to the wild type situation. This means that with the linear pJA2 substrate, $\sim 1/5$ th of the recombinants produced by RecA occur in a RecB, RecF and RecG independent manner. The recombination assay with linear pJA2 could not be performed in *uvrD* mutant, due to the instability of plasmid pIsceI in this context, nor in the *recB recF* double mutant due to its low viability.

In general, the mutations tested had a similar effect on the yield of Lac^+ and Lac^- clones, compared to the wild type strain. An interesting exception is the *recD* mutation. The *recD* mutation inactivates the dsDNA exonuclease (ExoV) activity of the RecBCD complex, while maintaining its recombination potential. The overall yield of $\text{Cm}^R \text{Lac}^+$ recombinants was increased by 4 fold,

Table 3. Properties of the HFIR clones obtained.

Strain name	Size of <i>lacZ</i> deletion, upstream of <i>lacZ</i> -end	Size of the pJA2 insert	Homology at the junction	Sequence of the joint ^a
MAC1343	1081 bp	1747 bp	None	5' ATGTCGGTTT CCGCGAGGTG 3' <i>lacZ</i> 5' TATCGATACC GTGACGCCG 3' pJA2
JAC233	1357 bp	1750 bp	None	5' TTTGCCGACC GCACGCCGA 3' <i>lacZ</i> 5' GCGTTCGAAA ATTCTCGTAC 3' pJA2
JAC228	245 bp	1989 bp	None	5' GCGCTGTACT GGAGGCTGAA 3' <i>lacZ</i> 5' CTTATCGATA CCGTGACGC 3' pJA2
JAC227	1385 bp	1974 bp	1 nt	5' GCCACTCGCT TTAATGATGA 3' <i>lacZ</i> 5' CGTACCACT TCTCTTGATC 3' pJA2

^aThe joint is composed of the concatenation of the two bold sequences from each parent.
doi:10.1371/journal.pone.0028876.t003

Table 4. Recombination frequencies with circular pJA2 in *E. coli* mutant derivatives.

MAC1308 Donor × Recipient strain ⁽¹⁾	Relevant genotype of recipient	Nb of experiments	Av. Cm ^R /recipient	Fold, relative to WT ⁽²⁾
MG1655	WT	5	3.3 (+/-1.1) × 10 ⁻⁴	1
JAC21	<i>recA</i>	4	2.7 (+/-2.3) × 10 ⁻⁷	0.0006 **
MAC1394	<i>recB</i>	4	2.7 (+/-0.9) × 10 ⁻⁴	0.58
MAC1354	<i>recD</i>	4	6.2 (+/-2.7) × 10 ⁻⁴	1.9
MAC1470	<i>recF</i>	4	4.8 (+/-1.6) × 10 ⁻⁴	1.45
MAC1503	<i>recO</i>	4	5.1 (+/-2.8) × 10 ⁻⁴	1.5
MAC1507	<i>recR</i>	4	3.5 (+/-1.6) × 10 ⁻⁴	1.04
MAC1350	<i>recJ</i>	4	4.0 (+/-2.0) × 10 ⁻⁴	1.2
MAC1511	<i>recB recF</i>	4	1.8 (+/-0.5) × 10 ⁻⁴	0.54 *
MAC1497	<i>ruvA</i>	4	5.6 (+/-4.4) × 10 ⁻³	17 *
MAC1476	<i>recG</i>	4	2.7 (+/-0.8) × 10 ⁻⁴	0.8
MAC1662	<i>uvrD</i>	4	7.2 (+/-2.9) × 10 ⁻³	22 *

⁽¹⁾Conjugations performed at 37°C.⁽²⁾*Significant below 5%, ** Significant below 1% with a Student test.

doi:10.1371/journal.pone.0028876.t004

but the proportion of Cm^R Lac⁻ was decreased 14 fold. These two effects are those expected for the exonuclease mutant: the 4-fold overall increase in recombination is due to the higher amount of linear DNA substrate that escapes ExoV degradation. It suggests that in a RecBCD+ background, with 4 fold less recombinants compared to the *recD* mutant, only ¼ of the linear molecules resist degradation, i.e. ¼ of the linear molecules are degraded. This degradation would have been even more pronounced if pJA2 had not been protected by its two Chi sites (see Methods). Therefore in the *recD* background, pJA2 linear molecules remain undegraded, can close up and integrate by SCO (see Figure 3: increase of the right pathway). The lower yield of Lac⁻Cm^R is also due to an escape from ExoV-mediated degradation of the *lacZ* end region: the *lacZ* segment remains complete and gives rise more often to a Lac⁺ recombinant (Figure 3, decrease of the left pathway). Actually, sequencing the junction in the three Lac⁻ clones obtained in the *recD* mutant revealed that none of them had suffered *lacZ* end degradation. These 3 clones were not produced by HFIR either, but rather resulted from a SCO between an undeleted pJA2 and the chromosome. Their Lac⁻ phenotype was

due to spontaneous mutations elsewhere in *lacZ*, either a 16 or 59 bp long deletion, or a 5 bp insertion in *lacZ*, but away from the junction (not shown). More generally, a PCR analysis of the Lac⁻ clones obtained in the various genetic backgrounds was done, to discriminate between true HFIR events and a *lacZ* duplication (oligonucleotides j18 and j19). Results are reported in Table 5, last column. We conclude that HFIR events were not drastically enriched in any of these backgrounds, as most events observed were again *lacZ* duplications.

The most striking conclusion of this genetics study is that SCO are almost independent of both the RecFOR and the RecBCD pathway, and inhibited by RuvA.

Discussion

During plasmid conjugation, in *E. coli*, HFIR at the *lacZ* locus occurs at a low frequency of 3 × 10⁻¹⁰, it is therefore not a frequent by-product of conjugation events. We can also exclude that HFIR is not detected due to the competition with a more efficient process removing the appropriate linear substrate. Indeed, the various steps

Table 5. Recombination frequencies with linear pJA2 in *E. coli* mutant derivatives.

MAC1306 Donor × Recipient strain	Relevant genotype of recipient	Nb of expe- riments	Av. LacZ ⁺ Cm ^R / recipient	Fold, relative to WT ⁽¹⁾	Frequency LacZ ⁻ Cm ^R / recipient ⁽²⁾	Fold, relative to WT	Proportion of HFIR events among LacZ ⁻
JAC7	WT	31	4.2 (+/-3.4) × 10 ⁻⁶	1	4.1 × 10 ⁻⁹	1	4/60
MAC1348	<i>recA</i>	12	1.3 (+/-0.8) × 10 ⁻⁸	0.003 **	<8.7 × 10 ⁻¹⁰	<0.21	NA
MAC1397	<i>recB</i>	12	7.0 (+/-7.0) × 10 ⁻⁷	0.17 **	<2.1 × 10 ⁻⁹	<0.26	NA
MAC1473	<i>recF</i>	12	1.1 (+/-0.6) × 10 ⁻⁶	0.26 **	1.8 × 10 ⁻⁹	0.31	1/9
MAC1357	<i>ruvA</i>	12	2.5 (+/-2.8) × 10 ⁻⁵	5.8 *	1.1 × 10 ⁻⁸	1.90	0/9
MAC1479	<i>recG</i>	12	7.5 (+/-6.3) × 10 ⁻⁷	0.18 **	1.5 × 10 ⁻⁹	0.26	0/2
MAC1352	<i>recD</i>	15	1.8 (+/-1.0) × 10 ⁻⁵	4.2 **	4.4 × 10 ⁻¹⁰	0.07	0/3
MAC1350	<i>recJ</i>	12	3.7 (+/-2.6) × 10 ⁻⁶	0.87	6.2 × 10 ⁻⁹	1.51	0/22

⁽¹⁾*Significant below 5%, ** Significant below 1% with a Student test.⁽²⁾Frequencies calculated with the first term of a Poisson law.

doi:10.1371/journal.pone.0028876.t005

of this study allow drawing a picture of the fate of linear DNA in *E. coli*: the pJA2 linear molecules are first degraded by RecBCD, and our measurements in the *recD* background suggest a frequency of degradation “D” of 75% of the molecules (this degradation would have been much higher in the absence of Chi sites). Among the molecules escaping degradation, a ligation step occurs, at a frequency “L” which we can estimate to be 10%, because the yield of recombinants with linear pJA2, in the *recD* background (2×10^{-5} , Table 5), is 10% that of recombinants with circular pJA2 in the wild type background (3.3×10^{-4} , Table 4). Finally, some of the circular molecules enter the chromosome by SCO, at a frequency “H” of 3×10^{-4} as deduced from the number of recombinants with circular pJA2 in the wild type background (Table 4, first lane). An even smaller fraction integrates the chromosome by illegitimate recombination (frequency $I = 5 \times 10^{-7}$, Figure 4 panel pJA1, value for uncut DNA). Therefore 22.5% of the linear molecules should remain linear and not degraded, providing sufficient substrate for HFIR. The remaining pJA2 molecules that do not recombine with the chromosome, both linear and circular, are probably diluted across generations, being non replicative.

A way to check whether estimations for D, L and H are correct, is to calculate the predicted frequency “F” of Lac⁺ clones obtained with pJA2 after linearization by double cutting: it should be the product of the probabilities “1-D” of not being degraded, “L” of being ligated, and “H” of recombining by SCO:

$$F = (1-D) \times L \times H,$$

This gives $F = (0.25) \times (0.1) \times (3 \times 10^{-4}) = 7.5 \times 10^{-6}$, a value not far from the observed value of 4.2×10^{-6} (Table 5).

We also detected events resulting from a partial RecBCD-dependent degradation of linear molecules, followed by ligation and SCO (Figure 3, right panel), which led to the very same phenotype as the HFIR events, and were 10 times more frequent than HFIR. The frequency of this event (Lac⁺ Cm^R) being 4×10^{-9} , we can deduce that partial degradation occurs at a frequency D’ such that:

$$D' \times L \times H = 4 \times 10^{-9}$$

Using the values estimated for L and H, this gives $D' = 1.3 \times 10^{-4}$. This illustrates the reported high processivity of RecBCD on its dsDNA substrate [13], once loaded, it does not unload frequently.

Our initial bioinformatics observation of abundant insertions coupled with deletions of short DNA fragments (also called dimorphic loci) among *E. coli* genomes is therefore unlikely to be accounted for by HFIR as suggested initially, and these rearrangements might be formed by more complex events [7]. Visual inspection of 47 dimorphic loci revealed that in 70% of the cases, the two supposedly different variable segments detected by multiple genomes alignment had in fact some level of similarity. Homeologous recombination between distantly related DNA

sequences could thus explain such situations [7]. The low efficiency of HFIR in *E. coli* during conjugation suggests that it is a process essentially restricted to naturally competent species, where the activities that process the incoming linear DNA might promote their efficient insertion coupled with a deletion in the chromosome. The observations reported in this work may even suggest that during transformation of these species, some protein might prevent the circularisation of linear double-strand DNA, to permit HFIR, or that DNA enters the chromosome in a single-strand state.

In the process of this work, we nevertheless came across a surprising observation, apparently unnoticed previously, that the exact mechanism by which SCO is taking place in *E. coli* is unknown at present. We report here that it is almost independent of both RecBCD and RecF pathways. Rather than supposing that the DNA substrate needed to initiate the recombination process is present on the incoming plasmid molecule, it may be that it is present on the recipient chromosome. For instance, a gap might be present in the process of some DNA repair (mismatch repair or nucleotide excision repair) or simply at a replication fork. Indeed, several reports point to the possible presence of RecA at stalled replication fork, with the need of a helicase such as UvrD to remove RecA from such forks [12,14,15]. In line with this hypothesis, we also report that SCO is increased in a *uvrD* background. However, RecA loading at the fork is thought to be *recF* dependent, an observation incompatible with the data reported here. The absence of any *recF*, *recO* or *recR* phenotype in our experimental set up remains puzzling, and suggest that RecA may use still other mediators, or none at all, to load onto certain substrates. We also report that RuvABC apparently inhibits SCO. This had already been reported in a different recombination assay based on replicative plasmids [16]. The proposed interpretation was that Ruv-mediated resolution might favour the non-cross-over products. Besides its role in resolution, RuvAB is also known to reverse some types of stalled replication forks [17]. It may be that by doing so, it removes the substrate with which circular plasmids usually interact for recombination. In conclusion, this work raises the interesting question of how SCO recombinants are produced in *E. coli*.

Acknowledgments

We thank Didier Mazel for providing the initial material for all conjugation experiments, and Marina Elez and Ariel Lindner for the gift of the pIsceI plasmid. We are indebted to François-Xavier Barre for decisive discussions at the onset of this work, and Colin R. Tinsley and Marianne de Paeppe for critical reading of the manuscript prior to publication.

Author Contributions

Conceived and designed the experiments: JAB MAP. Performed the experiments: JAB MG MAP. Analyzed the data: JAB MG MAP. Contributed reagents/materials/analysis tools: JAB MAP. Wrote the paper: MAP.

References

1. Prudhomme M, Libante V, Claverys JP (2002) Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* 99: 2100–2105.
2. de Vries J, Wackernagel W (2002) Integration of foreign DNA during natural transformation of *Acinetobacter* sp. by homology-facilitated illegitimate recombination. *Proc Natl Acad Sci U S A* 99: 2094–2099.
3. Meier P, Wackernagel W (2003) Mechanisms of homology-facilitated illegitimate recombination for foreign DNA acquisition in transformable *Pseudomonas stutzeri*. *Mol Microbiol* 48: 1107–1118.
4. Harms K, Schon V, Kickstein E, Wackernagel W (2007) The RecJ DNase strongly suppresses genomic integration of short but not long foreign DNA fragments by homology-facilitated illegitimate recombination during transformation of *Acinetobacter baylyi*. *Mol Microbiol* 64: 691–702.
5. Lovett ST, Kolodner RD (1989) Identification and purification of a single-stranded-DNA-specific exonuclease encoded by the *recJ* gene of *Escherichia coli*. *Proc Natl Acad Sci U S A* 86: 2627–2631.
6. Lovett ST, Kolodner RD (1991) Nucleotide sequence of the *Escherichia coli* *recJ* chromosomal region and construction of *recJ*-overexpression plasmids. *J Bacteriol* 173: 353–364.
7. Touzain F, Denamur E, Medigue C, Barbe V, El Karoui M, et al. (2010) Small variable segments constitute a major type of diversity of bacterial genomes at the species level. *Genome Biol* 11: R45.

8. Demarre G, Guerout AM, Matsumoto-Mashimo C, Rowe-Magnus DA, Marliere P, et al. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalph) conjugative machineries and their cognate *Escherichia coli* host strains. *Res Microbiol* 156: 245–255.
9. Ferrieres L, Hemery G, Nham T, Guerout AM, Mazel D, et al. (2010) Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J Bacteriol* 192: 6418–6427.
10. Loenen WA (2003) Tracking EcoKI and DNA fifty years on: a golden story full of surprises. *Nucleic Acids Res* 31: 7059–7069.
11. Persky NS, Lovett ST (2008) Mechanisms of recombination: lessons from *E. coli*. *Crit Rev Biochem Mol Biol* 43: 347–370.
12. Veaute X, Delmas S, Selva M, Jéusset J, Le Cam E, et al. (2005) UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in *Escherichia coli*. *EMBO J* 24: 180–189.
13. Roman LJ, Eggleston AK, Kowalczykowski SC (1992) Processivity of the DNA helicase activity of *Escherichia coli* recBCD enzyme. *J Biol Chem* 267: 4207–4214.
14. Flores MJ, Sanchez N, Michel B (2005) A fork-clearing role for UvrD. *Mol Microbiol* 57: 1664–1675.
15. Lestini R, Michel B (2007) UvrD controls the access of recombination proteins to blocked replication forks. *EMBO J* 26: 3804–3814.
16. Lovett ST, Hurley RL, Sutera VA, Jr., Aubuchon RH, Lebedeva MA (2002) Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics* 160: 851–859.
17. Baharoglu Z, Petranovic M, Flores MJ, Michel B (2006) RuvAB is essential for replication forks reversal in certain replication mutants. *EMBO J* 25: 596–604.

Résumé

Les génomes de bactériophages ont une capacité remarquable d'évolution. L'objectif de ma thèse a été d'étudier le rôle des recombinaisons de bactériophages dans cette évolution. Notre hypothèse était que les recombinaisons phagiques diffèrent de la recombinaison bactérienne RecA par une fidélité relâchée lors de la réaction de recherche d'homologie, qui pourrait expliquer en partie la très grande plasticité des génomes de bactériophages. Nous avons tout d'abord utilisé une approche bio-informatique basée sur la recherche d'homologies lointaines pour prédire un maximum de gènes de recombinaisons dans les génomes entièrement séquencés, puis nous avons confirmé l'activité de recombinaison de certaines d'entre elles, par un test d'appariement d'ADN simple brin entre séquences identiques *in vivo*. Ceci nous a permis de conclure qu'il existait trois superfamilles de recombinaisons chez les bactériophages, de type Rad52, Rad51/RecA et Gp2.5, présentes dans 42% des 465 génomes analysés. Dans un deuxième temps, nous avons comparé six de ces recombinaisons à RecA et avons montré que toutes étaient capables de faire de l'appariement simple brin *in vivo*, contrairement à RecA. Pour deux d'entre elles, Red β de Lambda (type Rad52) et Sak4 de HK620 (type Rad51/RecA), nous avons observé que l'appariement simple brin continuait à se produire, avec une efficacité diminuée, jusqu'à 13% de divergence entre les séquences. L'appariement d'ADN simple brin est donc une propriété commune aux recombinaisons de bactériophages qui les distingue de RecA, et semble pouvoir se maintenir pour un niveau élevé de divergence, ce qui soutient l'hypothèse d'une recombinaison homologue différente et plus relâchée dans sa fidélité chez les bactériophages.

Summary

Bacteriophage genomes have a remarkable ability to evolve. The aim of my thesis was to study the role of bacteriophage recombinases in this evolution. Our hypothesis was that such recombinases differ from the bacterial RecA recombinase by a relaxed fidelity during homology search, which may partly explain the high plasticity of bacteriophage genomes. We first used a bioinformatics approach based on remote homology search to predict a maximum of recombinase genes in completely sequenced genomes, and confirmed the recombination activity of some of them by a single-strand DNA annealing assay between identical sequences *in vivo*. This allowed us to conclude that there were three superfamilies of bacteriophage recombinases, Rad52-like, Rad51/RecA-like and Gp2.5-like, which were present in 42% of the 465 genomes analyzed. In a second step, we compared six of these recombinases to RecA and showed that all were able to anneal single-stranded DNA *in vivo*, in contrast to RecA. For two of them, Red β of Lambda (Rad52-like) and Sak4 of HK620 (Rad51/RecA-like), we also observed that they were able to anneal non identical (13% of divergence) single-stranded DNA, with a reduced efficiency. We conclude that the single-stranded DNA annealing is a property common to recombinases of bacteriophages, which is absent in RecA, and seems to tolerate diverged sequences. This supports the hypothesis of a different and more relaxed recombination in bacteriophages.