

2m11.2668.10

Université de Montréal

Clonage et caractérisation du gène *xerD* de *Proteus mirabilis*

par

Manuela Villion

Département de microbiologie et immunologie

Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M. Sc.)
en microbiologie

Septembre 1998

© Manuela Villion, 1998



W

4

U58

1999

V.034

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Clonage et caractérisation du gène *xerD* de *Proteus mirabilis*

présenté par :

Manuela Villion

a été évalué par un jury composé des personnes suivantes :

Président du jury : Dr Marc Drolet

Directeur de recherche : Dr George Szatmari

Membre du jury : Dr Josée Harel

Mémoire accepté le : 99.01.29

Sommaire

Le maintien des réplicons (chromosome et plasmides) chez *Escherichia coli*, est assuré par un système de recombinases spécifique de site, le système Xer. Les deux recombinases impliquées (XerC et XerD) font partie de la famille des intégrases du phage λ . Elles agissent de façon coopérative sur des loci particuliers que l'on retrouve sur le chromosome (site *dif*) et sur les plasmides (ex : *cer* de ColE1, *psi* de pSC101, etc.).

Tous les sites d'action ont une région centrale variant de 28 à 30 pb. Elle se subdivise en trois parties : deux séquences de 11 pb flanquent le cœur de la région où se fait l'échange. La partie « gauche » sert de site de liaison à XerC, tandis que la « droite » sert de site de liaison à XerD.

Ces recombinases nécessitent des facteurs accessoires pour résoudre les multimères plasmidiques qui surviennent au cours de la recombinaison homologue. Pour la recombinaison au site *cer* de ColE1, ces facteurs sont ArgR et PepA, tandis que pour la recombinaison au site *psi* de pSC101, les facteurs sont PepA et le système ArcA/ArcB. Dans les deux cas, ils permettent la sélectivité pour la résolution. La recombinaison se fait alors de façon intramoléculaire entre des sites directement répétés. Par contre, la recombinaison au site *dif* peut être aussi bien intramoléculaire qu'intermoléculaire et elle ne nécessite pas de facteurs accessoires.

La réaction de recombinaison aux sites *cer* et *psi* est séquentielle, en ce sens que XerC effectue le premier clivage entre son site de liaison et la région centrale. Cela mène à la formation d'une jonction de Holliday. Dans le cas de la recombinaison au site *psi*, XerD effectue le deuxième échange pour obtenir le produit recombinant, tandis que cette réaction semble être effectuée par un autre système (non encore déterminé) au niveau du site *cer*, et ce, même si la présence de XerD y est obligatoire.

Compte tenu du rôle important de ce système pour la stabilité de l'information génétique d'*E. coli*, l'on pouvait fortement supposer qu'il devait exister chez d'autres espèces bactériennes. Sa présence a d'ailleurs été montrée chez certaines *Enterobacteriaceae* (Sirois et Szatmari, 1995) et des recombinases similaires ont été séquencées chez plusieurs autres espèces, autant gram-négatives que positives.

L'observation de Goebel et Helinski (1968) de la multimérisation de ColE1 dans *Proteus mirabilis* avait laissé supposer que le système Xer ne devait pas exister chez cette espèce, ou du moins, être non fonctionnel. Cependant, l'activité Xer au site *cer* a été détectée par le laboratoire (Sirois et Szatmari, 1995). Il devenait donc intéressant de cloner et de séquencer les gènes de cette espèce qui pourraient encoder des protéines similaires à XerC et XerD d'*E. coli*. Les méthodes de PCR, dégénéré et inverse, ont permis de réaliser cet objectif et le mémoire présenté ici, fait part principalement du clonage et de la caractérisation du gène *xerD* de *P. mirabilis*. Malgré un assez bon pourcentage d'identité entre la séquence en acides aminés de XerD de cette espèce et celle d'*E. coli*, les études de complémentation ne se sont pas révélées aussi concluantes que ce à quoi on aurait pu s'attendre. Par contre, des tests de retardement sur gel ont permis de montrer que la protéine encodée par le gène séquencé était bel et bien apte à se lier au site *cer*.

Mots-clés :

Recombinaison spécifique de site\ Intégrase\ XerD\ XerC\ *Proteus mirabilis*.

Table des matières

IDENTIFICATION DU JURY	ii
SOMMAIRE.....	iii
TABLE DES MATIÈRES	v
LISTE DES TABLEAUX	viii
LISTE DES FIGURES	ix
LISTE DES SIGLES ET ABRÉVIATIONS	x
DÉDICACE	xii
CHAPITRE I- INTRODUCTION.....	1
1. LA RECOMBINAISON SPÉCIFIQUE DE SITE	2
1.1 Généralités.....	2
1.2 Les résolvases/invertases.....	3
1.3 Les intégrases	4
1.4 Points communs parmi les intégrases	5
2. LES RECOMBINASES DE TYPE XER	6
2.1 XerC	7
2.2 XerD	7
2.3 Rôles de XerC et XerD.....	8
2.4 Activité topoisomérase.....	9
2.4.1 <i>Les topoisomérases d'E. coli</i>	9
2.4.2 <i>Points communs aux topoisomérases et aux intégrases</i>	10
3. MÉCANISME D' ACTION DES INTÉGRASES : ON CLIVE EN CIS OU EN TRANS ?	12
4. SITES D' ACTION.....	15
4.1 Sites <i>dif</i> , <i>cer</i> et <i>psi</i>	15
4.1.1 <i>Site dif</i>	15
4.1.2 <i>Site cer</i>	19
4.1.3 <i>Site psi</i>	19
4.2 Principales différences entre les sites <i>dif</i> , <i>cer</i> et <i>psi</i>	20
4.3 Détermination de la sélectivité des sites	20
4.4 Autres sites	22
5. FACTEURS ACCESSOIRES	23

5.1 ArgR et PepA	24
<u>5.1.1 XerA= ArgR</u>	24
<u>5.1.2 XerB=PepA</u>	25
<u>5.1.3 Rôle des deux protéines dans la recombinaison au site cer</u>	26
5.2 ArcA/ArcB	28
6. ÉTUDE DES INTERMÉDIAIRES DE LA RECOMBINAISON	29
6.1 Jonction de Holliday.....	29
<u>6.1.1 Structure de la jonction</u>	29
<u>6.1.2 La structure dans la recombinaison</u>	31
<u>6.1.3 Modèle de la réaction de recombinaison</u>	31
<u>6.1.4 Action de XerC et XerD</u>	33
6.2 Les substrats suicides	35
6.3 Formes des substrats pour la recombinaison par Xer.....	35
7. STRUCTURE TRIDIMENSIONNELLE DE CERTAINES INTÉGRASES	36
7.1 Structure de XerD.....	37
7.2 Structure de l'Int de λ	38
7.3 Structure de l'Int de HP1	39
7.4 Structure de Cre.....	40
7.5 Comparaison de la structure de XerD aux autres structures d'intégrases.....	41
8. CARACTÉRISATION DE XERD	42
9. UBIQUITÉ DES RECOMBINASES DE TYPE XER	45
10. PROJET DE MAÎTRISE.....	47
CHAPITRE II- ARTICLE	48
ABSTRACT.....	49
1. INTRODUCTION.....	49
2. MATERIALS AND METHODS.....	50
2.1 Bacterial strains and plasmids	50
2.2 Growth conditions and DNA manipulations	50
2.3 PCR conditions	50
2.4 Sequence analysis	51
2.5 Protein overexpression and DNA-binding assays	51
3. RESULTS.....	51
3.1 Cloning of <i>xerD</i> from <i>P. mirabilis</i>	51

3.2 Characteristics of the sequence.....	53
3.3 Complementation analysis.....	53
3.4 Expression and partial purification of the <i>P. mirabilis</i> XerD recombinase	54
4. DISCUSSION.....	54
ACKNOWLEDGMENTS.....	55
REFERENCES.....	56
CHAPITRE III- DISCUSSION	57
1. CARACTÉRISATION DE LA SÉQUENCE.....	58
2. EXPRESSION DE XERD.....	59
3. ÉTUDES DE COMPLÉMENTATION ET DE RETARDEMENT SUR GEL	60
3.1 Liaison à l'ADN	60
3.2 Clivage et activation.....	61
3.3 Coopérativité	63
4. À SUIVRE.....	66
BIBLIOGRAPHIE	68
ANNEXES	xiii
I ALIGNEMENT DE PLUSIEURS RECOMBINASES DE TYPE XER	xiii
II COMPARAISON DES ORGANISATIONS GÉNOMIQUES DE <i>XERC</i> ET <i>XERD</i> D' <i>E. COLI</i> ET DE <i>P. MIRABILIS</i>	xvi
III CARACTÉRISTIQUES DE XERD SUITE À L'INSERTION ALÉATOIRE DE PENTAPEPTIDES	xviii
IV ESPÈCES CHEZ LESQUELLES DES GÈNES ENCODANT DES RECOMBINASES HOMOLOGUES À XERC OU XERD ONT ÉTÉ SÉQUENCÉS	xx
V SÉQUENCE DE XERC DE <i>P. MIRABILIS</i>	xxii
REMERCIEMENTS	xxiv

Liste des tableaux

Chapitre I- Introduction

TABLEAU I- Rôles présumés de certains résidus ou régions de XerD	45
--	----

Liste des figures

Chapitre I- Introduction

FIGURE 1- Résultats possibles de la recombinaison spécifique de site	2
FIGURE 2- Schémas des diverses possibilités de forme de clivage s'offrant aux intégrases.....	13
FIGURE 3- Alignement de plusieurs sites d'action ayant une homologie avec le site <i>cer</i>	16
FIGURE 4- Formes possibles de l'enroulement autour d'ArgR et de PepA.....	26
FIGURE 5- Modèles pour le complexe synaptique entre deux sites <i>cer</i>	27
FIGURE 6- Formes que peut prendre la jonction de Holliday	30
FIGURE 7- Mécanisme d'action des recombinases XerC et XerD au site <i>dif</i> selon le modèle de « strand swapping-isomerization »	32
FIGURE 8- Structure cristalline de XerD d' <i>E. coli</i>	37
FIGURE 9- Structure cristalline du domaine C-terminal de l'Int de λ	38
FIGURE 10- Structure cristalline du domaine C-terminal de l'Int de HP1	39
FIGURE 11- Structure cristalline complète de Cre	40

Chapitre II- Article

FIGURE 1- Alignment of <i>E. coli</i> and <i>P. mirabilis</i> XerD proteins.....	52
FIGURE 2- Complementation analysis of an <i>E. coli xerD</i> mutant with <i>P. mirabilis</i> XerD	53
FIGURE 3- Overexpression and purification of <i>P. mirabilis</i> XerD.....	53
FIGURE 4- DNA binding assay.....	54

Chapitre III- Discussion

FIGURE 12- Vecteur utilisé pour les études de complémentation	62
---	----

Sigles et Abréviations

Acides aminés :

A : Ala : Alanine
 C : Cys : Cystéine
 D : Asp : Acide aspartique
 E : Glu : Acide glutamique
 F : Phe : Phénylalanine
 G : Gly : Glycine
 H : His : Histidine
 I : Ile : Isoleucine
 K : Lys : Lysine
 L : Leu : Leucine
 M : Met : Méthionine
 N : Asn : Asparagine
 P : Pro : Proline
 Q : Gln : Glutamine
 R : Arg : Arginine
 S : Ser : Sérine
 T : Thr : Thréonine
 V : Val : Valine
 W : Trp : Tryptophane
 Y : Tyr : Tyrosine

Acides nucléiques :

A : Adénine
 C : Cytosine
 G : Guanine
 T : Thymine

Unités de mesure :

Å : Angström : 10^{-10} mètre
 Da : Dalton
 g : pour la centrifugation : représente
 $9,8 \text{ m/s}^2$
 g : gramme
 h : heure
 kb : kilobase(s)
 kDa : kilodalton
 µg : microgramme

µg/ml : microgramme par millilitre
 min : minute
 µl : microlitre
 ml : millilitre
 µm : micromètre
 mM : millimolaire ou millimole/litre
 ng : nanogramme
 nm : nanomètre
 ° : Degré
 °C : degré Celsius
 pmol : picomole
 s : seconde
 V/cm : volt par centimètre

Autres :

3'OH : 3 prime hydroxyle
 α : Alpha
 aa : acide aminé ou amino acid
 ADN : Acide désoxyribonucléique
 AMPc : Adénosine monophosphate
 cyclique
 Ap : Ampicilline
 ARN : Acide ribonucléique
 β : Bêta
 C-terminal : Carboxy-terminal
 δ ou Δ : Delta
 DNase : Désoxyribonucléase
 EDTA : Acide
 éthylènediaminetétraacétique
 Fig. : Figure
 γ : Gamma
 Int : Intégrase
 L-Arg : Arginine lévogyre
 Mg²⁺ : Ion magnésium
 Mn²⁺ : Ion manganèse
 N-terminal : Amino-terminal
 ORF : open reading frame (cadre
 ouvert de lecture)
 pb : paires de bases

PCR : Polymerase chain reaction (réaction de polymérisation en chaîne)
PM : Poids moléculaire
Réf. : Référence
RNase : Ribonucléase
sp. : species
Tc : Tétracycline
Tn : transposon
Zn²⁺ : Ion zinc

Abréviations spécifiques au chapitre II :

bp : base pairs
DIG : digoxigenin
DNA : deoxyribonucleic acid
IPTG : Isopropylthiogalactoside
LB : Luria-Bertani
NEB : New England Biolabs
NSERC: Natural Sciences and Engineering Research Council of Canada (en français : CRSNG)
OD₆₀₀ : Optical density at 600 nanometers (densité optique à 600 nm)
pers. comm. : personal communication
ref. : reference
SDS-PAGE : Polyacrylamide gel electrophoresis in sodium dodecyl sulfate (gel d'électrophorèse de polyacrylamide contenant du SDS)
Tn10 : transposon encodant la résistance à la tétracycline
Tn9 : transposon encodant la résistance au chloramphénicol
Tp : trimethoprim
UV : ultraviolet
X-gal : 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

À Quinou, Yanou et Kano (!!!)

« Enfin, quand il eut épuisé toutes ses ruses. [Proteus] revint à sa forme ordinaire, et donna à Ménélas les éclaircissements qu'il lui demandait. »

P. Commelin,
Mythologie grecque et romaine.

« [Proteus] symbolise la matière originelle qui servit à créer le monde. »

Dictionnaire culturel de
la mythologie gréco-romaine,
Sous la direction de R. Martin.

CHAPITRE I :

INTRODUCTION

Chapitre I- Introduction

1. La recombinaison spécifique de site

1.1 Généralités

Les réactions impliquant des réarrangements de l'ADN se divisent en trois catégories (Sadowski, 1993):

- la recombinaison générale,
- la transposition et,
- la recombinaison spécifique de site.

La recombinaison spécifique de site se distingue des deux autres mécanismes par le fait qu'elle est conservatrice. Elle n'implique ainsi aucun gain ou perte de nucléotides, il y a absence de réplication et elle ne nécessite

nullement de cofacteurs énergétiques supplémentaires. Elle est, par le fait même, réciproque. De plus, il y a une spécificité et une orientation (polarité) dans les sites où elle s'effectue. Comme le nom l'indique, elle est en fait une réaction au cours de laquelle des brins d'ADN sont brisés puis échangés en des positions précises, et ce, dans le but de permettre certaines fonctions biologiques particulières (Nash, 1996).

La recombinaison spécifique de site peut donner lieu à plusieurs résultats, selon qu'elle est

intramoléculaire, c'est-à-dire si l'action se produit entre deux sites situés sur la même molécule, ou qu'elle est intermoléculaire, si la recombinaison s'effectue entre des sites situés sur deux molécules différentes, comme par exemple deux plasmides. Dans le premier cas, cela peut aboutir à une inversion des segments d'ADN situés entre les deux sites quand ils sont inversement répétés, ou à une résolution ou excision, lorsque

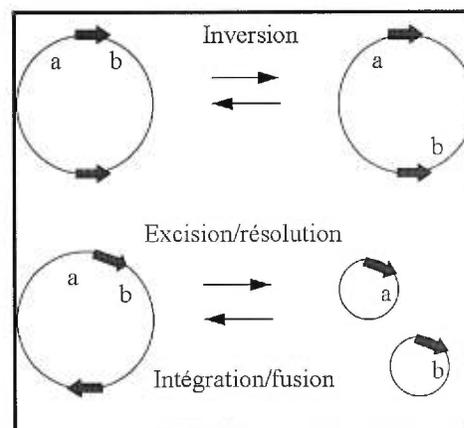


Figure 1- Résultats possibles de la recombinaison spécifique de site. Les flèches indiquent les sites d'action et leur orientation. (Fig. adaptée de Hallet et Sherratt, 1997).

les deux sites sont directement répétés. Dans le second cas, il peut y avoir une intégration ou une fusion de deux molécules (fig. 1) (Hallet et Sherratt, 1997). La recombinaison permet ainsi de procurer de la variabilité génétique par les substitutions ("switch") dans l'expression génique et l'épissage de certains gènes, de résoudre des réplicons dimériques pour assurer leur maintien stable sous forme de monomères, elle sert à l'intégration et à l'excision de certains phages dans le chromosome bactérien, etc. (Stark *et al.*, 1992).

Les recombinases qui effectuent ce type de réaction se divisent en deux grandes classes : les résolvases/invertases et les intégrases de la famille de l'Int du phage λ (qu'on nommera seulement intégrases tout au long du texte). Ces deux familles se distinguent par leur mécanisme d'action : alors que les résolvases/invertases interagissent de façon covalente avec l'ADN par l'intermédiaire d'une sérine située dans le domaine N-terminal, le résidu catalytique des intégrases est une tyrosine retrouvée dans le domaine C-terminal de la protéine. Dans les deux cas cependant, la réaction se produit par des mécanismes de transestérification en deux étapes. Dans les deux cas également, il peut y avoir nécessité de séquences et de protéines accessoires pour mieux contrôler la recombinaison. Seules les résolvases ne semblent en avoir nul besoin (Stark *et al.*, 1992). De plus, le surenroulement de l'ADN peut être requis pour l'assemblage et/ou l'activité du complexe synaptique.

1.2 Les résolvases/invertases

Les résolvases les plus connues font partie des transposons Tn3 et $\gamma\delta$, tandis que les systèmes Gin/gix du bactériophage Mu et Hin/hix de *Salmonella sp.* sont des invertases. Dans le cas des invertases, un activateur est requis, et la protéine Fis (facteur pour la stimulation de l'inversion) est nécessaire pour son activité (Van de Putte et Goosen, 1992). La première attaque nucléophile contre le lien phosphodiester survient par le groupement hydroxyle de la sérine de l'enzyme. Il y a alors libération d'un groupement 3'OH du désoxyribose et la sérine se retrouve attachée au côté 5'phosphate. C'est le groupement hydroxyle libéré qui attaquera à son tour le lien

phosphoséryle du brin partenaire pour donner lieu à la ligation¹ dans une deuxième étape. Ces deux étapes sont effectuées par quatre recombinases qui sont chacune liée à leur site respectif sur les deux paires de brins d'ADN qui seront échangées. Les quatre brins se voient donc clivés et reliés en même temps. Avant l'étape de religation, il y a une rotation de 180° vers la droite de l'une des paires de brins (revue dans Hallet et Sherratt, 1997). La région centrale entre les deux points de clivage d'une paire de brins est de 2 pb.

La structure du dimère de $\gamma\delta$ lié au site d'ADN révèle un motif hélice-tour-hélice pour la liaison à l'ADN, dans le domaine C-terminal de la protéine, et l'ADN est courbé de 60° (Yang et Steitz, 1995). Enfin, les résidus catalytiques de la résolvasse $\gamma\delta$ agissent en *cis*.

1.3 Les intégrases

La famille des intégrases est plus divergente que celle des résolvasse/invertases. En plus de l'Int du phage λ , elle comprend l'Int du phage HP1, la protéine Cre du phage P1, la protéine Flp du plasmide 2μ de *Saccharomyces cerevisiae*, les recombinases XerC et XerD d'*E. coli*, les protéines FimB et FimE d'*E. coli*, etc.

La première attaque du lien phosphodiester par les intégrases survient par une tyrosine. Il y a ainsi formation d'un lien 3'phosphotyrosyle et libération d'un groupement 5'OH. La deuxième étape consiste en une attaque nucléophile par le groupement 5'OH libéré du duplexe partenaire pour donner lieu à la jonction de Holliday (voir la section 6 de la présente introduction). Dans cette première réaction, seuls les brins dits supérieurs ont été échangés. Les brins inférieurs sont ensuite échangés pour donner les produits recombinants, et cela se fait par le même processus que précédemment. La région centrale séparant le point de clivage du brin supérieur de celui du brin inférieur varie de 6 à 8 pb. Le mécanisme exact par lequel intervient

¹ Le terme ligation est, semble-t-il, préférable en français. Cependant, compte tenu de l'emploi plus fréquent de « ligation », ce dernier sera utilisé tout au long du mémoire. De plus, il est accepté dans le Dictionnaire anglais-français des sciences médicales et paramédicales de W.J. Gladstone (4^e édition, Edisem-Maloine, Canada, 1996).

l'homologie de la région centrale où se fait l'échange des brins, est inconnu, mais il est su qu'elle doit jouer un rôle dans la plupart des cas. Deux modèles ont été proposés : le modèle de la migration de jonction ("branch-migration") et le modèle de l'isomérisation suivant l'échange des brins ("strand swapping-isomerization"). Ces deux modèles sont discutés au point 6.1.3 de l'introduction. Il semblerait cependant que le dernier modèle soit le plus probable (Hallet et Sherratt, 1997 et voir la fig.7).

Contrairement aux résolvasés/invertases, il y a une divergence au sein des membres des intégrases, quant à savoir par quelle sous-unité de recombinaise se feront les réactions de clivage dans les différents systèmes. C'est le paradoxe *cis-trans* (voir la section 3).

Enfin, la structure de quatre intégrases a été obtenue dans la dernière année et révèle un repliement nouveau dans le domaine C-terminal catalytique, conservé chez les quatre recombinaises (Cre de P1, Int de Hp1, Int de λ et XerD) (voir la section 8).

1.4 Points communs parmi les intégrases

Comme déjà mentionné, la famille des intégrases est assez hétérogène. En faisant une étude comparative de sept systèmes bactériophagiques différents, Argos *et al.* (1986) ont pourtant réussi à obtenir un alignement de la région C-terminale des séquences dans lequel deux régions montraient plus de conservation que le reste de la séquence. Dans la seconde région, trois résidus étaient absolument conservés. Il s'agissait d'une histidine et d'une arginine dans un motif H-X-X-R et d'une tyrosine (Argos *et al.*, 1986). Il a été supposé que ces trois résidus devaient faire partie du site actif des enzymes, et la tyrosine a effectivement été montrée comme le résidu catalytique chez l'Int de λ (Pargellis *et al.*, 1988). Une deuxième arginine, située un peu plus en N-terminal que les trois autres résidus, s'est également révélée absolument conservée lors de l'analyse des séquences de 28 intégrases (Abremski et Hoess, 1992).

Par la suite, d'autres alignements ont été effectués avec un nombre croissant de séquences d'intégrases. L'histidine n'est alors plus absolument conservée. La

première analyse, regroupant la partie C-terminale de 81 séquences, révèle un motif K-G-H qui est très bien conservé parmi toutes les intégrases, dans la région où l'on retrouve la tyrosine catalytique (Esposito et Scocca, 1997).

La seconde analyse, également des domaines catalytiques, mais cette fois-ci de 105 membres, a tenu compte des structures cristallines obtenues pour l'Int de λ , l'Int de Hp1, Cre de P1 et XerD (Nunes-Düby *et al.*, 1998). Cela a permis l'identification de nouveaux motifs impliqués dans la structure et dans les activités biologiques. La conservation de R-H-R est également observée au niveau de la structure secondaire des quatre recombinases mentionnées plus haut. La boîte II ayant H-R-Y est très conservée parmi les procaryotes, mais les résidus entourant la tyrosine sont plutôt divergents. Puisque les résidus hydrophobes sont assez bien conservés dans la région couvrant la boîte I à la boîte II, il est supposé que toutes les intégrases adopteraient la même structure secondaire à ce niveau.

La figure en annexe 1 montre un alignement des recombinases de type Xer. Chez ces recombinases, en plus des deux autres motifs retrouvés chez les intégrases, un troisième motif a été révélé (Cao *et al.*, 1997); il est indiqué sur la figure.

2. Les recombinases de type Xer

La découverte des recombinases de type Xer est survenue à la suite de l'étude de la stabilité plasmidique de ColE1. Les premières protéines identifiées comme ayant un rôle dans le maintien de ce plasmide sous forme de monomère se sont révélées être accessoires (voir la section 5 de l'introduction). Il restait donc à découvrir les protéines essentielles à ce processus, à savoir celles qui effectuent la recombinaison en tant que telle. À la fin de l'année 1990, un troisième gène impliqué dans la recombinaison au site *cer* (site retrouvé sur ColE1), était révélé : le gène *xerC* (Colloms *et al.*, 1990). En 1993, un autre gène, *xerD*, a également été impliqué dans la recombinaison au site *cer* et au site *dif* (site du chromosome d'*E. coli*) (Blakely *et al.*, 1993). Ces deux gènes encodent des protéines montrant de l'homologie aux autres membres de la famille des intégrases.

2.1 XerC

Le gène *xerC* est localisé à 3993,9 kb sur le chromosome d'*E. coli* (numéro d'accèsion de GenBank: U00096) et code pour la protéine XerC de 298 acides aminés, avec un poids moléculaire de 33,8 kDa (Colloms *et al.*, 1990). Le gène *xerC* se situe dans une unité multicistronique ayant le gène *dapF* en amont. Ce dernier code pour la diaminopimélate épimérase qui est impliquée dans la biosynthèse de la lysine et du peptidoglycane (Richaud *et al.*, 1987). Deux cadres ouverts de lecture (ORF: "open reading frame") flanquent *xerC* et ils coderaient pour deux protéines ayant 235 et 238 acides aminés immédiatement en amont et en aval de *xerC* respectivement (voir l'annexe 2). Un peu plus en aval, on retrouve le promoteur P1 du gène *uvrD*.

La séquence peptidique de XerC est identique à 32% avec les séquences de FimB et FimE ; ceux-ci sont impliqués dans la variation des fimbriae de type I chez *E. coli*, en médiant l'inversion d'un segment qui comprend le promoteur de *fimA* (Klemm, 1986), et font partie de la famille des intégrases du phage λ . D'ailleurs, la séquence de XerC révèle deux domaines qui sont similaires aux deux domaines conservés de cette famille. Elle contient les trois acides aminés absolument conservés de la famille dans le 2^e motif, à savoir une histidine, une arginine et une tyrosine (Colloms *et al.*, 1990) et également la 2^e arginine du motif I (Abremski et Hoess, 1992).

2.2 XerD

Lors du séquençage de la région chromosomique comprenant *recJ*, un cadre ouvert de lecture nommé *xprB* a montré de l'homologie avec la région C-terminale conservée de plusieurs intégrases (Lovett et Kolodner, 1991). Ce gène est situé à 3036,9 kb sur le chromosome d'*E. coli* (numéro d'accèsion de GenBank: U00096). Tout comme XerC, la séquence peptidique de XprB (298 aa) avait le plus fort pourcentage d'identité avec FimB et FimE. En comparant cette séquence avec celle de XerC, Blakely *et al.* (1993) ont supposé, et par la suite confirmé, que *xprB* codait pour une seconde recombinase du système Xer. Les deux protéines ont 37%

d'identité, ce qui est le plus fort pourcentage déterminé entre deux membres de la famille des intégrases. De plus, XprB contient les deux motifs conservés de la famille de l'intégrase de λ . Le gène, alors renommé *xerD*, est le premier gène d'une unité multicistronique qui en compte trois (voir l'annexe 2). Le second est *xprA* et le troisième est *recJ*. Immédiatement en amont de *xerD*, on retrouve un gène codant probablement pour la flavodoxine (FldB), mais il est transcrit en sens inverse (Hayes *et al.*, 1997). Le gène *xprA* a été renommé *dsbC* par la suite (Missiakas, *et al.*, 1994) et il code pour une oxydo-réductase périplasmique impliquée dans la formation de ponts disulfures de plusieurs autres protéines périplasmiques.

2.3 Rôles de XerC et XerD

Lorsque l'une des positions conservées de la famille des intégrases est supprimée ou mutée chez XerD ou XerC, un plasmide ayant deux sites *cer* n'est plus résolu. En fait, ces deux protéines sont catalytiquement impliquées dans la recombinaison spécifique de site (Blakely *et al.*, 1993). Par contre, les gènes cotranscrits avec *xerC* ou *xerD* ne semblent pas avoir de rôle dans ce genre d'action.

Les deux protéines se lient au site *dif* de façon individuelle. Cependant, si l'on compare le complexe formé de XerC (ou XerD) seule liée à *dif* au complexe formé lorsque les deux protéines sont mises ensemble avec un site *dif* marqué, la mobilité électrophorétique de ce dernier en est réduite. En fait, elles se lient au site de façon coopérative, tout comme c'est le cas au site *cer* (Blakely *et al.*, 1993) et donc il y a interaction entre ces deux protéines. De plus, la liaison de XerC est plus faible (Blakely et Sherratt, 1994). Alors que XerC se fixe à la portion gauche de la région centrale (région où se fait l'échange des brins), XerD se lie à la portion droite de cette région (Blakely *et al.*, 1993).

Comme on le constate, le système XerC/D est particulier en ce sens qu'il utilise deux recombinases au lieu d'une seule, comme c'est le cas pour les autres membres de la famille des intégrases. Dans ces systèmes, l'asymétrie de la région centrale sert à aligner les sites correctement, ce qui n'est pas le cas pour Xer. Il a été supposé que la

liaison précise des régions séparées par XerC et XerD pouvait procurer l'asymétrie recherchée (Sherratt *et al.*, 1993). Une autre raison avancée était que l'emploi de deux recombinases permettrait aux échanges des brins d'être sous des contrôles biochimique et génétique séparés, dans lesquels les recombinases catalyseraient chacune un échange de brin différent. C'est d'ailleurs ce qui a été révélé par la suite (voir les points 6.1.5 et 6.2).

La revue de littérature présentée dans ce mémoire fait voir l'importance de ces recombinases et ce qui est connu de leur mécanisme d'action au sein de la cellule bactérienne.

2.4 Activité topoisomérase

Les recombinases XerC et XerD, tout comme certaines autres intégrases, ont une activité topoisomérase de type I, qui est cependant beaucoup plus faible que l'activité de recombinaison. Cette fonction nécessite les résidus catalytiques et des sites de recombinaison ; les recombinases sont donc des topoisomérases spécifiques de site. Par contre, l'activité de relaxation ne requiert pas de synapse entre les sites d'action qui sont distants l'un de l'autre. De plus, l'activité de chacune des recombinases est inhibée par la recombinase partenaire, et ce, même si elle n'est pas catalytiquement active (Cornet *et al.*, 1997).

Cette activité avait également été retrouvée chez Cre de P1 (Abremski *et al.*, 1986) et chez l'Int de λ (Kikushi et Nash, 1979). Bien que dans le cas de l'Int de λ , l'activité ne semblait pas forcément être exécutée sur des sites spécifiques, il a été montré par la suite que c'était bel et bien une topoisomérase spécifique de site (Craig et Nash, 1983).

2.4.1 Les topoisomérases d'*E. coli*

« Les topoisomérases sont des enzymes qui agissent sur l'ADN pour modifier le niveau de surenroulement, aussi bien en "caténant" ou en "décaténant" les

chromosomes » (Luttinger, 1995). Elles catalysent le passage de brins simples d'ADN ou d'hélices doubles les uns au travers des autres (Wang, 1996). Il y a quatre topoisomérases connues chez *E. coli*, réparties en deux classes : les topoisomérases I et III sont de type I, tandis que les topoisomérases II (nommée gyrase) et IV sont de type II. Tandis que la gyrase introduit du surenroulement négatif, la topoisomérase I relaxe l'ADN qui est surenroulé négativement. La première classe a comme activité la création d'une coupure monocaténaire et le passage d'un autre morceau d'ADN simple brin à travers la brèche. Elle peut également passer un morceau d'ADN double brin, en autant que l'un des deux brins soit déjà brisé. Quant à la seconde classe, elle brise de façon transitoire un segment double brin pour faire le passage de brins doubles-hélices à travers ce bris (Luttinger, 1995).

Tout comme les intégrases, les topoisomérases d'*E. coli*, interagissent par l'intermédiaire d'une tyrosine pour effectuer les réactions de clivage et de religation (Wang, 1996). La réaction se fait en deux étapes de transestérification. Dans la 1^{re} étape, il y a liaison covalente de la tyrosine au phosphate venant d'être clivé à une extrémité de la cassure, tandis qu'il y a libération d'un groupement 5'hydroxyle de l'autre côté. La 2^e étape permet de rejoindre les deux brins en libérant l'enzyme (revue dans Wang, 1996).

Les topoisomérases sont impliquées dans plusieurs processus biologiques. Ainsi, elles jouent un rôle dans la réplication et dans la transcription, en prévenant un surenroulement excessif de l'ADN. Elles seraient également impliquées dans le maintien de la stabilité génomique et dans la ségrégation plasmidique (Wang, 1996).

2.4.2 Points communs aux topoisomérases et aux intégrases

Les topoisomérases de type IB -c'est-à-dire les topoisomérases de type I eucaryotes et celles codées par certains poxvirus comme le virus de la vaccine²-, et les

² Les topoisomérases d'*E. coli* sont de type IA. Cette famille englobe en fait plusieurs autres topoisomérases procaryotes et eucaryotes. Les deux familles (IA et IB) se distinguent dans la façon de procéder à l'étape du passage des brins et dans la spécificité de substrat (Wang, 1996).

recombinases spécifiques de site de la famille des intégrases, partagent certaines caractéristiques communes.

Tout récemment, il a été montré que le domaine C-terminal de la topoisomérase I du virus de la vaccine avait une structure similaire au domaine catalytique de la protéine Cre et d'autres intégrases (Cheng *et al.*, 1998). Cette similarité est par contre restreinte à la structure tertiaire, puisque la séquence en acides aminés est plutôt divergente. Le groupe de Shuman en a tiré l'hypothèse que les domaines catalytiques des deux types d'enzymes (recombinases spécifiques de site et topoisomérases de type IB) proviennent probablement d'une transférase ancestrale commune agissant au phosphate en 3' de la coupure du site d'action, et que la spécificité biologique en tant que telle serait due aux domaines N-terminaux qui divergent dans les deux familles d'enzymes. Il est intéressant de remarquer que le résidu impliqué dans le pouvoir catalytique des topoisomérases de type IB (la lysine 167 chez la topoisomérase du virus de la vaccine) est situé à la même position chez les intégrases dont la structure a été déterminée (XerD, Int de λ , Int de Hp1 et Cre de P1), c'est-à-dire dans la boucle reliant les feuillets $\beta 2$ et $\beta 3$ (voir la section 7). De plus, tout comme ce qui est retrouvé chez XerD, un changement conformationnel serait nécessaire pour permettre à la tyrosine catalytique de la topoisomérase du virus de la vaccine d'attaquer le phosphate susceptible d'être clivé ("scissile phosphate") (Cheng *et al.*, 1998).

En plus du point de vue structurel, la topoisomérase de la vaccine peut résoudre des jonctions de Holliday synthétiques *in vitro*, tout comme c'est le cas pour les intégrases (Sekiguchi *et al.*, 1996) et elle agit en des sites spécifiques. De plus, les étapes de la réaction impliquant la tyrosine ne nécessitent pas de cofacteurs dans les deux cas. Par contre, contrairement aux intégrases qui nécessitent quatre sites de liaison autour du point d'enjambement ("crossing-over"), la résolution peut se faire avec seulement deux sites de liaison. Il y a nécessité de quatre recombinases pour effectuer la réaction de recombinaison complète, tandis qu'une seule topoisomérase suffit pour la réaction de modification de la topologie de l'ADN (Sherratt et Wigley, 1998). Cet état multimérique des recombinases pourrait peut-être expliquer comment

une enzyme est déterminée à agir préférentiellement en tant que recombinase. Il existe en effet un domaine d'oligomérisation conservé chez les intégrases, ce qui n'est pas le cas chez les topoisomérases.

Enfin, une activité RNase est également retrouvée chez les deux sortes d'enzymes (du moins chez Flp pour ce qui est des intégrases), et elles peuvent utiliser d'autres résidus nucléophiles que la tyrosine (Sherratt et Wigley, 1998).

3. Mécanisme d'action des intégrases : on clive en cis ou en trans ?

Il existe un paradoxe en ce qui concerne la façon dont clivent les recombinases de la famille des intégrases : le paradoxe *cis-trans* (Jayaram, 1997). Comme mentionné précédemment, quatre protomères sont nécessaires pour effectuer l'échange des brins. La question est alors de savoir lequel de ces protomères effectue le clivage du lien phosphodiester activé. Quatre possibilités s'offrent ainsi :

- Clivage en *cis* : le protomère clive au site d'action où il est lié. Il faut donc, pour cette forme de clivage, que le protomère ait en lui-même un site actif complet.
- Clivage *trans*-horizontal : le protomère situé de l'autre côté de la région d'espacement effectue le clivage, sur le même brin d'ADN.
- Clivage *trans*-vertical : c'est le protomère de la paire de brins partenaire, situé du même côté que le protomère situé au site de clivage, qui exerce son action à ce site.
- Clivage *trans*-diagonal : le clivage se fait par le protomère situé sur l'autre paire de brins, en position opposée par rapport au protomère situé le plus près du site de clivage.

La figure 2 de la page suivante schématise les diverses possibilités.

Il a même été suggéré que le clivage pouvait être asymétrique et se faire en dualité avec le passage d'un clivage *trans*-horizontal à *trans*-vertical dans le cas de la protéine Flp (Qian et Cox, 1995), et même d'un clivage en *cis* à un clivage en *trans*

selon l'étape de la réaction comme cela pourrait peut-être être le cas de l'Int de λ (Stark et Boocock, 1995 et Jayaram et Lee, 1995).

Pour la protéine Flp provenant de *Saccharomyces cerevisiae*, il a été suggéré qu'il devait y avoir un site actif partagé : la tyrosine catalytique proviendrait d'un protomère, tandis que les trois autres résidus constituant le site actif et probablement impliqués dans l'activation des liens phosphodiester, seraient fournis

par une seconde sous-unité (revue dans Jayaram, 1994). En fait, le groupe de Jayaram a montré que le résidu tyrosine est procuré par le protomère situé sur le même

site où se fait la réaction de clivage, mais à l'extrémité opposée du pont de clivage. Il s'agirait donc d'une réaction *trans*-horizontale (Lee *et al.*, 1994). Cependant, un nouveau modèle fut suggéré (asymétrique cette fois-ci) lorsque les résultats d'un autre groupe montrèrent que trois protomères de Flp étaient suffisants pour les réactions de clivage et d'échange. Il y aurait ainsi possibilité de la réaction *trans*-horizontale suivie de la réaction *trans*-verticale mentionnée plus haut (Qian et Cox, 1995).

Dans tous les cas donc, les expériences effectuées avec Flp s'accordent pour montrer un clivage en *trans*. D'un autre côté, les études faites avec l'intégrase du phage λ ont présenté des résultats contradictoires. Dans une étude, la sous-unité située en *cis* du site semblait procurer les deux arginines conservées, tandis que la tyrosine semblait fournie par le monomère en *trans* (Han *et al.*, 1993). Par contre, en étudiant les interactions protéiques entre l'intégrase du phage λ et l'intégrase du phage lambdaïdoïde HK022, lesquelles se distinguent seulement par la reconnaissance de sites centraux différents, le résidu catalytique semblait plutôt fourni par le protomère situé

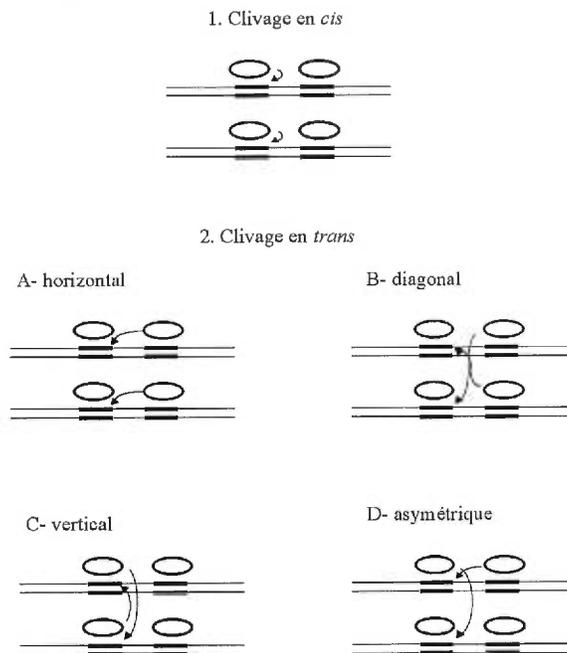


Figure 2- Schémas des diverses possibilités de forme de clivage s'offrant aux intégrases. Les lignes foncées représentent les sites de liaison des protéines. Les flèches indiquent les attaques nucléophiles par la tyrosine catalytique de chaque recombinaison. (Figure tirée de Qian et Cox, 1995).

en *cis* du site de clivage (Nunes-Düby *et al.*, 1994). La structure du domaine catalytique de l'Int de λ n'a pas pu résoudre le débat, puisque les deux modes de clivage peuvent convenir selon la configuration adoptée (Kwon *et al.*, 1997).

Les expériences faites avec le système Xer étaient facilitées puisque deux enzymes différentes sont employées. Elles montrent avec un peu plus de certitude une activité en *cis* (le modèle *trans*-vertical ne peut par contre être totalement éliminé) (Arciszewska et Sherratt, 1995). Comme il a été montré que XerC clivait le brin supérieur d'un substrat ayant une jonction de Holliday *in vitro*, et que XerD, malgré sa liaison au site d'action (ici, le site *cer*), n'exerce pas d'activité catalytique, les modèles *trans*-horizontal et *trans*-diagonal sont éliminés immédiatement. De plus, comme un XerCR243Q (muté dans le site arginine conservé) n'a pu compléter un XerCY275F (n'ayant pas le résidu nucléophile), il semble plus probable que le mécanisme s'effectue en *cis*. D'ailleurs, la structure de XerD confirme ce type de clivage: la tyrosine catalytique se situe près des autres résidus conservés (R-H-R) (Subramanya *et al.*, 1997).

Le modèle de clivage en *cis* s'avère également plus probable pour l'Int de HP1, selon ce qui est révélé par la structure du domaine C-terminal (Hickman *et al.*, 1997).

En supposant que l'ensemble des intégrases procaryotes avait tendance à effectuer un clivage en *cis*, il avait été proposé que le choix du mode de clivage serait déterminé par l'espace retrouvé entre la tyrosine et les autres résidus conservés. La grandeur de cet espace est différent selon que l'intégrase est procaryote ou eucaryote (Blakely et Sherratt, 1996a). Cette dernière hypothèse ne peut plus s'avérer fondée puisqu'il semblerait que Cre clive en *trans* (Shaikh et Sadowski, 1997).

Cependant, la structure de Cre laisse supposer un clivage en *cis*, mais il est malgré tout possible que l'hélice α M (voir fig. 11) soit fournie en *trans* (Guo *et al.*, 1997). Les auteurs suggèrent alors la possibilité d'un mécanisme similaire au mode asymétrique discuté dans le cas de Flp.

4. Sites d'action

Pour exercer leur action sur les réplicons (plasmides ou chromosomes), les recombinaisons de type Xer doivent agir en des sites précis retrouvés sur ces réplicons.

Tous les sites retrouvés jusqu'à maintenant sont organisés de la même façon. La région centrale, suffisante pour la recombinaison, est constituée de 28 à 30 pb. Deux régions de 11 pb flanquent la région d'espacement où a lieu l'échange des brins, laquelle varie de 6 à 8 pb. La région à gauche du site d'enjambement constitue le site de liaison à XerC, tandis que la région à droite est le site de liaison de XerD (Blakely *et al.*, 1993). La figure 3 montre l'alignement de ces différents sites. Comme on peut le constater, le site de liaison à XerD est beaucoup plus conservé que le site de liaison à XerC. Rappelons également que la liaison des deux recombinaisons à l'ADN est coopérative (Blakely *et al.*, 1993).

4.1 Sites *dif*, *cer* et *psi*

4.1.1 Site *dif*

a) Caractéristiques du site :

Le site *dif* (pour **d**ele**t**ion-induced **f**ilamentation) est situé à 1588,8 kb dans la région de terminaison de la réplication sur le chromosome d'*E. coli*, entre les sites d'arrêt de la réplication *TerA* et *TerC* (Kuempel *et al.*, 1991 et numéro d'accèsion de GenBank: U00096). Il est adjacent à l'opéron *hip* (**h**igh **p**ersistence locus; Black *et al.*, 1991).

La délétion de ce site induit les mêmes phénotypes que les mutations dans *xerC* (Blakely *et al.* 1991) et/ou *xerD* (Blakely *et al.* 1993). Il y a alors filamentation d'une certaine proportion des cellules de la population bactérienne, présence de structures nucléoïdes aberrantes et induction de la réponse SOS. Il y a également un ralentissement du taux de croissance et une réduction du nombre d'unités viables (Blakely *et al.*, 1991, 1993). La filamentation n'est par contre pas observée dans une

	XerC	Espacement	XerD
ColoDF13 (<i>parB</i>)	GGTACCGATAA	..GGGATG	TTATGGTAAAT
ColA (<i>car</i>)	GGTGCGTACAA	..CGGATG	TTATGGTAAAT
ColE1 (<i>cer</i>)	GGTGCGTACAA	TTAAGGGA	TTATGGTAAAT
ColE2	GGGGCGTACAA	..CGGGAG	TTATGGTAAAT
ColE2-CA42	GGTACGTACAA	..CGGGAG	TTATGGTAAAT
ColE2-GEI602	GGGGCGTACAA	..CGGGCG	TTATGGTAAAT
ColE2imm-K317	GGTGCGTTTAA	..CGGGAG	TTATGGTAAAT
ColE3 (<i>cer3</i>)	GGTGCGTACAA	..CGGGAG	TTATGGTAAAT
ColE4-CT9	GGTGCGTACAA	..CGGGAA	TTATGGTAAAT
ColE5-099	GGTACGTACAA	..CGGGAG	TTATGGTAAAT
ColE6-CT14	GGTGCGTACAA	..CGGGAG	TTATGGTAAAT
ColE7-K317	GGTGCGTACAA	..CGGGAG	TTATGGTAAAT
ColE8-J	GGTACGTACAA	..CGGGAA	TTATGGTAAAT
ColE9-J	GGTACGTACAA	..CGGGAG	TTATGGTAAAT
ColK (<i>ckr</i>)	GGTGCGTACAA	TTAAGGGA	TTATGGTAAAT
ColN	GGTGCGTACAA	.TAAGGGA	TTATGGTAAAT
pHS-2 (<i>hmr</i>)	GGTGCGCGCAA	.TAAAATG	TTATGTTAAAT
plasmide R1	GGTGCGCATAA	..TGTATA	TTATGTTAAAT
pMB1	GGTGCGTACAA	TTAAGGGA	TTATGGTAAAT
pNTP16 (<i>nmr</i>)	GGTGCGCGTAA	.TGAGACG	TTATGGTAAAT
pSC101 (<i>psi</i>)	GGTGCGCGCAA	..GATCCA	TTATGTTAAAC
pPvu1	GGTTCCGATAA	..GACCCA	TTATGTTAAAC
<i>E. coli</i> (<i>dif</i>)	GGTGCGCATAA	..TGTATA	TTATGTTAAAT
<i>X. campestris</i> *	¹⁴ ATTTCGCATAA ⁻⁴	..TGTATA ⁻⁷⁺¹	⁺⁴ TTATGTCAAAT ⁻¹⁴
Consensus	<u>GT</u> T A GGAT GC TT AA AGG <u>T</u> CCGC		<u>GT</u> T TTATG AAA TC C

Figure 3- Alignement de plusieurs sites d'action ayant une homologie avec le site *cer*. Cette figure est une modification de l'annexe 2, provenant du mémoire de Stéphane Sirois (Détection des recombinases XerC et XerD chez les *Enterobacteriaceae*, Département de microbiologie et immunologie, Université de Montréal, 1994). La séquence du plasmide R1 est celle indiquée au numéro d'accèsion X57116, locus ECSPSSR ; la séquence parue dans Clerget *et al.* (1991) diffère de deux bases. Les chiffres indiqués sous le site *dif* correspondent aux positions des nucléotides telles que définies par Hayes et Sherratt (1997). * : Chez *Xanthomonas campestris* pv. *citri*, ce site correspond à un site d'attachement pour le phage cfl6-v1, mais il pourrait également être lié par le système Xer (Clerget *et al.*, 1991). Enfin, les bases du consensus en caractères gras indiquent un consensus absolu, tandis que les bases soulignées sont conservées dans tous les sites, à l'exception de celui retrouvé chez *X. campestris*.

souche ayant une mutation dans *recA* ou dans *recBC*. Même si cela pouvait donner à penser que la réponse SOS pouvait induire la filamentation, ce n'en était pas la cause première (Kuempel *et al.*, 1991). En fait, l'échange des chromatides-soeurs par la recombinaison homologue (formant ainsi des dimères chromosomiques) causerait la filamentation.

Le site *dif* contient une séquence de 33 pb qui est similaire à celle du site *cer* que l'on retrouve sur le plasmide ColE1. Lorsque présente dans le vecteur pUC18, cette région s'avère un substrat pour la multimérisation et la résolution *in vivo* (Blakely *et al.*, 1991), et XerC et XerD peuvent s'y lier *in vitro* (Blakely *et al.*, 1993). Ce 33 pb peut subséquentement être réduit à 28 pb qui constituent en fait le cœur même du site (Leslie et Sherratt, 1995).

b) Rôles du site *dif* dans le chromosome bactérien

La région de 33 pb, lorsqu'insérée dans une souche dont la région entière de *dif* est supprimée, permet l'abolition du phénotype de filamentation (Tecklenburg *et al.*, 1995 et Sherratt *et al.*, 1995), et ce, peu importe son orientation (Leslie et Sherratt, 1995). Ainsi, les régions flanquantes de cette région ne semblent pas nécessaires pour la fonction de *dif*. Toutefois, une région de 231 pb (qui inclut des séquences adjacentes aux 33 pb) donne un taux plus élevé de recombinaison (Kuempel *et al.*, 1996). De plus, ce site fonctionne même lorsqu'il n'a que deux sites de liaison à XerD, mais il nécessite malgré tout la présence de XerC (Hayes et Sherratt, 1997).

Son emplacement dans le chromosome semble cependant important, puisque le phénotype Dif⁺ apparaît si *dif* est localisé ailleurs (Leslie et Sherratt, 1995 et Tecklenburg *et al.*, 1995). Il y a en fait une zone précise dans laquelle *dif* est actif : DAZ (pour *dif* activity zone) (Cornet *et al.*, 1996). Les résultats obtenus par Cornet *et al.* (1996) semblent montrer que la terminaison de la réplication ne serait pas nécessaire pour l'activité de *dif*, puisque la délétion de *TerC* (situé près de *dif* sur le chromosome) et de *tus* (gène codant pour une protéine qui arrête la progression de la fourche de réplication; Khatri *et al.*, 1989) ne montrent pas de phénotype cellulaire

particulier. Il semblerait que ce soit plutôt le processus de la recombinaison spécifique de site en tant que tel qui serait impliqué dans la ségrégation normale du chromosome. En effet, *dif* peut être remplacé par *loxP*, (en autant que la protéine Cre soit fournie), et ce, sans que le phénotype des cellules Dif n'apparaissent (Leslie et Sherratt, 1995). Ceci s'avère également vrai avec le remplacement de *dif* par *psi* (Cornet *et al.*, 1994). Par contre, ce remplacement n'a pu se faire avec les sites *res* (site pour la résolvasse Tn3), *cer* ou *cer6* (site *cer* avec deux paires de bases en moins) (Leslie et Sherratt, 1995).

On croyait que la recombinaison de type Xer au site *dif* devait servir à contrer la dimérisation des chromosomes due à la recombinaison homologue survenue pendant ou après la réplication; elle assurerait une forme monomérique avant la ségrégation (Sherratt *et al.*, 1993). Cependant, un vecteur ayant un site *dif* ne tend pas nécessairement à rester monomère: il n'y a pas de sélectivité pour la recombinaison intramoléculaire (résolution plasmidique) (Blakely *et al.*, 1991). De plus, ce site n'est pas encadré de séquences accessoires impliquées dans la recombinaison comme c'est le cas de *cer* (voir le point 4.1.2). Ainsi, pour assurer la monomérisation des dimères chromosomiques probablement formés par l'échange des chromatides-soeurs à la fin de la réplication, le sens de la réaction devrait être assuré par un aspect quelconque de la ségrégation du chromosome lui-même (Tecklenburg *et al.*, 1995). Par un essai de marquage par densité, il a été montré récemment que si la formation du septum était bloquée, que ce soit lors d'une étape initiale de la formation de l'anneau septal, ou à un stade plus éloigné de croissance du septum, la résolution des chromatides-soeurs ne survient pas (Steiner et Kuempel, 1998). Ainsi, la division cellulaire serait impliquée dans la recombinaison au site *dif*.

Un site similaire à *dif* serait aussi retrouvé dans la région terminale de la réplication de *Haemophilus influenzae* et il serait un substrat pour le système Xer (Hayes et Sherratt, 1997). De plus, un site situé sur le chromosome de *Xanthomonas campestris* pv. *citri* s'avère homologue à *dif* (Dai *et al.*, 1988). Ce site sert à l'attachement de l'ADN du phage filamenteux Cf16-v1, qui possède un site similaire,

comme c'est le cas des sites *attB* et *attP* qui servent à l'intégration du phage λ dans le chromosome d'*E. coli*.

4.1.2 Site *cer*

On peut pratiquement affirmer que l'histoire de la recombinaison spécifique de site de type Xer a débuté avec la découverte du site *cer*. En 1984, Summers et Sherratt publiaient un article montrant que la stabilité plasmidique de ColE1 était due essentiellement à la présence d'un déterminant essentiel pour la conversion des multimères en monomères (Summers and Sherratt, 1984). Ce site a été désigné *cer* pour ColE1 resolution.

Contrairement à ce qui est observé avec ColE1, plusieurs dérivés de ce plasmide ont tendance à multimériser dans les souches d'*E. coli* RecA⁺ (pouvant effectuer la recombinaison homologue). Ils finissent par être rapidement perdus des souches hôtes, parce qu'ils ne contiennent pas le site *cer*. Ils sont par contre stables dans les souches qui sont *rec*⁻ (Sherratt *et al.*, 1986).

Le site *cer* fait environ 250 pb. La frontière à gauche de la région d'enjambement a été localisée entre les paires de bases 3733 et 3737 de ColE1. La région de crossing-over s'avère être la plus conservée parmi les sites similaires retrouvés chez d'autres plasmides (Summers et Sherratt, 1988). Ce site montre de la sélectivité pour la recombinaison intramoléculaire et il nécessite la présence des séquences retrouvées en amont de la région centrale, ainsi que des protéines accessoires: ArgR et PepA (Stirling *et al.*, 1988a, 1988b, 1989).

4.1.3 Site *psi*

Un autre site a également été caractérisé dans le plasmide pSC101, le site *psi* (pSC101 stabilized inheritance). Il est localisé entre les paires de bases 6783 et 6810, après le gène *repA* servant à la répllication du plasmide. Les séquences flanquantes de *psi* sont différentes de celles entourant le site *cer* dans ColE1. Selon le contexte

génétique dans lequel se trouve *psi*, la recombinaison peut être contrainte ou non, (nécessiter ou non des facteurs accessoires). Contrairement à *cer*, *psi* peut supprimer les effets d'une délétion de *dif* dans le chromosome d'*E. coli*, et, alors, il n'y a pas de nécessité de séquences accessoires. Cependant, quand ce site est cloné dans un autre vecteur, la présence de ces séquences s'avère essentielle (Cornet *et al.*, 1994).

4.2 Principales différences entre les sites *dif*, *cer* et *psi*

La principale différence entre les sites *cer*, *psi* et *dif*, vient du fait que les deux premiers nécessitent des facteurs accessoires. Il y a donc une sélectivité pour la résolution. Le site *dif* par contre, peut aboutir aussi bien à une fusion de deux molécules, qu'à une résolution. Par contre, bien que PepA soit nécessaire dans les deux premiers cas (Stirling *et al.*, 1988a, Cornet *et al.*, 1994), la seconde protéine accessoire est différente pour chacun. La recombinaison au site *cer* a besoin d'ArgR (Stirling *et al.*, 1988b), tandis que la recombinaison à *psi* nécessite ArcA/ArcB (Colloms *et al.*, 1998).

Il y a également une différence au niveau des produits finaux obtenus *in vitro*. La recombinaison au site *psi* donne lieu à des caténanes, tandis que la recombinaison au site *cer* s'arrête à la jonction de Holliday (voir la section 5). Pour ces cas, c'est la séquence de la région centrale du site qui détermine le produit obtenu, tandis que les séquences accessoires déterminent l'utilisation d'ArgR (Colloms *et al.*, 1996).

4.3 Détermination de la sélectivité des sites

Comme le site *cer* a une région centrale de 8 pb, tandis que la région centrale de *dif* n'a que 6 pb, on aurait pu croire que la longueur de cette région pouvait expliquer en partie le fait que certains sites nécessitent des facteurs accessoires pour orienter la réaction de recombinaison vers une résolution.

En fait, par la construction d'un site *dif* ayant 8 pb, Blakely et Sherratt (1996b) ont pu démontrer que la sélectivité pour la résolution et la nécessité des protéines et

séquences accessoires sont liées à l'efficacité du clivage et au degré de courbure de l'ADN. La géométrie du complexe recombinases/ADN est changée lorsque le site contient ces 8 pb dans sa région centrale (au lieu de 6 pb), et ainsi, XerC clive moins bien et la coopérativité entre XerC et XerD diminue, bien que l'affinité individuelle des recombinases pour leur site ne soit pas affectée. La courbure de l'ADN est, quant à elle, plus grande.

Cependant, le site *psi* fait 6 pb et montre malgré tout une préférence pour la recombinaison intramoléculaire, lorsqu'il y a présence de facteurs accessoires (Cornet *et al.*, 1994); il en est de même pour un autre site dénommé *parB* (voir le point 4.4.). Il y aurait donc «autre chose» qui déterminerait le comportement de ces sites.

Lorsque la recombinaison est effectuée entre *parB* et *cer*, les deux hybrides obtenus ont des comportements complètement différents. L'hybride de type I montre de la sélectivité comme les parents, tandis que l'hybride de type II agit plutôt comme un site *dif*, puisque la recombinaison donne lieu à des multimères. Les deux sites ne diffèrent seulement que par deux nucléotides, l'hybride de type II ayant deux T de moins dans la région d'enjambement (Summers, 1989). Cet hybride contient la région centrale de *parB* et le site de liaison à XerC de *cer*. Le site de liaison de XerD est le même pour les deux sites d'origine (voir la fig. 3). Ainsi, la détermination pour la sélectivité du site *parB* devait résider dans son site de liaison à XerC, et il ne se différencie de celui du site *cer* que par quatre nucléotides.

Des hybrides entre les sites *psi* et *dif* ont par la suite été construits. Ces deux sites ne se distinguent que par deux nucléotides dans leur site de liaison à XerC et par un nucléotide dans le site de liaison à XerD, comme on peut le constater avec l'alignement des sites de recombinaison. Ces constructions ont permis de constater que la sélectivité pour la résolution est influencée par la séquence des sites de liaison des recombinases (Blake *et al.*, 1997). La recombinaison peut d'ailleurs être modifiée par le changement d'un seul nucléotide. Tout comme ce qui avait été montré précédemment, la géométrie du complexe ADN-recombinase est changée. Ainsi, et la

longueur et la séquence de la région centrale influencent la façon dont se produit la réaction de recombinaison.

Un autre niveau de sélectivité dans les sites se situe dans la reconnaissance spécifique par XerC et XerD des deux séquences qui flanquent la région centrale des sites, où se fait l'échange des brins. Ces séquences sont pratiquement palindromiques et malgré tout, la liaison des recombinases est spécifique. Cette spécificité serait reliée aux terminaisons externes de ces séquences, du moins pour le site *dif*. En fait, le nucléotide -10 s'avère le plus important pour la spécificité de la liaison de XerC à son site, tandis que le nucléotide en position +9 serait important pour la liaison spécifique de XerD (les positions sont indiqués sur la fig. 3) (Hayes et Sherratt, 1997). La partie interne des sites de liaison, où la symétrie est plus grande, servirait à l'interaction avec une structure protéique commune aux deux recombinases, comme par exemple, une région du domaine C-terminal qui est mieux conservé chez les deux recombinases (Blakely et Sherratt, 1994).

4.4 Autres sites

D'autres sites plasmidiques se sont révélés similaires aux sites *cer*, *dif* et *psi* précédemment mentionnés. Ainsi, le site *parB* de CloDF13 contient un site pour la résolution du plasmide à 4773 pb du plasmide (Hakaart *et al.*, 1984 et Nijkamp *et al.*, 1986). Une région *par* a également été identifiée chez ColN (Kolot, 1990) et ArgR s'est avérée essentielle pour la résolution des multimères. ColK présente un site *ckr* (Col K resolution) qui a, non seulement la même organisation génomique que le site *cer*, mais également la même région centrale pour la recombinaison. Le site *ckr* est situé entre les gènes pour la mobilisation du plasmide et pour la production de colicine (Summers *et al.*, 1985). ColA, qui origine de *Citrobacter freundii* CA31, a un site *car* dont la région d'échange des brins est de 6 pb, comme chez *parB* de CloDF13 (Morlon *et al.*, 1988). Entre les régions susceptibles d'être impliquées dans la stabilité des réplicons ColE2 et ColE3, il n'y a qu'une différence d'une paire de bases au niveau du site de liaison à XerC, et là encore, la région centrale contient 6 pb (Yasueda *et al.*, 1989). Cette région se retrouve d'ailleurs chez plusieurs autres

plasmides de la famille de ColE2 (Hiraga *et al.*, 1994). Une forte région d'homologie avec *dif*, *parB*, *cer* et le système d'intégration du phage Cf16-v1-*Xanthomonas campestris*, se trouve également chez le plasmide R1 (Clerget, 1991). Enfin, les plasmides pNTP16, provenant de *Salmonella typhimurium*, pPvuI de *Proteus vulgaris*, ainsi que pHS-2 de *Shigella flexneri* semblent également posséder une région similaire aux sites déjà mentionnés (Cannon et Strike, 1992, Calvin/Koons et Blumenthal, 1995, Réhel et Szatmari, 1996). D'ailleurs, la recombinaison spécifique de site a été réalisée entre deux sites *nmr* (NTP16 multimer resolution) séparés de 6 pb sur un même plasmide, et même, la recombinaison a pu être faite entre un site *cer* et un site *nmr* (Zakova et Szatmari, 1995). Des tests de recombinaison ont également été effectués avec deux sites *hmr* (pHS-2 multimer resolution) (Réhel et Szatmari, 1996), mais ils restent à faire pour le site similaire de pPvuI.

5. Facteurs accessoires

Il avait été présumé que la recombinase spécifique agissant à *cer* serait encodée par l'hôte, puisque le « potentiel codant » de *cer* était trop limité. C'est pourquoi le(s) locus (loci) susceptibles d'encoder la recombinase ont été désignés *xer* (**chromosomal ColE1 recombination function(s)**). Deux groupes de complémentation ont alors été révélés : *xerA* et *xerB*, lesquels étaient bel et bien situés sur le chromosome de l'hôte (Stirling *et al.*, 1988a). C'était une situation nouvelle compte tenu du fait que toutes les recombinases caractérisées jusqu'alors avaient les sites de recombinaisons adjacents aux gènes codant pour les protéines recombinantes. Ainsi, même si les protéines codées par ces gènes devaient se révéler accessoires par la suite, elles étaient tout de même absolument nécessaires pour la recombinaison au site *cer* (Stirling *et al.*, 1988a).

5.1 ArgR et PepA

5.1.1 XerA= ArgR

Le gène, initialement dénommé *xerA*, est identique à *argR*. Il est localisé à 3382,3 kb du chromosome d'*E. coli* (numéro d'accèsion de GenBanK; U00096). De plus, le répresseur de l'arginine de *Bacillus subtilis* peut compléter une souche d'*E. coli* mutante pour *xerA*, et ce, malgré que les deux protéines ne présentent que 27% d'identité (Stirling *et al.*, 1988b).

ArgR, codée par *argR*, est une protéine impliquée dans la biosynthèse de l'arginine. On la retrouve sous forme hexamérique et le poids moléculaire d'un monomère est de 16 500 Da. Le domaine C-terminal, dont la structure a été déterminée, est impliqué dans l'hexamérisation, ainsi que dans la liaison de la L-arginine (revue dans Maas, 1994). Cet acide aminé est essentiel au bon fonctionnement du répresseur.

ArgR interagit avec deux boîtes ARG adjacentes, qui sont en palindromes (Lim *et al.*, 1987). Ces boîtes ARG ont 18 pb arrangées de telle sorte que 9 pb sont symétriques aux 9 pb restantes. Les deux boîtes, retrouvées dans la région promotrice des gènes impliqués dans la biosynthèse de l'arginine, sont séparées par 2-3 pb. Une boîte ARG est également retrouvée dans le site *cer* à environ 100 pb de la région d'échange des brins et XerA (ArgR) s'y lie. Ainsi, Stirling *et al.* (1988b) ont supposé qu'ArgR avait peut-être un lien avec l'organisation du complexe synaptique pour permettre la recombinaison intramoléculaire, puisqu'on ne pouvait alors plus supposer qu'ArgR avait le rôle de recombinaison au site *cer*.

Comme mentionné plus haut, ArgR est dépendante de la L-arginine. Il avait été montré que l'on pouvait muter une région impliquée dans la liaison à la L-Arg sans que la capacité de réprimer l'expression des gènes impliqués dans la biosynthèse de la L-arginine ou que la recombinaison au site *cer* ne semblent affectées (Burke *et al.*, 1994). Cependant, en absence de L-Arg, ArgR a moins d'affinité pour son site et sa

liaison se fait sous forme de trimère. Cet acide aminé stabilise en fait les hexamères en contrôlant leur éventuelle conversion en trimères, et il est nécessaire à la liaison des hexamères d'ArgR (sans mutation) à l'ADN (Chen *et al.*, 1997).

5.1.2 XerB=PepA

xerB, situé à 4482,0 kb sur le chromosome d'*E. coli* (numéro d'accèsion de GenBank: U00096), près des gènes *pyrB* et *argI*, s'est révélé l'homologue de *pepA* de *Salmonella typhimurium* et d'*E. coli* : l'activité aminopeptidase a d'ailleurs été montrée (Stirling *et al.*, 1989). Pour cette activité, l'enzyme encodée par ce gène nécessite alors du Mn^{2+} , mais est inhibée par le Zn^{2+} et l'EDTA ; elle est stable à la chaleur. Cette activité aminopeptidase n'est par contre pas nécessaire pour la recombinaison spécifique de site, puisque lorsque PepA est mutée pour éliminer ce type d'action, la souche demeure Xer⁺ ; la réaction *in vitro* n'a alors pas besoin de Mn^{2+} (McCulloch *et al.*, 1994a). Le polypeptide fait 503 acides aminés, et a un poids moléculaire de 55,3 kDa. De plus, la position relative du gène sur le chromosome d'*E. coli* (à 4560 kb) est la même que celle de l'aminopeptidase A de *S. typhimurium*. En fait, c'est le pourcentage d'identité de 52 entre la portion C-terminale de XerB et la portion C-terminale de l'aminopeptidase de leucine de lentille bovine qui a mené à la recherche d'aminopeptidases similaires chez les bactéries entériques et qui a révélé l'identité entre XerB et PepA d'*E. coli* (Stirling *et al.*, 1989).

Le rôle de PepA dans la recombinaison au site *cer* était alors difficile à définir. Il n'y avait aucune évidence que PepA pouvait lier l'ADN. Cependant, il a été montré par la suite que le produit encodé par *carP* est identique à PepA et que CarP lie la région promotrice de l'opéron *carAB* et la région contrôle de *carP* (Charlier *et al.*, 1995). L'opéron *carAB* code pour les deux sous-unités (CarA et CarB) de la carbamoylphosphate synthétase qui sert à la formation du carbamoylphosphate requis pour la synthèse de l'arginine et des résidus pyrimidines. PepA est donc impliquée dans la régulation de la biosynthèse des pyrimidines, en se liant au promoteur P1 de l'opéron *carAB*. Elle est également impliquée dans la régulation de sa propre synthèse puisqu'elle se lie à une région contenant un segment en amont du promoteur P1 et le

P1. (En fait, la transcription de *PepA* est initiée à trois sites : P1, P2 et P3, lesquels sont situés à 150, 161 et 188 nucléotides en amont du codon d'initiation, respectivement.)

Ainsi, tout comme *ArgR*³, *PepA* serait impliquée dans deux processus cellulaires différents : 1) la recombinaison spécifique de site et, 2) la régulation de l'expression génique.

5.1.3 Rôle des deux protéines dans la recombinaison au site *cer*

Comme déjà mentionné aux points 4.2 et 4.3, la recombinaison à *cer* nécessite absolument la présence de ces deux protéines et cette nécessité est également rencontrée avec le site *parB* (Summers, 1989). Ils ont tous deux une spécificité topologique, ce qui n'est pas retrouvé avec le site *dif* et les hybrides dits de type II (qui proviennent de la recombinaison entre *cer* et *parB*. Voir le point 4.3). Une nouvelle classe de sites a été déterminée lors de l'analyse du comportement de *cer3* (site de *ColE3*), et de sites provenant de la recombinaison entre un site *cer* et *cer3* et entre *cer* et *car* (de *ColA*) (Guhathakurta *et al.*, 1996). Ces sites sont contraints selon les conditions dans lesquelles se fait la recombinaison. Ainsi, ils ne sont pas contraints s'il y a absence de protéines accessoires. Ces sites ont donc permis de déterminer que les facteurs accessoires ont un double rôle à *cer* : -ils bloquent la recombinaison entre les sites en *trans* (recombinaison intermoléculaire), et -ils activent les sites pour la recombinaison en *cis* (recombinaison intramoléculaire).

Colloms *et al.* (1997), suite à leurs résultats, ont postulé que la recombinaison ne pouvait survenir que dans un complexe nucléoprotéique défini, dans lequel les séquences accessoires de deux sites participants

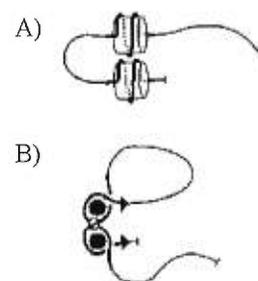


Figure 4: Formes possibles de l'enroulement autour d'*ArgR* et de *PepA*: A): Surenroulement solénoïdal, B): Enroulement plectonémique. (Figure tirée de *Alén et al.*, 1997).

³ *ArgR* est également impliquée dans la régulation de l'opéron *carAB*, puisqu'elle se lie au promoteur P2 impliqué dans la biosynthèse de l'arginine (Piette *et al.*, 1984).

s'enroulent de façon plectonémique, dans un sens tourné vers la droite, autour des protéines accessoires, en attrapant trois supertours négatifs. L'enroulement plectonémique signifie que deux segments d'hélice d'ADN sont enroulés en se croisant (voir fig. 4). La topologie est importante pour faire la différence entre des réactions de fusion et de résolution, puisque la recombinaison de type Xer survient entre des sites identiques.

Àlen *et al.* (1997) ont ensuite montré que PepA pouvait se lier spécifiquement au site *cer*. Elle agit en coopération avec ArgR pour former un complexe nucléoprotéique contraint. Toutes deux imposent un changement dans le nombre de supertours négatifs sur des substrats ayant deux sites *cer* directement répétés. L'enroulement autour des deux protéines s'est vraiment avéré être de type plectonémique, et non pas solénoïdal (forme que l'on retrouve avec l'ADN dans les nucléosomes) (fig. 4). Finalement, ils ont vérifié par un test de protection à la DNase I qu'ArgR se liait bien à une région de 18 pb, tandis que le patron de protection avec les

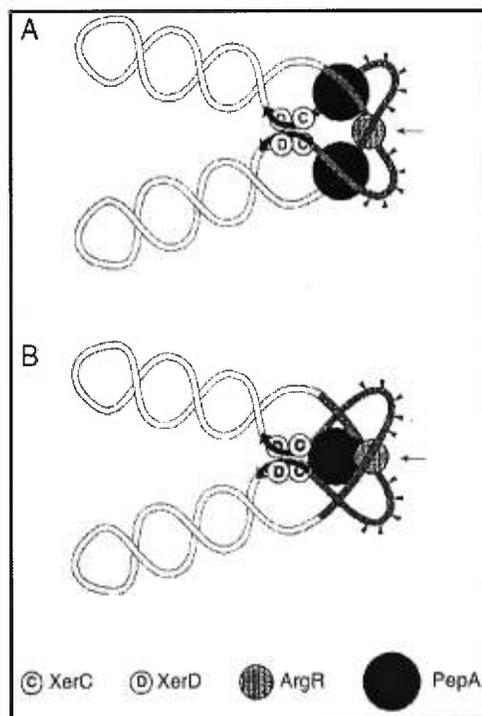


Figure 5- Modèles pour le complexe synaptique entre deux sites *cer* avec la présence d'ArgR et, A) deux hexamères de PepA ou B) un hexamère de PepA. Les triangles indiquent des régions plus susceptibles à la DNase I. (Figure tirée de Àlen *et al.*, 1997).

deux protéines s'étendait sur 180 pb de séquences accessoires. Le même patron est observé avec PepA employée seule, quoiqu'il soit moins fort.

De ces résultats, deux modèles ont été proposés pour la formation du complexe synaptique aux sites *cer* avec les protéines accessoires. Il avait déjà été montré qu'ArgR induisait une courbure de l'ADN de 65° quand elle était liée à un seul site *cer* (et donc, sur une seule boîte ARG) (Burke *et al.*, 1994). Les modèles présentés (voir fig. 5) ne diffèrent que par le nombre de molécules PepA impliquées ;

dans les deux cas, un hexamère d'ArgR attrape un croisement entre les deux sites *cer*, en se liant à une boîte ARG de chaque site. Dans le 1^{er}

modèle, deux hexamères de PepA dont chacun se lie à un segment d'ADN de chaque site et stabilise un croisement entre les deux sites, tandis que le 2^e montre un seul hexamère faisant contact avec quatre segments d'ADN et stabilisant alors deux croisements entre les deux sites. Ces modèles laissent également une soixantaine de paires de bases entre les deux boîtes ARG et les sites de liaison à XerC/D, ce qui correspond à ce qui avait été constaté lors de leurs tests de protection contre la DNase I ; les cinq régions plus susceptibles observées sont séparées par 10-11 pb les unes des autres.

Finalement, pour expliquer la nécessité absolue de ces protéines pour la recombinaison au site *cer*, les auteurs supposent que des contacts directs avec XerC/D seraient effectués pour activer la réaction d'échange, ou encore, que ces protéines favoriseraient un alignement des sites ou une conformation qui ferait en sorte de, soit lier les recombinases à leur site d'action, soit activer le premier échange (Àlen *et al.*, 1997).

5.2 ArcA/ArcB

Le site *psi* ne contient pas de boîte ARG (Cornet *et al.*, 1994). ArgR n'est donc pas nécessaire pour la recombinaison à ce site. Seule PepA était essentielle (Colloms *et al.*, 1996). Cependant, il a été montré tout récemment que deux autres protéines accessoires stimulent la recombinaison à ce site: il s'agit du système ArcA/ArcB (Colloms *et al.*, 1998).

Ce système contrôle l'expression de certains gènes d'*E. coli* dans des conditions d'anaérobiose (Arc: aerobic respiration control). ArcA (29 kDa) est une protéine régulatrice, tandis qu'ArcB (82 kDa) est une kinase sensorielle. Le changement dans la concentration en oxygène est donc détecté par ArcB. Dans ces conditions, ArcB, une fois phosphorylée, phosphoryle à son tour ArcA, qui peut alors se lier à la région promotrice de plusieurs gènes pour réprimer ou activer leur

transcription. Une revue de ces mécanismes est présentée dans Iuchi et Weiner (1996).

Pour la recombinaison au site *psi*, ArcA serait presque essentielle, tandis que ArcB ne le serait pas du tout; elle ne ferait que l'assister. La recombinaison peut être effectuée en absence d'ArcA, mais une plus grande quantité de PepA devient alors nécessaire. Une boîte ArcA est située à environ 100 pb du point d'échange des brins ; cela correspond à la position de la boîte ARG en amont du site de la région d'enjambement de *cer*. Cette boîte est liée de façon équivalente, peu importe qu'ArcA soit phosphorylée ou non. Il est supposé qu'ArcA pourrait jouer un rôle similaire à celui d'ArgR dans le cas du site *cer* (Colloms *et al.*, 1998).

6. Étude des intermédiaires de la recombinaison

6.1 Jonction de Holliday

Les jonctions de Holliday sont des structures intermédiaires retrouvées lors de la recombinaison, qu'elle soit homologue ou spécifique de site. Elles sont également retrouvées dans certains événements reliés à la réplication (Lilley et Clegg 1993). En plus des enzymes qui forment ces jonctions de Holliday, il a été montré que certaines protéines ont seulement une activité de liaison à ces jonctions et donc, auraient un rôle architectural (Zlatanova et Van Holde, 1998). Ainsi, il existe plusieurs sortes d'enzymes qui vont agir à un niveau ou à un autre : certaines vont être impliquées essentiellement dans la reconnaissance et la liaison spécifique à la structure formée, tandis que d'autres vont effectuer la catalyse pour la résolution de la jonction en tant que telle. Dans la recombinaison homologue, il y a, en plus, des enzymes impliquées uniquement dans le mouvement de la migration des branches (Duckett *et al.*, 1995).

6.1.1 Structure de la jonction

Les premières structures à quatre voies étudiées étaient cruciformes. Elles étaient formées par des séquences inversement répétées dans l'ADN surenroulé

(revue dans Lilley et Clegg, 1993). L'introduction de jonctions à quatre voies dans le centre d'un fragment d'ADN ralentit la mobilité de l'ADN en question dans un gel de polyacrylamide, et cette mobilité dépend de la concentration et du type de cations présents.

En présence de cations métalliques, la jonction adopte une forme en X (fig. 6). Cette forme provient d'une jonction à quatre voies qui a empaqueté les bras hélicaux de façon coaxiale en paire. Deux terminaisons des bras s'approchent l'une de l'autre, et c'est la même chose du côté opposé. Cette structure est antiparallèle et présente deux côté inégaux. D'un côté, un sillon majeur continu passe en dessous d'un bras, à travers le point d'échange, et continue sur un bras opposé du paquet d'hélices. De l'autre côté, un sillon mineur analogue continu est observé. Les quatre brins sont ainsi séparés en deux types : deux brins sont continus et les deux autres sont croisés ou échangeants. La structure est de plus tournée vers la droite.

Dans cette forme compactée, il existe deux isomères, c'est-à-dire deux formes alternatives possibles, selon le choix des brins qui doivent être croisés; ils se retrouvent en équilibre. Les populations relatives seraient déterminées par la séquence des bases s'étendant au moins aussi loin que la position suivant celle du centre (Lilley, 1997).

En absence d'ions métalliques, la structure est étendue et non empaquetée (fig. 6); les quatre bras sont dirigés vers les coins d'un carré.

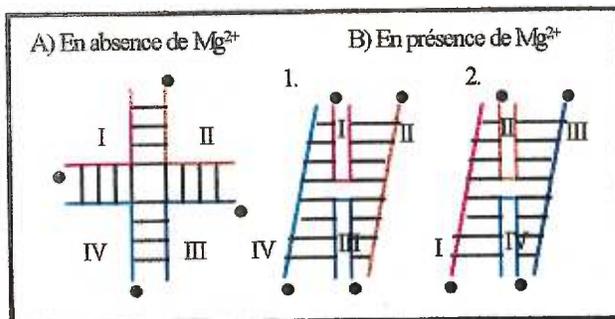


Figure 6- Formes que peut prendre la jonction de Holliday, en absence (A) ou en présence (B) de cations métalliques. Dans le second cas, il existe deux isomères qui diffèrent par le choix des brins croisés. Les • indiquent le côté 5' des brins marqués de I à IV. (Fig. tirée de Lilley, 1997).

6.1.2 La structure dans la recombinaison

L'accumulation de jonctions de Holliday a été observée *in vivo* lorsque la recombinaison a été effectuée entre deux sites *cer* sur une même molécule d'ADN (McCulloch, *et al.*, 1994b). Ces jonctions étaient formées par l'échange entre les brins supérieurs. Pour la formation de cette structure, il y a nécessité d'ArgR et de PepA. Par contre, malgré que RuvC peut cliver les jonctions de Holliday formées par le système Xer, la présence de l'enzyme n'est pas nécessaire pour l'évolution de la réaction. Durant la recombinaison homologe, c'est RuvC qui est l'endonucléase responsable de la résolution des jonctions de Holliday formées (Shinagawa et Iwasaki, 1996). La résolution de cette jonction est normalement très rapide, du moins lorsque l'action se fait par l'Int de λ (Nunes-Düby *et al.*, 1987). Cependant, cela ne semble pas être le cas au niveau du site *cer*, puisque le groupe de McCulloch (1994b) a pu en observer une grande quantité. Dans les deux cas, on peut au moins se persuader que la formation de cette structure doit s'avérer l'une des étapes intermédiaires de la réaction spécifique de site.

6.1.3 Modèle de la réaction de recombinaison

La première étape de la recombinaison spécifique de site est la liaison de deux protomères à chacun des deux duplexes d'ADN. Il y a ensuite la synapse des deux sites d'ADN occupés, laquelle est suivie par la formation de cette fameuse jonction de Holliday survenue à la suite de l'activation et du clivage d'un lien phosphodiester sensible, avec un échange réciproque des brins d'ADN. Enfin, il y a clivage de la deuxième paire de liens phosphodiester sensibles, puis échange des brins pour aboutir à la résolution de la jonction de Holliday. On obtient ainsi les produits recombinants (Blakely *et al.*, 1997 et fig. 7).

On croyait que la jonction de Holliday se formait au point de la première paire d'échange des brins, ce qui aurait nécessité par la suite une migration des branches à travers la région d'enjambement pour que le deuxième échange puisse s'effectuer (voir Stark *et al.*, 1992). Cependant, l'utilisation de jonctions de Holliday dans lesquelles la

mobilité se voyait réduite par l'introduction de substitutions de certaines paires de bases, n'a pas pu supporter le modèle de migration de jonction ("Branch migration") (Arciszewska *et al.*, 1995). La jonction de Holliday n'avait pas besoin d'être à la position de clivage par XerC, puisque ce clivage s'effectuait de toutes les façons à la position attendue, entre le dernier nucléotide de la région de liaison de XerC et le premier nucléotide de la région de crossing-over. Leurs résultats laissaient suggérer que le point de jonction optimal était au moins à deux nucléotides en 3' du point de clivage de XerC.

Cela semblait en accord avec le modèle de l'isomérisation suivant l'échange des brins ("strand swapping-isomerization") proposé par Nunes-Düby *et al.* (1995) (voir fig. 7). Dans ce modèle, le point de branchement pour l'échange n'est pas localisé au site de clivage et de ligation. En fait, la résolution la plus efficace est survenue lorsque le point de jonction était près de la moitié de la région de 7 pb où se fait l'échange. Il n'y a alors plus besoin de la migration qui était observée dans la

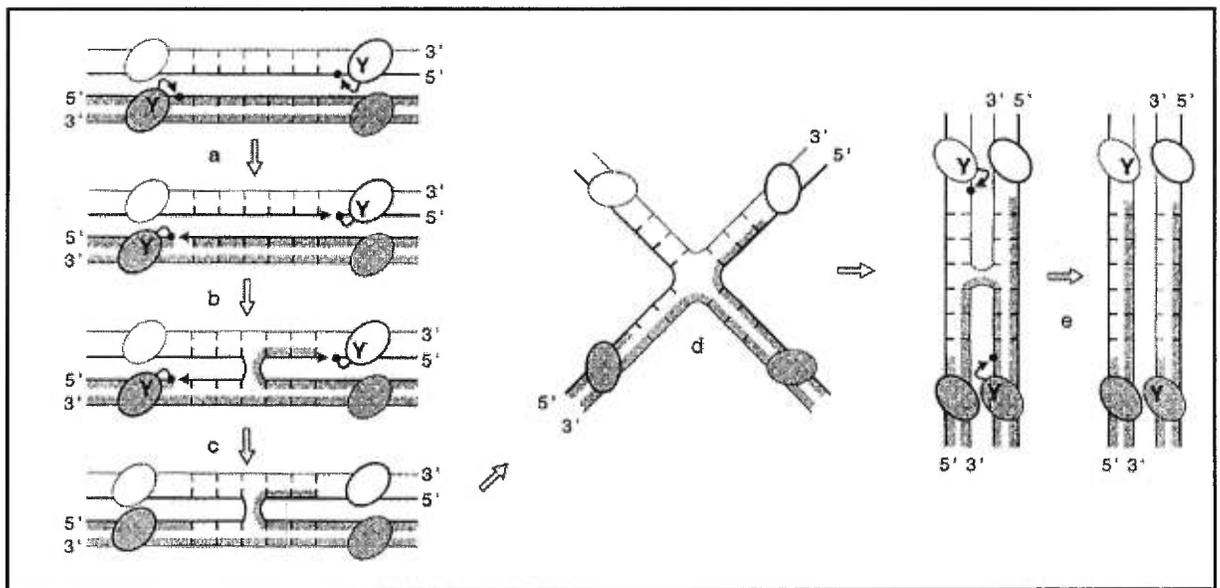


Figure 7- Mécanisme d'action des recombinases XerC et XerD au site *dif* selon le modèle de « strand swapping-isomerization » proposé par Nunes-Düby (1995) (fig. tirée de Hallet et Sherratt, 1997). **a**) Les brins supérieurs (traits épais) sont clivés en premier par la tyrosine catalytique de XerC (Y). Le clivage s'effectue entre le premier nucléotide de la région d'espace de 6 pb, et le dernier nucléotide du site de liaison de la recombinase. **b**) Il y a ensuite échange entre les deux partenaires, puis **c**) religation. Les brins supérieurs sont croisés et le point de branchement de la jonction est exactement au centre de la région d'espace. **d**) Pour permettre l'échange des brins inférieurs (traits minces) par XerD, il y a isomérisation de la jonction de Holliday ainsi formée. **e**) Finalement, les mêmes étapes (a à c), sont exécutées, mais par l'autre recombinase, ce qui donne lieu à la résolution de la jonction de Holliday et à la formation des produits recombinants.

recombinaison homologue, et qui nécessitait des changements structuraux entre les deux étapes pour l'échange des brins. Selon ce modèle, les clivages des brins supérieurs donnent lieu à la jonction de Holliday simplement en faisant « fondre » deux ou trois nucléotides des brins clivés et en les échangeant entre les sites actifs des deux protomères d'Int partenaires. Le point de jonction se situe donc approximativement au centre de la région d'enjambement pour l'échange médié par l'Int de λ , mais exactement au centre pour l'échange médié par le système Xer.

Ce modèle est également supporté par la structure de la protéine Cre à son substrat (Guo *et al.*, 1997). En effet, les quatre sites actifs de la synapse sont tous symétriques après le premier clivage de leur phosphate susceptible d'être clivé ; il y a adoption d'une conformation planaire pseudo-carrée par les bras des demi-sites *loxA* (*loxA*: site *loxP* modifié pour permettre l'emprisonnement de Cre à son site, voir le point 7.4) où la distance entre tous les sites est semblable.

6.1.4 Action de XerC et XerD

Des substrats ayant une jonction de Holliday synthétique ont été employés pour mieux définir le rôle de chacune des recombinases dans la réaction d'échange des brins (Arciszewzka et Sherratt, 1995). Cette jonction de Holliday a été construite avec quatre oligonucléotides de longueurs différentes, pour permettre de distinguer la résolution des brins supérieurs de celle des brins inférieurs. C'est l'utilisation de ces substrats qui a permis de montrer que le clivage des brins dits supérieurs se faisait entre le dernier nucléotide du site de liaison à XerC et le premier nucléotide de la région centrale où a lieu l'enjambement. L'utilisation de mutants de XerC et de XerD dans les résidus impliqués dans la catalyse de la réaction et de protéines fusionnées avec la protéine liant le maltose (MBP : maltose-binding protein) a montré que XerC était impliqué dans cette 1^{re} étape de transestérification.

Le même type de substrat a été employé par la suite pour montrer l'action des recombinases sur ceux-ci, et savoir ce que liaient les recombinases (Arciszewska *et al.*,

1997). Les protéines sont probablement plus près du point de branchement, puisqu'une distorsion locale de l'ADN survient lorsque les recombinaisons s'y lient.

L'expérience d'Arciszewska *et al.* (1997) a permis de montrer que XerD ne pouvait agir sur des jonctions de Holliday ayant le site *dif* que lorsque les brins inférieurs étaient croisés. XerD a donc une spécificité de substrat. Encore une fois, le modèle de Nunes-Düby *et al.* (1995) semble bien concorder avec les résultats, puisque alors le croisement des brins inférieurs ne se fait que par un processus inverse à celui qui a permis l'échange des brins supérieurs, et ce, sans nécessiter trop de changements conformationnels. L'isomérisation permet ainsi à XerD de pouvoir cliver les brins inférieurs.

Cette préférence pour l'isomère dont les brins sont croisés a également été montrée avec l'Int de λ (Azaro et Landy, 1997). L'intégrase est donc capable de distinguer les deux isomères de type empaqueté. Leurs résultats ont également montré que l'échange des brins supérieurs et celui des brins inférieurs étaient symétriques, mais différents facteurs vont influencer l'ordre des événements de la recombinaison, comme le surenroulement de l'ADN.

Quant à XerC, bien que son action ait été plus efficace lorsque les brins supérieurs étaient croisés, elle pouvait également agir sur le substrat ayant des brins continus (ce sont alors les brins inférieurs qui sont croisés). Dans ce cas, il est possible qu'il n'y ait pas nécessité d'isomérisation advenant le cas où l'échange serait initié par XerD pour certains substrats, autres que *psi*, *cer*, ou *dif* (Arciszewska *et al.*, 1997).

Dans une étude (Zerbib *et al.*, 1997), il a été suggéré que des contraintes topologiques empêchent ou ralentissent l'isomérisation de la jonction. Elles influencent donc peut-être la façon dont la recombinaison survient, puisque les deux formes d'isomères peuvent exister. Au niveau du site *cer*, contrairement à ce qui est retrouvé chez *dif* ou *psi*, la conformation dans laquelle les brins supérieurs sont croisés est probablement adoptée préférentiellement, car sa région centrale est riche en purines sur le brin supérieur ; Azaro *et al.* (1997) ont montré que la forme isomère observée était celle dans laquelle les brins croisés avaient un maximum de purines.

6.2 Les substrats suicides

Un autre type de substrats très employés pour l'étude des intermédiaires de la recombinaison sont les substrats suicides: ils permettent l'accumulation de ces intermédiaires. Ces substrats ont une brèche dans la région d'échange des brins, ce qui modifie la réaction de recombinaison, puisque l'attaque nucléophile par le résidu 5'OH libéré lors de la liaison de la tyrosine catalytique de la recombinase à son site d'action, ne peut s'effectuer. L'énergie du lien phosphodiester n'est donc plus disponible pour la recombinase et c'est ce qui permet de « retenir » des intermédiaires dans lesquels la recombinase est liée de façon covalente à l'ADN (Nunes-Düby *et al.*, 1987).

Pour connaître le rôle de XerC et XerD au site *dif*, de tels substrats ont été construits avec une brèche située à trois nucléotides du côté 3' du site de clivage de la recombinase. Ainsi, une fois le clivage effectué, il y a perte des trois nucléotides ayant le 5'OH qui aurait pu relier le substrat d'ADN dans une réaction inverse (Blakely *et al.*, 1997).

Ces substrats ont permis de montrer que XerD était responsable du clivage des brins inférieurs du site *dif*, et ce, entre le dernier nucléotide de son site de liaison et le premier nucléotide de la région centrale, comme c'est le cas pour XerC (Blakely *et al.*, 1997). Il avait été établi que XerC pouvait cliver si XerD n'était pas catalytiquement actif (Arciszewska et Sherratt, 1995), mais il s'avère que XerC peut même cliver en absence de XerD au niveau du site *dif*, en autant qu'une quantité suffisante de XerC soit employée. XerC pouvait même se lier au site de liaison de XerD sur un site *dif* « suicide », grâce à une interaction coopérative entre deux protomères de XerC, de la même sorte que celle qui survient entre XerC et XerD. Il doit donc y avoir un domaine d'interaction commun aux deux protéines (Blakely *et al.*, 1997).

6.3 Formes des substrats pour la recombinaison par Xer

Les plasmides ayant les sites *psi* ou *cer* doivent être surenroulés pour qu'agissent les recombinases de type Xer *in vitro* (la recombinaison n'a pas été

observée entre des sites *dif*) (Colloms *et al.*, 1996). Le surenroulement serait peut-être nécessaire lors de l'échange des brins, mais il n'est pas requis pour la formation de la synapse (Ålen *et al.*, 1997). Les produits obtenus par la recombinaison sur les sites *psi* et *cer* sont différents. La recombinaison sur un plasmide ayant deux sites *cer in vitro* (avec ArgR, PepA, XerC et XerD) donne lieu à l'accumulation de jonctions de Holliday. La recombinaison au site *psi*, par contre, permet l'obtention de caténanes (Colloms *et al.*, 1996). Un caténane consiste en deux anneaux interreliés⁴. La recombinaison au site *psi* est donc complète, tandis qu'elle cesse après le premier échange des brins au niveau du site *cer*. La formation de la jonction de Holliday est exécutée par XerC dans les deux cas. Tandis que XerD n'a pas de rôle catalytique au niveau du site *cer*, son action doit être plutôt rapide au niveau du site *psi*, puisque très peu de produits secondaires sont alors observés. Comme la recombinaison au site *cer in vivo* donne lieu à des produits recombinants complets (McCulloch *et al.*, 1994b), il a été supposé qu'un système autre que Xer devait être impliqué pour le deuxième échange, mais il n'a pas encore été déterminé (Colloms *et al.*, 1996).

Bien que le résultat final soit différent pour les deux sites, l'intermédiaire formé par XerC est une figure en 8 caténée à 4 noeuds. Pour le site *psi*, la topologie du produit final a été définie : il s'agit d'un caténane à 4 noeuds tourné vers la droite, dont les sites *psi* sont antiparallèles (Colloms *et al.*, 1997).

7. Structure tridimensionnelle de certaines intégrases

Jusqu'en 1997, aucune structure de recombinases de la famille de l'intégrase du phage λ n'était connue. La seule structure résolue d'une recombinase spécifique de site était celle de la résolvasse $\gamma\delta$, faisant partie de la famille des résolvasse/invertases (Rice et Steitz, 1994: pour la résolvasse intacte, et Yang et Steitz, 1995 : pour la résolvasse complexée au site de clivage). La cristallisation de différentes recombinases de type intégrase fut un grand pas vers l'avant, puisqu'elle permet, entre autres, de mieux comprendre les mécanismes d'action des enzymes et de visualiser les

⁴ Pour les différentes formes que peut prendre l'ADN, voir Sinden, (1994).

domaines impliqués dans différentes actions telles que la liaison à l'ADN, les interactions protéine-protéine, etc.

7.1 Structure de XerD

La structure de XerD d'*E. coli* a été déterminée à une résolution de 2,5Å (Subramanya *et al.*, 1997).

XerD est constitué de deux domaines distincts, reliés par un segment de 26 résidus (voir fig. 8):

- Le domaine N-terminal (résidus 1 à 107) est constitué de quatre hélices α : deux hélices parallèles (nommées αA et αB) sont perpendiculaires aux deux autres (αC et αD), parallèles entre-elles également.

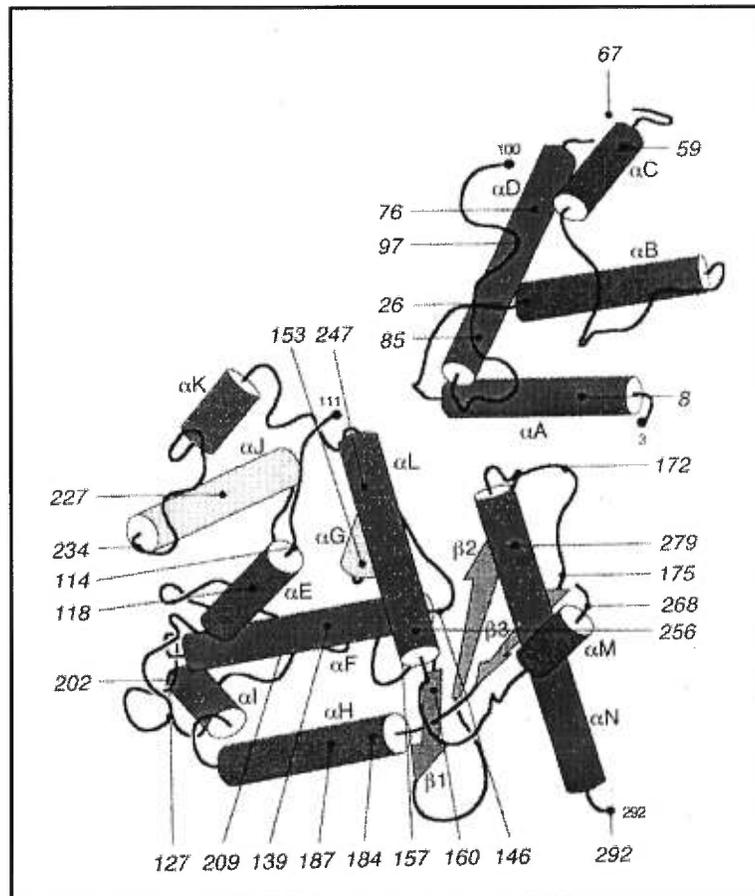


Figure 8- Structure cristalline de XerD d'*E. coli*. Cette figure est tirée de Cao *et al.* (1997). Les positions indiquées font référence aux différents points d'insertion de pentapeptides dans le cadre de l'expérience de cette équipe, pour tenter de définir la fonction de ces régions.

- Le domaine C-terminal (résidus 108 à 298) est constitué de 10 hélices α (E à N) et de trois feuillets β antiparallèles (1 à 3).

Il faut noter que trois régions n'ont pu être définies : elles contiennent les résidus 64-70, 101-110 et 269-270, en plus des résidus aux extrémités N- et C-terminales.

Le repliement du domaine C-terminal s'avère unique. Il n'a été retrouvé que chez les autres intégrases cristallisées jusqu'à présent, c'est-à-dire Int du phage λ (Kwon *et al.*, 1997), Int de HP1 (Hickman *et al.*, 1997), et Cre de P1 (Guo *et al.*, 1997). La tyrosine catalytique est située dans l'hélice α N qui est précédée d'un tour. La structure primaire de cette région est la plus conservée parmi les recombinases de type Xer, et même, parmi les autres intégrases. Pourtant, elle est très divergente du point de vue structural (voir le point 7.5 de ce chapitre). Dans le cas de XerD, la tyrosine catalytique serait camouflée par le domaine N-terminal, ce qui imposerait la nécessité d'un changement de conformation pour qu'elle soit disponible pour l'attaque du lien phosphodiester activé du site d'action.

7.2 Structure de l'Int de λ

La structure de l'Int de λ ne comprend que le domaine C-terminal, qui représente le domaine catalytique minimal (fig. 9). Elle comprend les résidus 170 à

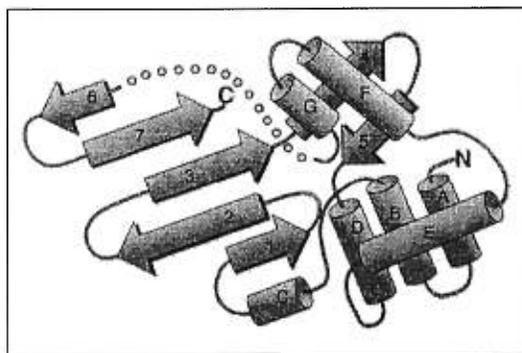


Figure 9- Structure cristalline du domaine C-terminal de l'Int de λ (résidus 170 à 356), déterminée à une résolution de 1,9 Å. (Figure tirée de Kwon *et al.*, 1997).

356 et a 7 hélices α et un plus grand nombre de feuillets β que XerD, soit 7 (Kwon *et al.*, 1997). Ce fut la première structure obtenue d'un membre de la famille des intégrases. Le cœur du domaine est formé en fait des hélices α A, α B et α C (qui correspondent aux hélices α F, α G et α H de Cre selon Yang et Mizuuchi, 1997), lesquelles sont encadrées par les hélices α D, α E et α F (correspondant à α I, α J et α K). La tyrosine

catalytique se situe sur une boucle de 17 acides aminés, qui est exposée et qui s'étend

entre l'hélice αG et le feuillet $\beta 6$. La tyrosine se voit donc faiblement liée au reste de la protéine. De plus, deux configurations différentes ont été obtenues pour deux protomères. Dans l'une d'elles, la tyrosine était non visible.

L'Int de λ se sépare en fait, tout comme XerD en deux domaines : le domaine N-terminal, comprenant les résidus 1 à 64, est le domaine de reconnaissance et de liaison des bras d'ADN, tandis que le domaine C-terminal (résidus 65 à 356) se lie au cœur de l'ADN en tant que tel et a l'activité topoisomérase. Ce domaine peut être réduit davantage à partir du 170^e résidu, et il constitue alors le domaine catalytique minimal pouvant cliver un substrat suicide ayant le site spécifique pour l'Int (*att*) (Tirumalai *et al.*, 1997). C'est ce domaine qui a été cristallisé.

7.3 Structure de l'Int de HP1

L'intégrase de HP1 permet l'intégration et l'excision de ce phage à l'intérieur et hors du chromosome bactérien d'*Haemophilus influenzae*, comme c'est le cas de l'Int de λ pour le bactériophage λ . Cette protéine possède deux domaines. Le domaine C-terminal (résidus 165 à 337) a été cristallisé à 2,7Å et la tyrosine catalytique est située sur une hélice α dans un genre de crevasse enlignée par des résidus basiques et dans laquelle on retrouve également les autres résidus actifs des intégrases (Hickman *et al.*, 1997) (fig. 10). La structure de ce domaine présente neuf hélices α et trois feuillets β ; la dernière hélice s'éloigne du reste du corps de la protéine. Un dimère est formé par la localisation de la queue en extrémité du domaine C-terminal d'un monomère dans la crevasse hydrophobe du second monomère.

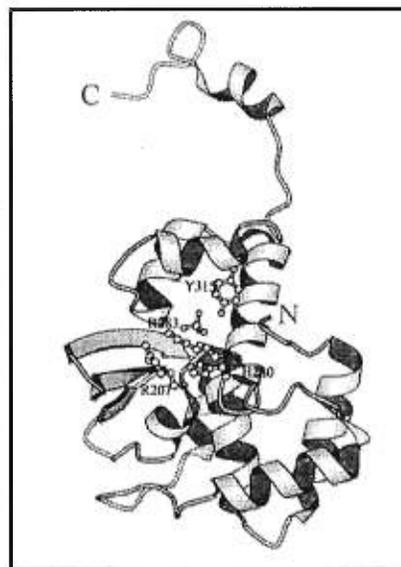


Figure 10- Structure cristalline du domaine C-terminal de l'Int de HP1. Les résidus du cœur catalytique sont indiqués. (Figure tirée de Hickman *et al.*, 1997).

7.4 Structure de Cre

La recombinase Cre sert au maintien sous forme de monomère du bactériophage P1 dans une cellule d'*E. coli*, lors de son état lysogène. Cette protéine contient deux domaines distincts (tout comme les autres recombinases mentionnées plus haut). La structure révèle que les deux domaines sont reliés par un court lien (Guo *et al.*, 1997). Le domaine N-terminal (résidus 20 à 129) contient cinq hélices α , disposées de telle sorte que trois hélices (α C, α D et α E) sont antiparallèles les unes

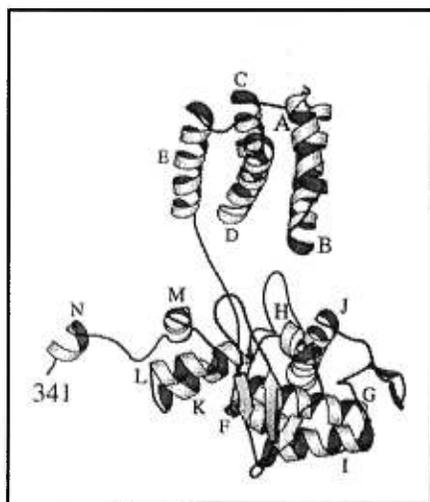


Figure 11- Structure cristalline complète de Cre, déterminée à une résolution de 2,4 Å, tirée de Guo *et al.* (1997). Les 19 premiers résidus, ainsi que les 2 derniers, n'ont pu être définis.

aux autres, tandis que les deux autres (α A et α B) sont quasi-perpendiculaires à ce paquet. Le domaine C-terminal (résidus 132 à 341) contient neuf hélices α et trois feuillets β ; la dernière hélice (α N) s'éloigne du corps de la protéine (fig. 11), tout comme c'est le cas chez l'Int de HP1.

La structure a également été obtenue avec Cre liée à un substrat suicide (*loxA*), lequel est le site spécifique de Cre (*loxP*) modifié. Les contacts avec l'ADN et les interactions protéine-protéine ont ainsi pu être visualisés. Les domaines N- et C-terminaux entourent l'ADN un peu sous la forme d'une pince. L'ADN se voit courbé d'environ 100° en son centre. Le domaine C-terminal fait surtout

des contacts avec le sillon mineur, tandis que les hélices α B et α D du domaine N-terminal ont des contacts avec le sillon majeur.

Il a été remarqué qu'il y avait une différence dans le domaine C-terminal entre un monomère qui exerce l'action de clivage et celui qui ne le fait pas. La conformation du segment reliant l'hélice α M à l'hélice α N est plus étendue lorsque Cre clive son substrat. Le contact entre les monomères a également été observé : la dernière hélice de la protéine s'enfonce dans la poche hydrophobe ("hydrophobic pocket") formée des hélices α L et α M de l'autre monomère. À la différence de l'Int

de HP1, où l'échange est réciproque, l'échange des hélices est cyclique (Guo *et al.*, 1997).

7.5 Comparaison de la structure de XerD aux autres structures d'intégrases

Selon Yang et Mizuuchi (1997), deux caractéristiques importantes peuvent ressortir de l'analyse des quatre structures des membres de la famille des intégrases :

- Seules la multimérisation correcte et l'association avec l'ADN permettent la formation du site actif complet.
- L'échange entre les molécules d'ADN ne nécessiterait que de légers changements conformationnels, d'après ce qui est révélé par la structure du tétramère cyclique de Cre sur lequel l'ADN forme une couche planaire-carrée.

La structure des quatre recombinases cristallisées (Cre de P1, Int de HP1, Int de λ et XerD d'*E. coli*) présentent des similarités, dans le domaine contenant le site actif de l'enzyme, c'est-à-dire le domaine C-terminal. Toutes ont trois feuillets β conservés qui correspondent chez XerD à la région qui suit immédiatement le motif I. Ce qu'il y a d'un peu particulier, c'est que la dernière portion des protéines, qui s'avère fortement conservée et contenant le 2^e motif (avec les trois autres résidus du site actif), sont différentes du point de vue structurel. L'Int de HP1 et Cre présentent toutes deux la dernière hélice qui s'éloignent du corps de la protéine (fig. 10 et 11). Pour l'Int de λ , il y a une boucle flexible (qui suit α G) et un feuillet β qui remplacent les hélices α M et α N trouvée chez Cre et le déplacement de l'extrémité C-terminale empêche la tyrosine catalytique d'être près des autres résidus conservés (fig. 9). Quant à XerD, la tyrosine catalytique ne peut pas attaquer le phosphate susceptible d'être clivé, puisqu'elle est camouflée par le domaine N-terminal ; l'hélice α M de l'extrémité C-terminale (correspondant à α L de Cre) tourne autour de α N (correspondant à la fusion de α M et α N de Cre) et en change la direction (fig. 8).

Au contraire, la structure primaire est peu conservée dans la région qui s'étend de αF à αJ chez Cre, mais la structure tertiaire, elle, est similaire chez les quatre recombinases.

Enfin, le cœur catalytique que l'on retrouve chez Cre, formé des hélices αK , αL et αM , n'est retrouvé ni chez XerD, ni chez l'Int de λ , tandis qu'il est modifié chez l'Int de HP1 (Yang et Mizuuchi, 1997).

8. Caractérisation de XerD

Une première caractérisation des recombinases se fait au niveau de la séquence primaire, et ce, afin de faire le lien avec la fonction de la région ciblée. Ainsi, une série de délétions ont été effectuées pour associer la séquence primaire de XerD à la capacité de liaison à l'ADN, et aux interactions protéine-protéine et protéine-ADN qui mènent à la liaison coopérative, à la synapse et à l'échange des brins des sites de recombinaison (Spiers et Sherratt, 1997). Une fois la structure tridimensionnelle de XerD déterminée, un autre genre d'expérience a consisté à y insérer des pentapeptides de façon aléatoire, également pour tâcher d'identifier la fonction de certaines régions de la protéine (Cao *et al.*, 1997). Dans ce dernier cas, la protéine de *Salmonella typhimurium* a été employée ; elle s'était révélée fortement identique à XerD d'*E. coli* (Hayes *et al.*, 1997).

Cette dernière technique peut par contre présenter un désavantage : elle peut favoriser un mauvais repliement de la protéine, ou encore un déplacement des éléments qui sont près du point d'insertion du pentapeptide, et alors, l'inactivité ne serait pas forcément due à la mutation dans la séquence primaire. Ainsi, une insertion près de la tyrosine catalytique empêche la recombinaison au site *cer* (Cao *et al.*, 1997), alors qu'une simple mutation ponctuelle n'affecte pas la réaction (Colloms *et al.*, 1996). En fait, « ces insertions peuvent modifier suffisamment la poche catalytique ("catalytic pocket") pour endommager les réactions d'échange des brins » (Cao *et al.*, 1997) ; l'activation de XerC par XerD ne peut plus se faire. Malgré tout, la technique

peut donner une assez bonne idée des régions impliquées dans les fonctions importantes de la protéine.

L'activité *in vivo* à *dif* et à *cer*, la liaison à *dif*, ainsi que la coopérativité à *dif* ont été évaluées et le tableau de l'annexe 3 présente l'ensemble des résultats obtenus par cette technique. Les points d'insertion des pentapeptides sont indiqués sur la fig. 8. Cela a permis de confirmer l'importance des motifs I et II dans l'activité de recombinaison et dans l'efficacité de la liaison au site *dif*. D'ailleurs, la délétion de ces motifs ne permettaient plus la résolution intramoléculaire entre deux sites *cer*, ni la recombinaison intermoléculaire entre des sites *dif* (Spiers et Sherratt, 1997). La coopérativité entre XerC et XerD ne semble pas impliquer le motif I, mais plutôt la région située près du deuxième R conservé faisant partie de la poche catalytique. L'insertion au niveau de l'histidine (résidu 268) de la séquence LGH du motif II, n'affecte ni la coopérativité, ni la liaison à l'ADN et pourtant la recombinaison est inhibée. Chez l'Int de HP1, il a été proposé que cette histidine maintiendrait peut-être l'intégrité de la poche catalytique (Hickman *et al.*, 1997).

Quant au troisième motif hautement conservé, sa fonction exacte n'est pas encore déterminée. Ce motif se situe dans le tour reliant les feuillets $\beta 2$ à $\beta 3$. Il contient la séquence GKGXKXG qui est retrouvée chez toutes les recombinases de type Xer (voir l'annexe 1). Les insertions dans ce motif, bien qu'elles n'altèrent ni la liaison à *dif*, ni l'interaction avec XerC, abolissent l'activité catalytique de XerD à ce site (Cao *et al.*, 1997). L'extrémité C-terminale de ce motif (résidus 278 à 283) permettraient la communication de XerD avec XerC, pour l'exécution de la catalyse et de l'échange des brins médiés par XerC (Spiers et Sherratt, 1997).

Selon le potentiel électrostatique de la surface des domaines, les deux domaines (N- et C-terminaux) révèlent des sites pour l'interaction avec l'ADN (Subramanya *et al.*, 1997). Pour le domaine 1, cela se ferait par l'intermédiaire des hélices αB et αD . Ces hélices du domaine N-terminal, dont les séquences sont peu conservées parmi les recombinases de type Xer, seraient peut-être impliquées dans l'interaction avec des segments palindromiques des sites de liaison de XerC et XerD.

Dans l'hélice αJ , on retrouve les résidus 220R et 221Q qui sont conservés chez toutes les recombinases de type XerD, à l'exception d'*Helicobacter pylori* (voir l'annexe 1). Subramanya *et al.* (1997) ont d'ailleurs proposé un modèle pour l'interaction du 2^e domaine de XerD avec l'ADN. Pour ce faire, ils se sont basés sur la liaison de la protéine CAP (catabolite gene activator protein) à une séquence de 30 pb. La protéine CAP, également nommée CRP (cAMP receptor protein), lorsqu'associée à l'AMPc, agit à des séquences d'ADN spécifiques en amont de différents opérons. Un pli de 40° est créé dans l'ADN lorsqu'il entoure un motif hélice-tour-hélice de la protéine CAP (Schultz *et al.*, 1991). Ce type de courbure correspond à ce qui a été déduit de la liaison de XerD avec *dif* (Blakely et Sherratt, 1996b). Cependant, le motif mentionné n'apparaît pas vraiment chez XerD, mais il est supposé que les hélices αG et αJ , séparées par environ 65 résidus, peuvent y correspondre. Ce modèle concorde bien avec les résultats obtenus précédemment par Hayes et Sherratt (1997), Blake *et al.* (1997) ou Blakely et Sherratt (1996b), où il y avait des contacts spécifiques avec certaines positions au site *dif*. L'hélice αJ serait l'hélice de reconnaissance de l'ADN. Alors, qu'une insertion de pentapeptides dans la région comprise entre les hélices αG et αJ affecte, en général, relativement peu la recombinaison *in vivo* aux sites *cer* et *dif*, l'hélice αH est fortement affectée par une telle insertion (Cao *et al.*, 1997). Les deux positions d'insertion sont en bordure C-terminale du motif III.

Enfin, outre les trois motifs cités et les deux hélices mentionnées du domaine N-terminal, il est d'autres régions qui ont rôle quelconque dans la recombinaison. Ainsi, les cinq derniers résidus de XerD seraient importants pour la formation de la synapse puisque leur délétion ne permet plus l'échange de XerC sur des substrats surenroulés (Spiers et Sherratt, 1997).

Les rôles présumés des régions ciblées par les deux expériences mentionnées sont résumés dans le tableau suivant.

Régions de XerD	Rôles présumés	Réf.
Région autour du résidu 26	Coopérativité avec XerC au site <i>dif</i>	3
Hélices α B et α D	Liaison à l'ADN	2,3
Position 59 (hélice α C)	Pour l'activité <i>in vivo</i> à <i>cer</i> , pas à <i>dif</i>	3
Position 76 (hélice α D)		
Hélice α E		
Segment reliant les deux domaines	Pas de rôle important connu	3
Motif I (domaine 2)	Liaison à l'ADN	3
Boucle reliant β 2 à β 3 (motif III)	Activité catalytique à <i>dif in vivo</i> ; pas vraiment à <i>cer</i> ; rôle dans la réaction d'échange des brins	3
Hélice α J	Reconnaissance pour la liaison à l'ADN	3
Hélice α L	Coopérativité avec XerC au site <i>dif</i>	3
R247 (R conservé du motif II)		
Résidu 256 (extrémité d' α L)	Pour l'activité <i>in vivo</i> à <i>cer</i> , pas à <i>dif</i>	3
Motif LGH (H en position 268)	L'histidine serait peut-être impliquée dans le maintien de l'intégrité de la poche catalytique ("catalytic pocket")	4
Y279	Activités topoisomérase et catalytique	1
Résidus 278-283 (Hélice α N)	Communication avec XerC pour la catalyse et l'échange	1
Cinq derniers résidus (294-298)	Formation du complexe synaptique	1

Tableau I- Rôles présumés de certains résidus ou régions de XerD. Réf.: références. 1: Spiers et Sherratt (1997), 2: Subramanya *et al.* (1997), 3: Cao *et al.* (1997), 4: Hickman *et al.* (1997). La position des résidus est indiquée sur la structure cristalline de XerD, fig. 8.

9. Ubiquité des recombinaisons de type Xer

Bien entendu, les premières recombinaisons de type Xer à avoir été définies, furent celles d'*E. coli*. Ce sont également celles qui sont le plus étudiées pour parvenir à saisir leur rôle et leur mécanisme d'action au sein de la cellule bactérienne. Cependant, compte tenu du rôle, justement, particulièrement important, on pouvait fortement supposer que ce système devait exister chez l'ensemble des espèces bactériennes, sinon du moins chez d'autres *Enterobacteriaceae*. Ce fut l'objet de l'étude du laboratoire du Dr Szatmari qui a révélé que des séquences homologues de ces deux recombinaisons existent chez *Klebsiella pneumoniae*, *Morganella morganii*, *Salmonella typhimurium*, *Serratia marcescens* et *Shigella flexneri* (Sirois et Szatmari, 1995). Les séquences de *S. typhimurium* ont d'ailleurs été publiées (Hayes *et al.*, 1997) et elles montrent un très fort pourcentage d'identité avec les protéines d'*E. coli*, de l'ordre des 90%. Elles sont de plus aptes à compléter une souche d'*E. coli* mutante pour les gènes respectifs. Enfin, l'organisation génomique est exactement la même que celle d'*E. coli*. Il en est de même pour XerC de *S. marcescens* qui montre

74% d'identité avec XerC d'*E. coli*. De plus, une insertion dans le gène codant pour cette protéine ou dans *uvrD* qui suit immédiatement l'*orf238*, mène à une surexpression des nucléases, probablement par l'induction du système SOS (Guynn *et al.*, 1998).

La protéine Sss de *Pseudomonas aeruginosa* est également homologue à XerC, mais avec un pourcentage d'identité plus faible, soit de 50% (Höfte *et al.*, 1994). Cependant, malgré la plus grande divergence entre les deux protéines, l'organisation génomique reste la même. Il en est de même pour la protéine Sss de *P. fluorescens* qui montre 48% d'identité à XerC (Dekkers *et al.*, 1998).

Ce type de recombinaisons a également été retrouvé chez des bactéries gram-positives telles que *Bacillus subtilis* (Slack, *et al.*, 1995), *Lactobacillus leichmannii* (Becker et Brendel, 1996) et plus récemment, *Staphylococcus aureus* (Tichat et Szatmari, résultats non publiés). Les deux recombinaisons de *B. subtilis* ont pour nom CodV et RipX, qui correspondent à XerC et XerD respectivement. Alors que l'organisation génomique de ces recombinaisons semble assez bien conservée chez les Gram-, elle apparaît complètement différente chez les Gram+. Chez *B. subtilis*, CodV est le premier gène d'un opéron qui en compte quatre et le gène qui suit, *codW* encode une protéine de choc thermique qui correspondrait à HslV retrouvé chez *E. coli* (Slack *et al.*, 1995). Le gène *xerC* de *L. leichmannii* est suivi aussi par un ORF similaire à *hslV* (Becker et Brendel, 1996).

Dans les banques de données, d'autres séquences montrent de fortes similarités aux deux recombinaisons et les espèces sont regroupées dans l'annexe IV avec les numéros d'accès correspondant. Des séquences similaires semblent être également retrouvées chez *Neisseria gonorrhoeae* (Cao *et al.*, 1997), *Streptococcus mutans* et *Actinobacillus actinomycetemcomitans*, mais les séquences n'ont pas encore toutes été vérifiées (site web : <http://www.genome.ou.edu/>).

10. *Projet de maîtrise*

Comme on peut le constater par l'étendue de ces recombinaisons au sein de diverses espèces bactériennes, on pouvait fortement supposer que ce système devait être ubiquitaire. Cependant, on pouvait supposer que pour *Proteus mirabilis*, tel ne devait pas être le cas, puisque le plasmide ColE1 avait tendance à multimériser dans cette espèce (Goebel et Helinski, 1968). De plus, la multimérisation d'un vecteur possédant les sites *cer* et *nmr* (pNCK-1) avait également été observée (Sirois et Szatmari, 1995). Ce dernier vecteur était d'ailleurs perdu après 40 ou 80 générations. Aucune autre expérience n'a été effectuée pour savoir quel pouvait être le problème dans le maintien de la stabilité de ces plasmides. C'est pourquoi l'étude de *P. mirabilis* semblait très intéressante. De plus, il avait été même montré qu'une sonde faite à partir des recombinaisons d'*E. coli* n'hybridait pas au génome de *P. mirabilis* (Sirois et Szatmari, 1995). Ceci pouvait signifier que ce système n'existait effectivement pas chez cette espèce, ou bien qu'il avait tellement divergé que les séquences n'étaient pas suffisamment similaires pour être détectées par une hybridation de type Southern. Cependant, la résolution de pNCK-1 ayant pu être observée en plus de sa multimérisation, cela a permis de suggérer alors que le système, soit ne parvenait pas à favoriser la monomérisation, soit favorisait ou ne pouvait empêcher la multimérisation.

Mon projet de maîtrise a permis de révéler que l'espèce *P. mirabilis* possède bel et bien les deux gènes qui encoderaient pour des recombinaisons équivalentes à XerC et XerD. L'étude a par la suite porté principalement sur la caractérisation du gène *xerD* et c'est ce qui a mené à une publication dans FEMS Microbiology Letters, laquelle est présentée au chapitre suivant.

CHAPITRE II :

ARTICLE



ELSEVIER

FEMS Microbiology Letters 164 (1998) 83–90

FEMS
 MICROBIOLOGY
 LETTERS

Cloning and characterisation of the *Proteus mirabilis xerD* gene

Manuela Villion, George Szatmari *

Département de Microbiologie et Immunologie, Université de Montréal, C.P. 6128 Succursale Centre-Ville, Montréal, Qué. H3C 3J7, Canada

Received 24 March 1998; revised 4 May 1998; accepted 5 May 1998

Abstract

The Xer site-specific recombination system is involved in the stable maintenance of replicons (certain plasmids and chromosomes) in *Escherichia coli* and other bacteria by converting multimers into monomers. This system requires a *cis*-acting DNA sequence (the chromosomal *dif* site or the ColE1 *cer* site) and two *trans*-acting factors: the XerC and XerD recombinases, which belong to the lambda integrase family of tyrosine site-specific recombinases. In addition, in order to resolve plasmid multimers into monomers, two additional factors are required: the ArgR and PepA proteins. We have previously shown the presence of *xerC* and *xerD* genes (and their function) by Southern hybridisation and by *in vivo* recombination in a wide variety of Enterobacteriaceae. We have now cloned and sequenced the *xerD* gene of *Proteus mirabilis* using degenerate and inverse PCR methods. This gene encodes a tyrosine recombinase which is highly similar to the *E. coli* XerD recombinase, is capable of complementing an *E. coli xerD* mutant, and displays sequence-specific DNA binding activity. © 1998 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

Keywords: Site-specific recombination; Tyrosine recombinase; *xerD*; *Proteus mirabilis*

1. Introduction

The stable inheritance of replicons (plasmids and chromosome) in *Escherichia coli* is assured by a site-specific recombination system, called the Xer system. This system comprises two recombinases, XerC and XerD, which form part of the lambda integrase family of site-specific recombinases [1]. These two proteins act in a cooperative manner at specific sites in the bacterial chromosome (the *dif* site), or in plasmids (i.e. *cer* in ColE1, *psi* in pSC101, etc.) [1,2]. In order to act on plasmid sites, additional factors (ArgR and PepA) are required to favourise intramo-

lecular recombination (resolution selectivity) [3,4]. However, in the case of the chromosomal *dif* site, no other additional factors are required [5], and the reaction can be either intra- or intermolecular *in vitro*.

All of the recombination sites contain a central region of 28–30 bp. This region can be further subdivided into three regions: two 11 bp sequences flanking a core region where the strand exchange takes place. The left part of the site contains the XerC binding site, while the right part is bound by XerD [6]. The central core region contains 6–8 bp of variable sequence. The top strand is cleaved by XerC between its binding site and the core region, while the bottom strand is cleaved by XerD between its binding site and the core [7,8].

* Corresponding author. Tel: +1 (514) 343-5767;
 Fax: +1 (514) 343-5701; E-mail: szat@ere.umontreal.ca

The Xer system, initially discovered in *E. coli*, has also been found in most Enterobacteriaceae by Southern hybridisation, and has also been shown to be functional by a plasmid-based *cer* recombination assay [9]. Similar recombinases have been sequenced in *Pseudomonas aeruginosa* [10], *Bacillus subtilis* [11], *Lactobacillus leichmannii* [12], *Salmonella typhimurium* [13], *P. fluorescens*, *Haemophilus influenzae*, *Mycobacterium tuberculosis*, *M. lepreae* and *Neisseria gonorrhoeae* [14], and in *Helicobacter pylori* (GenBank accession numbers, AE000580 for *xerC* and AE000608 for *xerD*) and *Serratia marcescens* (GenBank accession number, AFO28736 for *xerC*).

The study of Xer recombination in *Proteus mirabilis* is of interest for two reasons. First, observations by Goebel and Helinski [15] showed ColE1 plasmids to be unstable in this bacterium due to excessive multimerisation, despite the occurrence of a multimer resolution site (*cer*) in these plasmids [9]. Our laboratory was nonetheless able to detect Xer/*cer* activity in this species [9], suggesting that the Xer recombination system in this bacterium was either ineffective in halting plasmid multimerisation, was altered in such a fashion as to favour the formation of multimers, or had lost its resolution selectivity (converted multimers to monomers as well as monomers to multimers). Second, this observation of Xer activity, coupled to the lack of DNA sequence homology between *P. mirabilis* and *E. coli* made the study of these genes interesting from an evolutionary standpoint. In this work, we described the cloning of the *xerD* gene of *P. mirabilis* by degenerate and inverse PCR, and the functional characterisation of this gene by complementation analysis and gel retardation.

2. Materials and methods

2.1. Bacterial strains and plasmids

The *E. coli* strains used in this study are all derivatives of DS941, a *recF lacI^Q lacZ ΔM15* derivative of AB1157 [16]. DS9029 is an *xerC xerD* mutant of DS941 (DS941 *xerD::Tp^R xerC::miniMu PR13* [8] and DS9008 is DS941 *xerD2::Tn10-9* [1]. For routine plasmid transformations, DH5α was used and for overexpression studies, strain HMS174(DE3) [17]

was used. The *Proteus mirabilis* strain used is a clinical isolate (UM 240-852), and has been described previously [9].

For the cloning of PCR products, pTZ-PC was used (D.C. Tessier and D.Y. Thomas, personal communication). This vector is a pTZ-19R derivative containing two *XcmI* sites, which generates 5'T overhangs after digestion, allowing for the cloning of PCR products containing 3'A extensions. For complementation analysis, plasmid pCS210 was used, which is a pACYC184-derivative encoding Tc resistance and contains two directly repeated *cer* sites flanking a *lacZ* gene [18]. For control experiments, pRM130, a pUC19 derivative containing the *E. coli xerD* gene [1] was used. For overexpressing and purifying XerD, plasmid pET-14b (Novagen) was used. Finally, for the preparation of specific DNA-binding fragments, plasmid pKS492, a pUC18 derivative containing the 280 bp *HpaII-TaqI cer* fragment in the *AccI* site of the polylinker [3], was used.

2.2. Growth conditions and DNA manipulations

Bacteria were routinely grown in LB broth or plated on LB agar, containing the appropriate antibiotics when required. Ampicillin was used at 100 μg/ml and tetracycline at 6.25 μg/ml. Restriction enzymes were obtained from New England Biolabs (NEB), and were used according to the supplier's conditions. All routine DNA manipulations were performed as described in [19] and [20]. DNA fragments were extracted from agarose gels using the BandPrep kit from Pharmacia. Southern hybridisations were done according to Sirois and Szatmari [9].

2.3. PCR conditions

PCR reactions were carried out using a Biometra or a Hybaid thermocycler using Taq (Boehringer Mannheim), Tbr (Amresco) or Vent polymerase (NEB), according to the suppliers' conditions. Cycling conditions for degenerate primer PCR were 94°C/30 s, 37°C/60 s, and 72°C/2 min for 30 cycles with a final extension of 5 min at 72°C using Taq polymerase and 250 pmol of the following primers: for reverse side, (primer C) RTGNGTRTADA-TYTGNGT; and for forward side, (primer A1) GNGAYMRNGCNHTNYTNGARBT.

Inverse PCR was performed using a protocol adapted from E. Campeau (personal communication), and is as follows: 4.5 µg of genomic DNA was digested with 45 units of *Bgl*II in a total of 55 µl, for 16 h at 37°C. The DNA was extracted with chloroform-isoamyl alcohol, precipitated in 70% ethanol+250 mM NaCl at -70°C for 0.5 h and centrifuged at 12 500×g at room temperature for 20 min. The pellet was resuspended in 50 µl of water. Dilutions of 1:10, 1:5 and 7:10 were prepared from the resuspended pellet and distributed into three tubes. Each tube was incubated overnight at 16°C with 400 units of T4 DNA ligase (NEB), in a total reaction of 50 µl. An amount of 10 µl of each ligation was subsequently used for inverse PCR. Cycling conditions were then: 94°C/30 s, 56°C/60 s, 72°C/2 min 30 s for 30 cycles with a final extension of 30 min at 72°C (using Tbr polymerase), with 100 pmol of the following primers: (PrD) GTTTTCCCTTGCATAAGG for the 5' end of the coding region, and (mirD) GATGTTATTGGGG-CACAG for the 3' end of the coding sequence.

2.4. Sequence analysis

To clone and sequence the *xerD* gene, template DNA was prepared by performing PCR on genomic DNA using 100 pmol of the following primers: (PmdF) GCCGCCATTTCGGTATAACAG and (PmdR) CCCACTATCGGTGTCGGTTG, and cycling conditions: 95°C/15 s, 62°C/30 s, 72°C/45 s for 30 cycles, with a final extension at 72°C for 5 min, using Vent polymerase. The amplified product was then sequenced using an ABI 373 autosequencer with the following internal primers: (PmXDFseq) GCCAGTTGCGCTTTAATAGC and (PmXDRseq) GGTC ACTATTGTACCAG. The *xerD* sequence was submitted to GenBank (accession number, AF033497). A similarly amplified PCR product (using Taq polymerase with PmdF and PmdR) was also used to obtain a clone of *xerD* in plasmid pTZ-PC. This plasmid, pMVD37, was used in complementation studies.

2.5. Protein overexpression and DNA-binding assays

Using PCR, an *Nde*I site was engineered into the N-terminus of the coding region of the *xerD* gene of

P. mirabilis using primers: (PmdNdeI) GAGTAAA-CATATGACCCAGACC and PmdR. Cycling conditions were the same as those used for PmdF and PmdR, except that the annealing was at 54°C and Taq polymerase was used. The amplified gene was then cloned into pTZ-PC, then sub-cloned into pET-14b. The resultant clone, named pMVD72 was then transformed into HMS174(DE3). For expression, 2 ml of an overnight culture were used to inoculate 50 ml LB Ap medium. Cells were incubated at 37°C until a OD₆₀₀ of 0.6–0.7, then induced with 0.1 mM IPTG for 3 h. Cell extracts were lysed by freeze-thaw, cleared by centrifugation at 12 500×g at 4°C for 25 min, filtered through a 0.22 µm filter and passaged through a His-trap (Pharmacia) column and eluted according to the manufacturer's instructions. Proteins were separated by SDS-PAGE on 12.5% gels [21] and visualised by Coomassie blue staining.

Specific DNA binding was determined by the gel retardation assay [1] using specific fragments labelled by PCR using biotin or digoxigenin-labelled nucleotides and gel purified. DNA (3.5 ng), protein (240 ng) and polyIdC (1 µg) were used. The buffer used for DNA binding-assay was TE with 5% glycerol. Protein concentrations were estimated by the Bradford method [22] and DNA concentrations were estimated by absorbance at 260 nm. Reactions were incubated at 37°C for 30 min and the protein-DNA complexes were separated by non-denaturing polyacrylamide gel electrophoresis [20] in Tris-glycine buffer at 7.5 V/cm. After electrophoresis, the gels were blotted onto Appligene positively charged nylon membranes, UV crosslinked, and detected using the Phototope-Star Detection kit (NEB) for biotin-labelled DNA, or the DIG detection protocol for chemiluminescence (Boehringer Mannheim), for DIG-labelled DNA. In both cases, final detections were done with CDP-Star (NEB), followed by exposure to Kodak XAR5 X-ray film.

3. Results

3.1. Cloning of *xerD* from *P. mirabilis*

Degenerate primers were designed from two conserved regions of the Xer recombinases, correspond-

```

PmD 1 MTQTKLSTQTTTDDNNQEDNDVIEQFLDSIWLEQGLSANTLSAYRLDLQA 50
      . . . . . | | | | | . : | | | . | . | | | . | | | | |
EcD 1 . . . . . MKQDLARIEQFLDALWLEKNLAENTLNAYRRDLSM 35

PmD 51 LSQWLVTQKLNWLSVTTLDLHAFLATRLDEGYKATSAARLLSTLRRFFQY 100
      : : | | . | . . . | | | | | : | | | | | . | | | | |
EcD 36 MVEWLHHRGLTLATAQSDDLQALLAERLEGGYKATSSARLLSAVRRLFQY 85

PmD 101 LYREKLRQDDPSALLSTPKLPKRLPKDLSEQQVENLLSAPCIDEFIELRD 150
      | | | | | : | | | | | . | | | | | . | | | | | | | | | | | | | | |
EcD 86 LYREKLFREDDPSAHLASPKLPQRLPKDLSEAQVERLLQAPLIDQPLELRD 135

PmD 151 KAMLEVLYACGLRVSELVGLSLSDISLRQGVLRVIGKGDKERLVPLGEEA 200
      | | | | | | | | | | | | | | | . : | | | | | | | | | | | | | | |
EcD 136 KAMLEVLYATGLRVSELVGLTMSDISLRQGVVVRVIGKGNKERLVPLGEEA 185

PmD 201 IYWLEQYLQYGRPALMQGKTDDIVFPSLRGQKMTROTTFWHRIKHYAVIAG 250
      : | | | | | | | : : | | | | : | . | : . | | | | | . | | | | | | | | | | | | | | |
EcD 186 VYWLETYLEHGRPWLLNGVSDIVLFPSQRAQQMTRQTFWHRIKHYAVLAG 235

PmD 251 IDSEKLSPHVLRHAFATHLLNHGADLRVVQMLLGHSDLSTTQIYTHVATE 300
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
EcD 236 IDSEKLSPHVLRHAFATHLLNHGADLRVVQMLLGHSDLSTTQIYTHVATE 285

PmD 301 RLKVLHQQHHPRG 313
      | | : | | | | | | |
EcD 286 RLRQLHQQHHPRA 298

```

Fig. 1. Alignment of *E. coli* and *P. mirabilis* XerD proteins. Alignments were done using the Gap program from the Wisconsin package version 9.1 of the Genetics Computer Group (GCG), copyright 1982-1997. Vertical lines (|) represent identity; double dots (:) represent a high degree of similarity; and single dots (.) represent less similarity between amino acids. According to the value given by the program, the proteins have 74.2% identity. As revealed by the alignment, maximum identity is found in the C-terminal region of the proteins. PmD, *P. mirabilis* XerD; EcD, *E. coli* XerD.

ing to conserved domains I and II. In domain I, the sequence RDXAXLEXXYXXGXR is highly conserved in both XerC and XerD (see Fig. 7 from [14]), and a degenerate primer was constructed based on the sequence RDXAXL. The second primer was constructed based on the sequences surrounding the active site tyrosine in domain II (TQIYTH). These sequences are shown in Section 2. These primers were used on *P. mirabilis* chromosomal DNA, and amplified a DNA fragment of 450 bp, corresponding to the expected distance between domains I and II in the Xer recombinases. The fragment was cloned into the T-overhang cloning vector pTZ-PC and sequenced. The sequence obtained from this fragment showed strong similarity to the C-terminal region of the Xer recombinases, and inverse PCR primers were synthesised based on the sequence of the 450 bp amplicon. This amplicon was also used to probe digested *P. mirabilis* DNA, in order to select the choice

of an appropriate restriction enzyme for inverse PCR. The enzyme *Bg*/II yielded a restriction fragment around 2.1 kb that hybridised with the 450 bp amplicon (not shown), and this enzyme was used for preparing the template for inverse PCR.

Utilizing the inverse PCR primers, a DNA fragment around 2.1 kb was obtained, cloned and sequenced. This completed the sequence of *xerD* and yielded the sequence of the genes adjacent to *xerD* in the *P. mirabilis* chromosome. In order to correct for possible polymerase errors, another set of primers were synthesised corresponding to positions immediately upstream and downstream of the coding sequence of *xerD*. These primers were used to amplify the *xerD* coding region using Vent polymerase. The amplified product was purified using a Microcon-50 cartridge and sequenced using a set of internal sequencing primers, which allowed for the correction of four Tbr polymerase errors.

3.2. Characteristics of the sequence

The *xerD* gene of *P. mirabilis* codes for a 313 aa protein that displays a high degree of similarity with the sequenced *xerD* gene of other Enterobacteriaceae. It shows 74.2% identity and 79.2% similarity with *E. coli* XerD (see Fig. 1) and 73.15% identity with *S. typhimurium*. The regions of the proteins that are the most similar are, not surprisingly, at the carboxy terminus where the active site tyrosine resides. However, the *P. mirabilis* recombinase is significantly longer than its *E. coli* and *S. typhimurium* counterparts by 15 amino acids. When the proteins are aligned, the extended length is found at the amino terminus of *P. mirabilis* protein (Fig. 1). A greater degree of divergence is seen with *xerD* from *H. influenzae* (65.4% identity) and the *ripX* gene of *B. subtilis* (44.56% identity).

The *xerD* gene of *P. mirabilis* also shows a similar genetic organisation when compared to *E. coli* and

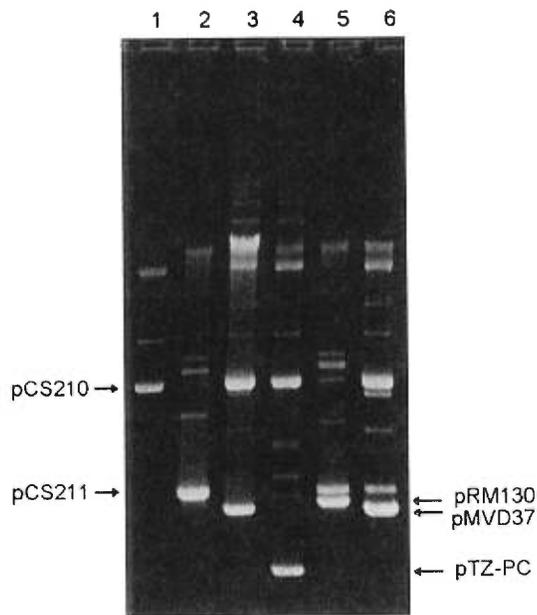


Fig. 2. Complementation analysis of an *E. coli* *xerD* mutant with *P. mirabilis* XerD. Lane 1, pCS210 (in DS9029 *xerC*⁻*xerD*⁻); lane 2, pCS211 (in DH5 α *xer*⁺) is pCS210 that has undergone *cer-cer* recombination; lane 3, pMVD37 in DS9029 that has a resident pCS210; lane 4, pTZ-PC in DS9008 *xerD*⁻ that has a resident pCS210; lane 5, pRM130 in DS9008 pCS210; lane 6, pMVD37 in DS9008 pCS210. pMVD37 is *P. mirabilis* *xerD* in pTZ-PC and pRM130 is *E. coli* *xerD* in pUC19 (see Section 2).

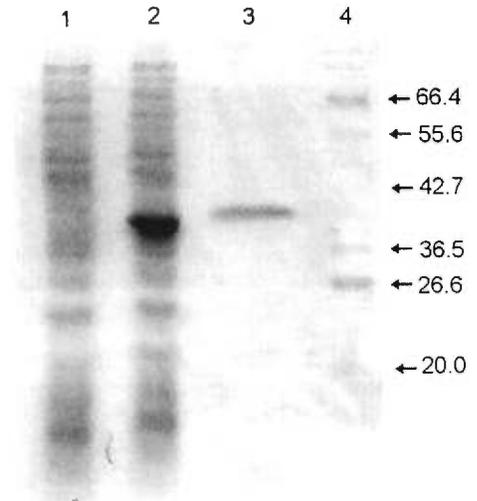


Fig. 3. Overexpression and purification of *P. mirabilis* XerD. Lane 1, non-induced crude extract; lane 2, 3 h induced crude extract; lane 3, partially purified XerD; lane 4, protein molecular weight standards. Expression and purification were done as indicated in Section 2.

S. typhimurium. In all three cases, XerD is part of an operon of at least three genes. Our preliminary sequencing data have indicated that the ORF immediately downstream of *xerD* displays significant similarity with the *E. coli* *xprA* (*dsbC*) gene [23]. An ORF with substantial similarity to the putative *E. coli* *fldB* is also found immediately upstream of the *P. mirabilis* *xerD* gene, and transcribed in the opposite orientation, in good agreement with the *fldB-xerD-xprA* gene order found in *E. coli* and *S. typhimurium* [13].

3.3. Complementation analysis

Upon site-specific recombination at the *cer*-sites, plasmid pCS210 loses the *lacZ* gene but retains the p15A plasmid origin and resistance to tetracycline (Tc) [18]. This can be observed *in vivo* by the presence of white colonies on X-gal Tc-containing media, or by the analysis of extracted plasmid DNA *in vitro*. Upon transformation of the *xerD* mutant strain DS9008 with plasmid pMVD37, containing the *P. mirabilis* *xerD* gene, blue colonies were obtained. Analysis of the plasmid DNA contained within these colonies showed the majority of the plasmid to be non-recombined, with a small amount of recombined

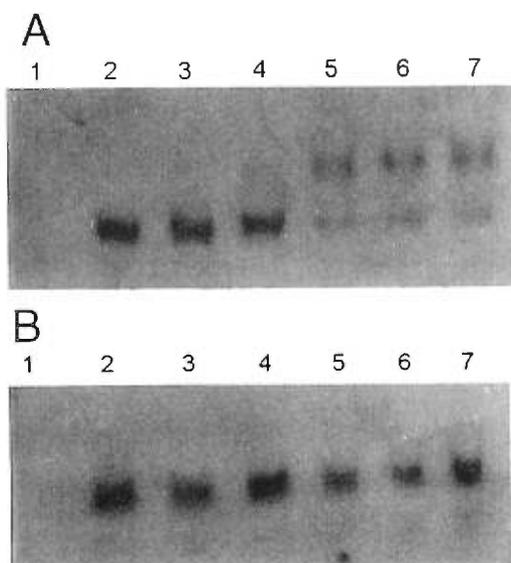


Fig. 4. DNA binding assay with 3.5 ng of ColE1 *cer* site (a) and 3.5 ng of *pepA* promoter region site (b). Lane 1, 240 ng of XerD without specific labelled DNA; lane 2, specific labelled DNA, *cer* site (a) or *pepA* promoter region (b) without any protein; lane 3, 30 ng of XerD with DNA; lane 4, 60 ng; lane 5, 120 ng; lane 6, 180 ng; lane 7, 240 ng.

product visible (Fig. 2, lane 6). This small amount of recombinant product was not observed when a control plasmid, pTZ-PC was present (Fig. 2, lane 4). To show that this small amount of recombination was not solely due to the overexpression of XerD acting alone at *cer*, plasmid pMVD37 was placed in DS9029, an *xerC**xerD* double mutant. No *cer* recombination was detected in this strain (Fig. 2, lane 3).

3.4. Expression and partial purification of the *P. mirabilis* XerD recombinase

The *xerD* coding region from *P. mirabilis* was cloned into pET-14b by modifying the 5' region of the gene to generate a *Nde*I site. This manipulation makes only minor changes in the N-terminus of the protein, and also adds a polyhistidine tail to this region. Upon transformation into HMS174(DE3), a protein with an apparent molecular weight of 39.6 kDa is observed after induction of a culture with 0.1 mM IPTG, a good agreement with the predicted molecular weight from the derived amino acid se-

quence of the protein which is 36.55 kDa. The expression and purification of the histidine-tagged XerD is shown in the Fig. 3.

The purified protein was utilised for DNA binding studies by the gel retardation assay, using the *cer* site present in plasmid pKS492 as a specific fragment, and the *E. coli pepA* promoter region as a non-specific fragment. Both fragments were labelled non-radioactively with either biotin or digoxigenin. Specific binding to the ColE1 *cer* site was observed with as little as 60 ng of protein, while equivalent amounts of protein did not retard the migration of *pepA* promoter fragment (Fig. 4).

4. Discussion

In this report, we describe the cloning, overexpression and purification of the *P. mirabilis xerD* gene and its product. The *P. mirabilis xerD* gene was cloned using degenerate primers corresponding to two highly conserved regions found in all Xer-type recombinases. Despite the high degree of degeneracy of the A1 primer (greater than 200 000), a product of the correct size and sequence was obtained under standard PCR conditions, thus indicating the usefulness of the degenerate PCR technique. The deduced amino acid sequence of the *P. mirabilis* XerD protein displays strong similarities to other sequenced XerD proteins, especially in the C-terminal region. Residues 123–313, which correspond to this region (see [24]), show 84% identity at the amino acid level with *E. coli* XerD, whereas the N-terminal domain displays only 56% identity. The gene appears to be organised as part of an operon analogous to the *E. coli recJ* operon [25], and also shows similar upstream and downstream genes in *S. typhimurium* [13]. No similarities are found with the flanking sequences of the *xerD* homologue of *B. subtilis* (also known as *ripX*) and *H. influenzae*.

The purified, histidine-tagged protein has a molecular weight of 39.6 kDa, in good agreement with its predicted molecular weight of 36.55 kDa. The purified protein displays strong, specific DNA-binding activity to the ColE1 *cer* site, showing retardation of 3.5 ng of *cer* DNA with as little as 60 ng of protein, comparable to the results of Blakely et al. [1]. Under similar conditions, no binding was ob-

served with a non-specific control fragment, the *E. coli pepA* promoter.

Surprisingly, the cloned *xerD* gene from *P. mirabilis* did not complement *E. coli xerD* mutants very effectively. This result is unexpected, considering that the cloned *sss* gene of *Pseudomonas aeruginosa* (an *xerC* homolog) was able to complement *E. coli xerC* mutants [10]. There are several possibilities as to why we did not observe efficient complementation: deficiencies in DNA binding, inefficient cooperativity of binding, inefficient cleavage, and inefficient activation. The first possibility can be ruled out, as we were able to demonstrate binding with the *P. mirabilis* XerD protein that was as effective as *E. coli* XerD binding (Fig. 4). The third and fourth possibilities are also unlikely for the following reasons: we were able to show efficient XerD recombination at *cer* with our clones when both the *P. mirabilis xerC* (GenBank accession number, AF033498) and *xerD* were co-expressed from the same plasmid (manuscript in preparation). Under these conditions, efficient *cer-cer* recombination (as measured by the resolution of the pCS210 plasmid) was observed in strain DS9029, an *xerC**xerD* mutant. Secondly, we can rule out inefficient activation of the *E. coli* XerC protein by the *P. mirabilis* XerD, because the regions implicated in this activation were mapped out by Spiers and Sherratt [26]: residues 263–268 of domain II of the *E. coli* XerD protein are either directly or indirectly implicated in its interaction with the XerC protein, and residues further toward the C-terminus are not necessary for this cooperativity. However, the C-termini of both *E. coli* and *P. mirabilis* show strong similarities in this region. The inability, or the reduced ability of *E. coli* XerC-*P. mirabilis* XerD recombination at *cer* suggests that these proteins do not interact very well with each other. Since the proteins are almost identical at their C-termini, the poor recombination activity of these proteins together must be due to differences in some other region(s). Furthermore, the region corresponding to α -helix L of XerD [24] implicated in XerC-XerD cooperativity [14] is identical to that of *E. coli*.

Thus, the inefficient complementation observed in our experiments may be due to inefficient cooperativity of binding of the XerC and XerD proteins to the *cer* site. The *P. mirabilis* protein is significantly longer than its *E. coli* counterpart, and this extra

length, which maps to the N-terminus, may present some steric hindrance to XerC binding at its site. Alternatively, some other protein-protein contacts may be involved. Spiers and Sherratt [26] found that deletions of the terminal 5 amino acids of the *E. coli* XerD protein resulted in a protein that was proficient for DNA binding, cooperativity and activation, but was still unable to perform site-specific recombination in vivo. As there are 3 amino acid differences in the last 10 amino acids of both XerD proteins, we cannot rule out this region as the cause of our poor complementation. Finally, other uncharacterised regions of dissimilarity could also account for this result. There is a large region of dissimilarity between residues 200 and 233 of *P. mirabilis* XerD (corresponds to 186–218 of *E. coli*).

Initial observations regarding the increased levels of multimerisation of ColE1 plasmids in *P. mirabilis* raised the possibility that this multimerisation may be due to inherent differences between the Xer proteins between these species [9]. Our initial results, however, do not lend support to this hypothesis. In the *E. coli xerC**xerD* double mutant, co-expression of the *P. mirabilis* Xer proteins does not result in any appreciable increase in plasmid multimerisation (manuscript in preparation). However, in order to completely rule out this possibility, one will have to construct *P. mirabilis xer* mutants, and observe the behaviour of ColE1 plasmids in these backgrounds. These experiments are currently being undertaken and should yield some interesting results on the stability of replicons in this bacterium.

Acknowledgments

The authors wish to thank Dr. David Thomas of the NRC Biotechnology Research Institute for plasmid pTZ-PC, and Dr. David Sherratt of Oxford University, for plasmid pRM130 and strains DS9008 and DS9029. We also wish to thank Patty Singer of the University of Maine for her assistance in performing automated sequence analysis. Eric Campeau of McGill University for his inverse PCR protocol, and Nancy St-Cyr and Urszula Wiatrowska for technical assistance. This work was supported by an operating grant from the Natural Sciences and Engineering Research Council of

Canada (NSERC). M.V. is supported by an NSERC graduate scholarship.

References

- [1] Blakely, G., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T. and Sherratt, D.J. (1993) Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell* 75, 351–361.
- [2] Cornet, F., Mortier, I., Patte, J. and Louarn, J.-M. (1994) Plasmid pSC101 harbors a recombination site, *psi* which is able to resolve plasmid multimers and to substitute for the analogous chromosomal *Escherichia coli* site *dif*. *J. Bacteriol.* 176, 3188–3195.
- [3] Stirling, C.J., Szatmari, G., Stewart, G., Smith, M.C.M. and Sherratt, D.J. (1988) The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus. *EMBO J.* 7, 4389–4395.
- [4] Stirling, C.J., Colloms, S.D., Collins, J.F., Szatmari, G. and Sherratt, D.J. (1989) *XerB*, an *Escherichia coli* gene required for plasmid ColE1 site-specific recombination, is identical to *pepA*, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. *EMBO J.* 8, 1623–1627.
- [5] Blakely, G., Colloms, S., May, G., Burke, M. and Sherratt, D.J. (1991) *Escherichia coli* XerC recombinase is required for chromosomal segregation at cell division. *New Biol.* 3, 789–798.
- [6] Blakely, G.W. and Sherratt, D.J. (1994) Interactions of the site-specific recombinases XerC and XerD with the recombination site *dif*. *Nucleic Acids Res.* 22, 5613–5620.
- [7] Arciszewska, L.K. and Sherratt, D.J. (1995) Xer site-specific recombination *in vitro*. *EMBO J.* 14, 2112–2120.
- [8] Colloms, S.D., McCulloch, R., Grant, K., Neilson, L. and Sherratt, D.J. (1996) Xer-mediated site-specific recombination *in vitro*. *EMBO J.* 15, 1172–1181.
- [9] Sirois, S. and Szatmari, G. (1995) Detection of XerC and XerD recombinases in Gram-negative bacteria of the family *Enterobacteriaceae*. *J. Bacteriol.* 177, 4183–4186.
- [10] Höfte, M., Dong, Q., Kourambas, S., Krishnapillai, V., Sherratt, D. and Mergeay, M. (1994) The *sss* gene product, which affects pyoverdinin production in *Pseudomonas aeruginosa* 7NSK2, is a site-specific recombinase. *Mol. Microbiol.* 14, 1011–1020.
- [11] Slack, F.J., Serron, P., Joyce, E. and Sonenshein, A.L. (1995) A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Mol. Microbiol.* 15, 689–702.
- [12] Becker, J. and Brendel, M. (1996) Molecular characterization of the *xerC* gene of *Lactobacillus leichmannii* encoding a site-specific recombinase and two adjacent heat shock genes. *Curr. Microbiol.* 32, 232–236.
- [13] Hayes, F., Lubetzki, S.A. and Sherratt, D.J. (1997) *Salmonella typhimurium* specifies a circular chromosome dimer resolution system which is homologous to the Xer site-specific recombination system of *Escherichia coli*. *Gene* 198, 105–110.
- [14] Cao, Y., Hallet, B., Sherratt, D.J. and Hayes, F. (1997) Structure–function correlations in the XerD site-specific recombinase revealed by pentapeptide scanning mutagenesis. *J. Mol. Biol.* 274, 39–53.
- [15] Goebel, W. and Helinski, D.R. (1968) Generation of higher multiple circular DNA forms in bacteria. *Proc. Natl. Acad. Sci. USA* 61, 1406–1413.
- [16] Summers, D.K. and Sherratt, D.J. (1988) Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional of the *cer* site. *EMBO J.* 7, 851–858.
- [17] Campbell, J.L., Richardson, C.C. and Studier, F.W. (1978) Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA. *Proc. Natl. Acad. Sci. USA* 75, 2276–2280.
- [18] Stirling, C.J., Stewart, G. and Sherratt, D.J. (1988) Multicopy plasmid stability in *Escherichia coli* requires host-encoded functions that lead to plasmid site-specific recombination. *Mol. Gen. Genet.* 214, 80–84.
- [19] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- [20] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987–1997) *Current Protocols in Molecular Biology*. Wiley, New York.
- [21] Dreyfuss, G., Adam, S.A. and Choi, Y.D. (1984) Physical change in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. *Mol. Cell. Biol.* 4, 415–423.
- [22] Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.
- [23] Missiakas, D., Georgopoulos, C. and Raina, S. (1994) The *Escherichia coli* *dsbC* (*xprA*) gene encodes a periplasmic protein involved in disulfide bond formation. *EMBO J.* 13, 2013–2020.
- [24] Subramanya, H.S., Arciszewska, L.K., Baker, R.A., Bird, L.E., Sherratt, D.J. and Wigley, D.B. (1997) Crystal structure of the site-specific recombinase, XerD. *EMBO J.* 16, 5178–5187.
- [25] Lovett, S.T. and Kolodner, R.D. (1991) Nucleotide sequence of the *Escherichia coli* *recJ* chromosomal region and construction of RecJ-overexpressing plasmids. *J. Bacteriol.* 173, 353–364.
- [26] Spiers, A.J. and Sherratt, D.J. (1997) Relating primary structure to function in the *Escherichia coli* XerD site-specific recombinase. *Mol. Microbiol.* 24, 1071–1082.

CHAPITRE III :

DISCUSSION

Chapitre III- Discussion

Malgré l'observation que ColE1 (Goebel et Helinski en 1968) et pNCK-1 (possédant les sites *cer* et *nmr*; Sirois et Szatmari, 1995) multimérisent dans *Proteus mirabilis*, cette espèce possède bel et bien deux gènes codant pour deux protéines fortement similaires aux recombinases XerC et XerD retrouvées chez *E. coli*. L'article présenté dans la section précédente relate le clonage et la caractérisation de *xerD*.

1. Caractérisation de la séquence

Comme il a été mentionné, le clonage de ce gène s'est fait grâce à l'utilisation d'amorces fortement dégénérées, lesquelles ont été conçues à partir des motifs conservés chez les membres de la famille des intégrases. La première amorce (désignée A1), correspond à une partie du motif I, tandis que la seconde amorce (désignée C) correspond au motif TQIYTH situé en C-terminal des protéines. Ces deux amorces ont permis d'obtenir un fragment interne de 450 pb qui semblait plus désigner une protéine de type XerD que de type XerC, de par le pourcentage d'identité avec cette région de XerD. Par la suite, des amorces spécifiques ont été désignées à partir de la séquence de ce fragment, pour permettre l'obtention du reste du gène par PCR inverse.

La séquence nucléotidique ainsi obtenue est de 42% en GC, ce qui correspond assez bien au pourcentage de l'ensemble du génome de *P. mirabilis* (évalué à $39,3 \pm 1,4\%$) (Holt, 1984-1989).

La séquence en acides aminés est fortement identique à celle de XerD d'*E. coli*: elle montre 74% d'identité globalement. Comme on pouvait s'y attendre, la portion la plus identique se situe dans la région C-terminale de la protéine, où l'on retrouve le site catalytique des intégrases. Le second domaine, correspondant aux résidus 108 à 298 chez *E. coli* (123 à 313 chez *P. mirabilis*), montre 84% d'identité, tandis que le domaine N-terminal n'en montre que 56%. De plus, la séquence de *P.*

mirabilis est plus longue de 15 acides aminés, lesquels se situent tous dans la portion N-terminale. Elle présente de plus les quatre résidus conservés des membres de la famille des intégrases (RHRY) (Argos *et al.*, 1986; Abremski et Hoess, 1992), ainsi que le motif GKGXKXR présent chez les recombinases de type Xer (Cao *et al.*, 1997). La séquence de *P. mirabilis* montre également une forte identité au XerD de *S. typhimurium* (73%), mais un peu moins avec celui d'*H. influenzae* (65%) ou RipX de *B. subtilis* (44,6%).

Le pourcentage d'identité avec la protéine XerC d'*E. coli* correspond à celui retrouvé entre les séquences de XerC et de XerD d'*E. coli*, soit d'environ 37%. Cela nous permet donc de croire fortement que la séquence identifiée pour *P. mirabilis* représente bel et bien une recombinase de type XerD.

De plus, l'organisation génomique semble être similaire à celle retrouvée chez *E. coli* et chez *S. typhimurium* (voir la figure présentée en annexe 2). Le PCR inverse a permis l'obtention d'une partie des séquences flanquantes de *xerD*. En aval de *xerD*, la séquence semble correspondre à *xprA* et une partie d'un autre cadre de lecture semble correspondre au début du gène codant pour RecJ. Immédiatement en amont, en position inverse, le début de la séquence d'un gène pourrait correspondre à *fldB*. Comme mentionné dans l'article, seule la séquence de *xerD* a été corrigée. Ainsi, les pourcentages des séquences flanquantes sont approximatifs, car les séquences n'ont pas été vérifiées et qu'il peut y avoir eu quelques erreurs créées par la polymérase Tbr employée. Cependant, les séquences obtenues permettent quand même de constater que *xerD* de *P. mirabilis* semble faire partie d'un opéron similaire à celui d'*E. coli*.

2. Expression de XerD

La protéine XerD de *P. mirabilis* a été surexprimée et purifiée grâce au vecteur pET-14b de Novagen, qui place en N-terminal de la protéine une queue d'histidine. Cela facilite la purification qui se fait par la suite à l'aide de colonnes d'affinité, tout en effectuant de faibles modifications à l'ensemble de la structure

protéique (du moins c'est ce que nous espérons!). De plus, la position en N-terminal de la protéine nous évite des désagréments potentiels compte tenu de l'emplacement du site catalytique en C-terminal des recombinaases de type Xer. D'ailleurs, des fusions avec la protéine liant le maltose (MBP) avaient été faites en N-terminal des protéines d'*E. coli*, et ce, sans que cela n'ait affecté les capacités de liaison ou de clivage (Arciszewska *et al.*, 1995).

Cela a permis l'obtention d'une protéine de 39,6 kDa. Comme il est mentionné au chapitre II, ce poids moléculaire correspond à celui qui était attendu, lequel est bien entendu plus grand que celui de XerD d'*E. coli* qui fait 33,8 kDa. Elle a été purifiée par la suite pour des études de liaison à l'ADN.

3. Études de complémentation et de retardement sur gel

Compte tenu du fort pourcentage d'identité entre les XerD de *P. mirabilis* et d'*E. coli*, on pouvait s'attendre à ce que le XerD de *P. mirabilis* soit apte à compléter une souche d'*E. coli* mutante pour le gène codant cette protéine. Elle aurait ainsi le potentiel d'interagir avec XerC d'*E. coli*, pour effectuer la recombinaison entre deux sites *cer* situés sur un même vecteur. Cependant, les tests de complémentation n'ont pas été aussi concluants que ceux faits avec les protéines de *S. typhimurium* (Hayes *et al.*, 1997). Une certaine complémentation a été observée, certes, mais nullement complète (chapitre II, fig. 2).

Plusieurs raisons pouvaient expliquer ce résultat: il pouvait y avoir des problèmes au niveau de la liaison à l'ADN, au niveau du clivage en tant que tel, ou de l'activation de XerD, ou encore, un problème de coopérativité avec XerC d'*E. coli*.

3.1 Liaison à l'ADN

Comme mentionné dans l'introduction, un modèle de la liaison de XerD à l'ADN a été déduit à partir de la structure de la protéine Cap liée à son site. Chez XerD, les hélices αG et αJ du domaine C-terminal formeraient le pseudo-motif hélice-

tour-hélice correspondant à ce qui est trouvé chez la protéine Cap (Subramanya *et al.*, 1997) et elles auraient un rôle dans la liaison de l'ADN. Les séquences de ces deux hélices sont conservées chez *P. mirabilis* (il n'y a qu'un seul acide aminé différent dans l'hélice α J).

La liaison à l'ADN impliquerait également le domaine N-terminal, comme cela a été révélé par le potentiel électrostatique de la structure de XerD (Subramanya *et al.*, 1997). Les hélices α B et α D seraient responsables de cette action. Il y a ici un peu plus de divergence entre les deux espèces.

Cependant, la liaison à l'ADN a pu être vérifiée par un test de retardement sur gel. Nous avons pu montrer que la protéine de *P. mirabilis* a été capable de retarder spécifiquement le site *cer* de ColE1, qui avait été marqué à la digoxygénine. Aux mêmes concentrations, la protéine n'a pu retarder un fragment d'ADN représentant le promoteur de PepA (chapitre II, fig. 4).

Le problème de complémentation ne semble donc pas se situer au niveau d'une liaison inefficace au site *cer*.

3.2 Clivage et activation

Un autre test de complémentation a ensuite été effectué en mettant les deux gènes de *P. mirabilis* sur un même vecteur, dans une souche d'*E. coli* mutante pour les deux recombinaisons. Contrairement à ce qui avait été observé auparavant, la complémentation a été ici complète. Le pCS210 (vecteur qui contient le gène *lacZ*, flanqué des deux sites *cer*; Stirling *et al.*, 1988a) a été résolu (fig. 12), c'est-à-dire que le gène *lacZ* a été perdu grâce à la recombinaison dépendante du système Xer, aux sites *cer*. Cela a pu être observé *in vivo* par l'apparition de colonies blanches sur un milieu contenant du X-gal, et *in vitro*, par l'obtention d'un vecteur de plus faible poids moléculaire; le pCS210 avait alors perdu 3 kb (résultat non montré).

Ainsi, ce test nous a permis de penser que les deux recombinases de *P. mirabilis*, une fois mises ensemble, étaient bien aptes à effectuer la recombinaison. XerD pouvait donc fort probablement activer la réaction pour que XerC effectue le clivage des premiers brins, tandis que XerD devait sûrement être apte pour effectuer le second clivage advenant le cas où ce serait effectivement lui qui exercerait cette action. En effet, il ne faut

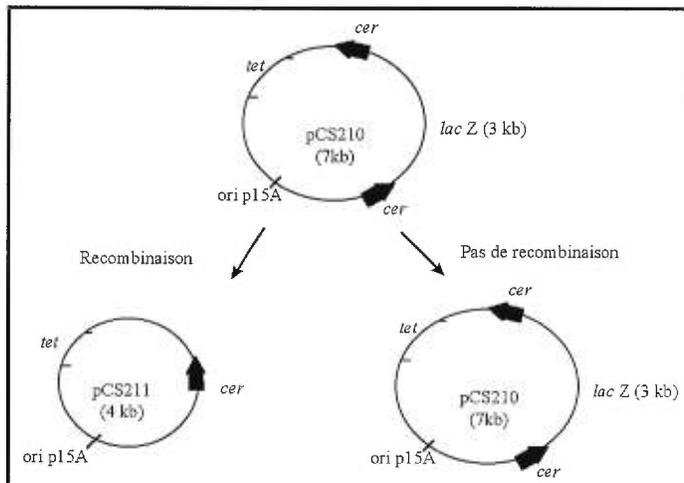


Figure 12- Vecteur utilisé pour les études de complémentation. Le pCS210 est un dérivé de pACYC184. Il contient deux sites *cer* qui flanquent le gène *lacZ*. Lorsqu'il y a recombinaison, ce gène est perdu et le vecteur ainsi obtenu (nommé pCS211) fait 3 kb de moins. Il ne possède alors qu'un seul site *cer*. Le vecteur conserve le gène de la résistance à la tétracycline (*tet*) et l'origine de répllication provenant de p15A (*ori p15A*). Lorsqu'il n'y a pas de recombinaison, le pCS210 est conservé intégralement.

pas oublier que le clivage du site *cer* par XerD d'*E. coli* n'a pas pu être montré et qu'il avait alors été supposé qu'un système autre que Xer devait compléter cette réaction (Colloms *et al.*, 1996). Il en est peut-être de même avec *P. mirabilis*, mais on peut au moins croire que XerD de cette espèce est quand même responsable de l'activation de la réaction, comme c'est le cas chez *E. coli*. Une interprétation des résultats obtenus par Spiers et Sherratt (1997) est que la communication avec XerC requise pour la catalyse et l'échange des brins impliquerait les résidus 278 à 283. Ces résidus sont absolument conservés chez *P. mirabilis*. Cette fonction ne doit donc pas être perturbée au niveau du premier test de complémentation.

Il est mentionné ici que nous avons employé les deux gènes de *P. mirabilis*. La séquence en acides aminés de XerC est présentée en annexe V et elle montre 37% d'identité avec la séquence en acides aminés de XerD de la même espèce. L'obtention du gène s'est fait de la même façon que pour *xerD*, en ce sens que les amorces dégénérées A1 et C ont mené à l'amplification d'un fragment de 450 pb. Cette fois-là

cependant, la séquence semblait plus correspondre à un *xerC* qu'à un *xerD*, et le PCR inverse fait à l'aide d'amorces spécifiques à cette séquence, a permis de confirmer ce soupçon. De plus, l'organisation génomique de *xerC* semble être la même que celle retrouvée chez *E. coli* (annexe II). Le gène *xerC* ainsi amplifié, puis cloné, a ensuite pu être utilisé pour tester son action *in vivo* avec *xerD* de la même espèce dans une souche DS9029 (*E. coli xerC⁻, xerD⁻*). Les tests ont donné les résultats mentionnés au début de cette section.

3.3 Coopérativité

La complémentation faible obtenue entre un XerD de *P. mirabilis* dans une souche d'*E. coli xerD⁻*, semblerait alors plus indiquer que l'interaction entre protéines d'espèces différentes s'effectue mal. Par contre, il faut noter que ce résultat ne correspond pas à ce qui a été observé avec d'autres espèces. Ainsi, les recombinases de *S. typhimurium* sont aptes à compléter des souches mutantes d'*E. coli* pour ces gènes (Hayes *et al.*, 1997). Même Sss de *P. aeruginosa* a pu compléter une souche mutante d'*E. coli* pour *xerC* (Höfte *et al.* 1994) et ce, malgré la plus grande divergence entre ces espèces, qu'entre les deux *Enterobacteriaceae* employées dans notre étude.

Les résidus identifiés comme ayant un rôle à jouer au niveau de la coopérativité entre les deux recombinases (résidus 263 à 268 chez *E. coli*) (Spiers et Sherratt, 1997) sont absolument conservés chez *P. mirabilis*. Il avait été ajouté que si le manque de coopérativité entre les deux recombinases était dû à l'ajout du fragment oméga qu'ils ont employé pour muter XerD, et non pas à la perte de ces résidus, la région impliquée s'étendrait des résidus 244 à 262. Même si cela s'avérait exact, cela ne peut pas expliquer le manque de coopérativité entre les deux recombinases d'espèces différentes, puisque toute cette région est absolument conservée chez les deux espèces dont il est question ici. En fait, une seconde région impliquée dans la coopérativité contient les résidus 256 à 258 (Subramanya *et al.*, 1997).

Selon le modèle fait par Subramanya *et al.* (1997) à partir de la structure de XerD, l'hélice située en extrémité C-terminale, laquelle contient la tyrosine catalytique, serait la plus importante pour les interactions XerC-XerD. La région correspondant à cette hélice et à la suite terminale de cette hélice ne contient que trois acides aminés différents chez *P. mirabilis*, mais cela peut malgré tout avoir influencé le niveau de faible complémentation observée dans nos études. Et même, l'un de ces résidus est le dernier de la protéine ; il est supposé que les cinq derniers résidus de XerD seraient importants pour la formation du complexe synaptique (Spiers et Sherratt, 1997).

De plus, les études de Cao *et al.* (1997) qui ont inséré des pentapeptides en différents endroits de XerD d'*E. coli* pour voir la fonction possible des régions où il y a eu insertion, ont montré qu'il y avait perte de coopérativité non seulement lorsque l'insertion se faisait dans le 256^e résidu, mais également lorsqu'elle est près de la position de l'arginine conservée dans le second motif, soit R247. La position 26 du domaine N-terminal était également affectée à ce niveau et ce, malgré que les fonctions de liaison à l'ADN et l'activation ne soient pas touchées (voir annexe III). Cette dernière position, correspondant à un résidu dans l'hélice α B, est, tout comme les deux autres, conservée chez *P. mirabilis*.

Il existe également des régions probablement impliquées spécifiquement à *cer* et d'autres pour la recombinaison à *dif*, puisque certaines insertions n'ont affecté que l'un des deux types de recombinaison (Cao *et al.*, 1997). Ainsi, la recombinaison au site *dif* était affectée (malgré que la liaison et la coopérativité étaient observées) lorsqu'une insertion était au niveau des résidus 172 et 175, mais celle au site *cer* n'était nullement affectée.

Au contraire, seule la recombinaison au site *cer* a été touchée lorsque les insertions étaient dans les résidus 59, 76, 114, 118 et 256 (annexe III). Le résidu 59 se situe dans l'hélice α C du domaine N-terminal; bien que ce résidu en tant que tel soit conservé chez *P. mirabilis*, la région qui l'entoure présente quelques différences. Si cette hélice est impliquée dans les interactions spécifiques avec XerC, ces divergences pourraient peut-être expliquer la mauvaise complémentation observée au

niveau des sites *cer* dans la figure 2 (puits 6; chapitre II). La même situation est retrouvée avec l'insertion au niveau du résidu 76 de l'hélice α D. Dans ce cas cependant, la liaison à l'ADN serait plus affectée si l'on en croit le potentiel électrostatique de la structure de XerD (Subramanya *et al.*, 1997). Enfin, les régions dans lesquelles se situent les trois autres points d'insertion sont conservées pour les deux espèces et ne marquent donc pas la spécificité des recombinases de chaque espèce.

Comme la plus grande divergence entre les deux protéines se situent au niveau N-terminal, il est peut-être possible qu'un domaine dans cette région serait responsable en partie du contact entre les deux protéines.

De plus, l'excédent de 15 acides aminés chez XerD *P. mirabilis* pourrait peut-être provoquer un encombrement stérique qui empêcherait un contact efficace de XerC d'*E. coli* à son site de liaison. Cet encombrement pourrait modifier la structure globale de la protéine qui pourrait faire en sorte de réduire la qualité des interactions entre les deux protéines. Malgré que les deux séquences soient fortement similaires, cela ne signifie pas forcément que les structures des deux XerD soient parfaitement identiques et cela pourrait être suffisant pour réduire l'efficacité de la recombinaison. La structure de XerD révèle que la tyrosine catalytique est camouflée par le domaine N-terminal et qu'un changement conformationnel serait nécessaire pour la rendre disponible à son site d'action (Subramanya *et al.*, 1997). Il serait alors possible que le changement conformationnel nécessaire, probablement induit par XerC, ne puisse être fait quand le XerC provient d'une autre espèce que *P. mirabilis*.

Il est à noter que le XerC de *P. mirabilis*, fait exactement le même nombre d'acides aminés que le XerD de la même espèce. Ceci pourrait favoriser la coopération entre les deux recombinases de la même espèce. L'encombrement stérique peut-être éprouvé à cause de cette longueur excédentaire ne se retrouverait pas chez Sss de *P. aeruginosa* (Höfte *et al.*, 1994) ni chez XerC de *S. marcescens* (Gwynn *et al.*, 1998) lesquels ont été capables de montrer une activité de recombinaison avec le XerD provenant d'*E. coli*. D'ailleurs, ces recombinases ne

montrent qu'une très légère différence dans la longueur au niveau N-terminal, de l'ordre de trois ou quatre acides aminés en moins ou en plus respectivement, par rapport à XerC d'*E. coli* (voir l'annexe IV). Il faudrait vérifier cette hypothèse par l'étude de la complémentation entre une protéine XerC ou XerD d'*E. coli* dont l'extrémité N-terminale aurait une dizaine de résidus en plus et ainsi voir si le manque de complémentation est réellement dû à un encombrement stérique. Il faudrait également faire une étude par mutagenèse dirigée, afin de cibler d'autres régions qui seraient impliquées dans la spécificité de la coopérativité entre les deux types de recombinases chez *P. mirabilis*.

Enfin, même si le domaine C-terminal est très bien conservé dans les deux espèces, il y a quand même une région qui montre un peu plus de dissimilarités. Cette région s'étend des résidus 201 à 232 inclusivement (186 à 217 chez *E. coli*); elle pourrait peut-être aussi expliquer le résultat du test de complémentation.

4. À suivre...

Pour expliquer la multimérisation excessive de ColE1 (Goebel et Helinski, 1968) et celle de pNCK-1 (Sirois et Szatmari, 1995) dans les souches de *P. mirabilis*, l'hypothèse avait été posée que les séquences entre les deux espèces étaient probablement très divergentes. Une hybridation avec une sonde faite du gène *xerD* d'*E. coli* n'avait d'ailleurs pas permis de détecter sa présence chez l'espèce étudiée ici (Sirois et Szatmari, 1995). Cependant, les recombinases existent bel et bien chez *P. mirabilis* et il ne semble pas qu'elles causent une multimérisation plus importante du vecteur rapporteur pCS210 dans les souches d'*E. coli* mutantes employées dans les tests de complémentation. Bien entendu, comme il est mentionné à la fin de l'article, il reste encore beaucoup à faire pour mieux comprendre les différences qui existent entre les recombinases de type Xer des deux espèces. Ainsi, des mutagenèses aléatoires et spécifiques permettraient de cartographier certaines autres régions impliquées dans les interactions protéiques. Il faudrait également comprendre exactement comment le système agit au sein de la cellule bactérienne, et savoir quels sont les facteurs qui peuvent influencer l'expression de ces protéines. Compte tenu de la présence qui

semble ubiquitaire chez l'ensemble des espèces bactériennes, il serait intéressant d'en comprendre les spécificités pour chaque espèce. Ainsi, cela permettrait peut-être de comprendre pourquoi le plasmide ColE1 s'est avéré instable chez *P. mirabilis* (Goebel et Helinski, 1968), alors que le système existe dans cette espèce et qu'il a pu résoudre un vecteur contenant deux sites *cer*, dans une souche d'*E. coli* mutante pour les deux gènes. Serait-il possible, par exemple, que les facteurs accessoires soient inefficaces pour la fonction de résolution au sein de *P. mirabilis* ? À suivre donc !...

BIBLIOGRAPHIE

Bibliographie

- Abremski, K., Wierzbicki, A., Frommer, B. et R.H. Hoess (1986). Bacteriophage P1 Cre-*loxP* site-specific recombination : Site-specific DNA topoisomerase activity of the Cre recombination protein. *J. Biol. Chem.* **261** : 391-396.
- Abremski, K.E. et R.H. Hoess (1992). Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng.* **5** : 87-91.
- Àlen, C., Sherratt, D.J. et S.D. Colloms (1997). Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination. *EMBO J.* **16** : 5188-5197.
- Arciszewska, L.K. et D.J. Sherratt (1995). Xer site-specific recombination *in vitro*. *EMBO J.* **14** : 2112-2120.
- Arciszewska, L., Grainge, I. et D. Sherratt (1995). Effects of Holliday junction position on Xer-mediated recombination *in vitro*. *EMBO J.* **14** : 2651-2660.
- Arciszewska, L.K., Grainge, I. et D.J. Sherratt (1997). Action of site-specific recombinases XerC and XerD on tethered Holliday junctions. *EMBO J.* **16** : 3731-3743.
- Argos, P., Landy, A., Abremski, K., Egan, J.B., Haggard-Ljungquist, E., Hoess, R.H., Kahn, M.L., Kalionis, B., Narayana, S.V.L., Pierson III, L.S., Sternberg, N. et J.M. Leong (1986). The integrase family of site-specific recombinases : regional similarities and global diversity. *EMBO J.* **5** : 433-440.
- Asueda, H., Horii, T. et T. Itoh (1989). Structural and functional organization of ColE2 and ColE3 replicons. *Mol. Gen. Genet.* **215** : 209-216.
- Azaro, M.A. et A. Landy (1997). The isomeric preference of Holliday junctions influences resolution bias by λ integrase. *EMBO J.* **16** : 3744-3755.
- Becker, J. et M. Brendel (1996). Molecular characterization of the *xerC* gene of *Lactobacillus leichmannii* encoding a site-specific recombinase and two adjacent heat shock genes. *Curr. Microbiol.* **32** : 232-236.
- Black, D.S., Kelly, A.J., Mardis, M.J. et H.S. Moyed (1991). Structure and organization of *hip*, an operon that affects lethality due to inhibition of peptidoglycan or DNA synthesis. *J. Bacteriol.* **173** : 5732-5739.
- Blake, J.A.R., Ganguly, N. et D.J. Sherratt (1997). DNA sequence of recombinase-binding sites can determine Xer site-specific recombination outcome. *Mol. Microbiol.* **23** : 387-398.

Blakely, G., Colloms, S., May, G., Burke, M. et D. Sherratt (1991). *Escherichia coli* XerC recombinase is required for chromosomal segregation at cell division. *New Biol.* **3** : 789-798.

Blakely, G., May, G., McCulloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T. et D.J. Sherratt (1993). Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E.coli* K12. *Cell* **75** : 351-361.

Blakely, G.W. et D.J. Sherratt (1994). Interactions of the site-specific recombinases XerC and XerD with the recombination site *dif*. *Nucleic Acids Res.* **22** : 5613-5620.

Blakely, G.W. et D.J. Sherratt (1996a). *Cis* and *trans* in site-specific recombination. *Mol. Microbiol.* **20** : 233-238.

Blakely, G. et D. Sherratt (1996b). Determinants of selectivity in Xer site-specific recombination. *Genes Dev.* **10** : 762-773.

Blakely, G.W., Davidson, A.O. et D.J. Sherratt (1997). Binding and cleavage of nicked substrates by site-specific recombinases XerC and XerD. *J. Mol. Biol.* **265** : 30-39.

Burke, M., Merican, A.F. et D.J. Sherratt (1994). Mutant *Escherichia coli* arginine repressor proteins that fail to bind L-arginine, yet retain the ability to bind their normal DNA-binding sites. *Mol. Microbiol.* **13** : 609-618.

Calvin/Koons, M.D. et R.B. Blumenthal (1995). Characterization of *pPvu1*, the autonomous plasmid from *Proteus vulgaris* that carries the genes of the *PvuII* restriction-modification system. *Gene* **157** : 73-79.

Cannon, P.M. et P. Strike (1992). Complete nucleotide sequence and gene organization of plasmid NTP16. *Plasmid* **27** : 220-230.

Cao, Y., Hallet, B., Sherratt, D.J. et F. Hayes (1997). Structure-function correlations in the XerD site-specific recombinase revealed by pentapeptide scanning mutagenesis. *J. Mol. Biol.* **274** : 39-53.

Charlier, D., Hassanzadeh Gh., G., Kholti, A., Gigot, D., Piérard, A. et N. Glansdorff (1995). *carP*, involved in pyrimidine regulation of the *Escherichia coli* carbamoylphosphate synthetase operon encodes a sequence-specific DNA-binding protein identical to XerB and PepA, also required for resolution of ColE1 multimers. *J. Mol. Biol.* **250** : 392-406.

Chen, S.-H., Merican, A.F. et D.J. Sherratt (1997). DNA binding of *Escherichia coli* arginine repressor mutants altered in oligomeric state. *Mol. Microbiol.* **24** : 1143-1156.

Cheng, C., Kussie, P., Pavletich, N. et S. Shuman (1998). Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell* **92** : 841-850.

Clerget, M. (1991). Site-specific recombination promoted by a short DNA segment of plasmid R1 and by a homologous segment in the terminus region of the *Escherichia coli* chromosome. *New Biol.* **3** : 780-788.

Colloms, S.D., Sykora, P., Szatmari, G. et D.J. Sherratt (1990). Recombination at ColE1 *cer* requires the *Escherichia coli xerC* gene product, a member of the lambda integrase family of site-specific recombinases. *J. Bacteriol.* **172**: 6973-6980.

Colloms, S.D., McCulloch, R., Grant, K., Neilson, L. et D.J. Sherratt (1996). Xer-mediated site-specific recombination *in vitro*. *EMBO J.* **15** : 1172-1181.

Colloms, S.D., Bath, J. et D.J. Sherratt (1997). Topological selectivity in Xer site-specific recombination. *Cell* **88** : 855-864.

Colloms, S.D., Alèn, C. et D.J. Sherratt (1998). The ArcA/ArcB two-component regulatory system of *Escherichia coli* is essential for Xer site-specific recombination at *psi*. *Mol. Microbiol.* **28** : 521-530.

Cornet, F., Mortier, I., Patte, J., J.-M. Louarn (1994). Plasmid pSC101 harbors a recombination site, *psi*, which is able to resolve plasmid multimers and to substitute for the analagous chromosomal *Escherichia coli* site *dif*. *J. Bacteriol.* **176** : 3188-3195.

Cornet, F., Louarn, J., Patte, J. et J.-M. Louarn (1996). Restriction of the activity of the recombination site *dif* to a small zone of the *Escherichia coli* chromosome. *Genes Dev.* **10** : 1152-1161

Craig, N.L. et H.A. Nash (1983). The mechanism of phage λ site-specific recombination : site-specific breakage of DNA by Int topoisomerase. *Cell* **35** : 795-803.

Cornet, F., Hallet, B. et D.J. Sherratt (1997). Xer recombination in *Escherichia coli* : Site-specific DNA topoisomerase activity of the XerC and XerD recombinases. *J. Biol. Chem.* **272** : 21927-21931.

Dai, H., Chow, T.-Y., Liao, H.-J., Chen, Z.-Y. et K.-S. Chiang (1988). Nucleotide sequences involved in the neolysogenic insertion of filamentous phage Cfl6-v1 into the *Xanthomonas campestris* pv. *citri* chromosome. *Virology* **167** : 613-620.

Dekkers, L.C., Phoelich, C.C., van der Fits, L. et B.J.J. Lugtenberg (1998). A site-specific recombinase is required for competitive root colonization by *Pseudomonas fluorescens* WCS365. *Proc. Natl. Acad. Sci. USA* **95** : 7051-7056.

Duckett, D.R., Murchie, A.I.H., Giraud-Panis, M.-J.E., Pöhler, J.R. et D.M.J. Lilley (1995). Structure of the four-way junction and its interaction with proteins. *Phil. Trans. R. Soc. Lond.* **347** : 27-36.

Esposito, D. et J.J. Scocca (1997). The integrase family of tyrosine recombinases : evolution of a conserved active site domain. *Nucleic Acids Res.* **25** : 3605-3614.

Goebel, W. et D.R. Helinski (1968). Generation of higher multiple circular DNA forms in bacteria. *Proc. Natl. Acad. Sci. USA* **61** : 1406-1413

Guhathakurta, A., Viney, I. et D. Summers (1996). Accessory proteins impose site selectivity during ColE1 dimer resolution. *Mol. Microbiol.* **20** : 613-620.

Guo, F., Gopaul, D.N. et G.D. Van Duyne (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389** : 40-46.

Guynn, L.J., Dai, W. et M.J. Benedik (1998). Nuclease overexpression mutants of *Serratia marcescens*. *J. Bacteriol.* **180** : 2262-2264.

Hakaart, M.J.J., van den Elzen, P.J.M. et J.J. Nijkamp (1984). Maintenance of multicopy plasmid CloDF13 in *E.coli* cells : Evidence for site-specific recombination at *parB*. *Cell* **36** : 203-209.

Hallet, B. et D.J. Sherratt (1997). Transposition and site-specific recombination : adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.* **21** : 157-178.

Han, Y.W., Gumpert, R.I. et J.F. Gardner (1993). Complementation of bacteriophage lambda integrase mutants : evidence for an intersubunit active site. *EMBO J.* **12** : 4577-4584.

Hayes, F. et D.J. Sherratt (1997). Recombinase binding specificity at the chromosome dimer resolution site *dif* of *Escherichia coli*. *J. Mol. Biol.* **266** : 525-537.

Hayes, F., Lubetzki, S.A. et D.J. Sherratt (1997). *Salmonella typhimurium* specifies a circular chromosome dimer resolution system which is homologous to the Xer site-specific recombination system of *Escherichia coli*. *Gene* **198** : 105-110.

Hickman, A.B., Waninger, S., Scocca, J.J. et F. Dyda (1997). Molecular organization in site-specific recombination : the catalytic domain of bacteriophage Hp1 integrase at 2.7Å resolution. *Cell* **89** : 227-237.

Hiraga, S.-I., Sugiyama, T. et T. Itoh (1994). Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J. Bacteriol.* **176** : 7233-7243.

Höfte, M., Dong, Q., Kourambas, S., Krishnapillai, Sherrat, D. et M. Mergeay (1994). The *sss* gene product, which affects pyoverdinin production in *Pseudomonas aeruginosa* TNSK2, is a site-specific recombinase. *Mol. Microbiol.* **14** : 1011-1020.

Holt, J.G.: Editor (1984-1989). *Bergey's manual of systematic bacteriology*. Williams & Wilkins. Baltimore. Vol.1: 964p.

Iuchi, S. et L. Weiner (1996), Cellular and molecular physiology of *Escherichia coli* in the adaptation to aerobic environments. *J. Biochem.* **120** : 1055-1063.

Jayaram, M. (1994). Phosphoryl transfer in Flp recombination : a template for strand transfer mechanisms. *Trends Biochem. Sci.* **19** : 78-82.

Jayaram, M. (1997). The *cis-trans* paradox of integrase. *Science* **276** : 49-51.

Jayaram, M. et J. Lee (1995). Return to sobriety after the catalytic party. *Trends Genet.* **11** : 432-433.

Khatri, G.S., MacAllister, T., Sista, P.R. et D. Bastia (1989). The replication terminator protein of *E. coli* is a DNA sequence-specific contra-helicase. *Cell* **59**: 667-674.

Kikuchi, Y. et H.A. Nash (1979). Nicking-closing activity associated with bacteriophage λ *int* gene product. *Proc. Natl. Acad. Sci. USA* **76** : 3760-3764.

Klemm, P. (1986). Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type I fimbriae in *Escherichia coli*. *EMBO J.* **5** : 1389-1393.

Kolot, M.N. (1990). Par site of the ColN plasmid: Structural and functional organization. *Mol. Gen. Genet.* **222** : 77-80.

Kuempel, P.L., Henson, J.M., Dircks, L., Tecklenburg, M. et Lim, D.F. (1991). *dif*, A RecA-independent recombination site in the terminus region of the chromosome of *Escherichia coli*. *New Biol.* **3** :799-811.

Kuempel, P., Hogaard, A., Nielsen, M., Nagappan, O. et M. Tecklenburg (1996). Use of a transposon (*Tndif*) to obtain suppressing and nonsuppressing insertions of the *dif* resolvase site of *Escherichia coli*. *Genes Dev.* **10** : 1162-1171.

Kwon, H.J., Tirulamai, R., Landy, A. et T. Ellenberg (1997). Flexibility in DNA recombination : structure of the lambda integrase catalytic core. *Science* **276** : 126-131.

Lee, J., Whang, I, Lee, J. et M. Jayaram (1994). Directed protein replacement in recombination full sites reveals *trans*-horizontal DNA cleavage by Flp recombinase. *EMBO J.* **13** : 5355-5360.

Lilley, D.M.J. (1997). All change at Holliday junction. *Proc. Natl. Acad. Sci. USA* **94** : 9513-9515.

Lilley, D.M.J. et R.M. Clegg (1993). The structure of the four-way junction in DNA. *Annu. Rev. Biophys. Biomol. Struct.* **22** : 299-328.

Lim, D., Oppenheim, J.D., Eckhardt, T. et W. K. Maas (1987). Nucleotide sequence of the *argR* gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor. Proc. Natl. Acad. Sci. USA. **84** : 6697-6701.

Lovett, S.T. et R.D. Kolodner (1991). Nucleotide sequence of the *Escherichia coli* *recJ* chromosomal region and construction of RecJ-overexpression plasmids. J. Bacteriol. **173** : 353-364.

Luttinger, A. (1995). The twisted 'life' of DNA in the cell : bacterial topoisomerases. Mol. Microbiol. **15** : 601-606.

Maas, W.K. (1994). The arginine repressor of *Escherichia coli*. Microbiol. Rev. **58** : 631-640.

McCulloch, R., Burke, M.E. et D.J. Sherratt (1994a). Peptidase activity of *Escherichia coli* aminopeptidase A is not required for its rôle in Xer site-specific recombination. Mol. Microbiol. **12** : 241-251.

McCulloch, R., Coggins, L.W., Colloms, S.D. et D.J. Sherratt (1994b). Xer-mediated site-specific recombination at *cer* generates Holliday junctions *in vivo*. EMBO J. **13** : 1844-1855.

Missiakas, D., Georgopoulos, C., S. Raina (1994). The *Escherichia coli* *dsbC* (*xprA*) gene encodes a periplasmic protein involved in disulfide bond formation. EMBO J. **13** : 2013-2020.

Morlon, J., Chartier, M., Bidaud, M. et C. Lazdunski (1988). The complete nucleotide sequence of the colicinogenic plasmid ColA. High extent of homology with ColE1. Mol. Gen. Genet. **211** : 231-243.

Nash, H.A. (1996). Site-specific recombination : integration, excision, resolution, and inversion of defined DNA segments. In *Escherichia coli* and *Salmonella typhimurium*-Cellular and molecular biology. Second Edition. Frederick C. Neidhardt, Ed. ASM Press, Washington. Vol.II : 2363-2376.

Nijkamp, H.J., de Lang, R., Stuitje, A.R., van den Elzen, P.J., Veltkamp, E. et A.J. van Putten (1986). The complete nucleotide sequence of the bacteriocinogenic plasmid CloDF13. Plasmid **16**: 135-160.

Nunes-Düby, S.E., Matsumoto, L. et A. Landy (1987). Site-specific recombination intermediates trapped with suicide substrates. Cell **50** : 779-788.

Nunes-Düby, S.E., Tirumalai, R.S., Dorgai, L., Yagil, E., Weisberg, R.A. et A. Landy (1994). λ Integrase cleaves DNA in *cis*. EMBO J. **13** : 4421-4430.

- Nunes-Düby, S.E., Azaro, M.A. et A. Landy (1995). Swapping DNA strands and sensing homology without branch migration in λ site-specific recombination. *Curr. Biol.* **5** : 139-148.
- Nunes-Düby, S.E., Kwon, H.J., Tirumalai, R.S., Ellenberger, T. et A. Landy (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26** : 391-406.
- Pargellis, C.A., Nunes-Düby, S.E., de Varga, L.M. et A. Landy (1988). Suicide recombination substrates yield covalent λ integrase-DNA complexes and lead to identification of the active site tyrosine. *J. Biol. Chem.* **263** : 7678-7685.
- Piette, J., Nyunoya, H., Lusty, C.J., Cunin, R., Weyens, G., Crabeel, M., Charlier, D., Glansdorff, N. et A. Piérard (1984). DNA sequence of the *carA* gene and the control region of *carAB* : tandem promoters, respectively controlled by arginine and the pyrimidines, regulate the synthesis of carbamoyl-phosphate synthetase in *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. USA* **81** : 4314-4138.
- Qian, X.-h. et M.M. Cox (1995). Asymetry in active complexes in FLP recombinase. *Genes Dev.* **9** : 2053-2064.
- Réhel, N. et G. Szatmari (1996). Characterization of the stable maintenance of the *Shigella flexneri* plasmid pHS-2. *Plasmid* **36** : 183-190.
- Rice, P.A. et T.A. Steitz (1994). Model for a DNA-mediated synaptic complex suggested by crystal packing of $\gamma\delta$ resolvase subunits. *EMBO J.* **13** : 1514-1524.
- Richaud, C., Higgins, W., Mengin-Lecreulx, D. et P. Stragier (1987). Molecular cloning, characterization, and chromosomal localization of *dapF*, the *Escherichia coli* gene for diaminopimelate epimerase. *J. Bacteriol.* **169** : 1454-1459.
- Sadowski, P.D. (1993). Site-specific recombination : hops, flips, and flops. *FASEB J.* **7** : 760-767.
- Schultz, S.C., Shields, G.C. et T.A. Steitz (1991). Crystal structure of a CAP-DNA complex : the DNA is bent by 90°. *Science* **253** : 1001-1007.
- Sekiguchi, J., Seeman, N.C. et S. Shuman (1996). Resolution of Holliday junctions by eukaryotic DNA topoisomerase I. *Proc. Natl. Acad. Sci. USA* **93** : 785-789.
- Shaikh, A.C. et P.D. Sadowski (1997). The Cre recombinase cleaves the *lox* site in *trans*. *J. Biol. Chem.* **272** : 5695-5702.
- Sherratt, D.J. et D.B. Wigley (1998). Conserved themes but novel activities in recombinases and topoisomerases. *Cell* **93** : 149-152.
- Sherratt, D.J., Summers, D.K., Boocock, M., Stirling, C. et G. Stewart (1986). Novel recombination mechanisms in the maintenance and propagation of plasmid genes, p.

263-273. *In* S.B. Levy et R.P. Novick (Ed.) Banbury Report 24 : Antibiotic resistance genes : ecology, transfer, and expression. Cold Spring Harbor Laboratory. New York.

Sherratt, D.J., Blakely, G., Burke, M., Colloms, S., Leslie, N., McCulloch, R., May, G. et J. Roberts (1993). Site-specific recombination and the partition of bacterial chromosomes. *In* The chromosome, Ed. :J.S. Heslop-Harrison et R.B. Flavell, Bios Scientific Publishers Limited, Oxford, p. 25-41.

Sherratt, D.J., Arciszewska, L.K., Blakely, G., Colloms, S., Grant, K., Leslie, N. et R. McCulloch (1995). Site-specific recombination and circular chromosome segregation. *Phil. Trans. R. Soc. Lond. B* **347** : 37-42.

Shinagawa, H. et H. Iwasaki (1996). Processing the Holliday junction in homologous recombination. *Trends Biochem. Sci.* **21** : 107-111.

Sinden, R.R. (1994). DNA structure and function. Academic Press. San Diego, Californie. 398p.

Sirois, S. et G. Szatmari (1995). Detection of XerC and XerD recombinases in gram-negative bacteria of the family *Enterobacteriaceae*. *J. Bacteriol.* **177** : 4183-4186.

Slack, F.J., Serror, F., Joyce, E. et A.L. Sonenshein (1995). A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Mol. Microbiol.* **15** : 689-702.

Spiers, A.J. et D.J. Sherratt (1997). Relating primary structure to function in the *Escherichia coli* XerD site-specific recombinase. *Mol. Microbiol.* **24** : 1071-1082.

Stark, W.M. et M.R. Boocock (1995). Gatecrashers at the catalytic party. *Trends Genet.* **11** : 121-123.

Stark, W.M., Boocock, M.R. et D.J. Sherratt (1992). Catalysis by site-specific recombinases. *Trends Genet.* **8** : 432-439.

Steiner, W.E. et P.L. Kuempel (1998). Cell division is required for resolution of dimer chromosomes at the *dif* locus of *Escherichia coli*. *Mol. Microbiol.* **27** : 257-268.

Stirling, C.J., Stewart, G. et D.J. Sherratt (1988a). Multicopy plasmid stability in *Escherichia coli* requires host-encoded functions that lead to plasmid site-specific recombination. *Mol. Gen. Genet.* **214** : 80-84.

Stirling, C.J., Szatmari, G., Stewart, G., Smith, M.C.M. et D.J. Sherratt (1988b). The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus. *EMBO J.* **7** : 4389-4395.

Stirling, C.J., Colloms, S.D., Collins, J.F., Szatmari, G. et D.J. Sherratt (1989). *XerB*, an *Escherichia coli* gene required for ColE1 site-specific recombination, is identical to

pepA, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. *EMBO J.* **8** : 1623-1627.

Subramanya, H.S., Arciszewska, L.K., Baker, R.A., Bird, L.E., Sherratt, D.J. et D.B. Dale (1997). Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **16** : 5178-5187.

Summers, D., Yaish, S. et D. Sherratt (1985). Multimer resolution systems of ColE1 and ColK : localisation of the crossover site. *Mol. Gen. Genet.* **201** : 334-338.

Summers, D.K. (1989). Derivatives of ColE1 *cer* show altered topological specificity in site-specificity recombination. *EMBO J.* **8** : 309-315.

Summers, D.K. and D.J. Sherratt (1984). Multimerization of high copy number plasmids causes instability : ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36** : 1097-1103.

Summers, D.K. et D.J. Sherratt (1988). Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the *cer* site. *EMBO J.* **3** : 851-858.

Tecklenburg, M., Naumer, A., Nagappan, O. et P. Kuempel (1995). The *dif* resolvase locus of the *Escherichia coli* chromosome can be replaced by a 33-bp sequence, but function depends on location. *Proc. Natl. Acad. Sci. USA.* **92** : 1352-1356.

Tirumalai, R.S., Healy, E. et A. Landy (1997). The catalytic domain of λ site-specific recombinase. *Proc. Natl. Acad. Sci. USA* **94** : 6104-6109.

Van de Putte, P. et N. Goosen (1992). DNA inversions in phages and bacteria. *Trends Genet.* **8** : 457-462.

Wang, J.C. (1996). DNA topoisomerases. *Annu. Rev. Biochem.* **65** : 635-692.

Yang, W. et K. Mizuuchi (1997). Site specific recombination in plane view. *Structure* **5** : 1401-1406.

Yang, W. et T.A. Steitz (1995). Crystal structure of the site-specific recombinase $\gamma\delta$ resolvase complexed with a 34 bp cleavage site. *Cell* **82** : 193-207.

Zakova, N. et G.B. Szatmari (1995). Site-specific recombination between ColE1 *cer* and NTP16 *nmr* sites in vivo. *Mol. Gen. Genet.* **247** : 509-514.

Zerbib, D., Colloms, S.D., Sherratt, D.J. et S.C. West (1997). Effect of DNA topology on Holliday junction resolution by *Escherichia coli* RuvC and bacteriophage T7 endonuclease I. *J. Mol. Biol.* **270** : 663-673.

Zlatanova, J. et K. Van Holde (1998). Binding to four-way junction DNA : a common property of architectural proteins ? FASEB J. **12** : 421-431.

ANNEXE I :

Alignement de plusieurs recombinaisons de type Xer.

PaxC 1 ~~~~~WRADDDALEHRSERQSAHTDGYRDDLLKYLALAEKA.G.....SDWNAIDTRSRRTDRLHQCGSSRSLARL
PfxC 1 ~~~~~ERQDDANCFHRSERQSPBTSAYRDDLLKYLIGWCQKQ.NE.....GSWAAIDIQRLSLRLHQCGSSRSLARL
EcxC 1 ~~~~~ETDVERLDHYTSERQSPITLNVQOQLDAIINASEN.G.....OSWQCCDVTMVRNVAVRSRKGLGASLAL
StxC 1 ~~~~~MTDVA..SQDESRELYLTERQSPITLNVQOQLDAIALAGET.G.....KSWQCCDAATVRSVAVRSRKGLGASLAL
SmxC 1 ~~~~~MTPFAPS..QPPVDAIRYLKERQSPITLNVQOQLDAIINASEN.G.....TDTTDAARHMLARSKRMGLQASLAL
FmxC 1 ~~~~~MSQIIDVPETSLADBSLSYDTERQSPVTEENYQOQIMTAQMVAI.K.....NWSLESQHRMLAKSHRSGLQASLAL
HixC 1 ~~~~~MLTANRWYTRERQSPBTTNYQOQOATIKIQAQO.D.....HSWQCTPSVSRFIAESKKQGLKESLAL
EcxD 1 ~~~~~VKQDLARFQELALWBERQSPBTTNAYRDDLSMVEV.HH.RG.....LTLTAQSDDDQALAEERLEGYKATSSARL
StxD 1 ~~~~~MEQDLARFQELALWBERQSPBTTNAYRDDLSMVEV.HH.RG.....KTLTAQADDQTLAEERVEGYKATSSARL
FmxD 1 ~~~~~VTOQTKLSTQTTDNDNQEDNVILQPTLSENEQCSANSSAARDLQANQSVT..QR.....LNNLSLTTLDHATLTRLDEGYKATSSARL
HixD 1 ~~~~~MKNLALHLPLNEYHBERQSPBTTNAYRDDLSMVEV.HH.RG.....LSLETDAVDLQGLERLEFGYKATSSARL
BaxC 1 ~~~~~MKDQKDFEYVMBERQSONTVSVERDLKYSLYTETLEW.....TDWNEATRIEIQKHLKDSGKSGKSSARL
BsxC 1 ~~~~~MENKFNFKLGYQOBERQSOYVNVVDSSEEFETLAVQ.G.....NGEAAAYQDTRIEATEAYGTSRRSSARL
LlxC 1 ~~~~~MTLEQOETSYLKNERSYSPRTLAVQKDTAAAKKQENGGE.....PGDQSSRRDLIPLATGQKLASSLSRI
MlxD 1 ~~~~~VTAAQTCQOCLLYLTERQSPBTTNAYRDDLSMVEV.HH.RG.....SSVRRDLIRYSKHSDRGIEDLAKVGEHDVSEFLVARRGDFDSGVALSAVSAARA
MtxD 1 ~~~~~MKTALQOCLLTERQSPBTTNAYRDDLSMVEV.HH.RG.....SSVRRDLIRYSKHSDRGIEDLAKVGEHDVSEFLVARRGDFDSGVALSAVSAARA
LlxC 1 ~~~~~VSDHVAIIEEDFIAARCGHSEETRAVLEDRSLSKHLEGRG.SRLDQLS.....LPVLRSLATAAGAGARTLARR
MtxC 1 ~~~~~VRRVDSGSRRHACQCGGQALDEEDVYLLCCGHSVETRAVLEDRSIFALDRG.SSLDALT.....LSVRSALATAGAGARTLARR
HpxC 1 MKHPLEELKDPFENLLLWIGRFLRYKCTSLNSQ.VKDQNVFECLNELNQACSSSQLEKVKCKARAGLLGENTLALPLKHFHEFSKARLTERLAFNSLKNDEVMALLEV.YTGETSLAEKKNY
HpxD 1 ~~~~MPHNKLTKSQRELFENLKAFLYTRAKNFPIQDVKDMALILDQDKILKCHNEQLKQLCHLHYQGIKE.TIMMCGTFLFFNFK.....DNLKLRSRMSEEQINDEFE.LAQNRPSSMANY

PaxC 74 LSAARGLVQLNRECHCDBDPAT.....APKSPRLPPTLDAORALQDLIG...AVEDDFIARRDCALLELYSGLRSELVGLD...LVRVCGKNVREPYGAAARQ.AL
PfxC 74 LSAARGLVQLNRECHCDBDPAT.....APKSPRLPPTLDAORALQDLIG...AVEDDFIARRDCALLELYSGLRSELVGLD...LVRVCGKNVREPYGAAARQ.AL
EcxC 77 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
StxC 79 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
SmxC 81 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
FmxC 84 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
HixC 74 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
EcxD 76 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
StxD 76 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
FmxD 91 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
HixD 75 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
BaxC 75 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
BsxC 77 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
LlxC 73 LSAARSTEDLSQNEKAKPAK...S.....APKAPRLPKNDVDSNLLLI...DINDP.LAARDRALENYAGLRSELVGLD...EVVVCGKSKERRP.G.NA.T.W
MlxD 86 LIAVRLGLHRAAEGCDELVDAR...R.....PFTFGRLEKSLTVDVLLSVGGESGADGPIV...RRALELYSTGRSEAVGLD...DTHAR...SVLPCGKGG...LVP...GAPVH.AL
MtxD 84 LIAVRLGLHRAAEGCDELVDAR...R.....PFTFGRLEKSLTVDVLLSVGGESGADGPIV...RRALELYSTGRSEAVGLD...DTHAR...SVLPCGKGG...LVP...GAPVH.AL
LlxC 77 TSAAFTAARWCGAADPAAR...Q.....VFAARTLPAVLRQDQALRAAAESCAEQGDPLARDRL...LLELYTGRSELVGLD...DTHAR...LVRVCGKN...TVPI...GAPVH.AL
MtxC 90 TSAAFTAARWCGAADPAAR...Q.....VFAARTLPAVLRQDQALRAAAESCAEQGDPLARDRL...LLELYTGRSELVGLD...DTHAR...LVRVCGKN...TVPI...GAPVH.AL
HpxC 129 RIALGLLSDNDQDENKSYIYNITLKNISGVNQSAGNLPTELNNKFKLSEDDKIEMSAK...VFAARTLPAVLRQDQALRAAAESCAEQGDPLARDRL...LLELYTGRSELVGLD...DTHAR...LVRVCGKN...TVPI...GAPVH.AL
HpxD 121 MYLROEEDLDR.....KARYSFDFTLKNLAFAKT...ESLPEBNDKPKSEKTLIDYKPTS.FEMRCL...LILGLRKCENK...KBOEONYSLECKGRKEFRAYESLPEPSL

PaxC 188 BAWP...L.RTOAEPEDCAFSG.SG.RPRAIQLRVQA...GQH.BPHLRRHSFATHLESSDLRAVOELLGHADLSTTOIYTHDFOLASVYAPIPRAKQGGRRGGNDP*
PfxC 188 BAWP...L.RALTNPADAFSG.CQG.RPRAIQLRVQA...GQN.BPHLRRHSFATHLESSDLRAVOELLGHADLSTTOIYTHDFOLATVYSAHPRAKRIKGDSE*
EcxC 190 BAWP...L.RDLFSEDAFSLG.RARNVQRTAE...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLASVYAAHPRAKRGK*
StxC 192 BAWP...L.RGLFSEDAFSLG.RARNVQRTAE...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
SmxC 194 BAWP...L.RDLFSEDAFSLG.RARNVQRTAE...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
FmxC 197 BAWP...L.RDYSPEDAFSG.SG.RPRAIQLRVQA...GQH.BPHLRRHSFATHLESSDLRAVOELLGHADLSTTOIYTHDFOLASVYAAHPRAKRGK*
HixC 187 BAWP...L.RALFNPKDAFSLG.RARNVQRTAE...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
EcxD 190 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
StxD 190 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
FmxD 205 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
HixD 189 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
BaxC 189 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
BsxC 191 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
LlxC 187 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
MlxD 204 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
MtxD 202 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
LlxC 195 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
MtxC 208 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
HpxC 258 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*
HpxD 242 BAWP...L.NVSI...GNNB.BPHLRRHSFATHLESSDLRAVOELLGHANLSTTOIYTHDFOLANVYAAHPRAKRGK*

Figure de la page précédente :

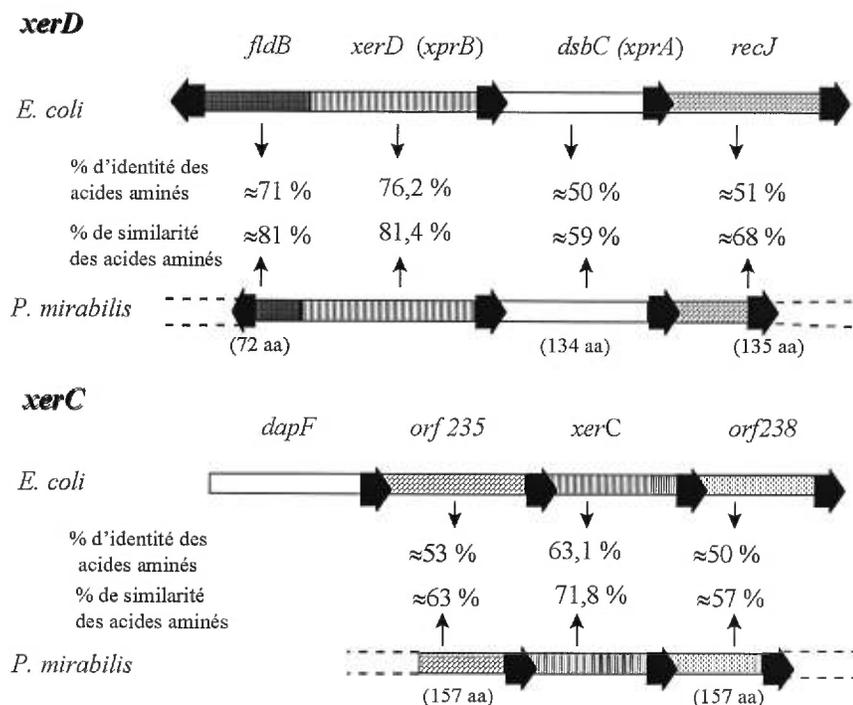
L'alignement comprend les recombinases du tableau présenté à l'annexe IV, à l'exception de *S. aureus* et *M. thermoautotrophicum*. PaxC : Sss de *Pseudomonas aeruginosa*, PfxC : Sss de *P. fluorescens*, EcxC : XerC d'*E. coli*, StxC : XerC de *Salmonella typhimurium*, SmxC : XerC de *Serratia marcescens*, PmxC : XerC de *Proteus mirabilis*, HixC : XerC de *Haemophilus influenzae*, EcxD : XerD de *E. coli*, StxD : XerD de *S. typhimurium*, PmxD : XerD de *P. mirabilis*, HixD : XerD de *H. influenzae*, BsxD : RipX de *Bacillus subtilis*, BsxC : CodV de *B. subtilis*, LlxC : XerC de *Lactobacillus leichmannii*, MlxD : XerD de *Mycobacterium lepreae*, MtxD : XerD de *M. tuberculosis*, MlxC : XerC de *M. lepreae*, MtxC : XerC de *M. tuberculosis*, HpxC : XerC de *Helicobacter pylori*, HpxD : XerD de *H. pylori*.

L'alignement a été fait grâce au programme pileup de GCG (Genetics Computer Group, Wisconsin package version 9.1, Madison, Wisconsin, © : 1982-1997), et le consensus a été mis en valeur par le programme BOXSHADE 3.21 (de K. Hofmann et M. Baron, http://www.isrec.isb-sib.ch/software/BOX_form.html). Les lettres blanches sur fond noir indiquent un consensus d'acides aminés identiques, tandis que les lettres noires sur fond gris indiquent un consensus d'acides aminés similaires. Les différentes boîtes correspondent aux motifs tels que nommés par Cao *et al.* (1997) : motifs I (cadre bleu), II (cadre vert) et III (cadre rouge). Les résidus conservés du site catalytique (RHRY), retrouvés chez tous les membres de la famille des intégrases, sont marqués d'un point.

ANNEXE II :

Comparaison des organisations génomiques de *xerC* et *xerD*
d'*E. coli* et de *P. mirabilis*.

Organisation génomique



Cette figure compare les organisations génomiques des régions entourant *xerD* et *xerC* de *P. mirabilis* avec celles que l'on retrouve chez *E. coli*. Les traits en pointillés indiquent des séquences incomplètes. Les pourcentages (%) d'identité et de similarité en acides aminés pour les séquences autres que les recombinases sont approximatifs (≈), puisque les séquences n'ont pas été revérifiées et sont souvent incomplètes. D'ailleurs, les chiffres entre parenthèses indiquent le nombre d'acides aminés correspondant à ces séquences partielles. Il est à remarquer aussi que DsbC fait 236 aa chez *E. coli*, la séquence obtenue pour *P. mirabilis* ne semble faire que 134 aa (mais il est possible qu'il y ait une erreur dans la séquence qui fait apparaître un codon stop trop tôt). Les flèches indiquent l'orientation de la transcription des gènes.

ANNEXE III :

Caractéristiques de XerD suite à l'insertion aléatoire de pentapeptides.

Position de l'insertion	Élément de la structure secondaire correspondant	Activité <i>in vivo</i> à:		Liaison à <i>dif</i> <i>in vitro</i>	Coopérativité avec XerC à <i>dif</i> <i>in vitro</i>
		<i>dif</i>	<i>cer</i>		
8 (a)	Hélice αA	++		Aucune	++
8 (b)	Hélice αA	++		Faible	++
26	Hélice αB	Aucune		Faible	Aucune
59	Hélice αC	++	Aucune	Faible	++
76	Hélice αD	++Aucune.....		
85	Hélice αD	++		Faible	++
114	Bordures de l'hélice αE	++	Aucune	++	Faible
118	Hélice αE	++	Moyenne	Faible	++
139	Hélice αFAucune.....			
146	Bordures de l'hélice αFAucune.....			
153	Hélice αG	++Aucune.....		
157	Boucle entre αG et $\beta 1$	Moyenne		Aucune	
160	Feuillet $\beta 1$	Moyenne	++	Aucune	++
172	Boucle entre $\beta 2$ et $\beta 3$	Aucune		++	
175	Boucle entre $\beta 2$ et $\beta 3$	Moyenne		++	
184	Hélice αHAucune.....			
187	Hélice αHAucune.....			
209	Boucle entre αI et αJ	++	Moyenne	Aucune	++
227	Hélice αJ	Moyenne	Aucune.....	
234	Boucle entre αJ et αK	++	Moyenne	Aucune	
247	Hélice αL	Aucune		++	Aucune
256	Hélice αL	++	Aucune	Faible	Aucune
268	Boucle entre αM et αN	Aucune		Faible	++
279	Hélice αN ($\cong Y$ catalytique)	Aucune		Faible	++
292	Bordures de l'hélice αN	Aucune		++	

Tableau des caractéristiques de XerD suite à l'insertion aléatoire de pentapeptides dans sa séquence. Les informations de ce tableau sont tirées du tableau 1 de Cao *et al.* (1997). Seules les insertions qui ont montré un phénotype différent de la protéine sauvage sont indiquées. L'activité *in vivo* est le taux de recombinaison : élevé (par des ++), moyen ou aucun. Les expériences *in vitro* correspondent au pourcentage de liaison ou de coopérativité par rapport au XerD sauvage : >50% (par des +), entre 25 et 50% (faible) et <25% (aucune). \cong signifie "correspondant à".

Il y a eu deux insertions par des séquences peptidiques différentes en position 8 : a et b. Les positions des insertions sont indiquées sur la structure de XerD (fig. 8).

ANNEXE IV :

Espèces chez lesquelles des gènes encodant des recombinases homologues à XerC et XerD ont été séquencés.

Espèces chez lesquelles des gènes encodant des recombinaisons
homologues à XerC ou XerD ont été séquencés

Espèce	Gène	Numéro d'accession	Taille (nombre d'acides aminés)
<i>Bacillus subtilis</i>	<i>codV</i> (\cong <i>xerC</i>)	U13634	304
<i>B. subtilis</i>	<i>ripX</i> (\cong <i>xerD</i>)	Z99116	296
<i>Escherichia coli</i>	<i>xerC</i>	M38257	298
<i>E. coli</i>	<i>xerD</i>	M54884	298
<i>Haemophilus influenzae</i>	<i>xerC</i>	U32750	295
<i>H. influenzae</i>	<i>xerD</i>	U32716	297
<i>Helicobacter pylori</i>	<i>xerC</i>	AE000580	362
<i>H. pylori</i>	<i>xerD</i>	AE000608	355
<i>Lactobacillus leishmannii</i>	<i>xerC</i>	X84261	295
<i>Methanobacterium thermoautotrophicum</i>	<i>xerC</i>	AE000666	311
<i>Mycobacterium leprae</i>	<i>xerC</i>	Z97369	302
<i>M. leprae</i>	<i>xerD</i>	U00021	316
<i>Mycobacterium tuberculosis</i>	<i>xerC</i>	Z74024	315
<i>M. tuberculosis</i>	<i>xerD</i>	Z98268	311
<i>Pseudomonas aeruginosa</i>	<i>sss</i> (\cong <i>xerC</i>)	X78478	302
<i>Pseudomonas fluorescens</i>	<i>sss</i> (\cong <i>xerC</i>)	Y12268	299
<i>Proteus mirabilis</i>	<i>xerC</i>	AF033498	307
<i>P. mirabilis</i>	<i>xerD</i>	AF033497	313
<i>Serratia marcescens</i>	<i>xerC</i>	AF028736	303
<i>Salmonella typhimurium</i>	<i>xerC</i>	U92524	300
<i>S. typhimurium</i>	<i>xerD</i>	U92525	298
<i>Staphylococcus aureus</i>	<i>xerD</i>	*	297

\cong : signifie "correspond à".

*: Le numéro d'accession n'est pas encore acquis
(Tichat et Szatmari, données non publiées).

ANNEXE V :

Séquence de XerC de *P. mirabilis*.

Séquence de XerC de *P. mirabilis*

Longueur : 307 acides aminés

```
1  MSQIIDVPET  LSLAIDSFLS  YIEVERRLSP  VTVENYQRQL  MTIAQMMVAI
51  KINQWSLLES  QHVRMLLAKS  HRSGLOPASL  ALRFSALRSF  LDWQVSQGML
101 AVNPAKGVRT  PKSGRHLPKN  MDVDEVSQLM  NIDLKDPLSV  RDRTMLEVMY
151 GAGLRLSELT  NLNINDIDLQ  EGEVRVLGKG  SKERKVPLGR  KAVEWLQHWF
201 AMRELYSPED  TAVFISTKSG  KRLSVRSVQK  RFELWGVKQG  LSSHVNPBKL
251 RHSFATHLLE  SSGDLRAVQE  LLGHANLSTT  QVYTHLDFQH  LAKVYDAAHP
301 RAKREKS
```

Remerciements

Profitons de cet espace pour remercier tout le monde !...

Premièrement, je ne peux m'empêcher de remercier chaleureusement mon directeur de recherche, le Dr George Szatmari, pour sa patience et pour toutes les questions auxquelles il a bien voulu se soumettre et tenter de répondre (!).

Je remercie tous les laboratoires du département de microbiologie et immunologie, en particulier ceux des Dr Drolet et Belhumeur pour l'emprunt de matériel très utile à l'avancement du projet, et également Pauline Phoenix et les étudiants de ces laboratoires qui ont bien voulu répondre à mes nombreuses questions d'ordre technique (et tous savent que j'en ai toujours une quantité industrielle !!). Je pense ici surtout à Sonia, Éric, Martin et Francis.

Je voudrais aussi souligner le support moral fourni gracieusement par Naïma Tichat et Man Hua Ngiep, deux collègues de travail : merci à Naïma pour m'avoir écouté déblatérer, merci à Man pour m'avoir expliqué les déboires possibles des manip reliées à l'expression protéique.

Une pensée toute spéciale est offerte à toute l'équipe du P-600, de la stérilisation et de la laverie, pour leur immense générosité et leur constante bonne humeur (☺).

Je remercie également Shona qui était là au tout début et qui m'a initiée à la recherche en bactériologie.

Finalement, je remercie bien évidemment le Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG) pour la bourse octroyée qui m'a permis d'effectuer ma maîtrise en évitant de penser à d'éventuels tracas financiers (Ouf !).

J'oublie quelque chose... Ah oui ! J'ajouterais un tout petit mot pour *Proteus*. Il ne faudrait quand même pas omettre de souligner ici son importance : j'avoue qu'il a aidé « un peu » à la réalisation de ce travail...