



HAL
open science

**Recombinaison spécifique de site chez les archaea.
Implication dans le cycle du virus SSV1 de Sulfolobus
shibatae**

Marie-Claude Serre

► **To cite this version:**

Marie-Claude Serre. Recombinaison spécifique de site chez les archaea. Implication dans le cycle du virus SSV1 de Sulfolobus shibatae. Biochimie [q-bio.BM]. Université Paris Sud - Paris XI, 2005. tel-00112419

HAL Id: tel-00112419

<https://theses.hal.science/tel-00112419>

Submitted on 8 Nov 2006

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

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

UNIVERSITE PARIS XI

NOTICE DE TITRES ET TRAVAUX

présentée par

MARIE-CLAUDE SERRE

en vue de l'obtention d'une

HABILITATION A DIRIGER LES RECHERCHES

**RECOMBINAISON SPECIFIQUE DE SITE CHEZ LES ARCHAEA
IMPLICATION DANS LE CYCLE DU VIRUS SSV1 DE *SULFOLOBUS SHIBATAE***

**INSTITUT DE GENETIQUE ET MICROBIOLOGIE, UMR8621
UNIVERSITE PARIS-SUD ORSAY**

REMERCIEMENTS

J'ai longtemps pensé que rien ne pourrait être pire que la rédaction de ma thèse, et m'étais jurée de ne pas reproduire l'erreur de la dernière minute avec la notice d'habilitation. Promesse bien évidemment oubliée, qui certes m'aura fait passer quelques nuits face à mon clavier, mais qui aura surtout obligé les membres de mon jury à eux aussi travailler pour « hier ». Que tous soient ici remerciés pour leur patience à mon égard, leur tolérance face à mes promesses de Gascon, ... et la rapidité avec laquelle ils ont dû juger ce travail, Lucienne Letellier, François Cornet, David Prangishvili et Herman Van Tilbeurgh pré-rapporteurs et/ou rapporteurs, Jean-Luc Pernodet et Jean-Michel Rossignol, examinateur et président de jury respectivement.

Il me faut aussi remercier pour les mêmes raisons, Philippe et Victorien qui doivent me maudire de toujours choisir la période des vacances pour m'atteler à la rédaction de ce manuscrit ou de sa version préliminaire...

L'exercice de rédaction aura finalement été bien plus enrichissant que je ne l'imaginais. En regardant les années passées, je réalise la chance que j'ai eu d'interagir avec des personnes aussi différentes que Pascale Cossart, Jean-Renaud Garel, Makkuni Jayaram, Mick Chandler, et Michel Duguet. Tous m'ont énormément apporté, scientifiquement et humainement, et ont su me transmettre leur curiosité et leur enthousiasme face à la science. J'espère n'avoir déçu aucun d'eux. Je souhaite remercier tout particulièrement Jean-Renaud qui par deux fois m'a fait confiance, et par deux fois a sû m'obliger à « grandir », et Michel qui a si gentiment accepté de m'accueillir dans son équipe avec ma propre thématique. Le projet SSVI leur doit beaucoup, j'en suis pleinement consciente, et je leur en suis grandement reconnaissante.

Je ne pourrai bien évidemment pas citer ici tous les collègues avec qui j'ai eu des discussions parfois animées, toujours enrichissantes, ni tous les étudiants qui ont contribué à l'avancement des différents sujets que j'ai étudié. Je ferai une exception en remerciant Claire Letzelter, qui a passé quatre années à mes côtés avant de partir voler de ses propres ailes. Claire est bien plus qu'une ancienne étudiante, c'est une véritable amie. Je terminerai en remerciant les membres passés et présents de l'équipe EAN pour tous les bons moments scientifiques et extra-scientifiques que nous aurons partagés, Christiane, Claire B., Claire L., Jeanine, Laila, Aude, Catherine, Marguerite, Murielle, Nadine, Evelyne, Yannick, Mauro, Thierry, Cong Zhao, Raynald, Fabrice et Michel qui a réussi l'exploit de faire cohabiter harmonieusement autant de forts caractères !

| | |
|--|-----------|
| CURRICULUM VITAE | 2 |
| ACTIVITES DE RECHERCHE | 9 |
| AVANT-PROPOS | 9 |
| I- FORMATION INITIALE (1985 – 1989) | 10 |
| II- RECOMBINAISON SPECIALISEE | 16 |
| 1. Recombinaison spécifique de site chez les eucaryotes | 17 |
| a. <i>Identification et modélisation du rôle des résidus conservés entre Flp et d'autres recombinases de levure</i> | 19 |
| b. <i>L'utilisation de demi-sites de recombinaison révèle les étapes précoces de la réaction de coupure et les propriétés du site actif de Flp</i> | 20 |
| 2. Transposition bactérienne : Propriétés fonctionnelles de la transposase d'IS1 | 22 |
| a. <i>Purification de la transposase d'IS1 et caractérisation d'une activité in vitro</i> | 23 |
| b. <i>Développement de tests d'activité in vivo</i> | 24 |
| c. <i>H-NS est impliquée dans la transposition d'IS1</i> | 25 |
| d. <i>Obtention et caractérisation de mutants de la transposase d'IS1</i> | 26 |
| 3. Recombinaison spécifique de site chez les archaea | 29 |
| a. <i>Analyse des propriétés structurales de l'intégrase de SSV1</i> | 31 |
| b. <i>L'intégrase de SSV1 coupe spécifiquement des substrats synthétiques</i> | 32 |
| c. <i>L'intégrase de SSV1 possède un site actif partagé</i> | 33 |
| III- PROJETS DE RECHERCHE | 35 |
| 1. Mécanisme de la recombinaison spécifique de site chez les archaea | 36 |
| 2. Rôle de la recombinaison spécifique de site dans le cycle viral de SSV1 | 38 |
| 3. Interactions hôte/virus lors de l'intégration virale | 39 |
| 4. Quelles fonctions pour quel cycle viral ? | 40 |
| 5. La recombinaison spécifique de site comme outil pour la thérapie génique | 41 |
| IV- EN GUISE DE CONCLUSION ... | 49 |
| V- BIBLIOGRAPHIE | 49 |

1990 - 1991 Chercheur Postdoctoral Associé. Etude du mécanisme de la recombinaison site-spécifique catalysée par la protéine Flp de *S. cerevisiae*. Department of Microbiology, University of Texas at Austin, Austin. Equipe M. Jayaram.

1986 - 1989 Boursière de Thèse du Ministère de la Recherche et de la Technologie. Etude des propriétés allostériques de la PFK d'*Escherichia coli*. Laboratoire d'Enzymologie, CNRS, Gif-sur-Yvette. Equipe J.R. Garel.

1985 - 1986 Boursière de DEA. Etude de la fixation du complexe CAP-AMP cyclique sur l'ADN. Unité de Biochimie Cellulaire, Institut Pasteur, Paris. Equipe G. Cohen.

III- FINANCEMENT DE LA RECHERCHE

- 2004** Programme Microbiologie du CNRS et appel d'offre BIOTOX
2000 ACI Microbiologie
1991 Bourse de la Fondation pour la Recherche Médicale
1990 - 1991 Bourse de la Fondation Philippe
1986 - 1989 Allocation de recherche, Ministère de la Recherche et de la Technologie

IV- COLLABORATIONS

A l'étranger:

2005 - ... Propriétés enzymatiques des intégrases de fuselloviridae. Dr Q. She, University of Copenhagen, Denmark.

2002 - Phylogénie des intégrases de fuselloviridae. Dr K. Stedman, Portland State University, USA.

En France:

2004 - ... Cristallisation de l'intégrase de SSV1 en complexe avec son site de recombinaison. Dr D. Gopaul, Institut Pasteur, Paris.

2001 - 2002 Interaction des bipyridyl-porphyrines avec la topoisomérase I. Dr C. Verchere, Laboratoire de Chimie Bioorganique et Bioinorganique, Orsay.

1996 - 1998 Cristallisation de l'intégrase de SSV1. Clonage et expression d'une protéine de capsid du virus de la Fièvre Jaune. Dr F. Rey, LEBS, Gif-sur-Yvette.

1994 - 1995 Rôle de la protéine H-NS dans la transposition d'IS1. Dr D. Lane, Laboratoire de Microbiologie et Génétique Moléculaire, UPR 9007, Toulouse.

1993 - 1995 Utilisation de transposons procaryotes comme outils pour la thérapie génique. Drs D. Noël et M. Piechaczyk, Institut de Génétique Moléculaire, Montpellier.

V- ENCADREMENT DE STAGES

- 2005** C. Floquet, stage volontaire de M1, mention GCDE. « Mise au point d'un test d'activité *in vivo* pour l'intégrase de SSV1 ». (100%)
- 2005** G. Qi, Stage obligatoire de M1, mention GCDE. « Analyse des propriétés d'une intégrase d'archae : obtention de mutants du domaine catalytique ». (100%)
- 2005** A. Simon, Stage obligatoire de 1^{ère} année de BTS Biotechnologie. « Clonage et mutagenèse d'une intégrase virale d'archae. Reconstruction de virus déficients dans l'activité de recombinaison ». (100%)
- 2003 - 2004** S. Ben Ghorbal, Stage obligatoire de fin d'école d'ingénieur, INSAT de Tunis. « Obtention et étude de mutants d'Int^{SSV} délétés dans la région N-terminale ». (100%)
- 2000 - 2003** C. Letzelter, Thèse de l'Université Paris XI. « La recombinaison spécifique de site chez les *archaea*: propriétés de l'intégrase du fusellovirus SSV1 ». (100%)
- 2001 - 2002** K. Guilloux, Stage de formation complémentaire d'initiative locale (formation complémentaire post-BTS). « Etude de l'organisation domaniale de l'intégrase de SSV1 ». (100%)
- 1999 - 2000** C. Letzelter, DEA expression génique chez les microorganismes, Université Paris XI. « Étude de la recombinaison spécifique de site chez les *archaea*: Analyse des propriétés de l'intégrase du virus SSV1 ». (100%)
- 1999** C. Letzelter, Stage volontaire de Maîtrise, Université Paris XI. (100%)
- 1997 - 1998** S. Bressanelli, DEA structure, fonction et ingénierie des protéines, Université Paris XI. « Cristallisation d'un complexe spécifique entre l'intégrase du virus SSV1 et son site ADN de reconnaissance ». (à 80 %, en co-encadrement avec Félix Rey laboratoire de Virologie moléculaire structurale)
- 1994 - 1995** S. Calgaro, DEA de Microbiologie et Biotechnologie, Université Toulouse III. « Obtention par mutagenèse régionalisée de mutants de la transposase d'IS1, et étude de leurs propriétés *in vivo* ». (100%)
- 1994** M.L. Bortolin, Stage volontaire de Maîtrise, Université Toulouse III. (100%)
- 1993** M.L. Bortolin, Stage volontaire de Licence, Université Toulouse III. (100%)
- 1991** J. Lee, 1^{ère} année de PhD, University of Texas at Austin. (50 %)
- 1987** H. Enslin, Stage obligatoire du module de Maîtrise « Biologie Moléculaire des Eucaryotes », Université Paris VI. (100%)

VI- ENSEIGNEMENT

2004 - ... A partir de l'année universitaire 2004-05, participation à l'enseignement du Master Sciences & Technologies, Mention Génomes, Cellules, Développement, Evolution. Université Paris XI.

UE Dynamique et fonctionnalité des acides nucléiques (M1 S2):

- Cours (2hr) : Interactions protéines – acides nucléiques. Disponible en ligne : <http://www.igmors.u-psud.fr/enseignement/Cours-IPAN/cours-IPAN.html>
- Cours (1h30) : Mécanismes de la transposition bactérienne. Disponible en ligne : <http://www.igmors.u-psud.fr/enseignement/Cours-ET/Cours-ET.html>
- TD (2 hr) : Topologie

UE Transactions de l'ADN (MI S2):

- Analyse d'articles (Recombinaison spécifique de site chez les bactéries). Encadrement d'étudiant et membre du jury d'examen.

2000 - 2002 Participation à l'enseignement du module de Maîtrise, Technologie des Macromolécules Informatives (TMI) à l'Université Paris XI. Encadrement des travaux pratiques (topologie de l'ADN), TD interprétation et analyse de données expérimentales, examen de TP-TD (50 heures).

1992 - 1995 Encadrement d'un atelier bibliographique (30 heures par an), au DEA de Microbiologie et Biotechnologie, Université Toulouse III. Responsabilité de l'atelier pour l'année 1994-95 sur le thème « Recombinaison spécifique de site » (organisation de séminaires, animation de sessions d'analyses d'articles, etc).

1992 Co-organisation et encadrement d'un cours pratique de Biologie Moléculaire dans le cadre du programme Tempus (Coopération entre l'Université Toulouse III et l'Université de Gdansk). « Role of the integration host factor (IHF) in the control of expression of the IPL promoter ».

1990 - 1991 Cours en Graduate School of Microbiology, University of Texas at Austin (4 heures). « Protein-DNA interactions » et « Retroviral integration ».

1987 - 1989 Vacataire d'enseignement à l'Université Paris VI. Travaux pratiques de la Maîtrise de Biochimie (100 heures / an). Tutorat au module de Maîtrise « Génétique *in vitro* » (séminaires et analyses d'articles), Université Paris VI.

1986 - 1987 Vacataire d'enseignement à l'Université Paris VI. Travaux pratiques de la Licence de Biochimie (100 heures).

1984 - 1985 Enseignante en Mathématiques, formation continue, Groupe d'Etablissements (GRETA) Paris Nord-Est (20 heures hebdomadaires sur l'année scolaire).

VII- PUBLICATIONS

- ♦ **14-** C. Letzelter, M. Duguet & **M.C. Serre*** (2004). Mutational analysis of the archaeal tyrosine recombinase Int^{SSV} suggests a mechanism of DNA cleavage in *trans*. *J. Biol. Chem.*, **279**, 28936-28944.
- ♦ **13-** C. Rouquette, **M.C. Serre** & D. Lane (2004). Protective role for H-NS protein in IS1 transposition. *J. Bacteriol.*, **186**, 2091-2098.
- ♦ **12-** **M.C. Serre** & M. Duguet. (2003). Enzymes that cleave and religate DNA at high temperature: the same story with different actors. In "Progress in Nucleic Acid Research and Molecular Biology" (Eds. K. Moldave). Vol 74, pp. 37-81.
- ♦ **11-** **M.C. Serre***, C. Letzelter, J.R. Garel & M. Duguet. (2002). Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. *J. Biol. Chem.*, **277**, 16758-16767.
- ♦ **10-** **M.C. Serre**, C. Turlan, M.L. Bortolin & M. Chandler. (1995). Mutagenesis of the IS1 transposase: Importance of a His-Arg-Tyr triad for activity. *J. Bacteriol.*, **177**, 5070-5077.
- ♦ **9-** **M.C. Serre**, L. Zheng & M. Jayaram. (1993). DNA splicing by an active site mutant of Flp recombinase. Possible catalytic cooperativity between the inactive protein and its DNA substrate. *J. Biol. Chem.*, **268**, 455-463.

* : corresponding author

- ♦ **8-** J. Lee, **M.C. Serre**, S. H. Yang, I. Whang, H. Araki, Y. Oshima & M. Jayaram. (1992). Functional analysis of Box II mutations in yeast site-specific recombinases Flp and R. Significance of amino acid conservation within the Int family and the yeast sub-family. *J. Mol. Biol.*, **228**, 1091-1103.
- ♦ **7-** **M.C. Serre** & M. Jayaram. (1992). Half-site strand transfer by step-arrest mutants of yeast site-specific recombinase Flp. *J. Mol. Biol.*, **225**, 643-649.
- ♦ **6-** **M.C. Serre**, B. Evans, H. Araki, Y. Oshima & M. Jayaram. (1992). Half-site recombinations mediated by yeast site-specific recombinases Flp and R. *J. Mol. Biol.*, **225**, 621-642.
- ♦ **5-** **M.C. Serre**, W. Teschner & J.R. Garel. (1990). Specific suppression of heterotropic interactions in Phosphofructokinase by the mutation of Leucine 178 to Tryptophan. *J. Biol. Chem.*, **265**, 12146-12148.
- ♦ **4-** W. Teschner, **M.C. Serre** & J.R. Garel. (1990). Introduction by site-directed mutagenesis of a tryptophan residue as a fluorescent probe for the folding of *E. coli* phosphofructokinase. *Biochimie*, **72**, 403-406.
- ♦ **3-** **M.C. Serre** & J.R. Garel. (1990). Role of the C-terminal region in the allosteric properties of *E. coli* Phosphofructokinase. *Eur. J. Biochem.*, **189**, 487-492.
- ♦ **2-** W. Teschner, **M.C. Serre** & J.R. Garel. (1990). Enzymatic properties, renaturation and metabolic role of Mannitol-1-Phosphate Dehydrogenase from *E. coli*. *Biochimie*, **72**, 33-40.
- ♦ **1-** P. Cossart, E. Groisman, **M.C. Serre**, M. Casadaban & B. Gicquel-Sanzey. (1986). *crp* genes of *Shigella flexneri*, *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.*, **167**, 639-646.

VIII- SEMINAIRES ET CONGRES

- 2005** ♦ Gordon Research Conference. Archaea: Ecology, Metabolism and Molecular Biology. Oxford (UK). **Poster**. « Site specific recombination in *Sulfolobales* : the SSV1 integrase shares unique properties with eukaryotic recombinases ».
 - ♦ Archaea, the first generation. Munich (All). **Poster**. « Site specific recombination in *Sulfolobales* : the SSV1 integrase shares unique properties with eukaryotic recombinases ».
- 2004** ♦ 8th Workshop on Site specific Recombination and Transposition, Woods Hole (USA). **Poster**. « Site specific recombination in archaea : catalytic properties of the thermophilic SSV1 integrase ».
 - ♦ Journées de l'I.G.M. **Communication orale**. « La conservation de fonction n'exclut pas la fantaisie: l'intégrase de SSV1, une tyrosine recombinase originale ».
- 2003** ♦ Institut de Génétique et Microbiologie. **Séminaire**. « Analyse fonctionnelle de la recombinaison spécifique de site chez *Sulfolobus* ».
- 2002** ♦ Extremophiles 2002, Naples (It). **Poster**. « Site-specific recombination in *Sulfolobales*: DNA recognition and cleavage by the SSV1 integrase ».
 - ♦ 7th Workshop on Site specific Recombination and Transposition, Oxford (UK). **Poster**. « Site-specific recombination in *Sulfolobales*: DNA recognition and cleavage by the SSV1 integrase ». (*Poster présenté par C. Letzelter*).
- 2001** ♦ L'ADN et ses machineries protéiques: aspects moléculaires et cellulaires. Figeac. **Poster**. « Etude fonctionnelle d'Int^{SSV}, une recombinase spécifique de site d'archae ».
 - ♦ Gordon Research Conference. Archaea: Ecology, Metabolism and Molecular Biology. Andover (USA). **Poster**. «Biochemical analysis of wild type and mutants SSV1 integrase». (*Poster présenté par C. Letzelter*).

- 1997** ♦ 2nd Colloque Biologie Moléculaire des Thermophiles. La Rochelle. **Communication orale**. « Reconnaissance et coupure de l'ADN par l'intégrase de SSV1 ».
♦ 2nd Colloque des 3R: Réplication, Recombinaison, Réparation, Villejuif, Juin 1997. **Poster**. « L'intégrase du virus SSV1 de l'archae *Sulfolobus shibatae* ».
- 1996** ♦ 1^{er} Colloque Biologie Moléculaire des Thermophiles, Gif-sur-Yvette. **Conférencier invité**. « Recombinaison chez les hyperthermophiles: l'intégrase du virus SSV1 de *Sulfolobus shibatae* ».
♦ 4th Workshop on Site specific Recombination and Transposition, Woods Hole (USA).
- 1995** ♦ 1^{er} Colloque des 3R, Réplication, Recombinaison, Réparation. Villejuif. **Poster**. « Caractérisation de mutants de la transposase d'*IS1* ».
♦ 4^{ème} Colloque Eléments Transposables, Versailles. **Poster**. « Caractérisation de mutants de la transposase d'*IS1* ».
- 1994** ♦ 3^{ème} Colloque Eléments Transposables, Clermont Ferrand. **Communication orale**. « Purification et propriétés de la transposase d'*IS1* ».
♦ Keystone Symposium, Transposition and Site-specific Recombination, Park City (USA). **Poster**. « Purification and properties of the native *IS1* transposase ».
- 1993** ♦ Département de Biologie Moléculaire, Université Science II de Genève. **Séminaire sur invitation**. « Mécanisme moléculaire de la recombinase FLP de levure ».
- 1992** ♦ 2nd Workshop on Site specific Recombination and Transposition, Berlin (RFA). **Poster**. « DNA splicing by an active site mutant of Flp recombinase: catalytic cooperativity between the inactive protein and its DNA substrate ».
♦ 1^{er} Colloque Eléments Transposables, Gif-sur-Yvette. **Communication orale**. « Mécanisme de la recombinaison site-spécifique médiée par Flp: substrats suicides et catalyse ».
- 1991** ♦ FASEB meeting on Genetic Recombination and genome rearrangements, Saxtons River (USA). **Poster**. « Half-site strand transfer reactions by Flp and step-arrest mutants of Flp ».
♦ Centre de Recherche de Biochimie et de Génétique Cellulaires, Toulouse. **Séminaire sur invitation**. « Mécanisme de la recombinaison site-spécifique catalysée par la protéine FLP de *S. cerevisiae* ».
♦ Laboratoire d'Enzymologie du CNRS, Gif-sur-Yvette. **Séminaire sur invitation**. « Mécanisme de la recombinaison site-spécifique catalysée par la protéine FLP de *S. cerevisiae* ».
- 1990** ♦ American Microbiology Society Meeting, Austin (USA). **Communication orale**. « Catalysis of site-specific recombination by the FLP recombinase from *S. cerevisiae* ».
- 1987** ♦ Protein Engineering Meeting, Oxford (UK). **Communication orale**. « Mutagenesis of Arg 180 of the *crp* gene of *Escherichia coli* K12 ».
- 1986** ♦ Unité INSERM U118, Paris. **Séminaire sur invitation**. « Utilisation du bactériophage M13 en Biologie Moléculaire ».

IX- PROCEEDINGS

- M.C. Serre, C. Turlan & M. Chandler.** (1994). Purification and properties of the native *IS1* transposase. *J. Cell. Biochem. Suppl.* 18B, 38.
- M.C. Serre, B. Gicquel-Sanzey & P. Cossart.** (1987). Mutagenesis of Arg 180 of the *crp* gene of *Escherichia coli* K12. *Prot. Engineer.*, **1**, 235.

P. Cossart, **M.C. Serre**, B. Gicquel-Sanzey, R. Ebright & J. Beckwith (1986). Classical genetics and site directed mutagenesis in the study of the specific interaction with DNA and CAP, the cyclic-amp receptor protein in *E. coli*-K12. *J. Mol. Graphics*, **4**, 230.

X- AUTRES ACTIVITES

Depuis 2004 ♦ Membre nommée de la commission de spécialistes de la 64-65^{ème} section de l'Université Paris-Sud XI.

Depuis 2002 ♦ Membre élue du conseil de l'IGM (UMR8621).
♦ Coordinateur de la "Fête de la Science" pour l'IGM (UMR8621).

Depuis 1996 ♦ Personne Compétente en Radio-Protection.
♦ Membre de la Société de Biochimie et Biologie Moléculaire.

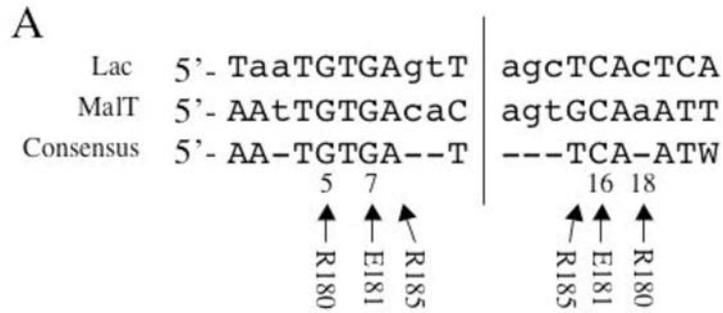
1993 - 1995 ♦ Membre élue du conseil Scientifique de l'UFR des Sciences de la Vie et de la Terre de l'Université Toulouse III (démissionnaire pour cause de mobilité).

AVANT PROPOS

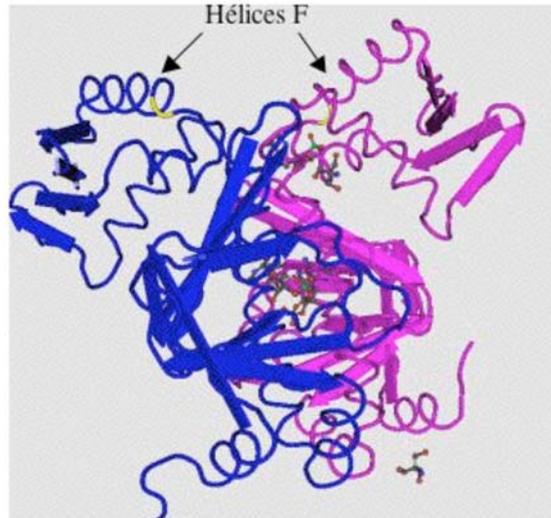
L'essor de l'ingénierie des protéines date du milieu des années 80. Cette nouvelle méthodologie est née de la conjonction de plusieurs technologies, voire de plusieurs disciplines, et a grandement bénéficié du développement de la technique de mutagenèse dirigée. C'est à la même période que je fis mes premiers pas dans la recherche. La perspective de pouvoir développer une approche moléculaire de l'enzymologie me sembla immédiatement attractive, et les années passées n'ont fait que conforter mon choix initial.

Après un DEA de Microbiologie (1986), sous la direction de Pascale Cossart, puis un doctorat de Microbiologie de l'université Pierre et Marie Curie (1989), sous la direction de Jean-Renaud Garel, je découvris le monde de la recombinaison spécialisée lors de mon stage post-doctoral chez Makkuni Jayaram à l'Université du Texas à Austin (1990-1991). C'est pendant ce séjour que j'ai été recrutée comme CR1 au CNRS (concours 1991) pour développer des approches *in vitro* de la transposition dans le laboratoire de Mick Chandler à Toulouse (1992-1995). Après ces trois années, j'ai souhaité pouvoir développer ma propre thématique sur la recombinaison spécifique de site, en choisissant comme modèle d'étude l'intégrase d'un virus d'archae hyperthermophile. Dans un premier temps, j'ai été hébergée dans l'équipe de Jean-Renaud Garel à Gif-sur-Yvette (1995-1998) puis en 1999 j'ai rejoint le laboratoire de Michel Duguet à Orsay.

Dans ce document je présenterai succinctement ma formation initiale (1985-1989), de manière plus détaillée mes activités dans le domaine de la recombinaison spécialisée (à partir de 1990), et je terminerai par les projets que je souhaite développer à moyen terme.



B



C

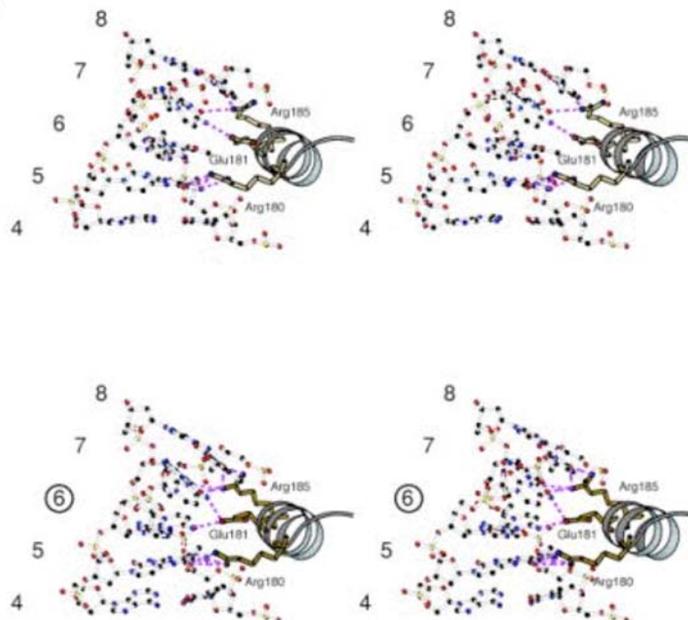


Figure 1 : Reconnaissance des sites de fixation par la protéine CAP.

A : Séquences de deux sites de fixation de CAP et du consensus déduit de l'alignement de différents sites. Les interactions proposées sont représentées **B** : Structure du complexe CAP-cAMP à 1.9 Å (PDB accession 1I5Z). La position Arg180 située sur l'hélice F est indiquée en jaune. **C** : Détail des interactions entre l'hélice F et le site de fixation (Chen *et al.*, 2001).

I- FORMATION INITIALE (1985 – 1989)

Unité de Biochimie Cellulaire, Institut Pasteur, septembre 1985 - septembre 1986

Laboratoire d'Enzymologie, CNRS Gif-sur-Yvette, octobre 1986 - décembre 1989

P. Cossart, E. Groisman, **M.C. Serre**, M. Casadaban & B. Gicquel-Sanzey. (1986). *crp* genes of *Shigella flexneri*, *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.*, **167**, 639 - 646.

M.C. Serre & J.R. Garel. (1990). Role of the C-terminal region in the allosteric properties of *Escherichia coli* Phosphofructokinase. *Eur. J. Biochem.*, **189**, 487 - 492.

M.C. Serre, W. Teschner & J.R. Garel. (1990). Specific suppression of heterotropic interactions in Phosphofructokinase by the mutation of Leucine 178 to Tryptophan. *J. Biol. Chem.*, **265**, 12146 - 12148.

W. Teschner, **M.C. Serre** & J.R. Garel. (1990a). Introduction by site-directed mutagenesis of a tryptophan residue as a fluorescent probe for the folding of *Escherichia coli* phosphofructokinase. *Biochimie*, **72**, 403 - 406.

W. Teschner, **M.C. Serre** & J.R. Garel. (1990b). Enzymatic properties, renaturation and metabolic role of Mannitol-1-Phosphate Dehydrogenase from *Escherichia coli*. *Biochimie*, **72**, 33 - 40.

J'ai effectué mon stage de DEA, dans l'unité dirigée par Georges Cohen à l'institut Pasteur, sous la direction de Pascale Cossart. Le projet consistait à étudier les interactions entre la protéine CAP (Catabolite gene Activator Protein) d'*E. coli* et ses sites spécifiques de fixation à l'ADN, en utilisant la mutagenèse dirigée pour générer une collection de mutants sur une position précise de la protéine. La protéine CAP et son cofacteur l'AMPc activent l'expression de nombreux opérons, en particulier des opérons cataboliques. L'activation intervient par la fixation de ce complexe sur des sites spécifiques localisés dans les régions promotrices des opérons régulés. Les sites de fixation de CAP ont une taille d'environ 22 pb, sont organisés autour d'un axe de symétrie (**Figure 1A**), et ont une position variable par rapport au démarrage de la transcription. La structure de CAP déterminée à 2.9 Å (McKay & Steitz, 1981) avait permis d'identifier un motif à deux hélices (α hélice - tour - α hélice) localisé dans le domaine C-terminal (**Figure 1B**) proposé interagir avec le grand sillon de l'ADN. D'autre part, l'étude de mutants obtenus au laboratoire avait permis de modéliser certaines interactions entre la protéine et ses sites (Ebright *et al.*, 1984a, 1984b). Ainsi, il avait été proposé que le résidu Glu181 interagisse avec les paires de bases 7 et 16 du site de fixation (Ebright *et al.*, 1984a) et que les résidus Arg180 et Arg185 contactent respectivement les paires de base 5 et 18 et 8 et 15 du site de fixation (**Figure 1A**). J'ai utilisé la mutagenèse dirigée pour tester la prédiction concernant la position Arg 180. Pour introduire le plus grand nombre possible de mutations à la position cible, je choisis d'utiliser un oligodéoxynucléotide dégénéré sur les trois bases du codon Arg180. Cette approche me permit d'obtenir six classes de mutants à cette position, correspondant au changement de l'Arg en Cys, Ser, Pro, His, Leu et Gly. Le taux médiocre de mutagenèse (11%) et la faible diversité des mutants obtenus peuvent être dûs à différents facteurs. D'une part, les méthodologies permettant d'augmenter le taux de mutagenèse n'étaient pas encore décrites (double mutagenèse avec marqueur de sélection, utilisation de souche *dut ung*, etc). D'autre part la séquence sauvage se trouvant dans le mélange d'oligonucléotides avait très probablement introduit un biais lors de l'hybridation à la cible. La répartition des mutants en fonction du nombre de bases

modifiées était d'ailleurs totalement biaisée, seuls 3 mutants sur 33 correspondaient à une double mutation, aucun triple mutant n'étant obtenu. Enfin, la redondance du code génétique a également contribué à diminuer le nombre de mutants exploitables. En effet, sur les 33 mutants obtenus, 4 correspondaient à des mutations silencieuses Arg (deux types de codons) et 5 à des mutants Leu (3 types de codons).

Dans un second temps, j'ai commencé le reclonage des différents mutants dans un plasmide permettant d'exprimer CAP sous contrôle de son propre promoteur. J'ai analysé le phénotype d'un de ces mutants (Arg180Cys) sur milieu indicateur McConkey (**Tableau 1**).

| Souche | Lac | Mal | Gal |
|-------------------|-----|-----|-----|
| <i>WT</i> | + | + | + |
| Δcrp | - | - | + |
| <i>crp</i> -R180C | +/- | - | + |

Tableau 1 : Analyse phénotypique de mutants de CAP.

crp : gène codant CAP. Les trois souches ont été striées sur des boîtes de milieu McConkey supplémentées avec du lactose (Lac), du maltose (Mal) ou du galactose (Gal) à une concentration finale de 1%. L'utilisation du galactose n'est pas CAP-dépendante.

Le phénotype observé pour le mutant Arg180Cys était intermédiaire entre celui d'une souche sauvage et celui d'une souche dépourvue de CAP. Des deux positions du site proposées être contactées par le résidu Arg180, une diffère entre le site *lac* et le site *malT*. Cette différence de base à la position 18 peut être à l'origine de la différence de phénotype sur lactose et maltose, et était compatible avec le rôle prédit pour Arg180 dans les interactions entre CAP et l'ADN (Serre *et al.*, 1987). La structure de CAP fixée à l'ADN a depuis été résolue (Schultz *et al.*, 1991) et l'analyse du complexe a validé les modèles d'interaction proposés (**Figure 1C**).

Parallèlement à ce travail, je participai à la détermination de la séquence des gènes *crp* de *Shigella flexneri* et *Salmonella typhimurium* qui était en cours dans le laboratoire (Cossart *et al.*, 1986). Le gène *crp* de *S. flexneri* ne présente que 4 paires de bases différentes de celui d'*E. coli*, et ces 4 modifications sont silencieuses au niveau protéique. Le gène *crp* de *S. typhimurium* est plus divergent au niveau nucléotidique (12,3%) mais conduit à une protéine très conservée, un seul acide aminé étant différent de la protéine CAP d'*E. coli* (Ala118Ser). La comparaison des séquences nucléotidiques des gènes *crp* et des séquences peptidiques déduites nous a permis de montrer que de toutes les protéines étudiées chez ces trois organismes (en 1986...), CAP était la plus conservée.

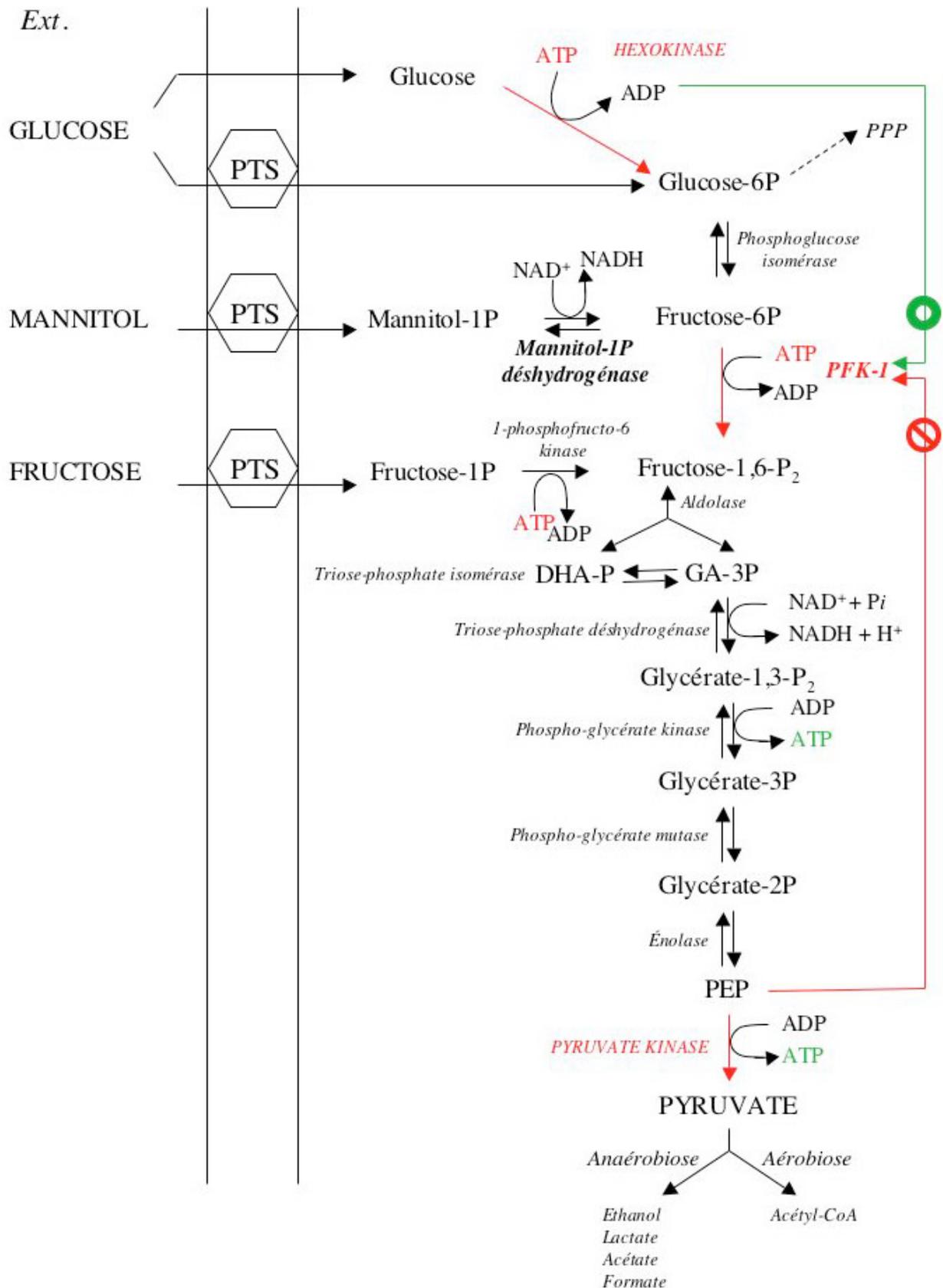


Figure 2 : Entrée dans la glycolyse de quelques hexoses chez *Escherichia coli*.

Les deux enzymes étudiées au cours de ma thèse sont indiquées en gras.

Les enzymes régulées allostériquement sont indiquées en rouge. Selon qu'elles soient consommées ou produites, les molécules d'ATP sont écrites en rouge ou en vert. Les effecteurs de la PFK sont signalés par des flèches rouge (inhibition) ou verte (activation).

La fin de mon stage de DEA a coïncidé avec le départ à la retraite de Georges Cohen, et la restructuration de son laboratoire. Pascale Cossart rejoignait l'unité de Biotechnologies dirigée par Julian Davies pour y démarrer l'étude de la pathogénicité des *Listeria*. Je choisis pour ma part d'effectuer ma thèse sous la direction de Jean-Renaud Garel qui partait s'installer au laboratoire d'Enzymologie à Gif-sur-Yvette. Il me semblait en effet essentiel d'acquérir des connaissances en structure des protéines et enzymologie, si je voulais pouvoir utiliser avec efficacité la mutagenèse dirigée, en d'autres termes prédire efficacement l'effet de mutations introduites sur l'activité biologique d'une protéine. Le groupe de Jean-Renaud étudiait la phosphofructokinase (PFK) d'*E. coli* et outre l'étude de ses propriétés enzymatiques, étudiait le repliement de cette protéine *in vitro*. Je devais pour ma part développer les méthodes de biologie moléculaire dans le laboratoire et obtenir par mutagenèse dirigée des mutants de la PFK affectés dans leurs propriétés allostériques.

Chez *E. coli*, la PFK est considérée comme l'enzyme clé de la glycolyse (voie d'Embden-Meyerhof), une des voies centrales du métabolisme intermédiaire du glucose (**Figure 2**). Deux facteurs importants nécessitent un contrôle strict du flux métabolique à travers cette voie. Le premier concerne les variations métaboliques intervenant lors de la balance entre aérobose et anaérobose, le flux glycolytique augmentant en anaérobose. Si l'on considère la glycolyse en terme de production nette d'ATP, l'augmentation du flux métabolique permet de compenser le plus faible rendement énergétique de la glycolyse anaérobie. Le second concerne la balance catabolisme/anabolisme, et donc la nécessité de réguler strictement glycolyse et gluconéogenèse qui ont en commun la plupart de leurs enzymes, à l'exception de la PFK et de la pyruvate kinase. Ces deux enzymes catalysent des réactions irréversibles et sont toutes deux soumises à une régulation allostérique de leur activité. L'activité de la PFK est sensible à deux facteurs contrôlant la glycolyse, d'une part le potentiel énergétique de la cellule (activation par l'ADP et rétro-inhibition par le PEP) et d'autre part la quantité de glucose disponible (coopérativité pour le substrat Fru-6P). La coopérativité de fixation du substrat fait intervenir des interactions homotropes (de site actif à site actif), et la modulation de l'affinité de l'enzyme pour ce substrat des interactions hétéotropes (de site effecteur à site actif). Des résultats antérieurs obtenus au laboratoire avaient mis en évidence l'importance de la région C-terminale dans la transmission des effets hétéotropes, une enzyme tronquée de ses 40 à 50 derniers résidus par protéolyse ménagée étant toujours active mais insensible aux effecteurs (Le Bras & Garel, 1982, 1986). Il avait été proposé que cette région de la protéine serait impliquée dans la formation des sites effecteurs, hypothèse confirmée par les données cristallographiques obtenues ultérieurement (Evans *et al.*, 1986; Shirakihara & Evans, 1988) (**Figure 3**). Par ailleurs, l'enzyme protéolysée (PFK*) montrait une coopérativité diminuée vis à vis du Fru-6P, bien que son affinité pour ce substrat soit inchangée. La région C-terminale de la protéine semblait donc être également impliquée dans la transmission des effets homotropes.

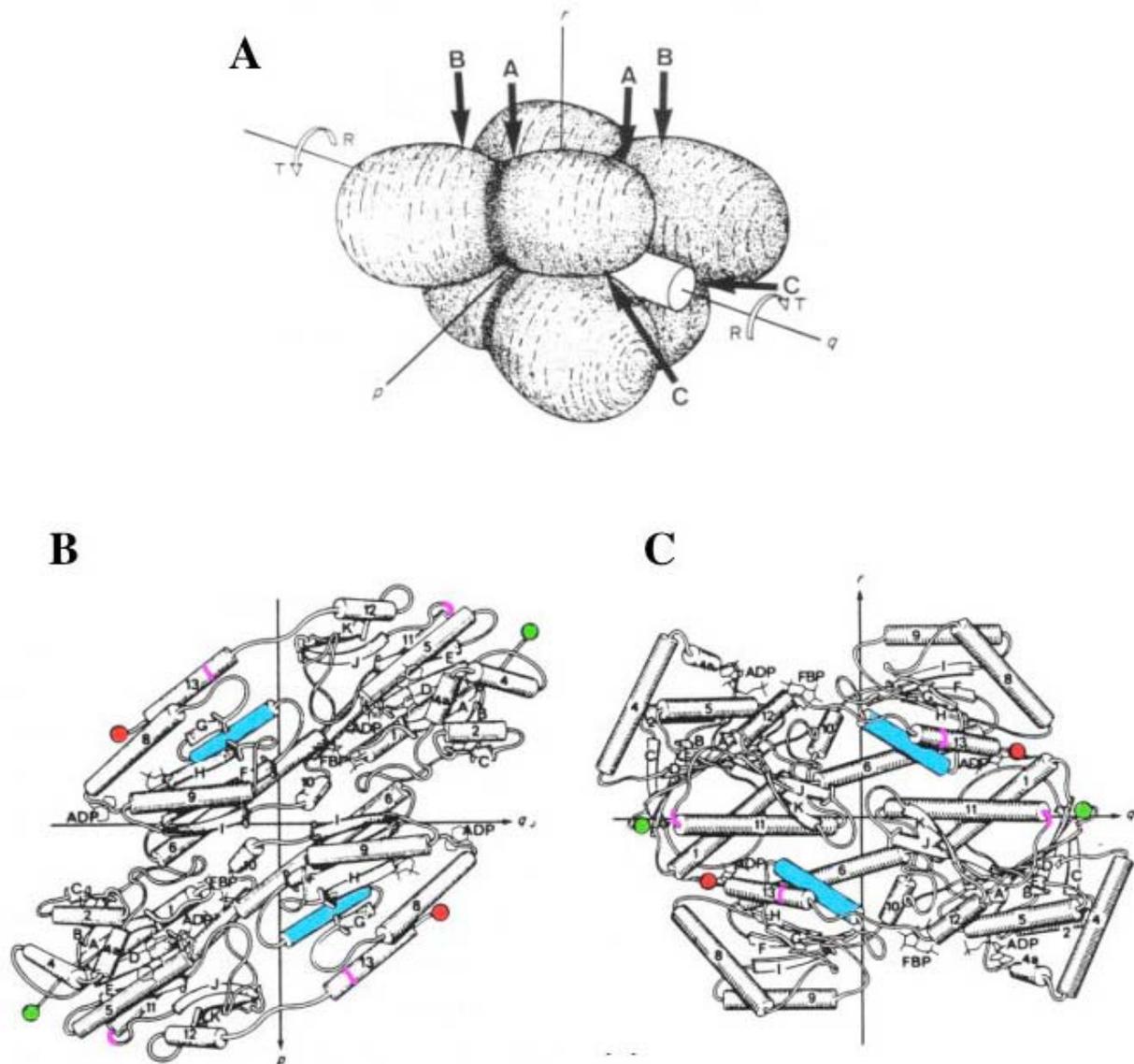


Figure 3 : Structure de la PFK d'*Escherichia coli*.

A: Représentation schématique du tétramère d'après (Evans & Hudson, 1979). Les sites A et B forment le site actif et lient respectivement le Fru-6P et l'ATP. Le site C est le site effecteur où se fixent l'ADP ou le PEP. La rotation d'un dimère par rapport à l'autre, effectuée lors de la fixation du PEP est également représentée. **B et C:** Vues des interfaces formant deux sites actifs (B) ou deux sites effecteurs (C). Le site actif est représenté en présence des produits de la réaction (ADP et Fru-1,6P₂) et le site effecteur en présence d'ADP (d'après (Shirakihara & Evans, 1988)). Les ronds vert correspondent à l'extrémité N-terminale et les ronds rouge à l'extrémité C-terminale des chaînes peptidiques. L'hélice α_7 est colorée en bleu. Les positions soulignées en rose correspondent aux extrémités des délétions C-terminales.

Enfin, la forme tétramérique de la PFK* n'était pas stable et se dissociait spontanément en dimères en absence de Fru-6P (Le Bras & Garel, 1985). *In vitro*, les 40 à 50 derniers acides aminés de la PFK semblaient donc largement contribuer à la stabilité de la protéine.

Afin d'étudier le rôle de la région C-terminale de la PFK dans la transmission des effets allostériques ainsi que dans la stabilité de la protéine *in vivo* et *in vitro*, j'ai créé deux délétions dans cette région en introduisant des codons stop par mutagenèse dirigée (**Figure 3B et C**). La première mutation conduisait à une délétion des 41 derniers acides aminés, et pouvait être considérée comme l'analogue *in vivo* de la PFK* active et non régulée obtenue *in vitro*. La seconde mutation conduisait à une délétion des 9 derniers acides aminés (Δ 311-319), qui devait permettre d'obtenir une protéine présentant des propriétés « intermédiaires » entre celles de la PFK et de la PFK*. En plus de la création d'un codon stop, les mutations introduites créaient des sites uniques de restriction, pouvant permettre par la suite de générer de nouveaux mutants de délétion dans la région C-terminale.

La délétion des 41 derniers acides aminés n'a pas permis d'obtenir *in vivo* une PFK active (phénotype *pfk*⁻). Plus précisément, il a été impossible de détecter la présence de cette protéine dans des extraits bruts, ce qui indiquait que la région C-terminale était essentielle pour le repliement ou l'assemblage de la PFK dans les conditions intracellulaires. A l'inverse, une délétion des 9 derniers résidus restaurait un phénotype *pfk*⁺ dans une souche Δ *pfk*, indiquant la présence d'une enzyme active *in vivo*. Il s'est avéré en cours de purification que ce mutant était instable en absence de Fru-6P, confirmant l'importance de la partie C-terminale de la chaîne peptidique dans la stabilité du tétramère. L'analyse des propriétés enzymatiques de ce mutant m'a permis de montrer que ni les propriétés catalytiques ni la transmission des effets homotropes n'étaient affectées par la délétion des 9 derniers acides aminés (Serre & Garel, 1990). L'étude de ses propriétés régulatrices a révélé que l'enzyme était toujours sensible à la présence des effecteurs allostériques, mais que son affinité pour l'un ou l'autre de ces effecteurs était très diminuée. Ce résultat était en accord avec les données cristallographiques indiquant que ce sont les mêmes résidus qui sont impliqués dans la fixation du GDP (ADP) et du PEP (Shirakihara & Evans, 1988). De manière plus surprenante, il s'est avéré que les propriétés catalytiques et régulatrices du mutant Δ (311-319) étaient en tout point comparables aux propriétés d'un mutant généré dans un groupe concurrent où seul le dernier résidu était délété (Δ 319, (Lau & Fersht, 1989)). Ainsi, l'effet de la délétion des 9 derniers acides aminés serait principalement de déstabiliser l'assemblage du tétramère, les altérations des paramètres enzymatiques pouvant être imputées au dernier résidu de la chaîne peptidique.

Certaines propriétés de la PFK, comme sa dissociation et sa réassociation mais également la fixation des effecteurs ou du substrat Fru-6P, pouvaient être suivies par des mesures de fluorescence. Le rapporteur de fluorescence était l'unique résidu Trp de la chaîne peptidique situé à

la position 311, et qui était donc absent dans les formes délétées de la PFK. Nous souhaitions pouvoir générer d'autres mutants de délétion de la région C-terminale pour identifier les structures impliquées dans la transmission des effets allostériques. Or de tels mutants n'auraient pu être soumis à une analyse par fluorescence. Nous avons alors décidé d'introduire un autre résidu Trp dans la chaîne peptidique, et avons défini un certain nombre de règles pour choisir la position de ce nouveau rapporteur de fluorescence : (1) le résidu cible ne devait pas être localisé dans les sites actifs ou effecteurs pour ne pas transmettre de signaux liés à la simple fixation des ligands. (2) Il devait cependant se situer dans une région qui variait entre les états « R » et « T » de la protéine. (3) Enfin, la mutation ne devait pas affecter la structure globale de la protéine. Le résidu choisi comme cible a été la Leu178, second acide aminé de l'hélice $\alpha 7$, située entre deux structures variant fortement entre les états libres et liés de la PFK (Rypniewski & Evans, 1989). Par ailleurs, aucun des résidus de cette hélice n'était impliqué dans la fixation des ligands (**Figure 3 B et C**), et l'unique résidu Trp de la PFK de *B. stearothermophilus* était situé dans l'hélice équivalente.

Les effets de la mutation Leu178Trp ont été évalués sur la PFK de taille sauvage. Cette mutation n'affecte ni le repliement ni l'assemblage d'une PFK active *in vivo*. Après avoir purifié ce mutant j'ai déterminé ses propriétés enzymatiques *in vitro*. La mutation Leu178Trp n'affecte pas les propriétés catalytiques ou les interactions homotropes de la protéine, et n'empêche pas la fixation des effecteurs allostériques GDP et PEP (Serre *et al.*, 1990). Cependant, l'activité enzymatique n'est plus sensible à la présence du PEP au site effecteur, suggérant que le résidu 178 (ou l'hélice dont il fait partie) serait impliqué dans la transmission des effets hétérotropes (**Tableau 2**).

| Ref. | Enzyme | [GDP] mM | [PEP] mM | k_{cat} (sec ⁻¹) | $S_{1/2}$ (Fru-6P, mM) | n_H |
|---------------------|--------|----------|----------|--------------------------------|------------------------|-----------|
| Serre <i>et al.</i> | WT | 0 | 0 | 120 | 0.5 | 3.8 ± 0.2 |
| | | | 10 | < 5 | 12 | 2 |
| | L178W | 0 | 0 | 60 | 0.45 | 3.3 ± 0.2 |
| | | | 10 | 60 | 0.62 | 3.2 ± 0.2 |
| Auzat <i>et al.</i> | WT | 0 | 0 | 120 | 0.35 | 3.8 ± 0.1 |
| | | 2 | | 137 | 0.041 | 1.1 ± 0.1 |
| | L178W | 0 | 0 | 60 | 0.23 | 2.7 ± 0.2 |
| | | 3 | | 95 | 0.098 | 2.7 ± 0.1 |
| | | | 10 | 60 | 0.45 | ND |
| | | | | | | |

Tableau 2 : Propriétés du mutant Leu178Trp vis à vis des effecteurs allostériques.

Les données regroupées dans ce tableau sont extraites de (Serre, 1989; Serre *et al.*, 1990; Auzat, 1993; Auzat *et al.*, 1997). ND, non déterminé.

Une analyse complémentaire du mutant Leu178Trp a été réalisée ultérieurement au laboratoire, montrant qu'il était sensible à l'activation par le GDP (Auzat *et al.*, 1997). L'activation observée est principalement due à une augmentation du k_{cat} plutôt qu'à une diminution du $S_{1/2}$, et comme cela avait été observé en présence de PEP, n'entraîne aucune modification de la coopérativité pour le Fru-6P (**Tableau 2**). Le mutant Leu178Trp ne pouvait donc pas être utilisé pour suivre les transitions allostériques induites par les effecteurs. Cependant, le Trp178 s'est avéré être un marqueur d'une étape tardive de la dénaturation de la PFK qui n'était pas accessible par des mesures de fluorescence chez l'enzyme sauvage. L'étude des propriétés de fluorescence du mutant Leu178Trp nous a ainsi permis de confirmer le mode de dissociation proposé pour la PFK (Deville-Bonne *et al.*, 1989) et de mettre en évidence que la dernière étape de cette dissociation est le dépliement complet de l'espèce monomérique (Teschner *et al.*, 1990a).

Une autre partie de mon travail de thèse a consisté à comprendre pourquoi la croissance sur mannitol comme seule source de carbone permettait de discriminer un phénotype *pfk*⁻ d'un phénotype *pfk*⁺. Le métabolisme du mannitol ne semble pas strictement dépendant de la conversion du Fru-6P en Fru-1,6-P₂ puisque l'interconversion entre le Fru-6P et le Glu-6P est une réaction réversible. Une souche *pfk*⁻ devrait pouvoir cataboliser le mannitol *via* le Glu-6P et le 6P-Gluconate (**Figure 2**). Nous avons purifié la mannitol-1-phosphate déshydrogénase (Mtl-1P-DH) qui catalyse la réaction



Deux enzymes portant cette activité avaient été purifiées par d'autres groupes, sans que leur régulation soit étudiée. L'une correspondait à une protéine monomérique de 40 kDa (Novotny *et al.*, 1984), l'autre à un dimère de deux peptides de 22 kDa (Chase 1986). L'enzyme que nous avons purifiée selon un nouveau protocole correspondait à un monomère de 45 kDa, ayant la même séquence N-terminale que celle purifiée par (Novotny *et al.*, 1984). L'analyse du génome d'*E. coli* révèle aujourd'hui que cet organisme ne possède qu'une seule Mtl-1P-DH, dont la masse moléculaire théorique est de 41 kDa. Nous avons ensuite analysé les propriétés enzymatiques de la Mtl-1P-DH, y compris son profil de pH dépendance, pour les réactions aller et retour. Nous avons ainsi montré que le pH optimal de la réaction d'oxydation du Mtl-1P en Fru-6P était de 9,5 alors que celui de la réaction de réduction du Fru-6P en Mtl-1P était de 7. Par ailleurs, pour une gamme de pH allant de 6,9 à 8,9, l'enzyme a 20 fois plus d'affinité pour le NADH que pour le NAD⁺, l'affinité pour les formes oxydées ou réduites du substrat étant identiques. Le pH intracellulaire d'*E. coli* a été estimé par de nombreux auteurs comme étant très constant, autour de 7.6 ± 0.2 , quelles que soient les conditions extérieures (Navon *et al.*, 1977; Ugurbil *et al.*, 1978; Slonczewski

et al., 1981). Dans les conditions physiologiques, le profil de pH dépendance de la réaction et la différence d'affinité de l'enzyme pour les formes réduites ou oxydées du coenzyme (Teschner *et al.*, 1990b) favoriseraient donc le déplacement de l'équilibre du Fru-6P vers le Mtl-1P, ce qui pourrait en absence d'une PFK fonctionnelle, empêcher la croissance sur mannitol comme seule source de carbone. Cette hypothèse est également renforcée par le fait que la phosphoglucose isomérase (**Figure 2**) fonctionnerait plus efficacement dans le sens Glu-6P -> Fru-6P que dans le sens Fru-6P -> Glu-6P (Schreyer & Bock, 1980).

A la fin de ma thèse, j'ai souhaité effectuer un séjour post-doctoral aux USA, pour acquérir une expérience différente tant sur le plan de la vie courante que sur le plan professionnel. Je désirais de plus travailler sur un système enzymatique utilisant l'ADN comme substrat, et un post-doctorat me semblait être idéal pour réaliser ce changement de thématique. C'est ainsi que je découvris le monde de la recombinaison spécialisée que je n'ai dès lors plus quitté.

II- RECOMBINAISON SPÉCIALISÉE

Les génomes de tous les organismes sont capables de promouvoir des réarrangements qui vont permettre de modifier ou diversifier (acquisition, délétion) le pool de leurs gènes et/ou le niveau de leur expression. Ces mécanismes de recombinaison génétique, essentiels à l'évolution, peuvent être classés en trois familles :

- Recombinaison homologue
- Recombinaison spécifique de site
- Recombinaison transpositionnelle

Les deux dernières sont souvent regroupées sous le terme de «recombinaison spécialisée». Ces trois voies se distinguent par les mécanismes mis en œuvre, les protéines impliquées et la nature de l'ADN recombiné. La recombinaison spécifique de site intervient entre deux sites ADN qui ont une courte région d'homologie (la séquence « core »). La coupure et l'échange de brin interviennent dans cette séquence sans qu'il y ait de perte ou de gain d'ADN ni de néo-synthèse d'ADN. A l'inverse, les réarrangements provoqués par des événements de recombinaison transpositionnelle ont un très fort caractère aléatoire. La majorité des éléments transposables ne requiert aucune similitude de séquence entre les différents partenaires ADN. Les éléments transposables

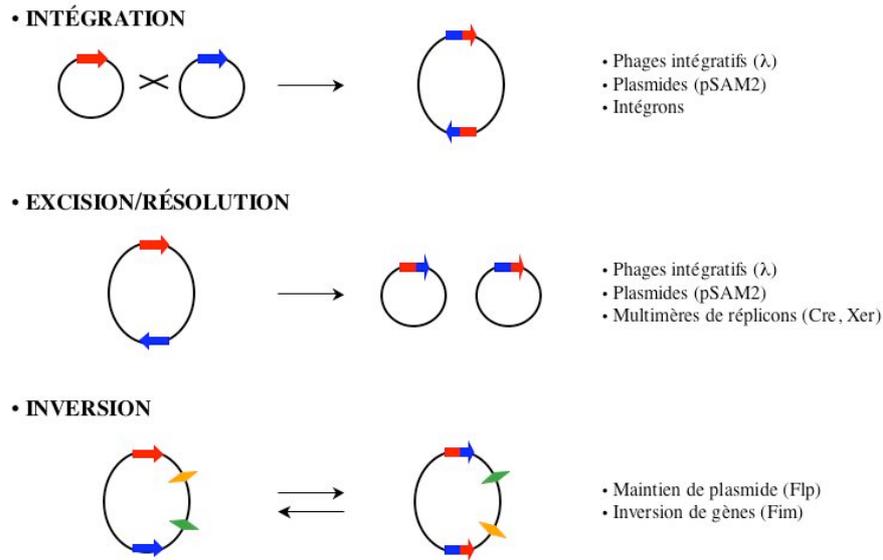


Figure 4 : Rôles biologiques de la recombinaison spécifique de site.

Des orientations relatives différentes entre les sites de recombinaison génèrent différents types de produits. Dans tous les cas, les mêmes évènements de coupure et d'échange de brins interviennent dans les régions d'homologies. Les sites « parentaux » sont représentés en rouge ou en bleu, les sites « recombinants » en rouge et bleu. Les évènements de recombinaison peuvent être intermoléculaire (intégration) ou intramoléculaires (inversion et excision/résolution). Lorsque les deux sites de recombinaison sont sur la même molécule, c'est leur orientation relative qui détermine le produit obtenu. S'ils sont en orientation directe, le fragment d'ADN qu'ils encadrent sera excisé/déleté/résolu sous forme circulaire. S'ils sont en orientation inverse, le fragment d'ADN qu'ils encadrent sera inversé.

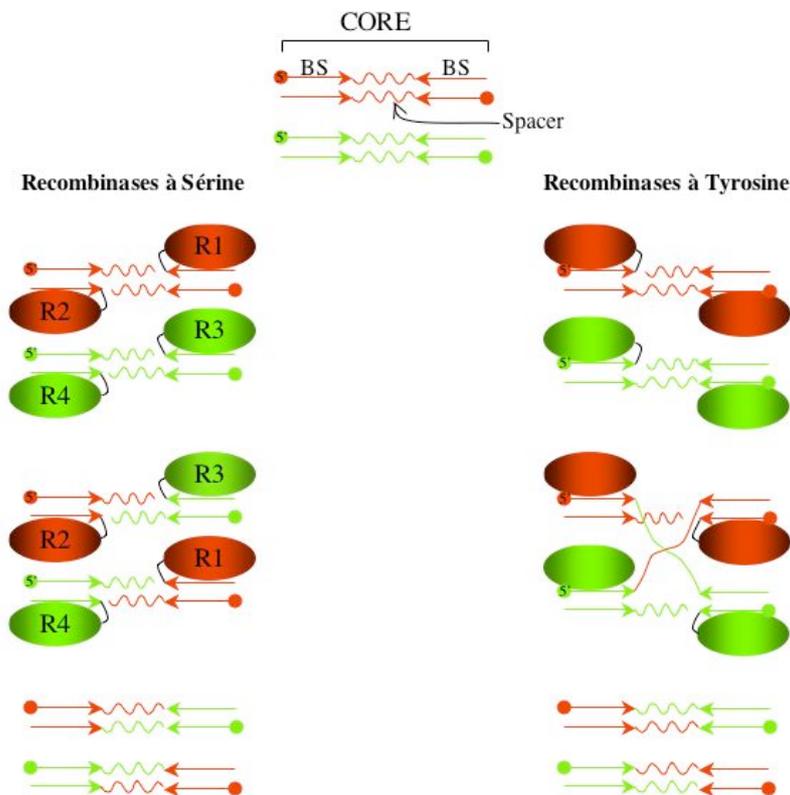


Figure 5 : Mécanismes de recombinaison spécifique de site.

L'organisation des sites core est similaire pour les deux familles de recombinaisons. Les produits de recombinaison sont différents. Les sites de fixation des recombinaisons (BS) sont représentés par des flèches qui encadrent la séquence spacer (représentée par une ligne ondulée).

représentent donc un moteur évolutif puissant, leur déplacement aléatoire étant une source de mutagenèse.

1- Recombinaison spécifique de site chez les eucaryotes

Université du Texas à Austin, janvier 1990 - décembre 1991

M.C. Serre, B. Evans, H. Araki, Y. Oshima & M. Jayaram. (1992). Half-site recombinations mediated by yeast site-specific recombinases Flp and R. *J. Mol. Biol.*, **225**, 621-642.

M.C. Serre & M. Jayaram. (1992). Half-site strand transfer by step-arrest mutants of yeast site-specific recombinase Flp. *J. Mol. Biol.*, **225**, 643-649.

J. Lee, **M.C. Serre**, S. H. Yang, I. Whang, H. Araki, Y. Oshima & M. Jayaram. (1992). Functional analysis of Box II mutations in yeast site-specific recombinases Flp and R. Significance of amino acid conservation within the Int family and the yeast sub-family. *J. Mol. Biol.*, **228**, 1091-1103.

M.C. Serre, L. Zheng & M. Jayaram. (1993). DNA splicing by an active site mutant of Flp recombinase. Possible catalytic cooperativity between the inactive protein and its DNA substrate. *J. Biol. Chem.*, **268**, 455-463.

J'ai rejoint le groupe de Makkuni Jayaram à l'Université du Texas à Austin, pour étudier le mécanisme de la recombinaison spécifique de site catalysée par la recombinase Flp de *S. cerevisiae*.

La recombinaison spécifique de site est un mécanisme conservé dans les trois règnes du vivant, qui permet de contrôler de nombreux aspects de la dynamique chromosomique (**Figure 4**). On peut ainsi citer les cycles d'intégration/excision des bactériophages ou des plasmides intégratifs, la conversion des dimères de réplicons (plasmides ou chromosomes) en monomères pour permettre leur ségrégation dans les cellules filles, le contrôle de l'expression de certains gènes, et le contrôle de la réplication de plasmides de levures.

La recombinaison spécifique de site est catalysée par des recombinases appelées intégrases, résolvasés ou invertases en fonction du type de recombinaison qu'elles catalysent. Ces recombinases se fixent sur des séquences spécifiques en orientation inverse comprises dans la séquence core, et séparées par une courte séquence d'ADN appelée « spacer » (**Figure 5**). Elles vont ensuite couper l'ADN de part et d'autre du spacer, chaque partenaire de recombinaison étant coupé à la même position, permettant ainsi un échange de brin réciproque dans la région d'homologie. Comme pour les topoisomérases, la coupure des brins d'ADN est accompagnée de la formation d'une liaison phosphodiester entre la recombinase et une extrémité du brin coupé. Cet intermédiaire covalent va permettre de conserver l'énergie de la liaison phosphodiester coupée et de la réutiliser lors de l'étape de religation de l'ADN. L'autre extrémité du brin coupé est libérée sous forme hydroxylée. Lors de l'échange de brin, c'est cette extrémité hydroxylée qui attaque la liaison phosphoprotéine du partenaire d'échange, permettant de régénérer un brin intact d'ADN et de recycler la protéine. Les recombinases spécifiques de sites sont regroupées en deux sous familles qui diffèrent par les chemins catalytiques suivis (**Figure 5**). Les recombinases à sérine forment un

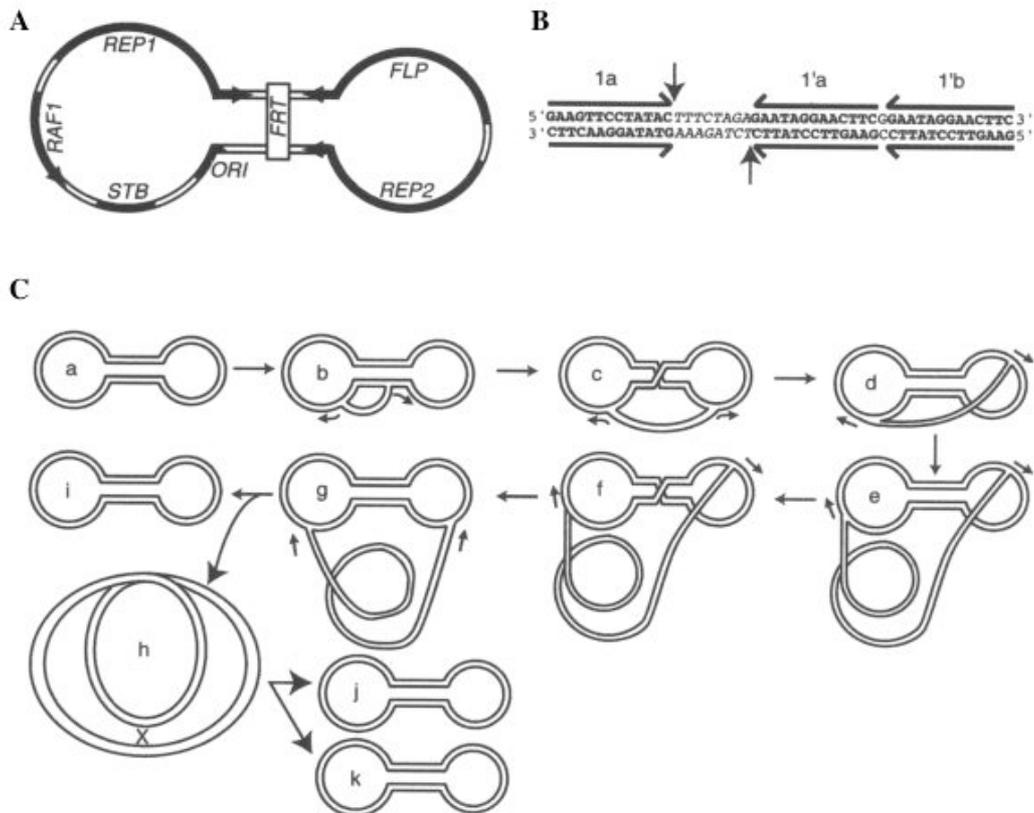


Figure 6 : Le plasmide 2 μ m de *Saccharomyces cerevisiae*.

A : Le plasmide 2 μ m est représenté de manière conventionnelle sous forme d'haltère, les lignes parallèles indiquant les séquences inversées répétées. Les gènes codant les protéines Rep1 et Rep2 ainsi que le locus STB sont les composants du système de partition du plasmide. La protéine Raf1 n'a pas été caractérisée, mais est très probablement un activateur de l'expression du gène codant la protéine Flp. Flp et ses sites cibles de recombinaison (FRT) constituent le système d'amplification du plasmide. L'origine de répllication du plasmide est également représentée (ORI). **B** : Séquence du site FRT. Les séquences permettant la fixation de Flp sont indiquées par des flèches horizontales tenant compte de leur orientation relative. La séquence du spacer est en italiques. Les flèches verticales indiquent la position des liaisons phosphodiester cibles de la coupure et de l'échange des brins. L'élément 1'b n'est pas essentiel pour la recombinaison. Le site FRT minimal est constitué des éléments 1a, 1'a et du spacer. **C** : Le modèle de Futcher pour l'amplification du plasmide est adapté de (Volkert & Broach, 1986). La propagation bidirectionnelle des fourches de répllication (b) est convertie par un événement de recombinaison Flp dépendant en propagation unidirectionnelle des fourches (c, d) qui va permettre l'amplification du plasmide (d à f). Un second événement de recombinaison (f) va terminer l'étape d'amplification (g). Les produits de répllication (h) et (i) sont des formes multimériques et monomériques du plasmide. La résolution de (h) va générer des monomères de plasmide (j, k, etc).

intermédiaire 5'-phospho-sérine et libèrent une extrémité 3'-OH. Elles coupent les quatre brins simultanément puis permutent les complexes covalents deux à deux. Les recombinases à tyrosine forment un intermédiaire 3'-phospho-tyrosine et libèrent une extrémité 5'-OH. Elles coupent et échangent les quatre brins séquentiellement, et génèrent une jonction de Holliday comme intermédiaire réactionnel.

La protéine Flp est une recombinase à tyrosine codée par le plasmide $2 \mu\text{m}$, un élément extrachromosomique multicopie présent dans de nombreuses souches de *Saccharomyces*. Ce plasmide contient deux copies d'une séquence de 600 pb en orientation inverse (**Figure 6A**). Les répétitions inverses contiennent les sites de recombinaison utilisés par Flp (**Figure 6B**) et divisent le plasmide en deux domaines. La réplication du $2 \mu\text{m}$ est soumise au contrôle du cycle cellulaire, un événement d'initiation intervenant par phase S. Néanmoins, le nombre de copies du plasmide est très élevé, et peut revenir à son niveau normal très rapidement en réponse à des événements qui diminuent son nombre de copies. Le modèle proposé par Futcher (Futcher, 1986) postule que l'amplification du nombre de copies plasmidiques soit due à la position asymétrique de l'origine de réplication par rapport aux sites de recombinaison de Flp (**Figure 6A**). Un événement de recombinaison entre les sites après qu'une des deux fourches de réplication ait passé le site proximal, conduirait à une progression monodirectionnelle de ces fourches, et par conséquent à la production de plusieurs copies de plasmide à partir d'une seule initiation de réplication (**Figure 6C**). La prédiction essentielle de ce modèle est que l'absence d'un système de recombinaison fonctionnel abolirait le processus d'amplification. Cette prédiction a été vérifiée, même si l'existence de l'intermédiaire de réplication par cercle roulant prédit par le modèle n'a jamais été rigoureusement démontrée (Volkert & Broach, 1986).

La protéine Flp appartient à la famille des recombinases à tyrosine (Argos *et al.*, 1986; Esposito & Scocca, 1997; Nunes-Düby *et al.*, 1998). Au moment où j'ai commencé ce travail, seuls trois résidus du site actif avaient été identifiés sur la base d'alignements de séquence (Argos *et al.*, 1986), la triade His-Arg-Tyr, et leur implication dans le mécanisme de recombinaison évaluée par mutagenèse dirigée (Prasad *et al.*, 1987; Parsons *et al.*, 1988; Parsons *et al.*, 1990). Il était alors proposé que His305 serait essentiel pour l'échange de brin mais pas pour la coupure, Arg308 interviendrait à l'étape de coupure, probablement en assurant le positionnement correct du phosphate cible, et que Tyr343 serait la tyrosine catalytique, hypothèse démontrée ultérieurement (Evans *et al.*, 1990). D'autres recombinases de plasmides apparentés au $2 \mu\text{m}$ avaient été identifiées (Utatsu *et al.*, 1987) dans des levures des genres *Zygosaccharomyces* et *Kluyveromyces*. L'alignement de leur séquence avec celle de la protéine Flp a montré l'existence de deux motifs fortement conservés (Box I et Box II). Le second de ces motifs comprenant la triade commune aux

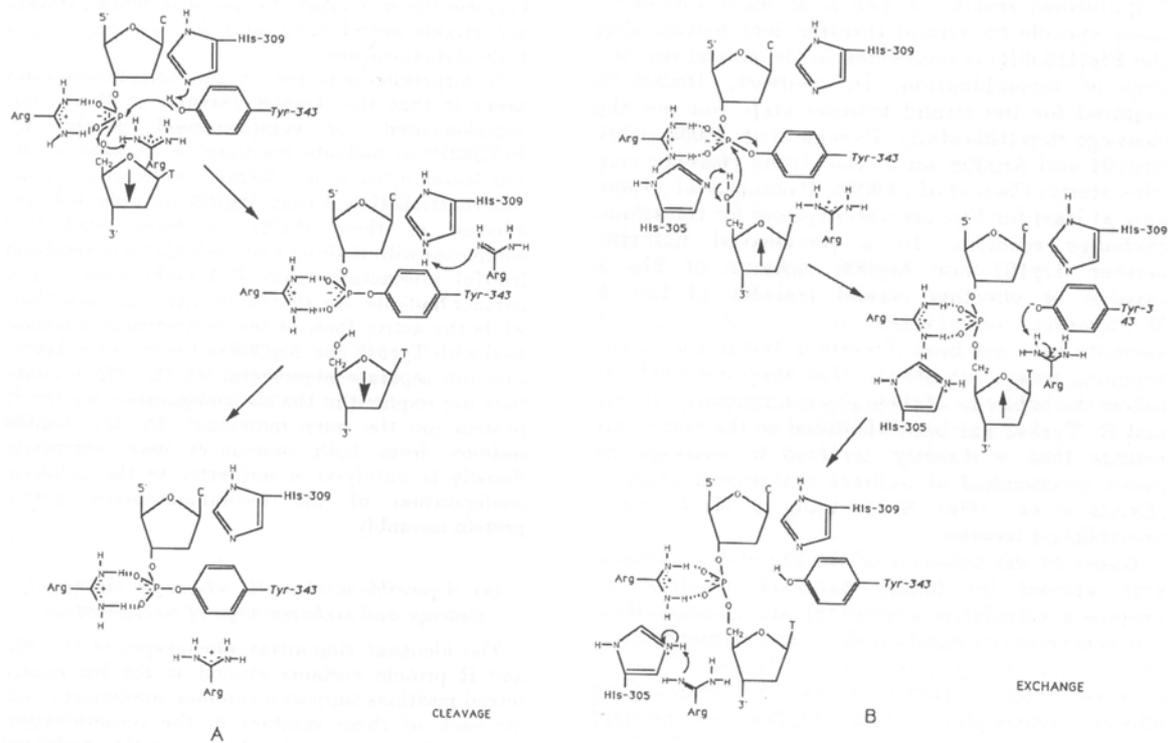


Figure 7 : Modélisation des réactions de coupure et d'échange de brins catalysés par Flp. (d'après Lee *et al.*, 1992).

Le rôle proposé pour les résidus Arg191, His305, Arg308, His309 et Tyr343 est représenté dans les étapes de coupure (A) et d'échange de brin (B). Les résultats des tests d'activité ne permettent pas de discriminer l'implication des deux Arg dans ce mécanisme. Leurs positions ne sont donc pas précisées dans les schémas. Les nucléophiles actifs sont respectivement l'hydroxyl de Tyr343 dans la réaction de coupure et le 5'-hydroxyl du brin coupé dans la réaction d'échange. Le résidu Tyr343 provient d'un monomère différent des autres résidus (Chen *et al.*, 1992).

recombinases à tyrosine, il était donc postulé que ces deux régions conservées contenaient des résidus du site catalytique.

J'ai abordé l'étude mécanistique de la réaction de recombinaison catalysée par Flp selon deux approches complémentaires. D'une part j'ai identifié de nouveaux résidus du site actif dans le motif 'Box II' conservé chez les recombinaisons de levures, et d'autre part j'ai développé de nouveaux substrats (demi-sites) de recombinaison. Ces deux approches m'ont permis de modéliser le mécanisme enzymatique de la réaction de recombinaison alors qu'aucune donnée structurale n'était encore disponible pour la famille des recombinaisons à tyrosine.

a- Identification et modélisation du rôle des résidus conservés entre Flp et d'autres recombinaisons de levure

L'alignement des séquences en acides aminés des protéines de différentes sous espèces de levures appartenant à la famille des recombinaisons a permis d'identifier deux régions d'homologies, l'une correspondant aux résidus 185 à 203 (Box I) et l'autre aux résidus 295 à 313 (Box II) (Utatsu *et al.*, 1987). Sur les 19 acides aminés inclus dans la seconde région, 9 sont strictement conservés pour toutes les protéines. Parmi ces 9 résidus, deux avaient déjà été identifiés comme membres du site actif, His305 et Arg308 (Parsons *et al.*, 1988; Parsons *et al.*, 1990). Il semblait raisonnable de considérer que la seconde région conservée chez les recombinaisons de levure fasse partie de leur site actif. J'ai modifié 5 des 7 autres positions strictement conservées par mutagenèse dirigée, et étudié les propriétés enzymatiques des différents mutants obtenus. Parallèlement Jehee Lee, étudiant que j'encadrais pendant sa 1^{ère} année de thèse, a réalisé la même étude sur la protéine R, recombinaison du plasmide pSR1 de *Zygosaccharomyces rouxii* (Araki *et al.*, 1992). Les mutants obtenus au cours de ce travail m'ont permis d'affiner le mécanisme de recombinaison site-spécifique proposé précédemment (Evans *et al.*, 1990).

Les propriétés des différentes protéines mutantes m'ont conduit à proposer que Phe296 et Leu310 seraient impliqués dans le positionnement correct du squelette sucre-phosphate de l'ADN dans la poche catalytique, par l'intermédiaire d'interactions de type hydrophobe. Ile298 pourrait également être engagé dans le même type d'interaction. Lys303 semblait également jouer un rôle dans la fixation de l'ADN, mais ce résidu chargé pourrait établir des interactions plus spécifiques, par exemple en établissant des liaisons hydrogènes avec un phosphate particulier du site de fixation.

De toutes les positions mutagénisées, seule His309 semblait être directement impliquée dans la catalyse de la réaction. Un autre acide aminé, Arg191, semblait également être directement impliqué dans le mécanisme catalytique (Abremski & Hoess, 1992; Chen *et al.*, 1992a). J'ai ainsi modélisé les réactions de coupure et d'échange de brin sous la forme d'une réaction de transfert de proton (**Figure 7**).

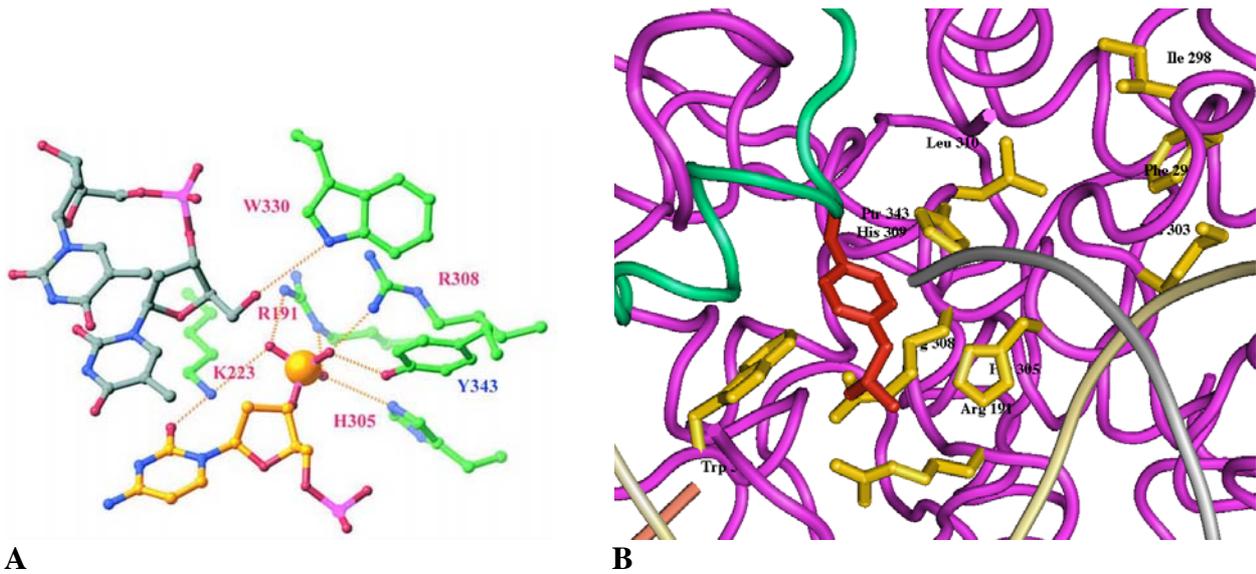


Figure 8 : Structure locale du site actif de Flp.

A : Résidus directement impliqués dans la catalyse (d'après Chen *et al.*, 2000).

B : Les chaînes latérales des résidus mutagénisés ou cités dans le texte sont représentées. Vert : Hélice d'une sous unité fournissant la tyrosine catalytique (représentée en rouge). Violet : structure de la seconde sous unité formant la poche du site actif. Trois des quatre brin d'ADN apparaissent dans cet agrandissement (représentés en gris, bronze et or).

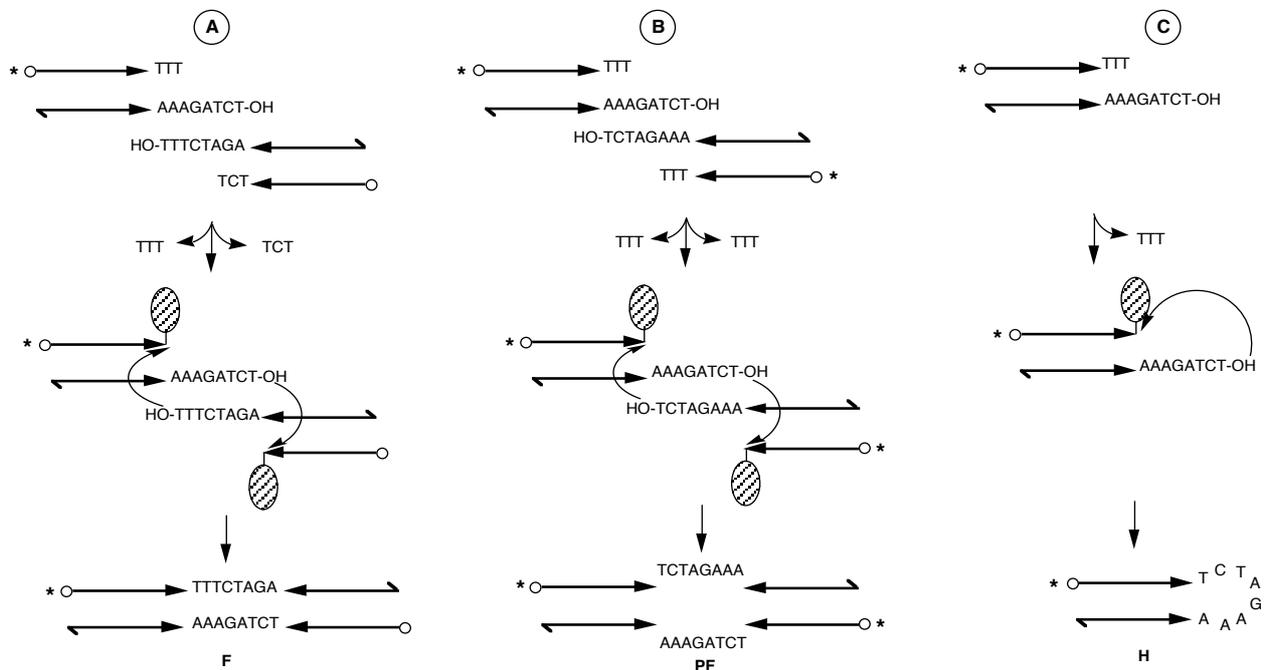


Figure 9 : Structure des demi-sites et produits de la réaction avec Flp. (d'après Serre *et al.*, 1992)

A : Recombinaison entre un demi-site droit et un demi-site gauche. Le produit de la réaction est un site complet (**F**). **B, C :** Recombinaison en présence d'un demi-site gauche. Une réaction intermoléculaire (**B**) conduit à l'obtention d'un pseudo-site complet (**PF**), alors qu'une réaction intramoléculaire (**C**) conduit à une structure en boucle (**H**).

L'analyse de la structure de Flp obtenue récemment (Chen *et al.*, 2000) permet de ré-évaluer le rôle de ces différents résidus (**Figure 8**). Le rôle joué par His309 n'est pas d'activer l'hydroxyl de la tyrosine catalytique comme nous l'avions modélisé (**Figure 7**), mais plutôt de positionner correctement ce résidu probablement par le biais d'interactions d'empilement (**Figure 8B** et Chen & Rice, 2003). Les deux arginine (Arg191 et Arg308) seraient bien impliquées dans la présentation et la stabilisation de la liaison phosphodiester cible (**Figure 8 A et B**), et His305 dans la stabilisation du brin sortant et/ou l'activation du brin entrant. Les autres positions communes aux recombinaisons de levure ne sont pas directement impliquées dans la catalyse mais sont vraisemblablement importantes pour le repliement correct du site actif (**Figure 8B**).

b- L'utilisation de demi-sites de recombinaison révèle les étapes précoces de la réaction de coupure et les propriétés du site actif de Flp

La mise au point de substrats synthétiques correspondant à des demi-sites de fixation m'a permis de procéder à une analyse fine du mécanisme de coupure. En présence de tels sites, Flp est capable de catalyser des événements de recombinaison intermoléculaire ou intramoléculaire, ce qui indiquait que la formation du complexe synaptique ne nécessitait pas de liaison physique entre chaque site de fixation de Flp (**Figure 9**). La position de la coupure s'est révélée strictement spécifique, et probablement déterminée par les contraintes de fixation de Flp sur son site. A l'inverse, l'étape d'échange de brin est moins stringente, puisqu'elle peut être initiée par différents types d'extrémités 5'-OH disponibles (intra ou intermoléculaires). Si l'extrémité 5'OH libre du spacer est phosphorylée (substrat "suicide"), Flp ne peut plus procéder à l'échange de brin, et la réaction s'arrête après coupure de la liaison phosphodiester cible, révélant que coupure et échange de brin sont des événements séquentiels plutôt que concertés. Enfin, la taille du spacer joue un rôle très significatif sur l'efficacité de la réaction, mais l'homologie entre les spacers de chaque substrat n'est pas requise pour que la réaction ait lieu (Serre *et al.*, 1992). Par ailleurs, j'ai également étudié le comportement des mutants de site actif en présence de ces sites et montré que la recombinaison entre demi-sites nécessitait la formation de complexes nucléoprotéiques stables (Serre & Jayaram, 1992).

L'utilisation des demi-sites m'a également permis de mettre en évidence que la fixation de la protéine sur son site rend la liaison phosphodiester à couper susceptible à l'attaque par un nucléophile. En effet un mutant dont le nucléophile normalement impliqué dans la réaction (Tyr343) est absent (Tyr343Phe) génère des produits de réaction avec des demi-sites si le brin du bas a un spacer suffisamment flexible et une extrémité 5'-OH disponible. Dans ces conditions, c'est l'ADN lui-même qui joue le rôle de catalyseur de la réaction de coupure. La conséquence majeure

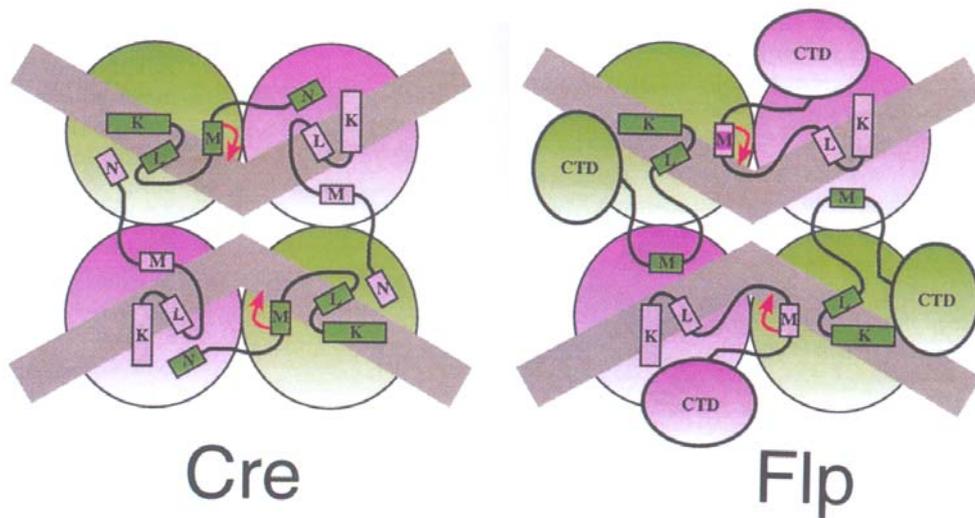


Figure 10 : Paradoxe *cis-trans* des recombinaisons à tyrosine. (d'après van Duyne, 2002)

Représentation schématique des tétramères de Cre et Flp dans les complexes synaptiques. Les sous-unités en vert sont celles qui sont actives pour la coupure du substrat. Pour Cre, l'hélice N est échangée en *trans* et la tyrosine catalytique portée par l'hélice M réalise la coupure en *cis*. Pour Flp, c'est l'hélice M qui est échangée en *trans* et qui permet donc à la tyrosine catalytique d'effectuer la coupure du substrat en *trans*. Le reste de la structure du domaine C-terminal (CTD) est replié en *cis* sur le monomère catalytique.

de cet événement est que dans ce cas les réactions de coupure et d'échange de brin sont concertées, alors qu'elles sont séquentielles en présence de la protéine sauvage (Serre *et al.*, 1993).

Le développement de ces substrats a ultérieurement permis de révéler une propriété unique de la recombinase Flp. En effet, leur utilisation a mis en évidence l'existence d'une complémentation fonctionnelle entre les mutants Arg191Ser, His305Leu ou Arg308Gln et Tyr343Phe (Chen *et al.*, 1992b). Cette complémentation fut interprétée comme résultant de l'assemblage du site actif de Flp à l'interface de deux protomères, et d'une attaque nucléophile de Tyr343 sur une liaison phosphodiester située en *trans* par rapport au site de fixation de son protomère d'appartenance. Ce modèle, bientôt connu sous l'appellation de paradoxe « *cis-trans* » (**Figure 10**), allait faire couler beaucoup d'encre dans les années qui ont suivi, opposant les tenants de la coupure en *cis* démontrée pour l'intégrase de lambda par l'équipe d'A. Landy (Nunes-Duby *et al.*, 1994), à ceux de la coupure en *trans* proposée également pour Cre par le laboratoire de P. Sadowski (Shaikh & Sadowski, 2000). La résolution de la structure de Cre complexée à l'ADN a révélé que Cre coupait son substrat en *cis* (Guo *et al.*, 1997; Gopaul *et al.*, 1998), renforçant ainsi l'opinion générale que la coupure en *trans* n'était qu'un artefact lié à l'utilisation des demi-sites et de mutants de Flp. L'équipe de Jayaram eu beau démontrer que le mécanisme en *trans* s'appliquait également à des substrats complets (Lee *et al.*, 1994) et à la protéine sauvage (Lee *et al.*, 1999), il fallu attendre la résolution de la structure de Flp par le groupe de P. Rice (Chen *et al.*, 2000) pour que la spécificité de Flp soit reconnue par la communauté, puis implicitement considérée comme étant une signature des recombinases à tyrosine eucaryotes. La comparaison des structures de Cre (*cis* active) et Flp (*trans* active) montre que l'hélice M qui porte la tyrosine catalytique n'adopte pas la même conformation pour ces deux protéines (**Figure 10**). Ainsi, chez Cre tous les résidus formant un site actif proviennent du même monomère, et la dernière hélice de la structure (hélice N) est distribuée en *trans*, assurant probablement une stabilisation du complexe synaptique. Chez Flp, c'est l'hélice M qui est distribuée en *trans*, le reste du domaine C-terminal se repliant sur son monomère d'appartenance. L'échange de l'hélice M, qui porte la tyrosine catalytique, conduit à la formation d'un site actif partagé entre deux monomères, et à une réaction de coupure opérant en *trans*. Il est intéressant de noter que le linker entre les hélices L et M est plus long pour Flp que pour Cre, ce qui pourrait permettre à l'hélice M de se positionner dans le site actif d'un autre monomère chez les recombinases de levure mais pas chez les recombinases bactériennes (Blakely & Sherratt, 1996).

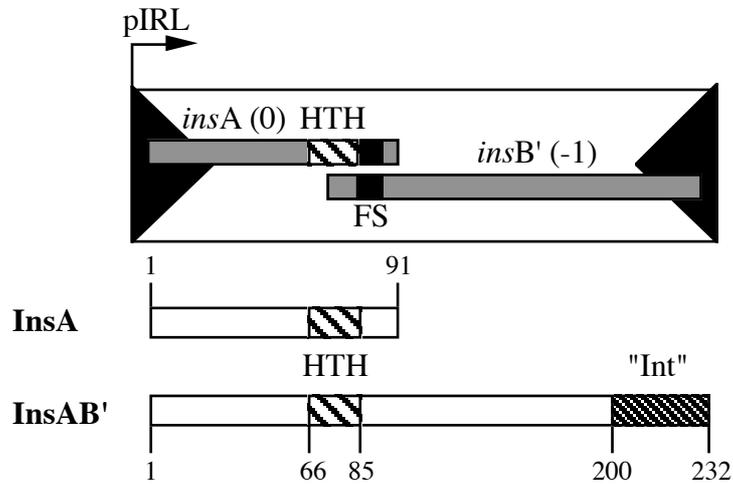


Figure 11 : Structure d'IS1 et de ses protéines.

Les extrémités inversées répétées sont représentées par des triangles. pIRL est le promoteur naturel. Les phases ouvertes de lecture *insA* et *insB'* sont représentées par des rectangles grisés. Les positions du motif potentiel hélice-tour-hélice (HTH) et du motif de décalage de phase (FS) sont indiquées respectivement en hachuré et en noir.

InsA (91 acides aminés) et InsAB' (232 acides aminés) présentent le motif HTH (positions 66 à 85). InsAB' présente de plus dans sa partie C-terminale un motif His-Arg-Tyr similaire à celui des recombinases à tyrosine ("Int").

2- Transposition bactérienne : Propriétés fonctionnelles de la transposase d'IS1

Laboratoire de Microbiologie et Génétique Moléculaire, UPR 9007, Toulouse, janvier 1992 - juin 1995

M.C. Serre, C. Turlan, M.L. Bortolin & M. Chandler. (1995). Mutagenesis of the IS1 transposase: Importance of a His-Arg-Tyr triad for activity. *J. Bacteriol.*, **177**, 5070-5077.

C. Rouquette, M.C. Serre & D. Lane (2004). Protective role for H-NS protein in IS1 transposition. *J. Bacteriol.*, **186**, 2091-2098.

En 1991 j'ai été recrutée comme CR1 au CNRS, et j'ai rejoint l'équipe de Mick Chandler à la fin de mon séjour post-doctoral. Ce groupe étudie le mode de transposition de différents éléments mobiles procaryotes. Mon projet était de développer l'étude du mécanisme de transposition de la séquence d'insertion bactérienne IS1. Dans le cadre de ce projet, j'ai encadré des stages d'étudiants en licence et maîtrise (100%) et un stage de DEA (100%).

IS1 a été l'une des premières séquences d'insertion bactériennes isolée et caractérisée. Elle a longtemps été considérée être spécifique aux entérobactéries, mais des variants apparentés ont été identifiés récemment dans d'autres génomes bactériens et dans certains génomes d'archaea (Mahillon *et al.*, 1999; Brugger *et al.*, 2002). IS1 est l'une des plus petites séquences d'insertion bactériennes isolées. Les 768 paires de bases de l'IS comprennent deux extrémités en répétition inverse d'environ 23 pb (IRL et IRR) et deux phases ouvertes de lecture partiellement chevauchantes, *insA* et *insB'*, dont l'intégrité est requise pour la transposition (Machida *et al.*, 1984; Jakowec *et al.*, 1988). Ces deux phases de lecture sont respectivement dans les phases de traduction 0 et -1 (**Figure 11**). Le transcript initié à partir du promoteur pIRL, partiellement localisé dans IRL (Wishart *et al.*, 1983), génère deux produits de traduction (Escoubas *et al.*, 1991). La protéine la plus abondante, InsA, se fixe spécifiquement aux IRs (Zerbib *et al.*, 1987; Zerbib *et al.*, 1990), et réprime la transcription à pIRL (Machida & Machida, 1989; Zerbib *et al.*, 1990). Elle inhibe également la transposition, probablement en entrant en compétition avec la transposase pour la fixation aux extrémités de l'élément (Escoubas *et al.*, 1991). La seconde protéine, InsAB', est la transposase de l'élément. Sa production résulte d'un mécanisme de décalage de phase traductionnel relativement inefficace entre les phases *insA* et *insB'*. Le site de décalage est localisé dans la partie 3' de la phase *insA* (Sekine & Ohtsubo, 1989; Escoubas *et al.*, 1991), en aval d'un motif potentiel hélice-tour-hélice probablement responsable de la spécificité de fixation d'InsA aux IRs. Le cadre en aval, *insB'*, fournirait les fonctions de coupure, de transfert de brins dans le site cible, et de religation. La protéine de fusion, InsAB' ou transposase, posséderait donc deux domaines, l'un N-terminal reconnaissant les extrémités de l'élément, l'autre C-terminal responsable de l'activité catalytique (**Figure 11**).

La fréquence naturelle de transposition d'IS1 est très faible (de l'ordre de 10^{-7} dans un test de mating out). La mise en phase des deux cadres de lecture, permettant une production constitutive de la transposase, ainsi que le remplacement du promoteur naturel d'IS1 par un promoteur exogène

inductible, augmentent le niveau de transposition de 10^3 à 10^4 fois *in vivo* (Galas & Chandler, 1981; Sekine & Ohtsubo, 1989; Escoubas *et al.*, 1991; Escoubas *et al.*, 1994). La transposition d'IS1 génère des insertions simples et/ou des cointégrats stables. L'existence de ces deux types de produits suggère fortement que deux voies de transposition puissent exister (Galas & Chandler, 1981). Cependant, malgré de nombreuses études génétiques, le mécanisme de transposition d'IS1 restait très peu connu au niveau moléculaire. Pour développer l'étude mécanistique de la transposition d'IS1, j'ai retenu deux types d'approches. D'une part, afin de pouvoir ultérieurement étudier le mécanisme de transposition *in vitro*, j'ai entrepris de purifier la transposase de l'élément. D'autre part, j'ai généré et étudié différents mutants de cette protéine, afin d'identifier les acides aminés importants pour la catalyse. Cette approche a nécessité le développement de nouveaux tests permettant de suivre *in vivo* les activités de la transposase.

a- Purification de la transposase d'IS1 et caractérisation d'une activité *in vitro*

La purification de la transposase d'IS1 s'est avérée être très difficile à réaliser. La protéine que j'ai finalement réussi à purifier ne présentait pas d'activité de transposition, l'ensemble de ce travail n'a donc pu être publié. Je décrirai brièvement les résultats que j'ai néanmoins obtenus.

La transposase d'IS1 est une protéine basique (pI théorique, 10.8) dont la chaîne peptidique a un poids moléculaire apparent de 29 kDa. Un problème majeur rencontré pour purifier InsAB' a été de l'obtenir en quantité détectable dans un extrait brut. De plus, en absence d'anticorps ou de test d'activité *in vitro*, le suivi de purification n'a pu être effectué qu'en analysant les différentes fractions en PAGE-SDS, et en observant la présence ou l'absence d'un signal peptidique à 29 kDa. Dans un premier temps, j'ai recherché un système d'expression permettant d'obtenir des quantités détectables de protéine dans les extraits bruts. L'utilisation du promoteur de l'opéron arabinose (Guzman *et al.*, 1995), m'a finalement permis d'obtenir un niveau d'expression de la transposase permettant sa détection par PAGE-SDS. La protéine recombinante obtenue a présenté un comportement agrégatif, et bien que de nombreuses conditions de solubilisation et de fractionnement aient été testées aucune n'a permis d'améliorer notablement sa solubilité. Après de nombreuses tentatives sur différents supports chromatographiques, j'ai finalement montré que la transposase était retenue sur un support héparine agarose, et qu'elle pouvait être spécifiquement éluée à haute force ionique. L'identité de la protéine a été confirmée par séquençage peptidique. La transposase purifiée a été soumise à une trypsinolyse, et les fragments tryptiques séparés en phase reverse. Un peptide de $PM_{app} = 1135$ a été séquencé et s'est avéré correspondre au fragment peptidique 123-130 de la transposase d'IS1. Quelques points sont cependant à noter: (i) une grande partie de la transposase ne se fixait pas sur le support héparine. Cette fraction non fixée pourrait correspondre à des micro-aggrégats insolubles ou à des formes partiellement dénaturées de la

protéine. (ii) Le rendement de purification en terme de protéines totales était mauvais. (iii) Les fractions obtenues après chromatographie étaient très diluées (30 $\mu\text{g/ml}$), et n'ont pu être concentrées par les méthodes classiquement utilisées.

A partir des faibles quantités de protéine purifiée dont je disposais, j'ai mis en évidence que la transposase présentait une activité endonucléasique transformant un substrat ADN circulaire surenroulé en circulaire relâché (coupure simple brin) et linéaire (coupure double brin). Cette réaction est strictement dépendante du Mg^{2+} . L'optimum de la réaction se situe à pH 8. A ce pH, les deux espèces (circulaire relâchée et linéaire) sont produites en quantité équivalente. A pH 8.5, la forme relâchée apparaît majoritairement, ce qui pourrait indiquer que la coupure du double brin d'ADN est effectuée de manière séquentielle. L'activité endonucléasique détectée ne présente pas de spécificité stricte de substrat. Quel que soit l'ADN super-enroulé fourni, on observe l'apparition de forme relâchée et linéaire. De plus, lorsque l'ADN linéaire purifié est soumis à une cartographie par restriction, l'analyse par électrophorèse révèle systématiquement un "smear", ce qui indique que les coupures générées par la transposase sont aléatoires. *IS1* ne présentant pas de spécificité stricte d'insertion, l'activité détectée pourrait être le reflet de la coupure non spécifique de la cible. Tous les essais de détection d'une activité de transposition, ou d'une fixation de la transposase sur ses cibles spécifiques (IRs de l'élément) ont été infructueux. De plus, aucun des nombreux essais ultérieurs visant à améliorer la purification pour obtenir une fraction soluble présentant une activité, au moins partielle, de transposition n'a abouti. J'ai donc finalement décidé d'arrêter cette approche biochimique, et de recentrer mon travail sur la recherche et l'étude de mutants de la transposase.

Le laboratoire de M. Chandler a réussi récemment à purifier la transposase d'*IS1* en fusion avec un tag histidine. Cette protéine fusion présente une activité de transposition *in vitro* (Ton-Hoang *et al.*, 2004). Il est très intéressant de noter que dans ce cas la transposase co-purifie avec une chaperone moléculaire, GroEL. Comme le signalent les auteurs : « *In spite of (or perhaps because of) the presence of GroEL, the InsAB' preparation was active in binding and catalysis in vitro* ». Ce résultat expliquerait *a posteriori* l'absence d'activité de la protéine que j'avais purifiée, probablement imputable à un problème de repliement de la transposase dans des conditions de surexpression.

b- Développement de tests d'activité in vivo

L'analyse fine de la réaction de transposition *in vivo* nécessitait d'obtenir des mutants de la transposase bloqués à différentes étapes du chemin catalytique. Quelle que soit la méthode choisie pour générer des mutants (aléatoire ou dirigée), il fallait pouvoir disposer de tests phénotypiques

soit pour les sélectionner soit pour analyser leurs propriétés. C'est dans ce but que j'ai développé ou adapté les tests d'activité suivants :

➤ Transposition intermoléculaire (test de papillation): J'ai adapté pour *IS1* un test décrit à l'origine pour *Tn5* (Krebs & Reznikoff, 1988) et *Tn10* (Huisman & Kleckner, 1987). Le substrat utilisé dans ce test est un plasmide dérivé de pSC101 portant un gène *lacZ* dépourvu de promoteur et encadré en 5' par une IRR et en 3' par une IRL. Lorsqu'il est introduit dans une souche Δlac , ce plasmide ne restaure pas un phénotype Lac^+ , ce qui indique que dans ce contexte la production de β -galactosidase est inexistante ou extrêmement faible. Des réarrangements liés à une activité de transposition qui placent *lacZ* sous contrôle d'un promoteur externe conduisent à l'activation du gène, détectée par l'apparition de papilles Lac^+ sur des colonies Lac^- (Serre *et al.*, 1995).

➤ Transposition intramoléculaire (test de délétion): Des événements de délétions adjacentes avaient été observés lorsque des plasmides surexprimant la transposase portent un transposon synthétique (Escoubas, 1992). Ces événements se produisent à haute fréquence lorsque la transposase d'*IS1* est exprimée sous contrôle du promoteur arabinose et peuvent être visualisés sur gel d'agarose (Serre *et al.*, 1995).

➤ Coupage des IRs (réponse SOS): Des mutants de la transposase d'*IS10* capables de catalyser l'excision du transposon mais défectueux pour une étape ultérieure de la réaction de transposition ont été isolés en utilisant un système basé sur le déclenchement de la réponse SOS (Haniford *et al.*, 1989). La transposase d'*IS1* est également capable d'induire la réponse SOS, si des extrémités de l'élément sont présentes sur le plasmide produisant la transposase. Cette induction traduit vraisemblablement l'activité de coupure plutôt que celle de transposition puisque des extrémités directement répétées (non productives en transposition) permettent d'observer une réponse (Lane *et al.*, 1994). J'ai donc construit des plasmides permettant de tester cette activité lorsque la transposase est sous contrôle du promoteur de l'opéron arabinose. Dans ce contexte, j'ai montré que l'induction de la réponse SOS pouvait être analysée qualitativement en mesurant l'activité β -galactosidase dans une souche indicatrice après induction de l'expression de la transposase (Serre *et al.*, 1995).

c- H-NS est impliquée dans la transposition d'IS1

En collaboration avec Dave Lane, je me suis intéressée au rôle qu'une protéine de l'hôte (H-NS) pourrait tenir dans la réaction de transposition *in vivo*. La participation de facteurs de l'hôte, et plus particulièrement de protéines histone-like, dans les processus de transposition est un phénomène largement décrit. En utilisant différents tests d'activité, j'ai mis en évidence que la transposition d'*IS1* est totalement inactivée dans une souche Δhns (phénotype papillation⁻). De plus, cette inhibition de la transposition semblait intervenir à une étape précoce puisqu'aucune activité de

coupure de brin n'est détectable *in vivo* en absence d'H-NS (phénotype SOS⁻). Attribuer un rôle à cette protéine dans la formation de la synapse entre la transposase, les extrémités du transposon et la cible de transposition semblait alors une hypothèse élégante. J'ai dans un premier temps analysé *in vitro* les propriétés de fixation de H-NS aux extrémités d'IS1 par retard sur gel, et déterminé que cette fixation était non spécifique. J'ai par ailleurs vérifié que la production de la transposase s'effectuait correctement dans une souche Δhns . Il s'est avéré que dans un tel environnement génique, la transposase d'IS1 n'était plus produite. Ces résultats suggéraient qu'en absence d'H-NS, l'expression d'une protéase d'*E. coli* serait déréprimée. Dave Lane a montré par la suite qu'une mutation dans *lon*, mais pas dans *ftsH* ou *clpP*, restaurait la synthèse de transposase dans les cellules Δhns . De plus une mutation dans *ssrA* restaure partiellement la synthèse d'InsAB', suggérant que l'absence d'H-NS affecte la traduction du messager codant la transposase et/ou conduise à la dégradation de la protéine produite (Rouquette *et al.*, 2004).

d- Obtention et caractérisation de mutants de la transposase d'IS1

Aucune similitude entre la transposase d'IS1 et d'autres transposases n'avait été mise en évidence, ce qui ne facilitait pas *a priori* une analyse fonctionnelle. De manière surprenante l'analyse de la séquence primaire de la transposase d'IS1 a révélé la présence d'une triade His-Arg-Tyr, caractéristique de la famille des recombinases à tyrosine (Argos *et al.*, 1986), dans la partie C-terminale de la protéine (Serre *et al.*, 1995). Un tel motif est également retrouvé dans la partie C-terminale d'EcoRII (Topal & Conrad, 1993). Pour ces deux enzymes les autres résidus catalytiques conservés chez les recombinases à tyrosine ne sont pas retrouvés.

Pour déterminer le rôle de la triade His-Arg-Tyr dans l'activité de la transposase d'IS1, j'ai généré des mutants à chacune de ces positions (His200Gln ; His200Leu ; Arg203Leu ; Tyr231Phe), et analysé leurs propriétés *in vivo* (Serre *et al.*, 1995). A l'exception de His200Gln, aucun des mutants n'est capable de promouvoir des évènements de transposition intermoléculaire ni de générer des produits de coupure à une fréquence induisant la réponse SOS. Enfin, aucun des mutants analysés ne génère de délétions adjacentes. D'une manière générale, ces résultats ont montré que des mutations introduites aux trois positions conservées chez les recombinases à tyrosine affectaient les activités de la transposase, probablement en diminuant voire en abolissant sa capacité à couper les IRs. Récemment, l'analyse *in vitro* des propriétés de ces mutants a montré qu'ils se fixaient aux IRs avec la même affinité que la protéine sauvage, mais qu'ils étaient très sévèrement affectés dans l'activité de transposition (Ton-Hoang *et al.*, 2004).

L'implication de la triade His-Arg-Tyr dans le mécanisme de transposition ne signifie pas que la transposase d'IS1 soit un membre de la famille des recombinases à tyrosine. D'une part, la signature complète du site actif des recombinases à tyrosine est constituée de six résidus strictement

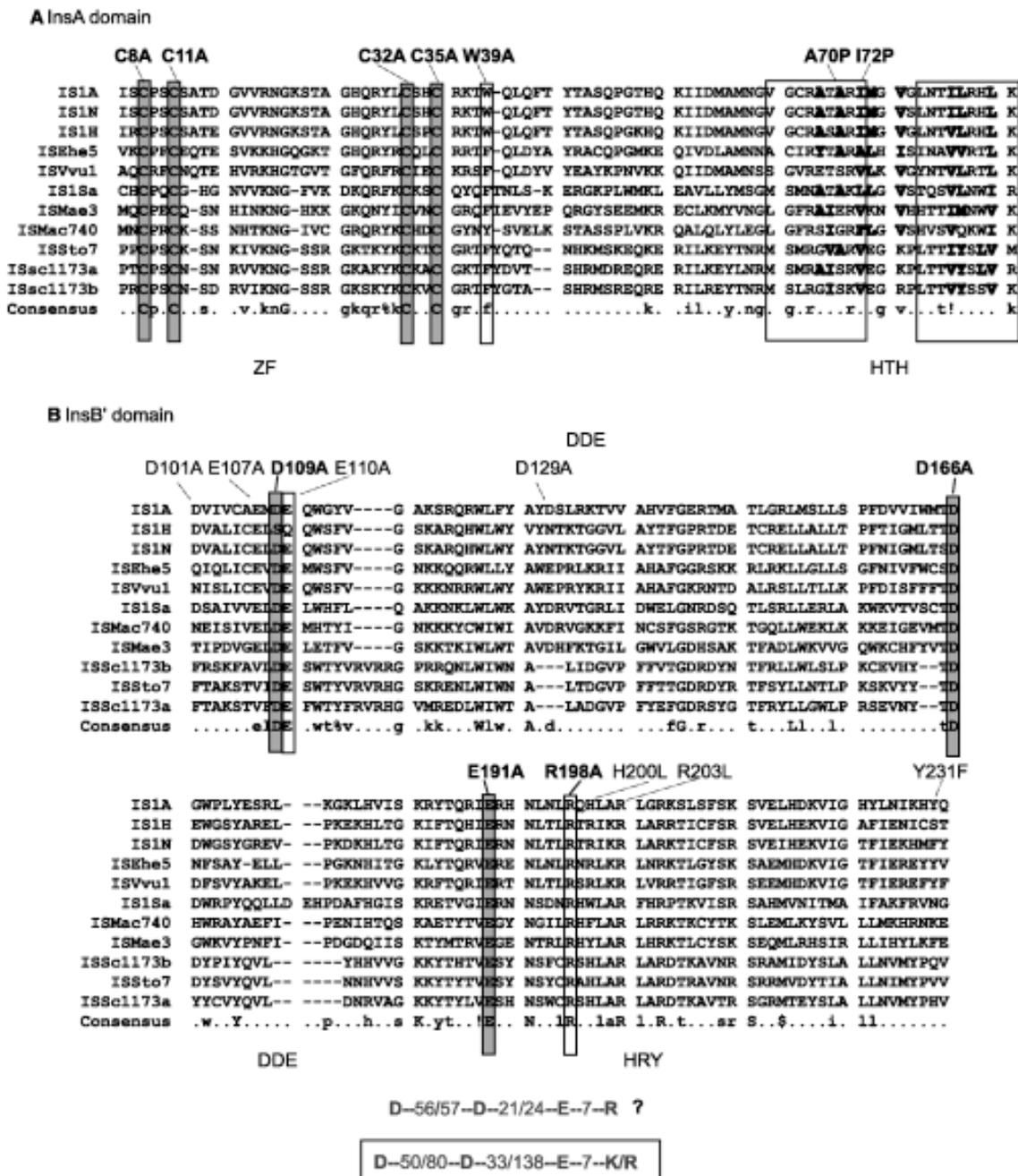


Figure 12 : Alignement des séquences des domaines InsA et InsB'. (d'après Ton-Hoang *et al.*, 2004)
A : Identification de motifs putatifs doigts de Zinc (ZF) et hélice-tour-hélice (HTH). **B** : Identification de motifs putatifs DDE et triade His-Arg-Tyr (HRY) des recombinases à tyrosine. Le consensus canonique du motif DDE des intégrases rétrovirales et des transposases est encadré.

conservés (Arg ... Lys ... His-(2)-Arg-(22/23)-His/Trp-(9/13)-Tyr). Seules trois positions sur six sont trouvées dans la transposase d'*IS1*. D'autre part, la polarité de la coupure et du transfert de brin sont différents dans les deux classes mécanistiques. Pour les recombinases à tyrosine la coupure de l'ADN génère une extrémité 5'OH libre alors que tous les systèmes de transposition analysés au niveau moléculaire à ce jour génèrent une extrémité 3'OH après coupure. Enfin, l'alignement de plusieurs séquences d'IS apparentées à *IS1* (**Figure 12**) a mis en évidence d'une part l'absence de conservation de la triade His-Arg-Tyr (même si l'Arg est conservée à 100%), et d'autre part l'existence d'un motif canonique « DDE » retrouvé chez les intégrases rétrovirales et de nombreuses transposases (Ton-Hoang *et al.*, 2004). Ce motif est directement impliqué dans la réaction de coupure de l'ADN (**Figure 13**). Bien que l'espacement entre les résidus du motif DDE soit non canonique pour les membres de la famille *IS1*, il semble fonctionnel chez la transposase d'*IS1*. En effet, des mutations ciblant les résidus le composant abolissent toute activité de coupure de l'ADN (Ton-Hoang *et al.*, 2004). L'ensemble de ces résultats fait que le rôle de la triade His-Arg-Tyr dans la catalyse de la transposition est actuellement très difficile à interpréter.

Afin d'obtenir des mutants défectueux en transposition mais capables de couper l'ADN, j'ai choisi d'utiliser une mutagenèse aléatoire. Le cadre de lecture *insAB'* a été soumis à une mutagenèse à l'acide nitreux, puis reclone dans le vecteur d'expression après amplification par PCR. Les mutants incapables de générer des papilles (*i.e.* déficients en transposition intermoléculaire) ont été sélectionnés, et leur capacité à induire la réponse SOS ou promouvoir des délétions adjacentes déterminée. Les mutations ont ensuite été identifiées par séquençage (**Tableau 3**). L'obtention et l'analyse de ces mutants ont été réalisées par S. Calgaro au cours d'un stage de DEA.

| Nom | Papillation | Induction du SOS ^a | Transposition Intramoléculaire | Protéine ^b | Mutation |
|-----|-------------|-------------------------------|--------------------------------|-----------------------|----------------------------|
| WT | +++ | +++ | +++ | 29 KDa | |
| A36 | - | + | - | non détectable | M73R, E191G |
| B31 | - | +++ | - | 29 KDa | G141D |
| B38 | - | +++ | - | 29 KDa | V92L, S94A |
| B39 | - | + | - | ND | mutation à la fin de l'IRR |
| D17 | - | ++ | - | 29 KDa | K133N, T147N |

Tableau 3: propriétés de quelques mutants obtenus par mutagenèse régionalisée.

a : l'induction de la réponse SOS a été observée sur milieu McConkey Lactose après 24 hr d'incubation à 37°C. ND, non déterminé. b: le poids moléculaire apparent a été déterminé par PAGE-SDS.

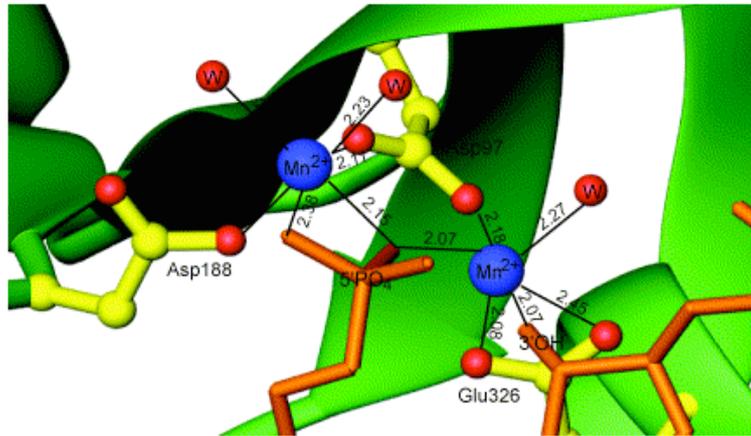
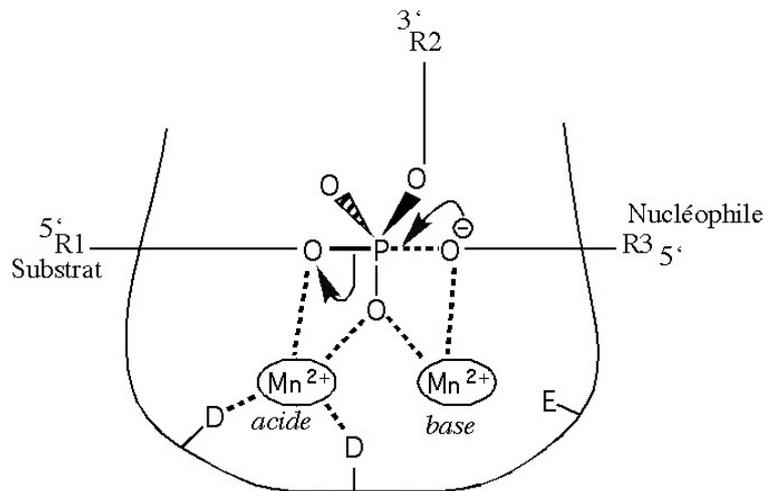
A**B**

Figure 13 : Rôle du motif DDE dans la coupure de l'ADN.

Historiquement, l'alignement des transposases de la famille IS3 avec les intégrases rétrovirales a mis en évidence l'existence d'un motif DD(35)E. Cette signature fut ensuite retrouvée dans de nombreuses autres transposases, dont celle de Tn5 (Tnp), et plus récemment dans celles de la famille d'IS1 (Ton-Hoang *et al.*, 2004).

A : Structure du site catalytique de la transposase de Tn5. (Steiniger-White *et al.*, 2004). Les résidus DDE coordonnent deux ions Mn^{2+} . La distance entre les deux ions est de 3.80 Å. Chaque Mn^{2+} a six ligands incluant les résidus catalytiques DDE, plusieurs molécules d'eau (représentées par les boules rouges), le 5'-PO4 du brin non transféré et le 3'-OH du brin transféré (l'ADN est en orange). La distance en angström entre chaque ion et ses ligands est indiquée.

B : Représentation de la triade catalytique DDE à l'état de transition. Le phosphate cible est dans un état pentacoordonné (pseudo-plan). Un des ions Mn^{2+} est proposé jouer le rôle d'acide de Lewis dans le mécanisme de substitution nucléophile. Il est coordonné par les deux résidus aspartate, et l'oxyanion du groupe partant. Le second Mn^{2+} est également coordonné par un oxygène du phosphate cible. Il va jouer le rôle de la base permettant de déprotoner le nucléophile entrant. Lors de l'excision de l'élément transposable, le nucléophile est une molécule d'eau. Lors de son intégration, l'extrémité 3'-OH de l'élément sera le nucléophile. Sur ce schéma, le substrat est la séquence 5'-R1-R2-3' et le produit de la réaction, 5'-R3-R2-3'.

L'intérêt principal de cette approche est qu'elle a permis d'obtenir des mutants dont les propriétés *in vivo* sont partiellement altérées (mutants « step »), certains mutants déficients en transposition étant toujours capables de couper les extrémités de l'IS (phénotype SOS⁺). Néanmoins l'efficacité de mutagenèse était trop élevée, ce qui se traduit par des mutations multiples dans le gène de la transposase (**Tableau 3**). Une analyse plus fine de ces mutants aurait nécessité de découpler les mutations par mutagenèse dirigée.

Parmi les mutations identifiées, une (Met73Arg) a été localisée dans le motif hélice-tour-hélice qui pourrait être impliqué dans la reconnaissance des extrémités de l'IS par la transposase. Cependant elle est associée à une autre mutation (Glu191Gly) qui pourrait également être responsable du phénotype observé. La position Glu191 a depuis été modifiée par mutagenèse dirigée (Ton-Hoang *et al.*, 2004) et s'est avérée correspondre au glutamate du motif catalytique DDE (**Figures 12 et 13**).

La majorité des mutations obtenues sont localisées dans le domaine insB', et plus particulièrement dans la région comprenant le motif DDE (**Figure 12**). La logique de ce ciblage n'est apparue que quelques années plus tard, avec l'identification de ce motif. Au moment où ce travail a été réalisé, ce ciblage préférentiel n'avait pas pu être interprété significativement. Les autres mutations que nous avons obtenu correspondaient à des décalages de phase traductionnelle, et généraient des protéines tronquées (détectables ou non dans les extraits bruts), malheureusement non informatives dans une recherche de positions impliquées dans la catalyse.

*Mes centres d'intérêts scientifiques étaient toujours orientés vers le déchiffrement des relations structure-fonction et l'analyse des mécanismes réactionnels d'enzymes modifiant l'ADN. L'échec de la purification de la transposase d'IS1 mettait abruptement un terme à toute possibilité d'évolution de mon projet de recherche dans cette direction. J'ai finalement pris la décision de quitter le laboratoire de Mick, et de revenir en région parisienne pour initier un projet de novo. Une tyrosine recombinase d'archae venait d'être identifiée, l'intégrase du virus SSV1, et pour la première fois une activité de recombinaison spécifique de site thermophile avait été décrite (Muskhelishvili *et al.*, 1993). Je fus rapidement convaincue de l'avantage que pouvait présenter un tel système. En effet, aucune structure de recombinase à tyrosine n'était résolue, et il me semblait judicieux de relever ce défi en utilisant une protéine thermophile. De plus, le laboratoire qui avait initié cette thématique ne poursuivait pas plus avant le projet, me permettant ainsi d'éviter dans un premier temps une compétition trop déséquilibrée. Jean-Renaud Garel a alors accepté de m'accueillir dans son équipe en me laissant toute latitude pour débiter ce projet. C'est donc avec beaucoup d'enthousiasme que j'ai réorienté mon travail de recherche vers le monde des extrêmophiles. Cette décision mûrement réfléchie allait cependant avoir bien plus de conséquences sur ma carrière que je n'aurai pu le supposer, puisqu'elle a entraîné une absence de publication de 6 ans. D'une part, le sujet que je quittais a été mis en stand-by pendant de longues années, et*



Figure 14 : Particules de SSV1. (d'après Martin *et al.*, 1984)

Les particules sont libérées dans le milieu environ 4 h après l'irradiation, sans provoquer la lyse de l'hôte. Certaines de ces particules se retrouvent attachées à ce qui semble être des fragments cellulaires (au centre de la photo).

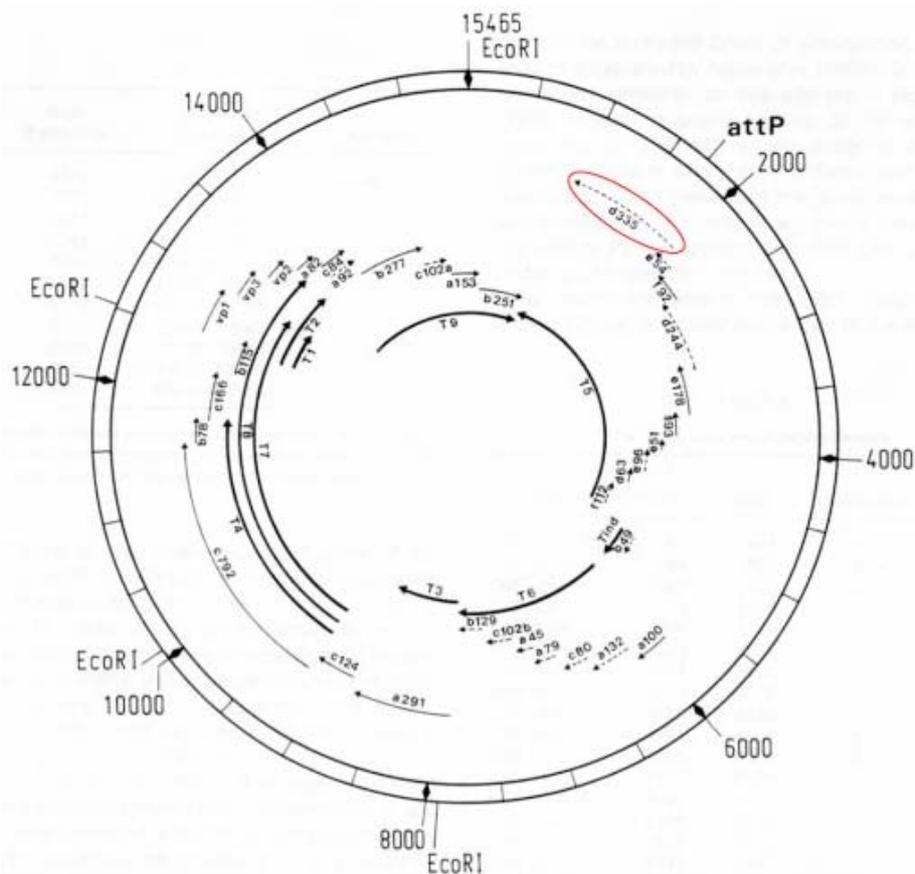


Figure 15 : Carte du génome de SSV1. (d'après Palm *et al.*, 1991).

La taille du génome est indiquée en paires de bases. La position du site *attP*, utilisée pour l'intégration dans le génome de *S. shibatae*, est indiquée. Les flèches en gras donnent les positions des transcrits, les flèches fines symbolisent les phases ouvertes de lecture. La phase correspondant à *Int^{SSV}* (d335) est entourée.

l'étude des derniers mutants obtenus abandonnée, et d'autre part comme je vais le présenter dans la section suivante, l'intégrase de SSV1 s'est révélée être plus « coriace » que je ne l'avais imaginé, et les travaux initialement publiés se sont avérés être non reproductibles...

3- Recombinaison spécifique de site chez les archaea (1996 - ...)

*Laboratoire d'Enzymologie et Biochimie Structurales, UPR9063, Gif-sur-Yvette, juillet 1995 - décembre 1998
Institut de Génétique et Microbiologie, UMR8621, Orsay, janvier 1999 - ...*

M.C. Serre, C. Letzelter, J.R. Garel & M. Duguet. (2002). Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. *J. Biol. Chem.*, **277**, 16758-16767.

M.C. Serre & M. Duguet. (2003). Enzymes that cleave and religate DNA at high temperature: the same story with different actors. In "Progress in Nucleic Acid Research and Molecular Biology" (Eds. K. Moldave). Vol 74, pp. 37-81.

C. Letzelter, M. Duguet & **M.C. Serre** (2004). Mutational analysis of the archaeal tyrosine recombinase Int^{SSV} suggests a mechanism of DNA cleavage in *trans*. *J. Biol. Chem.*, **279**, 28936-28944.

J'ai initié l'étude de la recombinaison spécifique de site chez les archaea dans le laboratoire de Jean-Renaud Garel au LEBS (Gif-sur-Yvette) et je l'ai ensuite développée dans le laboratoire de Michel Duguet à l'IGM (Orsay). Dans le cadre de ce projet, j'ai encadré un stage post-BTS (100%), deux DEA (80 et 100%), un stage de fin d'école d'ingénieur (100%), une thèse (100%), ainsi que des stages d'étudiants en M1 mention GCDE (100%).

Sulfolobus shibatae est une archae thermoacidophile isolée d'une source sulfureuse chaude au Japon (Yeats *et al.*, 1982). Ses conditions de croissance sont extrêmes (80°C, pH 2.5). L'intérêt majeur de cette souche est qu'elle possède un virus, SSV1. L'irradiation aux UV de cultures de *S. shibatae* induit la production et la libération de particules de type viral (Martin *et al.*, 1984). Ces particules ont une forme fusellée, une taille approximative de 100 x 60 nm et semblent posséder plusieurs fibres à une de leurs extrémités (**Figure 14**). Ce morphotype était totalement nouveau, et a conduit à la création d'une nouvelle famille virale, les *fuselloviridae*, ayant une spécificité d'hôte réduite aux crénarchaea. La taille des particules de SSV1 est comparable à celle des particules du phage T₄ qui possède cependant un génome de plus grande taille (168 kpb). Le génome de SSV1 a une taille de 15.5 Kpb (**Figure 15**), et est maintenu sous deux formes dans la cellule: une forme plasmidique, et une forme prophagique intégrée dans le génome de l'hôte (Yeats *et al.*, 1982). L'ADN contenu dans la particule virale est surenroulé positivement, alors que dans la cellule sa forme plasmidique présente une large distribution de topoisomères (Nadal *et al.*, 1986). La séquence complète de l'ADN viral a été déterminée (Palm *et al.*, 1991). Tous les transcripts ainsi que leurs promoteurs et terminateurs ont été identifiés (Reiter *et al.*, 1987b; Reiter *et al.*, 1988). Les gènes codant les protéines de capsid ont été localisés sur le génome de SSV1 (Reiter *et al.*, 1987a), ainsi que les gènes codant l'intégrase et l'excisionase potentielle du virus (Palm *et al.*, 1991). Ces

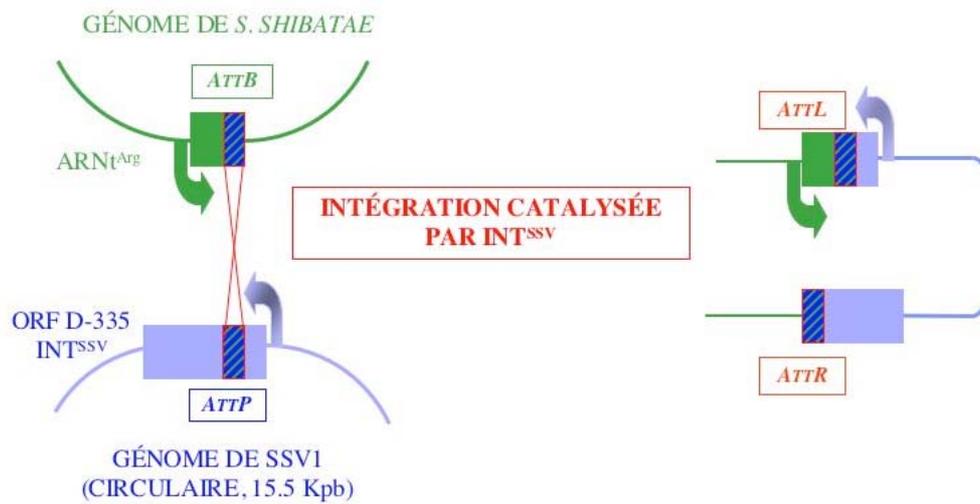


Figure 16 : Représentation de l'intégration du génome de SSV1 dans celui de son hôte.

La position des sites *att* où interviennent les événements de recombinaison spécifique est indiquée par une boîte hachurée. Cette séquence est commune aux quatre sites, et correspond aux 44 pb terminales du gène de l'ARN^{t^{Arg}}. L'intégration de SSV1 ne perturbe donc pas ce gène.

deux gènes sont co-transcrits (**Figure 15**). La fonction des autres ORFs identifiées reste inconnue. De manière surprenante, la qualité de virus a été accordée à SSV1 assez tardivement. Les tentatives de guérison de *S. shibatae* ont été infructueuses et la démonstration du caractère infectieux de SSV1 n'a été établie que 10 ans après sa découverte (Schleper *et al.*, 1992). Les souches *Sulfolobus solfataricus* P1 et P2 sont naturellement dépourvues de SSV1. Lorsqu'elles sont mises en présence de particules virales ou de cellules lysogènes de *S. shibatae*, l'infection a lieu et l'on retrouve une copie intégrée de SSV1 dans leur chromosome (**Figure 16**). Comme pour *S. shibatae*, l'intégration intervient spécifiquement dans un gène potentiel d'ARNt^{Arg} (Reiter *et al.*, 1989; Schleper *et al.*, 1992) et la production de virions peut être induite en irradiant aux UV les cellules infectées, sans pour autant entraîner la lyse de l'hôte (Martin *et al.*, 1984; Schleper *et al.*, 1992). L'intégration du génôme de SSV1 dans le chromosome de son hôte est spécifique de site et le déclenchement du cycle lytique répond à des facteurs induisant des lésions sur l'ADN, propriétés comparables à celles du bactériophage lambda d'*Escherichia coli*.

L'intégration de SSV1 dans le génôme de son hôte (**Figure 16**) fait intervenir au moins une protéine (l'intégrase du virus) et deux sites spécifiques, *attP* (site de SSV1) et *attB* (site chromosomique). Sur la base d'homologies de séquence peptidique, l'intégrase de SSV1 (Int^{SSV}) se rattache à la famille des recombinases à tyrosine (Palm *et al.*, 1991; Nunes-Düby *et al.*, 1998). Après intégration, le prophage est encadré par deux sites recombinants (*attL* et *attR*) qui, par analogie avec les bactériophages intégratifs, seraient utilisés lors de l'excision. L'étude de l'intégration de l'ADN viral dans sa cible a été initiée *in vitro* en utilisant une protéine de fusion GST-Intégrase exprimée chez *E. coli* (Muskhelishvili *et al.*, 1993). Cette étude restait cependant incomplète sur de très nombreux points, en particulier ceux concernant le mécanisme catalytique de l'intégration et de l'excision.

J'avais comme projet de résoudre la structure de l'intégrase de SSV1. Aucune structure de recombinase à tyrosine n'avait été résolue, et la polémique autour du paradoxe « *cis-trans* » (voir page 21) était à son apogée, rendant particulièrement excitante cette approche structurale. Je voulais en parallèle analyser le mécanisme de la recombinaison spécifique de site chez les archaea en générant des mutants de l'intégrase et en étudiant leurs propriétés en utilisant en particulier le test de recombinaison *in vitro* qui avait été récemment décrit (Muskhelishvili *et al.*, 1993). L'étude des relations structure-fonction de cette protéine aurait ensuite continué en exploitant les données de structure.

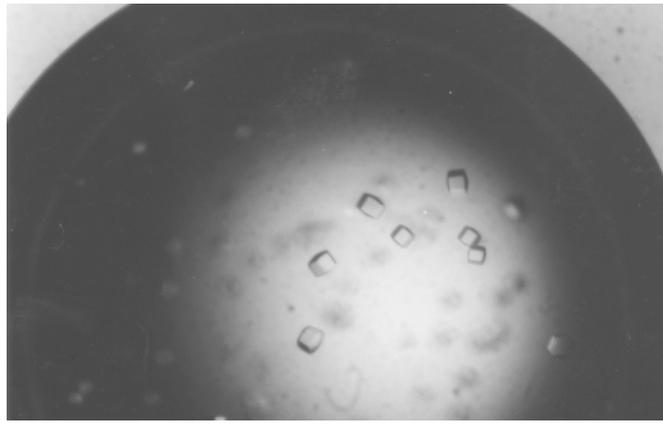


Figure 17 : Cristaux de Int^{SSV} obtenus en présence du substrat YTB.

Les cristaux ont été obtenus par la technique de la goutte en suspension, dans un tampon 0,1 M PIPES pH 6,8, 2 mM MgCl₂, 3,7% (w/v) sulfate de streptomycine. L'intégrase et son site étaient dans un rapport molaire de 1 :3. Dans l'exemple montré, les cristaux ont poussé en 5 jours pour atteindre une taille de 80 x 60 x 60 μm³. Dans les mêmes conditions, il a été possible d'obtenir des cristaux ayant une taille de 300 x 200 x 200 μm³. La séquence du site ADN utilisé dans ces tests est :

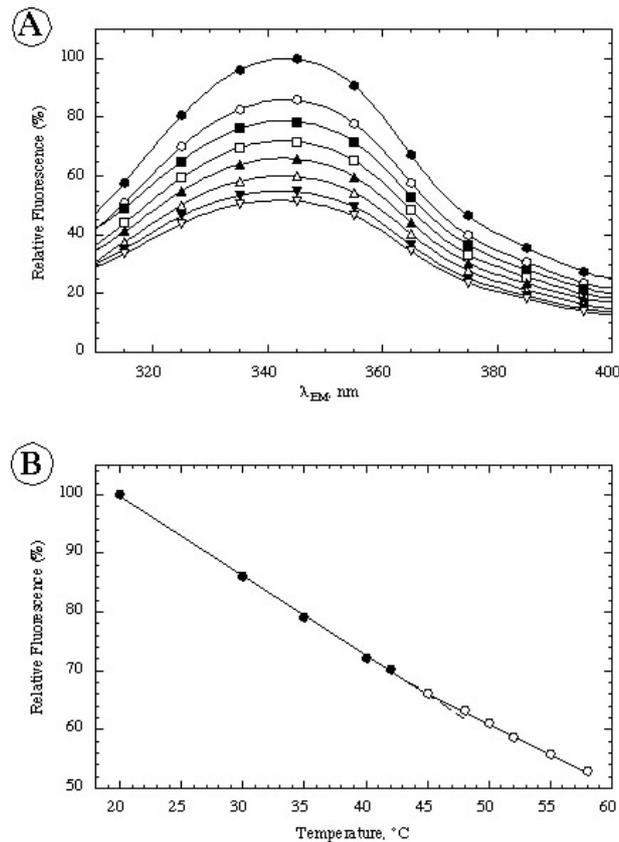
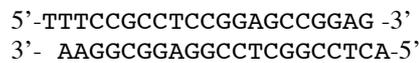


Figure 18 : Température-dépendance de la fluorescence d'Int^{SSV}. (Serre *et al.*, 2002)

A : Spectres d'émission de fluorescence d'Int^{SSV} à différentes températures après excitation à 295 nm. Les données ont été normalisées en prenant la valeur de λ_{EM} max à 20°C comme 100%. 20°C : ●; 30°C : ○; 35°C : ■; 40°C : □; 45°C : ▲; 50°C : △; 55°C : ▼; 58°C : triangles blancs inversés.

B : Les valeurs de λ_{EM} max sont représentées en fonction de la température, révélant une discontinuité dans la décroissance du signal de fluorescence.

a- Analyse des propriétés structurales de l'intégrase de SSV1

L'initiation d'un projet de biologie structurale nécessitait de disposer de grandes quantités de protéine purifiée, si possible sans addition d'étiquette. La protéine qui avait été produite dans le laboratoire de W. Zillig était obtenue en faible quantité sous forme de fusion avec la GST et ne pouvait donc pas être utilisée pour ce projet. J'ai cloné le gène d'Int^{SSV} sous contrôle du promoteur de l'opéron arabinose, en exploitant le vecteur d'expression que j'avais construit pour exprimer la transposase d'IS1 (Serre *et al.*, 1995). La purification de cette protéine recombinante s'est avérée très laborieuse, Int^{SSV} restant fixée à la plupart des supports chromatographiques testés. J'ai néanmoins réussi à développer une méthode de purification me permettant d'obtenir des quantités de protéine pure compatibles avec une approche cristallographique (Serre *et al.*, 2002). J'ai établi une collaboration avec Félix Rey, et nous avons recruté un étudiant de DEA (Stéphane Bressanelli) pour démarrer l'analyse structurale. Le début du stage de Stéphane a coïncidé avec la publication de la structure de l'intégrase de λ (Kwon *et al.*, 1997). Nous perdions la primeur d'une structure de recombinase à tyrosine, et avons donc décidé de rechercher directement des conditions de cristallisation d'Int^{SSV} sur son site. Nous avons réussi à identifier des conditions où Int^{SSV} co-cristallise avec un oligonucléotide comportant une séquence minimale du site de recombinaison (**Figure 17**). Malheureusement, les cristaux se sont révélés trop instables pour résister au rayonnement synchrotron. Tout aussi malheureusement, la structure de Cre en complexe avec l'ADN fût publiée à la même période (Guo *et al.*, 1997). La résolution de nouvelles structures de recombinases à tyrosine m'a finalement décidé à ne pas poursuivre cette approche cristallographique (Hickman *et al.*, 1997; Subramanya *et al.*, 1997).

L'étude cristallographique d'Int^{SSV} est redevenue importante après que nous ayons proposé que son site actif serait assemblé en *trans* comme celui de Flp (voir paragraphe *c*). La proposition qu'une recombinase procaryote puisse suivre un mode de fonctionnement eucaryote a une fois de plus ravivé les débats autour du paradoxe « *cis-trans* ». Il me semble essentiel de pouvoir valider le modèle enzymatique par une approche structurale. C'est dorénavant le groupe de D. Gopaul (Institut Pasteur) qui est en charge de ce projet.

Par ailleurs, j'ai utilisé différentes méthodologies pour étudier la structure en solution d'Int^{SSV} (ultracentrifugation analytique, chromatographie d'exclusion, DQEL, fluorescence). Les résultats obtenus à ce jour indiquent que l'état oligomérique de la protéine est fonction de la force ionique et qu'il s'accompagne d'un changement conformationnel. D'un état monomérique globulaire à haute force ionique, l'intégrase va s'associer en tétramère oblate à basse force ionique. De plus, l'analyse par fluorescence a révélé que la température induisait également des changements conformationnels d'Int^{SSV} (Serre *et al.*, 2002). Après excitation à 295 nm, les spectres d'émission de fluorescence enregistrés entre 20 et 58°C présentent un maximum d'émission à 343 nm (**Figure 18**). Lorsque la

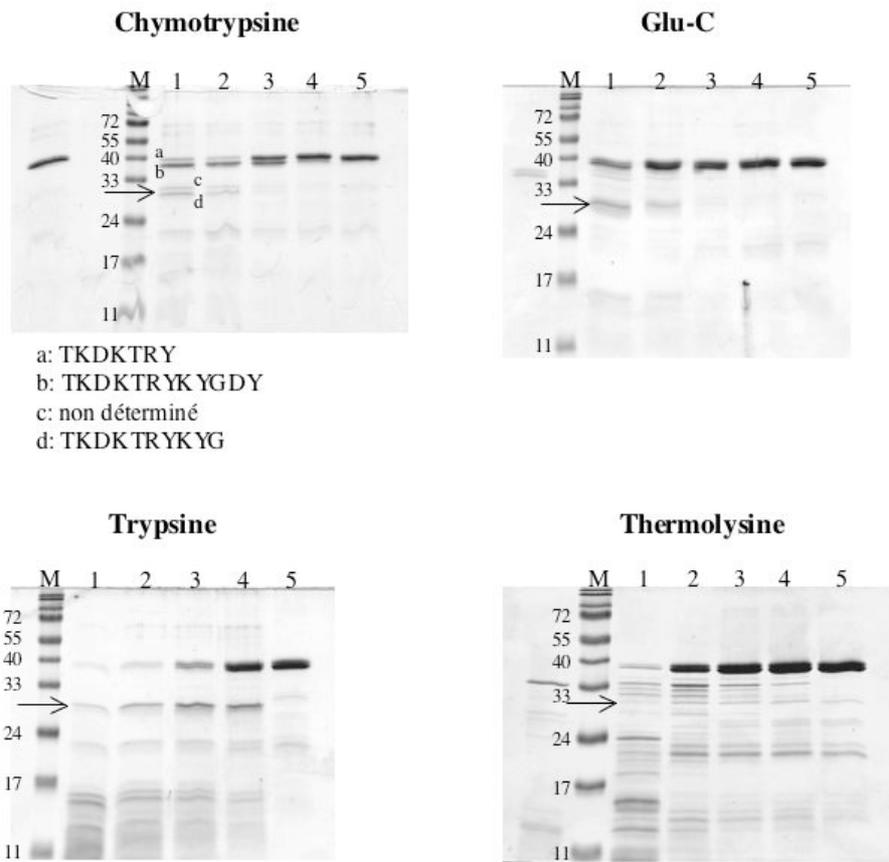


Figure 19 : Protéolyse ménagée d'Int^{SSV}.

M : marqueur de masse moléculaire. L'intégrase a été incubée 1 h à 37°C en présence de différentes protéases dans les rapports protéase :Int^{SSV} suivants :

Chymotrypsine : 1 : 1/1 ; 2 : 1/2 ; 3 : 1/4 ; 4 : 1/10 ; 5 : 1/20

Glu-C : 1 : 1/4 ; 2 : 1/8 ; 3 : 1/20 ; 4 : 1/40 ; 5 : Int seule

Trypsine : 1 : 1/80 ; 2 : 1/133 ; 3 : 1/200 ; 4 : 1/400 ; 5 : Int seule

Thermolysine : 1 : 2/1 ; 2 : 1/1 ; 3 : 1/2 ; 4 : 1/4 ; 5 : Int seule

La flèche indique l'apparition des fragments protéolytiques d'environ 30kDa. Certains produits de protéolyse générés par la chymotrypsine ont été séquencés (a à d). Les résultats du séquençage sont indiqués sous le gel correspondant.

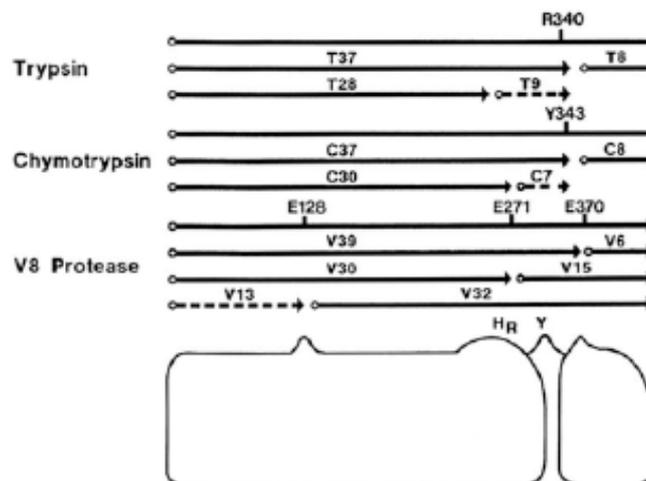


Figure 20 : Organisation structurale de F1p. (d'après (Evans *et al.*, 1990)

Positions de coupure de différentes protéases sur F1p, et représentation schématique de l'organisation domaniale de F1p. La sensibilité de l'analyse est très bonne, si l'on compare ce modèle à la structure d'un monomère (voir Figure 10).

température augmente, l'intensité de la fluorescence émise diminue jusqu'à atteindre, à 58°, 40% de l'intensité enregistrée à 20°C. Cette diminution du signal de fluorescence est totalement réversible, et traduit une diminution de la flexibilité des chaînes latérales des résidus aromatiques, conséquence probable d'une augmentation des interactions hydrophobes. Le renforcement des interactions hydrophobes en réponse à l'augmentation de température pourrait compenser l'augmentation de l'agitation thermique, et donc prévenir la dénaturation thermique de l'intégrase.

L'analyse domaniale par protéolyse ménagée suggère qu'Int^{SSV} serait organisée en deux domaines distincts, séparés par une charnière accessible aux protéases (**Figure 19**). La séquence N-terminale du plus gros peptide généré par la chymotrypsine a été déterminée, et s'avère correspondre à la séquence N-terminale de la protéine native. La charnière de sensibilité aux protéases se situe donc dans la région C-terminale de la protéine, sa localisation précise nécessitera une analyse par spectrométrie de masse. Ce résultat est différent de ce qui est observé avec l'intégrase de lambda, où une protéolyse ménagée permet de séparer un petit peptide N-terminal d'un grand peptide C-terminal portant les résidus catalytiques (Moitoso de Vargas *et al.*, 1988). Il est en revanche très similaire à ce qui est obtenu dans le cas de Flp (**Figure 20**), où une analyse similaire avait permis de détecter une charnière accessible aux protéases dans la région C-terminale de la protéine (Evans *et al.*, 1990). Il est à noter que la tyrosine catalytique est située dans cette région charnière, ce qui est en parfait accord avec la structure résolue ultérieurement. Bien que dans le cas d'Int^{SSV} je n'ai pas encore démontré que la charnière de sensibilité corresponde à la région équivalente, il est intéressant de rapprocher ce résultat des propriétés d'organisation du site catalytique que nous avons mises en évidence (voir paragraphe c).

b- L'intégrase de SSV1 coupe spécifiquement des substrats synthétiques

Il est rapidement apparu que les résultats publiés décrivant une activité de recombinaison *in vitro* associée à Int^{SSV} n'étaient pas reproductibles, même en utilisant la protéine recombinante et les substrats utilisés par (Muskhelishvili *et al.*, 1993)¹. Pour entreprendre l'étude des propriétés enzymatique d'Int^{SSV} j'ai dû développer de nouveaux tests d'activité (fixation aux substrats, réaction de coupure de l'ADN, relaxation). L'analyse de la réaction de coupure de petits substrats synthétiques m'a permis d'identifier la tyrosine catalytique (Tyr314), de mettre en évidence la formation d'un intermédiaire covalent protéine-ADN par le biais d'une liaison 3'-phospho-tyrosine, et de localiser précisément les sites de coupure (Serre *et al.*, 2002). L'analyse cinétique a de plus révélé que l'étape limitante de la réaction était la formation d'un complexe non covalent entre Int^{SSV} et le substrat. Au cours de sa thèse, Claire Letzelter a réalisé une mutagenèse du substrat sur les positions voisines des sites de coupure (positions +1, -1 et -2). Chacune de ces positions a été

¹ A ce jour, quatre laboratoires indépendants ont fait face au même problème...

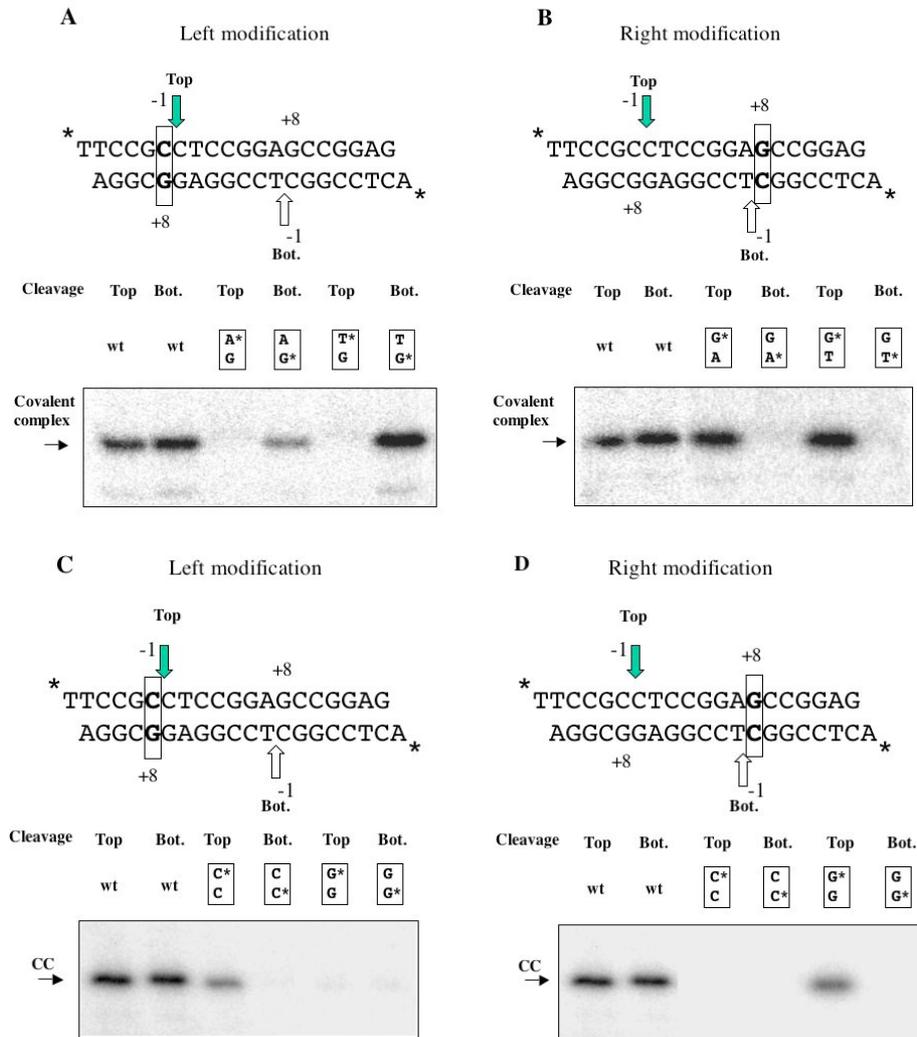


Figure 21 : Effet des mésappariements aux positions -1 sur la réaction de coupure.

La capacité d'Int^{SSV} à former un complexe covalent avec les substrats modifiés aux positions -1 gauche et droite a été analysée. L'effet des modifications a été déterminée pour chaque brin du substrat. Certains mésappariements affectent uniquement la coupure proximale (panneaux A et B), d'autres affectent les deux positions de coupure (G :G du panneau C, ou C :C du panneau D). Le mésappariement C :C en position -1 gauche (panneau C) affecte plus la position distale que la position proximale. Ce cas de figure est également rencontré pour les modifications à la position +1 droite.

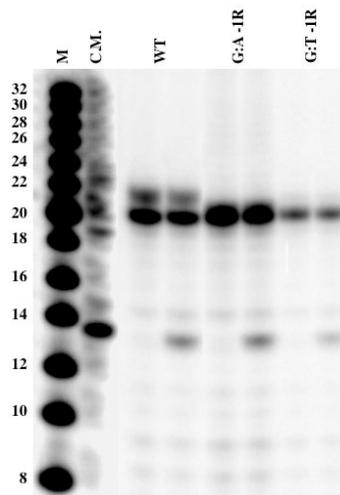


Figure 22 : Spécificité de la position de coupure de substrats mésappariés.

modifiée de manière à introduire toutes les paires de base possible, ainsi que différents types de mésappariements. La capacité d'Int^{SSV} à fixer et couper chacun de ces mutants a été évaluée. L'analyse de ces différents substrats a montré qu'Int^{SSV} présente de nombreuses similitudes avec d'autres recombinaisons à tyrosine dans ses contacts avec le substrat (Letzelter *et al.*)². Aucune des modifications introduites n'altère la capacité d'Int^{SSV} à se fixer aux substrats, suggérant que la nature des bases autour du site de coupure n'est pas un élément important pour la fixation d'Int^{SSV} sur la séquence core. En revanche, chacune de ces positions s'avère importante pour la réaction de coupure. Les deux sites de coupure localisés de part et d'autre du spacer ne sont pas indépendants. En effet, des modifications introduites sur un site de coupure se répercutent sur les deux réactions de coupure (**Figure 21**). Les effets des mésappariements aux positions -1 et -2 sont plus importants lorsque la séquence du brin supérieur est modifiée, suggérant que les contacts entre Int^{SSV} et le site core qui fournissent des informations de positionnement sont asymétriques. Nous avons de plus vérifié que les positions de coupure ne sont pas affectées par les mésappariements lorsque ceux-ci n'inhibent pas la réaction (**Figure 22**).

Les propriétés des mutants affectant la position +1 suggèrent que l'ouverture de la région spacer est une étape limitante de la réaction. La majorité des mésappariements introduits dans le spacer améliorent la fixation d'Int^{SSV} sur le site et certains de ces mésappariements augmentent fortement l'efficacité de la coupure du brin inférieur alors qu'ils diminuent la coupure du brin supérieur. Ce résultat suggère également que la première réaction de coupure interviendrait sur le brin inférieur.

Enfin, l'analyse des différents mutants a montré l'importance de la conservation des enchaînements de pyrimidine sur le site recombinant. L'existence de tels biais de séquence se retrouve également sur les sites d'autres recombinaisons (**Figure 23**), ce qui suggère qu'au-delà de la nature des bases composant le site, la structure locale de la double hélice aurait une influence sur la réaction de recombinaison.

Cette analyse systématique nous a permis d'identifier des mutants de site qui abolissent toute activité de coupure de l'enzyme. Ces mutants seront exploités pour étudier l'importance de la réaction d'intégration dans le cycle viral de SSV1 (voir chapitre III-2).

c- L'intégrase de SSV1 possède un site actif partagé

La famille des recombinaisons à tyrosine regroupe plus de 150 membres. Les données cristallographiques révèlent que la structure de différentes recombinaisons est très conservée, et qu'elle présente de nombreuses similitudes avec la structure des topoisomérases de type IB, au moins dans la région des sites actifs. En revanche les séquences primaires de ces protéines sont très

² Letzelter, C., Duguet M. & M.C. Serre (2005) *manuscript en préparation*.

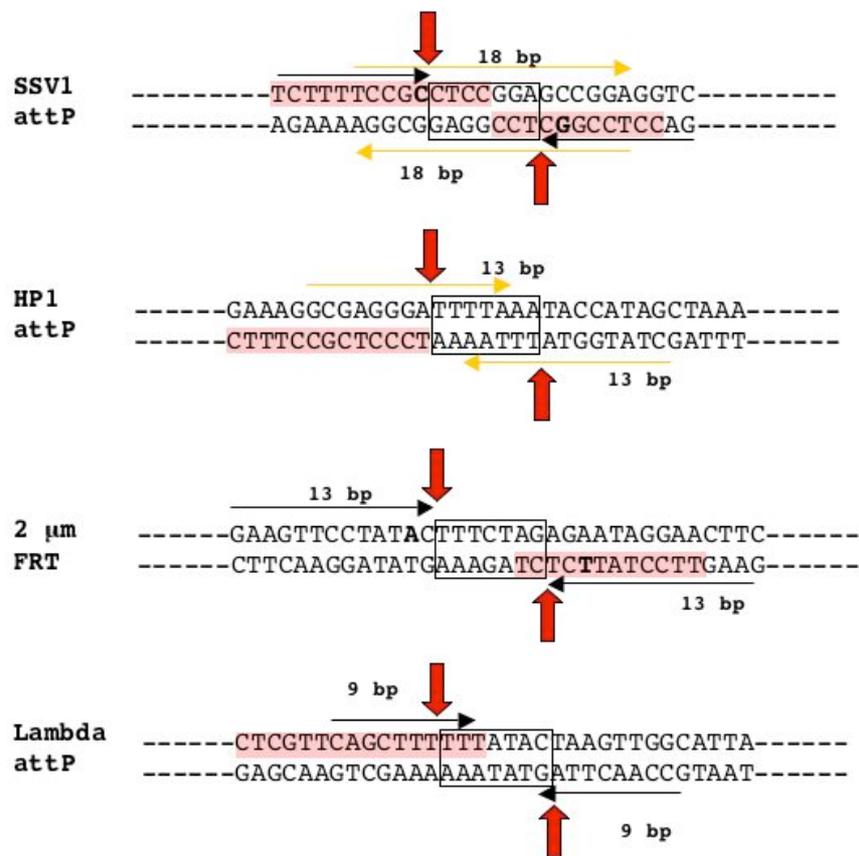


Figure 23 : Comparaison des séquences core de différentes recombinases à tyrosine.

Les positions de coupure sont indiquées par des flèches rouges, et les séquences spacer sont encadrées. Les séquences surlignées en rouge correspondent à des enchaînements de pyrimidine. Les flèches jaunes indiquent les régions protégées dans des expériences de footprint. Les flèches noires indiquent les sites de fixation des protéines (théorique dans le cas de SSV1). HP1 et SSV1 s'intègrent dans des gènes d'ARNt chez leurs hôtes respectifs, et dans les deux cas, la séquence comprise entre les deux positions de coupure correspond à la boucle anticodon.

divergentes. Les recombinases à tyrosine se caractérisent par la conservation stricte d'une tétrade Arg-His-Arg-Tyr (Argos *et al.*, 1986; Abremski & Hoess, 1992). Deux autres résidus conservés ont été ensuite identifiés, une Lys (notée Lys β) et une His/Trp (Redinbo *et al.*, 1998; Cao & Hayes, 1999). Parmi ces 6 résidus, cinq sont également conservés chez les topoisomérases IB. Ces six résidus sont impliqués dans le mécanisme catalytique partiellement commun entre les recombinases et les topoisomérases IB (**Tableau 4**).

| Tyr Rec consensus | R_I | <i>D/E</i> | K_{β} | H_{II} | R_{II} | <i>H</i> | H/W | Y |
|--------------------------|----------------------|-------------|---------------------------------------|-----------------------|-----------------------|-------------|-------------|-------------|
| Cre | R173 | E176 | K201 | H289 | R292 | V293 | W315 | Y324 |
| Lambda Int | R212 | D215 | K235 | H308 | R311 | S312 | H333 | Y342 |
| XerD | R148 | E151 | K172 | H244 | R247 | H248 | H270 | Y279 |
| Flp | R191 | D194 | K223 | H305 | R308 | H309 | W330 | Y343 |
| Topo IB consensus | R | <i>K</i> | K_{β} | K | R | <i>T</i> | H | Y |
| Vaccinia Topo IB | R130 | K133 | K167 | K220 | R223 | T224 | H265 | Y274 |
| Human Topo IB | R488 | N491 | K532 | K587 | R590 | T591 | H632 | Y723 |
| Int^{SSV} | R211 | E214 | R240 | K278 | R281 | K282 | R304 | Y314 |
| Mutations | L/Q | K/N | K/L | H/L | L | H/L | H/L | F/E |

Tableau 4 : Résidus catalytiques des recombinases à tyrosine et des topoisomérases IB.

Les positions strictement conservées chez les recombinases à tyrosine (Tyr Rec) et les topoisomérases IB (Topo IB) sont marquées en gras. Les résidus Asp/Glu et His en italiques sont conservés à 80% chez les recombinases à tyrosine. Les résidus correspondant chez les topoisomérases IB sont également indiqués. Les résidus d'Int^{SSV} présents à toutes ces positions sont indiqués ainsi que sur la dernière ligne les mutations qui ont été introduites. Les positions divergentes du consensus sont grisées.

L'alignement de la séquence d'Int^{SSV} avec le consensus des recombinases à tyrosine et celui des topoisomérases IB a révélé des différences au niveau des positions strictement conservées (**Tableau 4**). Ces modifications du site actif pouvaient soit être la signature d'une adaptation aux hautes températures, soit traduire une différence de mécanisme dans la réaction catalysée par Int^{SSV}. De fait, Int^{SSV} est le membre le plus éloigné de la famille des recombinases à tyrosine, et l'analyse de son mode de catalyse semblait essentiel pour déterminer si la recombinaison spécifique de site catalysée par les recombinases à tyrosine était un mécanisme conservé dans les trois domaines du vivant.

Nous avons réalisé une analyse mutationnelle d'Int^{SSV} (**Tableau 4**). L'analyse des propriétés enzymatiques des différents mutants obtenus nous a permis de montrer que l'intégrase virale était un membre à part entière de la famille des tyrosine recombinases (Letzelter *et al.*, 2004). Cette étude nous a permis de mettre en évidence le caractère hybride du mécanisme de recombinaison

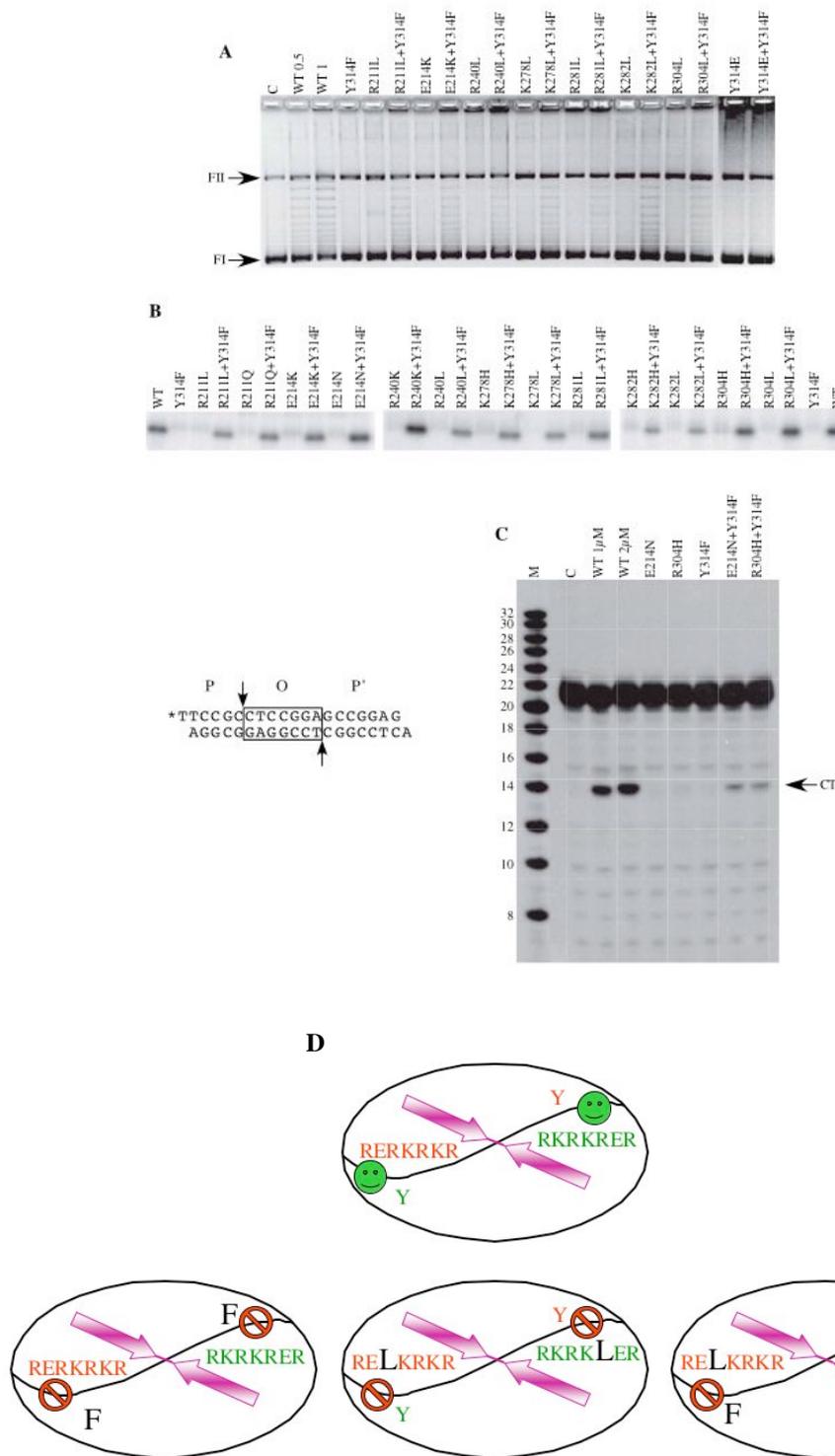


Figure 24 : Mise en évidence de l'existence d'un site actif partagé chez Int^{SSV}. (d'après Letzelter *et al.*, 2004)

A : Complémentation fonctionnelle entre mutants d'Int^{SSV} dans un test de relaxation. Seules les combinaisons présentées permettent de restaurer l'activité, à l'exception de la co-incubation des mutants Tyr314Phe et Tyr314Glu.

B : La formation d'un complexe covalent entre Int^{SSV} et un substrat synthétique est également restaurée par complémentation.

C : La position de la coupure obtenue dans les tests de complémentation est identique à celle observée pour la protéine sauvage.

D : Modélisation du concept de site actif partagé. L'hypothèse la plus simple, ne faisant intervenir que deux protomères est présentée, bien qu'on ne puisse exclure que l'assemblage d'un tétramère soit requis pour obtenir quatre sites fonctionnels. Dans les tests de complémentation, seul l'assemblage d'un hétérodimère conduit à la restauration d'un site actif fonctionnel.

spécifique de site chez les archaea. En effet, si l'organisation des sites de recombinaison est *a priori* similaire à celle de systèmes de bactériophages (**Figure 23**), celle du site actif de la recombinaison est de type eucaryote, puisqu'il est assemblé à l'interface de deux protomères apportant chacun une partie des résidus intervenant dans la catalyse (**Figure 24**) et (Letzelter *et al.*, 2004). S'il s'avère qu'une recombinaison à tyrosine d'archaea utilise un site actif partagé, alors le positionnement de la tyrosine catalytique pourrait avoir évolué différemment dans les domaines bactériens et eucaryotes. Dans ce dernier cas, l'addition d'un petit domaine à l'extrémité C-terminale de la protéine pourrait avoir contribué à stabiliser l'assemblage des protomères.

La détermination de la structure d'Int^{SSV} reste donc un objectif important à atteindre pour valider la modélisation effectuée sur la base des données enzymatiques. L'ensemble des données de structure en solution pourra également être exploité, mais de manière plus indirecte.

III- PROJETS DE RECHERCHE

Je souhaite désormais élargir ma thématique de recherche à l'étude du cycle viral de SSV1, comme système modèle permettant d'évaluer le rôle des virus intégratifs dans la plasticité des génomes d'archaea. L'étude des virus d'archaea, la manière dont ils sont capables d'infecter leurs hôtes et éventuellement de réaliser le transfert de certains gènes est d'intérêt pour mieux comprendre les mécanismes moléculaires qui ont permis le brassage de l'information génétique dans le phylum des archaea. En particulier, les virus infectant les hyperthermophiles sont des organismes hautement originaux, tant sur le plan fonctionnel que biochimique et moléculaire, qui n'ont pas d'homologues dans les autres domaines du vivant (Prangishvili *et al.*, 2001; Rachel *et al.*, 2002; Prangishvili, 2003). De plus, ces organismes peuvent être associés à des phénomènes tels que le transfert génétique entre cellules hôte, et ont probablement participé aux premières diversifications du vivant. Cependant les caractéristiques de ces virus (cycle viral, adaptation aux températures et pH extrêmes, relation avec l'hôte...) restent largement incomprises. Leur étude présente donc, en soi, un intérêt scientifique majeur et sera certainement à l'origine de la découverte de nouvelles fonctions biologiques.

Le cycle viral de SSV1 est actuellement très peu connu, mais comprend l'intégration du génome viral dans celui de l'hôte à un site spécifique et la production de particules virales, sans lyse cellulaire, lors d'irradiation UV des cultures infectées. Une originalité du virus SSV1 est que le

gène de l'intégrase est disrupté lors de l'étape d'intégration (**Figure 16**). Cette particularité, conservée chez tous les fusellovirus identifiés (Wiedenheft *et al.*, 2004), ouvre de nombreuses hypothèses quant au rôle de la recombinaison spécifique de site dans leur cycle répliatif. La compréhension du mode de répliation, du maintien et de la mobilisation des fusellovirus lors de signaux environnementaux est essentielle non seulement pour évaluer leur rôle dans la plasticité des génomes de leurs hôtes, mais également pour leur utilisation future comme outils génétiques chez les *Sulfolobales*.

J'ai comme projet d'étudier le rôle tenu par la recombinaison spécifique de site dans le cycle viral de SSV1 lors de l'infection de son hôte *S. solfataricus*. Je rechercherai également des partenaires potentiels de l'intégrase virale. Par ailleurs, je souhaite déterminer la fonction des autres protéines virales au cours du cycle de vie de SSV1. Ce projet devrait permettre de mieux comprendre les relations que les fusellovirus établissent avec leurs hôtes hyperthermophiles.

1. Mécanisme de la recombinaison spécifique de site chez les archaea

Collaboration avec le Dr D. Gopaul, Institut Pasteur Paris : résolution de la structure d'Int^{SSV} en complexe avec l'ADN.

Collaboration avec le Dr Q. She, University of Copenhagen, Denmark : enzymatic properties of integrases from fuselloviruses.

Les résultats que nous avons obtenus en étudiant des mutants du site actif ouvrent de nouvelles perspectives quant à l'étude du mécanisme de la recombinaison spécifique de site. Ainsi nous sommes actuellement en train de tester notre modèle de site actif partagé en construisant de nouveaux mutants d'Int^{SSV}. En effet, ce modèle pourrait être renforcé si certaines prédictions sont vérifiées. Par exemple, on peut s'attendre à ce qu'un double mutant Arg211Leu-Arg240Leu soit toujours complété par le mutant Tyr314Phe, alors qu'un double mutant Arg240Leu-Tyr314Phe ne le sera pas (**Figure 24**). Par ailleurs, j'espère pouvoir obtenir prochainement des informations sur la structure d'Int^{SSV}. J'ai développé des substrats symétriques, et testé les propriétés de fixation d'Int^{SSV} sur ces sites ainsi que sa capacité à les couper. Nous pouvons utiliser deux substrats, l'un de 27 pb l'autre de 33 pb, dans les essais de cristallisation. L'utilisation de tels substrats permettra d'améliorer la résolution de la structure de l'ADN dans les co-cristaux, et permet également d'augmenter les conditions d'obtention de structures cristallines.

Nous n'avons toujours pas réussi à reproduire la réaction de recombinaison *in vitro*. Si l'hypothèse du partenaire protéique manquant est correcte (voir paragraphe 3), Int^{SSV} devrait malgré tout présenter une activité de résolution des jonctions de Holliday. De telles structures sont en effet des intermédiaires réactionnels naturels des réactions d'intégration et d'excision. J'ai donc modélisé des oligonucléotides permettant de créer une jonction de Holliday par hybridation. Le marquage

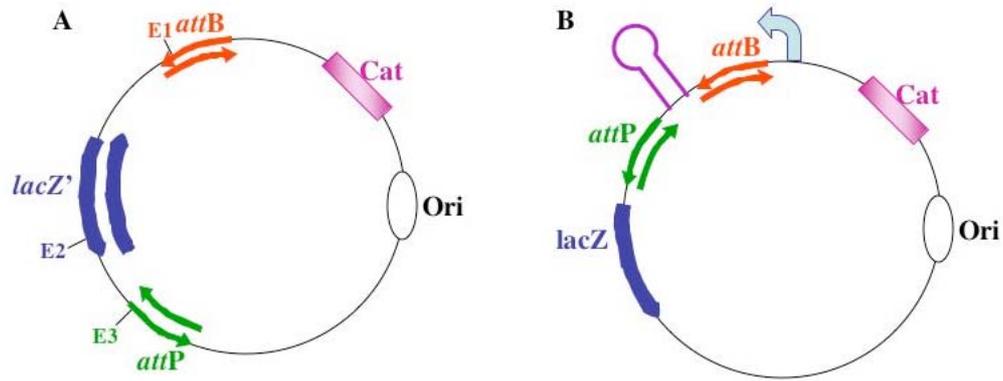


Figure 25 : Structure des plasmides rapporteurs de recombinaison.

Des plasmides portant différentes orientations relatives des sites *attB* et *attP* sont construits pour tester l'activité d'Int^{SSV}. Les deux types de construction portent l'origine de répllication de p15a et un gène de résistance au chloramphénicol.

A : Les sites *attB* et *attP* encadrent le gène *lacZ* sous contrôle de son propre promoteur. Le gène *lacZ* est présent dans l'une ou l'autre des orientations indiquées. Les différentes combinaisons possibles des sites de recombinaison et du marqueur phénotypique conduisent à huit plasmides différents. Selon l'orientation relative des sites *attP* et *attB*, un événement de recombinaison conduira soit à l'inversion de la région qu'ils délimitent, soit à l'excision de cette région. Dans ce cas, les bactéries transformées par les produits de recombinaison auront un phénotype Lac.

B : Les sites *attB* et *attP* encadrent un terminateur fort de transcription localisé entre un promoteur et le gène rapporteur (ici *lacZ*). Un événement de recombinaison conduira à l'excision du terminateur, et donc à l'expression du gène rapporteur.

alternatif de chacun des quatre brins permettra de suivre les produits de la réaction. L'analyse de la taille des oligonucléotides en fin de réaction nous indiquera si Int^{SSV} a résolu la structure, et plus précisément quel type d'échange de brin a été réalisé. Le retour aux sites parentaux, mimerait la réaction d'excision, alors que la production des sites recombinants traduirait l'intégration. De nombreuses recombinases à tyrosine sont actives sur de tels substrats, et c'est ce type d'activité qui a permis de révéler que les recombinases XerC et D étaient fonctionnelles *in vitro* (Arciszewska & Sherratt, 1995). Une autre hypothèse non exclusive concernant l'absence d'activité de recombinaison d'Int^{SSV} *in vitro*, est que les substrats ADN portant les sites *attB* et *attP* sont mal définis. J'ai développé des substrats permettant de suivre l'activité de recombinaison dans des tests *in vitro/in vivo* ou totalement *in vivo* en hôte hétérologue. L'avantage d'une étape *in vivo* est qu'elle devrait permettre de détecter des événements intervenus à une faible fréquence à condition que la sélection phénotypique soit suffisamment sensible. Les plasmides rapporteurs d'activité seront incubés *in vitro* en présence d'intégrase, et les produits de la réaction utilisés pour transformer *E. coli*. La première génération de ces plasmides rapporteurs utilise le gène *lacZ* comme marqueur phénotypique (**Figure 25 A**). Les essais préliminaires ont montré la limite de ce système qui consiste à chercher des clones Lac⁻ parmi des Lac⁺. La deuxième génération que nous allons construire permettra de rechercher des clones Lac⁺ parmi des Lac⁻, ou dans une version à plus haute sensibilité des clones résistants à un antibiotique (**Figure 25 B**). La version utilisant le gène *lacZ* sera également utilisée comme rapporteur d'activité dans le projet biotechnologique présenté au paragraphe 5.

Enfin, je viens de démarrer une collaboration avec Qunxin She de l'université de Copenhague. Son équipe étudie le virus SSV2, fusellovirus isolé de *Sulfolobus islandicus*. Nous souhaitons développer une étude conjointe des intégrases des deux virus, l'étude biochimique étant réalisée à Orsay et l'étude *in vivo* à Copenhague. Dans un premier temps nous allons construire des protéines chimères composées de la région N-terminale (comprenant le site *attP*) de l'une et de la région C-terminale de l'autre. Les propriétés des protéines chimères seront étudiées *in vitro* et *in vivo*. Notre hypothèse est que nous obtiendrons une spécificité d'intégration correspondant à l'origine virale du domaine N-terminal (SSV2 s'intègre spécifiquement dans un gène d'ARNt^{Gly}). S'il s'avère que nous pouvons altérer la spécificité d'intégration, nous construirons de nouvelles chimères permettant de cibler de nouveaux gènes d'ARNt, ne correspondant pas à une intégration naturelle. Notre objectif est de pouvoir ainsi obtenir plusieurs virus recombinants ayant tous une spécificité d'intégration différente. A plus long terme, nous espérons pouvoir cibler d'autres gènes d'intégration que les gènes d'ARNt, ce qui permettrait d'utiliser les fusellovirus comme outils de disruption génique chez les *Sulfolobales*.

2. Rôle de la recombinaison spécifique de site dans le cycle viral de SSV1

Un des principaux problèmes dans la compréhension du cycle viral de SSV1 est la coexistence dans l'hôte de deux formes virales, une forme plasmidique et une forme intégrée dans un gène ARNt^{Arg} de l'hôte. Si nous savons qu'après irradiation UV SSV1 est présent sous forme plasmidique et sous forme intégrée dans les cellules hôtes, il n'a pas été clairement démontré que la forme plasmidique soit présente en permanence dans chaque cellule. Il est en effet possible que la forme plasmidique n'apparaisse que lors de l'irradiation UV. Dans cette hypothèse, un évènement de recombinaison excisionnelle est indispensable pour régénérer (au moins) une copie plasmidique virale à partir de la forme intégrée. La forme intégrée serait donc *a priori* responsable de la réplication du virus et de la production de particules virales par la cellule hôte, lorsque celle-ci est exposée à une irradiation U.V. Cependant, cette hypothèse est remise en cause par la localisation du site viral *attP* à l'intérieur même de la séquence codante de l'intégrase (Palm *et al.*, 1991). Le gène codant l'intégrase est donc disrupté lors de l'intégration du virus, la séquence correspondant aux 75 acides aminés N-terminaux se trouvant dans le site *attL* et celle correspondant aux 260 acides aminés C-terminaux (contenant les résidus catalytiques) dans le site *attR*. Cette particularité rend problématique l'excision du génome viral de celui de son hôte en absence d'une source exogène d'intégrase fonctionnelle.

Je souhaite exploiter les résultats obtenus *in vitro* pour évaluer le rôle de l'intégration dans la stabilité du virus. Des mutations inactivant soit l'intégrase (Tyr314Phe) soit le site viral de recombinaison seront replacées dans le génome de SSV1. La construction des virus mutants est en cours en utilisant comme matériel un plasmide recombinant qui porte le génome complet de SSV1 cloné à un site unique de pBR322. Le génome viral entier peut ainsi être reconstitué après digestion et religation. Le devenir des virus mutants réintroduits dans *Sulfolobus solfataricus* sera évalué en collaboration avec l'équipe de David Pranghisvili à l'Institut Pasteur, et nous permettra de déterminer si l'intégration du génome viral dans celui de son hôte est essentielle au maintien et à l'amplification du virus.

En parallèle, une analyse biochimique réciproque consistera à déterminer si une forme tronquée de l'intégrase (correspondant au produit de la disruption intégrative) est fonctionnelle pour catalyser l'excision. La directionnalité des évènements d'intégration et d'excision peut en effet reposer sur l'existence de deux formes actives de l'intégrase (native pour l'intégration et tronquée pour l'excision). Nous avons déjà généré deux mutants du gène *Int* conduisant à une délétion des 75 premiers codons ou une délétion des 111 premiers codons. Ces mutants ont été clonés dans le vecteur d'expression utilisé pour produire la protéine sauvage mais le taux d'expression des protéines tronquées est extrêmement faible. Nous savons par ailleurs que l'intégrase sauvage ayant

une étiquette His₆ en N-terminal a les mêmes propriétés enzymatiques (et la même stabilité thermique) que la protéine non étiquetée. Nous allons donc cloner les gènes mutants dans un vecteur qui permet de placer une étiquette His₆ en N-terminal de la protéine d'intérêt, tout en gardant le système d'expression d'origine. Nous pourrions alors purifier les protéines tronquées par chromatographie d'affinité, ce qui nous permettra de nous affranchir des problèmes d'expression. Nous espérons par ce biais pouvoir déterminer si la partie N-terminale d'Int^{SSV} est essentielle à son activité, et si la particularité de l'intégration des fusellovirus dans le génome de leurs hôtes correspond à un mode de régulation original de la directionnalité des réactions d'intégration et d'excision.

3. Interactions hôte/virus lors de l'intégration virale

Bien que l'intégrase de SSV1 présente une activité biologique significative *in vitro* (Serre *et al.*, 2002; Letzelter *et al.*, 2004), la réaction complète de recombinaison spécifique de site médiée par Int^{SSV} n'a toujours pas été observée de manière reproductible *in vitro*. Ceci suggère que des interactions entre l'intégrase et d'autres facteurs cellulaires ou viraux soient nécessaires à cette activité. L'identification de ces facteurs permettrait donc de comprendre le mode d'intégration du virus dans le génome de son hôte. Le fait que la séquence du génome de *S. solfataricus* soit disponible permet d'envisager les approches suivantes.

L'identification des partenaires éventuels d'Int^{SSV} sera réalisée en utilisant des approches biochimiques classiques dans différentes conditions de croissance induisant ou non la production virale. Nous pourrions par exemple immobiliser l'intégrase portant une étiquette His₆ en N-terminal sur un support Ni-NTA. Cette protéine immobilisée servira à piéger d'éventuels partenaires en l'incubant avec des extraits cellulaires de *S. solfataricus* obtenus dans différentes conditions (avant ou après infection par SSV1, avant ou après irradiation aux UV, etc). Les extraits pourront de plus être supplémentés avec des sites *attP* pour favoriser des interactions de type synaptique, et les incubations réalisées à haute température pour s'affranchir autant que possible de modifications conformationnelles des protéines à basse température. Les complexes formés pourront alors être élués, et les partenaires protéiques piégés identifiés par séquençage peptidique. Une autre approche consistera à réaliser des co-immunoprécipitations puisque nous disposons d'anticorps polyclonaux dirigés contre l'intégrase. Nous pourrions alternativement réaliser des pontages chimiques *in vivo* dans différentes conditions physiologiques et piéger les complexes synaptiques formés à l'aide des anticorps dirigés contre l'intégrase. Les protéines copurifiées dans ces complexes seront isolées sur gel et identifiées par séquençage peptidique après que les pontages aient été réversés.

Les partenaires que nous aurons identifiés pourront alors être clonés et purifiés et leur implication dans la réaction d'intégration modélisée dans des tests d'activité *in vitro*.

4. Quelles fonctions pour quel cycle viral ?

Outre le rôle de l'intégration dans le cycle viral, je souhaite évaluer dans une approche plus globale le rôle des différentes protéines codées par le virus. En effet, sur les 34 protéines potentiellement produites par SSV1 seules 4 ont une fonction identifiée. Ainsi, les protéines de capsid VP1 et VP3 ont été purifiées et leurs gènes localisés sur le génome de SSV1 (Reiter *et al.*, 1987a). Dans la même étude, une troisième protéine a été isolée des particules virales et son gène identifié. Il a été proposé que VP2 serait une protéine se fixant à l'ADN (Reiter *et al.*, 1987a). L'intégrase de SSV1 avait été identifiée sur la base de similitudes de séquences avec les membres de la famille des recombinases à tyrosine (Palm *et al.*, 1991). Ses propriétés enzymatiques ont été déterminées au laboratoire (Serre *et al.*, 2002; Letzelter *et al.*, 2004). La recherche de protéines homologues aux autres protéines codées par SSV1 ne produit aucun résultat si l'on excepte les protéines potentielles des autres fusellovirus. Il faut cependant noter une exception, E. Koonin ayant proposé que la phase de lecture B251 de SSV1 coderait une protéine apparentée à DnaA (Koonin, 1992). Aucune étude de fonction n'est venue valider cette prédiction. Récemment, une approche de génomique structurale a conduit à la résolution de la structure de deux protéines de fonction inconnue, D63 et F93 (Kraft *et al.*, 2004a; Kraft *et al.*, 2004b). La structure de D63 suggère que cette protéine pourrait être un adaptateur dans des assemblages macromoléculaires. F93 est un homodimère en solution, et sa structure la place dans la superfamille des protéines fixant l'ADN par le biais d'un motif hélice-tour-hélice. Sa ressemblance structurale avec des protéines à motif « winged helix », pourrait traduire une fonction de facteur de transcription.

Par ailleurs, l'analyse comparative des différents génomes de fusellovirus montre que seules 18 ORFs (dont l'intégrase, DnaA, VP1 et VP3) sont communes à tous ces virus (Wiedenheft *et al.*, 2004). Les protéines correspondantes pourraient assurer les fonctions minimales essentielles au développement viral. A l'opposé, VP2 n'est retrouvée que dans SSV1, et D63 et F93 ne sont communes qu'à trois des quatre virus identifiés (Wiedenheft *et al.*, 2004). Ces protéines ne joueraient pas *a priori* un rôle essentiel commun à tous les fusellovirus.

Je souhaite définir la fonction des protéines conservées chez tous les fusellovirus. A l'exception de DnaA, aucune d'entre elle n'a d'homologue dans le vivant. Chacune de ces protéines putatives sera inactivée en insérant dans le gène correspondant une cassette poly-proline. Cette stratégie devrait nous permettre de nous affranchir d'effets secondaires transcriptionnels liés à l'organisation polycistronique du génome viral. Les différentes constructions seront réalisées en utilisant le

plasmide portant le génome complet de SSV1, ou certains de ces dérivés que nous avons construit au laboratoire. L'effet de l'inactivation de chaque ORF sera évalué chez *S. solfataricus* en prenant en compte différentes étapes du développement viral (stabilité du génome dans la cellule hôte, production de particules virales, infectivité...). Les phases de lecture correspondantes pourront également être clonées dans un vecteur d'expression, et une activité *in vitro* recherchée sur la base de leur fonction *in vivo*. J'espère ainsi pouvoir définir la fonction de ces protéines qui n'ont pour l'heure aucun homologue dans le vivant.

5. La recombinaison spécifique de site comme outil pour la thérapie génique : INTEGRA

(Projet Européen NEST ADVENTURE et ANR Innovation Biotechnologie)

Je participe à l'élaboration d'un projet européen visant à développer de nouveaux vecteurs ayant une spécificité d'intégration. Les applications de tels outils sont multiples, de la thérapie génique au développement de grands animaux transgéniques. Nous venons de passer la première étape de sélection, et soumettrons le projet final pour le 15 septembre. Ce projet a également été soumis à l'ANR dans le cadre de l'appel d'offre 'Innovation Biotechnologies'. Le consortium est composé des groupes suivants :

P1 : Equipe Jacques Mallet, Hopital Pitié Salpétrière

P2 : Pharmavectys

P3 : MC. Serre, Université Paris-Sud Orsay

P4 : Equipe Alain Fischer, Hopital Necker

P5 : Andrew Griffiths, Université Louis Pasteur, Strasbourg

P6 : Mark Isalan, EMBL

P7 : BruceWhitelaw, Roslin institute

Ma contribution à ce projet est décrite dans la partie « **milestone 2** », et j'assurerai la coordination du WP correspondant à l'étude des recombinaisons sélectionnées avant et après évolution dirigée. Il est prévu que je recrute un post-doctorant pour mener à bien ce projet.

INTEGRA

New tools for genetic engineering using targeted-integration vectors : application to agronomy, food safety and gene therapy

OUTLINE PROPOSAL

STREP

(Specific Targeted Research project)

Date of preparation : 14th of April 2005

Duration of the project : 3 years

Proposal Abstract : The overall objectives of the INTEGRA project is to develop new genetic engineering tools that will permit targeted integration of a transgene into the genome of the desired eukaryotic organism.

This project will allow the easy and rapid generation of transgenic, “knock-out” or “knock-in” animals, even in large animals, without the need of “embryonic stem cells”. This will have visible repercussions on fundamental and applied research, particularly in pharmaceuticals, agronomy and food sciences. At the same time, it will strengthen European research in the field of gene therapy with the development of safe viral vectors based on non-integrative or targeted-integration vectors adequate for clinical use.

We plan to meet the objectives by two alternative means. The first one will combine the use of a lentiviral episomal vector with a phage site-specific integration system to direct the integration of the transgene into a unique locus in the host genome.

The alternative method will use lentiviral vectors for which the integrase specificity will be modified. HIV integrases will be generated whose DNA binding domain will be replaced by a synthetic DNA binding domain able to recognize any given DNA sequence of 9 or 18 base pairs. Such specific integrase will be incorporated into a gene transfer vector system that will allow the genetic modification of somatic cells and zygotes.

These technologies will be used by the partners to generate transgenic livestock with modified traits (improved growth traits, disease resistance) and to generate “sensor” animals to monitor contamination of livestock by toxic agents. On the other hand, gene therapy preclinical studies will be performed using the secured vectors.

The agronomic applications presented in the project clearly lie outside the Thematic Priority areas of FP6, whereas the “gene therapy” and “food quality and safety” applications cut across the Thematic Priority 1 and 5 of FP6.

I- OVERALL OBJECTIVES

The overall objectives of the INTEGRA project is to develop new genetic engineering tools that will permit site-directed integration of a transgene into the genome of the desired eukaryotic organism.

This project will allow the easy and rapid generation of «knock-out» (KO) or «knock-in» (KI) animals, even in large animals, without the need of «embryonic stem cells» (ES cells). This will have visible repercussions on fundamental and applied research, particularly in pharmaceuticals,

agronomy and food sciences. At the same time, it will strengthen European research in the field of gene therapy with the development of safe viral vectors based on non-integrative or site-specific integration vectors, adequate for clinical use in both non-dividing and dividing cells.

II- BACKGROUND AND PERTINENCE FOR THE ADVENTURE INSTRUMENT

The generation of KO and KI mice using homologous recombination in ES cells has revolutionized the life sciences. It has allowed the study of many biological processes and the development of animal models for numerous human diseases. However, this breakthrough is limited by the high costs and the technical difficulties to obtain the mutant animals and the intractable methods currently available to generate KO or KI in other higher vertebrate animals (mainly because of the lack of ES cells for most species). The recent discovery of siRNAs could be an alternative means to attain such achievements, but they are limited by many technical obstacles linked to the reversible nature of the system. Indeed, the knock-down of a gene is dependant on the level of expression of an adequate siRNA (if available) that is itself dependant on the strategy of gene transfer. KO and KI sheep and pigs have been generated using somatic cell gene targeting and nuclear transfer (cloning), but the methodology is excessively difficult and in effect will have only limited uptake.

The development of a new technology that could allow the site-targeted integration of a transgene could be used for the generation of targeted *null* mutations in any cell type (*in vitro* or *in vivo*) and would thus open the way to non-reversible somatic and germinal mutagenesis. The range of applications for such a technology would be very wide, for example:

- directed mutagenesis of a gene in cultured cells (cell lines or primary cultures) for the study of biological processes (e.g. transduction pathways, gene expression, chromatin structure) or for the generation of cellular models of pathologies,
- directed mutagenesis in somatic cells *in vivo* for the same purposes,
- directed mutagenesis in germinal cells or zygotes to create KO or KI animals in virtually any species for biological (e.g. gene function studies), medical (e.g. generation of animal models of diseases), agronomical (e.g. generation of stocks with modified traits) or environmental applications (e.g. production of «sensor» animals that could reveal the presence of toxic agents in food or environment).
- moreover, it will allow the production of «classical» transgenic animals with a reproducible expression pattern due to the integration of the transgene into a selected locus, lowering the number of animals needed and consequently reducing the costs of stock rearing.

All these applications will have a large impact on Europe's scientific and technological capability. This clear progress well beyond the state-of-the-art will in the long-term benefit both the society and the economy. Moreover, the innovative engineering tools will have a high impact on an emerging field of medicine: gene therapy.

Only 16.2% of clinical protocols in gene therapy are performed in Europe, whereas 81.1% are performed in North America. These statistics clearly demonstrate that Europe is behind in this emerging field. The lack of safe vectors able to efficiently transfer a therapeutic gene to patient cells is probably the major obstacle to a massive penetration of clinical protocols by Europe. Recent findings demonstrating the potential harmfulness of the most used viral vectors (retrovirus) underline the crucial need for the development of secured vectors. 50.2% of the patients involved in gene therapy clinical trials are treated with retroviral vectors having the advantage of being able to stably integrate the genome of the infected cells. However, it is this property of retroviral vectors that is hazardous because of the potential random insertional mutagenesis they could generate.

The development of vectors unable to integrate the genome but nonetheless having a great transduction efficiency and no adverse effects would be ideal for gene therapy of quiescent cells

such as neurons or retinal cells. On the other hand, integrative vectors for which the integration is directed to a «neutral locus» would be ideal for gene transfer in dividing cells such as adult stem cells.

Our objective is to develop such a technology by generating new viral vectors and to demonstrate its potential in different applications that best take advantage of such new vectors. Although the range of applications is not limited to the ones we shall perform, the proof of principle for the potential of each new vector will be established thus opening the way to their general use for genetic engineering.

The development of such vectors and the demonstration of their potential requires a critical mass of expertise in diversified fields of biology such as virology, phage biology, neurobiology, somatic stem cell research, directed evolution and transgenesis in large animals. To the best of our knowledge there are no European researchers working on such vectors. Some US teams possess partial expertise to develop the same vectors and could merge their activities to attain the same goals. The technical and scientific breakthroughs that could be achieved with these new tools make it all the more important that Europe be the first to develop them.

In conclusion, the unlimited range of applications of the technology presented herein, combined with the highly innovative developments required, may open new avenues in genetic engineering with a high impact in science and technology. The technology will be exploited in the fields of agronomy, food sciences and gene therapy by our consortium. Thus, the proposed project clearly fits within the scopes of the ADVENTURE instrument:

- the research is very novel, very ambitious, with high impact through the development of new technologies which merge new areas of knowledge;
- although the technical risk is high, it will be reduced by the simultaneous development of alternative and/or complementary strategies to overcome the technical hurdles; and the project will be carried out in a way that realistic milestones with independent applications could be achieved;
- the agronomic applications presented in the project clearly lie outside the Thematic Priority areas of FP6, whereas the «gene therapy» and «food quality and safety» applications cut across the Thematic Priority areas of FP6 (respectively Thematic priority 1 and Thematic priority 5).

III- RESEARCH STRATEGY

We plan to meet the objectives by two alternative means. The first one will combine the use of an episomal vector with a phage site-specific integration system to direct the integration of the episomal genome into a unique locus (or a small number of loci) in the host genome. The choice of the episomal vector has been guided by the necessity to use a very efficient vector in terms of both infection and absence of any viral gene. Thus, the use of episomal vectors derived from commonly used adenoviral or herpetic vectors has been excluded. We will use a lentiviral episomal vector derived from HIV-1. To our knowledge, the approach of using episomal lentiviral vectors is unique and original.

The alternative method will use lentiviral vectors for which the integrase specificity will be modified. HIV integrase will be generated whose DNA binding domain will be replaced by a synthetic DNA binding domain able to recognize any given DNA sequence of 9 or 18 base pairs. A specific integrase able to recognize a desired DNA sequence of the genome will be generated and incorporated into a gene transfer vector system that will allow the genetic modification of somatic cells and zygotes.

III.1- Development of the new genetic engineering tools

Milestone 1: Episomal lentiviral vectors

HIV and other lentiviral vectors are able to transduce both dividing and non-dividing cells through their ability to cross the nuclear membrane. Their easy high-titer stock production, their high transduction efficiency for most eukaryotic cells and the possibility to pseudotype their envelope to modify their infection specificity make them powerful tools for gene transfer.

During the infectious cycle of lentiviruses, the viral genome (RNA) is retrotranscribed in a linear double-stranded DNA (proviral genome). At the nuclear transfer stage the proviral DNA can be circularised and generates two types of covalently closed circles containing either one or two LTRs (1LTR and 2LTR circles). The linear form can integrate into the host genome to form the integrated provirus. The integration process is governed mainly by the LTRs and the viral integrase (IN).

It has been shown that mutations in the integrase gene of wild-type HIV can prevent integration of the provirus into the host genome without preventing the formation of nuclear circles (1LTR and 2LTR).

The P1 team has generated mutations of the integrase gene affecting the catalytic activity of the enzyme. This mutated integrase gene has been introduced into the transcomplementation plasmid in place of the wild-type integrase gene. With this plasmid it is possible to produce lentiviral vectors (derived from HIV-1) containing a mutated IN (mIN vectors).

These vectors have been tested and compared to integrating lentiviral vectors. The mIN vectors are able to efficiently transduce dividing cell lines and primary cells. However, the expression of the transgene is abolished following the division of the transduced cells, as opposed to what happens with integrating vectors. This expected result suggests that the mIN vectors are able to transduce cells but that they remain in an episomal form (1LTR and 2LTR circles) that is lost by dilution through cell divisions. In non-dividing cells (primary cells and *in vivo*), the expression of the transgene is stable (longest time-point tested : 2 months).

P1 and P2 will demonstrate the lack of residual integrase activity in the newly developed vectors *in vitro* and *in vivo*. These episomal vectors will be used for designing site-targeted integration vectors by incorporating a modified integrase derived from other micro-organisms (phage integrases) or from the HIV-1. In addition these episomal vectors will be used as such for gene therapy pre-clinical studies by P1 and P2.

Milestone 2: Site-targeted integration vectors using hybrid phage/lentiviral vectors

It has been shown that site-specific phage integrases can be efficient in eukaryotic cells. For example, the integrase of the PhiC31 phage is able to mediate site-specific integration in mouse cells (using a plasmidic system) in two genomic loci that are partially homologous to its bacterial attB integration site.

Identified phages undergoing site-specific integration are numerous (more than 200). Each phage encodes an integrase that targets DNA specific sites, a phage site *attP* and a bacterial target site of integration *attB* (about 30 bp length). The P3 team will screen *in silico* all the known site-specific integrases of phages to select a series of the most suited integrases to attain the objectives. Then they will subclone the selected phage integrase genes and their corresponding *att* sites, and check for recombination in a bacterial system.

The P1 team will place an attP site into the previously described episomal lentiviral vectors (see milestone 1) and express the phage integrase into transduced cells. The integrase should then induce the integration of the episome into a unique site or a limited number of sites (referred as to Hm-attB) in human cells. The P3 team will test the specificity of integration in the Hm-attB sequence in a bacterial system to finally retain the best candidates. In order to design phage

integrases able to recognize any chosen locus in the genome, directed evolution technologies are envisaged.

Directed evolution is a technology allowing to modify a protein and confer it with new properties. Europe is behind the USA in this technology and should put forth efforts in developing it and its multiple potential applications. Directed evolution is a powerful tool for the engineering of site-specific recombinases. *In vivo* screening and selection of mutant libraries have been used to generate an improved FLP recombinase and to alter the site specificity of Cre recombinase. Completely *in vitro* selection systems, however, have many advantages over *in vivo* selection systems, not least in allowing the selection of much larger gene libraries.

The P5 team has developed a completely *in vitro* selection based on the novel concept of linking genotype to phenotype by compartmentalizing reactions in aqueous microdroplets (of ~2 μm diameter and 5 femtolitre volume) formed in water-in-oil emulsions. This *in vitro* compartmentalization (IVC) system is especially well suited to the directed evolution of enzymes as it allows direct and simultaneous selection for all enzymatic features (i.e. for substrate recognition, product formation, rate acceleration and multiple turnover), which enables the generation of extremely efficient enzymes. IVC will be used to improve the activity of the phage integrases isolated by the P3 team towards the Hm-*attB* homologues. Moreover, it will be used to modify the integrase specificity to target a sequence that is not homologous to the bacterial *attB* site. The resulting integrases will be tested for recombination by the P3 team in bacterial systems and by the P1 team in eukaryotic systems.

Milestone 3: Site-targeted integration vectors using ZFP-integrase fusions

The recent development of synthetic DNA binding domains or ZFPs (synthetic zinc-finger proteins) able to bind any desired sequence of 9 or 18 bp may allow the generation of site-specific integrases. The fusion of a DNA binding domain made of a synthetic ZFP to the integrase of HIV could lead to the creation of a site-specific HIV integrase.

Partner P6 will design and produce the synthetic ZFP able to target any selected 9 or 18 bp sequence. The ZFP cDNA will be fused to the lentivirus integrase cDNA and will be incorporated into the lentivirus production system by P1 and P2. These new vectors should have a specificity of integration directed toward the selected sequence.

IVC is also useful for the selection and directed evolution of DNA-binding proteins containing zinc-fingers (ZFPs). The binding of ZFPs designed by partner P6 will be optimized by selection using IVC. Furthermore, IVC will be used to directly optimize site-specific integrases generated by fusion of the ZFPs to the HIV

III.2- Application to genetic engineering of stocks and to gene therapy

Milestone 4: Use of episomal lentiviral vectors for gene therapy of the retina and the brain

The vectors developed in milestone 1 would represent a novel episomal vector devoid of toxic effects (because no viral gene remains in the lentiviral vectors), easy to produce and efficient for many cell types. For gene therapy its use would be restricted to non-dividing cells. P1 and P2 will use episomal lentiviral vectors (mIN vectors) for preclinical studies in Parkinson's disease and retinal degeneration animal models.

HIV-1 derived episomal vectors allowing the expression of neurotrophic genes (GDNF, neurturin) will be used for preventing degeneration of the dopaminergic neurons or the photoreceptor cells in animal models of the Parkinson disease (6-OHDA treated rats and MPTP treated monkeys) and retinitis pigmentosa models (RCS rats and Rd10 mice).

Degenerative diseases represent a social and medical burden in Europe, with no effective treatment in most cases. Gene therapy may open the way to new treatments since safety is guaranteed to the patients, for instance with the use of non-integrative vectors.

Milestone 5: Use of site-targeted integrative lentiviral vector in human hematopoietic stem cells

The vectors that will be developed in milestones 2 and 3 will provide tools to safely modify stem cells in view of gene therapy. Indeed, the major obstacle remaining for a safe and successful gene/cell therapy using somatic stem cells is the potential insertional mutagenesis that can be caused by integrative vectors. If the integration locus is not a coding sequence and it allows expression of the transgene, then a new safe site-specific integration vector will have been created that could be used for gene therapy. These new vectors will allow the integration of the vector into a site that does not compromise the cell survival and behavior.

These new site-directed integration vectors will be used by partner P4 for gene transfer into human hematopoietic stem cells. The validity of the strategy will have high impact for the treatment of rare genetic diseases, and many other acquired diseases where the technology could be transferred.

Milestone 6: Use of site-targeted integrative lentiviral vector for production of stocks with improved traits

Genetic engineering offers the possibility of incorporating in a livestock breeding population genes or specific alleles that are not normally present in that population or species. Thus, improved or new traits can be introduced into livestock species. Although transgenesis is relatively easy in rodents, it presents many technical hurdles that impose huge financial constraints for larger animals such as pigs or cattle. This includes low integration frequency and poor embryo survival. In addition, random transgene integration is often associated with low penetrance of transgene expression including transgene silencing. Finally, «classical» gene targeting methodology in livestock is at best extremely difficult. These inefficiencies ensure high-cost and low-uptake of the technology. The generation of a robust approach to integrate the transgene in a locus permissive for expression of the transgene will considerably reduce the costs of the generation of transgenic livestock, with a high repercussion on European economy.

The proof-of-principle of the technology will be first tested by partner 7 with the generation of pigs over-expressing a gene involved in growth, e.g. myostatin, SOCS2. The choice of species reflects our current efforts to develop lentiviral transgenic technology and this target allows assessment of phenotype in the founder generation allowing rapid evaluation of the technology. Another application will be the generation of transgenic livestock with enhanced disease resistance through the over-expression of specific allelic versions of genes central to the NFkB intracellular signalling response.

The current effort, with a large European involvement, to identify quantitative trait loci (QTL) in livestock will provide in the near future further candidate genes that could be exploited for the generation of livestock with improved agriculturally important traits, such as meat quality, reproduction and immune status

Milestone 7: Generation of «sensor» animals for the detection of toxic agents in the food

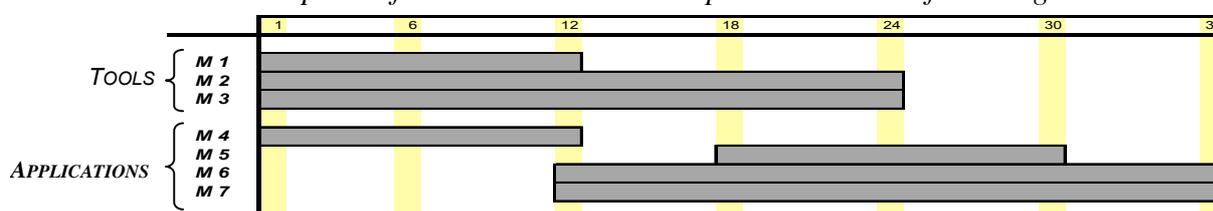
Food safety is an important issue in Europe. Moreover, the recent events concerning quality of stocks have modified consumer trust in the agronomic industry with high repercussion on economy. Preventing the contamination of stocks with toxic agents will further need the development of several strategies, comprising means for early detection of the contamination.

A way to detect the presence of a toxic agent in the stocks will be to have «sensor» animals grown with the other animals that would reveal the presence of the toxic agent by a simple mean. Partner 7 will generate pigs (or minipigs) transgenic for a reporter gene under the control of a toxic-stress inducible gene promoter. For example a promoter activated by elevated heavy metal concentration or one that responds to infection. These "sensor or sentinel animals" will express the reporter gene in the presence of the appropriate stimuli. Partner 7 has developed a novel reporter that is excreted into the urine and detected by a simple Elisa assay, thus enabling non-invasive and quazi-realtime monitoring of the degree of stress the animal is exposed to.

The ongoing development of pharmacogenetics and transcriptomics will provide many other candidate genes that can be induced by toxic agents that could further be used for generating «sensor» animals for these agents.

III.3- Gantt chart

The overall description of the research work is represented on the following Gantt chart :



IV- CONCLUSION: DOMAINS OF COMPETENCE OF THE PARTICIPATING TEAMS AND RESOURCES

The development of genetic engineering tools allowing the targeted integration of a transgene in a host genome will benefit all the areas of life sciences, with repercussions on health, biotechnologies, pharmaceuticals, agronomy and many other fields. The innovative technologies presented here will be applied by the partners to food safety, stocks quality and gene therapy. Partners 1, 2, 3, 5 and 6 will participate in the design of the engineering tools, whereas partners 1, 2, 4 and 7 will use these tools in the aforementioned applications.

| Participants | Domains of Competence | Duties and involvement | Costs (€) | Requested grant (€) |
|------------------|--|--|------------------|---------------------|
| P1 | vectorology, gene therapy of neurological diseases | Management , Milestones: 1, 2, 3 & 4 | 814 320 | 417 160 |
| P2 (SME) | lentivirus production, gene therapy of retina | Milestones: 1, 2, 3 & 4 | 601 680 | 305 840 |
| P3 | microbiology, phage biology, site-specific phage integrase | Milestones: 2 | 240 200 | 240 200 |
| P4 | human hematopoietic stem cell gene therapy | Milestones: 5 | 186 000 | 96 000 |
| P5 | protein directed evolution | Milestones: 2 & 3 | 723 976 | 371 988 |
| P6 | design and production of synthetic DNA binding domains | Milestones: 3 | 455 000 | 232 500 |
| P7 | large animal transgenesis | Ethics, Milestones: 6 & 7 | 1 309 834 | 664 917 |
| TOTAL (€) | | | 4 331 010 | 2 328 605 |

EN GUISE DE CONCLUSION ...

Tout au long de ma carrière j'ai favorisé la mobilité thématique et/ou géographique plutôt que le choix parfois rassurant de la stabilité sur le long terme. Seules ces dernières années auront été entièrement consacrées à l'étude de l'intégrase de SSV1. Si je souhaite continuer, au moins sur le court terme, l'étude de cette protéine je ne peux me satisfaire de ne suivre que ce chemin. Les projets que je souhaite développer devraient me permettre de relever de nouveaux défis.

Bien qu'il soit parfois frustrant de quitter un sujet au moment où l'histoire décolle, je ne regrette aucun de mes choix passés. En effet, chaque sujet m'a permis d'acquérir de nouvelles compétences, qui aujourd'hui me conduisent à envisager différents types d'approches pour répondre à une même question. Par ailleurs, le passage dans plusieurs équipes m'a obligé à savoir m'adapter à différentes conditions de travail, et à interagir avec des personnes ayant des parcours différents du mien, et par la même des modes de pensée différents. L'aventure scientifique s'est donc doublée d'une aventure humaine particulièrement enrichissante. Mon souhait le plus cher est que cela dure le plus longtemps possible...

IV- BIBLIOGRAPHIE

- Abremski, K. E. & Hoess, R. H. (1992). Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng.* **5**, 87-91.
- Araki, H., Nakanishi, N., Evans, B. R., Matsuzaki, H., Jayaram, M. & Oshima, Y. (1992). Site-specific recombinase, R, encoded by yeast plasmid pSR1. *J Mol Biol.* **225**, 25-37.
- Arciszewska, L. K. & Sherratt, D. J. (1995). Xer site-specific recombination in vitro. *EMBO J.* **14**, 2112-2120.
- Argos, P., Landy, A., Abremski, K., Egan, J. B., Haggard-Lindquist, E., Hoess, R. H., Kahn, M. L., Kalionis, B., Narayana, S. V. L., Pierson III, L. S., Sternberg, N. & Leong, J. M. (1986). The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.*, **5**, 433-440.
- Auzat, I. (1993). Coopérativité et régulation allostérique de la phosphofructokinase d'*Escherichia coli*, Université René Descartes, Paris V.
- Auzat, I., Le Bras, G. & Garel, J. R. (1997). Allosteric activation increases the maximum velocity of *E. coli* phosphofructokinase. *J Mol Biol.* **267**, 476-480.
- Blakely, G. W. & Sherratt, D. J. (1996). Cis and trans in site-specific recombination. *Mol Microbiol.* **20**, 234-237.
- Brugger, K., Redder, P., She, Q., Confalonieri, F., Zivanovic, Y. & Garrett, R. A. (2002). Mobile elements in archaeal genomes. *FEMS Microbiol Lett.* **206**, 131-141.

- Cao, Y. & Hayes, F. (1999). A newly identified, essential catalytic residue in a critical secondary structure element in the integrase family of site-specific recombinases is conserved in a similar element in eucaryotic type IB topoisomerases. *J Mol Biol.* **289**, 517-527.
- Chen, J. W., Evans, B. R., Yang, S. H., Araki, H., Oshima, Y. & Jayaram, M. (1992a). Functional analysis of box I mutations in yeast site-specific recombinases Flp and R: pairwise complementation with recombinase variants lacking the active-site tyrosine. *Mol Cell Biol.* **12**, 3757-3765.
- Chen, J. W., Lee, J. & Jayaram, M. (1992b). DNA cleavage in trans by the active site tyrosine during Flp recombination: switching protein partners before exchanging strands. *Cell.* **69**, 647-658.
- Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. & Rice, P. A. (2000). Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol Cell.* **6**, 885-897.
- Chen, S., Vojtechovsky, J., Parkinson, G. N., Ebright, R. H. & Berman, H. M. (2001). Indirect readout of DNA sequence at the primary-kink site in the CAP-DNA complex: DNA binding specificity based on energetics of DNA kinking. *J. Mol. Biol.* **314**, 63-74.
- Chen, Y. & Rice, P. A. (2003). The role of the conserved Trp330 in Flp-mediated recombination. Functional and structural analysis. *J Biol Chem.* **278**, 24800-24807.
- Cossart, P., Groisman, E. A., Serre, M. C., Casadaban, M. J. & Gicquel-Sanzey, B. (1986). *crp* genes of *Shigella flexneri*, *Salmonella typhimurium*, and *Escherichia coli*. *J. Bacteriol.* **167**, 639-646.
- Deville-Bonne, D., Le Bras, G., Teschner, W. & Garel, J. R. (1989). Ordered disruption of subunit interfaces during the stepwise reversible dissociation of *Escherichia coli* phosphofructokinase with KSCN. *Biochemistry.* **28**, 1917-1922.
- Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984a). Molecular basis of DNA sequence recognition by the catabolite gene activator protein: detailed inferences from three mutations that alter DNA sequence specificity. *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7274-7278.
- Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984b). Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. *Nature.* **311**, 232-235.
- Escoubas, J. M., Prere, M. F., Fayet, O., Salvignol, I., Galas, D., Zerbib, D. & Chandler, M. (1991). Translational control of transposition activity of the bacterial insertion sequence IS1. *EMBO J.* **10**, 705-712.
- Escoubas, J. M. (1992). Etude de l'élément transposable bactérien IS1: mise en évidence de la protéine impliquée dans la réaction de transposition, Université Paul Sabatier, Toulouse III.
- Escoubas, J. M., Lane, D. & Chandler, M. (1994). Is the IS1 transposase, InsAB', the only IS1-encoded protein required for efficient transposition? *J Bacteriol.* **176**, 5864-5867.
- Esposito, D. & Scocca, J. J. (1997). The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.* **25**, 3605-3614.
- Evans, P. R. & Hudson, P. J. (1979). Structure and control of phosphofructokinase from *Bacillus stearothermophilus*. *Nature.* **279**, 500-504.
- Evans, P. R., Farrants, G. W. & Lawrence, M. C. (1986). Crystallographic structure of allosterically inhibited phosphofructokinase at 7 Å resolution. *J. Mol. Biol.* **191**, 713-720.
- Evans, B. R., Chen, J. W., Parsons, R., Bauer, T., Teplow, D. & Jayaram, M. (1990). Identification of the active site Tyrosine of FLP recombinase. *J. Biol. Chem.* **265**, 18504-18510.

- Futcher, A. B. (1986). Copy number amplification of the 2 micron circle plasmid of *Saccharomyces cerevisiae*. *J Theor Biol.* **119**, 197-204.
- Galas, D. J. & Chandler, M. (1981). On the molecular mechanisms of transposition. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4858-4862.
- Gopaul, D. N., Guo, F. & Van Duyne, G. D. (1998). Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *EMBO J.* **17**, 4175-4187.
- Guo, F., Gopaul, D. N. & van Duyne, G. D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature.* **389**, 40-46.
- Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol.* **177**, 4121-4130.
- Haniford, D. B., Chelouche, A. R. & Kleckner, N. (1989). A specific class of IS10 transposase mutants are blocked for target site interactions and promote formation of an excised transposon fragment. *Cell.* **59**, 385-394.
- Hickman, A. B., Waninger, S., Scocca, J. J. & Dyda, F. (1997). Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution. *Cell.* **89**, 227-237.
- Huisman, O. & Kleckner, N. (1987). A new generalizable test for detection of mutations affecting Tn10 transposition. *Genetics.* **116**, 185-189.
- Jakowec, M., Prentki, P., Chandler, M. & Galas, D. J. (1988). Mutational analysis of the open reading frames in the transposable element IS1. *Genetics.* **120**, 47-55.
- Koonin, E. V. (1992). Archaeobacterial virus SSV1 encodes a putative DnaA-like protein. *Nucleic Acids Res.* **20**, 1143.
- Kraft, P., Kummel, D., Oeckinghaus, A., Gauss, G. H., Wiedenheft, B., Young, M. & Lawrence, C. M. (2004a). Structure of D-63 from *Sulfolobus* spindle-shaped virus 1: surface properties of the dimeric four-helix bundle suggest an adaptor protein function. *J Virol.* **78**, 7438-7442.
- Kraft, P., Oeckinghaus, A., Kummel, D., Gauss, G. H., Gilmore, J., Wiedenheft, B., Young, M. & Lawrence, C. M. (2004b). Crystal structure of F-93 from *Sulfolobus* spindle-shaped virus 1, a winged-helix DNA binding protein. *J Virol.* **78**, 11544-11550.
- Krebs, M. P. & Reznikoff, W. S. (1988). Use of a Tn5 derivative that creates lacZ translational fusions to obtain a transposition mutant. *Gene.* **63**, 277-285.
- Kwon, H. J., Tirumalai, R., Landy, A. & Ellenberger, T. (1997). Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science.* **276**, 126-131.
- Lane, D., Cavaille, J. & Chandler, M. (1994). Induction of the SOS response by IS1 transposase. *J Mol Biol.* **242**, 339-350.
- Lau, F. T. & Fersht, A. R. (1989). Dissection of the effector-binding site and complementation studies of *Escherichia coli* phosphofructokinase using site-directed mutagenesis. *Biochemistry.* **28**, 6841-6847.
- Le Bras, G. & Garel, J. R. (1982). A proteolyzed derivative of *Escherichia coli* phosphofructokinase is no longer sensitive to allosteric effectors and still shows cooperativity in substrate binding. *Biochemistry.* **21**, 6656-6660.
- Le Bras, G. & Garel, J. R. (1985). Fructose 6-phosphate prevents the proteolyzed derivative of *Escherichia coli* phosphofructokinase from dissociation and inactivation. *J. Biol. Chem.* **260**, 13450-13453.

- Le Bras, G. & Garel, J. R. (1986). Limited proteolysis at the carboxy end modifies interactions between the subunits of *Escherichia coli* phosphofructokinase. *Biochemistry*. **25**, 2490-2493.
- Lee, J., Serre, M. C., Yang, S. H., Whang, I., Araki, H., Oshima, Y. & Jayaram, M. (1992). Functional analysis of Box II mutations in yeast site-specific recombinases Flp and R. Significance of amino acid conservation within the Int family and the yeast sub-family. *J. Mol. Biol.* **228**, 1091-1103.
- Lee, J., Whang, I. & Jayaram, M. (1994). Directed protein replacement in recombination full sites reveals trans-horizontal DNA cleavage by Flp recombinase. *EMBO J.* **13**, 5346-5354.
- Lee, J., Jayaram, M. & Grainge, I. (1999). Wild-type Flp recombinase cleaves DNA in trans. *EMBO J.* **18**, 784-791.
- Letzelter, C., Duguet, M. & Serre, M. C. (2004). Mutational analysis of the archaeal tyrosine recombinase SSV1 integrase suggests a mechanism of DNA cleavage in trans. *J. Biol. Chem.* **279**, 28936-28944.
- Machida, C. & Machida, Y. (1989). Regulation of IS1 transposition by the insA gene product. *J Mol Biol.* **208**, 567-574.
- Machida, Y., Machida, C. & Ohtsubo, E. (1984). Insertion element IS1 encodes two structural genes required for its transposition. *J Mol Biol.* **177**, 229-245.
- Mahillon, J., Leonard, C. & Chandler, M. (1999). IS elements as constituents of bacterial genomes. *Res Microbiol.* **150**, 675-687.
- Martin, A., Yeats, S., Janekovic, D., Reiter, W.-D., Aicher, W. & Zillig, W. (1984). SAV1, a temperate U.V.-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. *EMBO J.* **3**, 2165-2168.
- McKay, D. B. & Steitz, T. A. (1981). Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature*. **290**, 744-749.
- Moitoso de Vargas, L., Pargellis, C. A., Hasan, N. M., Bushman, E. W. & Landy, A. (1988). Autonomous DNA binding domains of lambda integrase recognize two different sequence families. *Cell*. **54**, 923-929.
- Muskhelishvili, G., Palm, P. & Zillig, W. (1993). SSV1-encoded site-specific recombination system in *Sulfolobus shibatae*. *Mol. Gen. Genet.* **237**, 334-342.
- Nadal, M., Mirambeau, G., Forterre, P., Reiter, W. D. & Duguet, M. (1986). Positively supercoiled DNA in a virus-like particle of an archaebacterium. *Nature*. **321**, 256-258.
- Navon, G., Ogawa, S., Shulman, R. G. & Yamane, T. (1977). High-resolution ³¹P nuclear magnetic resonance studies of metabolism in aerobic *Escherichia coli* cells. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 888-891.
- Novotny, M. J., Reizer, J., Esch, F. & Saier, M. H., Jr. (1984). Purification and properties of D-mannitol-1-phosphate dehydrogenase and D-glucitol-6-phosphate dehydrogenase from *Escherichia coli*. *J. Bacteriol.* **159**, 986-990.
- Nunes-Duby, S. E., Tirumalai, R. S., Dorgai, L., Yagil, E., Weisberg, R. A. & Landy, A. (1994). Lambda integrase cleaves DNA in cis. *EMBO J.* **13**, 4421-4430.
- Nunes-Duby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T. & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* **26**, 391-406.
- Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W. D. & Zillig, W. (1991). Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. *Virology*. **185**, 242-250.

- Parsons, R., Prasad, P., Harshey, R. & Jayaram, M. (1988). Step-arrest mutants of FLP recombinase: implication for the catalytic mechanism of DNA recombination. *Mol. Cell. Biol.* **8**, 3303-3310.
- Parsons, R. L., Evans, B. R., Zheng, L. & Jayaram, M. (1990). Functional analysis of Arg-308 mutants of Flp recombinase. Possible role of Arg-308 in coupling substrate binding to catalysis. *J Biol Chem.* **265**, 4527-4533.
- Prangishvili, D., Stedman, K. & Zillig, W. (2001). Viruses of the extremely thermophilic archaeon *Sulfolobus*. *Trends Microbiol.* **9**, 39-43.
- Prangishvili, D. (2003). Evolutionary insights from studies on viruses of hyperthermophilic archaea. *Res Microbiol.* **154**, 289-294.
- Prasad, P. V., Young, L. J. & Jayaram, M. (1987). Mutations in the 2-microns circle site-specific recombinase that abolish recombination without affecting substrate recognition. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2189-2193.
- Rachel, R., Bettstetter, M., Hedlund, B. P., Haring, M., Kessler, A., Stetter, K. O. & Prangishvili, D. (2002). Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch Virol.* **147**, 2419-2429.
- Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J. & Hol, W. G. J. (1998). Crystal structures of human topoisomerase I in covalent and non covalent complexes with DNA. *Science.* **279**, 1504-1513.
- Reiter, W. D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W. & Grampp, B. (1987a). Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. *Mol. Gen. Genet.* **206**, 144-153.
- Reiter, W. D., Palm, P., Yeats, S. & Zillig, W. (1987b). Gene expression in archaeobacteria: physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. *Mol. Gen. Genet.* **209**, 270-275.
- Reiter, W. D., Palm, P. & Zillig, W. (1988). Analysis of transcription in the archaeobacterium *Sulfolobus* indicates that archaeobacterial promoters are homologous to eukaryotic pol II promoters. *Nucleic Acids Res.* **16**, 1-19.
- Reiter, W. D., Palm, P. & Yeats, S. (1989). Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucl. Acids Res.* **17**, 1907-1914.
- Rouquette, C., Serre, M. C. & Lane, D. (2004). Protective role for H-NS protein in IS1 transposition. *J Bacteriol.* **186**, 2091-2098.
- Rypniewski, W. R. & Evans, P. R. (1989). Crystal structure of unliganded phosphofructokinase from *Escherichia coli*. *J. Mol. Biol.* **207**, 805-821.
- Schleper, C., Kubo, K. & Zillig, W. (1992). The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: Demonstration of infectivity and of transfection with viral DNA. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 7645-7649.
- Schreyer, R. & Bock, A. (1980). Phosphoglucose isomerase from *Escherichia coli* K 10: purification, properties and formation under aerobic and anaerobic condition. *Arch Microbiol.* **127**, 289-298.
- Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science.* **253**, 1001-1007.
- Sekine, Y. & Ohtsubo, E. (1989). Frameshifting is required for production of the transposase encoded by insertion sequence 1. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4609-4613.

- Serre, M. C., Gicquel-Sanzey, B. & Cossart, P. (1987). Mutagenesis of Arg180 of the *crp* gene of *Escherichia coli* K12. *Prot. Engineer.* **1**, 235.
- Serre, M. C. (1989). Régulation de la glycolyse chez *Escherichia coli*: obtention et étude de mutants de la phosphofructokinase I, Université Pierre et Marie Curie, Paris VI.
- Serre, M. C. & Garel, J. R. (1990). Role of the C-terminal region in the allosteric properties of *Escherichia coli* phosphofructokinase-1. *Eur. J. Biochem.* **189**, 487-492.
- Serre, M. C., Teschner, W. & Garel, J. R. (1990). Specific suppression of heterotropic interactions in phosphofructokinase by the mutation of leucine 178 into tryptophan. *J. Biol. Chem.* **265**, 12146-12148.
- Serre, M. C., Evans, B. R., Araki, H., Oshima, Y. & Jayaram, M. (1992). Half-site recombinations mediated by yeast site-specific recombinases Flp and R. *J. Mol. Biol.* **225**, 621-642.
- Serre, M. C. & Jayaram, M. (1992). Half-site strand transfer by step-arrest mutants of yeast site-specific recombinase Flp. *J. Mol. Biol.* **225**, 643-649.
- Serre, M. C., Zheng, L. & Jayaram, M. (1993). DNA splicing by an active site mutant of Flp recombinase. Possible catalytic cooperativity between the inactive protein and its DNA substrate. *J. Biol. Chem.* **268**, 455-463.
- Serre, M. C., Turlan, C., Bortolin, M. L. & Chandler, M. (1995). Mutagenesis of the IS1 transposase: Importance of a His-Arg-Tyr triad for activity. *J. Bacteriol.* **177**, 5070-5077.
- Serre, M. C., Letzelter, C., Garel, J. R. & Duguet, M. (2002). Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase. *J. Biol. Chem.* **277**, 16758-16767.
- Shaikh, A. C. & Sadowski, P. D. (2000). Trans complementation of variant Cre proteins for defects in cleavage and synapsis. *J Biol Chem.* **275**, 30186-30195.
- Shirakihara, Y. & Evans, P. R. (1988). Crystal structure of the complex of phosphofructokinase from *Escherichia coli* with its reaction products. *J. Mol. Biol.* **204**, 973-994.
- Slonczewski, J. L., Rosen, B. P., Alger, J. R. & Macnab, R. M. (1981). pH homeostasis in *Escherichia coli*: measurement by ³¹P nuclear magnetic resonance of methylphosphonate and phosphate. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6271-6275.
- Steiniger-White, M., Rayment, I. & Reznikoff, W. S. (2004). Structure/function insights into Tn5 transposition. *Curr. Opin. Struct. Biol.* **14**, 50-57.
- Subramanya, H. S., Arciszewska, L. K., Baker, R. A., Bird, L. E., Sherratt, D. J. & Wigley, D. B. (1997). Crystal structure of the site-specific recombinase, XerD. *EMBO J.* **16**, 5178-5187.
- Teschner, W., Serre, M. C. & Garel, J. R. (1990a). Introduction by site-directed mutagenesis of a tryptophan residue as a fluorescent probe for the folding of *Escherichia coli* phosphofructokinase. *Biochimie.* **72**, 403-406.
- Teschner, W., Serre, M. C. & Garel, J. R. (1990b). Enzymatic properties, renaturation and metabolic role of mannitol-1-phosphate dehydrogenase from *Escherichia coli*. *Biochimie.* **72**, 33-40.
- Ton-Hoang, B., Turlan, C. & Chandler, M. (2004). Functional domains of the IS1 transposase: analysis in vivo and in vitro. *Mol Microbiol.* **53**, 1529-1543.
- Topal, M. D. & Conrad, M. (1993). Changing endonuclease EcoRII Tyr308 to Phe abolishes cleavage but not recognition: possible homology with the Int-family of recombinases. *Nucleic Acids Res.* **21**, 2599-2603.

- Ugurbil, K., Rottenberg, H., Glynn, P. & Shulman, R. G. (1978). ³¹P nuclear magnetic resonance studies of bioenergetics and glycolysis in anaerobic *Escherichia coli* cells. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2244-2248.
- Utatsu, I., Sakamoto, S., Imura, T. & Toh-e, A. (1987). Yeast plasmids resembling 2 micron DNA: regional similarities and diversities at the molecular level. *J Bacteriol.* **169**, 5537-5545.
- van Duyne, G. D. (2002). A structural view of tyrosine recombinase site-specific recombination. In *Mobile DNA II* (Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M., eds.), pp. 93-117. ASM Press, Washington, D.C.
- Volkert, F. C. & Broach, J. R. (1986). Site-specific recombination promotes plasmid amplification in yeast. *Cell.* **46**, 541-550.
- Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A. K., Zoeller, L., Snyder, J., Douglas, T. & Young, M. (2004). Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae viruses. *J Virol.* **78**, 1954-1961.
- Wishart, W. L., Machida, C., Ohtsubo, H. & Ohtsubo, E. (1983). *Escherichia coli* RNA polymerase binding sites and transcription initiation sites in the transposon Tn3. *Gene.* **24**, 99-113.
- Yeats, S., McWilliam, P. & Zillig, W. (1982). A plasmid in the archaeobacterium *Sulfolobus acidocaldarius*. *EMBO J.* **9**, 1035-1038.
- Zerbib, D., Jakowec, M., Prentki, P., Galas, D. J. & Chandler, M. (1987). Expression of proteins essential for IS1 transposition: specific binding of InsA to the ends of IS1. *EMBO J.* **6**, 3163-3169.
- Zerbib, D., Prentki, P., Gamas, P., Freund, E., Galas, D. J. & Chandler, M. (1990). Functional organization of the ends of IS1: specific binding site for an IS 1-encoded protein. *Mol Microbiol.* **4**, 1477-1486.