UNIVERSITÉ FRANÇOIS - RABELAIS

DE TOURS

ÉCOLE DOCTORALE Santé, Sciences, Technologies

GICC – UMR CNRS 6239

THÈSE **présentée par :**

Nicolas BOUCHET

soutenue le : **23 octobre 2009**

pour obtenir le grade de **Docteur de l'université François - Rabelais**

Discipline : Sciences de la Vie

Mécanismes de transposition et de régulation de la transposase de l'élément *mariner Mos1*

THÈSE dirigée par :

Madame AUGE-GOUILLOU Corinne	MCU/HDR, Université François-Rabelais de Tours

RAPPORTEURS :

Monsieur BESSEREAU Jean-Louis	DR2/HDR, Ecole Normale Supérieure de Paris
Madame BETERMIER Mireille	DR2/HDR, Centre de Génétique Moléculaire de Gif/Yvette

---

JURY :

Madame AUGE-GOUILLOU Corinne	MCU/HDR, Université François-Rabelais de Tours
Monsieur BESSEREAU Jean-Louis	DR2/HDR, Ecole Normale Supérieure de Paris
Madame BETERMIER Mireille	DR2/HDR, Centre de Génétique Moléculaire de Gif/Yvette
Monsieur MALLET Jacques	DR1/HDR, Hopital de la Pitié Salpêtrière de Paris
Monsieur PARISSI Vincent	CR2, Université de Bordeaux II

"*Nous piétinerons éternellement aux frontières de l'Inconnu, cherchant à comprendre ce qui restera toujours incompréhensible. Et c'est précisément cela qui fait de nous des hommes.*" Isaac Asimov

Remerciements

Je tiens en premier lieu à remercier Yves Bigot de m'avoir fait confiance en m'accueillant dans son laboratoire pour mon stage de master 2, puis de m'avoir renouvelé cette confiance pour ma thèse.

Un grand merci à mon encadrante Corinne Augé-Gouillou qui m'a laissé la liberté de mener mes travaux de recherche, tout en me donnant de précieux conseils pour que je ne m'égare pas et que je ne baisse pas les bras face aux nombreuses difficultés de manips et autres désillusions !

Merci aux membres de mon jury qui ont accepté de porter un regard critique sur mon travail de thèse : mes rapporteurs Mireille Bétermier et Jean-Louis Bessereau, ainsi que Jacques Mallet et Vincent Parissi.

Six mois de stage en master 2 et trois années de thèse m'ont permis de côtoyer de nombreuses personnes au sein du laboratoire et beaucoup ont eu à supporter mon humeur changeante ! Un énorme merci à Guillaume, le grand sachem de la protéine, qui m'a énormément apporté tant sur le plan expérimental que sur le plan humain. Merci à Jérem, avec qui nous avons porté les bactéries compétentes à une compétence jamais égalée ! Merci à tous les autres qui sont passés par l'équipe 'transposase', notamment Julien, qui a changé ma vision de la transposition, à Stéphanie, qui m'a judicieusement encadré pendant mon master, à Claire, qui est finalement parti du bon côté de la loi, et à Benji, qui m'a ouvert les portes du monde merveilleux des transposons à mon arrivée au laboratoire.

Mais il n'y avait pas que notre petite équipe transposase au sein du laboratoire ! L'arrivée de Patrick Gaudray a ouvert de nouveaux champs de recherche au sein du labo, et m'a permis d'avoir d'intéressantes discussions sur la recherche en général, et les prix Nobel en particulier ! Un grand merci à lui pour sa lecture de mon manuscrit et ses mots d'encouragements pour mon après-thèse. Merci à toutes les autres personnes du labo pour m'avoir supporté si longtemps, particulièrement Sophie, Solenne, Jeanne et Sylvie, que je n'ai pas épargnées, Fabien 'Dinozzo' pour m'avoir bien fait rire avec ses histoires de brevets, et les trois enseignantes que j'avais eues plus jeune et que j'ai retrouvées au labo : Flo, qui avait osé me gronder en cours, Agnès, qui m'avait encouragé à faire de la recherche, et Sylvaine,

qui m’a permis de goûter aux ‘joies’ de l’enseignement et de m’apercevoir que ça n’était pas pour moi ! Et bon courage aux autres thésards, vous verrez finalement c’est pas si dur que ça la rédaction ! Et pour rester dans le domaine universitaire, merci aussi à Géraldine et Stéphane, qui ont eu eux à me supporter depuis mes débuts à la fac !

La thèse, c’est beaucoup de temps passé au labo, une certaine tension continue et de nombreux soucis, mais il y a heureusement des à-côtés dans la vie pour s’aérer l’esprit. J’ai eu la chance d’avoir l’athlétisme et tous les athlètes que j’ai entraînés ces quatre dernières années pour penser à autre chose et passer des moments riches en émotions. Merci à Agathe, Margox, Coline, Lolo, Malika, Jeff, Simon, Adèle, les deux Alex, Kevin ‘Togo’, Salomé, Derrick ‘Gabon’, Prisca, Arnaud, Lulu, Leena, Julie, Faustine et tous les autres pour m’avoir fait connaître de nombreuses joies en tant qu’entraîneur, et à Lindsay, Jo et Erwan avec lesquels je me suis tant entraîné !

Merci à mes parents de m’avoir soutenu pendant mes études, à mon frère, même s’il est passé du côté obscur de la science en choisissant les mathématiques, à mes amis internationaux Branka, Edgar, Cristina, Ah-Kyung et Tom de m’avoir encouragé notamment sur la fin, à Gayane et sa petite fille Maria pour m’avoir montré combien la vie pouvait être miraculeuse. Le projet que j’ai pu mener en parallèle avec les enfants de *Suplacu de Barcau* en Roumanie a changé ma vie, et je les porterai toujours dans mon cœur, en particulier Abi, Ruth, Deny et Tudor, et merci aux jeunes du village de m’avoir accueilli comme l’un des leurs, en particulier Emi et Mihai. Sans oublier Carmen, qui a initié ce superbe projet !

Si j’ai oublié quelqu’un, ce qui est probable, toutes mes excuses, c’est le stress de fin de thèse !

Résumé

Les éléments transposables sont des petits fragments d'ADN capables de se déplacer dans les génomes, transposant d'un locus à un autre. Considérés comme un acteur majeur de l'évolution des génomes, ils jouent un rôle dans la création de nouveaux gènes chez leurs organismes hôtes. Les transposases, enzymes codées par les transposons et qui leur permettent de se déplacer dans les génomes, sont l'objet d'un grand nombre de travaux et d'études. Parmi elles, les enzymes à triade catalytique DDE présentent des caractéristiques communes avec les intégrases rétrovirales et les protéines RAG impliquées dans la recombinaison V(D)J. Au-delà de l'intérêt fondamental que représente l'étude de ces enzymes, les systèmes transposons/transposases peuvent être utilisés comme outils pour la transgénèse et/ou l'analyse des génomes.

Mes travaux de thèse ont consisté en l'étude de la transposase de l'élément transposable *Mos1*, sous différents aspects. *Mos1* est naturellement actif chez la drosophile et appartient à la famille de transposons à ADN la plus répandue dans les génomes eucaryotes.

Mes résultats complètent le modèle de transposition précédemment établi au laboratoire et apportent une nouvelle vision de la formation du complexe synaptique, le complexe qui permet l'excision du transposon du site donneur. Mes travaux ont également permis d'identifier le domaine de liaison à l'ADN de la transposase comme un domaine de type CRO, en parfaite adéquation avec le modèle de formation du complexe synaptique que je propose. La régulation de l'activité de la transposase par des modifications post-traductionnelles a aussi été étudiée, et il a été mis à jour que la transposase de l'élément *Mos1* était phosphorylée. Ces études sur la régulation ainsi que sur le mécanisme de transposition m'ont permis d'élargir le modèle de transposition de *Mos1* au contexte cellulaire eucaryote.

Des travaux d'ingénierie de la protéine ont été également menés afin d'obtenir des transposases hyperactives et fonctionnelles en cellules de mammifères. Ces études montrent la cytotoxicité des transposases hyperactives, et posent la question de l'impact des facteurs d'hôtes lors de la transposition de *Mos1*.

Enfin des inhibiteurs de la transposase de *Mos1* ont été identifiés et leur mode d'action caractérisé. Certains de ces inhibiteurs inhibent également l'intégrase du VIH-1 et/ou la

transposase de l'élément bactérien Tn*5*, ce qui démontre l'efficacité croisée de ces composés sur des enzymes possédant un mécanisme d'action similaire. Quatre inhibiteurs constituent une nouvelle famille de composés dans le domaine des enzymes à triade catalytique DDE et possèdent une activité inhibitrice sur l'intégrase du VIH-1.

Mes travaux ouvrent de nouvelles perspectives de recherche dans le domaine de la mécanique de transposition des enzymes à DDE. Des questions restent en suspens au niveau des ultimes étapes de la transposition et devront être résolue pour compléter le modèle de transposition de *Mos1*. De solides bases sur l'étude de la régulation de *Mos1* ont été posées dans ce mémoire, et il convient de les exploiter pour comprendre l'impact réel des modifications post-traductionnelles sur l'activité de *Mos1*. Enfin, des projets peuvent également être développés autour des inhibiteurs de transposases/intégrases. L'activité croisée de ces composés permet en effet d'utiliser la transposase de *Mos1* comme un modèle pour rechercher des inhibiteurs de l'intégrase du VIH-1.

Mots clés : transposon, *Mos1*, enzymes à triade catalytique DDE, mécanisme de transposition, modifications post-traductionnelles, inhibiteurs de transposases/intégrases

Résumé en anglais

Transposable elements are small DNA fragments able to move in the genomes. They contribute to the genomes evolution, taking part to the creation of new genes. Transposases are the enzymes encoded by the transposons and they allow them to move in the genomes; many studies about these enzymes are carrying out. Among them, DDE enzymes share common features with retroviral intégrases and RAG proteins, which are involved in V(D)J recombination. Beyond the fundamental interest of these enzymes, transposons/transposases systems are used as tools for gene transfer as well as genome analysis.

My work consisted in studying the transposase of transposable element *Mos1*, under different aspects. *Mos1* is naturally active in Drosophila and belongs to the most widespread family of DNA transposons in eukaryotic genomes.

My results complete the model of transposition previously established in the laboratory and bring a new vision of the formation of synaptic complex, the complex that allows excision of the transposon donor site. The DNA-binding of *Mos1* transposase has also been identified as a CRO-like domain. The regulation of transposase activity by post-translational modifications was also investigated, and it appeared that the *Mos1* transposase is phosphorylated. These studies on the regulation and the mechanism of transposition allowed me to expand the model of *Mos1* transposition in a eukaryotic cell context.

The engineering of the protein were also conducted in order to produce hyperactive transposases, which could also be functional in mammalian cells. These studies show the cytotoxicity of hyperactive transposases, and question about the impact of host factors in *Mos1* transposition.

Inhibitors of *Mos1* transposase have been identified and characterized. Some of them also inhibit HIV-1 integrase and/or the transposase of the Tn*5* bacterial transposon. It demonstrates the cross-activity of these compounds. Four inhibitors belong to a new family of DDE enzymes inhibitors and are active against HIV-1 integrase.

My works open new research prospects in the field of DDE enzymes mechanism. There are still some questions about the ultimate step of transposition, and these questions must be resolved to complete the transposition model of *Mos1*. Studies on regulation on *Mos1*

activity must be exploited to understand the real impact of post-translational modifications on the *Mos1* activity. Finally, projects can be developed on transposases and integrases inhibitors. The cross-activity of these compounds allows to use the *Mos1* transposase as a surrogate for an HIV-1 integrase inhibitors screen.

Keywords : transposon, *Mos1*, DDE enzymes, transposition mechanism, post-translationnal modifications, transposases/integrases inhibitors

Table des matières

Remerciements 3
Résumé 5
Résumé en anglais 7
Table des matières 9
Liste des figures 13
Liste des tableaux 17
Liste des annexes 18
Introduction 19
1/ La classification des éléments transposables : un débat toujours d'actualité 23
1.1 Transposons eucaryotes 23
1.2 Transposons procaryotes 26
2/ Les éléments de classe I : les rétrotransposons 28
3/ Les transposons bactériens 30
4 / Les transposons à ADN eucaryotes 31
4.1 Les transposons à mécanisme 'couper-coller' 31
4.1.1 La superfamille IS*630-Tc1-mariner* 32
4.1.2 La superfamille *hAT* 41
4.1.3 La superfamille *Merlin* 42
4.1.4 La superfamille *Transib* 43
4.1.5 La superfamille *Mutator* 43
4.1.6 La superfamille *PIF/Harbinger* 44
4.1.7 La superfamille *piggyBac* 45
4.1.8 L'élément *P* 46
4.1.9 Les éléments CACTA 47
4.1.10 L'ordre des éléments Cryptons 47
4.2 Transposons à mécanisme 'copier-coller' 47
4.2.1 Les *hélitrons* : transposition par cercle-roulant 48
4.2.2 Les *polintons/Mavericks* : des transposons complexes 49
5/ Mécanismes de transposition 51
5.1 Les éléments modèles Tn*5* et Tn*10* 53
5.1.1 Tn*5* : 'un modèle pour comprendre la transposition' 53

5.1.2 Tn*10* : un modèle de transposition conservative avec implication de protéines d'hôte 57
5.2 IS*911* : transposition avec un intermédiaire circulaire 61
5.3 Les éléments *ITm* 62
5.3 SETMAR : une transposase *mariner* domestiquée 72
5.4 RAG et la recombinaison V(D)J 75
5.5 L'intégrase du VIH-1 78
6 Régulation de la transposition 83
6.1 Expression de la transposase 83
6.2 Facteurs physico-chimiques influençant la transposition 85
6.2.1 La température 85
6.2.2 Les cations divalents 86
6.2.3 L'élément *P* et le GTP 86
6.3 Le transposon 87
6.4 Facteurs d'hôte et modifications post-traductionnelles 88
6.4.1 Facteurs d'hôte protéiques 88
6.4.2 Modifications post-traductionnelles 89
Résultats 91
1/ *Mos1* : la transposase 94
1.1 Le domaine de liaison à l'ADN : un ou deux HTH ? 94
1.1.1 Un ou deux HTH : point sur la controverse et conséquences sur l'assemblage du complexe synaptique 94
1.1.2 Transposases chimériques 95
1.1.3 Conclusion 98
1.2 Ingénierie de MOS1 99
1.2.1 Transposases hyperactives 99
1.2.2 Bilan de l'ingénierie de la transposase de *Mos1* 103
2/ *Mos1* : Transposition et régulation 105
2.1 Transposase procaryote versus transposase eucaryote 106
2.2 Mécanique de transposition 110
2.2.1 SEC et PEC 110
2.2.2 Stœchiométrie du complexe A 111
2.2.3 Le PEC1 est un SEC2 ! 113
2.2.3 Capture de cible par un SEC2 ? 114

2.3 MOS1 et phosphorylation 117
2.3.1 MOS1 est phosphorylée 118
2.3.2 Action des phosphatases sur MOS1 120
2.3.3 Transposase cytoplasmique versus transposase nucléaire 123
2.3.4 Pattern de phosphorylation : spectrométrie de masse 124
2.3.5 Design d'une transposase non-phosphorylable 125
2.3.6 Phosphorylation : conclusion 132
2.3.7 Autres modifications post-traductionnelles ? 133
2.4 Ponts disulfures chez MOS1 134
2.4.1 Rappels sur les ponts disulfures 134
2.4.2 La PPase2A met sur la piste de ponts disulfures chez MOS1 136
2.4.3 Approche directe : spectrométrie Raman 140
2.4.4 Ponts disulfures et stabilité de MOS1 143
2.4.5 Ponts disulfures et interactions protéine-protéine 145
2.4.5 Mutants des ponts disulfures 146
2.4.6 Ponts disulfures : conclusion 147
3/ Inhibiteurs de transposase 148
3.1 Criblage de composés chimiques sur MOS1 148
3.2 Inhibiteurs non retenus 153
3.3 Inhibition *in vivo* ? 154
3.4 Mode d'action des inhibiteurs 155
3.4.1 Inhibition de l'excision 155
3.4.2 Inhibition de l'assemblage des complexes ITR-transposase 157
3.4.3 Effet des dérivés maléimides sur le transfert de brin 161
3.5 Effet des dérivés maléimides sur l'intégrase du VIH-1 163
3.6 Inhibiteurs : conclusions 165
3.6.1 Inhibiteurs de Tn*5* et polyamides 165
3.6.2 Les dérivés maléimides 166
Discussion & Perspectives 169
Références bibliographiques 185
Annexes 211

Liste des figures

Figure 1. Modes de transposition des éléments de deux grandes classes d'éléments transposables. 23
Figure 2. Classification 'universelle' des éléments transposables proposée par Wicker 25
Figure 3. Classification 'universelle' des éléments transposables proposée par Kapitonov & Jurka 26
Figure 4. La transposition 'couper-coller' schématisée 32
Figure 5. Organisation des éléments IS*630*-Tc*1*-*mariner* 34
Figure 6. Structure du domaine de liaison à l'ADN de la transposase de Tc*3* 36
Figure 7. Schéma simplifié de la structure de l'élément *mariner Mos1* 38
Figure 8. Organisation schématisée de la transposase MOS1 40
Figure 9. Structure cristallographique de la transposase MOS1 40
Figure 10. Schéma simplifié de la transposition de *piggyBac* 46
Figure 11. Mécanisme de transposition en cercle-roulant des *hélitrons* 49
Figure 12. Mécanisme de transposition des *polintons* 50
Figure 13. Cœurs catalytiques de quelques enzymes RISF 51
Figure 14. Le motif DDE est conservé chez transposases et intégrases 52
Figure 15. Structure moléculaire du complexe synaptique de Tn*5* 53
Figure 16. Schéma de transposition de Tn*5* 54
Figure 17. Excision de Tn*5* 55
Figure 18. Clivage séquentiel des extrémités de Tn*5* 56
Figure 19. Structure de l'élément Tn*10* 57
Figure 20. Organisation schématique d'un OE de l'élément Tn*10* 57
Figure 21. Mécanisme de transposition de Tn*10* 58
Figure 22. Formation du transpososome de Tn*10* avec IHF 59
Figure 23. Schéma simplifié de la transposition de Tn*10* 60
Figure 24. Transposition de IS*911* 61
Figure 25. Schéma simplifié de la transposition de l'élément *Mos1* 63
Figure 26. Formation du complexe synaptique de *Mos1* 64
Figure 27. Clivage des extrémités du transposon *Mos1* 65
Figure 28. Coupure imprécise chez les éléments *mariner* 66
Figure 29. Chimie du clivage du brin transféré 67
Figure 30. Rôle des cations divalents lors du clivage 68

Figure 31. Réaction de transfert de brin chez les enzymes à DDE 70
Figure 32. Chimie du transfert de brin et de la désintégration 70
Figure 33. Rôle des cations divalents dans la catalyse 71
Figure 34. Formation des complexes ADN-transposase pendant les premières étapes de la réaction 71
Figure 35. La protéine SETMAR 72
Figure 36. La recombinaison V(D)J 75
Figure 37. Schéma d'un RSS 76
Figure 38. Clivage de l'ADN chez *Hermes* et lors de la recombinaison V(D)J 77
Figure 39. Les trois étapes de la recombinaison V(D)J 78
Figure 40. Les cations pontent la triade catalytique de l'intégrase et l'ADN 79
Figure 41. Structure des cœurs catalytiques RNaseH-*like* d'enzymes à DDE 80
Figure 42. Formation du PIC et passage dans le noyau 81
Figure 43. L'intégration rétrovirale vue sur le plan biochimique 81
Figure 44. Instabilité génétique chez le maïs 84
Figure 45. Impact de la taille du transposon sur sa fréquence de transposition en bactérie 87
Figure 46. Protéines chimériques LZ[34-345] et Gal4[35-345] 96
Figure 47. Analyse de la transposition *in vivo* des protéines tronquées 98
Figure 48. Fréquence de transposition des mutants ponctuels retenus 100
Figure 49. Fréquence de transposition des sept mutants multiples d'intérêt 101
Figure 50. Analyse cytotoxique avec les mutants multiples 102
Figure 51. Transposition *in vitro* de MOS1, et des mutants FET et FETY 102
Figure 52. Phosphorylation de MOS1 *in vitro* 106
Figure 53. Stabilité de la protéine eucaryote 107
Figure 54. Profil de liaison à l'ADN des transposases procaryotes et eucaryotes 108
Figure 55. Liaison aux ITR - gammes de transposases 109
Figure 56. Profil de liaison de la transposase procaryote sur ses ITR 110
Figure 57. Un dimère de transposase procaryote interagit avec les ITR de *Mos1* 111
Figure 58. Le complexe A contient deux transposases 111
Figure 59. Principe de la méthode ITR court/long 112
Figure 60. ITR court/long avec la transposase eucaryote 112
Figure 61. ITR court/long avec la transposase procaryote, d'après Augé-Gouillou *et al* 113
Figure 62. ITR court/long avec la transposase procaryote 114
Figure 63. Nouveau profil de liaison de la transposase procaryote à l'ADN 115

Figure 64. Différents ITR 3’ 116
Figure 65. Profil de liaison de MOS1 à différents ITR 3’ 117
Figure 66. La MBP-transposase produite en système eucaryote est phosphorylée 118
Figure 67. La transposase produite en système eucaryote est phosphorylée 119
Figure 68. Action de différentes phosphatases sur la transposase eucaryote 121
Figure 69. Impact de la déphosphorylation sur la liaison de MOS1 à l’ADN 122
Figure 70. Impact de la λPPase sur la stabilité de la transposase eucaryote 122
Figure 71. La transposase nucléaire est phosphorylée 123
Figure 72. Pattern partiel de phosphorylation de MOS1 125
Figure 73. Comparaison de la structure de quatre acides aminés 128
Figure 74. Conservation de résidus phosphorylables dans le motif HTH des éléments *mariner* 130
Figure 75. Ponts disulfures 134
Figure 76. Formation d’un pont disulfure par échange thiol-sulfide 135
Figure 77. Les trois conformations possibles des ponts disulfures 136
Figure 78. Impact de la déphosphorylation par la PPase2A sur la liaison de MOS1 à l’ADN 137
Figure 79. Impact de la déphosphorylation par la PPase2A sur la stabilité de la transposase 137
Figure 80. Impact du β-mercaptoéthanol sur la formation des complexes ITR-Tnp 139
Figure 81. Structure 3D de la partie C-terminale de MOS1 139
Figure 82. Structure 3D de la partie C-terminale de MOS1 140
Figure 83. Spectre Raman de MOS1 142
Figure 84. Impact du β-mercaptoéthanol sur la stabilité de MOS1 143
Figure 85. Impact du DTT sur l’interaction transposase-transposase 145
Figure 86. Impact des mutations C136S et C336S sur la formation des complexes ITR-Tnp 146
Figure 87. Les 10 composés inhibiteurs de MOS1 153
Figure 88. Activité inhibitrice du composé 9 sur l’excision du transposon 155
Figure 89. Activité inhibitrice des 10 composés 156
Figure 90. Impact des inhibiteurs 5 à 10 sur la formation du complexe ITR-transposase 157
Figure 92. Les dérivés maléimides n’interagissent pas avec l’ADN 159
Figure 93. Impact des dérivés maléimides sur la formation des complexes ITR-MOS1 160
Figure 94. Impact des dérivés maléimides sur le transfert de brin 162

Figure 95. Activité inhibitrice du composé 4 sur les activités de *3'-processing* et de transfert de brin de l'intégrase du VIH-1 164
Figure 96. Nouveau modèle de formation du complexe synaptique de *Mos1* 175
Figure 97. Cinétique de formation des complexes 176
Figure 98. Modèle de la formation du transpososome chez *Mos1* 177
Figure 99. Modèle de transposition de *Mos1* dans un contexte cellulaire eucaryote 181

Liste des tableaux

Tableau 1. Fréquence de transposition des transposases tronquées 97
Tableau 2. Fréquence de transposition des différentes transposases eucaryotes 124
Tableau 3. Résidus prédits phosphorylés chez MOS1 126
Tableau 4. Fréquences de transposition des mutants du domaine de dimérisation de MOS1 128
Tableau 5. Fréquences de transposition des mutants non-phosphorylables du motif HTH de MOS1 131
Tableau 6. Effet de différentes facteurs sur le stabilité de MOS1..... 144
Tableau 7. Facteur d'inhibition (FI) et pourcentage d'inhibition (PI) des composés MCV .. 150
Tableau 8. Facteur d'inhibition (FI) et pourcentage d'inhibition (PI) des composés CG, D et R 151
Tableau 9. Effet des dérivés maléimides sur l'intégrase du VIH-1 164

Liste des annexes

1/ Publications et communications 213
2/ Matériel et Méthodes 263
2.1 Techniques courantes 263
2.2 Composés chimiques testés sur la transposase 266
2.3 Analyses protéiques 268
2.4 Tests de transposition 270
2.5 Etudes des complexes ADN-protéine 273
2.6 Etude de la phosphorylation de MOS1 275
3/ Catalogue des mutants de MOS1 276

Introduction

Dans l'immense complexité qui la caractérise, la vie a permis la diffusion à travers tous les organismes vivants des éléments transposables (ET), également connus sous le nom de transposons. Ces ET, finalement peu connus par les physiologistes, sont des petits fragments d'ADN capables de se déplacer dans les génomes, transposant d'un locus à un autre. Ces ET ont été découverts dans les années 40 par la généticienne américaine Barbara McClintock, qui étudiait l'instabilité génétique chez le maïs (McClintock, 1948 et 1950) ; dans les années 60, les avancées en génétique moléculaire permettent l'identification des séquences d'insertion (IS), les éléments transposables bactériens, et la résistance bactérienne promue par les transposons. C'est dans les années 70 que la réalité de ces '*jumping gene*s' est définitivement établie, et McClintock est nobélisée en 1983 pour ses travaux.

Souvent présentés comme des parasites génétiques ou de l'ADN inutile ('*junk DNA*'), les ET sont désormais considérés comme un élément majeur de l'évolution des génomes, jouant notamment un rôle dans la création de nouveaux gènes (Biémont & Vieira, 2006). Les programmes de séquençage à grande échelle des génomes ont permis de montrer que les ET constituaient une part importante du matériel génétique chez la plupart des eucaryotes ; représentant environ 20% du génome de la drosophile, les ET constituent 44% du génome humain (Lander *et al*, 2001), 60% de celui du maïs et jusqu'à 77% de celui de la grenouille (Biémont & Vieira, 2006). La découverte de nouveaux transposons est d'actualité, tant chez les animaux (élément *Passport* chez les poissons – Clark *et al*, 2009), chez les plantes (élément *Lullaby* chez le riz – Picault *et al*, 2009) que chez les bactéries (IS*1642* chez *Mycobacterium avium* – Piao *et al*, 2009).

D'abord essentiellement étudiés d'un point de vue de leur évolution au sein des génomes, les éléments transposables sont de plus en plus étudiés pour l'impact qu'ils ont sur les génomes. Les transposons eux-mêmes peuvent avoir une fonction bénéfique pour l'hôte, tels les rétrotransposons HeT-A et TART qui participent au maintien des télomères chez les drosophiles, se substituant au rôle d'une télomérase de l'hôte (Pardue & DeBaryshe, 2003), ou les éléments LINE-1 qui jouent un rôle dans l'inactivation du chromosome X (Lyon, 1998 et 2000 ; Bailey *et al*, 2000). Mais la majorité des cas connus de domestication d'éléments transposables a conduit à la création de nouveaux gènes pour l'organisme hôte (Volff, 2006 ; Dooner & Weil, 2007 ; Feschotte, 2008 ; Sinzelle *et al*, 2009), néogènes qui remplissent diverses fonctions pour leur hôte. Certains sont mêmes devenus des gènes vitaux pour l'organisme ; l'exemple le plus connu est sans aucun doute la protéine RAG1, impliquée dans la recombinaison V(D)J. Cet élément clé du système immunitaire chez les mammifères dérive

des éléments *Transib* (Kapitonov & Jurka, 2005). Un autre exemple se retrouve chez *Arabidopsis thaliana,* où le gène *Daysleeper* code une protéine similaire aux transposases *hAT*, protéine essentielle pour un développement normal de la plante et qui intervient dans la régulation globale de l'expression des gènes (Bundock & Hooykas, 2005). Avec le développement de l'analyse des génomes, les exemples de néogènes abondent chez tous les organismes. L'un des exemples les plus étudiés actuellement est la protéine SETMAR, connue également sous le nom de Metnase, qui est une protéine fusion entre un domaine SET et une transposase *mariner* (Cordaux *et al*, 2006 ; Liu *et al*, 2007 ; Miskey *et al*, 2007). Cette protéine a une fonction d'histone méthyl-transférase ; elle est impliquée dans les mécanismes de réparation de l'ADN et augmente la résistance aux radiations (Lee *et al*, 2005 ; Roman *et al*, 2007 ; Beck *et al*, 2008).

L'impact sur les génomes n'est pas le seul sujet d'étude sur les éléments transposables ; les transposases, enzymes codées par les transposons et qui leur permettent de se déplacer dans les génomes, sont source d'un grand intérêt. Ces enzymes sont en effet capables d'exciser le transposon de son locus, et de l'intégrer dans un autre locus. Ce système transposon/transposase est maintenant relativement bien connu pour certaines familles d'éléments transposables, et peut être utilisé comme outil pour analyser les génomes et/ou pour faire de la transgénèse.

1/ La classification des éléments transposables : un débat toujours d'actualité

Les éléments transposables sont traditionnellement répartis en deux classes en fonction de leur mécanisme de transposition (figure 1) : les éléments de classe I, ou rétrotransposons, qui transposent *via* un intermédiaire à ARN, et les éléments de classe II, également appelés transposons à ADN, qui transposent *via* un intermédiaire à ADN (Finnegan, 1989).

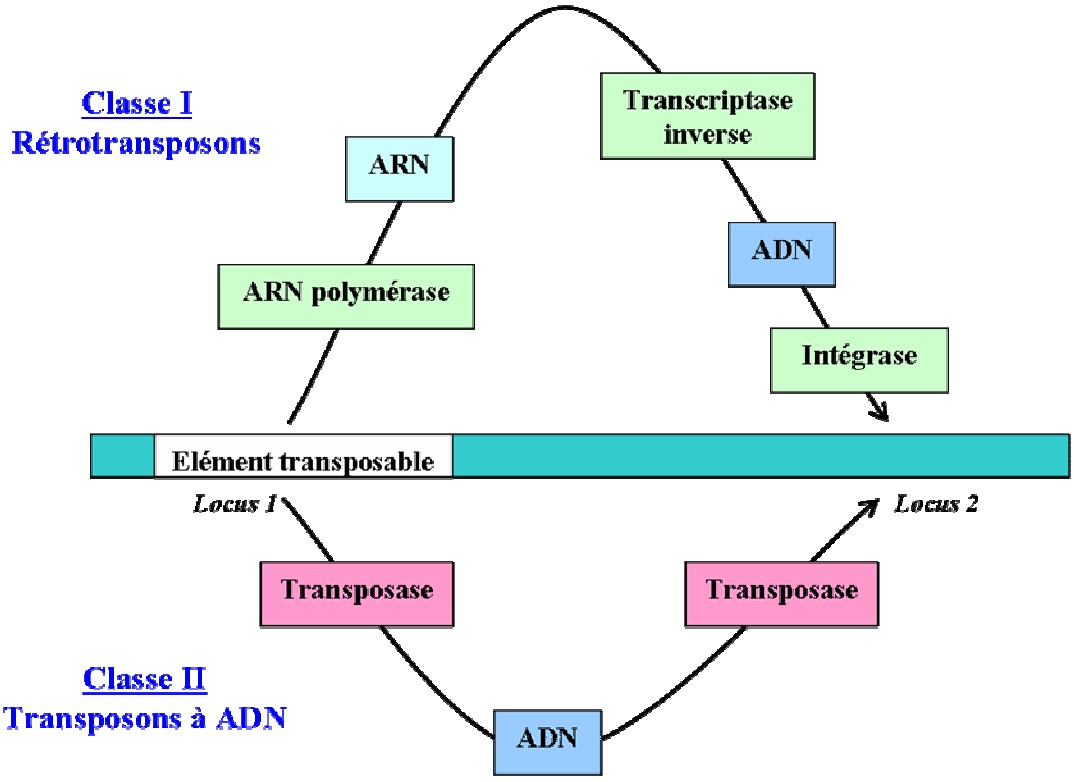


Figure 1. Modes de transposition des éléments de deux grandes classes d'éléments transposables.

1.1 Transposons eucaryotes

La profusion de nouveaux transposons, conséquence de l'analyse à grande échelle des génomes, a conduit à nommer et classer les nouveaux ET parfois sans vraiment suivre de règles de nomenclature. Pour remédier à cela, et pour permettre une annotation plus simple des génomes, un groupe de chercheurs (Wicker *et al*, 2007) a récemment proposé une mise à

jour de la classification (figure 2). Les deux grandes classes d'éléments (rétrotransposons et transposons à ADN), qui diffèrent par la présence ou non d'un intermédiaire à ARN, sont retrouvées en premier lieu. Suivent alors, comme pour les classifications phylogénétiques des organismes vivants, des sous-classes, des ordres, des superfamilles et familles, divisions qui sont fonction du mécanisme de transposition, des enzymes impliquées dans la transposition, et des similitudes de séquence.

Les rétrotransposons sont divisés en cinq ordres, d'après leur mécanisme de transposition et la phylogénie de leur transcriptase inverse (RT) : les rétrotransposons LTR, les éléments *DIRS-like*, les éléments *Penelope-like*, les LINE et les SINE (pour plus de détails, voir chapitre 2, *Les éléments de classe I : les rétrotransposons*).

Les éléments de la classe II sont, quant à eux, divisés en deux sous-classes, qui diffèrent en fonction du nombre de brin d'ADN coupé pendant la transposition. Dans la première sous-classe, on trouve les éléments qui transposent *via* un mécanisme dit 'couper-coller', répartis en deux ordres ; le premier, l'ordre ITR (pour *Terminal Inverted Repeat*), où l'on retrouve les transposons les plus connus et les plus étudiés (éléments *Tc1-mariner*, l'élément *P*, les transposons *piggyBa*c…). Le deuxième ordre est uniquement composé des éléments *Cryptons*, éléments exotiques peu étudiés, identifiés seulement chez les champignons. Dans la seconde sous-classe sont regroupés les *hélitrons* et les *polintons* (également connus sous le nom de *Maverick*) qui transposent par des mécanismes totalement différent des autres transposons de la classe II (mécanismes qui seront décrits plus loin).

Avec cette classification, les éléments de classe II comptent 12 superfamilles de transposons bien caractérisées (figure 2).

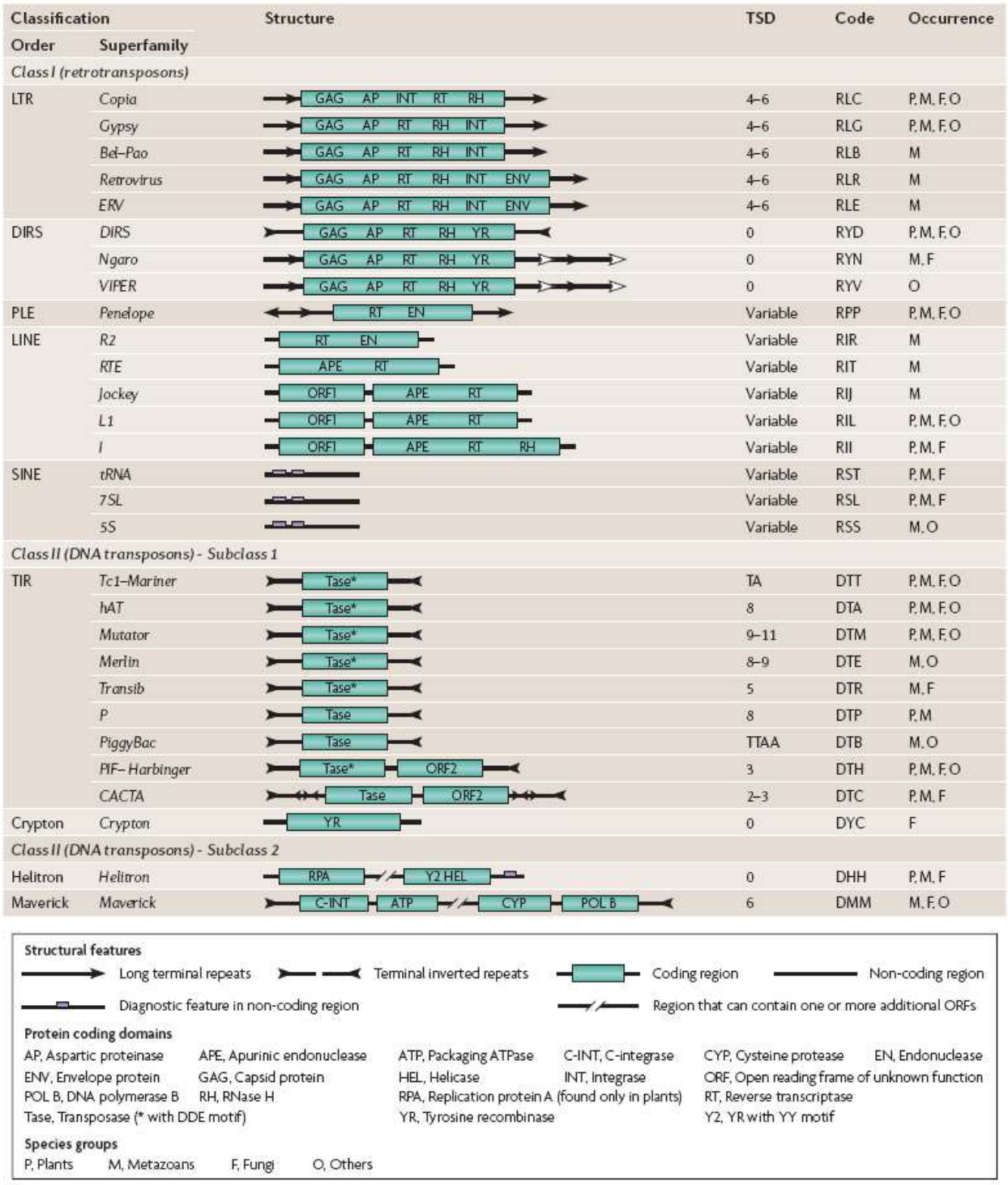
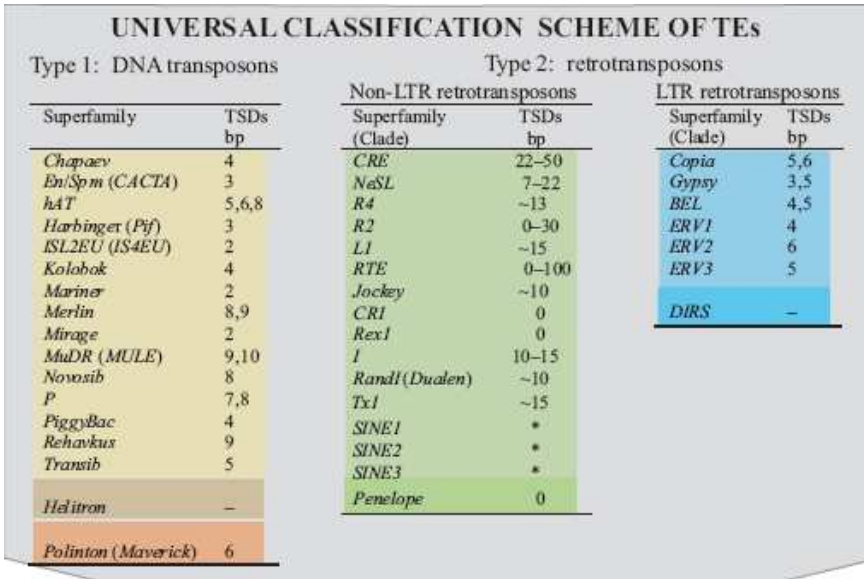


Figure 2. Classification 'universelle' des éléments transposables proposée par Wicker *et al*, 2007.

En réponse à cette 'classification universelle' proposée par Wicker *et al* en 2007, deux autres chercheurs, Vladimir Kapitonov et Jerzy Jurka, publient en 2008 leur propre classification, qui s'appuie sur Repbase, une base de données informatique des éléments transposables eucaryotes qui contient plus de 7600 séquences d'ET et de séquences répétées

(Kapitonov & Jurka, 2008). Cette classification est basée sur l'enzymologie, les similitudes structurelles et les relations entre les séquences d'ADN ; on retrouve les deux grandes classes d'éléments (rétrotransposons et transposons à ADN) divisés ensuite en cinq classes puis 40 superfamilles, avec des éléments classés à part sans raisons véritablement justifiées dans leur article (figure 3).



UNIVERSAL CLASSIFICATION SCHEME OF TEs

Type 1: DNA transposons

Superfamily	TSDs bp
Chapaev	4
En/Spm (CACTA)	3
hAT	5,6,8
Harbinger (Pif)	3
ISL2EU (IS4EU)	2
Kolobok	4
Mariner	2
Merlin	8,9
Mirage	2
MuDR (MULE)	9,10
Novosib	8
P	7,8
PiggyBac	4
Rehavkus	9
Transib	5
Helitron	–
Polinton (Maverick)	6

Type 2: retrotransposons

Non-LTR retrotransposons

Superfamily (Clade)	TSDs bp
CRE	22–50
NeSL	7–22
R4	~13
R2	0–30
L1	~15
RTE	0–100
Jockey	~10
CR1	0
Rex1	0
I	10–15
RandI (Dualen)	~10
Tx1	~15
SINE1	*
SINE2	*
SINE3	*
Penelope	0

LTR retrotransposons

Superfamily (Clade)	TSDs bp
Copia	5,6
Gypsy	3,5
BEL	4,5
ERV1	4
ERV2	6
ERV3	5
DIRS	–

Figure 3. Classification 'universelle' des éléments transposables proposée par Kapitonov & Jurka, 2008. On retrouve les deux grandes classes d'éléments (type 1 et type 2), avec 15 superfamilles de transposons à ADN, plus les hélitrons et les polintons classés en marge. Les rétroéléments sont classés en deux grands groupes, les non-LTR, avec 15 superfamilles, plus les éléments Penelope en marge, et les rétrotransposons à LTR avec 6 superfamilles plus les éléments DIRS eux aussi classés en marge.

La classification présentée par Wicker et ses collaborateurs semble plus simple et plus abordable que celle proposée par Kapitonov et Jurka, basée quasi-exclusivement sur des analyses de génomes, là où Wicker nous propose une classification basée plus sur le mécanisme de transposition. Ma thèse s'articulant autour du mécanisme de transposition des éléments *mariner*, c'est la classification de Wicker qui me servira de fil rouge pour la suite de mon introduction bibliographique.

1.2 Transposons procaryotes

Assez curieusement, les transposons procaryotes ont été oubliés des deux classifications précédemment décrites. Ils sont traditionnellement classés à part selon leur structure et les enzymes impliquées dans le mécanisme de transposition, et pas sur le mode de transposition lui-même. Il existe quelques publications références pour classer les éléments

bactériens, avec des règles maintenant bien établies de nomenclature (Toussaint & Merlin, 2001 ; Roberts *et al*, 2008). Les transposons bactériens ne seront évoqués que brièvement dans la présentation générale des ET, et je m'attarderai plus sur trois d'entre eux dans le chapitre consacré au mécanisme des transposases et enzymes apparentées.

2/ Les éléments de classe I : les rétrotransposons

Les éléments de classe I transposent *via* un intermédiaire ARN ; il s'agit d'un mode de transposition réplicatif, aussi appelé 'copier-coller', par opposition à celui de la majorité des transposons à ADN qui se déplacent dans les génomes *via* un mode dit 'couper-coller'. Les éléments de classe I transposent de la manière suivante : un intermédiaire ARN est transcrit à partir d'une copie de l'élément présente dans le génome, puis cet intermédiaire est rétrotranscrit en ADN par une transcriptase inverse (RT), codée ou non par l'élément, ce qui donne une nouvelle copie de l'élément à chaque cycle de réplication.

Les rétrotransposons sont divisés en cinq ordres, en fonction de leur mécanisme de transposition et des liens phylogénétiques de leur RT :

- les **rétrotransposons à LTR**, très présents chez les plantes et peu retrouvés dans le règne animal, sont très proches des rétrovirus en terme d'organisation. Ils codent la protéine structurelle GAG, une intégrase, une RNaseH et une RT. Les LTR (pour *Long Terminal Repeat*) sont de longues séquences d'ADN flanquant les gènes de l'élément et sont utilisés dans le processus d'intégration au génome de l'hôte. Il est proposé que certains rétrovirus aient évolué à partir des rétrotransposons à LTR *Gypsy* en ayant acquis le gène *env* codant la protéine d'enveloppe rétrovirale. L'inverse est également vrai, des rétrovirus perdant leur capacité de mobilité extracellulaire pouvant devenir des rétroéléments à LTR, ce sont les rétrovirus endogènes (Bannert & Kurth, 2006).

- les **éléments *Penelope-like*** (PLE), découverts chez la drosophile (Evgen'ev *et al*, 1997 et 2005), sont présents chez les plantes, les champignons et les animaux. Ils ont la particularité de coder une protéine de fusion entre une RT et une endonucléase (Pyatkov *et al*, 2004). L'élément *Penelope* est responsable de la dysgénésie des hybrides chez *Drosophila virilis*, syndrome causé par un fort taux de mutations dans les cellules germinales.

- les **éléments *DIRS-like***, très répandus chez les eucaryotes (Goodwin & Poulter, 2004), codent une recombinase à tyrosine à la place d'une intégrase. Ils sont

classés parmi les éléments de classe I du fait qu'ils transposent *via* un intermédiaire à ARN.

- les **éléments LINE** sont retrouvés chez tous les eucaryotes où ils ont été recherchés et ont la particularité de ne pas posséder de LTR. Parmi ces LINE, on retrouve l'élément L1, qui représente environ 20% du génome humain, plus de 99% d'entre eux étant inactifs. (Lee *et al*, 2007).

- les **SINE** sont des petits éléments non autonomes qui se multiplient dans les génomes en utilisant la machinerie enzymatique des éléments LINE autonomes ; ils possèdent néanmoins un promoteur pour leur transcription par l'ARN polymérase III (Kramerov & Vassetzky, 2005). L'élément *Alu*, dont on retrouve plus d'un million de copies dans le génome humain (Bennett *et al*, 2008), est un SINE. Certains de ces éléments *Alu* sont actifs chez l'homme (Mills *et al*, 2007) et leur transposition est source de maladies, 43 au moins ayant été identifiées (Belancio *et al*, 2008), comme l'hémophilie A ou B ou certains cancers du sein.

Les rétrotransposons ont un impact certain sur les génomes de leur hôte, et notamment chez l'homme, où ils représentent quelques 42% du génome, la majorité étant cependant inactifs. Les éléments L1 sont un élément majeur de la plasticité du génome humain (Cordaux, 2008), et sont impliqués dans des cassures de l'ADN (Farkash & Prak, 2006), et des cancers (Schulz, 2005). Ces éléments seraient également impliqués dans l'inactivation du chromosome X chez les mammifères (Lyon, 2000 et 2006). Chez la drosophile, les rétrotransposons HET-A et TART se sont carrément substitués à la télomérase de l'hôte pour permettre la maintenance des télomères (Pardue & DeBaryshe, 2003).

3/ Les transposons bactériens

Les transposons bactériens qui transposent *via* un intermédiaire à ADN sont retrouvés dans tout le spectre bactérien ; ils représentent moins de 2% du génome de *Escherichia coli*, mais jusqu'à 25% du génome de *Enterococcus faecalis* (Paulsen *et al*, 2003) et plus de 40% de celui de *Orientia tsutsugamushi* (Nakayama *et al*, 2008). Ils sont classés selon leur structure et le type d'enzymes impliqués dans leur transposition (Roberts *et al*, 2008).

Les plus connus des transposons procaryotes sont les **séquences d'insertion (IS)** ; il s'agit de petits éléments autonomes, de taille inférieure à 2,5 kb, codant une transposase (et parfois une seconde protéine, peut-être un régulateur de la transposition – Sekine *et al*, 1997), ce gène étant flanqué par des séquences répétées inversées de 9 à 40 pb (Siguier *et al*, 2006). Certains de ces IS sont classés avec des transposons à ADN eucaryotes, comme l'élément IS*630*, qui appartient à la grande superfamille des éléments IS*630-Tc1-mariner*, puisqu'il code une transposase à triade catalytique DDE et s'insère dans les génomes au niveau d'un dinucléotide TA, comme les éléments *Tc1-mariner* (Tenzen *et al*, 1990).

Chez ces transposons bactériens, on trouve également des **éléments** dits **composites**, qui sont constitués de deux IS dégénérés en répétition directe ou inversée, contenant une séquence d'ADN exogène entre les deux IS, séquence qui code souvent pour un gène de résistance aux antibiotiques. Chez ces éléments, seul un des deux IS code pour une transposase active (de la Cruz *et al*, 1993). Les éléments Tn*5* et Tn*10*, dont le mécanisme de transposition sera détaillé plus loin, sont des transposons composites.

Les **éléments non composites** sont quant à eux des transposons plus complexes que les éléments composites. Ils possèdent des ITR et codent une transposase ainsi que des protéines accessoires de la transposition, utilisées dans l'assemblage du complexe synaptique ou lors de la capture cible dans la transposition. Les exemples les plus connus de ces éléments sont le transposon Tn*7* et Mu (Craig, 2002 ; Chakonas & Harshey, 2002).

4 / Les transposons à ADN eucaryotes

Les transposons à ADN eucaryotes transposent *via* un intermédiaire à ADN, le plus souvent à l'aide d'une enzyme unique capable de promouvoir toutes les étapes de la transposition, la transposase (Tnp). Ces éléments sont divisés en deux sous-classes, selon qu'il y ait ou non une coupure double-brin de l'ADN donneur lors de la transposition.

4.1 Les transposons à mécanisme 'couper-coller'

Cette sous-classe d'éléments est divisée en deux ordres : l'ordre ITR et l'ordre Crypton (figure 2). L'ordre ITR comprend neuf superfamilles, ces éléments étant caractérisés par des ITR de longueur variable. La taille des ITR, la taille et la séquence de la duplication d'ADN au niveau du site d'insertion après l'insertion du transposon, ainsi que l'analogie de séquences entre les éléments sont autant de critères permettant de différencier ces superfamilles. Le deuxième ordre des transposons 'couper-coller' est la superfamille *Crypton*, dont les éléments ne possèdent pas d'ITR mais transposent *via* un intermédiaire à ADN avec une coupure double-brin de l'ADN donneur.

Ces éléments ont une organisation assez simple : des séquences d'ADN de longueur variable, codant la transposase (Tnp) qui est la protéine nécessaire à leur mobilité, flanquées par des ITR de longueur variable (exception faite des éléments *Cryptons*). La plupart de ces éléments possèdent une transposase caractérisée par une triade catalytique d'acides aminés de type DDE, avec une structure analogue à celle des intégrases rétrovirales et de la RNaseH (Rice & Baker, 2001 ; Hickman *et al*, 2005 ; Nowotny, 2009).

Sont présentées ici ces dix superfamilles, avec leurs principales caractéristiques structurales, avec tout d'abord les éléments de l'ordre ITR avec ceux possédant une transposase avec une triade catalytique DDE, puis les autres, et enfin les éléments *Cryptons*.

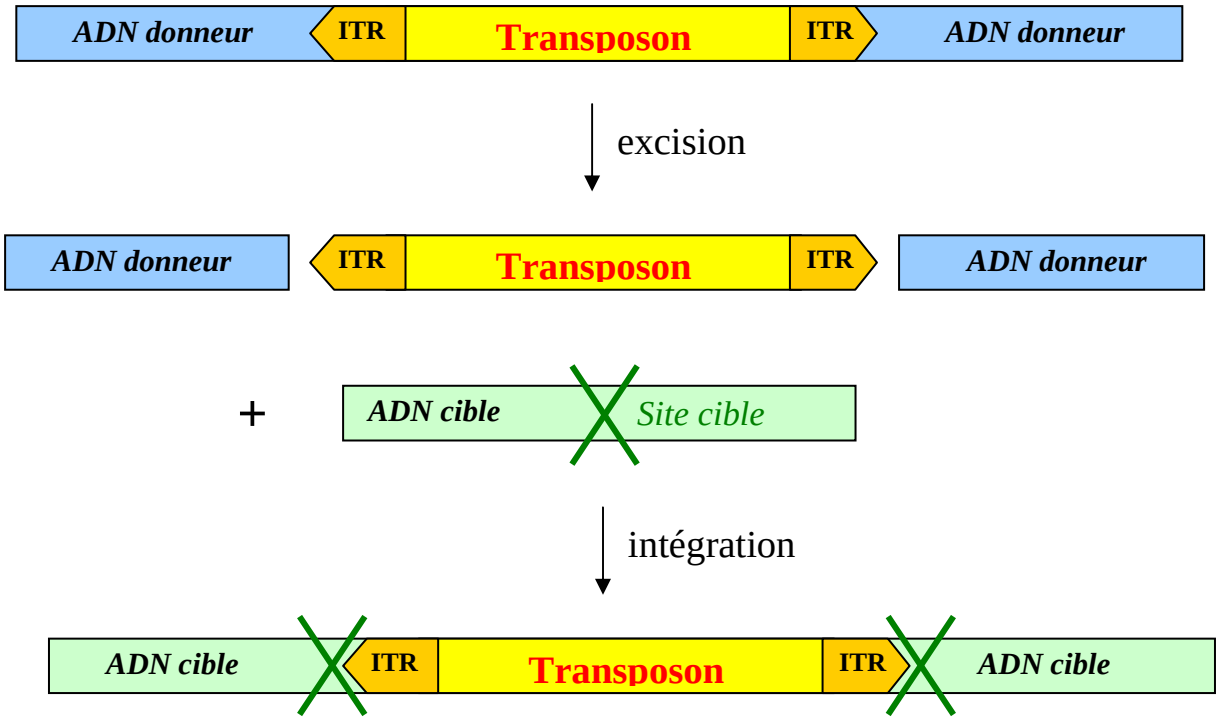


Figure 4. La transposition 'couper-coller' schématisée. *Le transposon est excisé de l'ADN donneur, puis inséré dans l'ADN cible à une séquence précise (X), qui est dupliquée après insertion du transposon.*

4.1.1 La superfamille IS*630-Tc1-mariner*

Les éléments IS*630-Tc1-mariner* (*ITm*) sont les transposons les plus répandus chez les eucaryotes ; ils sont retrouvés chez les champignons, les plantes, et dans toutes les lignées animales, des ciliés aux mammifères en passant par les nématodes, arthropodes et les poissons (Augé-Gouillou *et al*, 1995 ; Plasterk 1999 ; Miskey *et al*, 2005). Cette famille était initialement appelée *Tc1-mariner*, mais elle est maintenant reliée aux IS bactériens, dont l'élément IS*630* qui code une transposase à triade catalytique DDE et s'insère dans les génomes au niveau d'un dinucléotide TA comme les éléments *Tc1-mariner* (Tenzen *et al*, 1990 ; Doak *et al*, 1994).

A partir de quelques copies mises en évidence par les généticiens, grâce aux phénotypes particuliers que leur mobilité confère, des éléments apparentés ont été recherchés dans d'autres génomes par des techniques différentes : PCR dégénérée sur des séquences spécifiques des ITR (Bigot *et al*, 1994 ; Augé-Gouillou *et al*, 1995) ou avec des séquences de motifs protéiques des Tnp (Robertson *et al*, 1993 ; Augé-Gouillou *et al*, 1995), recherche bioinformatique de similitudes de séquences nucléotidiques ou protéiques dans les bases de données (Lam *et al*, 1996 ; Robertson & Asplund, 1996), et parfois par hasard, comme

l'élément *Tes1* chez la myxine (Heierhorst *et al*, 1992). Aujourd'hui, des centaines d'éléments *ITm* différents sont décrits dans la littérature.

Toutefois, dans leur immense majorité les transposons découverts sont inactifs, à cause de mutations dans les gènes qui codent la transposase. Il existe seulement sept transposons naturellement actifs dans cette superfamille : les éléments *Tc1*, *Tc3* et *Tc4* chez *Caenorhabditis elegans* (Moerman & Waterston, 1984 ; Collins *et al*, 1989 ; Moerman *et al*, 1991 ; Yuan *et al*, 1991), *Mos1* et *Minos* chez le genre *Drosophila* (Medhora *et al*, 1988 ; Franz *et al*, 1994), le transposon *Famar1* chez le forficule (Barry *et al*, 2003) et *Mboumar9* chez la fourmi *Messor bouvieri* (Palomeque *et al*, 2006 ; Munoz-Lopez *et al*, 2008).

Les transposons de la superfamille IS*630-Tc1-mariner* partagent les caractéristiques structurales suivantes, avec une taille variable de 1200 à 2300 pb (figure 5) :

- une ORF codant une transposase d'une taille d'environ 350 acides aminés ; cette transposase possède une triade catalytique DDxE/D (x correspondant au nombre d'acides aminés séparant les deux derniers acides aminés de la triade) nécessaire à sa transposition et situé dans sa partie C-terminale (Lohe *et al*, 1997 ; Shao & Tu, 2001). La partie C-terminale de la Tnp et sa triade sont conservées parmi les éléments *ITm* (Shao & Tu, 2001 ; Brillet *et al*, 2007). La partie N-terminale possède un domaine de liaison à l'ADN et elle est peu conservée au sein de la superfamille.

- cette ORF est flanquée par des séquences non traduites appelées UTR (*UnTranslated Regions*), qui pourraient avoir un rôle dans la régulation de la transposition.

- l'ORF et les UTR sont encadrées par les ITR, d'une taille variable de 19 à 300 pb. Ces séquences contiennent un motif de fixation de la transposase.

- les éléments *ITm* ont la particularité de toujours s'intégrer au niveau d'un dinucléotide TA, exception faite de l'élément *TBE1*, qui s'insère au niveau d'un trinucléotide TNA (Plasterk *et al*, 1999). Ces bases sont dupliquées lors de l'insertion de l'élément.

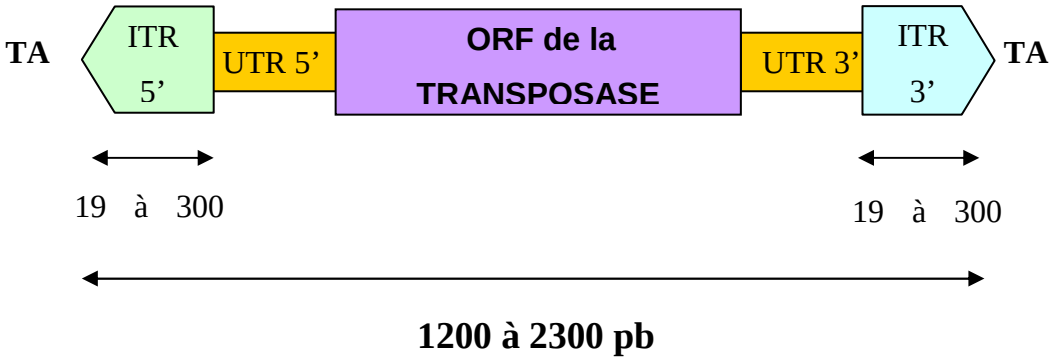


Figure 5. Organisation des éléments IS*630*-Tc*1*-*mariner*.

Les éléments *ITm* sont classés en fonction de leur triade catalytique qui est soit trois acides aspartiques (DDD) soit plus communément deux acides aspartiques et un acide glutamique (DDE). Cette triade est utilisée pour identifier la famille de l’élément (Shao & Tu, 2001) ; on parle alors de DDxD/E, où x correspond à la distance en acide aminé entre le deuxième et le troisième acide aminé impliqué dans la triade. Une analyse phylogénétique de la partie catalytique des Tnp des éléments *ITm* (Brillet *et al*, 2007) a permis de classer ces éléments en trois grands groupes :

- les éléments *ITmD37D* (ou *maT*) et *ITm D37E*, identifiés respectivement chez *Caenorhabditis* et *Bombyx mori* pour les premiers (Claudianos *et al*, 2002) et chez les moustiques et les rotifères pour les seconds (Shao & Tu, 2001 ; Arkhipova & Meselson, 2005). Ces éléments ne seront pas présentés plus en détail par la suite.

- les *TLE*, pour *Tc1-like elements*

- le troisième groupe comprend les *MLE* (pour *mariner-like elements*), les éléments de la famille *Gambol* (*ITmD34E-bis*), identifiés récemment chez les moustiques (Coy & Tu, 2005), ainsi que la famille des éléments *ITmD39D*, isolés seulement dans les génomes des plantes.

Vont maintenant être présentés en détail les éléments *TLE* et *MLE*, qui sont bien caractérisés d’un point de vue biochimique, contrairement aux autres éléments de la superfamille. L’élément *mariner Mos1*, sujet d’étude de ma thèse, appartient aux *MLE* et ses caractéristiques seront présentées dans le sous-chapitre consacré aux *MLE*.

4.1.1.1 Les *TLE* / famille *ItmD34E*

Cette famille, caractérisée par une triade catalytique DD34E, est le plus souvent appelée *TLE* (pour *Tc1-like elements*) à cause du premier élément identifié, *Tc1*, découvert chez *Caenorhabditis elegans* (Emmons *et al*, 1979 et 1983). Ces *TLE* ont depuis été trouvés chez les insectes, les mammifères, les nématodes, les amphibiens et les poissons, les plus connus étant les éléments actifs du ver *C. elegans Tc1* et *Tc3*, et l'élément *Minos*, actif chez *Drosophila Hydei* (Franz *et al*, 1994). A ceux-là s'ajoutent deux éléments 'ressuscités' qui ont été reconstruits moléculairement : *Sleeping Beauty*, qui provient de séquences de poissons (Ivics *et al*, 1997) et *Frog Prince*, reconstruit à partir de séquences d'amphibiens (Miskey *et al*, 2003). Ces deux éléments sont très utilisés comme vecteurs de transfert de gènes (Miskey *et al*, 2005).

La taille des ITR des *TLE* est très variable, allant de 54 pb pour *Tc1* à une taille exceptionnellement longue de 462 pb pour l'élément *Tc3* (Collins *et al*, 1989). Ils codent tous une unique transposase, caractérisée en son domaine C-terminal par la triade catalytique DD34E (exception faite de l'élément *Maya* du xénope, qui possède une triade DD36E – Sinzelle *et al*, 2005) et par des motifs protéiques très conservés autour de deux des trois acides aminés de la triade catalytique. On retrouve dans son domaine N-terminal le domaine de liaison bipartite à l'ADN, composé de deux motifs HTH (hélice-tour-hélice) de type PAIRED (Vos & Plasterk, 1994 ; Colloms *et al*, 1994 ; Brillet *et al*, 2007). Ceci a été confirmé par la cristallisation de la partie N-terminale de la Tnp de l'élément *Tc3* (Watkins *et al*, 2003). Les deux HTH sont impliqués dans la liaison à l'ADN ; PAI est un sous-domaine de reconnaissance spécifique de l'ADN, tandis que le sous-domaine RED, incapable de se lier seul à l'ADN, stabilise l'interaction ADN-protéine. Ces deux sous-domaines sont reliés par un crochet AT (*AT-hook*) (figure 6). L'étude cristallographique a confirmé que ces transposases avec un domaine de type PAIRED se fixaient à l'ADN sous forme monomérique.

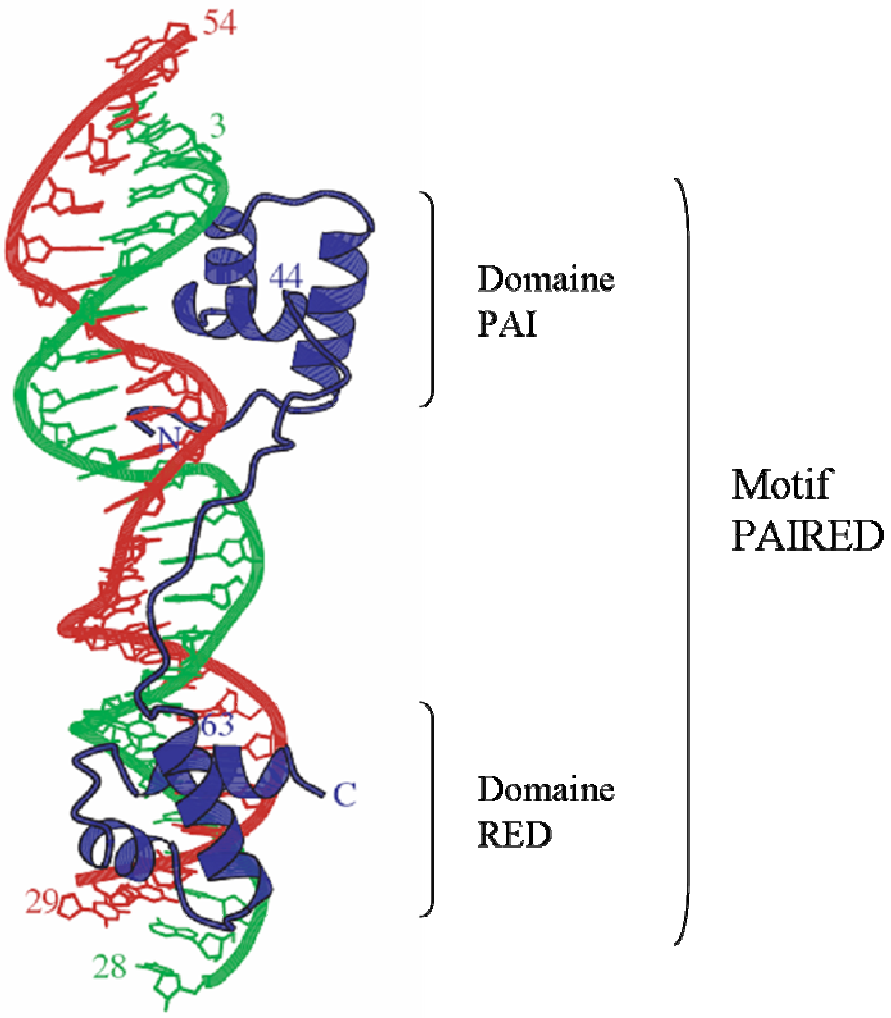


Figure 6. Structure du domaine de liaison à l'ADN de la transposase de Tc*3*, d'après Watkins *et al*, 2003. *La liaison à l'ITR de l'élément Tc3 se fait par les deux motifs HTH représentés en bleu, ces deux motifs, reliés par un linker (en bleu) formant le domaine PAIRED.*

4.1.1.2 Les *MLE* / famille *ItmD34D*

Traditionnellement appelés famille *mariner*, les *MLE* (pour *mariner-like elements*) sont les éléments de classe II les plus répandus chez les eucaryotes (Robertson & Lampe, 1995a et b) et sont retrouvés dans tout le règne animal. Ils sont classiquement divisés en cinq sous-familles, qui sont nommées en fonction de l'espèce de drosophile chez laquelle les premiers éléments ont été découverts : *cecropia*, *elegans/briggsae*, *irritans*, *mauritiana* et *mellifera/capitata*. L'existence d'une nouvelle famille est suggérée avec la découverte de l'élément *Tvmar1* chez *Trichomonas vaginalis*, premier élément *mariner* retrouvé chez une espèce de protiste (Silva *et al*, 2005 ; Bui *et al*, 2008).

Le premier élément *mariner* découvert est le transposon *Dmmar1*, communément appelé *Mos1* (Medhora *et al*, 1988). Cet élément est naturellement actif et autonome chez la drosophile, tout comme l'élément *Famar1* chez le forficule (Barry *et al*, 1994) et l'élément *Mboumar9* caractérisé récemment chez la fourmi *Messor bouvieri* (Palomeque *et al*, 2006 ; Munoz-Lopez *et al*, 2008). L'élément actif *Himar1*, très bien caractérisé biochimiquement, a été reconstruit moléculairement à partir de séquences consensus dérivées d'éléments inactifs présents chez la mouche *Haematobia irritans* (Lampe *et al*, 1996). Il est à noter que des éléments *mariner* ont également été identifiés chez l'homme, mais ils sont inactifs. Il s'agit des éléments *Hsmar1* et *Hsmar2* (Augé-Gouillou *et al*, 1995 ; Oosumi *et al*, 1995 ; Robertson *et al*, 1996 ; Robertson & Martos, 1997).

Les éléments de cette superfamille sont des séquences d'ADN d'environ 1300 pb, avec une seule ORF sans intron codant une Tnp, enzyme nécessaire et suffisante pour promouvoir la transposition (Lampe *et al*, 1996 ; Tosi & Beverley, 2000). Le gène de la Tnp est flanqué des deux UTR et de deux ITR 5' et 3' longs de 28 à 30 pb, sauf exceptions (19 pb pour l'élément *Acmar1* (Sumitani *et al*, 2002), 69 pb pour *Ahmar1* (Yoshiyama *et al*, 2001) et 355 pb pour les ITR du transposon *Mcmar1* (Leroy *et al*, 2003)).

L'élément modèle de cette famille est **l'élément *Mos1***, isolé chez *Drosophila mauritiana* par l'étude d'une mutation dans le gène *white* de la couleur des yeux de la drosophile (Jacobson & Hartl, 1985). Cette mutation donne une couleur pêche aux yeux de la mouche par l'insertion du transposon en aval du promoteur du gène *white*. Il s'agissait en fait d'une copie distincte de *Mos1* qui sera appelée *peach* par la suite. La transposition de cet élément *peach* est en fait non autonome (Capy *et al*, 1990), et requiert la présence d'un autre élément *mariner* qui se trouve être *Mos1* (Medhora *et al*, 1988, Bryan *et al*, 1990). Onze nucléotides différencient les séquences des éléments *peach* et *Mos1*, ce qui correspond au final à quatre acides aminés différents entre les transposases de *peach* (non active) et de *Mos1* (active). La leucine en position 344 de la protéine de l'élément *peach*, contre une phénylalanine chez MOS1, serait impliquée dans la perte d'activité (Maruyama *et al*, 1991 ; Medhora *et al*, 1991).

Cloné à partir du génome de *Drosophila simulans* (Medhora *et al*, 1991), l'élément *mariner Mos1* est une séquence de 1286 pb flanquée de deux UTR et de deux ITR imparfaits de 28 pb ; il code une transposase unique de 345 acides aminés (figure 7). Les UTR de *Mos1* ne sont pas de la même longueur, l'UTR 3' faisant 52 pb, l'UTR 5' 145 pb ; ces UTR

semblent être impliqués dans une activité de régulation de l'activité de transposition (Tosi & Beverley, 2000) mais ne sont pas essentiels à la transposition.

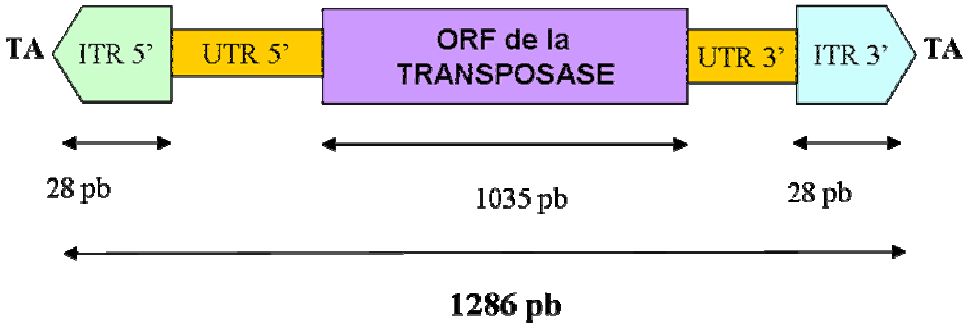


Figure 7. Schéma simplifié de la structure de l'élément *mariner Mos1*. *L'ORF codant la transposase est flanquée de deux régions non traduites (UTR) et de deux répétitions terminales inversées (ITR). L'ITR 5' diffère de l'ITR 3' de 4 pb.*

Les ITR sont dits imparfaits parce qu'ils diffèrent par quatre nucléotides. Cette différence a de l'importance puisque l'ITR 3' est un meilleur substrat pour la transposase que son homologue 5' (Augé-Gouillou *et al*, 2001a ; Pledger & Coates, 2005). Il est intéressant de noter que la construction utilisée dans notre laboratoire comprend le gène de la transposase flanqué par deux ITR 3', ce qui nous donne des fréquences de transposition identiques à l'élément *Mos1* sauvage, soit le gène de la Tnp, des UTR 5' et 3', et les ITR 5' et 3' (Casteret *et al*, 2009).

La **transposase MOS1** est donc une protéine de 345 acides aminés caractérisée par sa triade catalytique DD34D. C'est l'enzyme nécessaire et suffisante pour assurer toutes les étapes de transposition de l'élément *mariner Mos1* (Tosi & Beverley, 2000). La présence de facteurs d'hôte n'a pas été décrite pour cette Tnp, mais a été suggérée à plusieurs reprises pour expliquer la faible activité de transposition de *Mos1* dans les espèces où il est utilisé comme vecteur de transfert de gènes (Lohe & Hartl, 2002 ; Wilson *et al*, 2003 ; Pledger *et al*, 2004). Cette transposase est composée de deux domaines (figure 8) :

- la **partie N-terminale**, domaine de liaison à l'ADN a un motif HTH prédit entre les positions 72 et 110 (Augé-Gouillou *et al*, 2001a ; Pietrokovski & Henikoff, 1997 ; Zhang *et al*, 2001). Deux modèles font actuellement débat parmi les spécialistes des transposons *mariner* :
    - le **premier modèle** propose que le domaine de liaison à l'ADN des éléments *mariner* est un domaine de type CRO, soit un seul HTH impliqué

dans la liaison aux ITR et une dimérisation de la Tnp (Augé-Gouillou *et al*, 2005b ; Brillet *et al*, 2007 ; Feschotte, 2008 ; Pietrokovski & Henikoff, 1997) sur l'ITR. En effet, il est montré que les protéines de type CRO interagissent avec l'ADN sous forme de dimères (Darling *et al*, 2000a et b) et que la cible contient un motif palindromique (Albright & Matthews, 1998). Ces caractéristiques sont décrites pour les éléments *mariner* (interaction sous forme de dimère – Augé-Gouillou *et al*, 2005a et Zhang *et al*, 2001 pour *Mos1* ; Buttler *et al*, 2006 pour *Himar1* ; séquence palindromique – Bigot *et al*, 2005)

- le **second modèle** propose que les *TLE* et *MLE* utilisent le même mode de liaison à l'ADN parce qu'ils sont de la même superfamille. Est alors émise l'hypothèse de l'existence d'un second motif HTH, créant ainsi un domaine de liaison à l'ADN appelé RED-PAI, appelé ainsi car il serait disposé en sens inverse du domaine PAIRED des éléments *Tc1-like* (Richardson *et al*, 2006). Dans ce cas, la liaison ITR-Tnp ne nécessite la présence que d'un monomère de Tnp.

- Cette controverse entre les tenants de l'un ou l'autre modèle n'est pas que théorique puisque le mode de liaison de la transposase conditionne la structure du complexe synaptique.

- la **partie C-terminale** est le second domaine. Il contient la triade catalytique DD34D caractéristique des éléments *mariner* et les séquences conservées YSDLAP et WVPHEL. Cette dernière séquence, positionnée entre les domaines N-terminal et C-terminal, joue un rôle dans la formation du transpososome (Augé-Gouillou *et al*, 2005b). La triade catalytique des éléments *mariner* a la particularité de posséder un troisième acide aspartique à la place du classique acide glutamique observé chez les autres transposases à DDE et les intégrases rétrovirales. Cet acide aminé joue un rôle essentiel dans l'activité de l'enzyme, puisque si on le remplace par un acide glutamique (pour retrouver une triade plus classique DDE), la protéine devient inactive (Lohe *et al*, 1997). La partie C-terminale a été cristallisée (figure 9) en utilisant un mutant fonctionnel (T216A) plus soluble que la forme sauvage de la protéine (Richardson *et al*, 2006). Il est à noter que la protéine entière n'a pas pu être cristallisée du fait d'une activité d'auto-protéolyse propre à

la transposase de *Mos1*. L'analyse cristallographique révèle une structure catalytique de type RNaseH, comme les intégrases rétrovirales et les transposases déjà cristallisées, comme celle de l'élément *Hermes*. Lors des réactions de transposition, la triade catalytique DDD fixe un cation divalent, manganèse ou magnésium (Tosi & Beverley, 2000 ; Richardson *et al*, 2006).

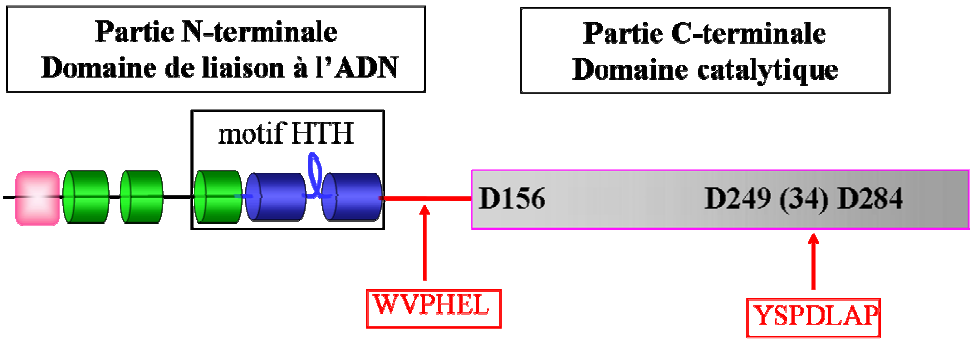


Figure 8. Organisation schématisée de la transposase MOS1. *Les hélices α sont représentées en vert et bleu, le feuillet β en rose. La partie N-terminale est le domaine de liaison à l'ADN et la partie C-terminale contient la triade catalytique DDE. Les motifs conservés des éléments mariner WVPHEL et YSPDLAP sont en rouge.*

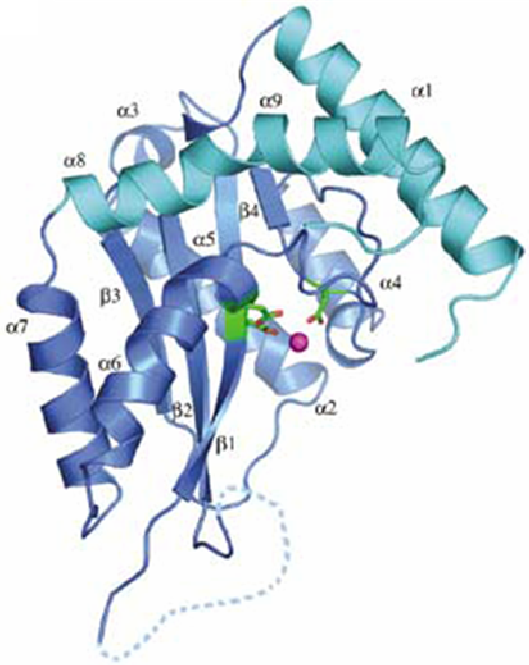


Figure 9. Structure cristallographique de la transposase MOS1, d'après Richardson *et al*, 2006. *Dans cette représentation, le domaine RNaseH est en bleu marine, les hélices N- et C-terminales en cyan. Le cation divalent est représenté en violet.*

L'élément *mariner Mos1*, modèle d'étude de notre laboratoire, est relativement bien caractérisé biochimiquement (Augé-Gouillou *et al*, 2005a et b ; Richardson *et al*, 2006 ; Bouchet *et al*, 2009 ; Carpentier *et al*, 2009). Il peut servir de modèle pour étudier les enzymes à triade catalytique DDE (communément appelées enzymes à DDE), comme les intégrases rétrovirales et nombre de transposases. Le mécanisme de transposition de *Mos1* sera présenté dans le chapitre consacré au mécanisme de transposition des enzymes à DDE. Il est également à noter que l'un des objectifs du laboratoire a été d'utiliser *Mos1* comme un vecteur de transfert de gènes en cellules de mammifères, une alternative aux vecteurs adénoviraux ou rétroviraux ; *Mos1* entre ici en compétition avec d'autres transposons, les plus connus étant l'élément *TLE Sleeping Beauty*, l'élément *Tol2*, et *piggyBac*. *Mos1* est par ailleurs un très bon vecteur pour effectuer du transfert de gènes chez les invertébrés (Robert & Bessereau, 2007).

4.1.2 La superfamille *hAT*

La superfamille *hAT* tire son nom de trois éléments transposables que sont l'élément *hobo*, identifié chez *Drosophila melanogaster* (McGinnis *et al*, 1983), l'élément *Activator* (ou *Ac*) chez le maïs, qui est le premier ET découvert (McClintock, 1948), et *Tam3* de la gueule-de-loup (Coen *et al*, 1986). L'origine commune de ces trois éléments a été mise à jour par l'analyse des séquences des Tnp de ces éléments (Calvi, 1991 ; Feldmar & Kunze, 1991 ; Rubin *et al*, 2001). L'étude d'Eitan Rubin (2001) retrace l'évolution des éléments *hAT* et leur conservation dans les trois grands règnes. Très répandus chez les plantes mais peu trouvés chez les animaux, les éléments *hAT* se transmettent de manière verticale à travers les grands règnes du vivant. Cela signifierait que les transposons *hAT* sont très anciens, existant avant la séparation des plantes, animaux et champignons, puisqu'il y a une absence totale d'éléments de transferts horizontaux entre ces trois règnes.

Les transposons *hAT* possèdent des ITR assez courts qui flanquent le gène codant une transposase à DDE de 500 et 800 acides aminés. Le site cible commun de ces éléments, qui est dupliqué après insertion, a une taille de 8 pb.

Outre les éléments *hobo*, *Ac* et *Tam3*, deux autres éléments 'célèbres' sont retrouvés dans cette superfamille :

- *Tol2*, un transposon de poisson qui est le seul ET de vertébrés à être naturellement actif (Koga *et al*, 1996 et 2000). Cet élément est très utilisé par les Japonais comme outil de mutagenèse (Kawakami *et al*, 1999 et 2004).

- l'élément *Hermes*, issu du génome de la mouche *Musca domestica*, possède la particularité d'avoir une transposase presque entièrement cristallisée (Hickman *et al*, 2005). Transposon d'une taille de 2,7 kb, l'élément *Hermes* a des ITR de 17 pb et code une Tnp de 612 acides aminés. L'assemblage d'un hexamère de transposases semble nécessaire à la transposition. Son domaine catalytique possède une conformation similaire à celle que l'on retrouve chez les intégrases rétrovirales et chez la transposase de l'élément bactérien Tn*5*.

L'étude récente du mécanisme de transposition de l'élément *Hermes* (Zhou *et al*, 2004) a permis de lier la transposition des éléments *hAT* à la recombinaison V(D)J, catalysée par les protéines RAG, les deux mécanismes présentant des similitudes. Un autre aspect qui rapproche les transposases *hAT* et RAG est la distance entre le second D et le E de la triade catalytique ; contrairement aux autres enzymes à DDE, ces deux acides aminés sont assez distants l'un de l'autre : 324 acides aminés pour *Hermes*, 254 pour RAG. Malgré le fait qu'aucun élément *hAT* n'ait été montré comme naturellement actif dans le génome humain, l'étude de la mécanique de ces transposases présente un grand intérêt. Ainsi au moins 26 gènes humains dérivent d'éléments *hAT* (Lander *et al*, 2001), certains étant des facteurs de transcription ou étant impliqués dans des maladies génétiques (Sinzelle *et al*, 2009).

4.1.3 La superfamille *Merlin*

Ces éléments ont été découverts par une étude d'analyse des génomes (Feschotte, 2004) et ne sont retrouvés que chez les animaux et les eubactéries. Il s'agit des premiers ET à ADN décrits chez *Schistosoma mansoni* et des reliques sont retrouvés dans le génome humain. Cédric Feschotte décrit une activité récente de ces transposons chez les vers et chez *S. mansoni* où il y aurait peut-être encore des évènements de transposition autonome.

Ces éléments ont des ITR d'une vingtaine à plus de 460 pb ; ils possèdent une transposase potentielle de 300 acides aminés, à triade catalytique DDE, qui présente de fortes similitudes avec celles du groupe IS*1016*. Feschotte (2004) décrit la superfamille

Merlin/IS*1016* comme un assemblage de transposons eucaryotes et procaryotes ayant un ancêtre commun.

4.1.4 La superfamille *Transib*

Découverts lors d'une analyse du génome de *Drosophila melanogaster* (Kapitonov & Jurka, 2003), les éléments *Transib* ont été nommés ainsi en référence au Trans-Sibérien qui relie Moscou à l'Océan Pacifique ! Ces éléments n'ont été décrits que chez les drosophiles, où ils partagent une séquence consensus qui varie entre 2000 et 3000 pb, avec des ITR de 40 à 45 pb, et chez l'anophèle, où l'élément est plus grand, 4 kb, avec des ITR de 9 pb (Kapitonov & Jurka, 2003 et 2005). La transposase potentielle codée par ces éléments est d'une taille de 450 à 700 acides aminés avec une triade catalytique DDE. Tous ces éléments sont inactifs, ne codant pas une transposase fonctionnelle. Le mécanisme de transposition est supposé être 'couper-coller' (Kapitonov & Jurka, 2005).

Le premier élément *Transib* intact et actif a très récemment été identifié chez *Helicoverpa zea*. Nommé *Hztransib*, il possède une ORF codant une transposase de 507 acides aminés (Chen & Li, 2008).

Il est maintenant acquis que la protéine RAG, impliquée dans la recombinaison V(D)J et élément clé du système immunitaire chez les mammifères, dérive des éléments *Transib* (Kapitonov & Jurka, 2005 ; Hemmrich *et al*, 2007 ; Volff, 2006 ; Sinzelle *et al*, 2009). Cette protéine et son mécanisme d'action seront présentés dans le chapitre consacré au mécanisme de transposition des enzymes à DDE.

4.1.5 La superfamille *Mutator*

Présents chez les champignons, les plantes et les bactéries (Lisch, 2002), les éléments *Mutator* ont été récemment découverts chez les protozoaires (Pritham *et al*, 2005). Il s'agit du transposon le plus actif et donc le plus mutagène chez les plantes ; sa fréquence de transposition peut atteindre les 100% chez le maïs, soit un évènement de transposition par génération (Alleman & Freeling, 1986).

L'élément modèle de la famille est le transposon *MuDR*. Trouvé chez le maïs, il est requis pour la transposition des éléments *Mutator* non autonomes. Contrairement aux transposons précédemment décrits, cet élément contient deux gènes (*mudrA* et *mudrB*), un

épissage alternatif faisant que chaque gène peut coder plusieurs protéines. Le transcrit le mieux connu de *mudrA* code une transposase de 120 kDa appelée MURA ; ce transcrit, ainsi que les autres issus de *mudrA*, codent des transposases ayant de nombreuses similitudes avec des Tnp procaryotes (Eisen *et al*, 1994). Ces Tnp ont un motif catalytique de type DD34E et sont indispensable à l'activité des éléments *Mutator*. Le gène *mudrB* code quant à lui une protéine de 23 kDa (MURB) ; cette protéine, à priori présente uniquement chez le maïs, serait requise pour l'intégration du transposon (Lisch *et al*, 1995 ; Raizada & Walbot, 2000). Chez le maïs où MURB est présente, la transposition est élevée. Dans les autres génomes où MURB est absente, la transposition des éléments *Mutator* existe mais est assez faible.

Ces éléments *Mutator* sont utilisés pour le séquençage et comme outil de mutagenèse chez le maïs (Walbot, 2000). A noter également la découverte chez le riz d'un élément *Mutator* nommé *Anaconda*, qui a évolué et s'est diversifié en acquérant des gènes de l'hôte (Ohtsu *et al*, 2005).

4.1.6 La superfamille *PIF/Harbinger*

Cette superfamille est la réunion des familles *Harbinger*/IS*5* et *PIF*, pour *P Instability Factor* (Jurka & Kapitonov, 2001).

L'élément *Harbinger* a été identifié chez *Arabidopsis thaliana* (Kapitonov & Jurka, 1999). Il est retrouvé chez les diatomées, les plantes, les insectes, les vers et les vertébrés (Kapitonov & Jurka, 2004). Transposon long de 5,3 kb et possédant des ITR de 25 pb, cet élément contient deux ORF ; la première code une Tnp, HABRT, similaire à celle des éléments bactériens IS*5*, avec une triade catalytique DDE. La seconde code une protéine de liaison à l'ADN, avec un motif spécifique très conservé ; cette protéine est nécessaire pour l'expression de la Tnp de l'élément *Harbinger*.

Les transposons *PIF* ont initialement été découverts chez le maïs (Walker *et al*, 1997). L'élément modèle *PIF*, long de 3,7 kb, code une Tnp de 331 acides aminés avec un motif catalytique DDE. La Tnp est là aussi similaire à celle des éléments IS*5*. Ces éléments PIF ont été domestiqués chez la drosophile pour former de nouveaux gènes (Casola *et al*, 2007).

Une étude récente a par ailleurs montré que les éléments *Harbinger* sont actifs en cellules humaines (Sinzelle *et al*, 2008).

4.1.7 La superfamille *piggyBac*

Retrouvés chez les champignons, les insectes, les crustacés, les oursins, et les vertébrés (Sarkar *et al*, 2003), les éléments *piggyBac* ont une séquence consensus de 2,5 kb avec des ITR de 13 pb ; ils dupliquent une séquence spécifique de quatre nucléotides TTAA après insertion (Fraser *et al*, 1996). L'élément type de la superfamille, *piggyBac*, a été identifié chez le lépidoptère *Trichoplusia ni* (Fraser *et al*, 1985).

Les éléments *piggyBac* codent une transposase de 64 kDa (Elick *et al*, 1996), récemment identifiée comme appartenant aux enzymes à DDE (Mitra *et al*, 2008). Le mécanisme de l'élément *piggyBac* était jusqu'alors très peu étudié. Une étude récente a montré que *piggyBac* transpose *via* une coupure double-brin de l'ADN (Mitra *et al*, 2008). La signature d'insertion est une séquence spécifique de quatre nucléotides TTAA, qui n'est pas dupliquée après insertion par synthèse d'ADN par les protéines de l'hôte comme c'est classiquement le cas pour les transposons à ADN de type 'couper-coller'. C'est là toute l'originalité de la transposition de *piggyBac*, un processus sans synthèse d'ADN au niveau du site donneur après l'excision tout comme au niveau du site receveur après l'insertion de l'élément. Ceci est la conséquence de la formation originale d'une structure en épingle à cheveux, communément appelée *hairpin*, aux extrémités du transposon avec la séquence cible TTAA (figure 10). Une fois résolue, l'*hairpin* libère ces quatre nucléotides flanquants qui sont transposés et insérés avec le transposon dans le génome, prévenant ainsi une synthèse d'ADN. C'est la première fois qu'une *hairpin* du côté du transposon, impactant sur la transposition, est observé chez les transposons eucaryotes. Ce phénomène est par contre bien connu chez les éléments bactériens Tn*5* et Tn*10*.

Dans la recherche de transposons pour le transfert de gènes et la thérapie génique non-virale, l'élément *piggyBac*, avec *Sleeping Beauty*, est en première ligne (Feschotte, 2006 ; Wilson *et al*, 2007).

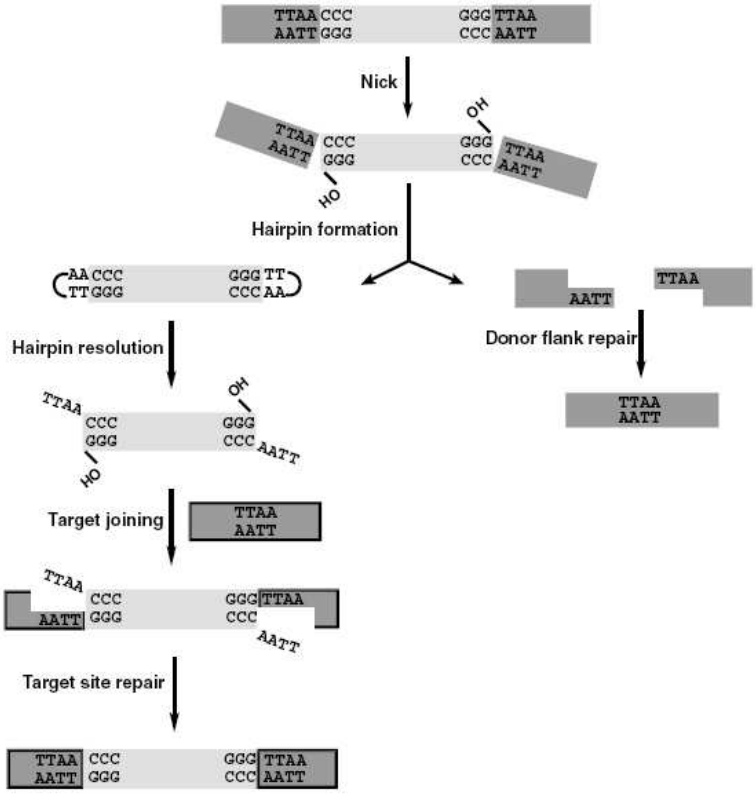


Figure 10. Schéma simplifié de la transposition de *piggyBac*, d'après Mitra *et al*, 2008. *La coupure du premier brin aux extrémités 3' de l'élément entraîne la libération d'un groupe 3'-OH ; ce dernier attaque ensuite l'autre brin d'ADN à quatre bases dans l'ADN flanquant. Se forme alors une hairpin avec ces quatre bases TTAA, tandis que le site donneur se répare par ligation, reformant ainsi la séquence TTAA. La Tnp résout l'hairpin de façon à libérer une extrémité 5' protubérante de quatre nucléotides (TTAA). Le groupe 3'-OH libre attaque la séquence cible TTAA en 5'. Il y a insertion et duplication du site cible TTAA sans synthèse d'ADN puisque l'ADN complémentaire est amené par le transposon.*

4.1.8 L'élément *P*

L'élément *P*, initialement découvert chez *Drosophila melanogaster* (Bingham *et al*, 1982), a été trouvé depuis chez d'autres insectes. Des séquences homologues à *P* existent dans le génome des vertébrés (Hagemman & Pinsker, 2001).

Le transposon *P* autonome et actif est d'une longueur de 2,9 kb, avec des ITR de 31 pb et un seul gène codant une transposase dite à DDED. Ce gène comporte quatre exons qui sont transcrits dans les lignées germinales et somatiques de manières différentes. Dans la ligné germinale, les quatre exons sont transcrits en un ARNm qui code une transposase de 751 acides aminés. Dans les cellules somatiques, seuls les deux premiers introns sont épissés, ce qui conduit à la production du *répresseur P*, protéine de 576 acides aminés, qui inhibe la transposition (Misra *et al*, 1990). La Tnp possède un motif de liaison au GTP qui agit en tant

que cofacteur de l'assemblage du complexe ADN-protéine (cf le chapitre consacré à la régulation de la transposition).

4.1.9 Les éléments CACTA

Transposons classiquement retrouvés chez les plantes, les transposons CACTA se nomment ainsi en raison de la conservation du motif CACTA aux extrémités des ITR (sauf exception avec parfois le motif CACTG). Ils ont été récemment identifiés chez les animaux et les champignons mais avec un motif CCC à la place de l'habituel CACTA, constituant de nouveaux éléments pour la mutagenèse insertionnelle chez les platyhelminthes (Demarco *et al*, 2006).

Les éléments CACTA sont principalement étudiés chez le riz (Langdon *et al*, 2003), mais le premier élément, nommé *En/Spm*, a été découvert chez le maïs (Peterson, 1953 ; McClintock, 1954). D'une longueur de 8,3 kb avec des ITR de 13 pb, ils codent plusieurs transcrits en raison d'un épissage alternatif (Masson *et al*, 1989). Deux de ces protéines, TNPA et TNPD sont essentielles pour l'excision du transposon (Frey *et al*, 1990) : TNPA active le promoteur méthylé du transposon tout en réprimant le promoteur actif non-méthylé (Schläppi *et al*, 1994 ; Cui *et al*, 2002), tandis que la protéine TNPD interagit avec TNPA en stabilisant la liaison à l'ADN (Raina *et al*, 1998). En 2003, Thomas Wicker a mis en évidence chez l'élément *Caspar* la présence d'un deuxième ORF codant pour une protéine à ce jour non identifiée (Wicker *et al*, 2003).

4.1.10 L'ordre des éléments Cryptons

Les éléments *Cryptons* sont un ordre à eux-seuls puisqu'ils diffèrent des autres éléments à ADN en ne possédant pas d'ITR. Ils transposent cependant *via* un intermédiaire à ADN avec une coupure double-brin, et ont une signature de transposition avec duplication du site cible après leur insertion. Retrouvés pour le moment uniquement chez les champignons, ces éléments ont la particularité de coder une tyrosine-recombinase, et s'apparentent ainsi aux rétroéléments DIRS (Goodwin *et al*, 2003 et 2004).

4.2 Transposons à mécanisme 'copier-coller'

Ces transposons à ADN ont la particularité de transposer *via* un mécanisme réplicatif avec le déplacement de seulement un brin d'ADN. Ils ne possèdent vraisemblablement pas

d'ancêtres communs avec les autres transposons à ADN, mais le fait qu'ils ne transposent pas *via* un intermédiaire à ARN les place dans la classe II des ET (figure 2).

Ils sont répartis en deux ordres, comportant chacun une seule superfamille éponyme, les *hélitrons* et les *polintons*. Leur mécanisme de transposition sera brièvement décrit ici. Ces éléments ne seront plus évoqués dans la suite de mon mémoire.

4.2.1 Les *hélitrons* : transposition par cercle-roulant

Découverts chez *Arabidopsis thaliana* et *Caenorhabditis elegans*, ces éléments sont également trouvés chez les plantes, les champignons, les arthropodes, et les vertébrés où ils sont appelés *helentrons* (Kapitonov & Jurka, 2001 et 2007 ; Poulter *et al*, 2003). Ils composent quelques 2% du génome du ver *C. elegans* et jusqu'à 3% de certaines espèces de chauve-souris (Pritham & Feschotte, 2007).

Ces ET sont de taille assez variable mais toujours bien plus grands que les autres éléments de classe II. Leur taille est le plus souvent supérieure à 15 kb et ils ne possèdent pas d'ITR. La caractéristique des *hélitrons* est la présence d'un dinucléotide TC en 5' et de quatre nucléotides CTRR en *hairpin* en 3' de la séquence du transposon. Ce sont des éléments autonomes avec plusieurs ORF, codant notamment une hélicase et une protéine RPA-*like* (Kapitonov & Jurka, 2001). L'absence de duplication au site d'insertion ainsi que l'homologie de séquence entre les transposases putatives des *hélitrons* et les transposases d'éléments bactériens transposant *via* un mécanisme de cercle roulant avaient conduit Kapitonov et Jurka à proposer un modèle de transposition réplicatif en cercle-roulant pour ces *hélitrons* (figure 11). Une étude très récente (Li & Dooner, 2009) montre que les hélitrons peuvent s'exciser de leurs sites chromosomiques dans l'ADN des cellules somatiques, suggérant ainsi que les hélitrons seraient capables de transposer soit par un mode réplicatif, soit par un mode conservatif, comme le transposon bactérien Tn*7* (Craig, 2002).

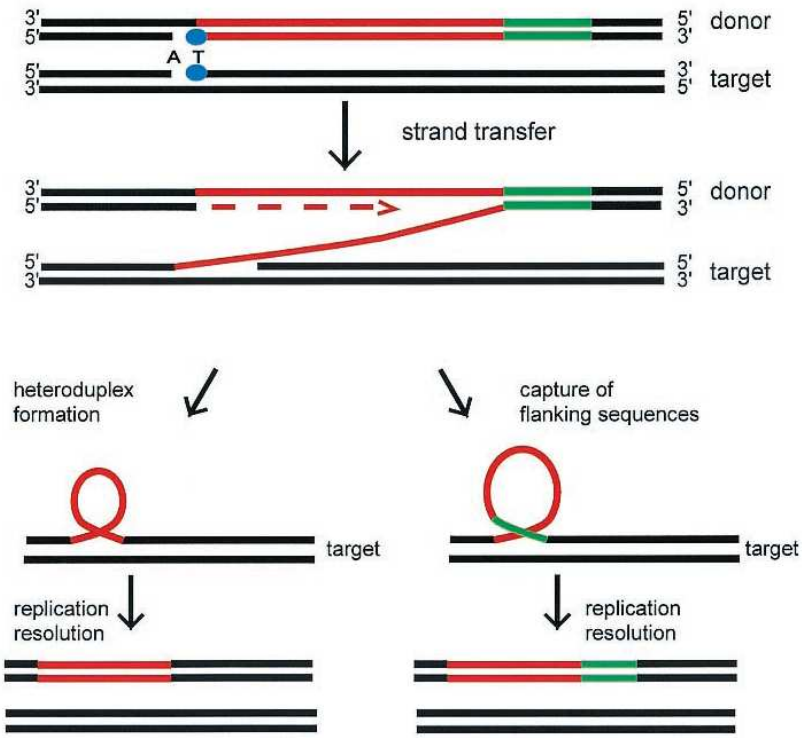


Figure 11. Mécanisme de transposition en cercle-roulant des *hélitrons*, proposé par Kapitonov, 2006. *Deux Tnp (en bleu) clivent un brin à l'extrémité de l'élément (représenté en rouge) et l'ADN cible. La réplication s'initie sur l'ADN donneur permettant le déplacement d'un brin de l'élément qui est à son tour répliqué. A gauche, l'hélitron a transposé seul, il y a eu clivage après la séquence CTRR. A droite, l'hélitron a transposé avec de l'ADN flanquant ce qui explique l'acquisition de nouvelles séquences par les hélitrons.*

4.2.2 Les *polintons/Mavericks* : des transposons complexes

En 2005, Cédric Feschotte et Ellen Pritham présentent les éléments géants *Mavericks*. Ceux-ci codent es c-intégrases retrouvées chez les nématodes et le poisson-zèbre et ils sont caractérisés par de longs ITR (Feschotte & Pritham, 2005). En 2006, Kapitonov et Jurka montrent que ces nouveaux transposons, qu'ils nomment *polintons*, sont des éléments auto-réplicatifs présents chez les protistes, les champignons et dans le règne animal, mais absents des génomes de plantes (Kapitonov & Jurka, 2006).

Il s'agit d'éléments d'une taille de 15 à 20 kb, avec des ITR de plusieurs centaines de pb, des duplications spécifiques au site d'insertion de cinq ou six pb, et codant jusqu'à dix protéines. Parmi elles, il y a toujours une intégrase de type rétrovirale (c-intégrase) et de cinq à neuf protéines dont une ADN polymérase B, une protéase à cystéine de type adénovirale, une ATPase ou encore une protéine putative de capside (figure 12).

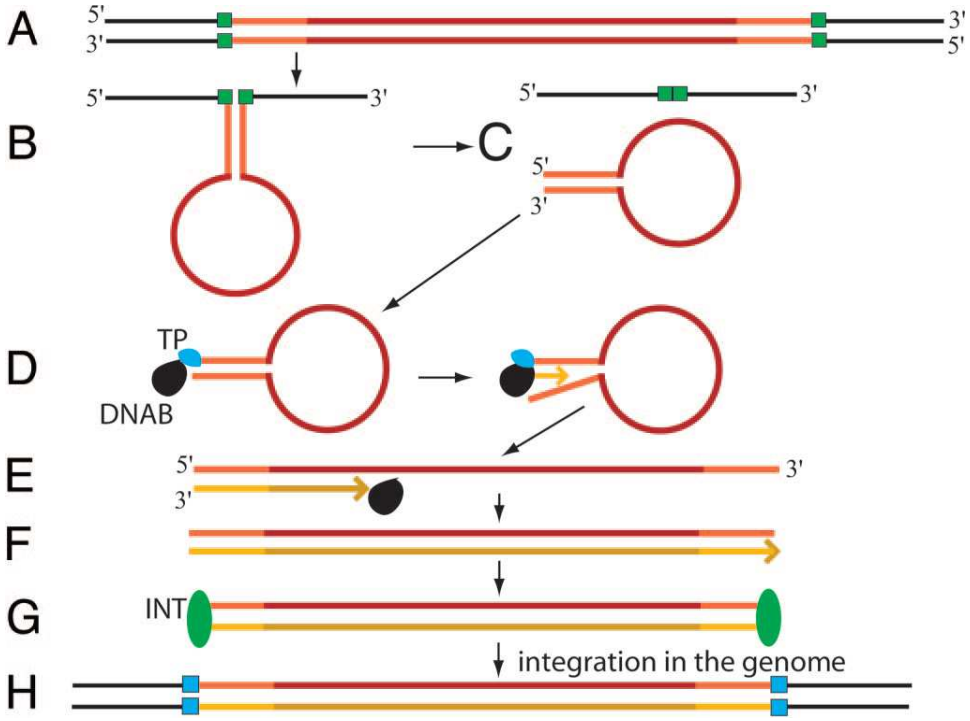


Figure 12. Mécanisme de transposition des *polintons*, proposé par Kapitonov & Jurka, 2006. *Un seul brin du transposon est excisé après avoir adopté une structure dite en 'raquette' (A à C). La polymérase B (DNAB) réplique ensuite l'élément extra-chromosomique avec l'aide de la protéine terminale (TP) qui initie la réplication en se fixant en partie 5' du polinton simple brin (D à F). Une fois l'élément sous forme double brin, l'intégrase (INT) codée par le polinton se fixe aux extrémités du transposon et l'intègre dans le génome de l'hôte (G et H).*

Dans la dernière publication consacrée à ces éléments, les auteurs proposent que ces éléments soient un maillon évolutif entre les éléments invasifs de l'ADN (bactériophages, adénovirus) et les plasmides linéaires eucaryotes (Pritham *et al*, 2007).

5/ Mécanismes de transposition

Ma thèse s'articulant autour du mécanisme de transposition de l'élément *mariner Mos1*, je vais m'intéresser ici aux différents modes de transposition des transposases à DDE aussi bien procaryotes (Tn*5*, Tn*10* et IS*911*) qu'eucaryotes (les éléments *ITm*), toutes très bien étudiées d'un point de vue de leur mécanique de transposition, ainsi qu'aux autres enzymes à DDE que sont l'intégrase du VIH-1 (Virus de l'Immunodéficience Humaine de type 1) et la protéine RAG impliquée dans la recombinaison V(D)J et dérivée du transposon *Transib*.

Marcin Nowotny (2009) classe toutes ces enzymes dans la grande superfamille des intégrases rétrovirales (RISF) dans laquelle il inclut les RNaseH et les transposases à DDE. Ces protéines RISF possèdent toutes un cœur catalytique de type RNaseH (figure 13) qui est la première enzyme de cette superfamille à DDE à avoir été cristallisée (Katayanagi *et al*, 1990 ; Yang *et al*, 1990).

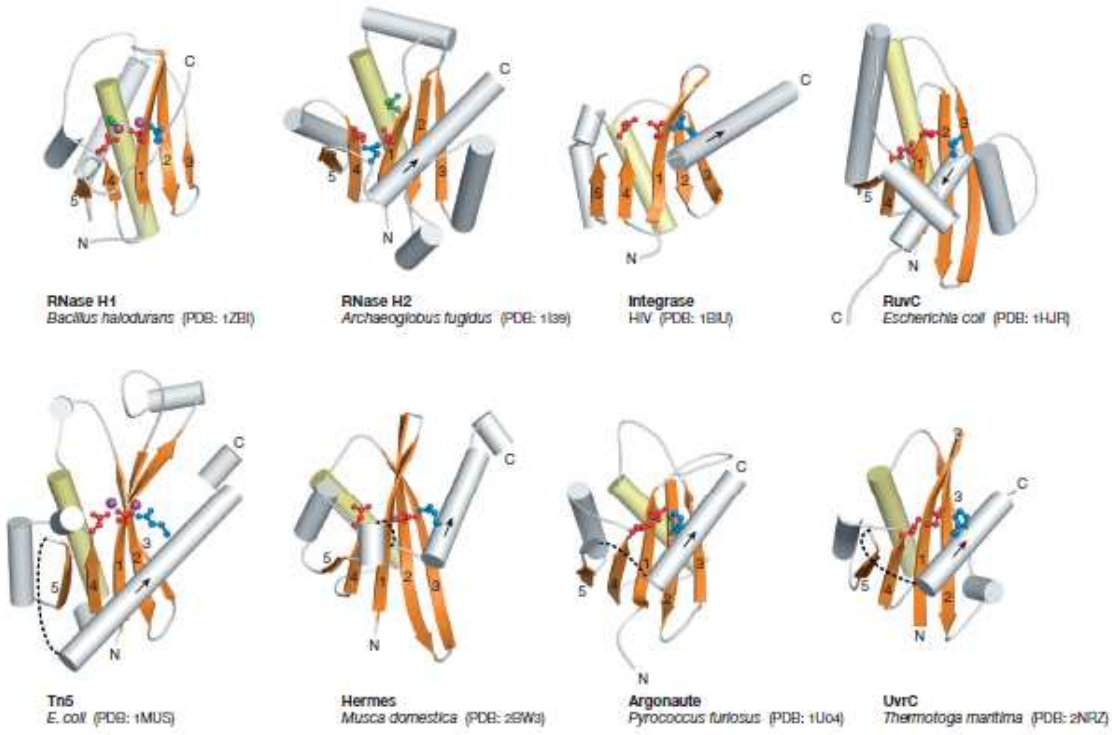


Figure 13. Cœurs catalytiques de quelques enzymes RISF, d'après Nowotny, 2009. *Les feuillets β centraux et l'hélice α, très conservés, sont représentés en orange et jaune. Les parties non conservées sont en gris. La similitude structurale existant entre ces enzymes est ici évidente.*

La triade catalytique DDE (DDD chez certains *ITm* dont les éléments *mariner*) est le deuxième point commun de toutes ces enzymes de la superfamille RISF. Ces acides aminés sont conservés chez toutes ces enzymes, la distance entre les acides aminés variant selon la protéine (figure 14).

VIH-1	**D**$_{64}$ (51) **D**$_{116}$ (35) **E**$_{152}$
ASV	**D**$_{64}$ (56) **D**$_{121}$ (35) **E**$_{157}$
Mos1	**D**$_{156}$ (92) **D**$_{249}$ (34) **D**$_{284}$
Tc*3*	**D**$_{144}$ (86) **D**$_{231}$ (34) **E**$_{266}$
IS*911*	**D**$_{207}$ (59) **D**$_{287}$ (35) **E**$_{396}$
Tn*7* (TnsA)	**D**$_{28}$ (85) **D**$_{114}$ (34) **E**$_{149}$
Tn*5*	**D**$_{97}$ (90) **D**$_{188}$ (136) **E**$_{325}$
Tn*10*	**D**$_{97}$ (63) **D**$_{161}$ (130) **E**$_{292}$
Hermes	**D**$_{180}$ (67) **D**$_{248}$ (323) **E**$_{572}$
RAG1	**D**$_{600}$ (107) **D**$_{708}$ (253) **E**$_{962}$

Figure 14. Le motif DDE est conservé chez transposases et intégrases. *La triade DDxE est le cœur du site catalytique de ces enzymes RISF. VIH-1 : Virus de l'Immunodéficience Humaine de type 1 ; ASV : Virus du Sarcome Aviaire.*

Ces enzymes ont en commun la formation d'un complexe synaptique qui permet une réaction concertée pour la coupure de l'ADN aux deux extrémités du transposon.

Il est intéressant de noter que la distance entre le D et le E est classiquement de 34 ou 35 acides aminés, exception faite des transposases de Tn*5*, Tn*10* et d'*Hermes* ainsi que la protéine RAG1, où la distance est beaucoup plus importante. Cette caractéristique est liée au fait que ces trois protéines catalysent la formation d'une *hairpin* lors du clivage de l'ADN (cf plus loin).

5.1 Les éléments modèles Tn*5* et Tn*10*

5.1.1 Tn*5* : ‘un modèle pour comprendre la transposition’

L’élément bactérien Tn*5* est l’un des premiers transposons à avoir été découvert (Berg *et al*, 1975). Son mécanisme de transposition a depuis été très bien disséqué, si bien que le spécialiste du sujet, William Reznikoff, le présente comme un modèle pour comprendre la transposition au niveau moléculaire (Reznikoff, 2003, 2006 et 2008).

Tn*5* est un transposon composite dans lequel se trouvent deux IS quasi-identiques, IS*50L* et IS*50R*, qui encadrent trois gènes de résistance aux antibiotiques. IS*50R* code une transposase fonctionnelle à cœur catalytique DDE de 477 acides aminés et un inhibiteur de transposition. IS*50L* code des versions tronquées et inactives de la Tnp et de l’inhibiteur. La Tnp de Tn*5* a été cristallisée (Davies *et al*, 2000) en version dimérique sur ses ES (séquences terminales où se fixe la Tnp en prélude à la transposition) (figure 15).

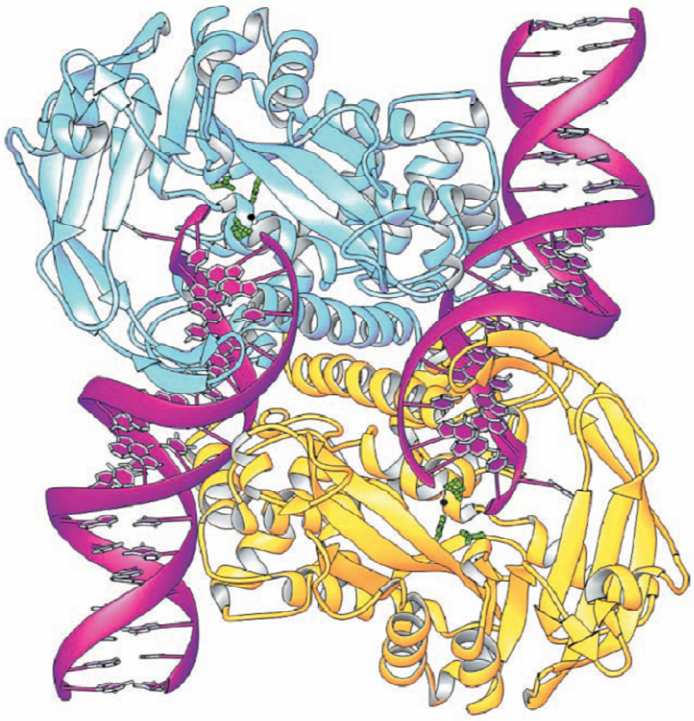


Figure 15. Structure moléculaire du complexe synaptique de Tn*5*, d’après Davies *et al*, 2000. *Les deux transposases sont en jaune et bleu, les deux ES de 20 pb en violet. Les résidus catalytiques sont en vert et les cations divalents* Mn^{2+} *sont représentés par de petites billes noires.*

Le transposon Tn*5*, comme les transposons à ADN eucaryotes présentés plus loin, transpose donc *via* un mécanisme ‘couper-coller’, avec excision de l’élément du site donneur, intégration dans une séquence cible et duplication du site d’insertion de la séquence cible (Reznikoff, 2002 ; figure 16).

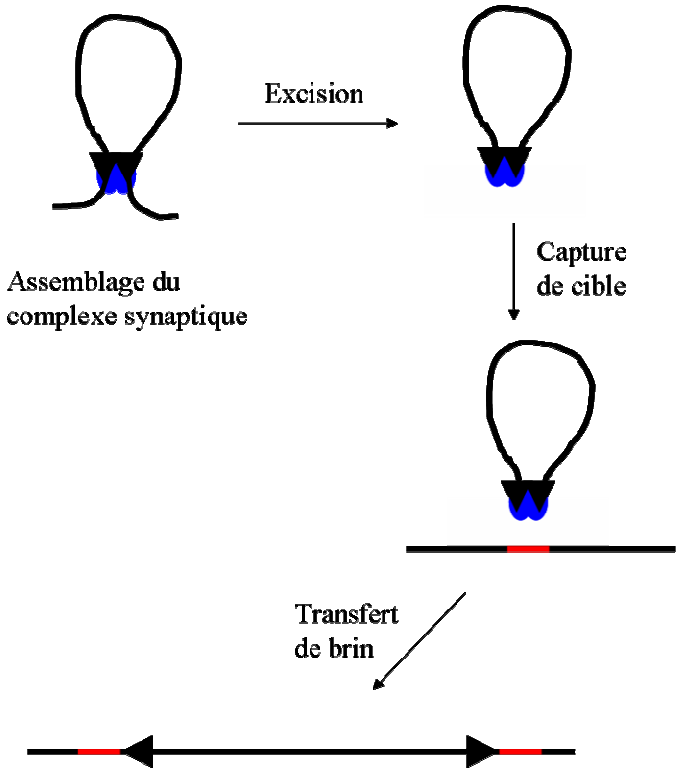


Figure 16. Schéma de transposition de Tn5. *Le complexe synaptique est responsable de l'excision de l'élément ce qui libère ainsi le transpososome. Après capture de cible et insertion, il y a duplication du site cible, ici représenté en rouge.*

La première étape de la transposition est la liaison de la Tnp aux ES du transposon. Le mécanisme de liaison n'est pas encore élucidé parfaitement, l'hypothèse qui prédomine est qu'il y a un couplage entre traduction et fixation de la transposase à l'ES ; il n'y aurait ainsi pas de monomère actif de Tnp en solution. La liaison de deux molécules de Tnp conduit à la formation du complexe synaptique. C'est ce complexe qui promeut l'excision de l'élément de manière séquentielle et dynamique (Steiniger *et al*, 2006 ; Czyz *et al*, 2007). Sur une extrémité, il y a d'abord clivage de la liaison phosphodiester du premier brin avec génération d'un groupement 3'-OH, l'eau jouant le rôle du nucléophile lors de cette réaction. Ce groupe 3'-OH libre attaque à son tour la liaison phosphodiester sur le second brin, formant ainsi une *hairpin (*Bhasin *et al*, 1999*)*. La transposase catalyse ensuite le clivage hydrolytique de cette *hairpin*, libérant ainsi le groupement 3'-OH du brin transféré, prêt pour la réaction de transfert de brin (figure 17-A).

La formation de l'*hairpin* lors de cette réaction permet à un seul monomère de transposase de promouvoir le clivage double brin à une extrémité de l'élément. L'analyse de la formation de l'*hairpin* a révélé qu'un retournement de base avait lieu à l'extrémité de l'ES et était nécessaire pour stabiliser la réaction (Ason & Reznikoff, 2002 ; Bischerour & Chalmers, 2007). La base qui est retournée est située sur le brin non transféré deux bases à l'intérieur de l'extrémité du transposon (en position +2). Cette base pivote et sort de l'hélice

d'ADN pour être fixée par le résidu tryptophane 298 de la transposase ; cette distorsion est nécessaire à l'établissement de l'*hairpin*, ceci permettant au groupe 3'-OH libre d'attaquer le brin non transféré (Guegen *et al*, 2005 ; Bischerour & Chalmers, 2009 ; figure 17-B).

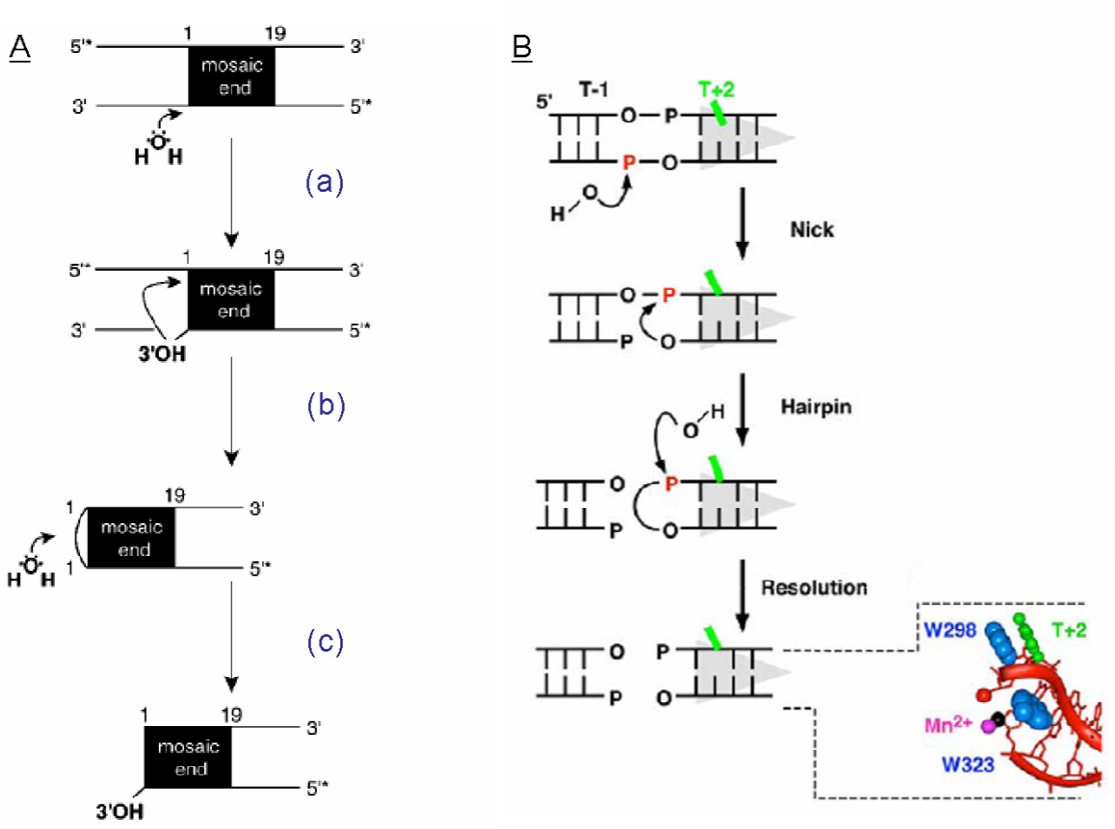


Figure 17. Excision de Tn*5*. (A) Clivage d'une extrémité de l'élément Tn*5*, d'après Bhasin *et al*, 1999. ***(a)*** *L'eau joue le rôle du nucléophile pour le clivage du premier brin, libérant ainsi une extrémité 3'-OH, responsable de l'attaque nucléophile sur le second brin en 5'* ***(b)****, ce qui conduit à la formation de l'hairpin. Cette dernière est résolue par clivage hydrolytique* ***(c)*** *catalysé par la transposase, libérant le 3'-OH, essentiel pour l'étape suivante qu'est le transfert de brin.* **(B) Formation de l'*hairpin* et retournement de base à l'extrémité de Tn*5*, d'après Bischerour & Chalmers, 2009.** *Les phosphates attaqués sont en rouge, l'extrémité du transposon en grisé. La liaison de la transposase créé une distorsion dans l'ADN ce qui déstabilise la base en T+2 (en vert) du brin non transféré. La première coupure qui expose le groupe 3'OH à l'extrémité du transposon provoque le retournement de la base T+2 de l'hélice, ce qui permet la formation de l'hairpin, puis sa résolution. Le petit insert en bas représente la structure de l'extrémité de Tn5, les résidus de la transposase liant l'ADN n'étant pas représentés, exception faite des deux résidus tryptophane. Un des résidus est inséré dans l'hélice d'ADN, tandis que l'autre stabilise la base retournée.*

Ces réactions de clivage dans le complexe synaptique sont séquentielles : clivage du premier, puis du second brin à une extrémité, ce qui donne un produit *Single End Break (SEB)*, puis le même processus a lieu pour la seconde extrémité de l'élément (figure 18), libérant ainsi le transpososome constitué du transposon et d'un dimère de Tnp.

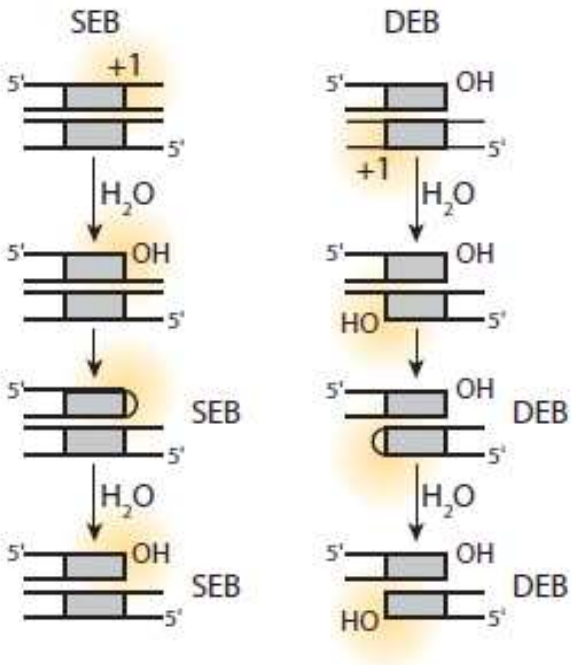


Figure 18. Clivage séquentiel des extrémités de Tn*5*, d'après Reznikoff, *2008*. *(SEB : Single End Break, DEB : Double End Break) La transposase est symbolisée par le rond orangé. La réaction a d'abord lieu à une extrémité avec résolution de l'hairpin. La même réaction à la seconde extrémité permet la libération du transpososome, dimère de Tnp fixé sur les deux extrémités du transposon qui a des extrémités 3'-OH.*

Le transpososome doit ensuite capturer la cible. Pour ce faire, il y a un changement de conformation à l'extrémité 5' du brin non transféré, positionnant le phosphate dans une configuration optimale pour le transfert de brin à suivre (Klenchin *et al*, 2008). Comment un fragment d'ADN est défini comme une cible et comment le transpososome reconnaît sa cible, voilà deux questions qu'il convient de résoudre. Le complexe de capture de cible formé, la dernière étape du processus de transposition peut prendre place. Il s'agit du transfert de brin au cours duquel les deux groupes 3'-OH présents aux extrémités des ES servent de nucléophiles pour la rupture de la liaison phosphodiester au niveau de l'ADN cible. Lors de ces réactions de trans-estérification, il a été montré *in vitro* (Czyz *et al*, 2007) que le transfert de brin des deux extrémités se déroulait l'un à la suite de l'autre. Là aussi une question reste en suspens : ces deux brins sont-ils transférés indépendamment ou s'agit-il d'évènements obligatoirement séquentiels liés par les modifications de conformation requises dans le complexe ?

5.1.2 Tn*10* : un modèle de transposition conservative avec implication de protéines d'hôte

L'élément bactérien Tn*10* est un transposon composite de 9147 pb constitué de deux séquences IS*10* (*left* et *right*) encadrant une séquence d'ADN contenant sept ORF, dont quatre sont impliqués dans la résistance à la tétracycline. La transposase est codée par l'IS*10*-right, l'IS*10-left* étant une copie dégénérée de l'IS*10-right* ne code pas de Tnp (Foster *et al*, 1981). Les extrémités de chaque IS*10* sont nommées OE et IE (pour *outside* et *inside end*) et contiennent les séquences nécessaires à la fixation de la transposase (figure 19). Les deux OE contiennent en plus un site de fixation pour la protéine d'hôte IHF (figure 20).

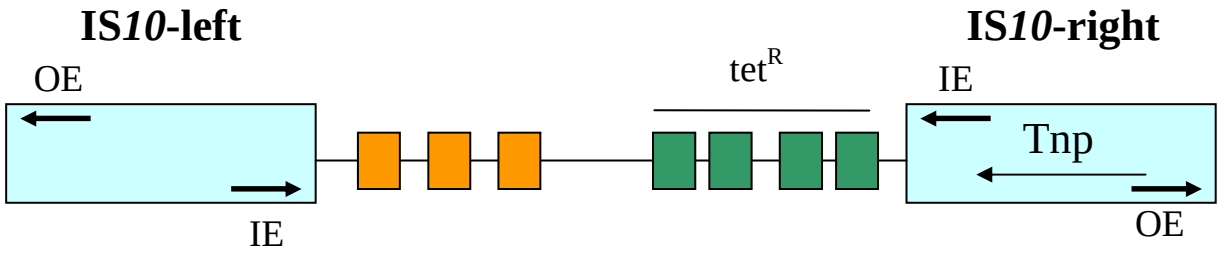


Figure 19. Structure de l'élément Tn*10*. *Les deux IS10 encadrent une séquence non-IS contenant sept ORF (représentés par des carrés) dont quatre sont impliqués dans la résistance à la tétracycline (en vert). L'IS10-right code la transposase (Tnp) et on retrouve à chaque extrémité des IS10 les OE (outside end) et IE (inside end).*

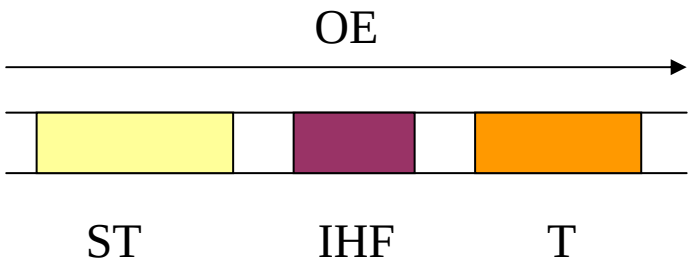


Figure 20. Organisation schématique d'un OE de l'élément Tn*10*. *T est la région terminale où se fixe initialement la transposase ; ST est la région sub-terminale où se fixe la transposase après que l'ADN ait été courbé par la protéine IHF, qui se fixe quant à elle sur la région colorée en violet (cf explications à la suite).*

Sur le plan du mécanisme de transposition au niveau moléculaire, il s'apparente à celui décrit pour l'élément Tn*5* : fixation d'un monomère de transposase aux extrémités du

transposon, dimérisation, excision avec formation et résolution d'une *hairpin*, capture de cible et insertion (figure 21).

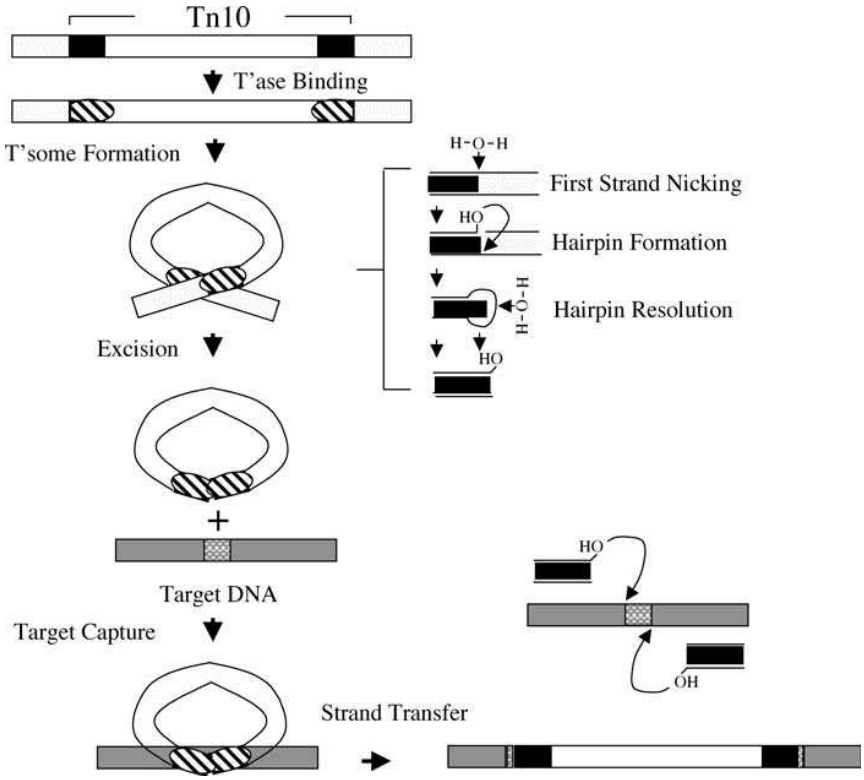


Figure 21. Mécanisme de transposition de Tn*10*, d'après Haniford, 2006. *T'ase pour transposase, T'some pour transpososome. La transposase se fixe sous forme monomérique à chaque extrémité du transposon, puis se dimérise avant l'étape d'excision. Comme pour Tn5, un groupe 3'-OH est généré à l'extrémité du transposon ce qui conduit au clivage du second brin et à la formation d'une hairpin du côté du transposon. L'hairpin est résolue et le complexe synaptique capture la cible. Le transfert de brin est catalysé par la transposase. L'hôte répare le trou de 9 pb créé lors du transfert de brin ce qui conduit à la duplication du site cible (petits rectangles à damiers).*

La particularité du mécanisme de transposition de Tn*10* est qu'il fait intervenir des protéines d'hôte dans le processus. Lors de la formation du complexe synaptique intervient la protéine IHF (pour *Integration Host Factor* – Sakai *et al*, 1995). C'est en utilisant un système expérimental où ne sont présents que les OE du transposon sous forme linéaire (Sakai *et al*, 1995 et 2000) qu'ont été étudiées en détail les conditions de formation du transpososome de l'élément Tn*10*. La protéine IHF intervient en courbant l'ADN de 180°, ce qui permet à la transposase qui s'était initialement liée au domaine T de l'OE de se fixer également au domaine ST (figure 22).

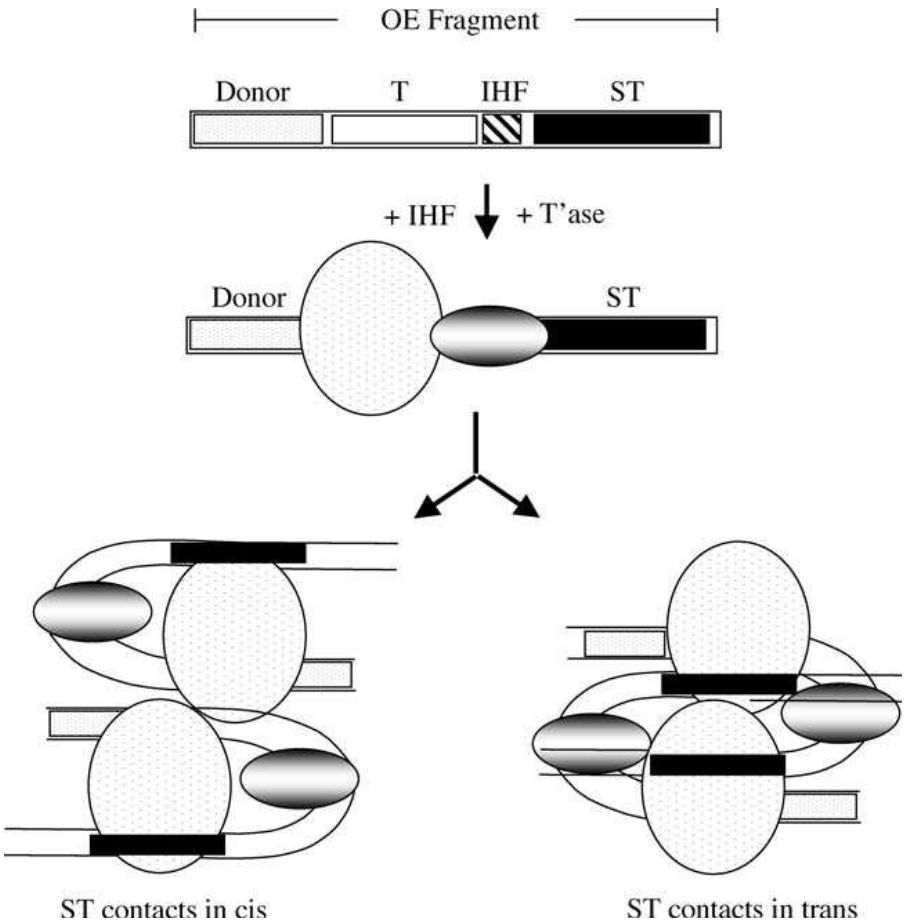


Figure 22. Formation du transpososome de Tn*10* avec IHF, d'après Haniford, 2006. *La Tnp (l'ovale) se fixe sur le domaine T de l'OE tandis que la protéine IHF (ovale) se fixe sur son site cible ; cette dernière courbe alors l'ADN de 180° permettant à la Tnp de se fixer à la région ST de deux manières possibles : soit en cis sur le domaine ST adjacent, soit en trans avec la région ST d'une autre OE.*

Il a également été montré que des protéines d'hôte, IHF et H-NS (pour *Histone-like nucleoid*) intervenaient dans la reconnaissance du site cible (Chalmers *et al*, 1998 ; Wardle *et al*, 2005 ; Singh *et al*, 2008). IHF est impliqué dans la transposition intramoléculaire (le transposon s'insère alors dans lui-même), à cause de la courbure qu'elle impose au transposon, tandis que H-NS promeut la transposition intermoléculaire (Ward *et al*, 2007 ; Singh *et al*, 2008), en agissant en opposition à IHF (figure 23), soit en éjectant IHF du transpososome soit en altérant la courbure de l'ADN induite par la protéine IHF (Wardle *et al*, 2005 ; Haniford, 2006).

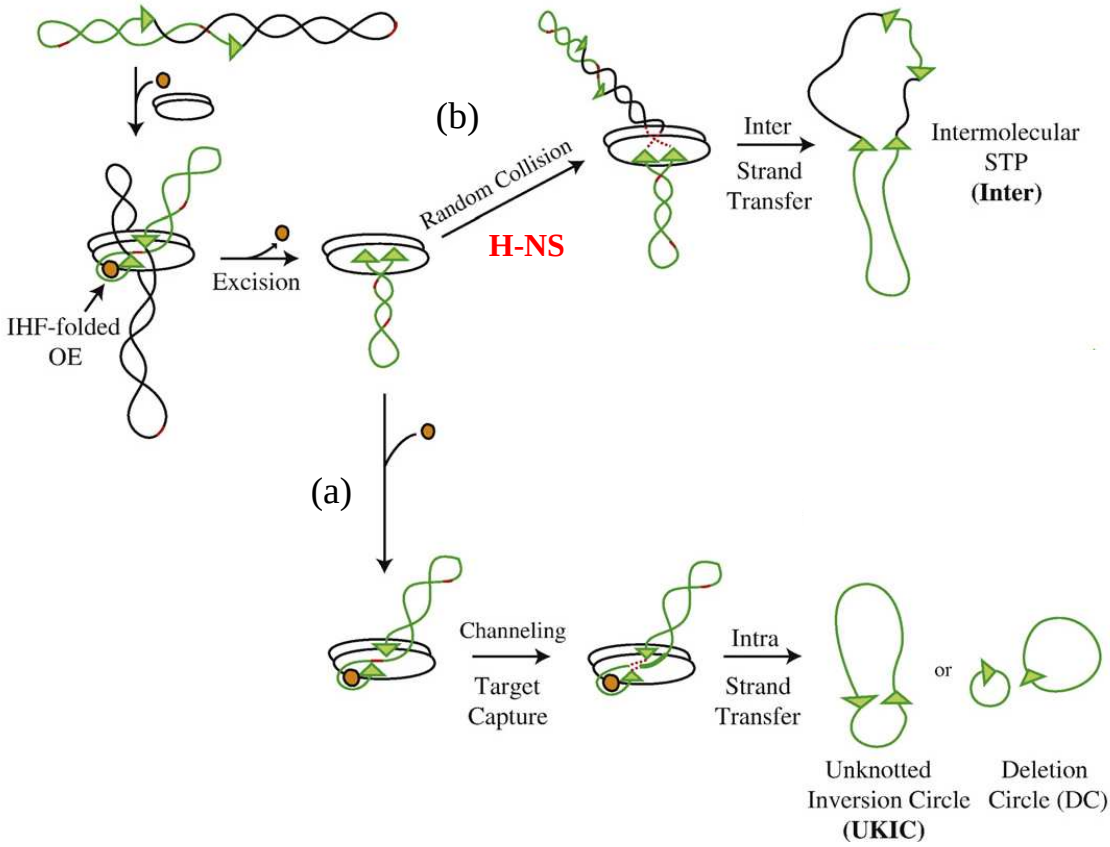


Figure 23. Schéma simplifié de la transposition de Tn*10*, adapté de Singh *et al*, 2008. *Le transposon est représenté en vert, l'ADN en noir. Après les étapes de liaison et d'excision (détaillées précédemment), il y a deux voies possibles. (a) la protéine IHF (en orange) maintient au moins un OE de courbé de 180° après l'excision, ce qui conduit l'insertion à se faire de manière intramoléculaire, dans une zone située près de l'OE (en rouge). Ceci conduit à des produits simples appelés UKIC ou DC. (b) la protéine H-NS agit en opposition de IHF, promouvant ainsi une transposition intermoléculaire et un produit de transfert de brin (STP) appelé ici Inter.*

Ces protéines IHF et H-NS peuvent être considérées comme des facteurs de régulation de la transposition puisque une transposition intramoléculaire est suicidaire pour le transposon.

5.2 IS*911* : transposition avec un intermédiaire circulaire

Appartenant à la famille de transposons bactériens IS*3* et isolé initialement chez *Shigella dysenteriae* (Prère *et al*, 1990), l'élément IS*911*, long de 1250 pb, est flanqué par deux ITR imparfaits de 36 pb appelés IRL et IRR (pour *Inverted Repeat Left* et R*ight*). Il contient deux ORF nommées *orfA* et *orfB* qui sont consécutives dans la séquence. Il exprime trois protéines que sont OrfA, OrfB et sa transposase OrfAB (Polard *et al*, 1991). Cette dernière est une protéine de fusion de 382 acides aminés qui regroupe le domaine de liaison à l'ADN de OrfA et la triade catalytique DD35E de OrfB.

IS*911* possède la particularité de transposer *via* un intermédiaire circulaire avec les deux extrémités de l'élément qui se retrouvent liés de manière covalente (figure 24). La transposase (ou les transposases car la stœchiométrie des complexes n'est pas connue) assemble le complexe synaptique avec les deux extrémités de l'élément, puis induit le clivage d'un premier brin à une extrémité, générant ainsi un groupe 3'-OH. Cette extrémité 3'-OH sert d'attaquant nucléophile sur la seconde extrémité ce qui permet l'établissement d'une liaison covalente simple brin entre les deux extrémités du transposon (Polard & Chandler, 1995). On obtient alors une première forme appelée 'figure en 8'. Ce sont ensuite les protéines cellulaires, et uniquement elles, qui résolvent cette 'figure en 8' pour donner un transposon en cercle qui reconnaît ensuite sa cible et s'insère avec l'aide de la transposase OrfAB.

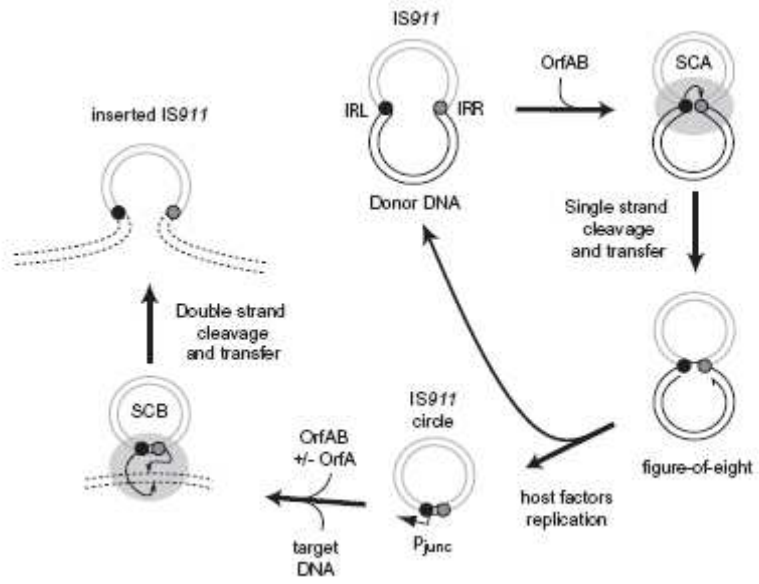


Figure 24. Transposition de IS*911*, d'après Gueguen *et al*, 2006. *Le transposon est ici en gris, l'ADN donneur en noir, et l'ADN cible en pointillés.*

5.3 Les éléments *ITm*

Plusieurs éléments ont été étudiés en détail du point de vue de leur mécanisme de transposition chez les éléments *Tc1-mariner* : les éléments du ver *Tc1* et *Tc3* (Vos & Plasterk, 1994 ; Vos *et al*, 1996 ; Colloms *et al*, 1994 ; Watkins *et al*, 2004), le transposon bactérien *ISY-100* (Feng & Colloms, 2008), celui de la fourmi *Mboumar9* (Munoz-Lopez *et al*, 2008), ainsi que les transposons *mariner Himar1* (Lampe *et al*, 1998 ; Lipkow *et al*, 2004a et b ; Butler *et al*, 2006) et *Mos1* (Augé-Gouillou *et al*, 2001a et b, 2005a et b ; Sinzelle *et al*, 2008). Ce dernier étant le sujet d'intérêt de ma thèse, c'est sur lui que je me concentrerai dans ce chapitre, tout en apportant des précisions sur les autres éléments lorsque le mécanisme diffère.

La transposition peut être divisée en quatre étapes principales (figure 25) : la liaison spécifique de la transposase aux ITR de l'élément, la formation du complexe synaptique, l'excision du transposon, et pour finir la reconnaissance et l'intégration au site cible.

Clivé du site donneur avec des extrémités 3'-OH libres, le transposon est ensuite transféré dans une cible par une réaction de trans-estérification. La transposase est l'unique protéine requise pour promouvoir toute la réaction de transposition, sans intervention de protéines d'hôte (Lampe *et al*, 1996).

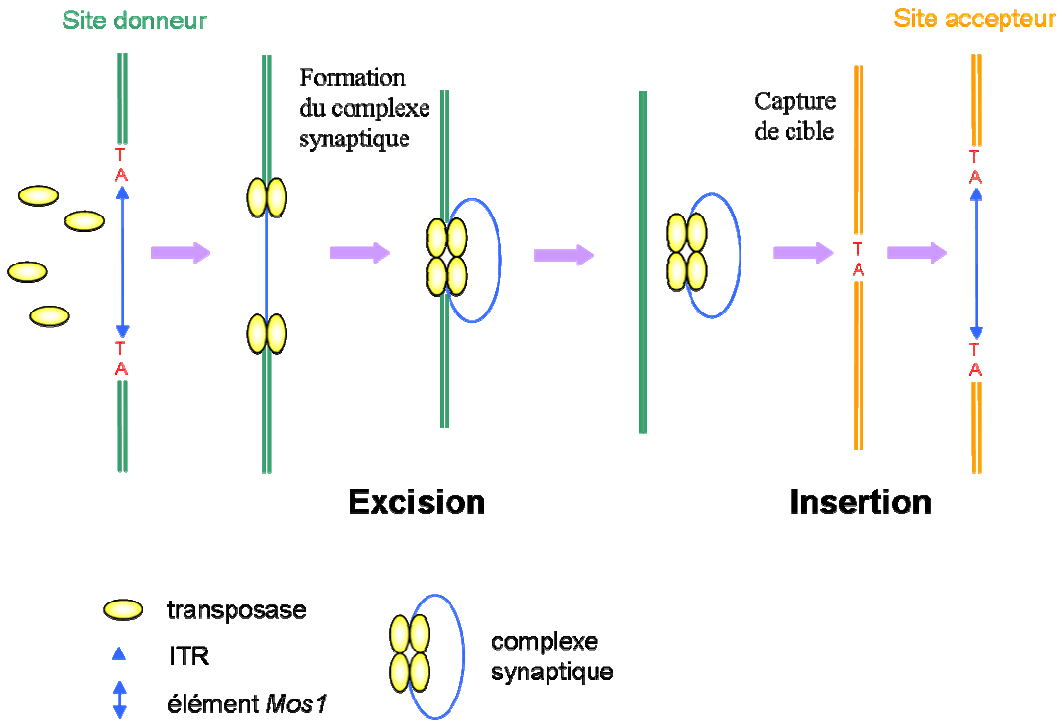


Figure 25. Schéma simplifié de la transposition de l'élément *Mos1*. *Un dimère de transposase se lie à chaque ITR du transposon Mos1, situé dans l'ADN génomique de l'hôte (en vert) ; ceci aboutit à la formation du complexe synaptique dans lequel un tétramère de transposase est lié aux ITR. Ce complexe s'excise de l'ADN, puis reconnaît un site accepteur du génome (en orange) au niveau d'un dinucléotide TA (capture de cible). Ce TA est dupliqué lors de l'insertion du transposon qui s'en suit.*

Rentrons maintenant dans le détail moléculaire de cette réaction (figure 26). La première étape du processus de transposition des éléments *Tc1-mariner* est la **fixation de la Tnp aux ITR** (Vos & Plasterk, 1994 ; Colloms *et al*, 1994 ; Ivics *et al*, 1997 ; Augé-Gouillou *et al*, 2001a ; Feng & Colloms, 2007). Cette liaison à l'ADN est différente selon les éléments. En effet, les ITR ne contiennent pas tous le même nombre de site de liaison pour la transposase ; un seul site de liaison a été identifié chez *Tc1* (Vos *et al*, 1996) contre deux chez *Mos1*, *Tc3* et *Sleeping Beauty* (Augé-Gouillou *et al*, 2005b ; Colloms *et al*, 1994 ; Ivics *et al*, 1997). Chez *Mos1*, la transposase, qui est capable de se dimériser en l'absence d'ADN (Zhang *et al*, 2001) et qui existe sous forme dimérique en solution (Richardson *et al*, 2006), se fixe probablement sur son ITR sous forme dimérique (Augé-Gouillou *et al*, 2005a et b). Cette interaction Tnp-Tnp se fait *via* la partie N-terminale de la protéine en impliquant les 35 premiers acides aminés (Augé-Gouillou *et al*, 2005a ; Butler *et al*, 2006).

Les dimères de transposases liés à chaque ITR se rapprochent pour former un complexe nucléoprotéique, le **complexe synaptique**. C'est une étape clé du processus de

transposition. En l'absence de cristallisation de Tnp fixées sur les ITR, ce sont les études biochimiques qui nous renseignent sur la stœchiométrie du complexe synaptique.

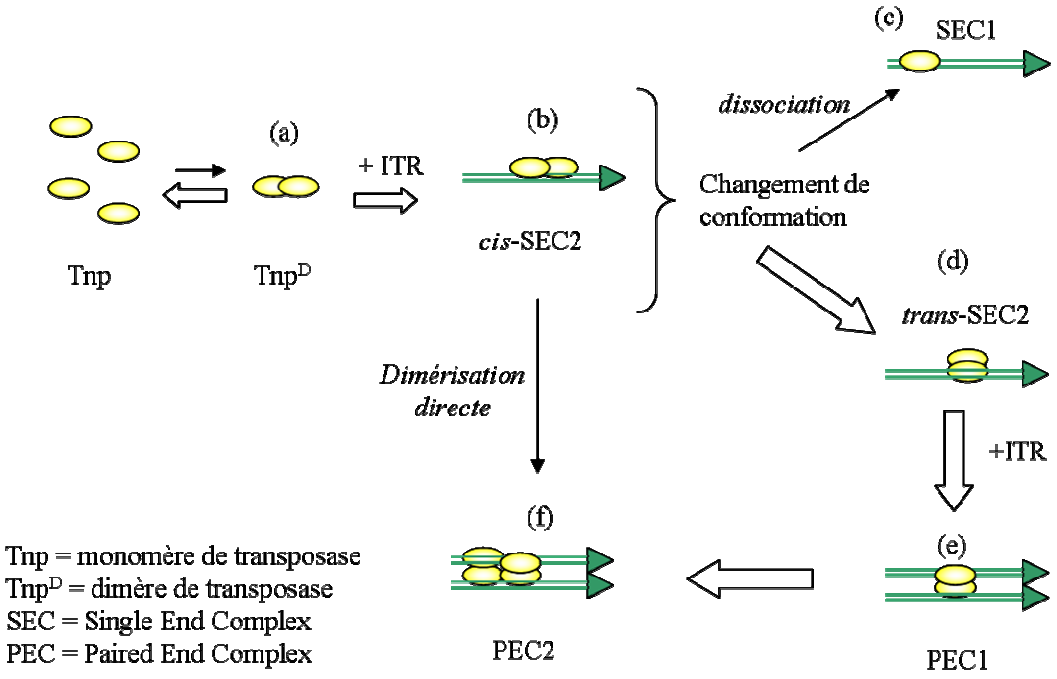


Figure 26. Formation du complexe synaptique de *Mos1*, adaptée d'Augé-Gouillou *et al*, 2005a. *Selon le modèle proposé en 2005, la transposase se dimérise (a), avant de se lier aux ITR (b), pour former un complexe appelé cis-SEC2. Ce complexe pourrait ensuite soit se dissocier, pour donner un SEC1 (c), ou alors changer de conformation pour former le trans-SEC2 (d), qui fixerait alors le deuxième ITR pour former le complexe PEC1 (e) ; la fixation d'un second dimère de Tnp donnerait le PEC2 (f), qui est le complexe synaptique capable de promouvoir l'excision. Ce PEC2 pourrait également être la conséquence d'une dimérisation directe du cis-SEC2.*

Pour *Sleeping Beauty* et *Tc3*, ce complexe serait aussi formé de deux ITR et d'un tétramère de Tnp dont la formation serait la conséquence d'une interaction deux à deux entre quatre monomères de Tnp (Izsvak *et al*, 2002 ; Watkins *et al*, 2004).

Pour *Mos1*, dimérisation directe ou passage par des complexes intermédiaires, la formation du PEC2 (*Paired End Complex*) reste sujette à question. De plus, dans ces deux hypothèses se pose la question de savoir comment se rejoignent les deux ITR de l'élément : scanning orienté du transposon par le complexe ITR-Tnp ou interaction au hasard entre les deux complexes ITR-Tnp ? La littérature est riche d'exemples indiquant que la fréquence de transposition diminue avec l'augmentation de la taille du fragment d'ADN compris entre les deux ITR, tant pour les éléments bactériens que les éléments eucaryotes comme les éléments

ITm Mos1 (Sinzelle *et al*, 2008) et *Himar1* (Lampe *et al*, 1998). Ces données font pencher la balance en faveur de l'hypothèse d'un scanning orienté (Brillet *et al*, 2007).

Une fois le complexe synaptique formé se déroule alors l'étape primordiale du processus de transposition, **l'excision de l'élément**. Cette étape est catalysée par des cations divalents (Vos *et al*, 1996 ; Lampe *et al*, 1998 ; Haren *et al*, 1999 ; le rôle des cations est expliqué plus loin). De la qualité de ce clivage dépend la suite du processus de transposition ; il doit être précis pour permettre la réintégration sinon il conduit à une excision abortive qui est délétère pour la transposition de l'élément. Il a été montré que chez *Mos1* l'excision qui autorise la réintégration s'effectue exactement sur le brin non transféré (NTS) du transposon et trois bases à l'intérieur de l'élément sur le brin transféré (TS) (Dawson et Finnegan, 2003 ; Julien Bischerour, communication personnelle ; figure 27).



Figure 27. Clivage des extrémités du transposon *Mos1*. *L'excision qui permet par la suite l'intégration du transposon s'effectue à l'extrémité de l'élément en 5' et trois bases à l'intérieur de l'élément en 3'. L'ADN flanquant est en noir, le TA en bleu, le transposon est en rouge foncé ; NTS : brin non transféré, TS : brin transféré.*

La coupure est loin d'être systématiquement précise chez les éléments *mariner* (figure 28). De nombreuses données *in vitro* et *in vivo* sont disponibles sur le clivage des éléments *Mos1*, *Himar1* et *Hsmar1*, et montrent que la transposase coupe tout autour et dans le TA à l'extrémité de l'élément (Lampe *et al*, 1996 ; Dawson & Finnegan, 2003 ; Augé-Gouillou *et al*, 2005b ; Miskey *et al*, 2007 ; Sinzelle *et al*, 2008 ; Julien Bischerour, communication personnelle).

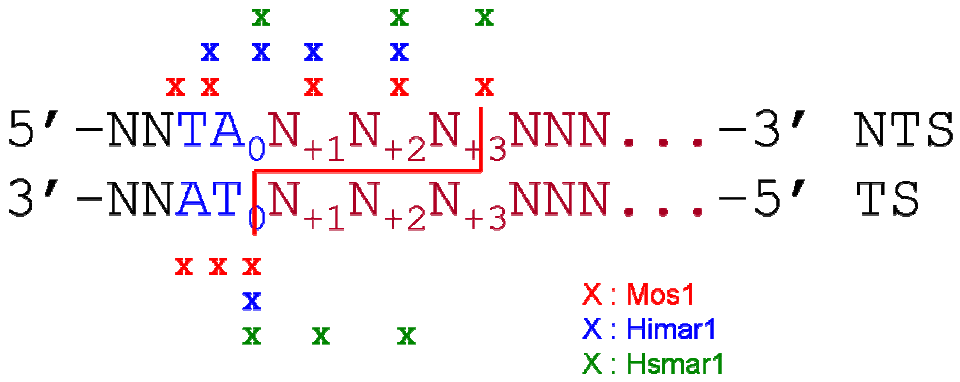


Figure 28. Coupure imprécise chez les éléments *mariner*. *Le trait rouge représente la coupure précise de Mos1 à une extrémité ; les X représentent les sites de coupure observés pour trois éléments mariner, d'après les données citées dans le texte.*

Chez *Mos1*, la coupure est tellement imprécise que cette étape devient limitante pour l'efficacité de transposition ; des résultats obtenus *in vitro* montrent que moins de 6% de la coupure du premier brin est précise. Pour le second brin, la transposase coupe de manière erronée à plus de 96% un ITR apporté sous une forme pré-clivée sur son premier brin (Julien Bischerour, communication personnelle). Moins de 6% sur le premier brin, moins de 4% sur le second, cela nous donne moins de 0,24% de coupures correctes à une extrémité. Or il y a la seconde extrémité du transposon à cliver, ce qui donnerait moins de 0,05% de coupures précises sur les deux ITR ! La réalité *in vitro* n'est évidemment pas la réalité *in vivo*, car cela ne permettrait pas le maintien du transposon chez son hôte. Pour étudier la coupure du second brin, le substrat donné à la transposase est un ITR déjà clivé sur le premier brin ; or, dans le contexte biologique, la transposase est déjà fixée sur l'ITR lors de la coupure du premier brin. Ceci conduit certainement à l'adoption d'une conformation particulière du complexe synaptique, l'hypothèse étant qu'une coupure précise sur le premier brin induit systématiquement une coupure précise sur le second brin grâce au positionnement du complexe sur l'ADN.

Deux différences majeures sont constatées entre *Mos1* et les éléments bactériens Tn*5* et Tn*10* présentés précédemment. Chez *Mos1*, l'ordre de coupure n'est pas le même et il y a absence de formation d'une *hairpin*. Contrairement aux éléments Tn*5* et Tn*10*, la Tnp de *Mos1* clive en premier le brin non transféré puis hydrolyse le second brin (Dawson & Finnegan, 2003 ; figure 29). La coupure du second brin ne nécessite pas ici la présence d'un groupe 3'-OH, ce qui implique que ce clivage s'effectue sans la formation d'une *hairpin*. La formation d'une *hairpin* lors de l'étape de clivage existe néanmoins chez d'autres éléments

eucaryotes. C'est le cas pour l'ADN flanquant d'*Hermes* (Zhou *et al*, 2004), dans le cadre de la recombinaison V(D)J catalysée par les protéines RAG (cf partie consacrée à RAG), ainsi qu'aux extrémités du transposon pour l'élément *piggyBac* (Mitra *et al*, 2008), ce qui n'était jusqu'alors pas retrouvé chez les ET eucaryotes. L'ordre des clivages fait que *Mos1* ne transpose pas de manière réplicative ne provoquant ainsi pas de réarrangement de l'ADN.

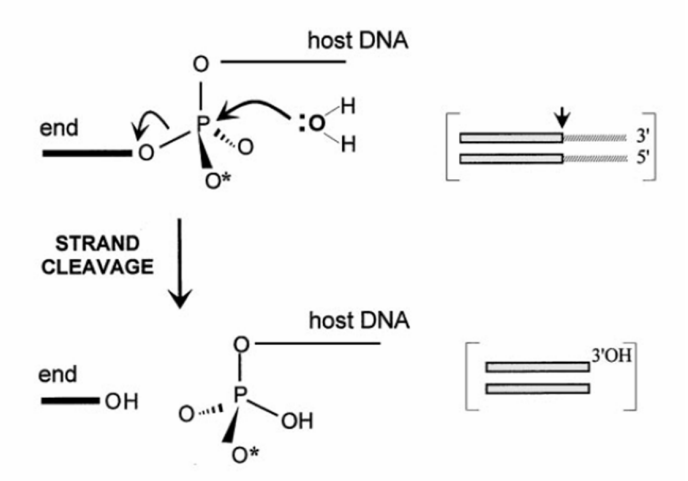


Figure 29. Chimie du clivage du brin transféré, d'après Haren *et al*, 1999. *L'eau joue le rôle du nucléophile pour l'attaque de la liaison phosphodiester représentée ici sous forme chirale. L'hydrolyse conduit à une inversion de la chiralité.*

Le clivage double brin n'est donc pas net chez les éléments *mariner*, et laisse derrière lui une cicatrice d'excision de trois bases chez les *MLE* (Bryan *et al*, 1990 ; Lampe *et al*, 1996 ; Luo *et al*, 1998), de deux bases chez les *TLE* (van Luenen *et al*, 1994 ; Vos *et al*, 1996). La cicatrice est ensuite réparée par les enzymes de l'hôte *via* les recombinaisons HR (*Homologous Recombinaison*) ou NHEJ (*Non Homologous End Joining*). Ces coupures non franches qui laissent derrière elles des cicatrices d'excision ne sont pas la propriété des éléments *ITm*, une cicatrice de 17 nucléotides est par exemple trouvé chez l'élément *P* (Rio, 2002).

La triade catalytique DDE joue un rôle primordial dans les réactions de trans-estérification qu'elle promeut, en organisant le positionnement dans l'espace des deux cations métalliques divalents nécessaires à la réaction, ceci dans le cas des transposases à DDE comme celui des intégrases à DDE (Haren *et al*, 1999 ; Yang *et al*, 2006). Les cations classiquement utilisés par les enzymes à DDE sont Mg^{2+} et Mn^{2+}, le premier étant préférentiellement utilisé *in vivo* (Delelis *et al*, 2008). L'un des cations joue le rôle d'une base de Lewis, augmentant la charge négative du nucléophile par déprotonation ; le second, coordonné par les deux acides aspartiques, joue le rôle d'un acide de Lewis et stabilise l'état de transition pentacovalent autour du phosphate. Quand la nouvelle liaison phosphodiester est

formée, la configuration du phosphate est inversé et le groupe 3'-OH qui reste est reprotoné (figure 30).

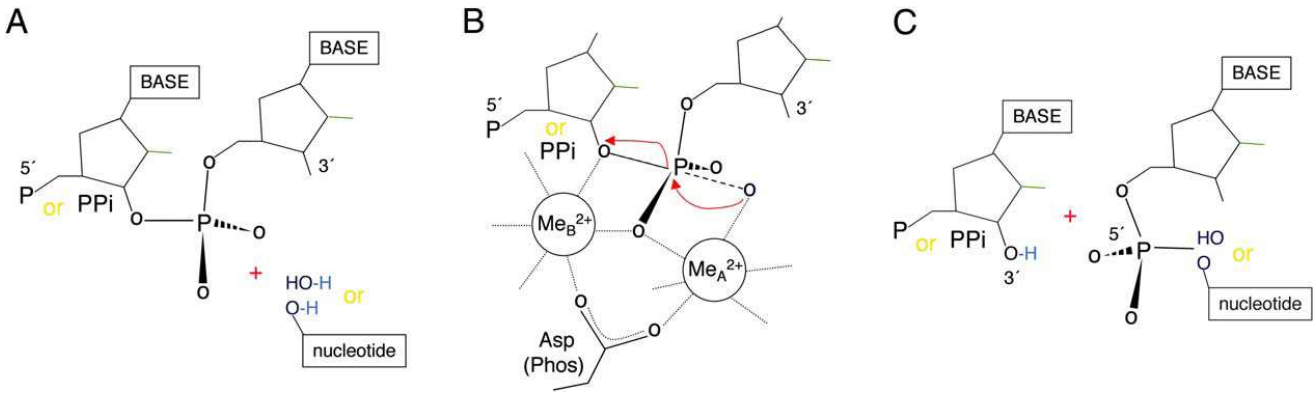


Figure 30. Rôle des cations divalents lors du clivage, d'après Yang *et al*, 2006. (A) *L'eau (dans le cas du clivage du brin transféré chez Mos1) ou le groupe hydroxyle d'un sucre (avec le H en bleu) doit être déprotoné pour devenir nucléophile.* **(B)** *Les deux cations métalliques (Me^{2+}) sont coordonnés par des oxygènes et l'acide aspartique pour stabiliser l'état de transition pentacovalent.* **(C)** *Une nouvelle liaison phosphodiester est formée, inversant la conformation du phosphate, et le groupe 3'-OH restant est reprotoné.*

Le complexe ADN-Tnp libéré après l'excision est appelé le **transpososome** ou **intermédiaire de transposition**. Ce complexe correspond à la forme excisée du complexe synaptique. Bien que peu étudié chez les éléments *Tc1-mariner*, le transpososome est un élément clé du processus de transposition car il va permettre l'intégration de l'élément, d'abord en reconnaissant puis capturant sa cible, avant l'étape proprement dite d'intégration. Pour les éléments *Tc1* et *Tc3*, des éléments extra-chromosomiques ont été détectés sous forme circulaire et linéaire chez le ver *Caenorhabditis elegans*, l'hôte naturel de ces transposons (Rose & Snutch, 1984 ; Ruan & Emmons, 1984 ; van Luenen *et al*, 1993). Rien n'a toutefois montré que ces éléments étaient capables de s'intégrer dans le génome. Il est intéressant de noter ici que le transpososome n'a donc pas formellement été identifié chez les *MLE* et *TLE* ; l'hypothèse qu'il n'y ait pas d'intermédiaire de transposition, mais un couplage des réactions d'excision et d'intégration ne peut pas être exclue en l'état actuel des connaissances sur la mécanique de transposition des éléments *ITm*.

La **reconnaissance et la capture de la cible** sont effectuées par le complexe ADN-Tnp par des mécanismes non élucidés. Les éléments *ITm* s'intègrent quasiment toujours au niveau d'un dinucléotide TA, avec quelques exceptions (comme TNA pour l'élément *TBE1* ; Doak *et al*, 1994). Souvent présentés comme des éléments pouvant s'intégrer de manière aléatoire dans le génome, les *ITm* présentent tout de même une certaine spécificité d'insertion.

En effet, tous les TA ne sont pas des sites cible. Une spécificité d'insertion existe aussi en fonction de la région dans laquelle ils se trouvent. Chez *Tc1*, il a été rapporté que certains TA contenus dans une région particulière de 1000 pb du gène *gpa-2* étaient plus fréquemment utilisés que d'autres (van Luenen & Plasterk, 1994 ; Vos *et al*, 1996). Pour *Mos1*, des *hot spots* d'intégration ont été observés dans le génome de la drosophile, son hôte naturel (Lohe & Hartl, 1996). Plus récemment, il a été publié que lors de tests de transposition *in vitro* et en bactérie, le gène de résistance au chloramphénicol (*cat*) était un *hot spot* d'insertion pour le transposon *Mos1* (Crénès *et al*, 2009). Contrairement aux éléments *TLE*, les auteurs montrent ici que la courbure de l'ADN n'a pas d'impact sur la préférence d'intégration. Par contre au sein du gène *cat* il y a une spécificité d'insertion importante puisque le transposon s'intègre dans des séquences TATA ou TANTA (N représentant n'importe quel nucléotide). *Mboumar9*, qui ne présente pas de biais d'intégration *in vitro*, n'est pourtant détecté que dans l'ADN satellite du génome de son hôte naturel, la fourmi *Messor bouvieri* (Munoz-Lopez *et al.*, 2008). Les *TLE* ont la particularité de s'intégrer dans un site proche du site d'excision (Luo *et al*, 1998 ; Fisher *et al*, 1999), et le *TLE Sleeping Beauty* s'intègre dans une région riche en AT, l'hypothèse étant que cela faciliterait la relaxation de l'ADN pendant l'intégration de l'élément (Vigdal *et al*, 2002).

Une fois la cible reconnue et capturée par le complexe ADN-Tnp, le groupe 3'- OH exposé à l'extrémité de l'élément joue le rôle du nucléophile pour attaquer la liaison phosphodiester de l'ADN cible lors d'une réaction de trans-estérification en une étape (Mizuuchi, 1992 ; Haren *et al*, 1999 ; Braam & Reznikoff, 2001 ; figures 31 et 32), le tout promu par le cœur catalytique DDE de la transposase. Cette réaction est similaire au clivage du brin transféré, à la différence qu'ici le nucléophile n'est pas l'eau mais le groupe 3'-OH libre situé à l'extrémité de l'élément. Il est à noter que les transposases, comme les intégrases d'ailleurs, sont capables de catalyser une réaction inverse au transfert de brin, appelée désintégration (Chow *et al*, 1992 ; Polard *et al*, 1996 ; Haren *et al*, 1999). Le groupe 3'-OH libérée sur l'ADN cible par le transfert de brin joue alors le rôle du nucléophile pour attaquer la liaison phosphodiester nouvellement formée entre le transposon et l'ADN cible (figure 32). Observée uniquement *in vitro*, il n'est pas certain que la désintégration ait un véritable rôle biologique.

Le complexe ADN-Tnp attaque l'ADN receveur sur ses deux brins, mais de manière légèrement décalée, conduisant ainsi à la formation d'un trou dans l'ADN cible (figure 31), trou ensuite comblé par les enzymes réparatrices de l'ADN de l'hôte. Cela conduit à une

duplication du site cible lors de l'insertion, signature des évènements de transposition des ET eucaryotes ; il s'agit d'un TA pour la majorité des éléments *ITm*.

La réaction de transfert de brin est également catalysée par des cations divalents, qui jouent le même rôle que décrit précédemment lors du clivage (Haren *et al*, 1999 ; figure 33).

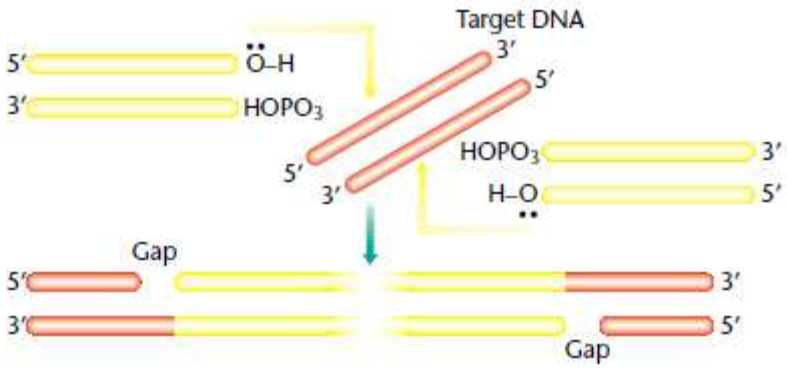


Figure 31. Réaction de transfert de brin chez les enzymes à DDE, d'après Braam & Reznikoff, 2001. *Le transposon est en jaune, l'ADN cible en rouge. Le groupe 3'-OH libre aux extrémités de l'élément rompt la liaison phosphodiester sur chaque brin de l'ADN cible. Les deux attaques nucléophiles étant décalées sur la cible, cela conduit à la formation d'un trou dans l'ADN (Gap) et à une duplication du site cible. Le gap est ensuite réparé par les enzymes de l'hôte.*

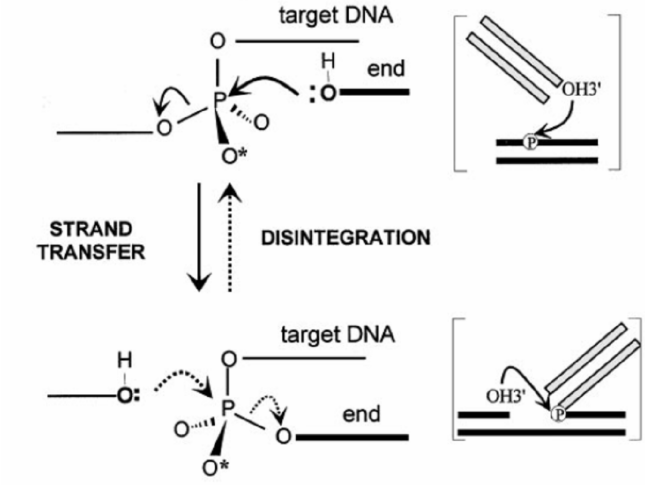
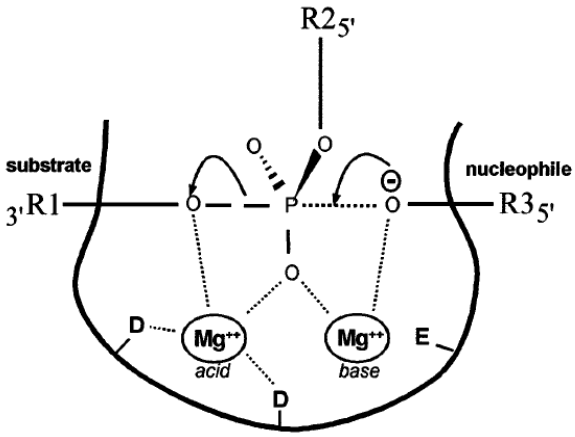
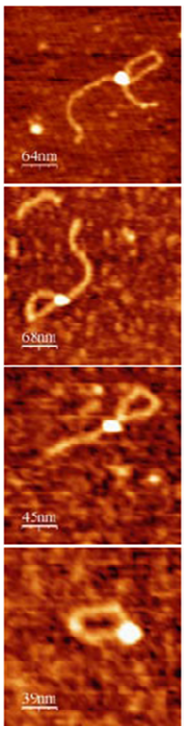


Figure 32. Chimie du transfert de brin et de la désintégration, d'après Haren *et al*, 1999. *Le groupe 3'-OH libre à l'extrémité du transposon joue le rôle du nucléophile dans la réaction de trans-estérification. La réaction provoque un changement de chiralité comme lors du clivage. La réaction inverse, appelée désintégration, est catalysée par la Tnp.*

Figure 33. Rôle des cations divalents dans la catalyse, d'après Haren *et al*, 1999. *Les cations divalents sont situés dans la poche catalytique, l'un étant coordonné aux deux acides aspartiques de la triade catalytique DDE en stabilisant l'état de transition pentacovalent autour du phosphate. Le second cation agit comme une base en déprotonant le nucléophile, augmentant ainsi sa charge négative partielle. L'ADN substrat est 3'R1-R25', le produit sera 3'R3-R25'.*



Dans l'étude du mécanisme de transposition des ET eucaryotes, une approche innovante a récemment été utilisée pour étudier la formation du complexe ADN-transposase de l'élément *P* (Tang *et al*, 2005 et 2007). Il s'agit de l'utilisation de la microscopie à force atomique (AFM) qui permet de visualiser directement les complexes formés lors des premières étapes de la transposition (figure 34). Les images obtenues montrent clairement la formation du complexe synaptique après fixation d'un tétramère préformé de transposases, puis le clivage séquentiel aux deux extrémités du transposon. Cette méthode de microscopie a également été utilisée pour étudier l'impact du GTP sur la formation du complexe synaptique de l'élément *P* (Tang *et al*, 2005 ; cf chapitre dédié à la régulation de la transposition).



Complexe synaptique

Complexes intermédiaires
avec une seule extrémité clivée

Complexe avec
les deux brins clivés

Figure 34. Formation des complexes ADN-transposase pendant les premières étapes de la réaction, d'après Tang *et al*, 2007. *Visualisation par microscopie à force atomique ; le complexe synaptique est formé, puis les deux extrémités sont clivées l'une après l'autre.*

5.3 SETMAR : une transposase *mariner* domestiquée

La protéine humaine SETMAR (connue également sous le nom de Metnase ; figure 35) est issue d'une fusion entre une méthylase de l'histone H3 et une transposase *mariner*. La composante transposase de SETMAR est entièrement codée par l'exon 3 du gène *setmar* (la protéine est communément appelée EXON 3). L'exon 3 est dérivé du transposon *mariner Hsmar1* présent en quelques 200 exemplaires dans le génome humain ; des analyses phylogénétiques montrent que la fusion a eu lieu il y a 50 millions d'années par insertion de *Hsmar1* dans le gène *SET*. Cette fusion est présente chez tous les primates (Robertson & Zumpano, 1997 ; Lee *et al*, 2005 ; Cordeaux *et al*, 2006).

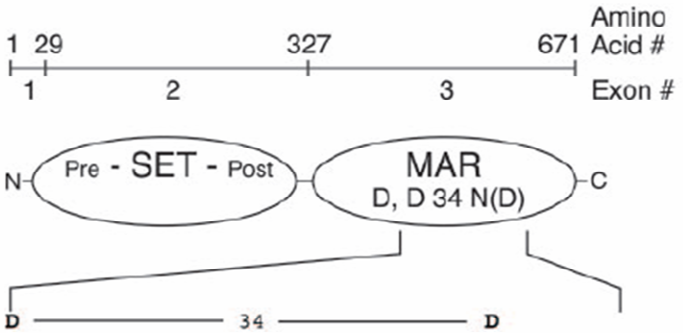


Figure 35. La protéine SETMAR. *Les exons 1 et 2 codent le domaine histone méthylase (domaine SET), et l'exon 3 code le domaine transposase (domaine MAR), dans lequel est retrouvé une triade catalytique DD(34)N, l'asparagine prenant ici la place du troisième D spécifique de la triade des éléments mariner.*

Le domaine SET possède une activité méthyl-transférase sur l'histone H3, augmentant ainsi la réparation de l'ADN par recombinaison non-homologue (NHEJ) et l'intégration d'ADN étranger au génome (Lee *et al*, 2005) ; la protéine intervient dans la réparation NHEJ en interagissant notamment avec la ligase IV, une des enzymes impliquées dans ce processus (Hromas *et al*, 2008).

Intéressons nous maintenant à l'activité transposase de SETMAR, étudiée avec l'élément *Hsmar1* comme 'partenaire' transposon. Autant la protéine pleine taille que EXON 3 seul sont capables de promouvoir des évènements de transposition (Liu *et al*, 2007), avec toutefois un taux relativement faible comparé à ce qui est trouvé chez Hsmar1 (Miskey *et al*, 2007) et chez les éléments *mariner Mos1* et *Himar1* (Liu *et al*, 2007).

Le détail du mécanisme de transposition montre que SETMAR et EXON 3 sont capables d'assembler un *Paired End Complex* (PEC), indispensable pour promouvoir l'étape suivante qu'est le clivage (Liu *et al*, 2007 ; Miskey *et al*, 2007). Danxu Liu et ses

collaborateurs ont facilement observé ce PEC en utilisant la technique de retard sur gel, ce qui leur fait dire qu'il est plus stable que celui des éléments *Mos1* et *Himar1*, chez lesquels il est délicat de l'observer (Dawson & Finnegan, 2003 ; Lipkow *et al*, 2004b ; Augé-Gouillou *et al*, 2005b ; observations personnelles).

Le complexe synaptique assemblé, l'étape de clivage peut avoir lieu comme décrit précédemment dans le chapitre consacré au mécanisme de transposition des éléments *ITm*. Il s'agit d'une réaction de trans-estérification au terme de laquelle un groupe 3'-OH est libéré sur le brin non transféré à chaque extrémité de l'ITR du transposon. Chez *Mos1*, la coupure qui permet la réintégration se fait précisément à l'extrémité de l'élément sur le brin NTS, et trois bases à l'intérieur de l'ITR sur le brin TS. Les coupures sont identiques avec EXON 3 (Liu *et al*, 2007) comme avec la protéine pleine taille (Miskey *et al*, 2007). La seule différence est que la coupure du brin transféré, qui libère le groupe 3'-OH libre nécessaire pour l'étape d'intégration, est très peu efficace. Il est à noter ici que l'utilisation d'un élément préalablement correctement clivé sur les deux brins permet d'augmenter l'activité de transposition, et que EXON 3 n'a aucun problème pour catalyser l'intégration des extrémités de *Hsmar1* à un dinucléotide TA.

Le clivage sur le brin transféré semble donc être l'étape limitante qui explique le faible taux de la transposition de l'élément *Hsmar1* promue par SETMAR. Sachant que la mutation de la triade catalytique DDD de la Tnp de *Mos1* en DDE donne une protéine non fonctionnelle (Lohe *et al*, 1997), il était logique de penser que la triade catalytique DDN de SETMAR soit un obstacle à la transposition promue par cette protéine. Une protéine EXON 3 mutée avec une triade DDD augmente la fréquence de transposition, mais ne restaure pas un clivage efficace sur le brin transféré à l'extrémité de l'élément *Hsmar1*. Cela peut s'expliquer par le fait que la mutation D en N est structurellement conservative et permet la coordination normale d'un des deux cations métalliques lors de la catalyse. D'autres facteurs entrent certainement en jeu pour expliquer la faible efficacité de la transposition promue par SETMAR ; n'oublions pas qu'il y a 19 acides aminés de différent entre la séquence de l'exon 3 et la séquence consensus de la Tnp de *Hsmar1* (Robertson & Zumpano, 1997).

Tenant compte du fait que SETMAR permet une coupure efficace du premier brin, mais pas sur le second brin, une hypothèse sur le rôle des acides aminés de la triade catalytique peut être émise. Les deux premiers D seraient suffisants pour promouvoir seuls le clivage du brin non transféré, en coordonnant sans problème un des deux cations métalliques.

Par contre, le clivage du brin transféré nécessiterait la coordination du second cation en impliquant le troisième acide aminé de la triade, et la présence d'un N ne permettrait pas cette coordination efficace, conduisant à un clivage du second brin très peu efficace.

Les contraintes évolutives ont préservé la fonction de liaison à l'ADN du domaine transposase de SETMAR (Cordaux *et al*, 2006), et ce domaine Tnp de la protéine est essentiel pour l'activité de la protéine dans la cellule (Lee *et al*, 2005). Ces deux éléments pris en compte amènent certains auteurs (Miskey *et al*, 2007) à émettre l'hypothèse que cette fonction serait liée à la présence dans le génome humain de nombreux site de liaison pour SETMAR. En effet, en plus des 200 copies de l'élément *Hsmar1* présentes dans le génome humain, il existe dans notre génome plus de 2000 copies de mini-éléments *Hsmar1*, appelés *Mihsmar1* (des éléments non autonomes), ainsi que de nombreux ITR seuls, ce qui représente au total un peu moins de 7000 sites de liaison potentiels (Cordaux *et al*, 2006 ; Liu *et al*, 2007). La liaison de SETMAR à ces sites pourrait entraîner des modifications de la chromatine, essentielles pour le fonctionnement normal de la cellule. Cette hypothèse est renforcée par une étude récente qui montre que SETMAR interagit, en augmentant son activité, avec la topoisomérase IIα, enzyme clé dans le processus de décaténation des chromosomes, (Williamson *et al*, 2008).

5.4 RAG et la recombinaison V(D)J

La recombinaison V(D)J est l'élément clé du système immunitaire des vertébrés, promoteur d'une grande diversité parmi les immunoglobulines et récepteurs antigéniques des lymphocytes T. Dans les ceux cas, c'est une recombinaison site-spécifique de l'ADN. Dans les lymphocytes B, elle intervient au niveau des gènes codant de la partie variable (V) des chaînes lourde et légère des immunoglobulines ; un exon variable est en effet constitué par les segments géniques V (pour variable), D (pour diversité) et J (pour jonction). Avec plusieurs dizaines de segments V et D, et moins d'une dizaine pour J, cela laisse possible la formation de plus d'un million de combinaisons différentes (figure 36).

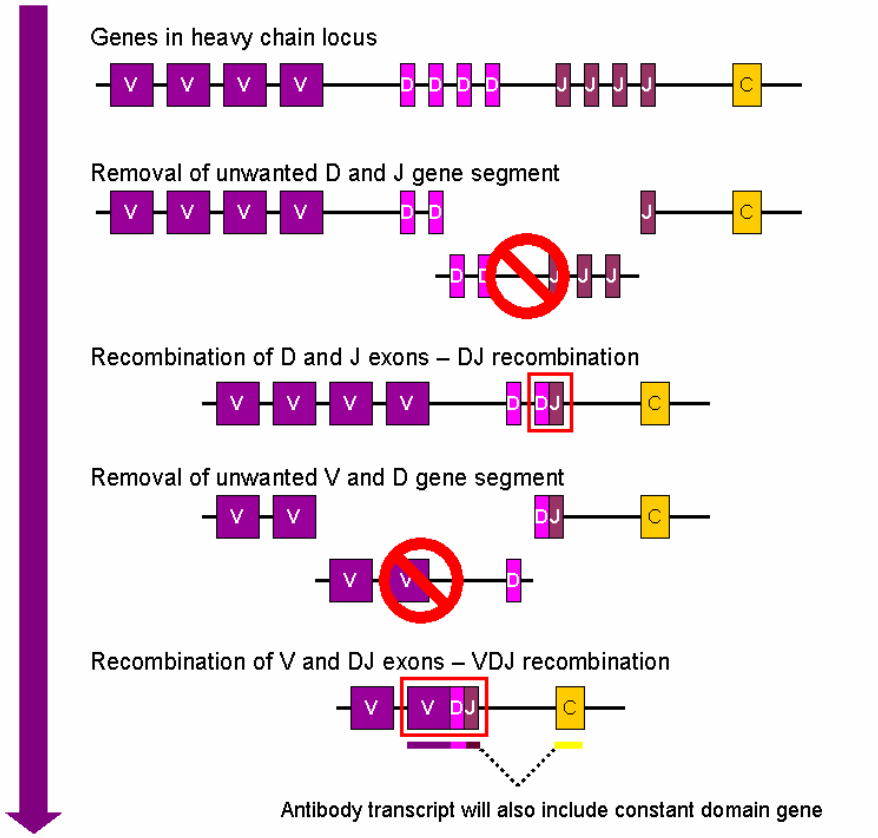


Figure 36. La recombinaison V(D)J. *Elle permet l'élimination aléatoire des segments D et J, puis des segments V, pour former une combinaison originale comprenant un segment V, D et J, qui s'associent avec un segment C codant la partie non-variable de l'anticorps.*

Le principe de la recombinaison V(D)J est le même pour les récepteurs antigéniques des lymphocytes T puisqu'il y a une recombinaison entre trois segments géniques V, D et J.

Intéressons nous plus précisément au mécanisme qui permet cette recombinaison ; il implique deux protéines, RAG1 et RAG2 (RAG pour *Recombination Activating Gene*), ainsi que des RSS (*Recombination Signal Sequences*) qui flanquent les segments géniques V, D et J, à l'image des ITR flanquant les gènes codant les transposases chez les éléments eucaryotes décrits précédemment. Ces RSS contiennent deux séquences nucléotidiques conservées, séparées par 12 ou 23 nucléotides (figure 37) ; pour que la recombinaison soit efficace, il faut qu'un RSS (12) recombine avec un RSS (23). Ces séquences RSS sont reconnues spécifiquement par le complexe enzymatique RAG1/2.

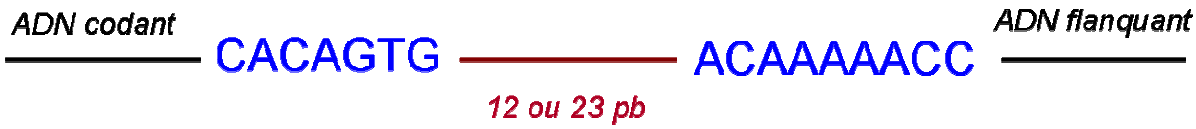


Figure 37. Schéma d'un RSS. *Le RSS, situé après l'ADN codant le segment génique, est constitué d'un heptamère palindromique et d'un nonamère. Ces deux séquences sont conservées et séparées par 12 ou 23 pb.*

Cette réaction est apparentée à la transposition et met en jeu la protéine RAG à cœur catalytique DDE. Les gènes codant la protéine RAG dérivent des éléments *Transib* (Kapitonov & Jurka, 2005 ; Hemmrich *et al*, 2007 ; Volff, 2006 ; Sinzelle *et al*, 2009). La recombinaison V(D)J implique une coupure double-brin de l'ADN et la formation d'une *hairpin* sur l'extrémité codante. Son mécanisme est proche de celle de l'élément *hAT Hermes* (Zhou *et al*, 2004 ; figure 38).

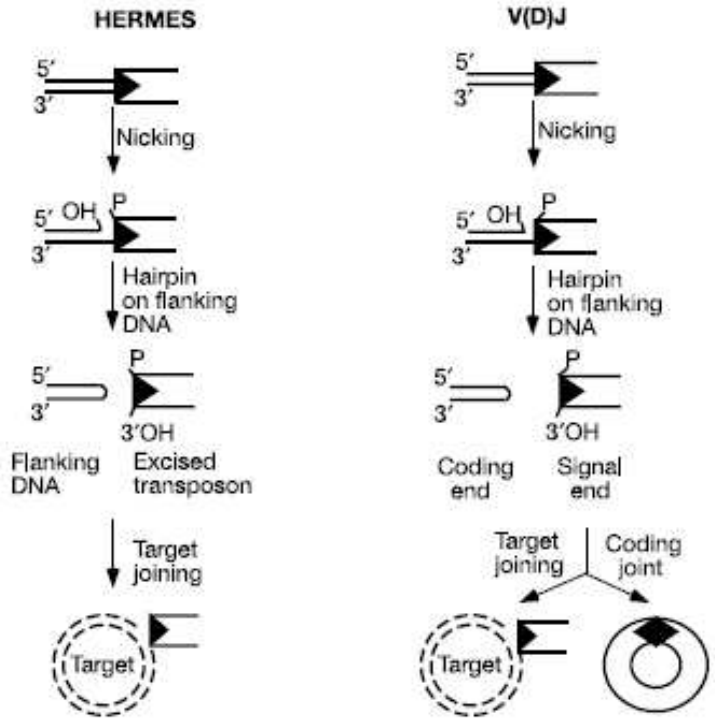


Figure 38. Clivage de l'ADN chez *Hermes* et lors de la recombinaison V(D)J, d'après Zhou *et al*, 2004. *Chez Hermes, l'hairpin se forme sur l'ADN flanquant non codant, tandis que dans la recombinaison V(D)J, l'hairpin se forme sur l'ADN codant. Le groupe 3'-OH libéré lors de la réaction est disponible pour une réaction de trans-estérification in vitro dans les deux cas.*

La recombinaison V(D)J implique également des protéines cellulaires, HMG1 ou HMG2. Le complexe composé des protéines RAG1, RAG2 et HMG1/2 fixe les RSS formant ainsi un complexe stable (Hiom & Gellert, 1998). Après le clivage de l'ADN, les protéines demeurent associées (Agrawal & Schatz, 1997 ; Hiom & Gellert, 1998). Ces protéines sont capables de promouvoir *in vitro* le transfert de brin de l'ADN non codant inutile pour la suite de la recombinaison V(D)J (Agrawal *et al*, 1998). Dans ce complexe, la protéine RAG1 seule peut reconnaître le RSS (Ciubotaru *et al*, 2003). Le rôle de RAG2 est notamment d'inhiber la liaison non-spécifique de RAG1 sur l'ADN (Zhao *et al*, 2009).

D'autres protéines cellulaires interviennent également dans la recombinaison V(D)J, notamment après la formation de l'*hairpin*. Les protéines Artémis et un complexe formé des protéines DNA-PK et Ku sont impliqués dans la résolution des *hairpins*, puis un autre complexe protéique, contenant la ligase, agit pour réunir les segments géniques (figure 39) par recombinaison de type NHEJ.

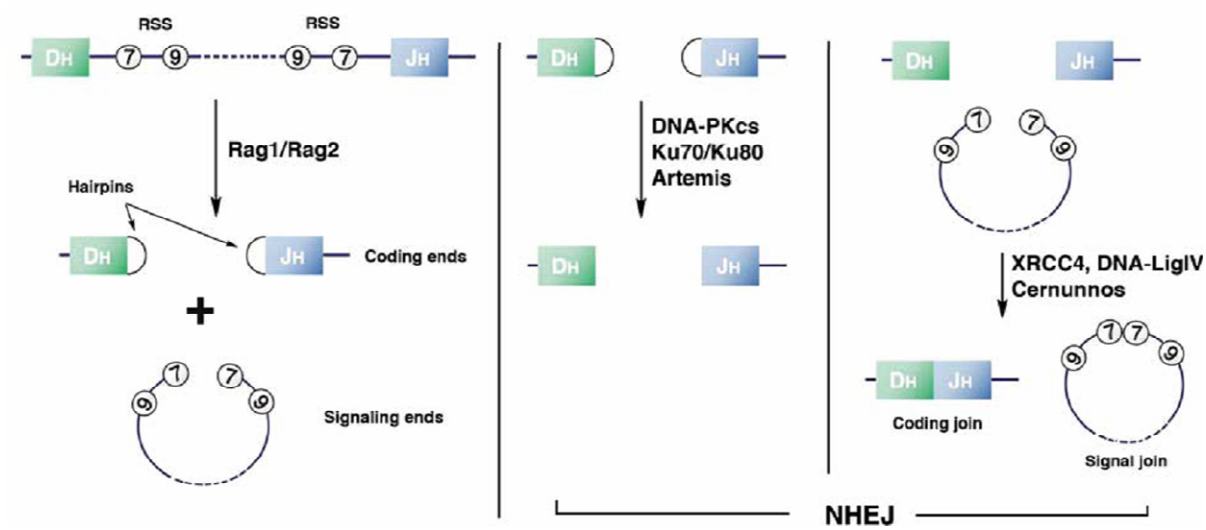


Figure 39. Les trois étapes de la recombinaison V(D)J, d'après Soulas-Sprauel *et al*, 2007. *Le complexe RAG1/2 clive au niveau des RSS des segments géniques ce qui conduit à la formation d'extrémités en hairpin. Les protéines Artémis et le complexe DNA-PK/Ku résout ensuite l'hairpin, avant qu'un autre complexe protéique ne ligue les segments géniques.*

Démontrant une fois de plus le lien unissant toutes ces enzymes à triade catalytique DDE, transposases, intégrases et recombinases RAG, Meni Melek et ses collaborateurs ont montré que des inhibiteurs de l'intégrase du VIH-1 étaient également efficaces sur les protéines RAG1/2 (Melek *et al*, 2002). Ces inhibiteurs agissent sur la liaison et le clivage de l'ADN par l'intégrase ou les protéines RAG1/2.

5.5 L'intégrase du VIH-1

Le Virus de l'Immunodéficience Humaine de type 1 (VIH-1) est le rétrovirus responsable du SIDA (Syndrome d'immunodéficience acquise), qui affaiblit le système immunitaire et rends ainsi l'organisme vulnérable aux infections opportunistes.

L'intégrase du VIH-1 est un élément clé du cycle rétroviral. Cette enzyme est fortement étudiée en raison de la pandémie de SIDA en cours depuis la fin des années 70. Après rétrotranscription de l'ARN rétroviral en ADN double brin *via* une transcriptase inverse virale, l'ADN viral est intégré dans l'ADN de l'hôte grâce à l'intégrase.

Les trois protéines nécessaires au cycle viral (protéase, transcriptase inverse et intégrase) sont codées par le gène *pol* du virus. Il exprime une polyprotéine ; l'intégrase est le

résultat du clivage de l'extrémité C-terminale de cette polyprotéine. Cette protéine de 228 acides aminés est constituée de trois domaines (Ceccherini-Silberstein *et al*, 2009) :

- le domaine N-terminal qui contient un motif en doigt de zinc impliqué dans la stabilisation du repliement de la protéine ; ce domaine assure l'oligomérisation de l'enzyme, processus clé de l'intégration.

- le cœur catalytique à triade DD35E. Ces trois acides aminés coordonnent l'action d'un ou deux cations divalents (Mg^{2+} ou Mn^{2+}) qui forment un pont avec l'ADN (Marchand *et al*, 2006 ; figure 40). Le cœur catalytique contient également d'autres sous-domaines et résidus fonctionnels, dont le motif de multimérisation responsable de la dimérisation de la protéine.

- le domaine C-terminal est responsable de la liaison non-spécifique à l'ADN ; ce domaine est aussi requis pour l'intégration. Il est impliqué dans l'oligomérisation de la protéine et dans la stabilisation du complexe formé avec l'ADN.

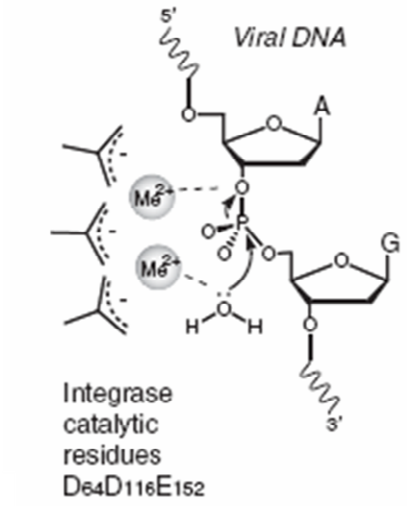


Figure 40. Les cations pontent la triade catalytique de l'intégrase et l'ADN, d'après Semenova *et al*, 2008. *Les cations divalents, Mg2+ ou Mn2+, servent de lien entre l'ADN, substrat de l'enzyme, et l'intégrase rétrovirale.*

Le cristal du cœur catalytique de l'intégrase du VIH-1, résolu dans les années 90, a permis de classer définitivement cette enzyme parmi les enzymes à cœur RNaseH-*like* (Dyda *et al*, 1994), comme les transposases à DDE (figure 41). La protéine complète n'a pas été cristallisée.

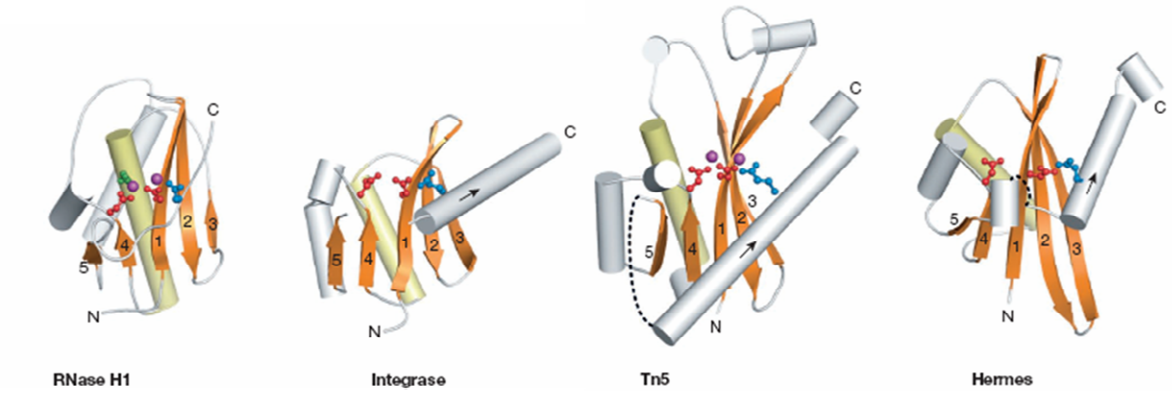


Figure 41. Structure des cœurs catalytiques RNaseH-*like* d'enzymes à DDE, d'après Nowotny, 2009. *Sont comparés ici les structures des cœurs catalytiques de la RNaseH1, de l'intégrase du VIH-1, et des transposases des éléments Tn5 et Hermes. Les feuillets β en orange ainsi que l'hélice α en jaune, sont conservés ; les parties en gris représentent des structures non conservés.*

Contrairement au système transposon/transposase, l'intégrase seule ne suffit pas pour promouvoir l'intégration de l'ADN rétroviral dans l'ADN cellulaire. L'intégrase se lie sous forme oligomérique à l'ADN juste après sa rétrotranscription. La formation du complexe de pré-intégration, ou PIC (*Pre-Integration Complex*), est crucial pour le passage de l'ADN dans le noyau et l'intégration de l'ADN viral dans le génome de l'hôte. Ce PIC contient un cœur de protéines rétrovirales (la transcriptase inverse, la matrice et la nucléocapside), associé à des protéines de l'hôte (Bukrinsky *et al*, 1993 ; Van Maele & Debyser, 2005 ; Van Maele *et al*, 2006). La plus connue est LEDGF/p75 qui se lie directement à l'intégrase (Busschots *et al*, 2007). Ces cofacteurs cellulaires sont primordiaux pour promouvoir l'intégration de l'ADN, le PIC étant un élément clé du mécanisme d'intégration (figure 42).

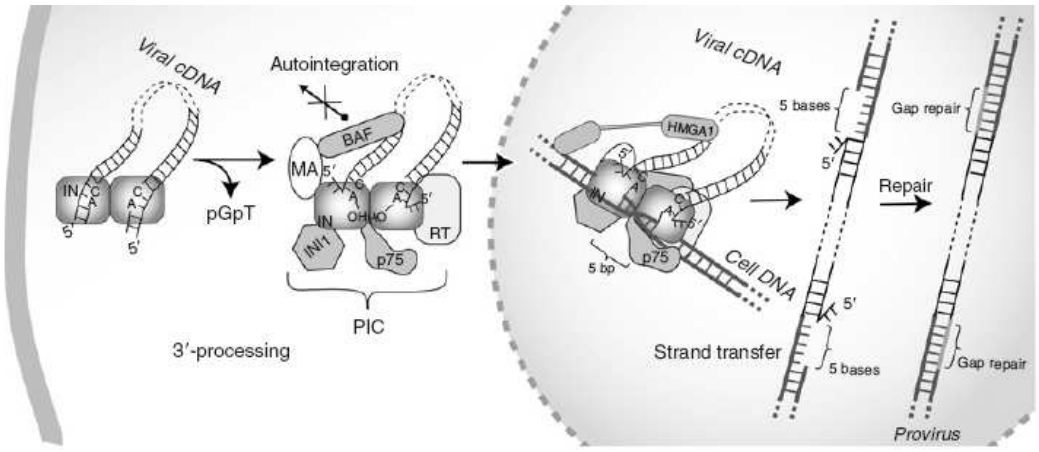


Figure 42. Formation du PIC et passage dans le noyau, d'après Semenova *et al*, 2008. *Le complexe de pré-intégration (PIC) est formé de l'intégrase, de protéines virales et de cofacteurs cellulaires. Sa formation est primordiale pour le passage de l'ADN viral rétrotranscrit dans le noyau.*

L'intégration de l'ADN promue par l'intégrase se déroule en deux grandes étapes : le *3'-processing* et le transfert de brin. Après s'être fixée aux LTR de l'ADN viral, l'intégrase sous forme dimérique catalyse une attaque nucléophile qui conduit à la libération d'un groupement 3'-OH, c'est l'étape de 3'-*processing*, qui prépare l'ADN rétroviral à la réaction covalente de transfert de brin (Faure *et al*, 2005, Guiot *et al*, 2006). Au cours de cette réaction, un dinucléotide GT est éliminé à chaque extrémité de l'ADN viral. Le PIC est transloqué dans le noyau, puis l'étape de transfert de brin sur le génome de l'hôte résulte d'une réaction de trans-estérification (figure 43) promue par un tétramère d'intégrases (Faure *et al*, 2005, Delelis *et al*, 2008).

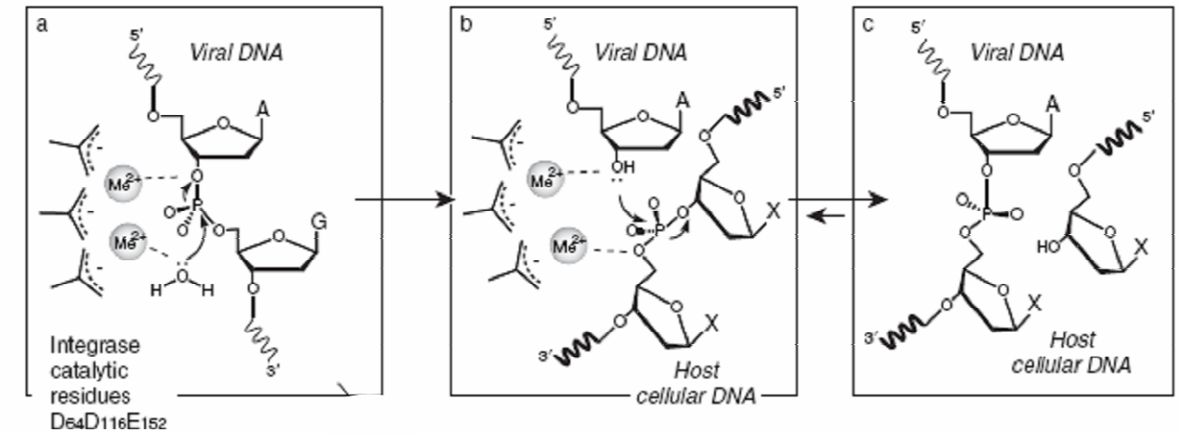


Figure 43. L'intégration rétrovirale vue sur le plan biochimique, d'après Semenova *et al*, 2008. (a) *Liée sur les LTR de l'ADN viral, l'intégrase promeut une attaque nucléophile sur cet ADN viral.* **(b)** *Une fois le PIC dans le noyau, c'est l'ADN cellulaire qui est la cible d'une attaque nucléophile, ce qui permet le transfert de brin* **(c)**.

Le *3'-processing* et le transfert de brin correspondent respectivement au clivage des ITR et au transfert de brin chez les transposons/transposases. Comme dans ces deux réactions, ils nécessitent la présence d'un cofacteur métallique, Mg^{2+} ou Mn^{2+}, le premier étant utilisé préférentiellement dans la cellule (Delelis *et al*, 2008). Les études biochimiques de ces deux réactions sont possibles *in vitro* en utilisant des oligonucléotides double brin mimant les LTR de l'ADN rétroviral (Delelis *et al*, 2008).

L'intégrase est une cible de choix dans la recherche d'inhibiteurs chimiques agissant contre le VIH-1. La recherche d'inhibiteurs de l'intégrase se développe depuis les années 90, et en 2007 un premier inhibiteur a été approuvé pour passer en tests cliniques, il s'agit du *Raltegravir* (Hazuda *et al*, 2009).

6 Régulation de la transposition

Nous avons vu dans les chapitres précédents la diversité des éléments transposables, retrouvés chez tous les organismes vivants où ils ont été recherchés. Il est facile d'imaginer que la transposition de ces éléments puisse être délétère pour les génomes, et la régulation de l'activité des ET essentielle pour empêcher d'importants taux de mutations. Pour préserver leur intégrité, les génomes ont donc mis en place divers systèmes de régulation des éléments transposables. De leurs côtés, les ET doivent se maintenir dans des régions où les gènes sont exprimés, en évitant une répression/élimination de la part de l'hôte. Leur activité peut également être bénéfique pour les génomes. Des conditions de stress peuvent déclencher une réponse adaptative dans laquelle les ET jouent un rôle en étant notamment domestiqués pour former des néogènes.

6.1 Expression de la transposase

Face à la transposition, les cellules peuvent tenter d'agir au niveau de la quantité active de transposase, en régulant le taux de transcription de la protéine, ou en agissant sur la protéine elle-même.

La **méthylation** de l'ADN au niveau des cytosines est l'une des modifications de l'ADN les plus répandues chez les eucaryotes, qui l'utilisent en tant que contrôle épigénétique (Bird, 2002). Il semble que chez les mammifères la méthylation soit l'arme majeure du système de défense mis en place par l'hôte pour lutter contre la menace que représente les parasites génétiques, éléments transposables en tête, en intervenant au niveau des promoteurs des ET (Yoder *et al*, 1997). Chez les éléments de plantes *Activator* et *Mutator*, la méthylation des ITR entraîne une inhibition de la transposition (Chandler & Walbot, 1986 ; Chomet *et al*, 1987 ; Slotkin *et al*, 2003).

Les histones peuvent également être méthylées, un processus connu dans le *silencing* des gènes (Lachner & Jenuwein, 2002). Un système de *silencing* des ET chez *Arabidopsis thaliana* a été mis en évidence avec une interdépendance entre la méthylation de l'ADN, celle des histones et le phénomène d'ARN interférence (RNAi) (Zilberman & Henikoff, 2004). Le phénomène de **RNAi** est décrit chez le ver *Caenorhabditis elegans*, où il provoque

l'inactivation de l'élément *Tc1* (Ketting *et al*, 1999 ; Vastenhouw & Plasterk, 2004). D'autres petits fragments ARN spécifiques des lignées germinales, appelés piRNA, sont également impliqués dans le *silencing* des ET chez la drosophile (Grivna *et al*, 2006 ; Vagin *et al*, 2006).

Peut également intervenir une régulation au cours du **développement de l'hôte,** appelée **régulation 'spacio-temporelle'**. Chez la drosophile, l'élément *P* n'est fonctionnel qu'en cellules germinales ; en effet, dans les cellules somatiques, il n'y a pas d'épissage du dernier intron, ce qui a pour conséquence la production d'une transposase tronquée non fonctionnelle et inhibitrice de la transposition (Laski *et al*, 1986 ; Misra & Rio, 1990). L'instabilité génétique observée chez le maïs (figure 44) est causée par des ET (*Mutator, Spm, Ac/Ds*) ; leur fréquence d'excision est considérablement augmentée pendant le développement du tégument du grain de maïs (Levy & Walbot, 1990).



Figure 44. Instabilité génétique chez le maïs. *Les éléments transposables induisent le changement de couleur des grains de maïs.*

L'inhibition par surproduction de la transposase (OPI pour *Over-Production Inhibition*) est un phénomène initialement décrit chez la drosophile pour l'élément non-autonome *peach*, excisé par la transposase de *Mos1* (Lohe & Hartl, 1996). La fréquence d'excision de l'élément *peach* diminue en présence d'un grand nombre de copie de *Mos1* ou de la surexpression d'une copie de *Mos1*, l'hypothèse étant que lorsqu'il y a un trop grand nombre de transposases la surproduction de la Tnp entraîne la formation d'oligomères inactifs. Ce phénomène a également été mis en évidence pour d'autres éléments de la

superfamille *Tc1-mariner* : *in vitro* pour la Tnp de l'élément *Himar1* (Lampe *et al*, 1998), et *in vivo* pour *Sleeping Beauty* (Geurts *et al*, 2003). Chez *Mos1*, les résultats sont différents *in vitro* ; la quantité de Tnp optimale pour la transposition est de 100 nM, et l'augmentation de la concentration de Tnp ne provoque pas de diminution du niveau du taux de transposition, qui se stabilise (Tosi & Beverley, 2000). Il est à noter que cette formation d'oligomères inactifs de transposases est retrouvé chez l'élément *Activator* du maïs (Kunze *et al*, 1993).

Les **éléments défectifs**, maintenus à travers l'évolution dans les génomes, pourraient jouer un rôle dans la régulation de la transposition des ET. Des copies défectives de transposons, avec des ITR normaux, peuvent amener des transposases actives à mobiliser ces éléments plutôt que des copies actives, comme cela a été décrit pour l'élément *P* (Simmons & Bucholz, 1985). Chez les éléments *mariner*, il a été montré que des transposases mutées ou délétées, produites par des copies mutées du gène, sont impliquées dans un phénomène dit de complémentation dominant-négative (Lohe *et al*, 1996 ; Lohe & Hartl, 1996 : Lohe *et al*, 1997). Ces protéines mutantes interviennent de différentes manières sur la régulation de la transposition : en s'associant avec des transposases actives, pour former des complexes synaptiques inactifs, ou induire l'OPI de Tnp actives, ou alors en se fixant sur les ITR du transposon, empêchant ainsi la fixation de transposase active.

6.2 Facteurs physico-chimiques influençant la transposition

Comme toutes les enzymes, les transposases ont besoin de conditions physico-chimiques adéquates pour fonctionner de manière optimale. Ce chapitre sera centré sur les connaissances actuelles de l'influence de ces facteurs sur la transposition des éléments *mariner*.

6.2.1 La température

Chaque enzyme possède une température optimale pour effectuer sa réaction, et les transposases ne diffèrent pas des autres enzymes sur cet aspect. Il avait été montré que la température optimale pour la transposase de *Himar1* était de 28°C et que son activité diminuait avec l'augmentation de la température (Lampe *et al*, 1998). Il en est de même chez *Mos1* où la Tnp a une activité optimale entre 28°C et 30°C (Sinzelle *et al*, 2008). Il est toutefois intéressant de noter que la liaison de la Tnp aux ITR de *Mos1* est indépendante de la température (Augé-Gouillou *et al*, 2005). Au contraire, la transposition est plus efficace à

30°C qu'à 37°C. Ceci s'explique par le fait qu'à 37°C les coupures sont inexactes et libèrent un transposon incapable de se réintégrer, tandis qu'à 30°C les coupures induites par la Tnp permettent une future réintégration (Sinzelle *et al*, 2008 ; Julien Bischerour, communication personnelle).

C'est l'inverse pour l'élément *hAT Tam3* chez le muflier *Antirrhinum majus* ; sa transposition est dépendante d'une température basse, inférieure à 15°C, et est très fortement réduite au dessus de 25°C (Carpenter *et al*, 1987). Ce phénomène est décrit sous l'acronyme LTDT (pour *Low-Temperature-Dependant Transposition*).

6.2.2 Les cations divalents

Enzymes avec une activité phosphatidyl-transférase, les transposases sur lesquelles j'ai axé mon précédent chapitre possèdent une triade catalytique DDxE/D. Comme les intégrases rétrovirales à DDE, elles ont un site de liaison de cations divalents, cofacteurs essentiels de l'activité enzymatique (Mizuuchi, 1992 ; Haren *et al*, 1999 ; Nowotny, 2009). Ces cations divalents sont le magnésium (Mg^{2+}) ou le manganèse (Mn^{2+}). Chez *Mos1* la structure cristalline de la Tnp a été observée avec un cation magnésium ou deux cations manganèse. La liaison du second cation métallique ne change pas la conformation de la transposase, excepté qu'il y a une rotation de la chaîne d'atomes de l'acide aspartique 284, le troisième acide aminé de la triade catalytique DDD (Richardson *et al*, 2006)

Il est tout d'abord à noter que ces cations n'ont aucun impact sur la liaison aux ITR puisque la Tnp se lie aussi bien à son ITR 3' en présence des cations ou d'un chélateur de cations comme l'EDTA (Augé-Gouillou *et al*, 2001). Le rôle des cations est important au niveau de l'activité phosphatidyl-transférase de la transposase ; chez *Mos1* et *Himar1*, il a été montré *in vitro* que le Mn^{2+} permettait d'obtenir une meilleure fréquence de transposition que le Mg^{2+} (Tosi & Beverley, 2000 ; Zhang *et al*, 2001 ; Lampe *et al*, 1996). Néanmoins, la spécificité d'insertion au niveau du dinucléotide TA est moins respectée avec le Mn^{2+}. Ainsi, le fait qu'*in vivo* les éléments *mariner* soient détectés uniquement flanqués de deux TA conduit à penser que le magnésium est le cation préférentiel de la réaction utilisé *in vivo*.

6.2.3 L'élément *P* et le GTP

La transposase de l'élément *P* présente la caractéristique unique parmi les transposases et intégrases d'utiliser du GTP (Guanosine Tri-Phosphate) pour effectuer la réaction de

transposition (Kauffman & Rio, 1992 ; Craig *et al*, 2002). Le GTP promeut l'assemblage du complexe synaptique initial ADN-transposase (Tang *et al*, 2005). L'hydrolyse du GTP n'est pas nécessaire pour effectuer l'assemblage du complexe ou promouvoir les réactions de coupure. En effet, la réaction est possible en présence de GTP non hydrolysable. Le rôle exact du GTP reste donc à préciser.

6.3 Le transposon

L'élément lui-même peut avoir une influence sur sa propre transposition, en fonction de sa taille et de sa séquence. Chez *Himar1* (Lampe *et al*, 1998) et *Mos1* (Lohe & Hartl, 1996 ; Sinzelle *et al*, 2008), la taille joue un rôle prépondérant dans l'efficacité de transposition. Dans le cas de *Mos1*, des pseudo-éléments où le gène codant la Tnp a été remplacé par des gènes de résistance aux antibiotiques, ont été utilisés pour étudier cette question. Les transgènes, d'une taille comprise entre 375 et 2500 pb, donnent des fréquences de transposition autour de 10^{-4}, ce qui est retrouvé classiquement avec le transposon sauvage. L'augmentation de la taille du transgène conduit à la diminution de la fréquence de transposition (figure 45).

Transgene size (bp)	Transposition frequencies
375	3.1×10^{-4}
423	0.9×10^{-4}
622	4×10^{-4}
1185	0.9×10^{-4}
1191	1×10^{-4}
2500	1.2×10^{-4}
5000	5×10^{-6}
7000	3.6×10^{-8}
12500	$<10^{-9}$

Figure 45. Impact de la taille du transposon sur sa fréquence de transposition en bactérie, d'après Sinzelle *et al*, 2008. *La fréquence de transposition d'un élément sauvage est de* 10^{-4} *; on note une diminution drastique de la fréquence de transposition avec une taille supérieure à 2500 pb.*

Si les transposases doivent scanner l'ADN à la recherche des ITR du transposon, il est logique de penser que la longueur influe sur la formation du complexe synaptique. Comme décrit précédemment, la littérature est riche d'exemples du fait que la fréquence de transposition diminue avec l'augmentation de la taille du fragment d'ADN compris entre les deux ITR aussi bien pour des éléments procaryotes (Chandler *et al*, 1982) qu'eucaryotes

(Lohe & Hartl, 1996 ; Lampe *et al*, 1998 ; Izsvak *et al*, 2000 ; Mates *et al*, 2007). Des exceptions existent néanmoins comme chez les éléments *Tol1* (Koga *et al*, 2007) et *Tol2*. Le second, long de 4,7 kb, transpose en effet aussi efficacement qu'un élément *Sleeping Beauty* long de 1,7 kb. De plus, un pseudo-élément de 10 kb a une efficacité de transposition égale à l'élément *Tol2* naturel (Balciunas *et al*, 2006 ; Hackett, 2007).

Au niveau de la séquence nucléotidique même de l'élément, il a été montré qu'il y avait une corrélation entre le taux de GC et la fréquence de transposition chez *Mos1* (Sinzelle *et al*, 2008). Toujours chez *Mos1*, les mêmes auteurs ont observé que la transposition était optimale quand celui-ci était sous forme super-enroulée ; l'intégration d'un donneur sous forme linéaire est 1000 fois moins efficace, ceci étant probablement la conséquence d'une excision non concertée.

6.4 Facteurs d'hôte et modifications post-traductionnelles

La mobilité des éléments transposables est souvent restreinte à un spectre d'hôtes. Notre modèle, l'élément *mariner Mos1*, est efficace en bactéries et très efficace en nématodes. Il transpose naturellement chez son hôte *Drosophila mauritiana* mais fonctionne moins bien dans d'autres organismes, voire pas du tout comme en cellules de mammifères. L'élément *Tc1-like Sleepling Beauty*, quant à lui, ne fonctionne pas en bactéries. Ces différences pourraient s'expliquer par la présence de facteurs d'hôte protéiques intervenant au moment de la transposition, et/ou des modifications post-traductionnelles, qui peuvent différer selon les organismes.

6.4.1 Facteurs d'hôte protéiques

Dans le chapitre consacré au mécanisme de transposition, j'ai montré que la particularité de l'élément bactérien Tn*10* est qu'il faisait intervenir des protéines d'hôte dans son mécanisme, les protéines IHF et H-NS. Ces protéines sont essentielles pour permettre la transposition de l'élément, IHF intervenant dans la formation du complexe synaptique. Les deux protéines agissent ensuite en promouvant soit une réaction de transposition intramoléculaire, soit une réaction intermoléculaire ; elles agissent également lors de la reconnaissance du site cible. Le fait de promouvoir une transposition intramoléculaire étant suicidaire pour le transposon, ces protéines peuvent être considérées comme régulatrices de la transposition de Tn*10*.

La protéine HMGB1 (impliquée dans la régulation des gènes et la recombinaison de l'ADN – Bustin, 1999) est un cofacteur de la transposition de *Sleeping Beauty*, nécessaire pour que la Tnp se fixe sur les ITR du transposon (Zayed *et al*, 2003). La Tnp de *Sleeping Beauty* est capable de fixer d'autres protéines, comme le facteur de transcription Miz-1, induisant ainsi un blocage du cycle cellulaire en phase G1, une action qui pourrait favoriser la transposition de l'élément (Walisko *et al*, 2006).

Autre protéine clé de la transposition de *Sleeping Beauty*, la sous-unité Ku de la DNA-PK (protéine kinase ADN dépendante), qui agit classiquement dans la réparation des cassures double-brin de l'ADN, *via* le mécanisme de réparation NHEJ (*Non Homologous End Joining*). La Tnp de *Sleeping Beauty* a la capacité de se lier à cette sous-unité Ku (Izsvak *et al*, 2004). Chez la drosophile, la protéine IRBP (*ITR Binding Protein*), qui présente de fortes homologies avec Ku, se lie aux ITR de l'élément *P* et serait impliqué dans la protection des ITR lors de la coupure double-brin de l'ADN (Rio & Rubin, 1988 ; Staveley *et al*, 1995).

6.4.2 Modifications post-traductionnelles

La littérature n'est pas riche à propos des modifications post-traductionnelles des transposases. Les seules décrites concernent la phosphorylation de la transposase de l'élément *P* ou d'intégrases de rétroéléments.

En effet, la transposase de l'élément *P*, phosphorylable par les ATM kinases (enzymes impliquées dans les processus de réparation des cassures double-brin de l'ADN), est régulée par phosphorylation, une action qui diminue l'activité de la transposase (Beall *et al*, 2002).

Ronald Plasterk a proposé que la phosphorylation puisse modifier la localisation des transposases des éléments *TLE* dans les cellules ; des sites consensus de phosphorylation par la CKII (*Casein Kinase II*) ont en effet été identifiés au niveau du signal de localisation nucléaire de ces Tnp (Plasterk *et al*, 1999).

Chez le rétrotransposon *Ty5* de la levure, la phosphorylation joue un rôle clé dans sa spécificité d'insertion (Dai *et al*, 2007), promue par un petit domaine de l'intégrase de *Ty5*, appelé TD (pour *Targeting Domain*). Ce domaine TD interagit directement avec une protéine de l'hétérochromatine (Sir4) ce qui permet l'attachement du complexe synaptique à la cible ; l'interaction entre TD et Sir4 n'a lieu que si le domaine TD est phosphorylé. De plus, la phosphorylation induit une insertion dans une région du génome pauvre en gènes, réduisant

ainsi les conséquences délétères de la transposition pour la cellule. L'absence de phosphorylation de TD conduit *Ty5* à une intégration aléatoire dans le génome, provoquant des mutations le plus souvent délétères pour la cellule. Ce processus de régulation va dans le sens de l'intégrité de la cellule et de la 'survie' de l'élément. Les ET peuvent en effet d'autant plus se maintenir dans les génomes s'ils s'intègrent dans des régions pauvres en gènes (Boeke & Devine, 1998) ; un compromis doit toutefois être trouvé car les ET doivent être exprimés pour ne pas disparaître. Dans le cas de *Ty5*, la phosphorylation permet à la cellule de contrôler l'élément, et à l'élément de persister dans le génome de son hôte.

Cette introduction a permis de faire l'état des connaissances sur la mécanique des enzymes à triade catalytique DDE et sur la régulation des éléments transposables. Mécanique et régulation de la transposase de l'élément *mariner Mos1* ont été deux aspects de mes travaux de thèse qui sont décrits dans la partie Résultats.

Résultats

Les résultats que j'ai obtenus au cours de ma thèse sont organisés en trois grandes parties. Le point commun entre ces trois parties est la transposase de l'élément *mariner Mos1* : son organisation structurelle, sa mécanique de transposition, sa régulation et son inhibition.

La **première partie** est consacrée à la transposase, avec l'étude spécifique de sa partie N-terminale qui contient son domaine de liaison à l'ADN et pour lequel deux modèles font débat parmi les spécialistes des éléments *mariner* : domaine de type CRO ou domaine de type PAI-RED ? Nous avons répondu à la question, ce qui n'a pas qu'une importante théorique, puisque le mode de liaison de la Tnp aux ITR conditionne la structure du complexe synaptique. Ces travaux font l'objet d'une publication en cours de soumission qui est présentée en annexe (Carpentier *et al*). J'ai également participé à l'étude de nombreux mutants de la transposase, étude qui s'inscrit dans un projet plus vaste d'ingénierie de MOS1 ; ces travaux sont publiés dans *Genetica* présentée en annexe (Germon *et al*, 2009).

La **deuxième partie** est consacrée à la mécanique de transposition de l'élément *mariner Mos1*. En me basant sur des résultats précédemment obtenus au laboratoire, mais non publiés, mon travail a initialement consisté à trouver les différences qui pouvaient exister tant sur le plan structural que sur un éventuel impact sur le mécanisme de transposition, entre une transposase MOS1 produite en système procaryote ou en système eucaryote. Les résultats obtenus lors de ces travaux, mis en relation avec la cristallographie de la partie catalytique de MOS1 (Richardson *et al*, 2006), m'ont conduit à m'interroger sur la présence de ponts disulfures chez la Tnp de *Mos1*, ces liaisons pouvant avoir une importance dans la mécanique de la transposition. Cette étude comparative m'a également amené à étudier en détail la formation du complexe synaptique et à remettre en cause certains résultats précédemment publiés.

Enfin la **troisième partie** décrit la découverte d'inhibiteurs de la transposase de *Mos1* et la caractérisation de leur mode d'action. Les inhibiteurs présentent un grand intérêt dans le cadre de l'étude du mécanisme de transposition avec la possibilité de bloquer la réaction à l'étape voulue. Ils peuvent également avoir une importance en cas d'utilisation du système transposon/transposase en transfert de gènes. Ces travaux font l'objet d'un article présenté en annexe, qui a été publié dans le journal *Mini Reviews of Medicinal Chemistry*.

1/ *Mos1* : la transposase

Ce premier chapitre des résultats est consacré plus spécifiquement à la transposase en tant que protéine. Il sera divisé en deux parties : la première traite du domaine N-terminal de la protéine qui contient le domaine de liaison à l'ADN, et la seconde présente des travaux d'ingénierie de la transposase dont le but était de produire des mutants hyperactifs. Ces deux thématiques font l'objet de deux publications (Carpentier *et al*, soumis ; Germon *et al*, 2009) présentées en annexe.

1.1 Le domaine de liaison à l'ADN : un ou deux HTH ?

Le projet d'étude du domaine de liaison à l'ADN a été mené au laboratoire par le Dr Guillaume Carpentier ; j'y ai contribué activement d'un point de vue théorique et technique et ai participé à la rédaction de l'article qui a synthétise ces travaux (Carpentier *et al*, en cours de soumission).

1.1.1 Un ou deux HTH : point sur la controverse et conséquences sur l'assemblage du complexe synaptique

Les transposases eucaryotes sont constituées de deux domaines principaux : un domaine de liaison aux ITR (IBD pour *ITR-Binding Domain*) et un domaine catalytique. Dans la famille *ITm*, les domaines catalytiques des *MLE* et des *TLE* sont similaires, contrairement aux domaines de liaison à l'ADN qui ne présentent aucune similitude de séquence.

Chez les *TLE*, les séquences et les analyses structurales ont montré que l'IBD était composé de deux motifs HTH (Hélice-Tour-Hélice). Ces motifs sont impliqués dans la liaison à l'ITR et apparentés au domaine de liaison à l'ADN (DBD pour *DNA-Binding Domain*) des protéines PAIRED-*like*. Ces protéines contiennent en effet deux motifs HTH impliqués dans la liaison à l'ADN : PAI est un sous-domaine de reconnaissance de l'ADN et RED a un rôle de stabilisateur de l'interaction avec l'ADN. Les données cristallographiques montrent que les transposases avec un IBD de type PAIRED interagissent avec l'ITR sous forme monomérique.

Chez les *MLE*, aucune structure cristallographique du domaine N-terminal n'a été résolue. Deux modèles font actuellement débat parmi les spécialistes des éléments *ITm* :

- *TLE* et *MLE* utilisent le même mode de liaison à l'ADN car ils appartiennent à la même superfamille d'éléments transposables. Ce modèle propose que l'IDB des éléments *mariner* soit organisé comme un domaine RED-PAI, similaire au domaine PAIRED, mais dans une orientation inversée (Richardson *et al*, 2006).

- le second modèle propose que les *MLE* contiennent un domaine de type CRO, c'est-à-dire un sous-domaine de dimérisation fonctionnel relié à un seul motif HTH impliqué dans la liaison à l'ITR. Ce modèle s'appuie sur le fait qu'un seul motif HTH est prédit entre les positions 72 et 110 pour MOS1 (Augé-Gouillou *et al*, 2001a ; Brillet *et al*, 2007 ; Feschotte, 2008 ; Pietrokovski & Henikoff, 1997 ; Zhang *et al*, 2001).

Cette controverse entre les tenants de l'un ou l'autre modèle n'est pas que théorique puisque le mode de liaison de la transposase conditionne la structure du complexe synaptique. En effet, un domaine de type RED-PAI implique une liaison sous forme monomérique de la transposase à son ITR, alors que la présence d'un domaine de type CRO implique une liaison à l'ITR sous forme dimérique. Les données de la littérature vont dans le sens de l'hypothèse CRO. Les protéines de type CRO interagissent avec l'ADN sous forme de dimères (Darling *et al*, 2000a et b) et leur cible contient un motif palindromique (Albright & Matthews, 1998). Ces deux caractéristiques ont été décrites pour les éléments *mariner* (interaction sous forme de dimère – Augé-Gouillou *et al*, 2005a et Zhang *et al*, 2001 pour *Mos1* ; Buttler *et al*, 2006 pour *Himar1* ; séquence palindromique – Bigot *et al*, 2005)

Pour démontrer que les *MLE* interagissaient avec l'ADN *via* un domaine de type CRO, et donc un seul motif HTH, nous avons utilisé des transposases chimériques pour mettre en relation la structure et la fonction du domaine N-terminal de la protéine.

1.1.2 Transposases chimériques

Le sous-domaine de dimérisation de MOS1 (les 34 premiers acides aminés) a été remplacé par des domaines de dimérisation totalement étrangers au monde des éléments transposables. Il s'agit de deux domaines qui ne présentent aucune propriété de liaison à l'ADN appartenant soit à la protéine de levure Gal4 (acides aminés 41 à 100 ; Li *et al*, 2007),

soit à la protéine de levure GCN4 (motif *leucine zipper* (LZ) ; acides aminés 253 à 278 ; Feldmann *et al*, 1993). Les protéines chimériques ainsi obtenues ont été nommées Gal4[35-345] et LZ[34-345] (figure 46).

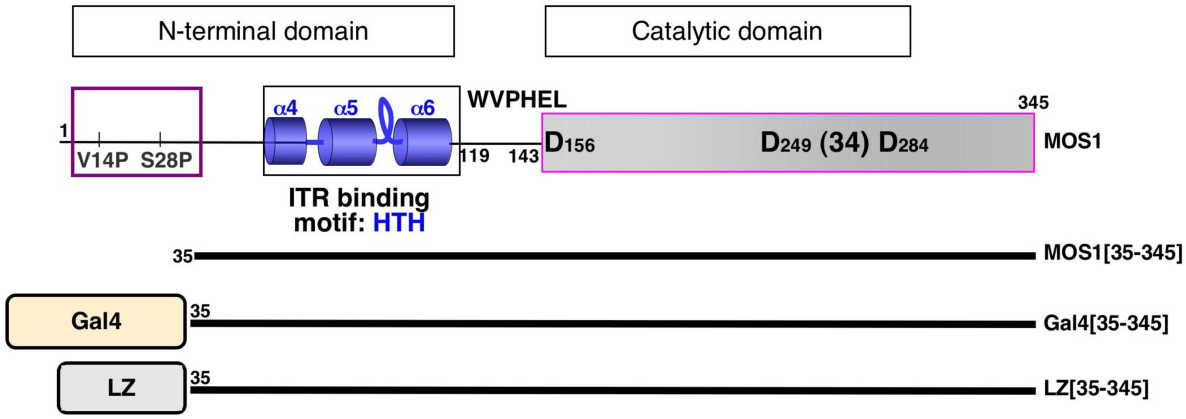


Figure 46. Protéines chimériques LZ[34-345] et Gal4[35-345], d'après Carpentier *et al*, soumis. *MOS1 est une protéine de 345 acides aminés; les barres noires représentent la partie tronquée de MOS1 et les rectangles colorés représentent les domaines de dimérisation fusionnés : Gal4 (59 acides aminés) en jaune et LZ (25 acides aminés) en gris. Le domaine de dimérisation de Mos1 est encadré en violet.*

L'hypothèse est que dans le cas d'un domaine de type PAIRED-*like* chez MOS1, ces protéines de fusion seraient incapables de promouvoir la transposition. Tandis que s'il s'agit d'un domaine de type CRO, la dimérisation serait prise en charge par les domaines de fusion, et la liaison à l'ADN par le motif HTH ; il y aurait donc transposition.

La capacité de ces protéines à promouvoir ou non la transposition de l'élément *Mos1* a été vérifiée par le test de transposition en bactérie. Ce test est le plus sensible dont nous disposons au laboratoire pour étudier la transposition. Son principe et son protocole sont présentés dans la partie Matériel et Méthodes. La présence indispensable des 34 premiers acides aminés de la protéine pour promouvoir la transposition a été vérifiée en utilisant d'une part une protéine tronquée de ses 34 premiers acides aminés (MOS1 [35-345]) et d'autre part deux mutants ponctuels du domaine de dimérisation : V14P et S28P (une proline a remplaçant la valine en position 14 et la sérine 28). Aucun évènement de transposition n'a été détecté avec ces trois protéines mutantes (tableau 1). Par contre, l'ajout des domaines de dimérisation étranger que sont LZ et Gal4 restaure la transposition. Toutefois les fréquences sont plus faibles qu'avec la protéine native (tableau 1) ce qui n'est pas surprenant étant donné que la

transposition requiert une organisation précise du complexe synaptique, qui n'est pas garantie par LZ et Gal4.

Transposase	Transposition frequency
no	Undetected*
MOS1	10^{-4} [$3x10^{-5}$]
V14P	Undetected*
S28P	Undetected*
[35-345]	Undetected*
Gal4-[35-345]	$6.8x10^{-9}$ [$7x10^{-9}$]
LZ-[35-345]	$2x10^{-8}$ [$4x10^{-9}$]

*Sensitivity threshold: $2x10^{-10}$

Tableau 1. Fréquence de transposition des transposases tronquées, d'après Carpentier *et al*, soumis. *Les fréquences de transposition indiquées sont la médiane d'au moins cinq tests indépendants ; les premiers quartiles sont notés entre crochets. La sensibilité du test est de $2x10^{-10}$.*

La transposition promue par les transposases tronquées LZ[34-345] et Gal4[35-345] a été analysée par restriction enzymatique et séquençage pour vérifier que l'insertion du transposon avait bien eu lieu au niveau d'un dinucléotide TA (figure 47) et qu'il était bien dupliqué après l'intégration. En effet, ce dernier point est la signature de transposition des éléments *mariner*.

Nos résultats indiquent que la substitution des 34 premiers acides aminés de MOS1 par des domaines de dimérisation étrangers n'abolit pas la transposition. Ils indiquent également que le motif HTH est capable d'assurer seul la liaison de la protéine à l'ITR, et de promouvoir précisément la suite de la réaction de transposition : transfert de brin, insertion au niveau d'un dinucléotide TA et duplication de ce dernier.

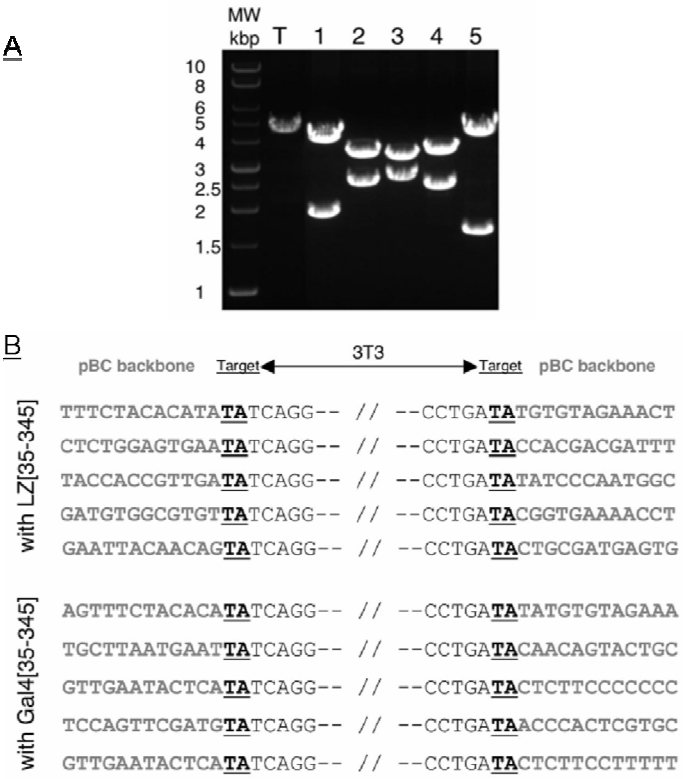


Figure 47. Analyse de la transposition *in vivo* des protéines tronquées, d'après Carpentier *et al*, soumis. (A) *Analyse des produits de transposition de Gal4[35-345] ; après purification, les plasmides sont digérés par HindIII et analysés par électrophorèse sur un gel d'agarose 1%. Le pBC3T3 est le témoin (T) et ne contient qu'un seul site de digestion par HindIII. Quand un pBC3T3 est utilisé comme une cible pour l'intégration, la digestion par HindIII donne deux bandes. Les lignes 1 à 5 sont des produits de transposition issus de cinq expériences indépendantes.* **(B)** *Séquences des sites d'insertion. Le dinucléotide TA est retrouvé flanquant tous les pseudo-transposons insérés, ce qui est la signature d'une vraie transposition.*

1.1.3 Conclusion

Les résultats présentés ici indiquent que le domaine de liaison à l'ITR des *MLE* est un domaine de type CRO, démontrant ainsi qu'un seul HTH est requis pour la transposition des *MLE*. L'hypothèse de la liaison d'un dimère de transposase à un ITR est donc démontrée. Ceci implique donc que le complexe synaptique de *Mos1* corresponde à l'assemblage de deux dimères soit un tétramère de Tnp, hypothèse déjà avancée par des études du mécanisme de transposition de *Mos1* (Augé-Gouillou *et al*, 2005b) et pour l'élément *Himar1* (Buttler *et al*, 2006). Une dernière question est de savoir si la dimérisation de MOS1 se fait avant la liaison à l'ITR ou sur l'ITR.

1.2 Ingénierie de MOS1

Le travail d'ingénierie de la transposase de *Mos1* s'inscrit dans le projet européen *SyntheGeneDelivery* dont le laboratoire était porteur et qui avait pour objectif la mise au point d'un système de transfert de gènes en cellules humaines. Actif chez les invertébrés (insectes et nématodes) et les vertébrés à sang froid, l'élément *mariner Mos1* s'est avéré être inactif dans les génomes de mammifères où il a été testé. Cela pose un problème majeur dans l'optique d'utiliser cet élément comme vecteur de transgénèse, et notamment pour développer un vecteur non-viral de transfert de gènes en cellules humaines. Pour contourner le problème causé par l'inactivité de *Mos1* dans les génomes de mammifères, la stratégie adoptée a été de mettre au point une transposase hyperactive et non phosphorylable.

Ces travaux font notamment l'objet d'un article publié dans la revue *Genetica* (Germon *et al*, 2009).

1.2.1 Transposases hyperactives

Le projet des transposases hyperactives a été mené au laboratoire par le Dr Stéphanie Germon ; j'y ai contribué activement d'un point de vue théorique et technique et ai participé à la rédaction de l'article qui a résumé ces travaux (Germon *et al*, 2009).

1.2.1.1 Etude préliminaire

Une approche rationnelle a été choisie dans cette étude en vue d'obtenir des Tnp hyperactives. A partir des données de la littérature disponibles sur les transposases des éléments *mariner Mos1* et *Himar1* (Lampe *et al*, 1999 ; Zhang *et al*, 2001 ; Butler *et al*, 2006 ; Keravala *et al*, 2006), plusieurs approches ont été développées pour trouver des mutations ponctuelles qui rendaient la Tnp hyperactive. Ces mutations uniques ont ensuite été combinées pour obtenir la meilleure combinaison possible.

Au final, une trentaine de mutants ont été analysés en test de transposition en bactérie. Seuls cinq ont fourni des données intéressantes (figure 48). Les mutants Q91R et T216A ont un impact positif sur la transposition de *Mos1* tandis que les mutants F53Y, E137K et Y237C ont un taux de transposition identique à celui obtenu avec la protéine sauvage. Nous avons vérifié par séquençage que ces mutants ponctuels faisaient de la transposition vraie (signature TA aux extrémités du transposon).

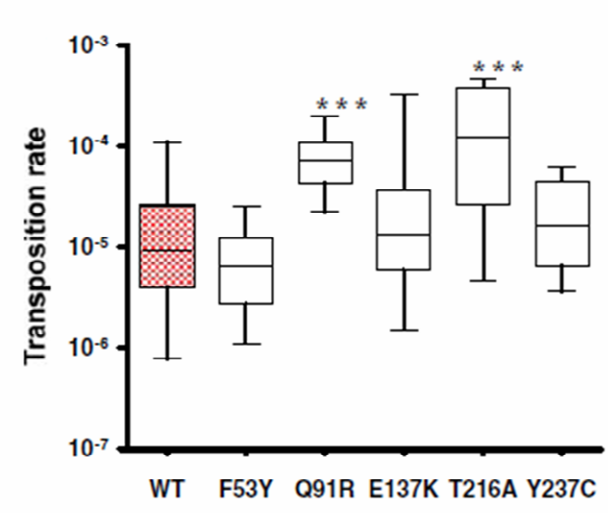


Figure 48. Fréquence de transposition des mutants ponctuels retenus, d'après Germon *et al*, 2009. *Les valeurs obtenues sur au moins trois clones de trois tests indépendants ont été représentés en diagramme de Tukey dit en boites-et-moustaches (logiciel GraphPad Prism). La ligne au milieu des rectangle représente la médiane, le haut et le bas du rectangle les troisième et premier quartiles, tandis que les 'moustaches' indiquent les valeurs minimale et maximale obtenues pour chaque mutant. La protéine sauvage est indiquée en rouge (WT = wild type). Les mutants Q91R et T216A présentent des différences significatives avec la protéine sauvage (***) d'après un test statistique de Mann & Whitney (P<0,05), les autres non.*

1.2.1.2 Mutants multiples : tests de transposition *in vivo*

Des mutants multiples ont été réalisés à partir de ces cinq mutants sélectionnés dans le but de trouver une combinaison qui augmenterait de manière importante la fréquence de transposition. Des combinaisons de doubles à quintuples mutants de MOS1 ont été produites.

Seules sept mutants[1] ont présenté un intérêt : FT, ET, TY, FTY, FQET, FETY et FQETY. Toutes contiennent la mutation T216A et conduisant à une augmentation d'au moins 20 fois la fréquence de transposition par rapport à la protéine sauvage (figure 49).

[1] Les différentes combinaisons de mutants ont été nommées avec les codes 'une lettre' des acides aminés mutés ; par exemple le mutant F53Y/T216A est nommé FT.

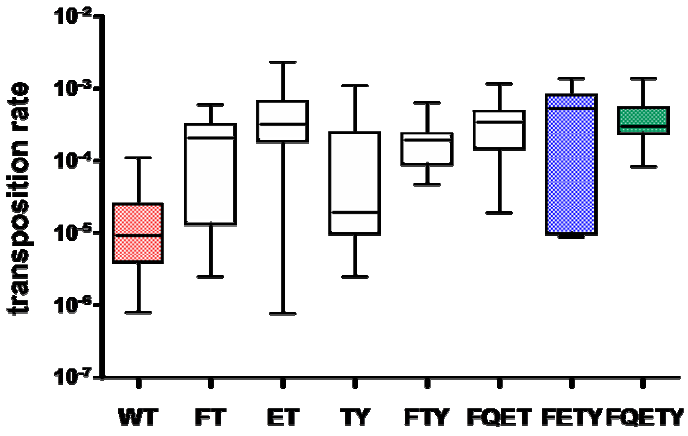


Figure 49. Fréquence de transposition des sept mutants multiples d'intérêt, d'après Germon *et al*, 2009. *Les valeurs obtenues sur au moins trois clones de trois tests indépendants ont été représentés en diagramme de Tukey dit en boites-et-moustaches (logiciel GraphPad Prism). La ligne au milieu des rectangle représente la médiane, le haut et le bas du rectangle les troisième et premier quartiles, tandis que les 'moustaches' indiquent les valeurs minimale et maximale obtenues pour chaque mutant. La protéine sauvage est indiquée en rouge (WT = wild type), MOS1 [F53Y / E137K / T216A / Y237C] en bleu (FETY) et le quintuple mutants MOS1 [F53Y / Q91R / E137K / T216A / Y237C] (FQETY) en vert. Ces sept mutants présentent des différences significatives (P<0,05) avec la protéine sauvage d'après un test statistique de Mann & Whitney.*

Outre ces sept combinaisons, le mutant QET est une protéine incapable de promouvoir la transposition, ceci démontrant l'antagonisme de certaines mutations. La protéine FET s'est révélée quant à elle être un cas déconcertant donnant selon les tests une perte presque totale de l'activité de transposition ou au contraire une augmentation importante de la fréquence de transposition (de 200 à 800 fois). Ces résultats fluctuants sont certainement dus à la toxicité de la transposase FET. La viabilité des bactéries lors du test a été mesurée, par calcul de la concentration de bactéries à la fin des tests de transposition *in vivo*, et en regardant s'il y avait une corrélation entre cette concentration en bactéries et l'activité des transposases sauvage, FET, FETY et FQETY (figure 50). Le résultat est que la fréquence de transposition de la protéine sauvage n'a aucun lien avec la concentration des bactéries en fin de test de transposition, ce qui montre que MOS1 n'est pas cytotoxique. Les protéines mutantes FETY et FQETY présentent une très faible cytotoxicité. Le comportement de la FET est différent : sa fréquence de transposition est inversement corrélée à la concentration en bactéries en fin de test. En effet quand la FET promeut de la transposition, la concentration en bactéries à la fin du test est 5000 à 10000 fois inférieure à celle obtenue avec la protéine sauvage. L'expression de la FET a donc un effet toxique sur les bactéries.

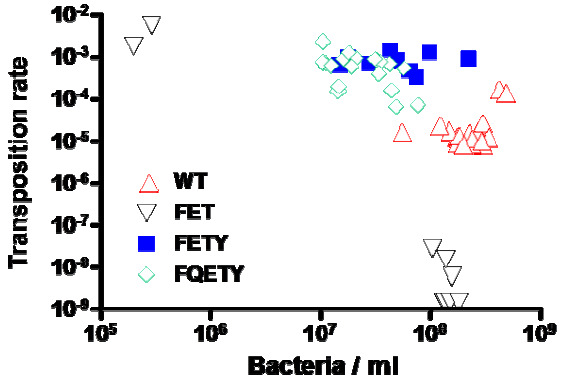


Figure 50. Analyse cytotoxique avec les mutants multiples, d'après Germon *et al*, 2009. *Le taux de transposition est représenté en fonction de la concentration en bactéries, chaque point correspondant à une analyse indépendante. La Tnp sauvage est représentée en rouge, la FET en noire, la FETY en bleu, la FQETY en vert.*

1.2.1.3 Mutants multiples : tests de transposition *in vitro*

En 1999, David Lampe montrait que les mutants hyperactifs de HIMAR1 étaient moins actifs *in vitro* qu'*in vivo* (Lampe *et al*, 1999). Les protéines FET et FETY ont été produites et purifiées (en fusion avec la MBP, comme pour la Tnp de *Mos1* utilisée usuellement au laboratoire) et testées en transposition *in vitro*. Ce test permet de déterminer une fréquence de transposition propre à chaque protéine, en utilisant un pseudo-transposon *Mos1* ; ce pseudo-transposon est porteur du gène de résistance à la tétracycline sans son promoteur à la place du gène codant la transposase. La fréquence de transposition est déterminée par l'acquisition de la résistance à la tétracycline (le principe du test est expliqué dans le Matériel et Méthodes). A toutes les concentrations testées (de 1 à 300 nM), la protéine sauvage est plus active que les deux mutantes (figure 51), ceci concordant avec ce qui avait été décrit pour HIMAR1.

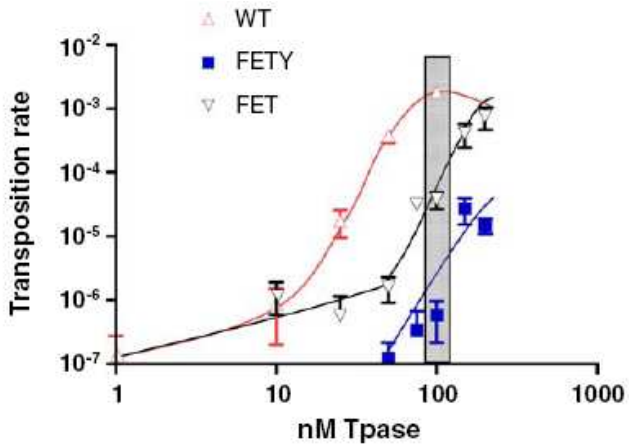


Figure 51. Transposition *in vitro* de MOS1, et des mutants FET et FETY, d'après Germon *et al*, 2009. *Le taux de transposition est en fonction de la concentration en protéine. La protéine sauvage (WT) est représentée par les triangles rouges, la FETY par les carrés bleus, la FET par les triangles noirs. Les médianes et les quartiles ont été calculés à partir de données provenant d'au moins cinq expériences indépendantes. A 100 nM (en gris), la fréquence de transposition des protéines mutantes est significativement différente (P<0,05) de celle de la protéine sauvage.*

En plus d'être hyperactive, la version de la transposase de *Mos1* qui pourra être utilisée comme vecteur de transgénèse en cellules de mammifères ne devra pas être sensible aux modifications post-traductionnelles. La seconde étape de l'ingénierie a été de concevoir une protéine mutante non-phosphorylable. Ces travaux ont été commencés avant mon arrivée au laboratoire, avant même d'avoir la preuve formelle que la transposase produite en système eucaryote était phosphorylée, et ont été poursuivis en parallèle des travaux portant sur la recherche d'un lien entre la phosphorylation et l'activité de la transposase.

Je décris ces travaux dans la seconde partie des résultats, dans le chapitre consacré à la phosphorylation de la transposase produite en système eucaryote.

1.2.2 Bilan de l'ingénierie de la transposase de *Mos1*

Les travaux d'ingénierie de la transposase MOS1 menés depuis plusieurs années au laboratoire ont permis l'obtention d'une très vaste collection de mutants de toute la protéine ; j'ai répertorié tous ces mutants, ainsi que ceux trouvés dans la littérature pour en faire un 'catalogue' présentant toutes les caractéristiques connues de chacun (ce 'catalogue' est présenté en annexe).

Ces travaux ont permis d'identifier des mutants hyperactifs de MOS1, avec toutefois une contrainte importante puisque ces mutants, à des degrés divers, sont cytotoxiques. Cette cytotoxicité, en plus d'induire une sous-estimation de la fréquence de transposition des mutants lors des tests de transposition *in vivo*, fait qu'il est délicat de produire correctement ces protéines mutantes en bactéries. Cela rend par conséquent le test de transposition *in vitro* inadéquat pour étudier l'hyperactivité des transposases mutantes produites en bactéries.

La production de transposases hyperactives par mutagenèse rationnelle est possible à partir d'une séquence de transposase naturellement existante. Toutefois, la barrière de l'espèce est un facteur à ne pas négliger. *Mos1* est très peu actif chez de nombreux invertébrés chez lesquels il a été testé, et ne transpose pas en cellules humaines. De nombreux facteurs peuvent avoir un impact sur l'activité de la transposase et la conception d'une transposase hyperactive dans un système ne signifie pas qu'elle le sera dans un autre système. Les études menées au laboratoire et celles présentées dans la littérature sur *Mos1* (Pledger & Coates, 2005) et sur *Himar1* (Lampe *et al*, 1999) laissent penser qu'une transposase hyperactive doit être produite dans la lignée cellulaire où elle va être utilisée. Cela semble également indiquer

que *Mos1* pourrait être sensible à certains facteurs d'hôte contrairement aux idées répandues dans la communauté des spécialistes des transposons eucaryotes.

La question de la pertinence du développement de transposases hyperactives sera traitée dans la partie Discussion & Perspectives.

2/ *Mos1* : Transposition et régulation

Bien que leurs hôtes naturels soient des arthropodes, donc des eucaryotes, les transposases des éléments *mariner Mos1* et *Himar1*, dont le mécanisme de transposition est le mieux caractérisé parmi les éléments *mariner*, ont toujours été étudiées après avoir été produites en système procaryote, en l'occurrence en bactérie (Lampe *et al*, 1996 et 1998 ; Lipkow *et al*, 2004a et b; Augé-Gouillou *et al*, 2005a et b). Même la cristallographie du domaine C-terminal de MOS1 est faite à partir d'un mutant de la protéine produit en bactérie (Richardson *et al*, 2006).

L'intérêt de produire les protéines en bactérie est évident : c'est simple et peu coûteux. Néanmoins, les caractéristiques physiologiques des bactéries sont différentes de celles des cellules eucaryotes, et la différence majeure entre une protéine produite en système procaryote et celle produite en système eucaryote portera sur les modifications post-traductionnelles. Bien qu'existantes en bactérie, les modifications majeures que sont la phosphorylation et la glycosylation (Kennelly, 2002 ; Abu-Qarn *et al*, 2008) n'interviennent pas sur les protéines exogènes produites dans la bactérie. Il peut donc être admis que l'étude d'une protéine eucaryote produite en système procaryote est susceptible de ne pas refléter ce qui se passe exactement chez l'hôte initial. C'est donc le cas pour la transposase de *Mos1*, lorsqu'elle est étudiée sous une forme 'bactérienne' alors que son hôte naturel est la drosophile. Un argument qui va dans ce sens est que la transposition de *Mos1* est peu efficace voire inexistante dans les organismes eucaryotes où *Mos1* est utilisé comme vecteur de transfert de gènes.

Le système d'expression baculovirus / cellules d'insectes s'est imposé de lui-même pour produire une version 'eucaryote' de la transposase de *Mos1*. Il assure des modifications post-traductionnelles et le *folding* des protéines complexes.

Des études préliminaires avaient déjà été menées au laboratoire, suite à la production de MOS1 en cellules d'insectes Sf21. Des différences au niveau de la liaison de la Tnp à ses ITR avaient été observées entre la protéine produite en système bactérien, qui sera nommée par la suite protéine 'procaryote' ou 'bactérienne', et celle produite en système baculovirus / cellules d'insectes, nommée par la suite protéine 'eucaryote' ou 'baculovirale'. De plus, il

avait été montré que la protéine eucaryote pouvait être phosphorylée *in vitro* ; ce résultat a d'ailleurs été publié dans un article présenté en annexe (Germon *et al*, 2009 – figure 52).

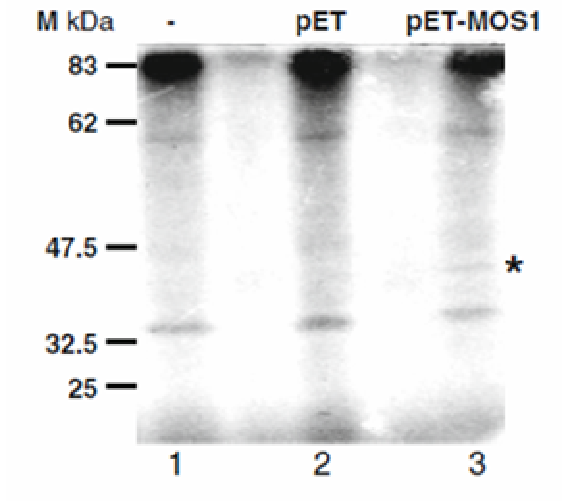


Figure 52. Phosphorylation de MOS1 *in vitro*, d'après Germon *et al*, 2009. *Analyse en gel SDS-PAGE d'échantillons d'une réaction couplée de transcription/traduction in vitro en cellules eucaryotes. (M) marqueur de poids moléculaire, (1) sans plasmide, (2) plasmide vide, (3) plasmide codant la Tnp. La synthèse a été faite en présence de $\gamma^{32}P$; le résultat est que la Tnp de Mos1 est faiblement phosphorylée (*).*

Ces résultats préliminaires nous ont poussés à poursuivre l'étude de la protéine eucaryote tout en la comparant à la protéine procaryote. En effet la protéine eucaryote est susceptible d'être modifiée de manière post-traductionnelle et cela pourrait avoir un impact sur le mécanisme de transposition et remettre en cause les connaissances acquises jusqu'alors avec la protéine procaryote.

2.1 Transposase procaryote versus transposase eucaryote

La transposase de *Mos1* fusionnée à la MBP (la protéine MBP-MOS1 sera nommée par la suite transposase ou MOS1, le tag étant toujours présent sauf mention) a été produite dans le système baculovirus / cellules d'insectes Sf21 au laboratoire par le Dr Guillaume Carpentier, grâce au système *BaculoGold* (*BD Sciences*). Le même protocole a été appliqué par la suite aux mutants de MOS1 produits en système eucaryote.

La première différence qui apparaît entre les deux protéines bactérienne et baculovirale se situe au niveau de la stabilité de la transposase. La protéine bactérienne n'est en effet pas stable, et seulement 70% de la protéine est sous forme pleine taille après purification. Les 30% restants apparaissent sous forme de produits de protéolyse caractérisés précédemment dans notre laboratoire (Corinne Augé-Gouillou, communication personnelle). La protéolyse (ou auto-protéolyse ?) se produit au niveau des acides aminés 80/81 d'une part et au niveau des acides aminés 101/102 d'autre part. Ces produits de protéolyse ne sont pas retrouvés après purification de la protéine produite en système eucaryote (figure 53) ; la

transposase apparaît alors sous sa forme pleine taille uniquement. Il est envisageable qu'une modification post-traductionnelle non présente chez les bactéries soit à l'origine de cette stabilité de la Tnp eucaryote.

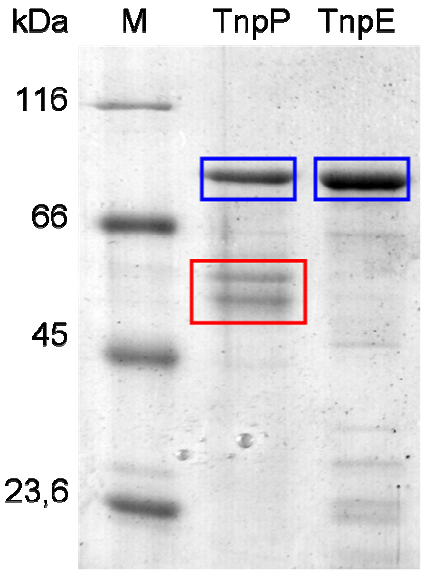


Figure 53. Stabilité de la protéine eucaryote. *(M = marqueur de poids moléculaire). Analyse en gel SDS-PAGE d'échantillons de Tnp produites soit en bactérie (TnpP) soit en système baculovirus / cellules d'insectes Sf 21 (TnpE). La MBP-Tnp plaine taille est encadrée en bleu, les produits de protéolyse en rouge.*

Une fois la différence de stabilité entre les deux protéines observée, leur activité de transposition a été comparée par un test de transposition *in vitro*, comme décrit précédemment. La conclusion est qu'il n'y a aucune différence entre les fréquences de transposition des protéines bactérienne et baculovirale. Les deux présentent une fréquence de transposition de 10^{-4} +/- [5 x 10^{-5}] lors de la réaction classiquement utilisée au laboratoire (30 minutes), et de 10^{-3} +/- [5 x 10^{-4}] lors d'une réaction de 4 heures. Cette absence de différence a été vérifiée par un test statistique de Mann & Whitney, réalisé avec 15 échantillons indépendants.

L'absence de différence sur le taux de transposition ne laissait pas entrevoir de différence de comportement des deux transposases dans leur mécanisme de transposition. Mais une différence est apparue en examinant leur mode de liaison à l'ADN. C'est par des expériences de retard sur gel, en utilisant des ITR 3' marqués au γ^{32}P, qu'est visualisée la fixation de la Tnp à ses ITR. Dans des conditions particulières d'incubation (15 minutes à 4°C en présence de magnésium), la Tnp produite en bactérie forme deux complexes de façon très majoritaire. L'un est identifié comme un SEC2 (*Single End Complex 2* : un ITR et deux Tnp) et l'autre comme un PEC1 (*Paired End Complex 1* : deux ITR et deux Tnp) (Augé-Gouillou *et al*, 2005b ; figure 54). Dans les mêmes conditions la protéine eucaryote ne forme quant à elle qu'un seul complexe, initialement nommé complexe A (figure 54).

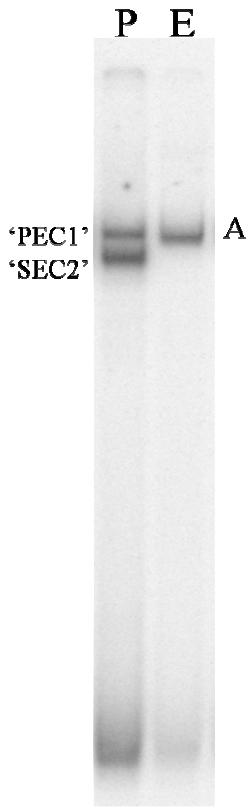
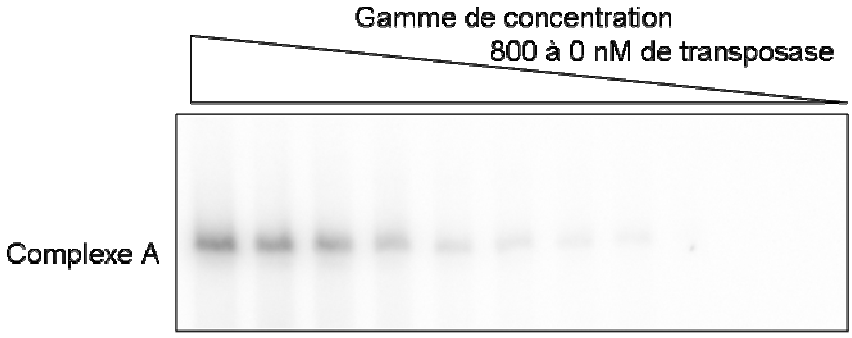


Figure 54. Profil de liaison[2] à l'ADN des transposases procaryotes et eucaryotes. *Retard sur gel en utilisant une sonde ITR 3' radiomarquée par du $\gamma^{32}P$; (P) Tnp 'procaryote', (E) Tnp 'eucaryote'. Les complexes sont formés à 4°C pendant 15 minutes. La Tnp produite en bactérie forme les deux complexes attendus, décrits comme un SEC2 et un PEC1 ; la protéine produite en système eucaryote ne forme elle qu'un seul complexe appelé A.*

Cette différence qualitative n'est pas la seule : l'affinité des deux Tnp pour les ITR n'est pas la même. Des gammes de concentration des deux protéines ont été effectuées et la liaison visualisée en retard sur gel. Quand la concentration de Tnp augmente, la fixation de la protéine eucaryote sur ses ITR est meilleure que celle de la protéine procaryote (figure 55). L'hypothèse la plus évidente est que la Tnp procaryote a une tendance plus importante à s'oligomériser quand la concentration augmente, ce qui la rend moins disponible pour la liaison.

[2] Cette première image de retard sur gel est présentée complète, les suivantes seront présentées sans les sondes libres, si cela n'affecte pas la qualité de l'image et son explication.



	[Tnp] (nM)	0	5	10	20	30	40	50	100	200	400	800
% de liaison	TnpP	0	1	1,1	1,3	1,4	1,5	2	7	12	31	48
	TnpE	0	1	1,1	1,3	1,5	1,6	2,2	8	16	40	78

Figure 55. Liaison aux ITR - gammes de transposases. *Expérience de retard sur gel en utilisant une sonde ITR 3' radiomarquée par du $\gamma^{32}P$ et des quantités variables de transposase. La Tnp et l'ITR radiomarqué ont été incubés 15 minutes à 4°C en présence de Mg^{2+}. Le gel présenté en exemple est celui de la protéine 'baculovirale', avec une concentration décroissante de Tnp (les valeurs sont dans le tableau du dessous). Le tableau représente le pourcentage de liaison de la Tnp (TnpP procaryote ou TnpE eucaryote) aux ITR. A forte concentration (200 nm et plus), la protéine eucaryote apparaît plus efficace pour lier l'ITR que la protéine procaryote. Cette différence est d'un facteur 2 à 800 nM.*

La différence de profil en gel retard a conduit à caractériser le complexe formé par la protéine eucaryote, toujours en comparant avec le profil de liaison de la protéine procaryote ; ce travail sera détaillé un peu plus loin.

La conclusion de cette étude préliminaire est l'absence de différence entre une Tnp procaryote et une Tnp eucaryote pour la transposition dans son ensemble. La seule différence observée est le profil de liaison de la protéine à ses ITR, ce qui peut paraître anodin puisque cela n'influence pas l'activité globale de transposition. Néanmoins examiner en détails cette différence de profil dans le but d'identifier le complexe A (combien de Tnp ? combien d'ITR ?) m'a conduit à remettre en cause des données publiées dans la littérature et à émettre de nouvelles hypothèses sur la formation du complexe synaptique et du transpososome.

2.2 Mécanique de transposition

2.2.1 SEC et PEC

Les données concernant la formation du complexe synaptique précédemment obtenues au laboratoire et publiées (Augé-Gouillou *et al*, 2005b) montrent que la transposase procaryote est capable de former quatre complexes distincts visualisables en retard sur gel. Il s'agit du SEC1 (une Tnp + un ITR), du SEC2 (deux Tnp + un ITR), du PEC1 (deux Tnp + deux ITR) et du complexe synaptique PEC2 (quatre Tnp et deux ITR) (figure 56).

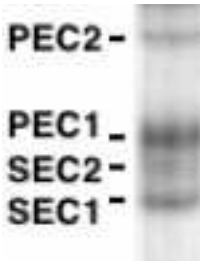


Figure 56. Profil de liaison de la transposase procaryote sur ses ITR, d'après Augé-Gouillou *et al*, 2005b. *La visualisation de ces quatre complexes est possible dans des conditions spécifiques non catalytiques en présence d'EDTA. La même expérience réalisée en présence de magnésium fait ressortir deux complexes majoritaires : le SEC2 et le PEC1 (figure 54).*

La présence de deux transposases bactériennes dans les complexes SEC2 et PEC1 a été confirmée en utilisant le fait que la protéine de fusion porte un site de clivage au facteur Xa situé entre la partie MBP et la partie MOS1. Les complexes ITR-Tnp ont été assemblés comme précédemment décrit (Augé-Gouillou *et al*, 2005b). Une cinétique de clivage au facteur Xa a ensuite été menée à 25°C (20, 60 et 150 minutes). Si le SEC2 et le PEC1 contiennent bien deux Tnp, un profil de trois groupes de bandes est attendu en retard sur gel : des bandes contenant le SEC2 et le PEC1 avec uniquement des MBP-MOS1, des bandes contenant dans le même complexe une MBP-MOS1 et une MOS1 sans le tag MBP, et des bandes contenant uniquement deux transposases sans MBP. C'est ce qui a été trouvé ici, confirmant donc la présence de deux transposases dans les complexes nommés SEC2 et PEC1 (figure 57).

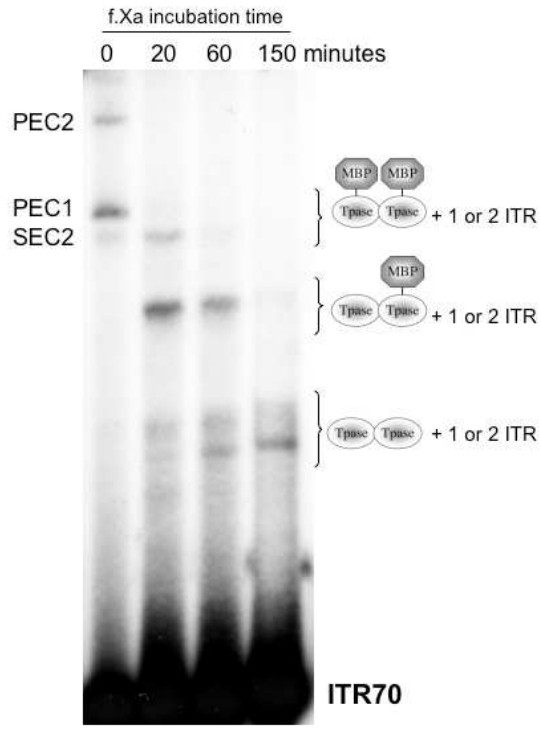


Figure 57. Un dimère de transposase procaryote interagit avec les ITR de *Mos1*. *Le clivage de la MBP a été mené une fois les complexes formés sur l'ITR (après 20 minutes d'incubation à 4°C en présence de magnésium). Le profil obtenu confirme la présence de deux Tnp dans les SEC2 et PEC1*

Les complexes ont été assemblés 20 minutes à 4°C en présence de magnésium, en utilisant 10 nM de sonde (ITR 3') et 400 nM de Tnp.

2.2.2 Stœchiométrie du complexe A

La première étape de caractérisation du complexe A formé par la protéine eucaryote a été de montrer que ce complexe contenait deux transposases. La méthode utilisée est celle décrite à l'instant pour l'identification des complexes SEC2 et PEC1 avec la protéine procaryote. Les complexes ITR-Tnp ont été formés, et une cinétique de clivage au facteur Xa a été effectuée. Le même profil à trois bandes que pour la protéine procaryote a été obtenu (figure 58) ; le complexe A contient deux transposases eucaryotes.

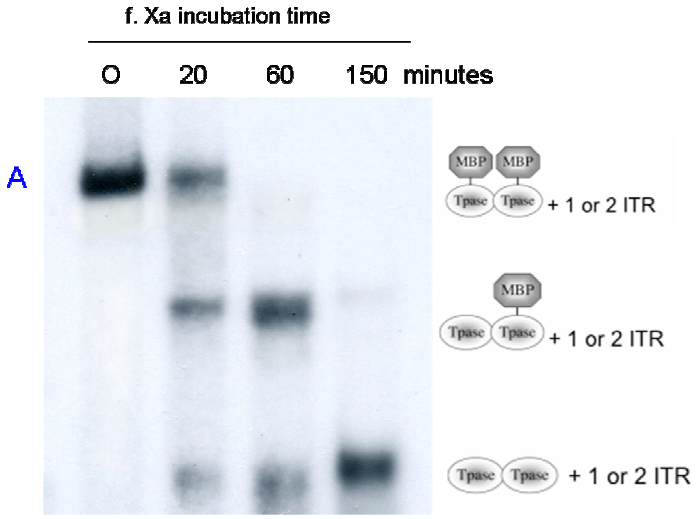


Figure 58. Le complexe A contient deux transposases. *Le clivage de la MBP a été mené une fois les complexes formés sur l'ITR (20 minutes d'incubation à 4°C en présence de magnésium,). Le profil obtenu confirme la présence de deux transposases dans le complexe A formé par la protéine eucaryote.*

Deux transposases donc dans ce complexe A, mais combien d'ITR ?

Pour déterminer le nombre d'ITR, une méthode dite ITR court/long a été utilisée. L'idée est de former les complexes avec deux ITR différents, un court (C) et un long (L), radiomarqués (*) ou non. Les complexes sont formés avec les combinaisons suivantes d'ITR : C*, C*/L, C/L* et L*. En présence d'un complexe contenant deux ITR, un profil dit en escalier doit être observé (figure 59).

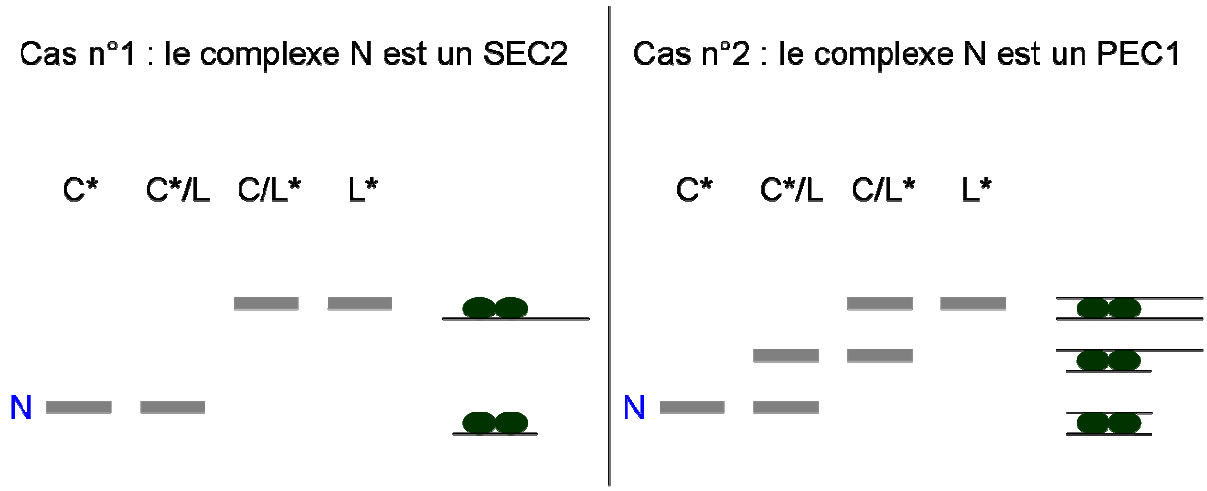


Figure 59. Principe de la méthode ITR court/long. *Cas n°1 : le complexe N est un SEC2, c'est-à-dire deux Tnp et un ITR ; ce profil est obtenu puisque dans les cas des pistes C*/L et C/L*, les SEC2 formés avec les ITR non radiomarqués ne sont pas visualisables. Cas n°2 : le complexe N est un PEC1, soit deux Tnp et deux ITR ; le profil dit en escalier est obtenu puisque les complexes formés avec un ITR radiomarqué et un ITR froid sont visualisables.*

Cette méthode a été employée avec la transposase eucaryote, en utilisant un ITR court de 70 pb et un ITR long de 260 pb. Le profil observé n'est pas celui en escalier, mais celui du cas n°1 présenté ci-dessus, ce qui signifie que le complexe A ne contient qu'un seul ITR (figure 60).

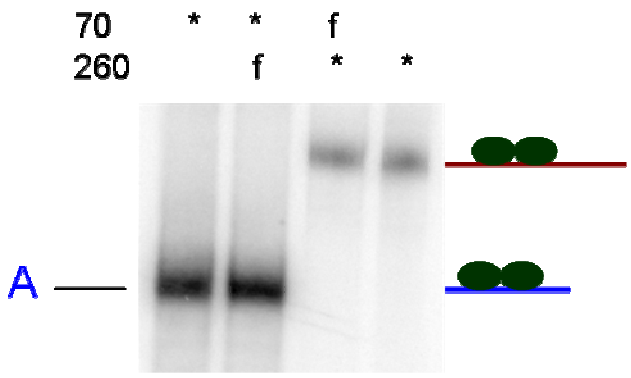


Figure 60. ITR court/long avec la transposase eucaryote. (*70 : ITR 70, 260 : ITR 260, * : ITR chaud, f : ITR froid). Retard sur gel avec la Tnp baculovirale, en présence d'ITR court ou long, chaud ou froid. Le profil obtenu est celui du cas n°1, dans lequel le complexe est constitué d'un seul ITR. Sur le schéma de droite, l'ITR court est en bleu, le long en rouge ; la Tnp est représentée par l'ovale vert foncé.*

Le clivage au facteur Xa de la MBP sur les complexes préformés a permis de montrer qu'il y avait deux transposases dans le complexe A ; l'expérience avec les ITR court/long a quant à elle démontré qu'il n'y avait qu'un seul ITR dans le **complexe A**, qui se trouve donc être un **SEC2**.

2.2.3 Le PEC1 est un SEC2 !

Dans leurs études sur la formation du PEC, Augé-Gouillou *et al* (2005b) ont utilisé la technique des ITR court/long pour déterminer le nombre d'ITR dans les différents complexes (figure 61).

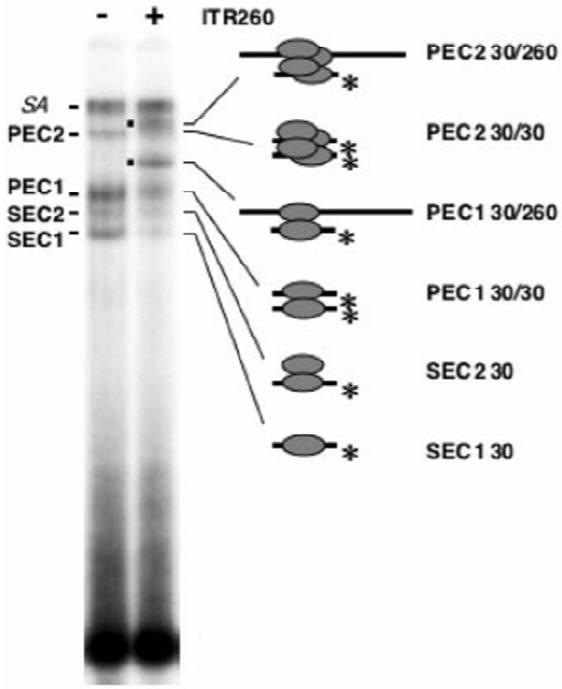


Figure 61. ITR court/long avec la transposase procaryote, d'après Augé-Gouillou *et al*, 2005b. *(30 = ITR 30, 260 = ITR 260, * = ITR chaud). Retard sur gel avec la Tnp procaryote, en présence d'ITR court ou long, chaud ou froid. Les auteurs concluent que le complexe nommé PEC1 est composé de deux ITR et de deux Tnp.*

Une critique peut néanmoins être faite sur l'utilisation de cette technique, puisqu'elle n'a été réalisée que partiellement : C* et C*/L. Les pistes complémentaires C/L* et L* n'ont pas été faites. Sans les pistes complémentaires, il semble toutefois difficile de conclure aussi formellement sur la nature du composé PEC1.

L'expérience complète ITR court/long a été réalisée avec la protéine procaryote, dans les mêmes conditions que celles utilisés avec la protéine baculovirale. Le profil en escalier, attendu si le complexe présenté comme un PEC1 était vraiment composé de deux ITR et deux transposases, n'est pas retrouvé (figure 62).

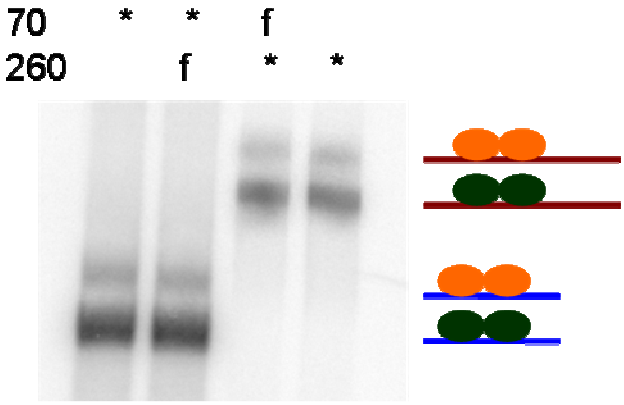


Figure 62. ITR court/long avec la transposase procaryote. (*70 : ITR 70, 260 : ITR 260, * : ITR chaud, f : ITR froid). Retard sur gel avec la Tnp procaryote, en présence d'ITR court ou long, chaud ou froid. Le profil obtenu est celui du cas n°1, dans lequel le complexe est constitué d'un seul ITR. Sur le schéma de droite, l'ITR court est en bleu, le long en rouge ; la Tnp est représentée par un ovale vert foncé ou orange.*

Le complexe initialement considéré comme un PEC1 est en réalité un SEC2, un complexe composé de deux transposases et d'un seul ITR. La formation des complexes en présence de magnésium autorise donc la visualisation de deux conformations différentes d'un SEC2 en retard sur gel. Le SEC2 initial est dès lors appelé SEC2 *bottom* (SEC2b) et celui qui était jusqu'alors considéré comme un PEC1 est nommé SEC2 *top* (SEC2t).

La nature du complexe nommé PEC1 30/260 (figure 61) est à déterminer. Il est possible que ce complexe soit un complexe de capture de cible (nommé STC pour *Strand Transfer Complex*) formé d'un SEC2 (deux transposases et un ITR) ayant capturé une cible, dans le cas de cette expérience un ITR 30 ou 260 radiomarqué. Ce point sera détaillé dans le chapitre suivant.

Ce résultat pourrait avoir un impact important sur la compréhension de la formation du complexe synaptique. En effet, aucun PEC1 (deux transposases et deux ITR) n'est dès lors visualisé en retard sur gel, ce qui pourrait signifier que ce complexe n'existe tout simplement pas. L'hypothèse d'une dimérisation directe du complexe SEC2 (figure 28) pour former le PEC2 est alors la plus probable parmi les deux voies proposées. La présence de deux conformations distinctes du SEC2 pourrait s'expliquer par le fait que pour se dimériser et former le PEC2, le SEC2 a besoin d'adopter une conformation particulière (SEC2 *top*), qui est différente que celle résultant de la liaison du dimère de Tnp à l'ITR (SEC2 *bottom*).

2.2.3 Capture de cible par un SEC2 ?

Notre connaissance actuelle sur la formation du complexe synaptique et du transpososome chez les éléments *mariner* en général, et chez *Mos1* en particulier, il peut paraître délicat d'admettre qu'un complexe de capture de cible puisse ne contenir que deux transposases fixés sur un ITR. Il est en effet décrit dans la littérature que c'est un tétramère de

transposases qui forme le complexe synaptique chez *Mos1* et *Himar1* (Augé-Gouillou *et al*, 2005b ; Butler *et al*, 2006).

La caractérisation du profil de liaison de la transposase à ses ITR présentée ici apporte une nouvelle interprétation à la figure issue du papier d'Augé-Gouillou *et al* (2005b) et présentée précédemment. Voici la même figure avec les complexes SEC2 et PEC1 justement renommés SEC2 *bottom* et SEC2 *top* (figure 63). Le complexe proposé comme étant un PEC1 formé avec un ITR 30 et un ITR 260 est donc certainement un autre type de complexe. Ce complexe pourrait être un complexe de capture de cible (nommé STC pour *Strand Transfer Complex*), un SEC2 contenant une cible. Mes travaux ne permettent pas de conclure quant aux complexes présentés comme des PEC2 ; le PEC2 30/260 peut être soit un PEC2 avec un ITR 30 et un ITR 260, soit un PEC2 contenant une cible.

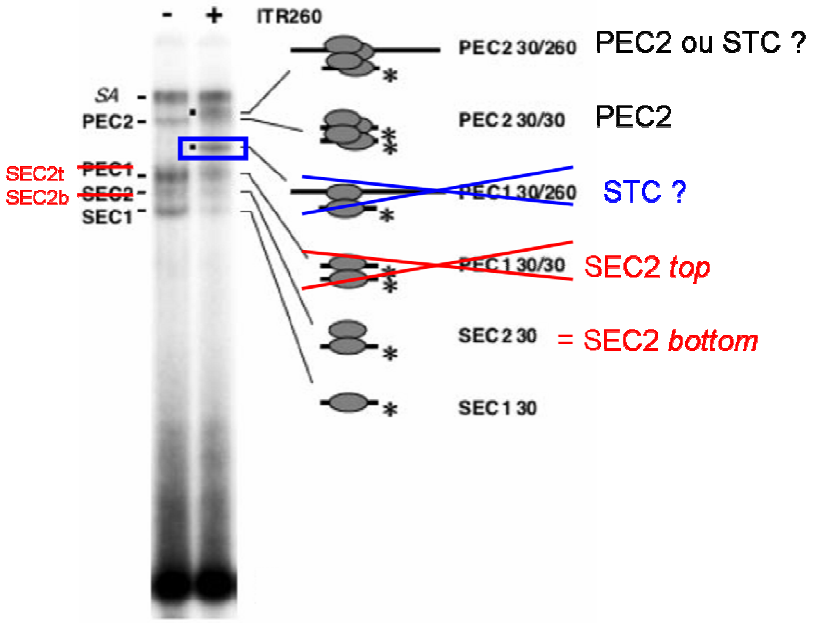


Figure 63. Nouveau profil de liaison de la transposase procaryote à l'ADN.

Ce complexe STC n'était jusqu'alors apparu en retard sur gel que dans cette expérience d'ITR court/long. L'utilisation de conditions particulières a permis d'observer à nouveau ce complexe STC. Il a fallu pour ceci utiliser des ITR particuliers, qui étaient déjà pré-coupés sur le premier brin ou pré-clivés (figure 64).

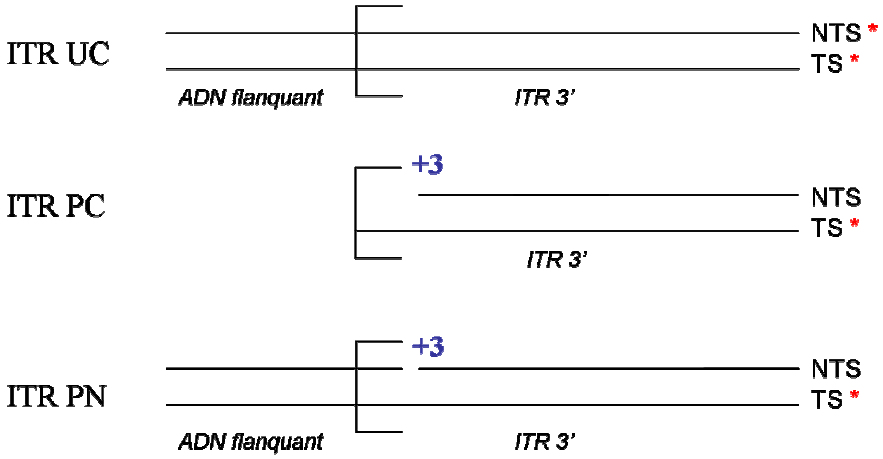


Figure 64. Différents ITR 3'. *L'ITR UC (pour UnCleaved) est l'ITR 3' classiquement utilisé dans les expériences de retard sur gel précédemment présentés. L'ITR PC (pour PreCleaved) mime l'ITR après coupure précise des deux brins par la transposase, c'est-à-dire à l'extrémité de l'ITR pour le brin transféré (TS) et en 3 bases à l'intérieur (+3) pour le brin non-transféré (NTS). L'ITR PN (pour PreNicked) mime l'ITR après coupure précise du brin non-transféré en +3. Les astérisques rouges représentent l'endroit où est marqué l'ITR par du* $\gamma^{32}P$ *: sur n'importe quel des deux brins pour l'ITR UC, et sur le brin transféré pour les ITR PC et PN.*

Les complexes ont été assemblés avec 200 nM de protéines (procaryote et eucaryote), avec 10 nM de chaque ITR et dans deux conditions d'incubation : 20 minutes à 4°C en présence de magnésium, et deux heures à 30°C en présence d'EDTA (figure 65). Le STC est visible lorsque les complexes sont assemblés en conditions catalytiques avec l'ITR PC et l'ITR PN, mais uniquement avec l'ITR PC lorsque les conditions sont non-catalytiques (EDTA). Il semblerait donc que le STC nécessite le clivage des deux brins pour être formé. En conditions catalytiques la Tnp peut cliver le second brin de l'ITR PN, mais pas en conditions non-catalytiques, où seul un ITR pré-clivé sur ses deux brins permet de visualiser le STC. Par contre, quelque soient les conditions, le complexe STC n'apparaît jamais avec l'ITR 3' non clivé (UC) ; ceci est logique en conditions EDTA, et s'explique en conditions catalytiques par le fait que le rendement de coupure du second brin est très mauvais (Augé-Gouillou *et al*, 2005b ; Julien Bischerour, communication personnelle), donc la configuration de l'ITR PC est rarement retrouvée.

Il a cependant été confirmé que le STC contenait bien une cible (Julien Bischerour, communication personnelle). Ce complexe disparaît notamment avec l'ajout de compétiteur, tel que du sperme de saumon soniqué.

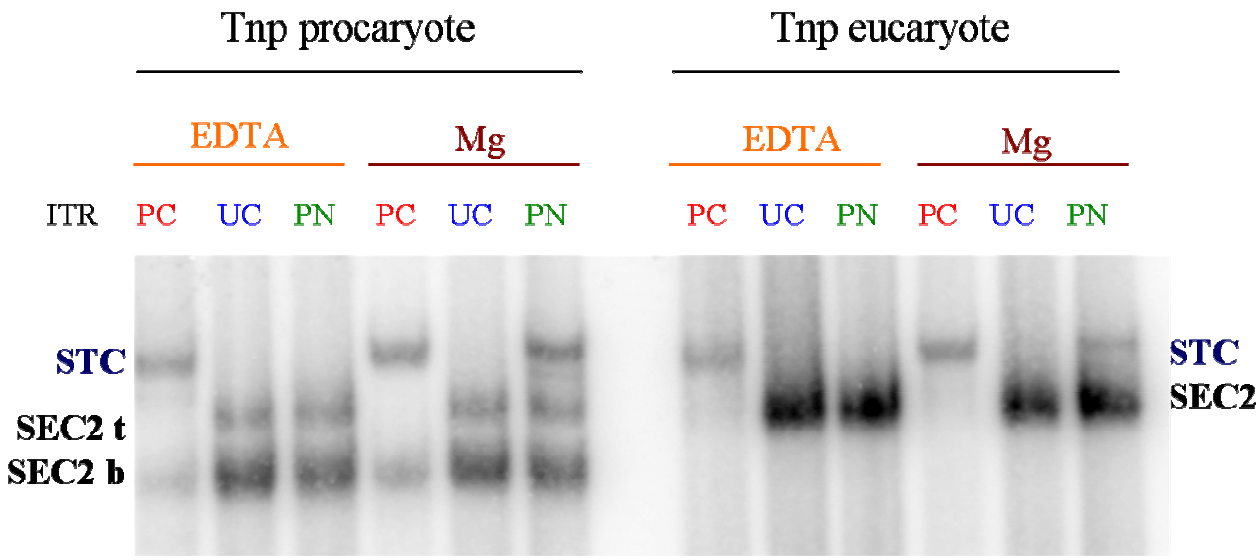


Figure 65. Profil de liaison de MOS1 à différents ITR 3'. *Les complexes ont été assemblés avec les transposases d'origine eucaryote et procaryote, en conditions catalytiques (avec du magnésium, Mg) ou non (EDTA) avec les trois types d'ITR : UC (non clivé), PC (pré-clivé sur les deux brins), PN (clivé sur le brin non transféré). Les complexes sont indiqués sur les côtés : STC (complexe de capture de cible), SEC2t (SEC2 top) et SEC2b (SEC2 bottom).*

Deux hypothèses sont alors envisageables. Le complexe STC ne serait finalement qu'un produit de dissociation du PEC2, dans lequel un dimère de transposases reste lié à un ITR avec la cible que le PEC2 avait capturé ; ou alors le STC serait le résultat d'un SEC2 ayant promu un clivage double-brin et capturé un ADN cible. Ce complexe pourrait n'être cependant qu'un artefact de réaction, incapable de promouvoir de la transposition.

Mon travail sur la mécanique de transposition de l'élément *Mos1* apporte de nouveaux éléments de compréhension de la transposition au niveau moléculaire. Le modèle de transposition est en pleine évolution et certains points restent à éclaircir, en particulier sur l'étape de capture de cible.

2.3 MOS1 et phosphorylation

Des études préliminaires précédant mon arrivée dans le laboratoire laissaient penser que la transposase eucaryote était phosphorylée. La protéine eucaryote a alors été étudiée en détail dans le but de déterminer l'impact de la phosphorylation sur l'activité de l'enzyme et d'établir son profil de phosphorylation.

Trois voies ont été suivies en parallèles, et sont présentées ici :

- des études biochimiques pour déterminer l'impact de la phosphorylation sur l'activité de la transposase

- une approche *in silico* pour déterminer les sites potentiels de phosphorylation, suivie d'une étude comparative de mutants pseudo- et non-phosphorylables des positions jugées les plus intéressantes

- une approche par spectrométrie de masse pour déterminer le profil de phosphorylation de la Tnp (projet développé par le Dr Guillaume Carpentier, et dont les résultats seront présentés brièvement ici).

2.3.1 MOS1 est phosphorylée

La confirmation que la transposase produite en système eucaryote est bel et bien phosphorylée est venue en utilisant le kit *proQ Diamond* (*Invitrogen*) ; il s'agit d'un kit permettant la détection sur gel de polyacrylamide des protéines phosphorylées, après coloration avec la solution *proQ* puis révélation sous fluorescence entre 532-560 nm à l'aide de l'analyseur d'images *Fuji LAS-4000*. La protéine baculovirale apparaît clairement phosphorylée, contrairement à la protéine bactérienne (figure 66 et 67).

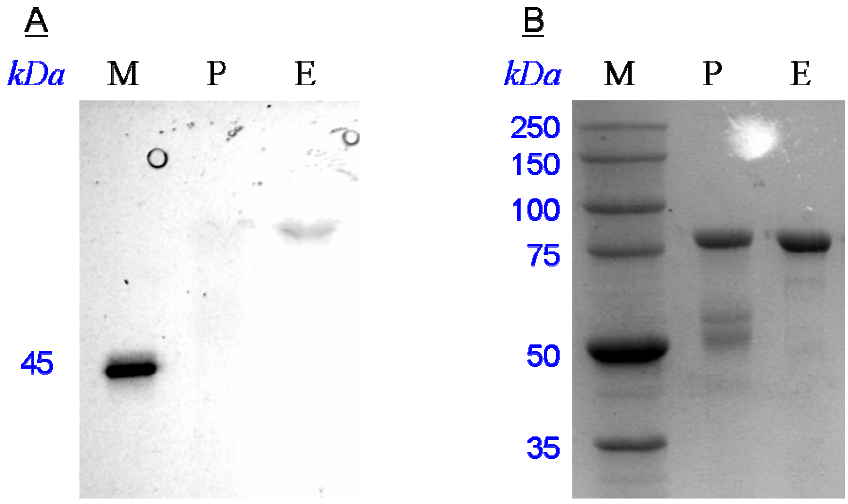


Figure 66. La MBP-transposase produite en système eucaryote est phosphorylée. *(M = marqueur de poids moléculaire ; P : Tnp procaryote ; E = Tnp eucaryote)* **(A)** *Le gel de polyacrylamide après coloration au proQ Diamond.* **(B)** *Les mêmes échantillons déposés sur un gel de polyacrylamide coloré au bleu de coomassie. La protéine eucaryote apparaît phosphorylée, contrairement à la protéine procaryote. Le témoin positif est l'ovalbumine (une protéine du marqueur phosphorylée) et les autres protéines du marqueur qui sont non phosphorylées servent de témoins négatifs.*

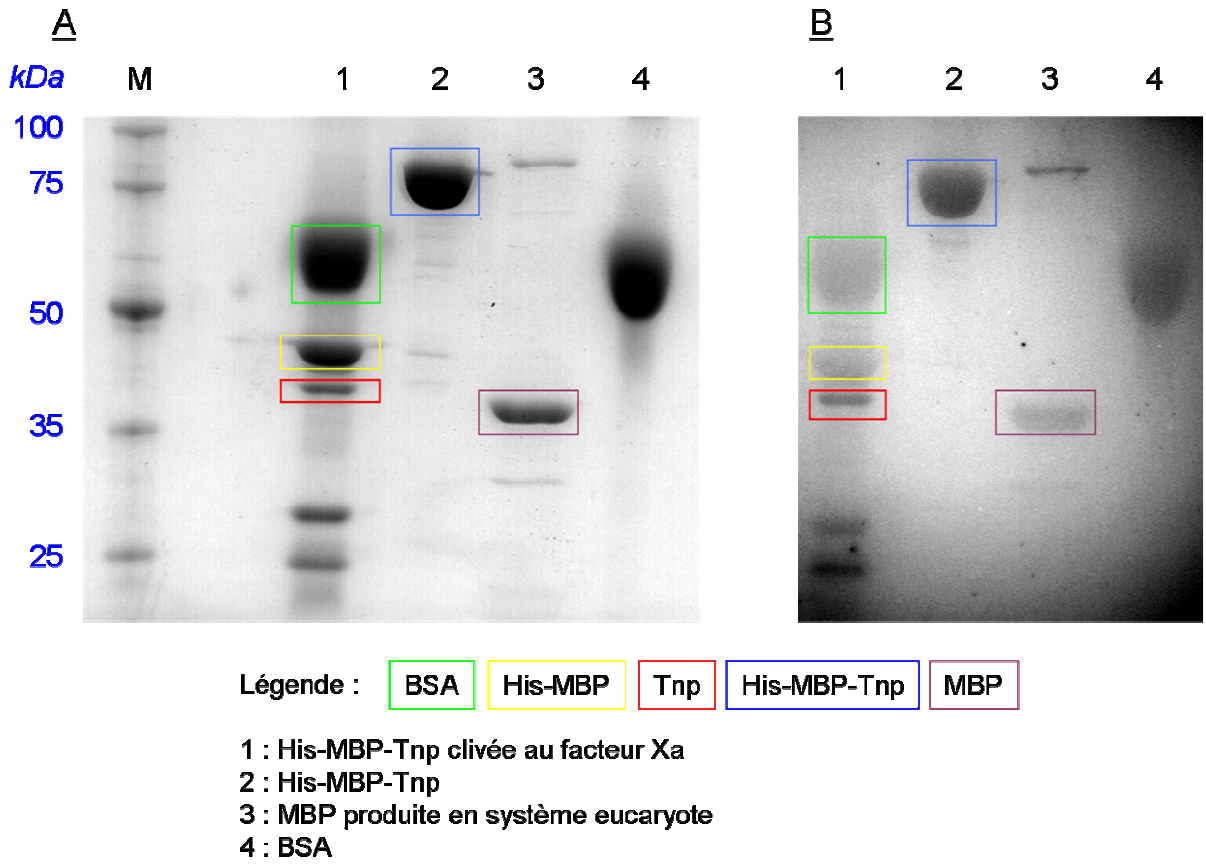


Figure 67. La transposase produite en système eucaryote est phosphorylée. *(M = marqueur de poids moléculaire) L'échantillon de Tnp utilisé ici est de la His-MBP-Tnp (ligne 2). Les échantillons ont été déposés sur un gel de polyacrylamide, d'abord coloré par le ProQ Diamond (gel B), puis au bleu de coomassie (gel A). La His-MBP-Tnp a été clivée au facteur Xa (NEB) pendant 2h à 25°C en présence de BSA (ligne 1). La Tnp (encadrée en rouge, ligne 1) est bien phosphorylée (gel B), de même que la MBP (encadrée en jaune et violet, lignes 1 et 3). La différence de taille pour la MBP entre la ligne 1 et 3 s'explique par la présence du tag His pour la ligne 1, alors qu'en ligne 3 la MBP a été produite en système baculovirus / cellules d'insectes sans tag His.*

Ces gel révèlent que la protéine produite en système baculovirus / cellules d'insectes est faiblement phosphorylée. Il en effet possible, selon le manuel d'utilisation du kit *proQ Diamond*, de calculer la quantité de phosphate par protéine, en fonction de l'intensité du signal et de la quantité de protéine déposée sur le gel. Cette méthode (présentée dans la partie Matériel et Méthodes) nous indique qu'il y a une demi-phosphorylation par protéine, ce qui signifie que 50% des protéines sont phosphorylées une fois, ou alors qu'une petite fraction de protéines est fortement phosphorylée.

Bien que le niveau de phosphorylation soit faible, quel est son impact sur l'activité de la Tnp ?

2.3.2 Action des phosphatases sur MOS1

Dans le but d'étudier l'impact des phosphorylations sur l'activité de la transposase eucaryote, la Tnp a été déphosphorylée par plusieurs phosphatases :

- la PPase1 et la PPase2A qui déphosphorylent les sérines et les thréonines
- la λPPase qui agit sur les sérines, thréonines, tyrosines et histidines
- la LAR PPase qui déphosphoryle uniquement les tyrosines.

Les conditions de déphosphorylation de la Tnp eucaryote sont indiquées dans la partie Matériel et Méthodes. La coloration par le *proQ Diamond* révèle la PPase1, la λPPase et la PPase2A déphosphorylent la Tnp eucaryote puisque le signal disparaît lors de la coloration au *ProQ Diamond* (figure 68-A). Le fait que la LAR PPase, qui n'agit que sur les tyrosines, ne semble n'avoir aucun effet sur la TnpE, et que la PPase1 déphosphoryle entièrement la transposase montrent que la protéine n'est pas phosphorylée uniquement sur des tyrosines. Des tyrosines peuvent néanmoins être phosphorylées dans le cadre d'une phosphorylation hétérogène de la protéine.

Le dépôt sur un gel de polyacrylamide coloré au bleu de coomassie des mêmes échantillons apporte un éclairage nouveau sur le rôle de la phosphorylation et l'action des phosphatases. L'action de la PPase2A entraîne l'apparition des produits de protéolyse observés initialement chez la protéine procaryote, alors que le traitement de la Tnp E par la λPPase et la PPase1 n'induit pas la protéolyse (figure 68-B).

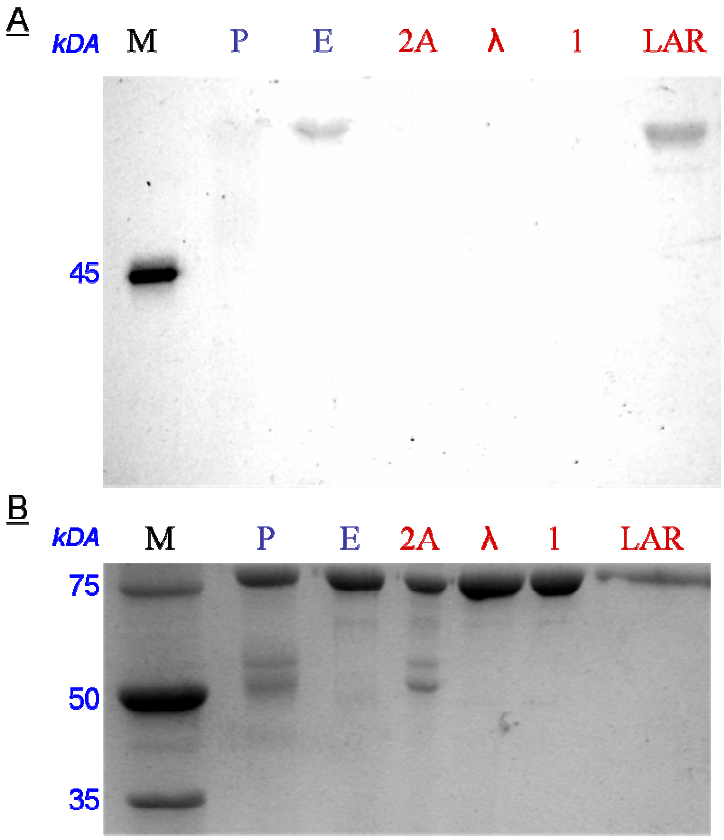


Figure 68. Action de différentes phosphatases sur la transposase eucaryote. (A) *(M = marqueur PeppermintStick ; P = Tnp procaryote ; E = Tnp eucaryote ; 2A, λ, 1 et LAR = Tnp E traitée à la PPase2A, à la λPPase, à la PPase1 et à la LAR PPase) La Tnp E a été incubée en présence de différentes phosphatases, et les échantillons déposés sur un gel de polyacrylamide par la suite coloré au ProQ Diamond. La PPase2A, la λPPase et la PPase1 déphosphorylent MOS1. La LAR PPase n'atténue pas le signal.* **(B)** *Les mêmes échantillons que dans le gel (A) déposés sur un gel de polyacrylamide par la suite coloré au bleu de coomassie. La PPase2A induit de la protéolyse, contrairement aux autres phosphatases.*

Les protéines déphosphorylées par la PPase2A (phosphatase qui déphosphoryle la Tnp et induit sa protéolyse) et la λPPase (qui déphosphoryle la Tnp sans impact sur sa stabilité) ont été comparées en test de transposition *in vitro*. La Tnp eucaryote native et la Tnp déphosphorylée par la λPPase ont des fréquences de transposition respectives de 10^{-4} +/- [5 x 10^{-5}] et 2 x 10^{-4} +/- [5,5 x 10^{-5}]. Un test de Mann & Whitney réalisé sur 10 échantillons indépendants a montré qu'il n'y avait aucune différence statistique entre ces deux valeurs (pour un risque de 5%). Par contre, la Tnp déphosphorylée par la PPase2A a une fréquence de transposition de 1,58 x 10^{-5} +/- [2 x 10^{-5}] ; un test de Mann & Whitney, réalisé sur 12 échantillons indépendants, montre que la différence est statistiquement significative entre les deux protéines (au risque de 5%). Les phosphatases n'ont donc pas le même impact sur l'activité globale de la protéine.

En parallèle des tests de transposition *in vitro*, une expérience de retard sur gel a été menée pour voir si la déphosphorylation avait des conséquences sur la liaison de la transposase à ses ITR. Un ITR 3' marqué par du γ^{32}P a été utilisé, comme décrit précédemment. Alors que la déphosphorylation de la protéine par la λPPase et la PPase1 n'a

aucune influence sur le profil de liaison de la Tnp eucaryote à son ITR, l'action de la PPase2A provoque l'apparition d'un second complexe, appelé B (figure 69).

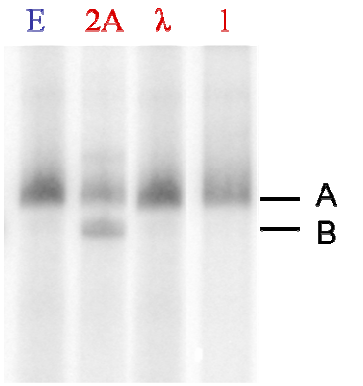


Figure 69. Impact de la déphosphorylation sur la liaison de MOS1 à l'ADN. *(E = Tnp eucaryote ; 2A, λ et 1 = Tnp E traitée à la PPase2A, à la λPPase et à la PPase1) La Tnp E ne forme qu'un seul complexe (A) et il en est de même une fois qu'elle est déphosphorylée par la λPPase et la PPase1. L'action de la PPase2A entraîne l'apparition d'un second complexe, nommé B.*

Il semble que la PPase2A ait une action supplémentaire à la déphosphorylation. Cette action a un impact sur la stabilité de la transposase, son profil de liaison à l'ITR et son taux global de transposition.

Pour vérifier la relation entre déphosphorylation et protéolyse de MOS1, une cinétique de déphosphorylation a été effectuée avec la λPPase (30', 1h30, 3h, 6h et 24h d'incubation), puis les échantillons ont été déposés sur un gel de polyacrylamide qui a été coloré au *ProQ Diamond*, puis au bleu de coomassie. La déphosphorylation de la Tnp eucaryote est totale après 30 minutes de réaction, et elle n'induit pas la protéolyse (figure 70).

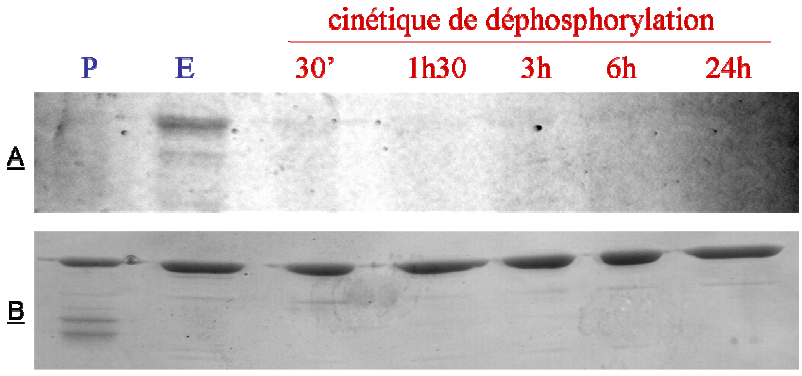


Figure 70. Impact de la λPPase sur la stabilité de la transposase eucaryote. *(A : gel coloré au proQ Diamond, B : le même gel coloré au bleu de coomassie ; P : Tnp procaryote ; E : Tnp eucaryote) La déphosphorylation par la λPPase n'induit pas de protéolyse, et la Tnp est déphosphorylée dès 30' d'incubation avec la phosphatase.*

Il semble donc que la phosphorylation n'ait aucun impact sur l'activité de la transposase eucaryote ; la déphosphorylation par la λPPase ne provoque pas de différence de stabilité, de transposition et de liaison à l'ITR.

2.3.3 Transposase cytoplasmique versus transposase nucléaire

La transposase produite en système baculovirus / cellules d'insectes présentée dans les études précédentes était d'origine cytoplasmique. La question s'est alors posée de savoir s'il y avait une différence entre une transposase extrait du cytoplasme et une transposase extraite du noyau de la cellule. Cette protéine nucléaire a été obtenue par le Dr Guillaume Carpentier au laboratoire.

Le premier constat est que la protéine nucléaire est phosphorylée d'après une coloration au *proQ Diamond* (figure 71). La transposase nucléaire est stable car elle ne présente pas de produits de protéolyse.

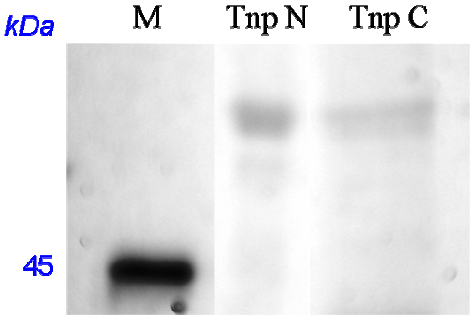


Figure 71. La transposase nucléaire est phosphorylée. *(M : marqueur de taille PepperMintStick ; Tnp N : transposase nucléaire ; Tnp C : transposase cytoplasmique) La transposase extraite du noyau apparaît phosphorylée.*

Un test de transposition *in vitro* a été réalisé avec les deux transposases, déphosphorylées par la λPPase ou non. Le bilan est qu'il n'y a aucune différence d'activité entre ces quatre protéines (tableau 2).

L'absence de différence entre protéines nucléaires et cytoplasmiques, phosphorylées ou non, et le fait que la Tnp eucaryote soit faiblement phosphorylée conduisent à penser que la population de protéines correctement phosphorylée était peut-être masquée par la fraction majoritaire de protéine partiellement ou non phosphorylée. Pour vérifier cette hypothèse, un kit d'enrichissement en phosphoprotéines a été utilisé (le *TALON PMAC Phosphoprotein Enrichment Kit* de chez *Clontech*), kit dont le principe est exposée dans la partie Matériel et Méthodes. Déposés sur un gel de polyacrylamide coloré au *ProQ Diamond*, ces lots de protéines enrichies apparaissent phosphorylées, mais pas plus que les lots non enrichis. Ces lots enrichis ont été testé en transposition *in vitro*, mais là encore il n'y a aucune différence avec la transposase cytoplasmique native (tableau 2).

	Fréquence de transposition
Tnp cytoplasmique	$1,9 \times 10^{-4}$ +/- [$1,2 \times 10^{-4}$]
Tnp cytoplasmique déphosphorylée	$2,2 \times 10^{-4}$ +/- [9×10^{-5}]
Tnp cytoplasmique enrichie	$1,8 \times 10^{-4}$ +/- [6×10^{-5}]
Tnp nucléaire	$9,5 \times 10^{-5}$ +/- [$1,3 \times 10^{-4}$]
Tnp nucléaire déphosphorylée	$1,4 \times 10^{-4}$ +/- [$2,7 \times 10^{-5}$]
Tnp nucléaire enrichie	$2,1 \times 10^{-4}$ +/- [$1,1 \times 10^{-4}$]

Tableau 2. Fréquence de transposition des différentes transposases eucaryotes. *Les fréquences de transposition présentées sont les médianes calculées sur au moins cinq échantillons indépendants ; les écart-types sont indiqués entre crochets.*

Un test statistique non-paramétrique de Kruskall & Wallis a été effectué sur les fréquences de transposition obtenues lors d'au moins cinq tests de transposition réalisés de manière indépendante. Le résultat est que les échantillons ont la même distribution (au risque de 5%), et donc qu'il n'y a pas de différence significative entre toutes ces fréquences de transposition.

2.3.4 Pattern de phosphorylation : spectrométrie de masse

En parallèle des études biochimiques visant à déterminer l'impact de la phosphorylation sur l'activité de la transposase eucaryote, des études de spectrométrie de masse ont été menées par le Dr Guillaume Carpentier, en collaboration avec le Dr Valérie Labas de l'INRA de Nouzilly. L'objectif était d'obtenir le pattern de phosphorylation de la protéine. Plusieurs méthodes de spectrométrie de masse ont été utilisées, et les résultats sont présentés ci-dessous.

Le premier résultat confirme que la transposase produite en système eucaryote est très peu phosphorylée : seulement 0,1 à 10 % de l'échantillon (Valérie Labas, communication personnelle). L'analyse par spectrométrie de masse a permis de déterminer un pattern partiel de phosphorylation et de détecter trois acides aminés phosphorylés : les sérines en position 99 (S99) et 170 (S170), et la thréonine en position 191 (T191). Quatorze autres positions apparaissent potentiellement phosphorylées, et d'autres analyses seront nécessaires pour confirmer ou non la présence d'une phosphorylation sur ces acides aminés ; ils sont présentés en bleu dans la figure ci-dessous (figure 72). A cause de sa séquence riche en acides aminés acides, la région allant de la sérine 55 à la lysine 90 est difficilement détectable en spectrométrie de masse. Pour cette raison, nous n'avons aucune information sur cette région.

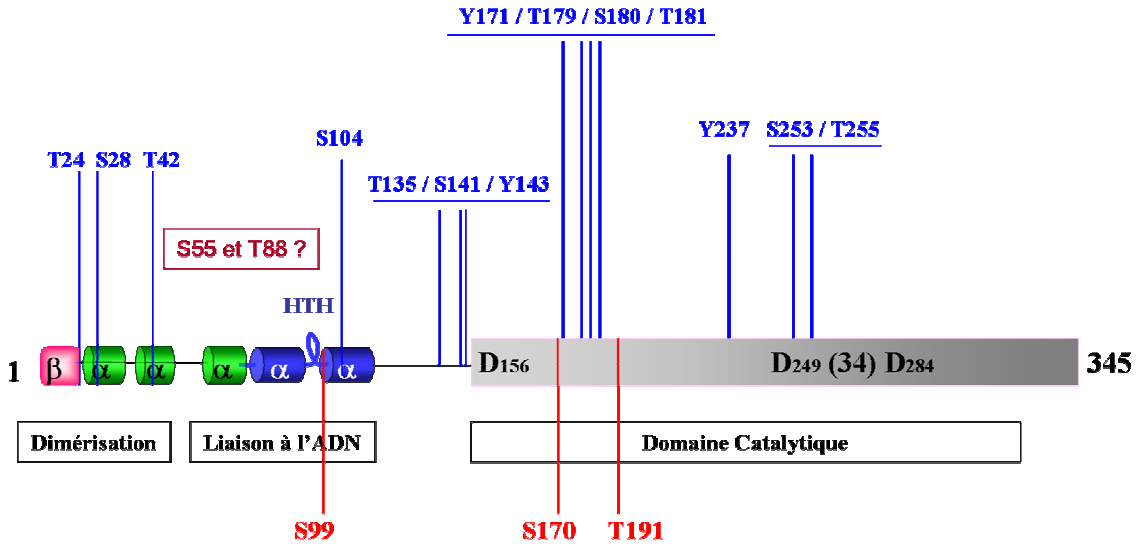


Figure 72. Pattern partiel de phosphorylation de MOS1. *La transposase est représentée sous forme schématique avec ces différents sous-domaines. Les trois acides aminés indiqués en rouge ont été détectés phosphorylés par spectrométrie de masse, les quatorze en bleu sont potentiellement phosphorylés. Pour des raisons techniques la région contenant les acides aminés S55 et T88 n'a pas pu être étudiée.*

Ces résultats apportent des éléments complémentaires aux études présentées précédemment sur le HTH. La sérine phosphorylée en position 99 est très conservée parmi les éléments *mariner*. Elle semble donc être une position critique dans la protéine, et importante pour son activité. Parmi les positions potentiellement phosphorylées est retrouvée la sérine 104, située en plein cœur du HTH, ainsi que plusieurs acides aminés situés dans la partie catalytique de la transposase. Le HTH étant impliqué dans la liaison de la Tnp à ses ITR, la phosphorylation pourrait jouer un rôle à ce niveau là.

La transposase MOS1 produite en système eucaryote est donc faiblement phosphorylée de manière hétérogène, mais l'échantillon semble inadéquat pour faire l'étude biochimique directe de la protéine. Cette approche ne permet pas de répondre à la question de savoir si la phosphorylation a un impact sur l'activité de la transposase.

2.3.5 Design d'une transposase non-phosphorylable

2.3.5.1 Etudes *in silico*

En lien avec les précédents résultats, des études *in silico* ont été menées dans le but d'identifier les positions dans MOS1 susceptibles d'être phosphorylées, en utilisant quatre logiciels de prédiction disponibles sur internet (tableau 3).

Résidus	GPS	Kinase Phos	NetPhos 2.0	Net PhosK	Domaines de MOS1
S3				x	Dimérisation
T11			x		
T13				x	
T24				x	
S28	x		x		
T42	x			x	
S55			x	x	
Y72		x	x		HTH – liaison à l'ADN
T88*	x		x	x	
S99*	x	x		x	
S104*	x		x	x	
T135*	x	x	x	x	Linker
S141				x	Catalyse
S147*	x		x	x	
T154*		x	x	x	
S170*	x	x	x	x	
Y171*	x	x	x		
T179	x				
S180*	x		x	x	
T181*	x		x	x	
S200				x	
Y219				x	
Y237*	x	x	x	x	
S253	x				
T262				x	
S277*	x		x	x	
Y283			x		
S302*	x		x	x	
Y303*	x		x	x	
S305*	x		x	x	
Y322			x		
Y345			x		

Tableau 3. Résidus prédits phosphorylés chez MOS1, adaptée de Germon *et al*, 2009. *Quatre serveurs ont été utilisés : GPS (http://gps.biocuckoo.org), KinasePhos (http://kinasephos.mcb.nctu.edu.tw), NetPhos 2.0 et NetPhosK (disponibles à l'adresse suivante : http://www.cbs.dtu.dk/services/). Les résidus prédis phosphorylés trois fois sont en gras avec un astérisque, ceux étudiés en détail par la suite en rouge.*

Notre attention s'est focalisée sur les premières étapes de la transposition : la dimérisation et la liaison aux ITR. C'est pourquoi nous ne nous sommes intéressés qu'à la partie N-terminale de la protéine, qui comprend le domaine de dimérisation et le domaine de liaison à l'ADN. L'étude s'est concentrée sur six acides aminés, qui étaient prédits

phosphorylés par deux ou trois serveurs avec une forte probabilité. Il s'agit des acides aminés T88, S99 (qui apparaît phosphorylé en spectrométrie de masse) et S104, qui sont prédits pour être phosphorylés par trois des quatre serveurs utilisés ; dans le domaine de dimérisation, T11, S28 et S55 présentent des pourcentages bien plus élevés que les autres acides aminés de ce domaine prédits pour être phosphorylés.

La mutagenèse de ces positions a été utilisée pour tester l'impact de la phosphorylation sur la transposition ; des mutants pseudo-phosphorylables ont tout d'abord été construits et testés en test de transposition *in vivo*. Les positions insensibles à la phosphorylation vont donner des mutants avec une activité de transposition similaire à la transposase sauvage. Les positions sensibles à la phosphorylation vont donner des mutants avec des fréquences de transposition différentes : augmentation de la fréquence de transposition pour un mutant d'une position positivement activée par la phosphorylation, diminution pour un mutant sur qui la phosphorylation aurait un impact négatif. En cas d'impact négatif de la pseudo-phosphorylation sur l'activité de la protéine, des mutants non-phosphorylables à cette position ont été construits. Trouver un mutant fonctionnel non-phosphorylable à ces positions est important dans l'optique de réaliser de la transposition en cellules de mammifères.

2.3.5.2 Mutants du domaine de dimérisation

Les mutants pseudo-phosphorylés ont été construits en remplaçant les sérines et les thréonines respectivement par des acides aspartiques et des acides glutamiques, des acides aminés qui miment la phosphorylation (Dai *et al*, 2007). Les résultats des tests de transposition sont présentés dans le tableau 4. Le mutant S55D[3] a une fréquence de transposition similaire à la protéine native ; ceci est montré par un test statistique non-paramétrique de Kruskall & Wallis qui indique que les échantillons ont la même distribution (au risque de 5%). La pseudo-phosphorylation à la position S55 n'affecte donc pas la transposition.

[3] Par commodités d'écriture, les mutants seront écrits par la suite de cette manière : AxxB, où A représente l'acide aminé muté, xx sa position dans la protéine, et B l'acide aminé par lequel il a été muté.

Protéine	Fréquence de transposition
MOS1	**5×10^{-5} / 10^{-4}**
T11E	**5×10^{-7}** +/- [$1{,}2 \times 10^{-7}$]
T11V	**$4{,}25 \times 10^{-5}$** +/- [9×10^{-6}]
S28D	**$1{,}6 \times 10^{-7}$** +/- [9×10^{-8}]
S28V	**$2{,}2 \times 10^{-5}$** +/- [$4{,}4 \times 10^{-6}$]
S55D	**4×10^{-5}** +/- [3×10^{-5}]

Tableau 4. Fréquences de transposition des mutants du domaine de dimérisation de MOS1. *Les fréquences de transposition sont les médianes calculées à partir d'au moins cinq clones indépendants issus de deux transformations distinctes ; les écart-types sont indiqués entre crochets. La transposase native est en bleu, les mutants n'ayant pas ou peu d'impact sur la transposition en vert, ceux entraînant une diminution significative de la fréquence de transposition sont en rouge.*

Les mutants T11D et S28E ont quant à eux des fréquences de transposition bien inférieures à la transposase native. Les positions T11 et S28 semblent donc sensibles à la phosphorylation. Des mutants non-phosphorylables de ces positions ont été construits. La valine et la cystéine sont les deux acides aminés les plus ressemblants de la sérine et de la thréonine (figure 73). Pour des raisons liées à la séquence de MOS1, c'est la valine qui a été choisie comme acide aminé non-phosphorylable pour remplacer la thréonine 11 et la sérine 28.

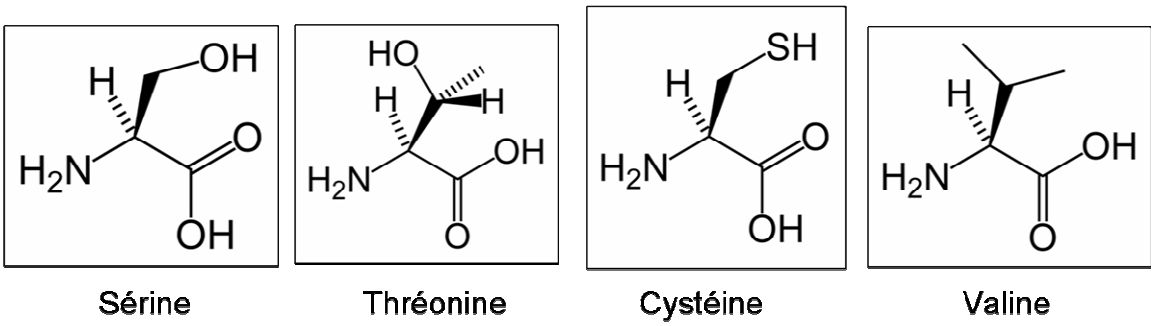


Figure 73. Comparaison de la structure de quatre acides aminés.

Les mutants T11V et S28V ont été testés en transposition *in vivo* (tableau 4). Le mutant T11V ne présente pas de différence significative avec la transposase sauvage, tandis qu'il y a une légère diminution de la fréquence de transposition pour le mutant S28V. La phosphorylation du résidu T11 semble donc avoir un impact négatif sur la transposition, puisqu'un mutant non-phosphorylable à cette position restaure une activité de transposition

identique à la transposase sauvage. La séquence entourant la position T11, prédite phosphorylée par le serveur NetPhos 2.0, ne correspond à aucune séquence consensus de site phosphorylable par des kinases.

La position S28 est quant à elle potentiellement phosphorylable par les PKC (*Protein Kinase C*) et les CK1 (*Casein Kinase 1*), deux familles de kinases présentes dans quasiment tous les types de cellules eucaryotes. Les CK1 sont généralement impliquées dans la régulation de la transduction de signaux dans les cellules, tandis que les PKC ont des fonctions multiples, dépendant du type cellulaire dans lequel elles se trouvent. Cette sérine en position 28, située dans le domaine de dimérisation de la transposase, semble être sensible à la phosphorylation, et il est intéressant de constater qu'elle est potentiellement phosphorylable par des kinases ubiquistes dans les cellules eucaryotes. Il est envisageable que la cellule contrôle par phosphorylation la dimérisation de la transposase, un pré-requis à la transposition.

2.3.5.3 Mutants du domaine de liaison à l'ADN

Les résidus T88, S99 (qui apparaît phosphorylé en spectrométrie de masse) et S104 étaient prédits phosphorylés par trois des quatre logiciels de prédiction utilisés lors de l'analyse *in silico* de la séquence protéique de MOS1. La mutation de ces résidus en des résidus pseudo-phosphorylés abolit complètement la transposition (tableau 5). Ces trois positions semblent donc impliquées dans la régulation négative par phosphorylation de MOS1. Ces trois acides aminés sont situés dans le motif HTH du domaine N-terminal de la protéine, motif impliqué dans la reconnaissance spécifique des ITR de *Mos1*.

L'idée a donc été de remplacer ces trois acides aminés par des résidus non-phosphorylables qui pourraient aider à restaurer l'activité de MOS1 en cellules humaines. Il est à noter qu'une mutation ponctuelle à chacune de ces positions n'entraîne pas d'altération du motif HTH (d'après par le logiciel *NPS@ Helix-Turn-Helix*). De nombreux mutants ponctuels ainsi que des combinaisons de doubles mutants ont été testés. Le but était de trouver des mutations altérant le moins possible la fréquence de transposition, pour ensuite pouvoir exprimer ces protéines mutantes en système eucaryote. Outre la position S99, qui s'avère être très conservée parmi les éléments *mariner* (figure 74), les glutamines en position 100 et 101 ont également été mutées ; le but était de faire disparaître le site consensus SQQ de phosphorylation par les ATM kinases. La même approche a été employée pour la position

104, pour laquelle le résidu arginine 106 a été muté pour casser le site consensus SNR de phosphorylation par la PKC.

```
(MOS1)       TQKQLAEQLEVSQQAVSNRLREMGKIQ
(HIMAR1)     KLIEIAEALKISKERVGHIIHQYLDMR
(DSECMAR1)   TQKQLAEQLEVSQQAVSNRLREMGKIQ
(DSIMAR1)    TQKQLAEQLEVSQQAVSNRLREMGKIQ
(DTESMAR1)   TQKQLAEQLEVSQQAVSNRLREMGKIQ
(MBMAR1)     TQKQLAEQLEVSQQAVSNRLREGGKIQ
(MDMAR1)     TQKQLAEQLEVSQQAVSNRLREMGKIQ
(BTMAR1)     TQQQLTDQLNVTREAVSIRLKAMGKIQ
(SINVMAR1)   TLLELSEQLNVTPIVVSKRLPAIGKIH
(MOMAR1)     TQSELAEQLKVTREAISIRLKKMGKIS
(CEMAR2)     SLREMEEKLECCHSTIARHLGRLGFTS
(CEMAR3)     ITRELSTTL-GSRQTNVRGWEPLGIKK
(CEMAR4)     TTRELASTFGVSHNSPDNMKL-LGMRK
(MCMAR1)     STRKLAEELGSSKDTVWRNMKEMELSY
(BYTMAR1)    KVEEISLEIRISHGSVCTIINHHLGMS
(CPMAR1)     KLIEIAEALKISKERVGHIIHQYLDMR
(AHMAR1)     KLSEIADIVKISKERVHHIIREYLGMR
(ARMAR1)     KLSEIADIVKISKERVHHIIREYLGMR
(SOYMAR1)    TVRTLA-VAMKTNTSAMYRLIQSGAI—
(RIZMAR1)    TIQDVATQLGVSTSKV-YRMKKEGAI-
(MPMAR1)     KLIEIADIVKISKERVHHIIREYLGYR
(DEMAR1)     ASRDIAQEMGVSHQTILNHLQKAGYKK
(AMMAR1)     TTREIAEKLHVSHTCIENHLKQLGYVQ
(HSMAR2)     TAETIANTTDISIGSAYTILTEKLKLS
(HSMAR1)     TTREVAEELNVDHSTVV-RHLKQIGKV
```

Figure 74. Conservation de résidus phosphorylables dans le motif HTH des éléments *mariner*. *Alignement de séquence de 25 transposases d'éléments mariner autour du résidu S99 de MOS1. La sérine (en rouge) est fortement conservée chez les éléments mariner, ou remplacée par une thréonine (en bleu) un autre résidu phosphorylable. Les résidus en vert ne sont pas phosphorylables, et ne sont retrouvés qu'occasionnellement à cette position.*

Les tests de transposition *in vivo* ont été effectués avec les différents mutants, et les médianes ont été calculées à partir des fréquences de transposition obtenues sur au moins cinq clones indépendants issus de deux transformations distinctes (tableau 5).

Protéine	Fréquence de transposition
MOS1	**5×10^{-5} / 10^{-4}**
T88E*	**0**
T88A*	**5×10^{-7}** +/- [4×10^{-7}]
T88C*	**2×10^{-6}** +/- [2×10^{-6}]
T88R	0
T88V	**10^{-5}** +/- [6×10^{-6}]
S99D*	0
S99A*	0
S99C*	**10^{-6}** +/- [$1{,}3 \times 10^{-6}$]
S99V	**10^{-7}** +/- [5×10^{-6}]
Q100H	0
Q100H/ Q101V	0

Protéine	Fréquence de transposition
S104D*	**0**
S104A*	**3×10^{-7}** +/- [1×10^{-6}]
S104C*	0
S104K	**3×10^{-6}** +/- [4×10^{-6}]
S104N*	0
S104P	0
S104V	**3×10^{-6}** +/- [4×10^{-6}]
S104Y*	0
R106I	0
T88C / S104V	**2×10^{-6}** +/- [3×10^{-6}]
T88V / S104V	0
T88C / S104A	**4×10^{-7}** +/- [5×10^{-7}]

Tableau 5. Fréquences de transposition des mutants non-phosphorylables du motif HTH de MOS1. *Les fréquences de transposition sont les médianes calculées à partir d'au moins cinq clones indépendants issus de deux transformations distinctes ; les écart-types sont indiqués entre crochets. Les mutants pseudo-phosphorylés sont soulignés, ceux n'ayant pas ou peu d'impact sur la transposition sont en vert et ceux entraînant une diminution significative de la fréquence de transposition sont en rouge. Les données avec un astérisque sont issues de la thèse de Benjamin Brillet (2006).*

Toutes les mutations présentées (exception faite de la T88V) altèrent l'activité de transposition promue par la transposase, certaines l'abolissant complètement. Trois combinaisons de mutants ont été testées ; deux d'entre elles, T88C/S104V et T88C/S104A donnent des fréquences de transposition identiques à celle des moins actifs des mutants ponctuels. La combinaison T88V/S104V abolit totalement la transposition, alors que les mutants simples présentaient une certaine activité ; cela indique un effet complémentaire des mutations.

L'hypothèse la plus simple pour expliquer le non fonctionnement de *Mos1* en cellules de mammifères suggère qu'exprimée en cellules de mammifères, la Tnp est phosphorylée en position S99 et peut-être sur T88 et/ou S104, ce qui inhibe la transposition à cause d'une capacité réduite de la Tnp à lier ses ITR. Cette hypothèse est renforcée par le fait que la

transposase de l'élément *mariner Himar1*, qui est active en cellules humaines, n'a pas de résidus phosphorylables aux positions correspondantes dans sa séquence. T88 et S104 chez MOS1 correspondent à K87 et G103 chez HIMAR1 (Butler *et al*, 2006), qui sont des acides aminés non phosphorylables, tandis que le résidu S99 de MOS1 a pour correspondance le résidu S98 de HIMAR1, un résidu prédit comme étant non-phosphorylé.

La position S104 est située dans un site consensus de phosphorylation par les PKC, kinases présentes dans la majorité des types cellulaires chez les eucaryotes. Les positions T88 et S99 sont phosphorylables par les ATM kinases et/ou les kinases DNA-PK, deux enzymes activées en cas de cassure double-brin de l'ADN dans les cellules de mammifères. Ce qui est le cas pendant la transposition. La phosphorylation pourrait intervenir à ce niveau pour réguler la liaison à l'ADN de la transposase, prévenant ainsi la transposition et donc des coupures double-brin de l'ADN.

2.3.6 Phosphorylation : conclusion

Nos résultats montrent que la transposase MOS1 produite en système baculovirus/cellules d'insectes est faiblement phosphorylée ; trois acides aminés (S99, S170 et T191) ont été observés phosphorylés, et quatorze autres au moins sont potentiellement phosphorylables. Dans nos conditions actuelles d'étude, l'impact de la phosphorylation sur l'activité de la transposase n'est pas mesurable, et il n'est pas possible d'associer avec certitude la phosphorylation d'un acide aminé particulier à une fonction. Néanmoins, il semble que la position S99 est primordiale pour l'activité de MOS1 ; elle est très conservée parmi les éléments *mariner*, apparaît phosphorylée dans les études de spectrométrie de masse et la mutation de ce résidu conduit à une baisse importante ou une abolition de la transposition.

Il existe une autre méthode de spectrométrie de masse en plus de celles que nous avons utilisées pour déterminer le pattern de phosphorylation partiel de MOS1 ; il s'agit de l'orbitrap, un type de spectrométrie qui permettrait d'étudier notre protéine entière, sans avoir à la cliver au préalable. Cette méthode serait intéressante pour étudier le domaine N-terminal de la protéine, et notamment les acides aminés S55 et T88, qui n'étaient pas accessibles pour des raisons techniques dans les analyses que nous avons faites. Or ces résidus sont phosphorylables, et leur mutation a donné des résultats intéressants, et ils sont situés dans des

régions primordiales pour la formation du complexe synaptique (dimérisation et motif HTH de liaison à l'ADN).

Les résultats obtenus avec les mutants du domaine N-terminal soulignent l'importance du domaine de dimérisation et du motif HTH dans la liaison à l'ADN. Les positions qui semblent sensibles à la phosphorylation dans le domaine N-terminal se situent toutes dans des sites consensus de phosphorylation par des kinases présentes dans la plupart des types cellulaires eucaryotes. La phosphorylation de certains résidus sérines et thréonines du HTH peut permettre d'expliquer pourquoi *Mos1* est si peu actif en dehors de son hôte naturel, voire pas du tout en cellules humaines.

2.3.7 Autres modifications post-traductionnelles ?

Notre travail s'est concentré sur la phosphorylation, une des modifications post-traductionnelles majeures chez les protéines eucaryotes, mais il faut garder à l'esprit qu'il existe de nombreuses autres modifications post-traductionnelles possibles sur les acides aminés, telles que l'acétylation ou les différentes glycosylations.

La séquence protéique de MOS1 a été analysée par divers logiciels de prédiction de modifications post-traductionnelles disponibles sur internet, dans l'objectif de voir si la transposase possédait des acides aminés susceptibles d'être soumis à d'autres modifications que la phosphorylation.

Le serveur NetOGlyc 3.1 prédit que trois acides aminés pourraient être O-glycosylés, il s'agit des thréonines en position 42 et 181 et de la sérine 180. Aucun site consensus pour les D, C et N-glycosylations n'existe sur MOS1 ; il s'agit en plus de modifications de taille importante et la différence de poids moléculaire aurait été observée en gel de coomassie. Il est donc improbable que la Tnp soit glycosylée. La sérine en position 3 présente quant à elle un site consensus d'acétylation, cette modification étant très répandue chez les eucaryotes ; plus de 80% des protéines humaines sont acétylées. A noter également que MOS1 présente de nombreux sites consensus de glycation au niveau des lysines, une réaction non enzymatique qui se produit généralement en présence d'une forte concentration de sucre.

Un élément à ne pas oublier est que MOS1, produite en cellules d'insectes, en bactéries et en plantes, a toujours le même poids moléculaire. Ceci implique qu'il n'y a pas de modifications post-traductionnelles de grosse taille sur la protéine.

2.4 Ponts disulfures chez MOS1

2.4.1 Rappels sur les ponts disulfures

Avant de regarder en détail si MOS1 contient ou non des ponts disulfures, voici un petit rappel sur le sujet. Un pont disulfure est une liaison covalente entre deux atomes de soufre provenant des chaînes latérales de deux résidus cystéines réduits (figure 75-A). C'est l'oxydation des deux groupements sulfhydriles (-SH) qui donne naissance à ce lien covalent fort. La réunion de ces deux cystéines est appelée une cystine, molécule où les deux carbones asymétriques centraux est comprise entre 4,5 et 7,2 angströms (figure 75-B).

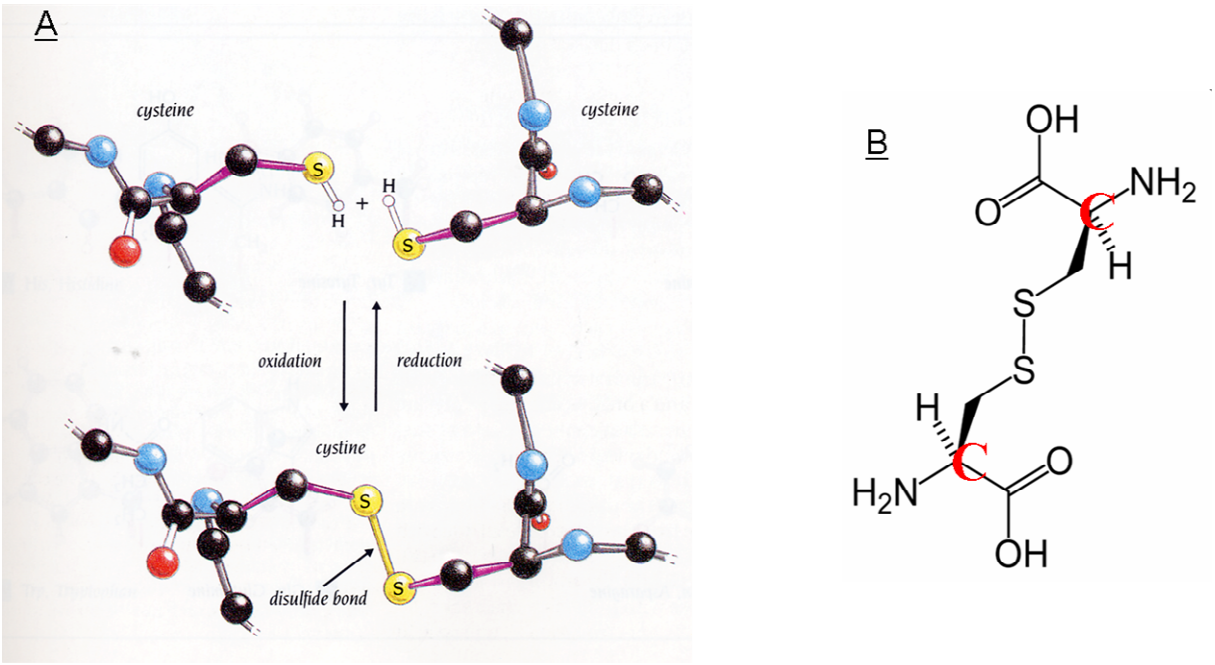


Figure 75. Ponts disulfures. (A) *Formation de ponts disulfures par oxydation de deux cystéines.* **(B)** *Cystine ; la distance entre les deux carbones représentés en rouge est comprise entre 4,5 et 7,2 angströms.*

La formation de ponts disulfures (ou ponts S-S) exige un environnement oxydant, ce qui explique qu'il n'y ait que peu de protéines intracellulaires avec des ponts S-S. Cette modification co- ou post-traductionnelle joue un rôle de stabilisation de la forme 3D des protéines, réduisant ainsi le nombre d'états conformationnels possibles pour la protéine. Ces liaisons permettent aux protéines de conserver leur structure fonctionnelle malgré les fluctuations de l'environnement ; elles peuvent même contribuer aux activités catalytiques chez certaines enzymes (Klink *et al*, 2000).

L'un des aspects les plus intéressants des ponts S-S est qu'ils peuvent connecter entre eux des domaines très éloignés d'un point de vue de la séquence primaire, imposant ainsi des contraintes structurelles fortes aux protéines.

Les ponts S-S peuvent se former de différentes manières :

- par catalyse enzymatique, à l'aide d'enzymes telles que la *Protein Disulfide Isomerase*, des oxidoréductases thiol:protéines ou des sulfhydryl-oxydases qui procèdent par oxydation directe de l'oxygène moléculaire.

- une formation spontanée est possible dans un environnement oxydatif (l'oxydation de l'air est assez rapide pour augmenter substantiellement le ratio de [-S-S-] par rapport à celui de [-SH] en quelques heures).

- par échange thiol-sulfide non catalysé à pH alcalin (figure 76) ; à un pH inférieur à 8, le groupement sulfhydrile [-SH] n'est pas réactif, la forme protonée du thiol étant favorisée par rapport à la forme déprotonée. Par contre, à un pH supérieur à 8, c'est la forme thiolate [-S^-] qui est favorisée, forme réactive qui peut promouvoir l'échange thiol-sulfide.

Il existe trois conformations différentes de ponts S-S, qui dépendent de l'orientation dans l'espace des chaînes des carbones adjacents aux atomes de soufre formant le pont disulfure (figure 77).

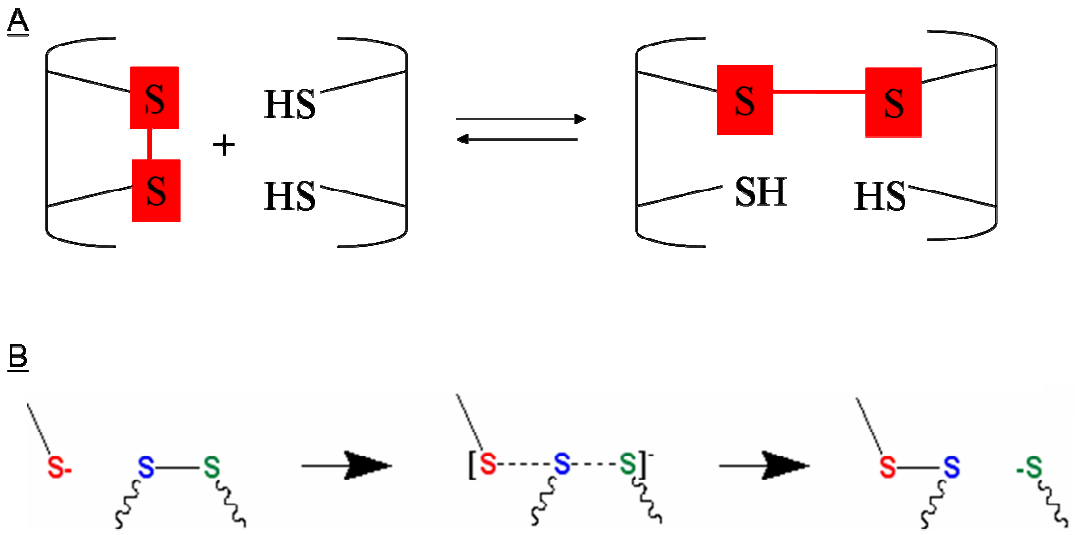


Figure 76. Formation d'un pont disulfure par échange thiol-sulfide. (A) *Exemple d'échange thiol-sulfide entre un pont S-S déjà existant et le groupement sulfhydrile d'une*

cystéine réduite. **(B)** *Echange thiol-sulfide au niveau moléculaire ; la charge négative est partagée entre les trois atomes de soufre dans l'état intermédiaire de la réaction.*

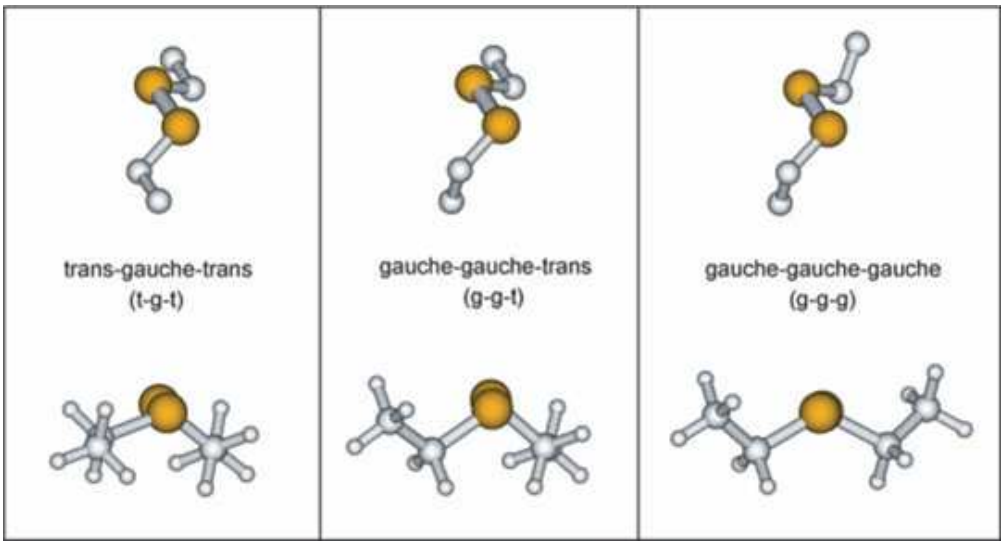


Figure 77. Les trois conformations possibles des ponts disulfures. *La conformation dépend de l'orientation dans l'espace des chaînes des carbones adjacents aux atomes de soufre (représentés en jaune) formant le pont disulfure. Ces conformations s'appellent trans-gauche-trans (tgt), gauche-gauche-trans (ggt) et gauche-gauche-gauche (ggg), et la stabilité augmente avec le nombre de g (stabilité : ggg > ggt > tgt).*

2.4.2 La PPase2A met sur la piste de ponts disulfures chez MOS1

Lors des études sur la déphosphorylation de *Mos1* par différentes phosphatases, le comportement de la PPase2A s'est avéré déconcertant. Cette phosphatase déphosphoryle la protéine mais induit en même temps la protéolyse de la protéine, une diminution du taux de transposition et la modification de son profil de liaison à l'ITR.

2.4.2.1 Impact de la PPase2A sur MOS1

La déphosphorylation de la transposase par la PPase2A provoque une modification du profil de liaison de la Tnp sur ses ITR (figure 69). Une déphosphorylation de la Tnp eucaryote a été réalisée par action de la PPase2A pendant 5, 30, 60 et 120 minutes. Les complexes ITR-Tnp ont ensuite été formés comme précédemment décrit et visualisés en retard sur gel. Dès cinq minutes de déphosphorylation, le complexe B apparaît en dessous du complexe A et devient majoritaire quand la durée du traitement augmente (figure 78).

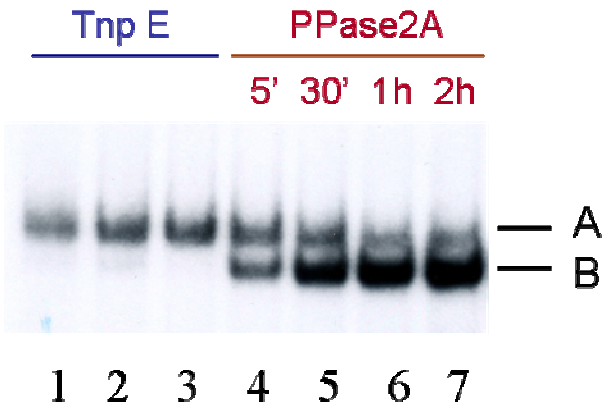


Figure 78. Impact de la déphosphorylation par la PPase2A sur la liaison de MOS1 à l'ADN. *(1 et 2 : témoins, Tnp E incubée à 30°C respectivement 5' et 2h ; 3 Tnp E ; 4 à 7 : Tnp E préalablement déphosphorylée à la PPase2A de 5' à 2h) Un seul complexe est formé en présence de Tnp E (complexe A), tandis que la déphosphorylation par la PPase2A induit la formation du complexe B qui devient majoritaire avec l'augmentation du temps de déphosphorylation.*

Suite à ce résultat, la stabilité de la protéine eucaryote déphosphorylée par la PPase2A a été vérifiée ; les protéines utilisées lors de l'expérience de retard sur gel ont été déposées sur un gel de polyacrylamide qui a été coloré au bleu de coomassie (figure 79). Il en ressort que le traitement à la PPase2A induit la protéolyse de la Tnp eucaryote ; les produits de protéolyse retrouvés chez la protéine produite en bactéries sont également retrouvés chez la protéine eucaryote déphosphorylée par la PPase2A, dès cinq minutes d'incubation avec la phosphatase.

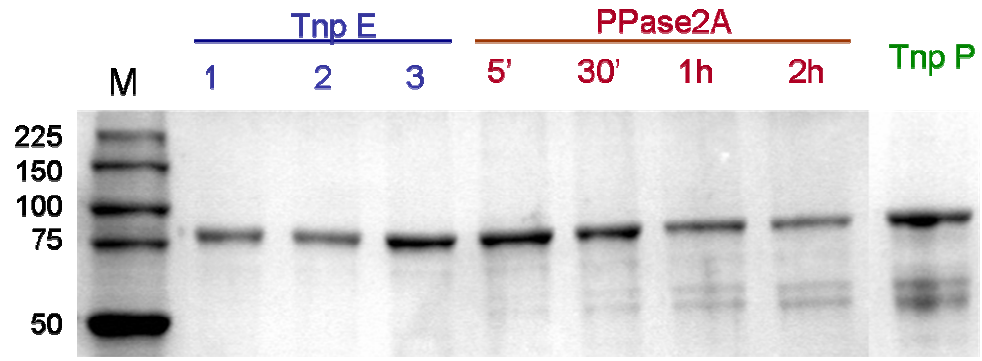


Figure 79. Impact de la déphosphorylation par la PPase2A sur la stabilité de la transposase. *(M : marqueur de poids moléculaire ; Tnp E 1 et 2 : protéine incubée 1h et 2h à 30°C respectivement ; 3 : Tnp E ; PPase2A : cinétique de traitement à la PPase2A de 5' à 2h ; Tnp P : Tnp procaryote avec ses produits de protéolyse) La Tnp eucaryote est stable, même incubée à 30° 1h ou 2h ; par contre, la déphosphorylation parla PPase2A induit la protéolyse dès cinq minutes de réaction.*

Les résultats obtenus après l'utilisation de la PPase2A laissaient penser que la phosphorylation avait un rôle primordial dans la stabilité de la protéine, mais d'autres résultats ont montré que la question était plus complexe. En effet, la déphosphorylation par la λPPase et la PPase1 n'a aucun impact sur la stabilité de la protéine et sur son activité. Dans ce cas-là, comment la PPase2A provoque-elle la mise en place du processus de protéolyse de la

protéine eucaryote ? La littérature ne décrit pourtant aucune activité autre que celle de phosphatase pour cette enzyme. La réponse se trouve dans le tampon de conservation de l'enzyme qui contient une quantité non négligeable de β-mercaptoéthanol (60 mM), un agent réducteur bien connu des ponts disulfures.

L'hypothèse est donc que la transposase de *Mos1* contiendrait un ou des ponts disulfures et que le β-mercaptoéthanol a un impact sur la stabilité et/ou l'activité de la transposase en le réduisant. Pour répondre à cette question, l'impact du β-mercaptoéthanol sur la liaison de la transposase à l'ADN a été testé par retard sur gel. Il est à noter que la protéine MBP, qui est en fusion avec la transposase MOS1 (d'origine procaryote ou eucaryote) ne contient aucune cystéine, et n'est donc par conséquent pas capable de participer à la formation de ponts disulfures.

2.4.2.2 Impact du β-mercaptoéthanol sur la formation des complexes ADN-Tnp

La transposase eucaryote a été préalablement incubée dans les mêmes conditions que celles utilisées pour la déphosphorylation par la PPase2A. Ont été comparées des protéines incubées avec de la PPase2A, avec le tampon de conservation de la PPase2A qui avait été reconstitué (cf partie Matériel et Méthodes), avec ce même tampon sans β-mercaptoéthanol et avec 60 mM de β-mercaptoéthanol, soit la quantité contenue dans le tampon de conservation de la phosphatase. Les complexes ITR-Tnp ont ensuite été formés avec un ITR 3' radiomarqué au en présence de magnésium, 20 minutes à 4°C, comme précédemment décrit. La Tnp eucaryote ne forme que le complexe A, comme la Tnp incubée avec le tampon de conservation de la PPase2A qui ne contenait pas de β-mercaptoéthanol. Les protéines incubées avec de la PPase2A, du β-mercaptoéthanol ou le tampon de conservation de la phosphatase présentent quant à elles un profil de liaison à l'ITR différent, avec l'apparition du complexe B (figure 80).

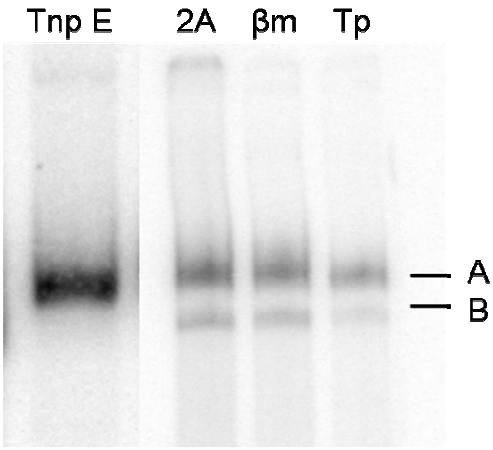


Figure 80. Impact du β-mercaptoéthanol sur la formation des complexes ITR-Tnp. *La Tnp E forme un seul complexe comme décrit précédemment ; la Tnp incubée avec le tampon de conservation de la PPase2A, sans β-mercaptoéthanol, a le même profil. Les Tnp incubées avec la PPase2A, avec le tampon de conservation de la phosphatase, ou avec du β-mercaptoéthanol ont un profil différent, avec l'apparition du complexe B.*

Ce résultat indique que le β-mercaptoéthanol a un impact sur la liaison à l'ADN de transposase eucaryote, renforçant ainsi l'hypothèse émise. Etant donné l'importance des ponts disulfures dans la conformation des protéines, il est logique de penser que la réduction d'un pont disulfure entraîne un changement de conformation et, par conséquent dans le cas de la transposase MOS1, un changement dans l'organisation du complexe formé sur l'ITR.

2.4.2.3 La structure tridimensionnelle rapproche deux cystéines

MOS1 contient cinq cystéines (en positions 18, 46, 136, 194 et 336) ; C18 et C46 ne sont visibles sur la partie C-terminale de MOS1 qui a été cristallisée (Richardson *et al*, 2006), mais les autres si. Et il apparaît au regard de la structure tridimensionnelle de la protéine que les cystéines en positions 136 et 336 sont très proches l'une de l'autre, les atomes de soufre n'étant distant que de 3,2 angströms (figure 81). La cystéine 194 se trouve quant à elle très éloignée des cystéines 136 et 336 (environ 25 angströms). La formation d'un pont disulfure entre les résidus cystéines 136 et 336 semblent dès lors possible.

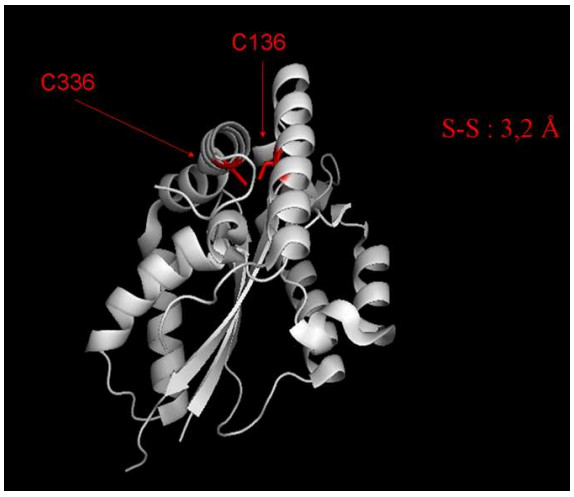


Figure 81. Structure 3D de la partie C-terminale de MOS1. *Les cystéines en positions 136 (C136) et 336 (C336) sont très proches l'une de l'autre.*

2.4.2.4 Approche *in silico*

Des logiciels de prédiction de ponts disulfures sont disponibles sur internet ; il s'agit de logiciels qui se servent de la structure protéique si elle est connue, ou alors ils effectuent une prédiction de la structure protéique avant de prédire la présence ou non de ponts disulfures. Les deux serveurs utilisés ont été DiANNA 1.1 et DIpro 2.0 ; les deux indiquent que toutes les cystéines peuvent former des ponts disulfures. Celui qui ressort dans les deux analyses est un pont entre les cystéines 194 et 336. Or ces deux résidus se situent dans une région cristallisée de la protéine, et il est impossible qu'un pont S-S se forme entre ces deux cystéines, dont les soufres sont distants de plus de 25 angströms (figure 82). De plus, il est apparu lors de la cristallographie du domaine C-terminal que le résidu 194 était enfoui dans la protéine. L'approche *in silico* de l'étude des ponts disulfures ne semble donc pas adaptée à l'étude de la transposase MOS1.

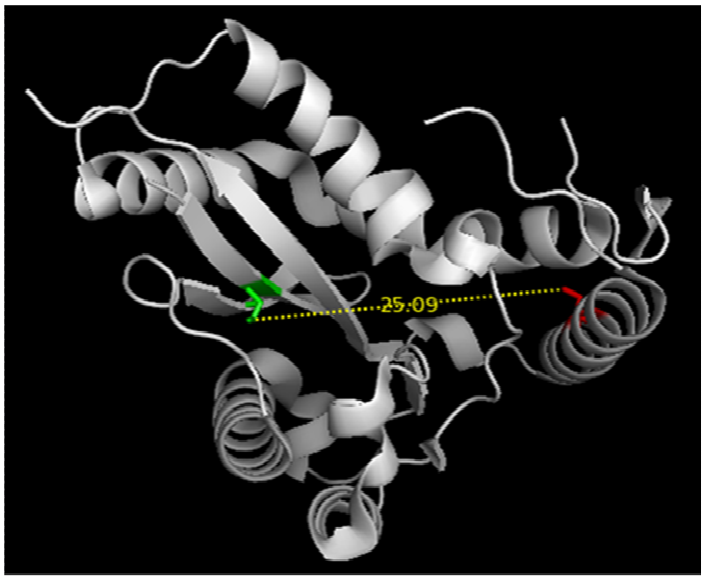


Figure 82. Structure 3D de la partie C-terminale de MOS1. *Le soufre de la cystéine 194 (en vert) est distant de plus de 25 angströms du soufre de la cystéine 336 ; un pont disulfure ne peut donc pas se former entre ces deux résidus cystéines.*

2.4.3 Approche directe : spectrométrie Raman

Parmi les approches directes pour confirmer la présence ou non de ponts disulfures dans une protéine, deux sont couramment utilisées : la spectrométrie de masse et la spectrométrie Raman.

Deux méthodes existent en spectrométrie de masse : une dite classique, qui consiste en un clivage chimique ou enzymatique de la protéine, puis un passage des peptides obtenus en spectrométrie de masse. Elle nécessite de cliver entre chaque cystéine et les conditions usuellement utilisées pour la digestion protéolytique (pH 8) promeut l'échange thiol-sulfide ;

par conséquent les ponts S-S déterminés peuvent être des artefacts. La seconde méthode est plus complexe et met en jeu des réactions de cyanylation préalables au clivage protéique ; elle permet toutefois une localisation efficace des ponts disulfures, mêmes si les cystéines sont très proches dans la séquence. Pour des raisons techniques, aucune de ces deux méthodes n'a pour le moment été testée avec notre transposase.

La spectrométrie Raman fait quant à elle appel à la diffusion Raman, un phénomène optique de diffusion inélastique du photon (par opposition avec la diffusion Rayleigh, qui est une diffusion élastique de la lumière). Il faut savoir que chaque type de liaisons moléculaires correspond à une vibration spécifique. La diffusion Raman se produit sur une infime partie de la lumière (approximativement un sur dix millions de photons), lorsque le milieu diffuse en modifiant légèrement sa fréquence de vibration, ce qui produit un échange d'énergie. Les photons ainsi diffusés par cette légère excitation ont une fréquence différente des autres photons, et c'est cette différence énergétique que mesure la spectrométrie Raman. Elle utilise ces propriétés pour caractériser les protéines et étudier la composition chimique de mélanges tels que des médicaments ou des explosifs. Dans le cadre de la caractérisation des protéines, elle permet de déterminer la structure chimique de la protéine testée, sa conformation avec une quantification relative possible en hélice α, feuillets β et *β-turn*, ainsi que les différentes liaisons entre les atomes (ponts disulfures, interactions de Van der Waals…).

Dans le cadre de la recherche de ponts disulfures, la spectrométrie Raman permet donc de détecter la présence ou non de ces liaisons, ainsi que leur conformation. En effet, chaque conformation de ponts S-S correspond à une fréquence particulière. C'est donc cette méthode que nous avons mis en œuvre pour la transposase de *Mos1*, en collaboration avec le Pr Igor Chourpa (Laboratoire de Chimie Analytique et Hydrologie de la Faculté de Pharmacie de Tours).

Deux transposases MOS1, celle produite en bactéries et celle produite en système eucaryote, ont été analysées en spectrométrie Raman (figure 83).

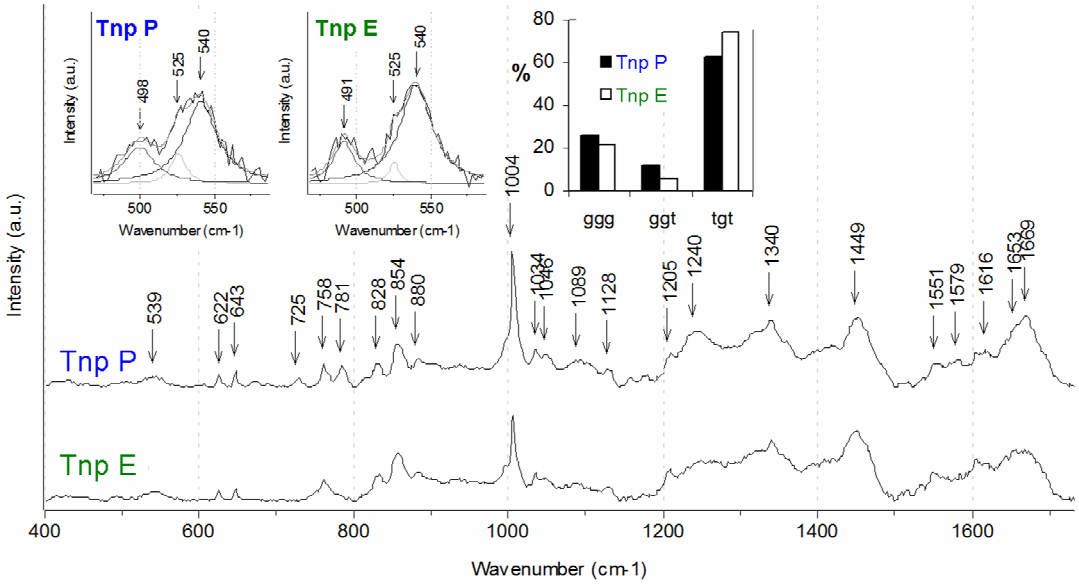


Figure 83. Spectre Raman de MOS1. *(Tnp P = Tnp procaryote ; Tnp E = Tnp eucaryote) Les grossissements au dessus des spectres principaux sont ceux de la région 550/560 nm, dans laquelle se trouvent les pics correspondant aux ponts disulfures. Le graphique représente une quantification relative des différentes conformations de ponts disulfures.*

Ces spectres donnent plusieurs informations :

- il y a une différence de conformation entre la transposase produite en système procaryote et celle produite en système eucaryote

- il y a présence d'au moins un pont disulfure dans la transposase, qu'elle soit produite en système procaryote ou en système eucaryote, mais MOS1 est globalement une protéine pauvre en ponts disulfures

- il y a la même quantité relative de ponts disulfures quelque soit l'origine de la protéine, la seule différence étant dans la conformation de ces ponts S-S (figure 77), où il semble y avoir une plus grande variabilité pour la protéine procaryote

La présence de ponts disulfures chez la protéine procaryote n'est pas une surprise. Lors de la production de la protéine, l'étape de lyse de la paroi bactérienne met la Tnp intracellulaire en contact avec les enzymes du périplasme, seul compartiment bactérien où se trouvent des protéines avec des ponts S-S, dû à la présence de disulfide-isomérases. De plus, lors de cette lyse, les conditions du milieu changent et la protéine pourrait se retrouver dans des conditions oxydantes favorisant la formation de ponts disulfures. Si cette seconde hypothèse s'avère exacte, alors la formation des ponts S-S dans la protéine bactérienne serait

non-enzymatique, conduisant peut-être à une variabilité dans la conformation. La formation des ponts S-S dans la cellule d'insecte se fait dans un milieu moins chaotique, conduisant à une plus grande homogénéité dans la conformation des ponts.

2.4.4 Ponts disulfures et stabilité de MOS1

Lors des études précédentes avec la PPase2A, il avait été montré que cette enzyme avait un impact sur la stabilité de la protéine, et sur le profil de liaison à l'ADN. Nous avons depuis montré que c'était le β-mercaptoéthanol présent dans le tampon de conservation de l'enzyme qui était responsable du changement de profil de liaison à l'ADN. Pour montrer s'il a également un impact sur la stabilité de la protéine, la transposase préalablement incubée avec du β-mercaptoéthanol a été déposé sur un gel de polyacrylamide qui a été coloré au bleu de coomassie (figure 84). Il apparaît que l'action du β-mercaptoéthanol n'a pas d'influence sur la stabilité de la transposase, qu'elle soit d'origine procaryote ou eucaryote.

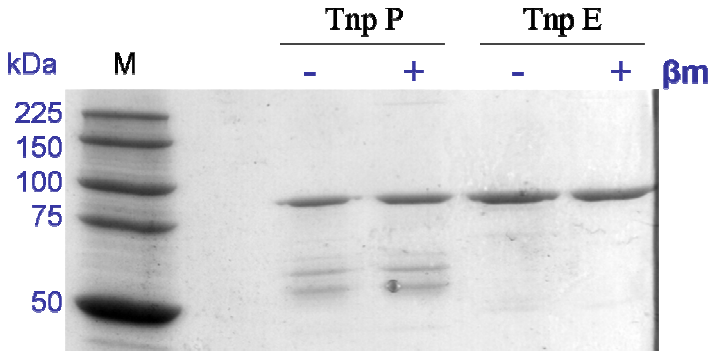


Figure 84. Impact du β-mercaptoéthanol sur la stabilité de MOS1. *(M = marqueur de poids moléculaire ; Tnp P = Tnp procaryote ; Tnp E = Tnp eucaryote ; βm = β-mercaptoéthanol) L'incubation de la Tnp avec 60 mM de β-mercaptoéthanol pendant 1h30 à 30°C n'induit pas de la protéolyse.*

L'incubation de la transposase eucaryote avec la PPase2A conduit à la protéolyse de la protéine. Comme nous avons montré que la déphosphorylation par d'autres phosphatases (λPPase et PPase1) n'induit pas de protéolyse de la Tnp, il était envisageable que ça ne soit que l'action du β-mercaptoéthanol présent dans le tampon de conservation de l'enzyme qui provoque la protéolyse de la Tnp. L'expérience présentée ici montre que ça n'est pas le cas.

Ces résultats sont également renforcés par ceux d'un test de transposition *in vitro* réalisé pour comparer la Tnp baculovirale native à une Tnp préalablement incubée avec du β-mercaptoéthanol. Dans ce test, la fréquence de transposition obtenue pour la Tnp eucaryote native est de $2,6 \times 10^{-4}$ +/- [$8,3 \times 10^{-5}$] et celle de la Tnp incubée avec du β-mercaptoéthanol de $2,9 \times 10^{-4}$ +/- [$1,1 \times 10^{-4}$]. Un test non-paramétrique de Mann & Whitney réalisé sur les fréquences de transposition obtenues sur cinq échantillons indépendants a montré que la

différence n'était pas significative, et que les échantillons avaient la même distribution. Il n'y a donc pas de différence entre la fréquence de transposition de la Tnp eucaryote native et celle de la Tnp eucaryote incubée avec du β-mercaptoéthanol. La différence de taux de transposition observée lors de la déphosphorylation par la PPase2A n'est pas retrouvée ici.

Le lot de PPase2A (*Promega*) semble être la seule cause de protéolyse de la protéine baculovirale. Des expériences complémentaires avec le β-mercaptoéthanol, des phosphatases et le facteur Xa ont été menées et sont résumées dans le tableau 6.

Composés	Protéolyse
PPase2A (β-mercaptoéthanol dans le tampon de conservation de l'enzyme)	oui
β-mercaptoéthanol	non
λPPase, PPase1	non
Facteur Xa (2h)	non
Facteur Xa (24h)	oui
β-mercaptoéthanol + λPPase	non
Facteur Xa (2h) + β-mercaptoéthanol	non
Facteur Xa (2h) + λPPase	non
Facteur Xa (2h) + PPase2A	oui
Facteur Xa (2h) + β-mercaptoéthanol + λPPase	non

Tableau 6. Effet de différents facteurs sur la stabilité de MOS1. *Le tampon de conservation de la PPase2A contient 60 mM de β-mercaptoéthanol, ce qui donne une concentration finale de 4 mM en présence de MOS1. Les différentes concentrations finales de β-mercaptoéthanol utilisées sont de 4, 6,6, 13, 33 et 66 mM.*

Le tag MBP joue le rôle de chaperon de la transposase qui, après 24h d'incubation en présence de facteur Xa, s'auto-protéolyse complètement. La transposase eucaryote se protéolyse toutefois plus lentement que la transposase procaryote, pour qui deux heures d'incubation en présence du facteur Xa sont suffisant pour une protéolyse complète. Cela peut s'expliquer par la différence de conformation entre les deux protéines et/ou des modifications post-traductionnelles.

La PPase2 est quant à elle capable d'induire la protéolyse rapide de la transposase MOS1 eucaryote. Deux hypothèses peuvent être émises :

- une conformation spécifique de cette phosphatase lui permet d'accéder à un site où la phosphorylation joue un rôle protecteur de la protéolyse, un site auquel ne peuvent pas accéder les autres phosphatases

- la présence d'une protéase dans les lots de PPase2A utilisés, mais cette hypothèse semble invalidée par le fait que les produits de protéolyse observés en présence de PPase2A sont les mêmes que lorsque la Tnp s'auto-protéolyse.

2.4.5 Ponts disulfures et interactions protéine-protéine

Un Western Blot réalisé avec des anticorps anti MBP sur la transposase laisse apparaître des oligomères de transposases (figure 85). Des échantillons de Tnp ont été préalablement incubés avec du DTT (1h30 à 30°C), un agent réducteur des ponts disulfures, en présence ou non d'iodoacétamide, un agent alkylant qui se fixe covalemment aux cystéines les empêchant par la suite de former des ponts disulfures. Ces échantillons ont été révélés par Western Blot avec le mélange d'anticorps anti MBP / anti Tnp (figure 85). La présence d'iodoacétamide n'a aucun effet sur les oligomères de Tnp, alors que le DTT fait disparaître les formes oligomériques, avec ou sans iodoacétamide.

L'hypothèse la plus probable est que les ponts disulfures intermoléculaires participent à la stabilisation des oligomères formés par la transposase.

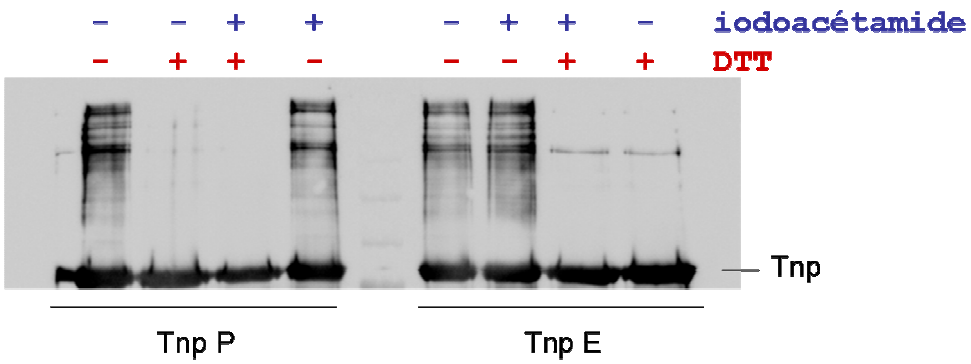


Figure 85. Impact du DTT sur l'interaction transposase-transposase. *(Tnp P = Tnp procaryote ; Tnp E = Tnp eucaryote) Des oligomères de Tnp sont détectables en western blot sur la Tnp native. L'iodoacétamide seul n'affecte pas la formation de ces oligomères. La pré-incubation avec du DTT, avec ou sans iodoacétamide, provoque la disparition de ces formes oligomériques.*

2.4.5 Mutants des ponts disulfures

Des travaux de mutagénèse sont actuellement en cours au laboratoire sur les cystéines de la transposase de *Mos1*. Les mutants C136S et C336S ont été produits en système baculovirus / cellules d'insectes. Les résultats préliminaires montrent que ces transposases mutantes ne sont pas protéolysées. Les deux mutants ont un pattern de liaison à l'ITR différent de celui de la Tnp eucaryote sauvage (figure 86). Le complexe formé par le mutant C136S est légèrement décalé vers le haut, tandis que le C336S se lie très mal à l'ITR. Le fait que les deux mutants n'aient pas le même comportement montre que ça n'est pas uniquement la réduction de l'éventuel pont disulfure entre C136 et C336 qui provoque ces modifications de profil de liaison. La région autour de la cystéine 336 semble importante pour la liaison du dimère de Tnp à son ITR et/ou la stabilisation de ce complexe.

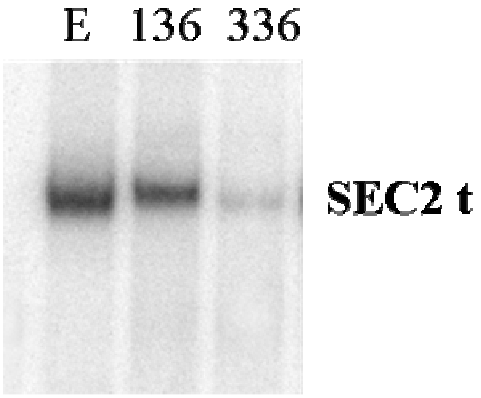


Figure 86. Impact des mutations C136S et C336S sur la formation des complexes ITR-Tnp. *(E : Tnp eucaryote ; 136 : mutant C136S ; 336 : mutant C336S) Le mutant C336S se lie mal sur l'ITR, tandis que la liaison du mutant C136S entraîne un léger déplacement du complexe vers le haut.*

Ceci indique également que le shift de complexe observé en présence de β-mercaptoéthanol (apparition du SEC2 *bottom* au détriment du SEC2 *top* – figure 80) n'est pas dû à la réduction d'un pont disulfure entre les cystéines 136 et 336. Ce shift peut donc être la conséquence de la réduction d'un autre pont disulfure chez MOS1.

2.4.6 Ponts disulfures : conclusion

La spectrométrie Raman a permis de confirmer la présence de ponts disulfures chez MOS1, que la Tnp soit produite en système eucaryote ou procaryote.

A la vue des résultats présentés ici, il s'avère qu'il pourrait y avoir deux types de ponts disulfures chez MOS1 :

- un pont intramoléculaire entre les cystéines 136 et 336 ; cette liaison aurait un rôle important dans la conformation de la protéine, et dissimulerait des acides aminés phosphorylés. Cet assemblage permettrait de stabiliser la protéine et d'empêcher sa protéolyse. Ceci n'empêcherait pas les trois autres cystéines de MOS1 (C18, C46 et C194) de former un autre pont intramoléculaire entre deux d'entre elles, mais l'absence de la structure cristalline de la partie N-terminale ne permet pas de poser une hypothèse à ce niveau.

- des ponts intermoléculaires impliqués dans des phénomènes d'oligomérisation de la transposase. La structure cristalline de la partie N-terminale permettrait de savoir si les résidus C18 et C46 sont exposés à la surface de la protéine ce qui les rendrait apte à former des ponts disulfures intermoléculaires.

Les ponts disulfures ne semblent pas avoir un impact sur l'activité globale de transposition. Ils semblent toutefois avoir un rôle dans l'assemblage des complexes ITR-Tnp, puisque la réduction de la Tnp par le β-mercaptoéthanol fait apparaître un nouveau complexe en retard sur gel, complexe qui correspond au SEC2 *bottom* de la protéine procaryote. La réduction d'un (ou plusieurs) pont disulfure provoque certainement un changement de conformation de la transposase, ce qui pourrait modifier l'assemblage du SEC2.

3/ Inhibiteurs de transposase

Cette partie décrit la découverte d'inhibiteurs de la transposase de *Mos1*, et la caractérisation de leur mode d'action. La recherche d'inhibiteurs de transposase présente plusieurs points d'intérêt :

- dans le cadre de l'étude de la mécanique de transposition, l'existence d'inhibiteurs de transposase pouvant bloquer la réaction à différentes étapes ouvrirait des perspectives intéressantes quant à l'étude moléculaire de la transposition

- avec l'objectif d'utiliser un système transposon/transposase en biotechnologies, l'existence d'inhibiteurs de transposase à DDE pourrait s'avérer utile pour contrôler l'intégration d'un ADN exogène

- enfin, la transposase de *Mos1* pourrait servir de modèle dans le cadre de la recherche d'inhibiteurs d'intégrase du VIH-1, en se basant sur la grande similitude structurale existant entre intégrases rétrovirales et transposases eucaryotes.

Avant mon travail, aucun inhibiteur n'existait pour une transposase eucaryote. Seuls des inhibiteurs de l'élément bactérien Tn*5* avaient été décrits (Ason *et al*, 2005). Une partie des résultats présentés ici ont fait l'objet d'une publication qui est présentée en annexe dans la revue *Mini-Review of Medicinal Chemistry* (Bouchet *et al*, 2009).

3.1 Criblage de composés chimiques sur MOS1

175 composés ont été rassemblés pour mener un crible initial :

- 111 composés sont issus des collections de produits de synthèse organique des laboratoires de chimie thérapeutique de la Faculté de Pharmacie de Tours ; 73 proviennent du laboratoire du Pr Marie-Claude Viaud (GICC – UMR CNRS 6239, équipe 5) et sont numérotés MCV1 à MCV73 et 38 ont été fournis par le Pr Cécile Enguehard-Gueiffier (laboratoire Physicochimie des Matériaux et des Biomolécules – PMBM – EA 4244) et sont numérotés CG1 à CG38.

- 27 composés sont présentés dans la publication d'Ason *et al* (2005) ; 20 sont décrits comme des inhibiteurs de la transposase de l'élément bactérien Tn*5*. Ils sont numérotés R1 à R20 et six d'entre eux inhibent également l'intégrase du VIH-1 *in vitro*. Sept autres sont des dérivés du composé R10, nommés R10-A à R10G, et présentent une activité inhibitrice *in vivo* sur l'intégrase du VIH-1.

- 37 composés ont des origines diverses. Certains possèdent des propriétés inhibitrices de l'intégrase du VIH-1 ou présentent des caractéristiques structurales similaires à des inhibiteurs de l'intégrase ; d'autres n'ont aucune activité inhibitrice connue. Ils sont nommés de D1 à D37.

Le crible de tous les composés s'est effectué sur la transposase produite en bactérie, la seule dont nous disposions au laboratoire quand a été commencée cette étude. Il s'est effectué par le test de transposition *in vitro* (cf le Matériel et Méthodes). Dans la littérature, les inhibiteurs d'intégrase présentent généralement des IC_{50}[4] inférieures à 80 µM ; c'est à partir de ce constat que nos composés ont été testés à une concentration de 80 µM sur la transposase. Tous les composés ont été au minimum testés à trois reprises et la médiane a été calculée. A partir de ces médianes obtenues, nous avons choisi de présenter nos résultats sous deux formes :

- un facteur d'inhibition (FI) qui est le rapport entre la fréquence de transposition du témoin et la fréquence de transposition obtenue avec le composé étudié. Le témoin correspond à la transposase en présence de DMSO, produit dans lequel ont été dilués tous nos composés, et qui n'altère pas la transposition de MOS1

- un pourcentage d'inhibition (PI) qui correspond au résultat de la formule suivante : 100 x (1-1/FI) ; un FI de 2 donne logiquement 50% d'inhibition.

Ces résultats sont présentés dans les tableaux 7 et 8.

[4] IC50 : concentration inhibitrice 50 ; il s'agit de la concentration pour laquelle est obtenu 50% d'inhibition.

Composés	FI	PI
MCV 1	4,78	79,08
MCV 2	1,43	30,23
MCV 3	0,83	0
MCV 4	0,88	0
MCV 5	0,35	0
MCV 6	0,99	0
MCV 7	1,53	34,5
MCV 8	2,24	55,28
MCV 9	4,15	75,92
MCV 10	0,14	0
MCV 11	0,64	0
MCV 12	367	99,73
MCV 13	6	83,33
MCV 14	3,34	70
MCV 15	2,72	63,24
MCV 16	12,74	92,15
MCV 17	9,34	89,29
MCV 18	1,64	39
MCV 19	33	96,97
MCV 20	1,45	31,14
MCV 21	0,77	0
MCV 22	13,66	92,68
MCV 23	22,12	95,48
MCV 24	0,66	0
MCV 25	0,45	0
MCV 26	3,6	72,2
MCV 27	3,1	67,74
MCV 28	7,68	87
MCV 29	9,22	89,15
MCV 30	2,12	52,91
MCV 31	6,07	83,54
MCV 32	0,78	0
MCV 33	1,77	43,56
MCV 34	0,28	0
MCV 35	6,34	84,23
MCV 36	8,69	88,49
MCV 37	0,43	0

Composés	FI	PI
MCV 38	1,03	3,09
MCV 39	2,57	61,14
MCV 40	5,94	83,17
MCV 41	100	99
MCV 42	0,13	0
MCV 43	3,81	73,77
MCV 44	0,16	0
MCV 45	6	83,33
MCV 46	75	98,67
MCV 47	1,22	17,7
MCV 48	5,76	82,63
MCV 49	0,44	0
MCV 50	5,34	81,26
MCV 51	1,2	16,72
MCV 52	0,6	0
MCV 53	2,08	52
MCV 54	23,45	95,73
MCV 55	3,08	67,53
MCV 56	0,69	0
MCV 57	49,22	98
MCV 58	1,11	10,18
MCV 59	8,83	88,67
MCV 60	1,25	20,08
MCV 61	9,44	89,4
MCV 62	7,75	87,1
MCV 63	1	0
MCV 64	5,83	82,84
MCV 65	3,79	73,59
MCV 66	2,42	58,75
MCV 67	2,34	57,23
MCV 68	1,51	33,91
MCV 69	0,61	0
MCV 70	0,85	0
MCV 71	1	1
MCV 72	0,52	0
MCV 73	1	0

Tableau 7. Facteur d'inhibition (FI) et pourcentage d'inhibition (PI) des composés MCV. *En rouge les composés retenus pour la poursuite de l'étude. MCV : composés provenant du laboratoire du Pr Marie-Claude Viaud (GICC – UMR CNRS 6239, équipe 5).*

Composés	FI	PI
CG 1	3,43	70,87
CG 2	0,16	0
CG 3	0,09	0
CG 4	0,03	0
CG 5	0,49	0
CG 6	0,05	0
CG 7	0,15	0
CG 8	0,05	0
CG 9	0,09	0
CG 10	0,07	0
CG 11	0,05	0
CG 12	0,14	0
CG 13	0,21	0
CG 14	1	0,56
CG 15	3,91	74,4
CG 16	2,71	63,14
CG 17	6,94	85,6
CG 18	2,89	65,44
CG 19	1,45	30,88
CG 20	2,03	50,63
CG 21	1,74	42,4
CG 22	2	48,8
CG 23	2	48,8
CG 24	2,66	62,43
CG 25	0,63	0
CG 26	3,26	69,28
CG 27	0,52	0
CG 28	16,17	93,81
CG 29	1,12	10,69
CG 30	5	80,1
CG 31	309,1	99,68
CG 32	321,23	99,69
CG 33	122,13	99,18
CG 34	385,87	99,74
CG 35	0,9	0
CG 36	3,13	68,05
CG 37	249,94	99,6
CG 38	6,6	84,86

Composés	FI	PI
D1	0,99	0
D2	0,77	0
D3	15,94	93,73
D4	2,29	56,26
D5	1,21	17,4
D6	1,96	48,88
D7	1,1	8,81
D8	1,99	49,71
D9	4,5	77,78
D10	0,53	0
D11	2,8	64,24
D12	4	75
D13	1,88	46,86
D14	2,12	52,76
D15	3,18	68,6
D16	1,41	29,14
D17	24,45	95,91
D18	2,42	58,67
D19	0,5	0
D20	48	97,92
D21	40	97,5
D22	0,75	0
D23	0,9	0
D24	4,33	76,92
D25	4,41	77,32
D26	1,25	20
D27	2,54	60,57
D28	0,6	0
D29	588,24	99,83
D30	2,1	52,4
D31	2,25	55,56
D32	3,5	71,43
D33	4,29	76,67
D34	5,63	82,22
D35	6	83,33
D36	6,41	84,4
D37	14,29	93

Composés	FI	PI
R 1	1,78	43,85
R 2	82,5	98,8
R 3	12,44	92
R 4	4,17	76
R 5	5,84	82,86
R 6	3,11	67,8
R 7	1,97	49,24
R 8	0,78	0
R 9	0,45	0
R 10	0,99	0
R 10A	0,66	0
R 10B	1,11	10,15
R 10C	0,31	0
R 10D	0,06	0
R 10E	1,11	10,15
R 10F	1,93	48,26
R 10G	5,6	82,15
R 11	79	98,73
R 12	19,76	94,94
R 13	3,22	69
R 14	2,45	63,66
R 16	25	96
R 17	2,58	61,19
R 18	6,06	83,51
R 19	145	99,31
R 20	5,15	80,6

Tableau 8. Facteur d'inhibition (FI) et pourcentage d'inhibition (PI) des composés CG, D et R. *En rouge les composés retenus pour la poursuite de l'étude. En bleu, des composés possédant des propriétés inhibitrices sur la transposase de Mos1, mais non exploités par la suite pour des raisons de propriété intellectuelle. En vert, le composé D29 dont le cas particulier est décrit plus bas. Les composés CG proviennent du laboratoire PMBM (EA 4244), les composés D sont d'origines diverses, et les composés R sont présentés dans la publication d'Ason et al (2005).*

La majorité des composés ont un effet sur la transposase de *Mos1*, aussi un seuil a dû être fixé pour la suite de l'étude et la sélection des 'meilleurs' composés chimiques. Nous avons arbitrairement choisi de ne retenir que les composés avec un facteur d'inhibition supérieur à 25 (ce qui correspond à 96% d'inhibition). Ce choix a permis de faire ressortir 16 composés :

- quatre sont des dérivés bis(furanyl)*N*-maléimides, produits de synthèse organique provenant de l'équipe 5 du GICC (UMR CNRS 6239) : MCV12, 19, 41 et 46. Ces composés sont de nouveaux inhibiteurs, jusqu'alors non décrits dans la littérature.

- deux sont des dérivés N-methylpyrrole polyamides connus pour interagir avec l'ADN (Dervan & Edelson, 2003) : D20 et D21.

- quatre sont des composés déjà décrits comme des inhibiteurs de la transposase de l'élément bactérien Tn*5* (Ason *et al,* 2005) : R2 11, 16 et 19. Le composé R2 possède également des activités inhibitrices sur l'intégrase du VIH-1.

- cinq sont des composés naturels extraits de la plante africaine *Ouratea sulcata* : CG 31 à 34 et CG37.

- le composé D29 est un cas particulier expliqué dans le chapitre suivant.

Sur ces 16 composés, seuls 10 ont été retenus pour la caractérisation de leur mode d'action sur MOS1. Il s'agit des dérivés maléimides, des polyamides et des inhibiteurs de la Tnp de Tn*5*. Par soucis de clarté, ils ont été renommés de 1 à 10 et sont présentés dans la figure suivante (figure 87). Le devenir des 6 autres inhibiteurs est expliqué dans le paragraphe suivant.

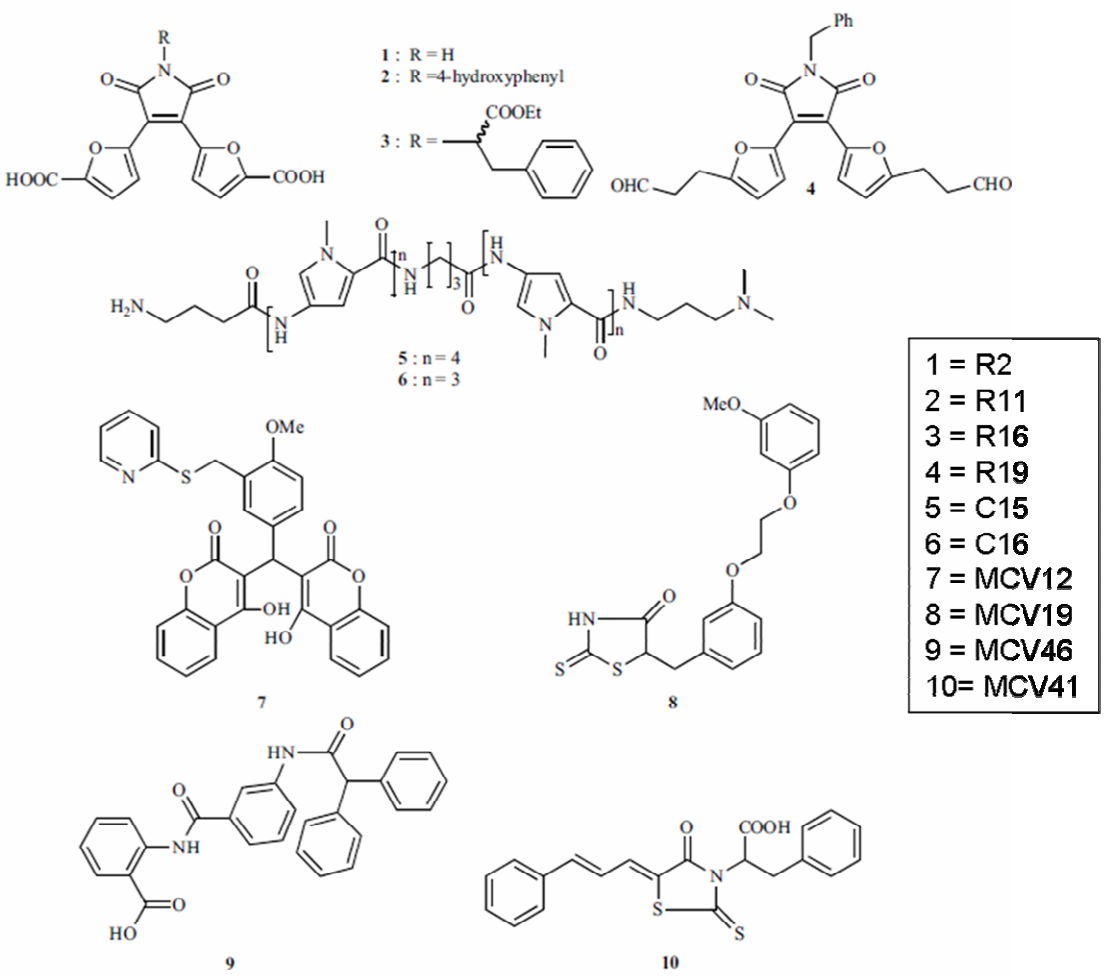


Figure 87. Les 10 composés inhibiteurs de MOS1.

3.2 Inhibiteurs non retenus

Malgré leur facteur d'inhibition supérieur à 25, six composés n'ont pas été retenus pour la suite de l'étude.

Pour les cinq composés extraits de la plante africaine *Ouratea sulcata* (CG31 à 34 et CG 37) il s'agit d'un problème de propriété intellectuelle, ces produits appartiennent à un laboratoire togolais et avait été inclus par erreur parmi les produits de synthèse organique issus du laboratoire PMBM (EA 4244) de la Faculté de Pharmacie de Tours.

Le cas du sixième composé (D29) est plus déconcertant ; tous les composés chimiques décrits comme des inhibiteurs de la transposase de Tn*5* (Ason *et al*, 2005) avaient été commandés à la société *Chembridge*. A l'issu du premier criblage, il ressortait que le composé R14 était un excellent inhibiteur de la Tnp de *Mos1*, inhibant presque totalement la

transposition (PI > 99,5%). Il avait été sélectionné avec les 10 autres inhibiteurs finalement retenus et un deuxième lot de R14 avait été commandé. Il s'est alors avéré que le composé ne possédait plus aucune activité inhibitrice sur MOS1, tout comme un troisième lot recommandé par la suite. Après vérification par RMN, il est apparu que les deuxième et troisième lots correspondaient bien à la formule chimique du R14, tandis que le premier lot (renommé alors D29) contenait un produit chimique contaminant, mais impossible à identifier ! La formule chimique de cet inhibiteur restera donc inconnue et il ne pourra malheureusement pas être exploité.

3.3 Inhibition *in vivo* ?

L'activité de transposition est évaluée *in vitro* ou en bactérie (cf le Matériel et Méthodes). Le crible initial pour la sélection des inhibiteurs a été réalisé par un test de transposition *in vitro*, puis nous avons voulu examiner leur action *in vivo* à t=0 et t=5h de réaction.

Pour le test *in vivo*, nous avons vérifié que le DMSO (dans lequel sont resuspendus tous les inhibiteurs) n'affectait ni la croissance bactérienne ni la fréquence de transposition *in vivo* aux concentrations utilisées, les inhibiteurs retenus (exceptés les composés 5 et 6 pour des raisons techniques) ont été testés à une concentration arbitrairement choisie de 400 µM pour une étude préliminaire.

Que ce soit à t=0 ou t=5h, aucun des composé ne présente une activité inhibitrice *in vivo* sur MOS1 à cette concentration de 400 µM. L'hypothèse que nous avons retenue est que les composés ne réussissent pas à pénétrer la paroi bactérienne et ne peuvent donc pas agir sur la transposase. Il est aussi envisageable que les bactéries puissent dégrader les composés chimiques s'ils arrivent à passer l'obstacle de la paroi bactérienne.

L'inhibition *in vivo* en bactérie de la transposase ne présentant pas d'intérêt dans le cadre de ma thèse, cet aspect n'a pas été étudié plus en détail. Toutefois, le fait de posséder des inhibiteurs de Tnp agissant en bactérie serait intéressant pour la production de transposases hyperactives et cytotoxiques.

3.4 Mode d'action des inhibiteurs

3.4.1 Inhibition de l'excision

Pour caractériser en détail le mode d'actions des 10 inhibiteurs sur la transposition de l'élément *Mos1*, la capacité des composés à inhiber l'excision du transposon a tout d'abord été testée en utilisant le pBC3T3 comme donneur de transposon. La Tnp promeut l'excision du transposon du plasmide en produisant deux fragments linéaires d'ADN, le transposon (3T3) et le *backbone*, qui est le plasmide sans le transposon (figure 88 – ligne 8). Suite à l'excision, le transposon peut être réinséré dans une cible. L'activité inhibitrice des composés chimiques sur l'excision a été mesurée en quantifiant le *backbone* qui est le produit final de l'excision. Le test d'excision effectué en présence du composé 9 est présenté en exemple (figure 88). En l'absence d'inhibiteur, la Tnp permet l'excision de l'élément, qui est visualisable sur gel d'agarose. A la plus haute concentration d'inhibiteur (1 mM dans cette expérience), l'excision du transposon est complètement abolie.

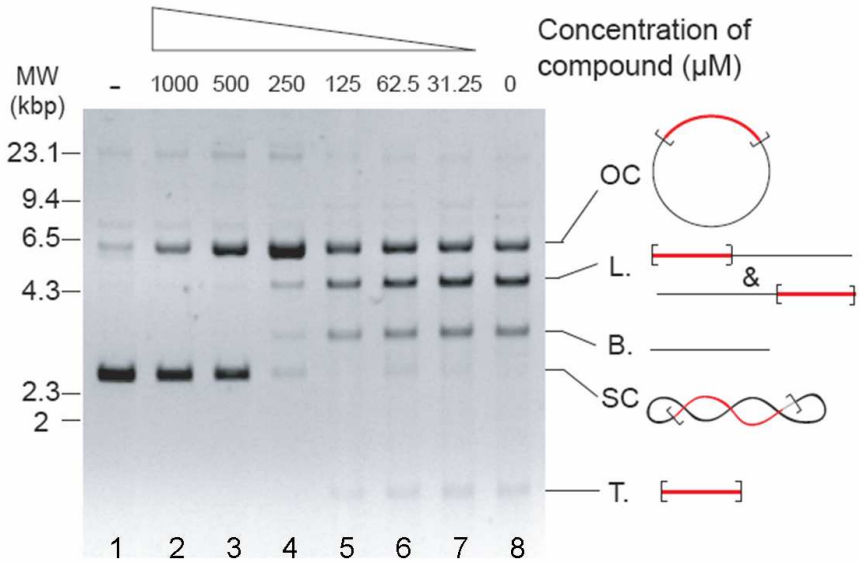
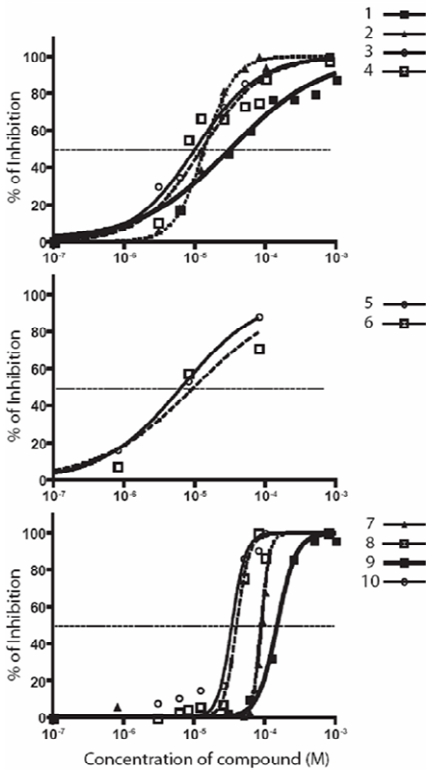


Figure 88. Activité inhibitrice du composé 9 sur l'excision du transposon, d'après Bouchet *et al*, 2009. *Le plasmide pBC3T3 sous forme super-enroulée (SC) est utilisé comme donneur de transposon (ligne 1). En présence de transposase (ligne 8) apparaissent les formes relaxé (OC, consécutive au clivage du premier brin) et linéaire (L, consécutive au clivage du second brin) du plasmide donneur. Le clivage aux deux ITR permet l'excision du transposon ce qui donne le transposon (T) et le backbone du plasmide (B). Tous ces produits ont été observés par électrophorèse sur un gel d'agarose. La réaction s'est effectuée en absence de Tnp (1), en présence de 80 nM de Tnp seule (8), et en présence de 80 nM de Tnp et d'une concentration croissante d'inhibiteurs (2 à 7).*

Pour déterminer la concentration de composé nécessaire pour inhiber 50% de la réaction (IC_{50}), les pourcentages d'inhibition mesurés lors d'au moins deux expériences distinctes ont été mis en relation avec les concentrations d'inhibiteurs utilisées. Les mêmes expériences ont été réalisées pour chacun des 10 inhibiteurs retenus dans notre étude, et les IC_{50} de chacun ont ainsi été obtenues (figure 89). Les composés 2 à 6 se sont avérés être les plus inhibiteurs de l'étape d'excision, avec une IC_{50} aux alentours de 10 µM. Les autres composés, dont les quatre inhibiteurs de la transposase de Tn*5*, agissent moins bien contre l'excision, leur IC_{50} étant comprises entre 35 et 150 µM.

Les 10 inhibiteurs de MOS1 ont un impact certain sur l'excision du transposon. Prenant en compte le fait que les quatre inhibiteurs de la Tnp de Tn*5* (composés 7 à 10) ont été décrits pour agir au niveau de la formation du complexe synaptique (Ason *et al*, 2005), et que les dérivés polyamides (composés 5 et 6) sont de puissants inhibiteurs de la liaison à l'ADN, ces six composés ont été testés pour leur capacité à inhiber la formation des complexes ITR-transposase, une étape qui précède l'excision.



Compound	IC50 Exc. (µM)	95% Confidence Intervals
1	33	18 < IC50<54
2	14	12 < IC50<15
3	10	7 < IC50<13
4	12	9 < IC50<18
5	6	5 < IC50<12
6	10	4 < IC50<22
7	89	86 < IC50<91
8	40	36< IC50<44
9	152	138 < IC50<166
10	34	29 < IC50<39

Figure 89. Activité inhibitrice des 10 composés, d'après Bouchet *et al*, 2009. *Le pourcentage d'inhibition des 10 composés a été mis en relation avec les concentrations testées, et les graphiques ont été réalisés avec le logiciel GraphPad Prism, ce qui a permis d'estimer les IC_{50} d'excision pour chaque inhibiteur, qui sont présentées dans le tableau ci-contre.*

3.4.2 Inhibition de l'assemblage des complexes ITR-transposase

L'étude de la formation des complexes entre la Tnp et l'ITR 3' est visualisable en retard sur gel. Les complexes ITR-Tnp ont été assemblés avec un ITR 3' (70) radiomarqué, en présence d'une concentration de composés équivalente à cinq fois l'IC_{50} d'excision, la concentration de Tnp utilisée pour ces tests étant de 400 nM (soit cinq fois celle utilisée lors des tests d'excision) (figure 90). Les concentrations des composés étaient donc de 65 µM pour les inhibiteurs 5 et 6, 440 µM pour le composé 7, 200 µM pour le 8, 750 µM pour le 9 et 170 µM pour le composé 10. Les pourcentages de liaison ont été quantifiés comme le pourcentage de sonde retenue dans les complexes. Le pourcentage d'inhibition correspond à la perte de liaison de la Tnp à l'ITR 3' en présence des inhibiteurs.

Il est à noter que les travaux sur les inhibiteurs ont été réalisés antérieurement aux travaux sur la mécanique de transposition présentés dans le premier chapitre de la partie Résultats. Ces travaux non publiés montrent que le complexe PEC1 est en fait une autre conformation du SEC2. Par conséquent, l'appellation des complexes SEC1, SEC2 et PEC1 présenté dans Bouchet *et al* (2009) et reproduit ici dans la figure 90 correspond à la nomenclature du profil de liaison ITR-Tnp publiée par Augé-Gouillou *et al* (2005b).

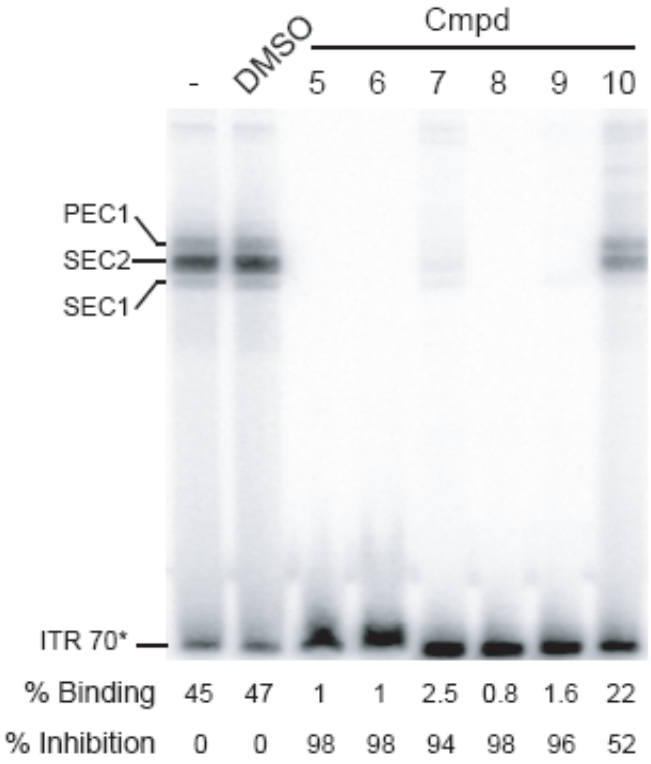


Figure 90. Impact des inhibiteurs 5 à 10 sur la formation du complexe ITR-transposase, d'après Bouchet *et al*, 2009. *(Cmpd = composé) Le test a été effectué avec 400 nM de transposase et 5nM d'ITR3' (70) radiomarqué. Les pourcentages de liaison et d'inhibition sont indiqués en dessous de l'image, les complexes sur le côté.*

Cinq composés montrent une action inhibitrice quasi-totale sur la formation du complexe Tnp-ITR (composés 5 à 9), tandis que le composé 10 a une action inhibitrice plus réduite bien qu'importante (plus de 50%).

Le résultat obtenu pour les deux dérivés polyamides (composés 5 et 6) n'est pas surprenant car ces composés chimiques sont connus pour interagir avec l'ADN. Cet effet se traduit ici par une légère modification du profil de migration de la sonde libre en bas du gel. Ces deux inhibiteurs ciblent donc l'ADN et agissent certainement comme compétiteurs avec la transposase pour lier l'ITR.

Le résultat obtenu pour les composés 7, 8, 9 et 10 est également en accord avec ce qui est décrit dans la littérature où ces composés sont présentés comme des inhibiteurs de la formation du complexe synaptique de Tn*5* (Ason *et al*, 2005).

L'étape suivante a été d'étudier les dérivés maléimides (composés 1 à 4), qui s'avèrent être des nouveaux inhibiteurs au regard de la littérature. Il a d'abord fallu vérifier qu'ils n'agissaient pas sur l'ADN comme les composés 5 et 6, mais bien directement sur la transposase. Deux méthodes ont été utilisées : une approche directe en testant la liaison des composés à l'ADN par un retard sur gel et une méthode indirecte en utilisant le test d'excision avec de l'ADN compétiteur en excès. La première expérience a été menée avec 10 nM d'ITR3' (70) radiomarqué et les inhibiteurs en excès (400 µM), avec les composés 5 et 6 en témoin positif et le DMSO en témoin négatif (figure 91-A). Le test d'excision a été réalisé comme décrit précédemment, la différence a été l'ajout d'ADN compétiteur en excès (cinq fois) et l'utilisation des inhibiteurs à une concentration de 40 µM (figure 92-B). Ces résultats indiquent que les quatre dérivés maléimides n'interagissent pas avec l'ADN, et qu'ils sont donc bien des inhibiteurs spécifiques de la transposase.

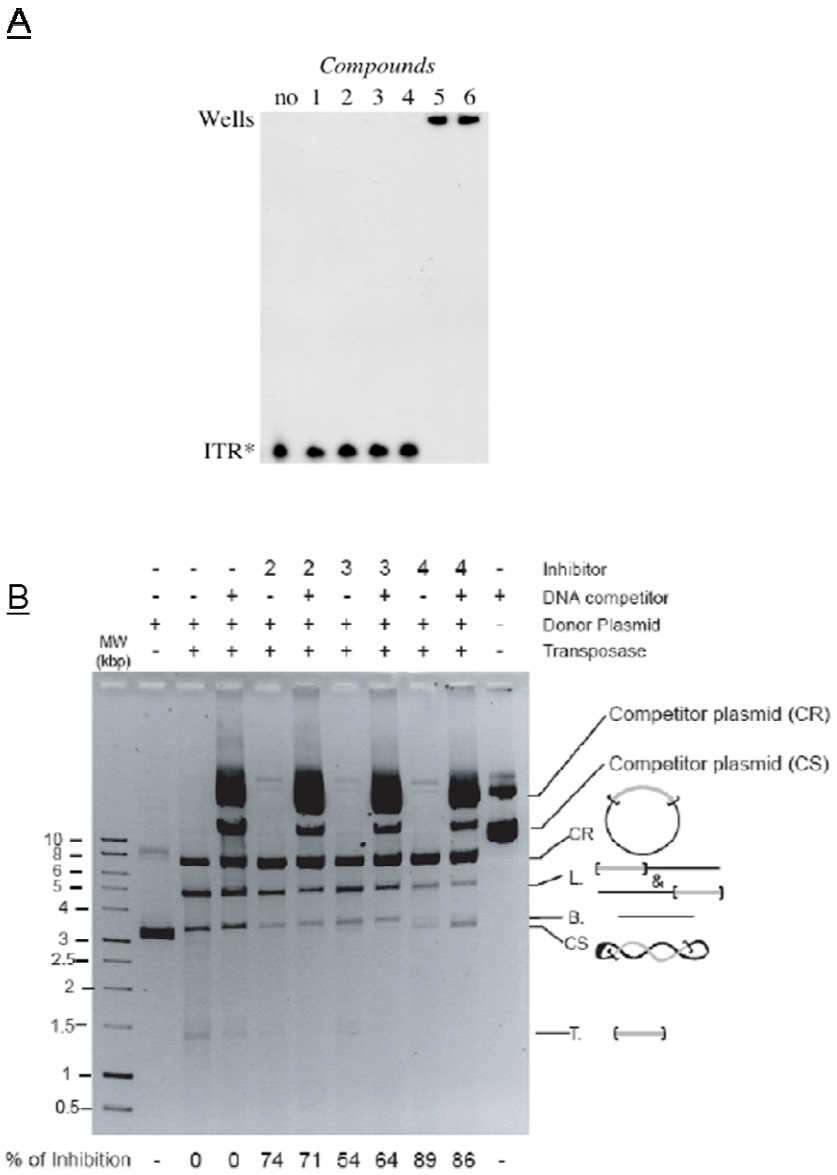


Figure 92. Les dérivés maléimides n'interagissent pas avec l'ADN, d'après Bouchet *et al*, 2009. (A) *Les composés 5 et 6 interagissent avec l'ADN, piégeant l'ITR dans les puits, ce qui n'est pas le cas pour les composés 1 à 4.* **(B)** *La présence d'ADN compétiteur ne modifie pas le pourcentage d'inhibition des dérivés maléimides, démontrant que ces composés chimiques n'interagissent pas avec l'ADN. Le même résultat a été obtenu avec le composé 1 (non montré ici).*

Cette démonstration effectuée, l'aptitude des composés 1 à 4 à inhiber la formation des complexes ITR-transposase a été testée. Les complexes ITR-MOS1 ont été formés avec un ITR3' (70) radiomarqué en présence d'une gamme de concentrations d'inhibiteurs allant de 0 à 400 µM puis analysés par retard sur gel (figure 93). Là encore la nomenclature SEC1, SEC2 et PEC1 fait référence à la publication d'Augé-Gouillou *et al* (2005b).

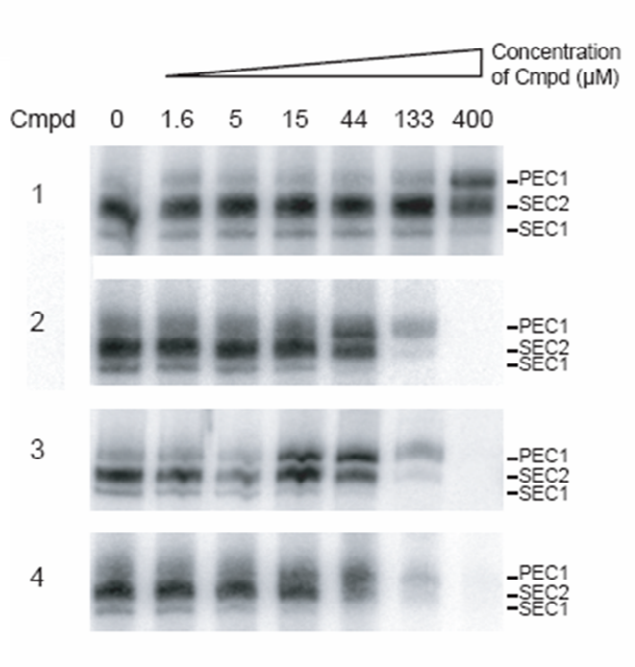


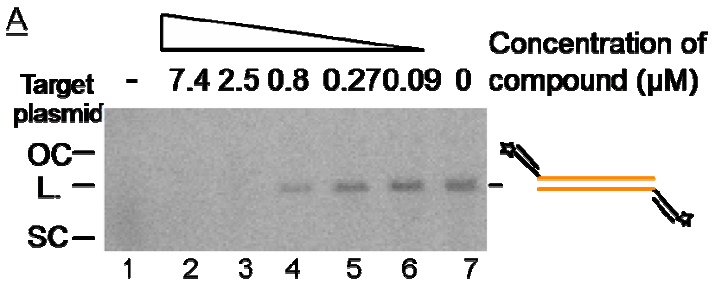
Figure 93. Impact des dérivés maléimides sur la formation des complexes ITR-MOS1, adapté de Bouchet *et al*, 2009. *Les complexes ont été formés en présence de concentrations différentes d'inhibiteurs (de 1,6 à 400 µM). Les numéros des composés sont indiqués sur le côté gauche de la figure. Les complexes sont indiqués sur la droite.*

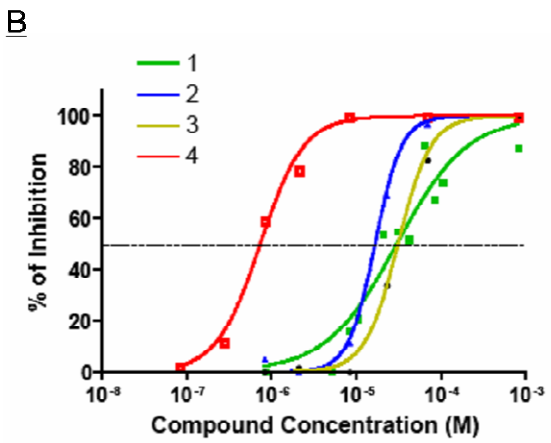
Le PEC1 correspond en réalité au SEC2 top, le SEC2 au SEC2 bottom.

A forte concentration, le composé 1 modifie le ratio des complexes ITR-MOS1, en augmentant la quantité du complexe PEC1 au détriment du SEC2. Le PEC1 étant en réalité une conformation du SEC2, appelée SEC2 *top*, cela signifie que cet inhibiteur induit un changement de conformation du *Single End Complex* quand il est apporté à forte concentration. Il est de même pour les composés 2 à 4, qui changent aussi le ratio des complexes mais à plus faible concentration : 15 µM pour le composé 3, 44 µM pour les composés 2 et 4. A des concentrations plus élevées, ces composés inhibent la formation des complexes ITR-MOS1 (133 et 400 µM). Ces données sont en accord avec les IC_{50} d'excision précédemment obtenues puisque les composés 2 à 4 étaient plus efficaces pour inhiber l'excision de *Mos1* que le composé 1. L'inhibition de la formation des complexes nécessite toutefois de fortes concentrations de composés ; aux IC_{50} d'excision ils n'inhibent donc pas la formation des complexes.

3.4.3 Effet des dérivés maléimides sur le transfert de brin

L'excision et le transfert de brin sont deux réactions consécutives dans la transposition, le transfert de brin étant également l'ultime étape de la transposition. Il a semblé intéressant de tester les nouveaux inhibiteurs à structure maléimide (composés 1 à 4) sur l'intégration. Pour ce faire, un ITR pré-clivé radiomarqué en 5' sur le brin transféré a été utilisé pour former le complexe d'intégration. Les inhibiteurs ont été ajoutés avant la formation des complexes MOS1-ITR. Les évènements d'intégration des ITR dans un plasmide cible ont été visualisés sur un gel d'agarose (montré pour le composé 1 dans la figure 94-A). Un produit de réaction, co-migrant avec la forme linéaire du plasmide, a été détecté en présence de MOS1 et du cofacteur cationique Mg^{2+}. Le pourcentage d'inhibition des nouveaux inhibiteurs a été déterminé par quatre expériences indépendantes puis a été représenté en fonction de la concentration en inhibiteurs (figure 94–B). Les composés 1 à 3 inhibent l'intégration de l'ITR avec des IC_{50} comparables à celles obtenues pour l'excision ; le composé 4 inhibe quant à lui quinze fois mieux le transfert de brin que l'excision (figure 94-C). Cela indique que le composé 4 a une activité supplémentaire par rapport aux composés 1 à 3. Cette activité pourrait être l'inhibition de la capture de cible ou une inhibition directe du transfert de brin lors de la catalyse. Cette activité est originale par rapport aux données de la littérature et revêt un aspect important puisque ces deux processus, capture de cible et transfert de brin, sont les étapes majeures de l'intégration. Cette étape ultime de la transposition est peu documentée dans le monde des transposases et correspond à un projet de recherche en cours de développement au sein du laboratoire.





C

Compound	IC_{50} (µM)	95% confidence intervals
1	30	19< IC_{50}<45
2	16	15< IC_{50}<18
3	30	27< IC_{50}<34
4	0.7	0.6< IC_{50}<0.9

Figure 94. Impact des dérivés maléimides sur le transfert de brin, d'après Bouchet *et al*, 2009. (A) *Activité inhibitrice du composé 4 sur le transfert de brin. Les réactions ont été menées avec un ITR pré-clivé radiomarqué en 5' sur le brin transféré ; le plasmide pBC-KS a été utilisé comme plasmide cible sous forme super-enroulée (SC). Après avoir été stoppées avec de l'EDTA et chauffées 10' à 75° les réactions ont été déposées sur un gel d'agarose 1%. L'intégration concertée des deux extrémités du transposon (deux ITR) au site cible donne un produit linéaire (L), tandis que l'hémi-intégration d'une extrémité du transposon (un ITR) aurait donné un plasmide relâché (OC). La ligne 1 est le témoin négatif (sans Tnp), la ligne 7 contient de la Tnp seule, les lignes 2 à 6 une concentration décroissante d'inhibiteur.* **(B)** *Activité inhibitrice des composés 1 à 4, chacun représenté dans une couleur distincte comme indiqué en haut du graphique, réalisé avec le logiciel GraphPad Prism.* **(C)** *IC_{50} obtenues pour l'intégration.*

3.5 Effet des dérivés maléimides sur l'intégrase du VIH-1

La notion de cross-activité des inhibiteurs est importante dans le monde des enzymes à triade catalytique DDE. Des composés inhibant la transposase de l'élément bactérien Tn*5* se sont avérés capables d'inhiber l'intégrase du VIH-1 (Ason *et al*, 2005) et également la transposase de *Mos1* (Bouchet *et al*, 2009). Les travaux présentés précédemment ont permis la découverte d'une nouvelle classe d'inhibiteurs, les dérivés maléimides, capables d'agir sur la transposase au niveau de l'excision et/ou de l'étape d'intégration. La capacité de ces nouveaux inhibiteurs à pouvoir agir sur une autre enzyme à DDE a été testée, en l'occurrence sur l'intégrase du VIH-1.

Ces essais ont été menés au Laboratoire de Biologie et Pharmacologie Appliquée de Cachan (LBPA - UMR CNRS 8113) en collaboration avec le Dr Olivier Delelis. Les nouveaux inhibiteurs ont été testés contre les activités de *3'-processing* et de transfert de brin de l'intégrase du VIH-1 (cf le chapitre consacrée à la mécanique des enzymes à DDE dans la partie Introduction). Les tests ont été réalisés en utilisant une intégrase recombinante purifiée et de courts oligonucléotides comme substrats, les produits étant visualisés sur un gel dénaturant (protocole décrit dans Craigie *et al*, 1991). Le gel est montré ici pour le composé 4 (figure 95-A) et la même expérience a été réalisée pour les trois autres dérivés maléimides. Les courbes représentant les pourcentages d'inhibition trouvés lors de trois expériences indépendantes pour chacun des quatre composés en fonction des concentrations d'inhibiteurs ont été tracées (figure 95-B et C). Ceci a permis le calcul des IC_{50} pour le *3'-processing* et le transfert de brin (tableau 9).

Le composé 1 qui est le moins efficace des quatre dérivés maléimides contre la Tnp de *Mos1* est à nouveau le moins actif des quatre contre l'intégrase du VIH-1. Il présente des IC_{50} de 116 µM et 58 µM pour le *3'-processing* et le transfert de brin respectivement. Au contraire, les composés 2, 3 et 4 sont efficaces contre les activités de *3'-processing* et de transfert de brin. Les IC_{50} sont inférieures à 30 µM pour la première activité et inférieure à 10 µM pour la seconde. Le composé 4 présente la particularité d'être cinq fois plus efficace contre le transfert de brin que contre le *3'-processing* (IC_{50} de 5,5 µM contre 29 µM), ce qui contraste avec les composés 2 et 3. Ce résultat est en accord avec ce qui avait été observé pour la transposase de *Mos1*, système dans lequel le composé 4 inhibait l'intégration quinze fois mieux que l'excision.

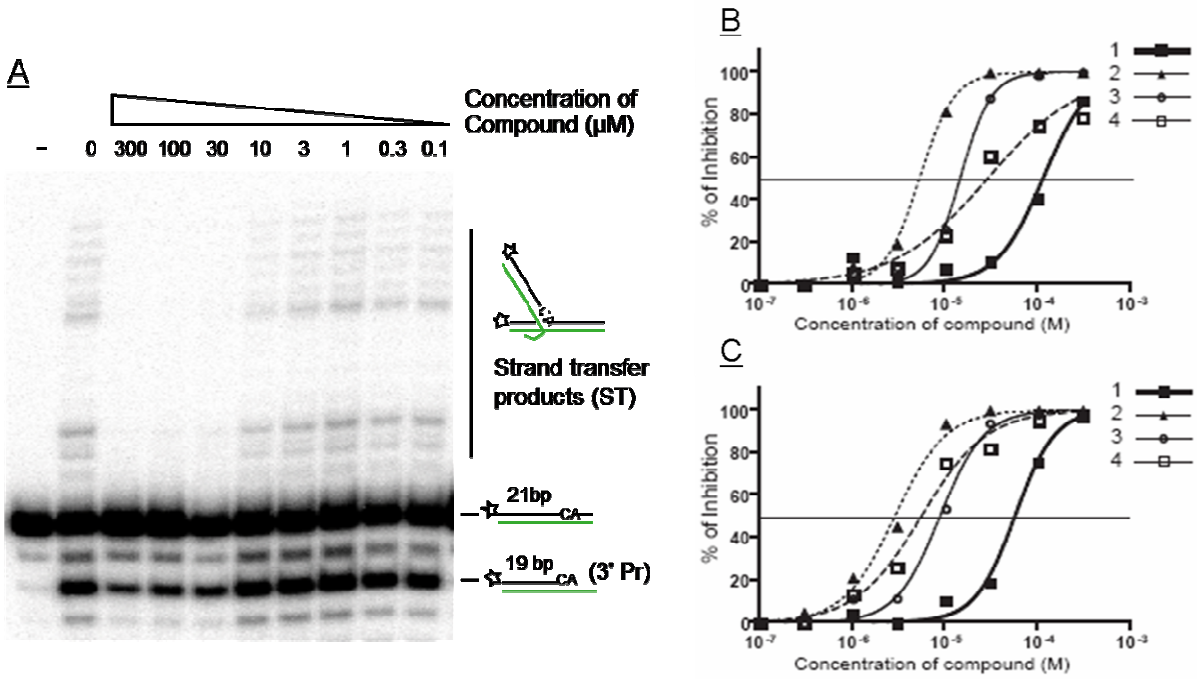


Figure 95. Activité inhibitrice du composé 4 sur les activités de *3'-processing* et de transfert de brin de l'intégrase du VIH-1, d'après Bouchet *et al*, 2009. (A) *Le test a été mené avec une intégrase du VIH-1 recombinante et purifiée et un oligonucléotide mimant l'extrémité LTR radiomarqué en 5' du brin processé. Les réactions ont été stoppées avec de l'EDTA et mis en électrophorèse sur un gel de polyacrylamide de 18% en présence de 7M d'urée. La réaction de 3'-processing donne un produit de 19 pb qui est le substrat pour la réaction de transfert de brin. Les réactions ont été faites en absence d'intégrase (-), en présence d'intégrase seule (0) et avec l'intégrase en présence de différentes concentrations d'inhibiteurs (de 0.1 à 300 µM).* **(B ; C)** *Les pourcentages d'inhibition obtenus ont été mis en relation avec la concentration des composés pour les quatre inhibiteurs, et les courbes ont été tracées en utilisant le logiciel GraphPad Prism.* **(B)** *Activité de 3'-processing des inhibiteurs.* **(C)** *Activité de transfert de brin des inhibiteurs.*

Composé	IC_{50} *3'-processing* (µM)		IC_{50} transfert de brin (µM)	
1	116	59< IC_{50}<228	58	35< IC_{50}<94
2	5,3	4,6< IC_{50}<5,1	3	2,3< IC_{50}<3,6
3	15	10< IC_{50}<19	9	6< IC_{50}<13
4	29	14< IC_{50}<60	5,5	4< IC_{50}<7,7

Tableau 9. Effet des dérivés maléimides sur l'intégrase du VIH-1, d'après Bouchet *et al*, 2009. *Les valeurs des IC_{50} pour le 3'-processing et le transfert de brin sont indiquées en µM, avec les intervalles de confiance à 95%.*

Les dérivés maléimides que nous avons décrits comme des inhibiteurs de la transposase de l'élément *mariner Mos1* présentent également des activités inhibitrices sur l'intégrase du VIH-1. Ceci souligne la cross-activité de ce nouveau groupe d'inhibiteurs parmi les enzymes à triade catalytique DDE.

3.6 Inhibiteurs : conclusions

Disposer d'inhibiteurs pour analyser et contrôler des mécanismes minutieux comme la transposition est d'une importance certaine. Le travail présenté dans ce chapitre correspond à la première caractérisation de composés chimiques inhibant la transposition de l'élément *mariner Mos1* qui appartient à la famille de transposons la plus répandus chez les eucaryotes. Après un premier crible, dix composés ont été retenus et leur mode d'action a été caractérisé.

3.6.1 Inhibiteurs de Tn*5* et polyamides

Quatre de ces dix composés (numérotés de 7 à 10) avaient été décrits dans la littérature comme des inhibiteurs de la transposase de Tn*5* (le 7 agissant également sur l'intégrase du VIH-1). Ces quatre composés présentent un mécanisme d'action similaire contre les Tnp de *Mos1* et Tn*5*. Cependant, ils moins actifs sur MOS1, ce qui laisse penser qu'il existe des différences entre l'organisation des complexes et la reconnaissance des ITR chez les deux enzymes.

Les composés 5 et 6 ont un mode d'action en accord avec celui décrit dans la littérature, à savoir qu'ils interagissent directement avec l'ADN, inhibant ainsi de manière indirecte la transposition de *Mos1*. Il s'agit de dérivés polyamides qui s'insèrent dans le petit sillon de l'ADN. Ils interagissent de manière spécifique avec une courte séquence nucléotidique de trois ou quatre A/T ; ces séquences sont localisées sur les ITR de *Mos1* ce qui explique les effets inhibiteurs de ces composés sur la transposition.

3.6.2 Les dérivés maléimides

Les dérivés bis-(hétéroaryl)maléimides (composés 1 à 4) constituent une nouvelle famille d'inhibiteurs des enzymes à DDE. Ce noyau maléimide semble être important dans l'inhibition. Ainsi les dérivés qui ne sont pas substitués ne présentent aucune activité inhibitrice ou alors relativement faible, ceci laissant penser que le noyau maléimide n'est pas le pharmacophore. La *N*-substitution ainsi que la substitution hétéroaryl ont un impact majeur sur le mécanisme d'action de ces composés.

Le **composé 1**, qui n'est pas *N*-substitué, n'inhibe pas l'assemblage du complexe ITR-MOS1. Il a un simple impact sur la conformation des complexes ; il présente des activités inhibitrices similaires sur l'excision et le transfert de brin, avec à chaque fois une IC_{50} supérieure aux autres dérivés maléimides testés.

Les **composés 2 et 3**, qui sont *N*-substitués, possèdent des IC_{50} d'excision et de transfert de brin identiques Ils agissent cependant différemment sur l'assemblage du complexe ITR-MOS1. Le composé 3 provoque un changement de la conformation du SEC2 à plus faible concentration que le composé 2.

Le **composé 4** est également *N*-substitué, mais avec en plus un groupement N-formyl contrairement aux composés 1 à 3 qui ont des fonctions acides. A l'inverse des composés 1 à 3, il est quinze fois plus efficace sur le transfert de brin que sur l'excision de *Mos1*. La recherche de nouvelles substitutions sur ce composé est engagée afin de préciser leur rôle sur l'action inhibitrice et surtout pour trouver un inhibiteur original spécifiquement ciblé sur le transfert de brin.

3.6.3 Cross-activité des inhibiteurs des enzymes à DDE

La capacité de composés chimiques à inhiber des enzymes reliées entre elles par un mécanisme d'action commun, comme les enzymes à DDE, a déjà été décrit dans la littérature. Dans le cas des enzymes à DDE, des composés ont montré des activités inhibitrices contre plusieurs intégrases rétrovirales (Lubkowski *et al*, 1998). Une autre étude avait montré la cross-activité des dikéto-acides contre l'intégrase du VIH-1 et les protéines RAG1/2 (Melek *et al*, 2002). La recherche d'inhibiteurs de la transposase de l'élément bactérien Tn*5* avait permis d'identifier des composés inhibant également l'intégrase du VIH-1.

Parmi les inhibiteurs de MOS1 mis à jour lors de ma thèse, certains présentent des propriétés inhibitrices de la transposase de Tn*5*, d'autres de l'intégrase du VIH-1, et un des trois enzymes. Il s'agit d'une nouvelle démonstration de la cross-activité des inhibiteurs des enzymes à DDE. Mon travail place la transposase de *Mos1* comme un modèle dans le cadre de la recherche d'inhibiteurs d'enzymes à DDE.

Discussion & Perspectives

Mes travaux de thèse ont abordé l'étude de la transposase de l'élément *mariner Mos1* par différents aspects qui sont toutefois étroitement imbriqués les uns dans les autres. Ils ont permis d'intégrer de nouvelles connaissances fondamentales sur la transposase MOS1 et sur le mécanisme de transposition des éléments *mariner*. La découverte d'inhibiteurs de MOS1 capables d'agir également sur la transposase de Tn*5* et l'intégrase du VIH-1 ouvre la voie à de nouveaux projets de recherche sur les enzymes à DDE. Quant aux nouveaux éléments sur la régulation de la transposase qui ont été mis à jour, ils permettent de poser les bases de nouveaux projets concernant l'étude de la régulation de *Mos1* par des modifications post-traductionnelles et d'éventuelles interactions avec des protéines spécifiques de l'hôte. Ces données remettent également en cause l'utilisation de *Mos1* comme vecteur de transfert de gènes chez les mammifères.

Inhibiteurs de MOS1

Mes travaux de thèse ont permis de mettre à jour **les premiers inhibiteurs d'une transposase eucaryote**. Parmi eux, des nouveaux inhibiteurs dans le domaine des enzymes à DDE qui inhibent également l'intégrase du VIH-1, ainsi que des inhibiteurs déjà décrits pour inhiber la transposase de Tn*5* et l'intégrase du VIH-1. Ces travaux démontrent l'activité croisée de ces composés, capables d'inhiber des enzymes reliées par leur mécanisme. Cette capacité de cross-activité sur les enzymes à DDE avait déjà été montrée entre des intégrases rétrovirales et les protéines RAG (Lubkowski *et al*, 1998 ; Melek *et al*, 2002), et entre l'intégrase du VIH-1 et la Tnp de l'élément bactérien Tn*5* (Ason *et al*, 2005).

L'aspect le plus intéressant de ce travail est **la découverte de nouveaux inhibiteurs d'enzymes à DDE** que sont les dérivés maléimides. Ceci ouvre la porte à l'utilisation de MOS1 comme un modèle dans le cadre de la recherche d'inhibiteurs d'enzymes à DDE dont les activités sur la physiologie des cellules (intégrase du VIH-1) ou la recombinaison d'ADN potentiellement associée à des pathologies ou désordres biologiques (phénomène de réparation d'ADN impliquant des néogènes) sont du domaine de la santé publique. L'utilisation de MOS1 comme modèle est intéressante pour plusieurs raisons. D'une part parce que la transposition de MOS1 se passe dans un environnement eucaryote (contrairement à Tn*5*, également proposé comme modèle – Barreca *et al*, 2007 ; Czyz *et al*, 2007) et que MOS1 et l'intégrase du VIH-1 ont des domaines catalytiques très proches. D'autre part parce

que le clivage de l'ADN promu par MOS1 est similaire à celui promu par les intégrases rétrovirales, sans formation d'un intermédiaire en *hairpin* comme c'est le cas lors du clivage chez Tn*5*. C'est pour ces raisons que **MOS1 constitue un meilleur modèle de l'intégrase du VIH-1 pour la recherche d'inhibiteurs** que la transposase de Tn*5*. Des différences existent toutefois entre MOS1 et l'intégrase du VIH-1, comme la structure des domaines de liaison à l'ADN. L'utilisation de composés présentant une cross-activité pourrait permettre d'identifier des similitudes et des différences entre ces deux enzymes. Enfin, il faudra toujours garder à l'esprit que ces inhibiteurs de l'intégrase du VIH-1 ne devront être utilisés en cellules humaines qu'après avoir été testés sur les protéines RAG et sur SETMAR, deux enzymes à DDE qui jouent des rôles importants dans la cellule. Il ne serait pas judicieux d'inhiber l'intégrase du VIH-1 en bloquant en même temps l'action vitale promue par les protéines RAG dans le cadre de la recombinaison V(D)J.

Parmi les nouveaux inhibiteurs mis à jour, le composé 4 présente une activité inhibitrice contre le transfert de brin bien plus importante que contre l'assemblage du complexe synaptique. La modulation des substitutions autour du noyau spécifique du composé 4 peut permettre la mise au point d'inhibiteurs spécifiques de la capture de cible et/ou du transfert de brin proprement dit, ce qui pourrait permettre d'enrichir nos connaissances sur cette ultime étape de la transposition peu documentée dans le monde des transposases.

MOS1 et mécanisme de transposition

L'intérêt de comprendre en détail le mécanisme de transposition de *Mos1* est double. D'une part, la cross-activité des inhibiteurs d'enzymes à DDE, à nouveau confirmée par mes travaux, permet d'utiliser MOS1 comme un modèle pour étudier la mécanique des intégrases rétrovirales et autres transposases à DDE. L'enzyme de premier intérêt est bien sur l'intégrase du VIH-1, avec qui la Tnp de *Mos1* partage plusieurs caractéristiques : dimère lié aux ITR ou LTR avant la formation du complexe synaptique impliquant un tétramère, coupure de l'ADN sans formation d'une *hairpin* comme c'est le cas chez les éléments bactériens Tn*5* et Tn*10*, chez l'élément eucaryote *Hermes* ou encore chez RAG. La caractérisation d'un mécanisme inconnu jusqu'alors chez les enzymes à DDE chez *Mos1* mériterait une étude sur cet aspect chez l'intégrase du VIH-1.

D’autre part, d’un point de vue très fondamental, la connaissance exacte du mécanisme de transposition pourrait apporter un nouvel éclairage sur le succès invasif des éléments transposables chez les eucaryotes, succès qui présente déjà des causes multiples (Pritham, 2009).

Lors de la transposition, le **mode de liaison de la transposase à ses ITR** conditionne la formation et la structure du complexe synaptique. Le débat est d’actualité parmi les spécialistes de la mécanique des transposons pour savoir si le domaine de liaison des transposases *mariner* est du même type que celui des *TLE*, c'est-à-dire de type PAIRED, ou de type CRO. L’organisation du domaine de liaison impacte logiquement sur la liaison à l’ADN ; un domaine de type PAIRED implique la liaison d’un monomère de transposase tandis qu’un domaine de type CRO entraîne la liaison d’un dimère. Plusieurs éléments allaient dans le sens d’un domaine de liaison de type CRO chez les éléments *mariner*, et nos travaux sur le domaine de dimérisation de MOS1 ont confirmé qu’il s’agissait bien d’un **domaine de type CRO**. Ce résultat renforce le modèle de transposition de *Mos1* proposé par le laboratoire, qui propose **la fixation d’un dimère de Tnp à l’ADN** (Augé-Gouillou *et al*, 2005a et b). D’autres données accréditent cette hypothèse :

- l’ITR possède deux sites de liaison organisés de façon palindromique, et la suppression d’un des deux sites se traduit par abolition de la transposition (Brillet, 2006)

- la Tnp se trouve sous une forme majoritairement dimérique en solution (Richardson *et al*, 2006) et le mutant T216A, qui n’est retrouvé que sous forme dimérique en solution, est une transposase hyperactive (Germon *et al*, 2009).

Dans un contexte cellulaire, cette **dimérisation a lieu dans le cytoplasme**. En effet, des travaux actuellement en cours au laboratoire montrent que la transposase est importée dans le noyau de la cellule sous forme dimérique. Ces données s’intègrent parfaitement dans le modèle global de transposition de *Mos1* que je propose, présenté plus loin dans la figure 99.

Ces travaux sur le domaine N-terminal montrent également qu’un seul HTH par transposase est requis pour la transposition de *Mos1*. Ce HTH assure la liaison spécifique de la transposase à l’ITR et le reste du processus de transposition. L’hypothèse la plus probable est que, lors de la dimérisation de MOS1, un HTH de chaque transposase du dimère est requis

pour former un domaine de liaison à l'ADN efficace. Ceci est similaire à ce qui est retrouvé dans la littérature pour les protéines CRO (Darling *et al*, 2000).

Les résultats obtenus sur le domaine de dimérisation vont également dans le sens de l'hypothèse largement répandue dans la communauté des spécialistes des transposons selon laquelle les *MLE* et les *TLE* résultent de l'association du même domaine catalytique à DDE avec des domaines de liaisons à l'ADN d'origine différente. Cela signifie qu'il pourrait y avoir des différences de liaison au sein même de la famille des *MLE*.

Mes travaux sur l'**assemblage du complexe synaptique** vont dans le sens de la littérature en ce qui concerne la fixation d'un dimère de Tnp aux ITR. Le complexe formé par deux Tnp et un ITR est appelé SEC2, et il s'agit du complexe observé majoritairement lors d'expériences de retard sur gel, tant avec la protéine procaryote qu'avec la protéine eucaryote.

Mes travaux remettent toutefois en cause une des voies de formation du complexe synaptique (PEC2) précédemment proposé (figure 26). Ce modèle proposait la formation d'un complexe appelé PEC1 comprenant deux Tnp et deux ITR (Augé-Gouillou *et al*, 2005b), consécutivement à la fixation du dimère de Tnp sur l'ITR. Ce complexe est en réalité un SEC2, puisqu'il n'est composé que d'un seul ITR et de deux Tnp. Il a été renommé SEC2 *top*, et il s'agit du seul complexe observé très majoritairement avec la protéine eucaryote, alors que deux conformations du SEC2 sont retrouvées avec la Tnp produite en système procaryote.

Le nouveau modèle de formation du complexe synaptique que je propose est présenté dans la figure 96. Un seul SEC2 étant observé avec la Tnp eucaryote, contre deux pour la Tnp procaryote, je propose que la Tnp procaryote se trouve à l'état monomérique sous une conformation différente de celle de la Tnp eucaryote. C'est une fois dimérisée et sur l'ITR qu'il y aurait un changement de conformation de la Tnp (passage du SEC2 *bottom* au SEC2 *top*), qui permet alors la formation du complexe synaptique (figure 96-B). Cela explique l'absence de différence de taux de transposition entre protéine procaryote et protéine eucaryote.

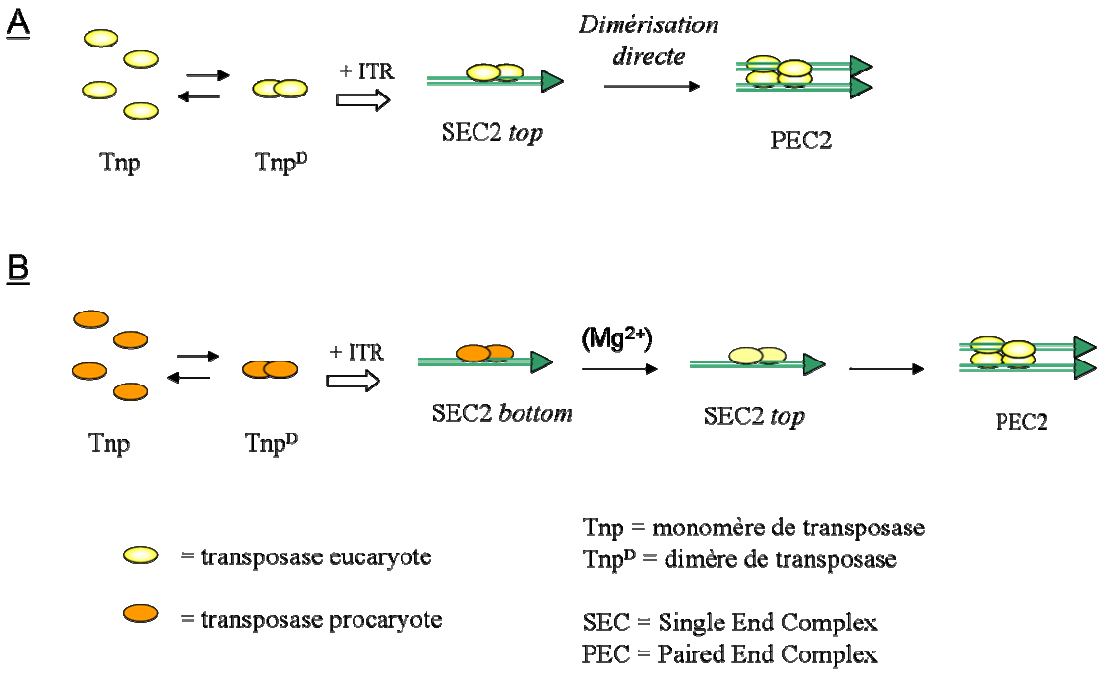


Figure 96. Nouveau modèle de formation du complexe synaptique de *Mos1*. (A) *Transposase eucaryote. La Tnp se dimérise, le dimère se fixe ensuite à l'ITR pour former le SEC2 top ; il y a ensuite dimérisation du SEC2 top pour donner le PEC2.* **(B)** *Transposase procaryote. La Tnp, qui a une conformation différente de la protéine procaryote, se dimérise puis se lie à l'ITR pour donner le SEC2 bottom. La Tnp change ensuite de conformation pour donner le complexe SEC2 top, qui suit ensuite la même voie qu'en* **(A)**. *Le Mg^{2+} contribue à accélérer ce changement de conformation (cf ci-dessous).*

Une nouvelle interprétation d'une expérience de retard sur gel présenté par Augé-Gouillou *et al* (2005b) laisse penser que le magnésium contribue au changement de conformation de la protéine procaryote une fois dimérisée et fixée sur l'ITR. En effet, la réaction de liaison à l'ITR menée en présence de Mg^{2+} amène à la disparition du complexe SEC2 *bottom* en 120 minutes, tandis qu'en présence d'EDTA le SEC2 *bottom* reste le complexe largement majoritaire à 120 minutes (figure 97).

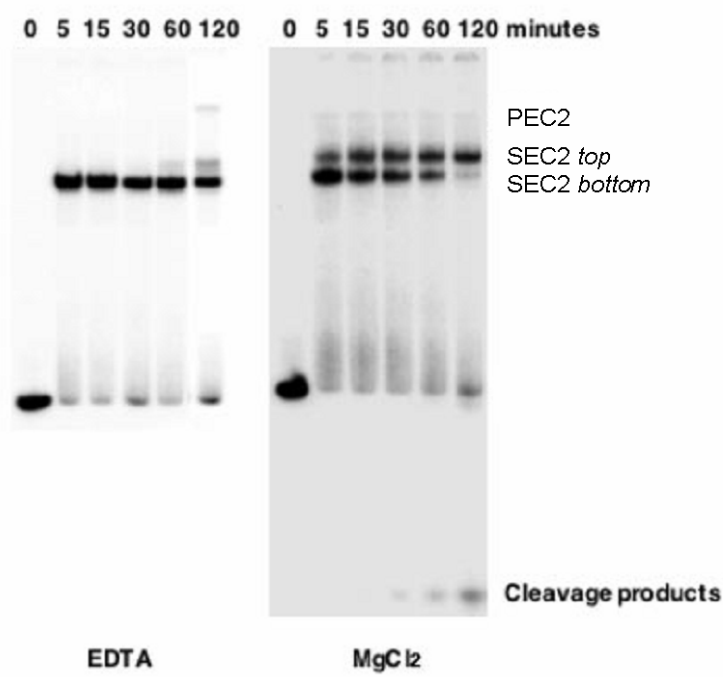


Figure 97. Cinétique de formation des complexes, adaptée d'Augé-Gouillou *et al*, 2005b. *L'apparition du SEC2 top ne se fait qu'après 1 heure de réaction en présence d'EDTA, et après 2 heures il est très minoritaire. Par contre, l'apport de Mg^{2+} favorise l'apparition rapide (dès 5 minutes) du SEC2 top, et provoque même la quasi-disparition du SEC2 bottom après 120 minutes de réaction.*

Les études en spectrométrie Raman de la transposase ont montré que la Tnp eucaryote et la Tnp procaryote avaient une conformation différente (Igor Chourpa, communication personnelle). Cela renforce le modèle de formation du PEC2 que je propose pour la protéine procaryote.

La cristallisation de la partie N-terminale de MOS1 permettra d'approfondir nos connaissances sur l'assemblage du complexe synaptique. Le problème reste néanmoins l'aptitude de la Tnp à se protéolyser par un mécanisme inconnu. Avec la Tnp eucaryote, nous disposons d'une protéine plus stable que sa version procaryote. Elle reste stable lorsqu'elle fusionnée à la MBP, mais se protéolyse en 24 heures lorsqu'elle est séparée de la MBP (contre 2 heures pour la Tnp procaryote).

Il reste à rentrer dans le nouveau schéma de formation du complexe synaptique le moment du clivage de l'ADN. Chez *Mos1*, le clivage qui permet par la suite la réintégration doit être précis : 3 bases à l'intérieur de l'ITR sur le brin non-transféré (NTS) et à l'extrémité exacte de l'ITR sur le brin transféré (TS) (figure 27). Ce clivage double-brin a lieu dans un ordre précis, le NTS étant clivé avant le TS. Il apparaît que le *SEC2 top* peut promouvoir la coupure du premier brin, mais très difficilement celle du second brin (Augé-Gouillou *et al*, 2005b). Par contre, le PEC2 est capable d'effectuer efficacement la coupure du brin transféré. Ces éléments permettent d'envisager que c'est le SEC2 *top* qui coupe le premier brin. Ceci permet alors à l'ADN de se courber plus facilement pour former le complexe synaptique

PEC2 par dimérisation des deux complexes SEC2 *top*. Il est possible que la dimérisation des SEC2 *top* ait lieu après un changement de conformation du complexe SEC2 *top* suite au premier clivage. Le PEC2 ainsi formé peut alors promouvoir le clivage du second brin et former le transpososome (figure 98). Le magnésium intervient pour permettre le clivage des deux brins et il a un certainement aussi un rôle dans la stabilisation du complexe SEC2 *top*. Cette dernière hypothèse est renforcée par le comportement de la protéine procaryote quand les complexes sont visualisés en retard sur gel, ainsi que par la formation par certains lots de protéine eucaryote du complexe SEC2 *bottom* (mais seulement de manière ultra-minoritaire en présence de magnésium et de manière minoritaire en présence d'EDTA).

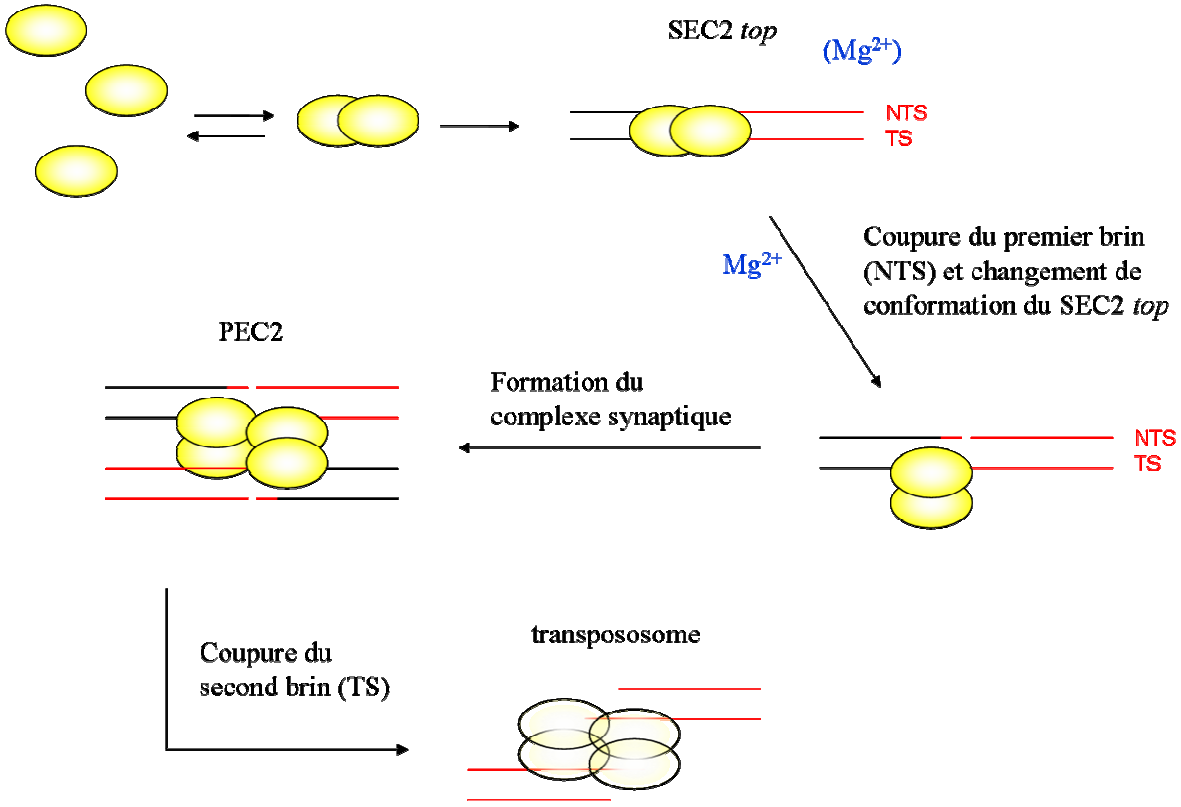


Figure 98. Modèle de la formation du transpososome chez *Mos1*. *(Les ITR sont représentés en rouge, l'ADN flanquant en noir, la transposase est représenté par les billes jaunes ; NTS : brin non-transféré ; TS : brin transféré) La Tnp sous forme dimérique se fixe sur un ITR (SEC2 top), ce complexe pouvant être stabilisé par du magnésium. S'il ne l'avait pas fait, le complexe coordonne le magnésium, ce qui permet la coupure du premier brin (NTS). S'ensuit alors un changement de conformation du complexe. Ces deux éléments permettent la formation du PEC2, et le tétramère de transposases peut promouvoir la coupure du second brin (TS) pour former le transpososome. Les transposases sont représentés transparentes pour permettre de visualiser les extrémités des ITR : la coupure se fait 3 bases à l'intérieur sur le brin NTS et à l'extrémité de l'ITR sur le brin TS.*

Il est à noter que le clivage promu par un dimère de protéines est retrouvé chez l'intégrase du VIH-1. L'étape du *3'-processing* est effectuée par un dimère d'intégrases avant que ne se forme le tétramère qui effectue la réaction de transfert de brin (Faure *et al*, 2005 ; Guiot *et al*, 2006).

Le clivage double-brin de l'ADN par la transposase est l'étape clé de la transposition. La coupure doit être précise (figure 28) pour permettre la réintégration, et l'hypothèse la plus probable est qu'une coupure précise sur le premier brin induit systématiquement une coupure précise sur le second brin, résultant d'un positionnement correct du complexe protéique sur l'ITR. Le pourcentage de coupures permettant la réintégration est très faible (Julien Bischerour, communication personnelle) ce qui fait du clivage l'étape limitante de la transposition. Il reste à déterminer si la coupure du second brin par le PEC2 se fait en *trans*, c'est à dire qu'une transposase fixée sur un ITR coupe l'autre ITR et inversement, comme il a été montré pour le phage *Mu* et la transposase de Tn*5*.

Une fois la coupure double-brin correcte effectuée, le transpososome va capturer sa cible et le transfert de brin peut se faire. Ces dernières étapes sont assez peu étudiées et notre compréhension du mécanisme exact de capture de cible est donc assez limitée. L'idée couramment admise est que le transpososome est un intermédiaire de réaction qui va capturer sa cible. Toutefois, aucun argument ne contredit la possibilité que le complexe synaptique capture sa cible avant de promouvoir le clivage du second brin. Nous aurions ainsi une réaction concertée entre clivage du second brin et capture de cible puis transfert de brin. Ceci n'est toutefois qu'une hypothèse de travail non étayée pour le moment.

Au cours de ma thèse, j'ai observé un complexe supposé être un SEC2 associé à une cible, complexe se formant préférentiellement en présence d'ITR préalablement clivés. Trois hypothèses existent quant à l'identité de ce complexe :

- produit de dissociation du 'vrai' complexe de capture de cible à savoir le transpososome (qui correspond au PEC2 avec deux ITR clivés sur leurs deux brins) ayant capturé une cible

- résultat d'un SEC2 ayant promu un clivage double-brin puis capturé un ADN cible. Il ne serait alors qu'un artefact de réaction, incapable de promouvoir de la transposition

- résultat d'un SEC2 ayant lié un ITR de manière non-spécifique *via* sa partie C-terminale

- intermédiaire de la réaction de transposition : le SEC2 permet la coupure double-brin et capture une cible, ceci permettant la dimérisation pour former le complexe à quatre transposases et deux ITR, avec la cible déjà capturée. Toute la réaction serait alors concertée.

Il ne s'agit certainement pas d'un produit de dissociation du transpososome ayant capturé une cible. En effet, dans l'expérience présentée dans la figure 97 et tirée d'Augé-Gouillou *et al* (2005b), on peut voir une accumulation de produits de clivage en présence magnésium, sans accumulation du PEC2. Or les auteurs ont montré dans la même publication que le PEC2 était à l'origine de la majorité des coupures du second brin. Les conditions sont réunies pour avoir un transpososome qui a capturé une cible (ici un autre ITR) après avoir promu le clivage du second brin. Mais on ne voit pas de nouveau complexe apparaître avec l'augmentation des produits de clivage. Je n'ai d'ailleurs observé ce complexe qu'avec des ITR préalablement clivés sur le premier brin ou sur les deux brins. On peut dès lors supposer que le complexe présenté comme un STC n'était qu'un artefact dû à la fixation dans une conformation particulière du dimère de Tnp sur un ITR particulier.

Régulation de MOS1

L'absence de transposition de *Mos1* en cellules de mammifères renforce l'idée selon laquelle les **modifications post-traductionnelles** ont un impact sur la transposition de *Mos1*. En effet, *Mos1* est capable de transposer chez les procaryotes, chez les invertébrés et chez les vertébrés à sang froid, mais pas chez les mammifères. Les modifications post-traductionnelles et les enzymes qui les provoquent peuvent différer d'un organisme à l'autre.

Les études menées sur le rôle de la **phosphorylation** de MOS1 n'ont pas été totalement concluantes dans la mesure où nos échantillons n'étaient pas adéquats pour étudier le réel impact de la phosphorylation sur la transposase. Nous savons toutefois que MOS1 est phosphorylée et un pattern de phosphorylation partiel a été obtenu, indiquant notamment que des positions clés sont phosphorylées (la sérine 99 située dans le HTH de liaison à l'ADN) ou potentiellement phosphorylés (positions situées dans le domaine de dimérisation ou dans le

HTH). Ce pattern partiel nous donne donc déjà de précieuses informations, et permet de relier la phosphorylation à la transposition de *Mos1*. La sérine en position 28, qui semble sensible à la phosphorylation d'après les études de mutagénèse, est phosphorylable par la PKC, une enzyme retrouvée dans le cytoplasme chez les mammifères. La dimérisation ayant lieu dans le cytoplasme avant l'internalisation nucléaire, la cellule pourrait contrôler la dimérisation par phosphorylation. Au niveau du HTH, il est intéressant de noter la différence entre HIMAR1 et MOS1. Les positions T88 et S104 de MOS1 correspondent à des acides aminés non-phosphorylables chez HIMAR1, tandis que la sérine 99, qui est phosphorylée chez MOS1, correspond à la sérine 98 de HIMAR1, qui est prédite non-phosphorylable. La différence entre les deux protéines est que HIMAR1 peut promouvoir de la transposition en cellules de mammifères tandis que MOS1 non. Les positions 88 et 99 de MOS1 sont phosphorylables par des ATM kinases, des enzymes mobilisées en cas de cassure double-brin de l'ADN, ce qui est le cas pendant la transposition. On peut dès lors envisager une régulation négative de la transposition de *Mos1* par phosphorylation des transposases non liées après l'excision d'un premier élément. La phosphorylation des Tnp pourrait aussi empêcher le transpososome de promouvoir la réintégration.

La spectrométrie Raman a montré la présence de ponts disulfures chez MOS1. Il reste à caractériser l'existence du ou des ponts intermoléculaires, qui pourrait avoir un rôle dans la conformation de la protéine. Des éléments tendent également à montrer que des ponts peuvent se former entre deux molécules de Tnp. Là encore la cristallographie de la protéine pleine taille nous permettra d'en apprendre d'avantage.

Toutes ces données et hypothèses permettent de proposer un **modèle général de la transposition de *Mos1* dans un contexte cellulaire** (figure 99).

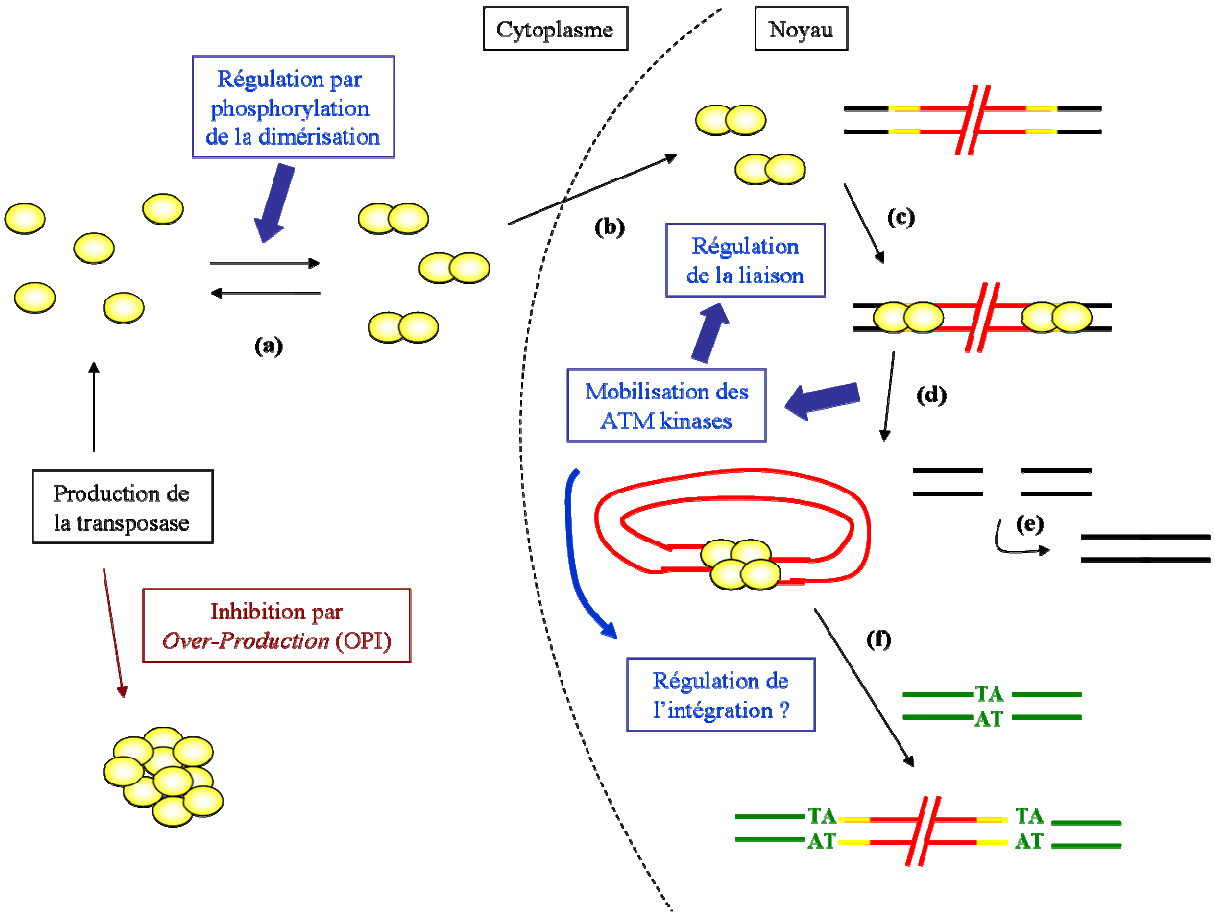


Figure 99. Modèle de transposition de *Mos1* dans un contexte cellulaire eucaryote. *(La transposase est représentée par des billes jaunes, le transposon en rouge, ses ITR en jaune ; l'ADN donneur est en noir, l'ADN receveur en vert) La transposase produite dans le noyau se dimérise* **(a)**, *une action régulée par phosphorylation. La régulation peut aussi se faire sous forme d'inhibition par surproduction (appelée OPI) ; la transposase surproduite s'oligomérise, devant incapable de promouvoir la transposition. C'est sous forme dimérique qu'elle importée dans le noyau* **(b)**, *puis le lie aux ITR du transposon* **(c)**. *Le dimère de Tnp clive le premier brin, ce qui entraîne la dimérisation du dimère pour former un tétramère de Tnp, capable de procéder au clivage du second brin, ce qui libère le transpososome* **(d)**. *L'ADN donneur est réparé par des enzymes cellulaires* **(e)**. . *Le transpososome reconnaît l'ADN cible et s'insert au niveau d'un dinucléotide TA* **(f)** *qui est dupliqué après l'intégration. La coupure double brin de l'ADN provoque la mobilisation des ATM kinases qui phosphorylent les Tnp non liées pour empêcher d'autres réactions de transposition. La phosphorylation par les ATM pourrait également réguler l'intégration.*

La recherche de modifications post-traductionnelles avait pour objectif le développement de transposases hyperactives et non régulées négativement dans les cellules pour y faire du transfert de gènes. D'un côté les travaux publiés dans Germon *et al* (2009) posent la question de la pertinence du développement de transposases hyperactives. En effet, les transposases les plus hyperactives sont également les plus toxiques pour les cellules où elles sont produites. L'utilisation de transposases hyperactives (et donc cytotoxiques) *in vivo* posent plusieurs problèmes. La cytotoxicité n'est pas prise en compte dans les études impliquant des Tnp hyperactives (Keravala *et al*, 2006), et l'utilisation de transposases hyperactives n'implique pas une augmentation de l'expression des transgènes. En effet, il a récemment été montré que des vecteurs basés sur le *TLE Sleeping Beauty* étaient sujets à l'extinction de l'expression des gènes (*gene silencing*) à plus de 94% après intégration (Garrison *et al*, 2007).

Tous ces travaux apportent un nouvel éclairage sur l'**absence d'activité des vecteurs *Mos1* en cellules de mammifères**. Le *silencing* observé avec *Sleeping Beauty* peut être possible avec *Mos1*, ce point n'a pour le moment jamais été recherché avec *Mos1*. Même si les résultats actuels ne nous permettent pas de conclure quant à l'impact des modifications post-traductionnelles sur MOS1, il est fortement envisageable que de telles modifications aient une action régulatrice sur la transposition. MOS1 pourrait également être sensible à des facteurs d'hôtes, et ceci pourrait expliquer les résultats obtenus chez *Caenorhabditis elegans*, avec la mise au point d'un outil très efficace (*Mos*TIC) basé sur *Mos1* pour manipuler le génome du nématode (Robert & Bessereau, 2007).

Contrairement à d'autres transposons (*Sleeping Beauty*, *piggyBac*) avec lesquels des résultats prometteurs ont été obtenus, il semble que *Mos1* ne soit pas la solution pour créer un vecteur non-viral de transfert de gènes chez les mammifères.

Perspectives

Mes travaux ont complété nos connaissances sur le processus de transposition de l'élément *mariner Mos1*. Plusieurs questions restent toutefois en suspens et elles doivent être résolues pour compléter le modèle de transposition de *Mos1*. Ces questions portent sur la formation du PEC2 et sur la capture de cible. Une fois le SEC2 *top* formé, comment est exactement assemblé le PEC2 ? Et qu'elle s'effectue par le PEC2 avant le clivage du second brin ou par le transpososome, la capture de cible reste non élucidée : comment la cible est-elle reconnue et capturée par le complexe protéique ?

De solides bases sur l'étude de la régulation de *Mos1* ont été posées dans ce mémoire. Il convient maintenant de les exploiter pour savoir si les modifications post-traductionnelles ont réellement un impact sur la transposition de *Mos1*. L'étude des ponts disulfures présente notamment un intérêt particulier dans l'étude de l'assemblage et du maintien du complexe synaptique.

A mon arrivée dans le laboratoire, *Mos1* se présentait comme un candidat prometteur pour en faire un outil biotechnologique de transfert de gènes, y compris en cellules de mammifères et à plus long terme de la thérapie génique *ex vivo*. L'ensemble des études menées depuis, notamment sur le mécanisme de transposition et la régulation de l'élément a montré que *Mos1* était loin d'être optimal pour faire du transfert de gènes en cellules de mammifères.

L'utilisation d'un transposon comme outil de transfert de gènes implique donc une connaissance approfondie de son mécanisme de transposition et de sa régulation *in vivo*. La description de plusieurs enzymes à triade catalytique DDE dans l'introduction de ce mémoire a montré la diversité des mécanismes de transposition, et tout nouveau transposon candidat pour devenir un vecteur de transfert de gènes devra voir son mécanisme de transposition élucidé.

S'il ne peut pas être utilisé en cellules de mammifères, *Mos1* fonctionne par contre chez les invertébrés, et on peut imaginer de l'utiliser à terme comme un vecteur antihelminthique, ces vers posant de nombreux problèmes de santé publique en particulier dans les pays en voie de développement.

Références bibliographiques

Abu-Qarn M, Eichler J, Sharon N, Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea, *Current Opinion in Structural Biology*, 2008, 18 (5), 544-550

Agrawal A & Schatz D, RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination, *Cell*, 1997, 89 (1), 43-53

Agrawal A, Eastman Q, Schatz D, Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system, *Nature*, 1998, 394 (6695), 744-51

Aiello F, Brizzi A, Garofalo A, Grande F, Ragno G, Dayam R, Neatami N, Synthesis of novel thiazolothiazepine based HIV-1 integrase inhibitors, 2004, *Bioorganic & Medicinal Chemistry*, 12 (16), 4459-66

Albright RA, Mossing MC, Matthews BW, Crystal structure of an engineered Cro monomer bound nonspecifically to DNA: possible implications for nonspecific binding by the wild-type protein, *Protein Science*, 2005, 7 (7), 1485-94

Alleman M & Freeling M, The Mu transposable elements of maize: evidence for transposition and copy number regulation during development, *Genetics*, 1986, 112, 107-19

Arkhipova IR & Meselson M, Diverse DNA transposons in rotifers of the class Bdelloidea, *PNAS*, 2005, 102 (33), 11781-11786

Ason B & Reznikoff WS, Mutational analysis of the base flipping event found in Tn*5* transposition, 2002, Journal of Biological Chemistry, 277, 11284–91

Ason B, Knauss DJ, Balke AM, Merkel G, Shalka AM, Reznikoff WS, Targeting Tn*5* Transposase Identifies Human Immunodeficiency Virus Type 1 Inhibitors, 2005 *Antimicrobial Agents Chemotherapy*, 49 (5), 2035-43

Augé-Gouillou C, Bigot Y, Pollet N, Hamelin M-H, Meunier M, Periquet G, Human and other mammalian genomes contain transposons of the *mariner* family, *FEBS Letters*, 1995, 368 (3), 541-546

Augé-Gouillou C, Hamelin MH, Demattei MV, Periquet G, Bigot Y, The ITR binding domain of the *Mariner Mos-1* transposase, *Molecular Genetics and Genomics*, 2001a, 265 (1), 58-65

Augé-Gouillou C, Hamelin MH, Demattei MV, Periquet M, Bigot Y, The wild-type conformation of the *Mos-1* inverted terminal repeats is suboptimal for transposition in bacteria, *Molecular Genetics and Genomics*, 2001b, 265 (1), 51-57

Augé-Gouillou C, Brillet B, Germon S, Hamelin MH, Bigot Y, *Mariner Mos1* transposase dimerizes prior to ITR binding, *Journal of Molecular Biology*, 2005a, 351 (1), 117-30.

Augé-Gouillou C, Brillet B, Hamelin MH, Bigot Y, Assembly of the *mariner Mos1* synaptic complex, *Molecular and Cellular Biology*, 2005b, 25 (7), 2861-70

Bailey JA, Carrel L, Chakravarti A, Eichler E, Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation : the Lyon repeat hypothesis, *PNAS*, 2000, 97 (12), 66348-6639

Balciunas D, Wangensteen KJ, Wilber A, *et al*, Harnessing a high cargo-capacity transposon for genetic applications in vertebrates, *PLOS Genetics*, 2006, 2 (11), e169

Bannert N & Kurth R, The evolutionary dynamics of human endogenous retroviral families, *Annual review of genomics and human genetics*, 2006, 7, 149-173

Barreca ML, Ortuso F, Iraci N, De Luca L, Alcaro S, Chimirri A, Tn*5* transposase as a useful platform to simulate HIV-1 integrase inhibitor binding mode, *Biochemical and biophysical research communications*, 2007, 363 (3), 554-60

Barry EG, Witherspoon DJ, Lampe DJ, A bacterial genetic screen identifies functional coding sequences of the insect *mariner* transposable element *Famar1* amplified from the genome of the earwig, *Forficula auricularia*, *Genetics*, 2004, 166 (2), 823-33

Beall EL, Mahoney MB, Rio DC, Identification and analysis of a hyperactive mutant form of *Drosophila P*-element transposase, *Genetics*, 2002, 162 (1), 217-27

Beck P, Park S, Lee Y, Roman Y, Hromas R, Lee S, Human Pso4 is a Metnase (SETMAR)-binding partner that regulates Metnase function in DNA repair, *The Journal of Biological Chemistry*, 2008, 283 (14), 9023-30

Belancio VP, Hedges DJ, Deininger P, Mammalian non-LTR rétrotransposons: for better or worse, in sickness and in health, *Genome Research*, 2008, 18 (3), 343-58

Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O, Devine SE, Active Alu rétrotransposons in the human genome, *Genome Research*, 2008, 18 (12), 1875-83

Berg DE, Davies J, Allet B, Rochaix J-D, Transposition of R factor genes to bacteriophage λ, 1975, *PNAS*, 72, 3628–32

Bhasin A, Goryshin IY, Reznikoff WS, Hairpin formation in Tn*5* transposition, 1999, *Journal of Biological Chemistry*, 274, 37021–29

Biémont C & Vieira C, Junk DNA as an evolutionary force, *PNAS*, 2006, 443 (5), 521-24

Bigot Y, Hamelin M-H, Capy P, Periquet G, *Mariner*-like elements in hymenopteran species: insertion site and distribution, *PNAS*, 1994, 368 (3), 541-546

Bigot Y, Brillet B, Auge-Gouillou C, Conservation of Palindromic and Mirror Motifs within Inverted Terminal Repeats of *mariner*-like Elements, *Journal of Molecular Biology*, 2005, 351 (1), 108-116

Bingham PM, Kidwell MG, Rubin GM, The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family, *Cell*, 1982, 29 (3), 995-1004

Bird A, DNA methylation patterns and epigenetic memory, *Genes & Development*, 2002, 16, 6-21

Bischerour J & Chalmers R, Base-flipping dynamics in a DNA hairpin processing reaction, *Nucleic Acids Research*, 2007, 35, 2584–95

Bischerour J & Chalmers R, Base-flipping in Tn10 transposition: an active flip and capture mchanism, *PLOS One*, 2009, 4 (7), e6201

Boeke JD & Devine SE, Yeast rétrotransposons: finding a nice quiet neighborhood, *Cell*, 1998, 93 (7), 1087-9

Bouchet N, Bischerour J, Germon S, Guillard J, Dubernet M, Viaud-Massuard M-C, Delelis O, Ryabinin V, Bigot Y, Augé-Gouillou C, First *mariner Mos1* transposase inhbitors, *Mini reviews in medicinal chemistry*, 2009, 9 (4), 431-39

Braam L & Reznikoff WS, DNA transposition: classes and mechanisms, *Encyclopedia of Life Science*, 2001

Brillet B, *Activité de la transposase de l'élément mariner Mos1: mise en place du complexe synaptique et regulation par phosphorylation*, thèse de doctorat, Université de Tours, 28/09/2006

Brillet B, Bigot Y, Augé-Gouillou C, Assembly of the *Tc1* and *mariner* transposition initiation complexes depends on the origins of their transposase DNA binding domains, *Genetica*, 2007, 130 (2), 105-20

Bryan G, Garza D, Hartl DL, Insertion and excision of the transposable element *mariner* in *Drosophila*, *Genetics*, 1990, 125 (1), 103-14

Bui QT, Casse N, Leignel V, Nicolas V, Chenais B, Widespread occurence of *mariner* transposons in coastal crabs, *Molecular Phylogenetics and Evolution*, 2008, 47 (3), 1181-89

Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M, Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection, *PNAS*, 1993, 90, 6125-9

Busschots K, Voet A, De Maeyer M *et al*, Identification of the LEDGF/p75 binding site in HIV-1 integrase, *Journal of Molecular Biology*, 2007, 365 (5), 1480-92

Bustin M, Regulation of DNA-dependent activities by the functional motifs of the highmobility-group chromosomal proteins, *Molecular and Cellular Biology*, 1999, 19 (8), 5237-46

Butler MG, Chakraborty SA, Lampe DJ, The N-terminus of *Himar1 mariner* transposase mediates multiple activities during transposition, *Genetica*, 2006, 127 (1), 351-66

Bundock P & Hooykaas P, An *Arabidopsis hAT*-like transposase is essential for plant development, *Nature*, 2005, 436 (7048), 282-84

Calvi BR, Hong TJ, Findley SD, Gelbart WM, Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: *hobo*, *Activator*, and *Tam3*, *Cell*, 1991, 66 (3), 465-71

Capy P, Chakrani F, Lemeunier F, Hartl DL, David JR, Active *mariner* transposable elements are widespread in natural populations of *Drosophila simulans*, *Proceedings. Biological Sciences / The Royal Society*, 1990, 242 (1303), 57-60

Carpenter R, Martin C, Coen E, Comparison of genetic behaviour of the transposable element Tam3 at two unlinked pigment loci in *Antirrhinum majus*, 1987, *Molecular and General Genetics* , 207, 82-89

Carpentier G, Adet J, Bouchet N, Demattei M, Renault S, Bigot Y, Augé-Gouillou C, *Mariner* mobility requires a single helix-turn-helix in the N-terminal transposase domain, soumis dans *EMBO Reports*

Casola C, Lawing M, Betran E, Feschotte C, PIF-like transposons are common in Drosophila and have been repeatedly domesticated to generate new host genes, *Molecular Biology and Evolution*, 2007, 24 (8), 1872-88

Casteret S, Chbab N, Cambefort J, Augé-Gouillou C, Bigot Y, Bonnin F, Physical properties of DNA components affecting the transposition efficiency of the *mariner Mos1* element, *Molecular Genetics and Genomics*, 2009, en cours de publication

Ceccherini-Silberstein F, Malet I, D'Arrigo R, Antinori A, Marcelin A, Perno C, Characterization and structural analysis of HIV-1 integrase conservation, *AIDS Reviews*, 2009, 11, 17-29

Chakonas G & Harshey RM, Transposition of phage Mu DNA, In Craig N *et al*, *Mobile DNA II*, ASM Press, 2002, 384-402

Chalmers R, Guhathakurta A, Benjamin H, Kleckner N, IHF modulation of Tn*10* transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring, *Cell*, 1998, 93, 897-908

Chandler V, Clerget M, Gallas DJ, The transposition frequency of IS1-flanked transposons is a function of their size, *Journal of Molecular Biology*, 1982, 154 (2), 229-43

Chandler VL & Walbot V, DNA modification of a maize transposable element correlates with loss of activity, *PNAS*, 1986, 83, 1767-71

Chen S & Li X, Molecular characterization of the first intact *Transib* transposon from *Helicoverpa zea*, *Gene*, 2008, 408, 51-63

Chomet PS, Wessler S, Dellaporta SL, Inactivation of the maize element Activator (Ac) is associated with its DNA modification, *EMBO Journal*, 1987, 6 (2), 295-302

Chow SA, Vincent KA, Ellison V, Brown PO, Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus, *Science*, 1992, 255, 723–6

Ciubotaru M, Ptaszek LM, Baker GA, Baker SN, Bright FV, Schatz DG, RAG1-DNA binding in V(D)J recombination. Specificity and DNA-induced conformational changes revealed by fluorescence and CD spectroscopy, *Journal of Biological Chemistry*, 2003, 278 (8), 5584-96

Clark K, Carlson D, Leaver M, Foster L, Fahrenkrug S, *Passport*, a native Tc*1* transposon from flatfish, is functionally active in vertebrate cells, *Nucleic Acids Research*, 2009, 37 (4), 1239-1247

Claudianos C, Brownlie J, Russell R, Oakeshott J, Whyard S, *maT*--a clade of transposons intermediate between *mariner* and *Tc1*, *Molecular Biology and Evolution*, 2002, 19 (12), 2101-9

Coen ES, Carpenter R, Martin C, Transposable elements generate novel spatial patterns of gene expression in *Antirrhinum majus*, *Cell*, 1986, 47 (2), 285-96

Collins J, Forbes E, Anderson P, The Tc3 family of transposable genetic elements in *Caenorhabditis elegans*, *Genetics*, 1989, 121 (1), 47-55

Colloms SD, van Luenen HG, Plasterk RH, DNA binding activities of the *Caenorhabditis elegans* Tc*3* transposase, *Nucleic Acids Research*, 1994, 22 (25), 5548-54

Cordeaux R, Udit S, Batzer M, Fescotte C, Birth of a chimeric primate gene by capture of the transposase gene from a mobile element, *PNAS*, 2006, 103 (21), 8101-6

Cordeaux R, The human genome in the LINE of fire, *PNAS*, 2008, 105, 19366-71

Coy MR & Tu Z, Gambol and Tc1 are two distinct families of DD34E transposons: analysis of the Anopheles gambiae genome expands the diversity of the IS630-Tc1-*mariner* superfamily, *Insect Molecular Biology*, 2005 14 (5), 537-46

Craig N, Tn*7*, In Craig N *et al*, *Mobile DNA II*, ASM Press, 2002, 423-56

Craigie R, Mozuuchi K, Bushman F, Engelman A, A rapid *in vitro* assay for HIV DNA integration, *Nucleic Acid Research*, 1991, 19, 2729-34

Crénès G, Ivo D, Herisson J, Dion S, Renault S, Bigot Y, Petit A, The bacterial Tn*9* chloramphenicol resistance gene: an attractive DNA segment for *Mos1 mariner* insertions, *Molecular Genetics and Genomics*, 2009, 281 (3), 315-28

Cui H, Fedoroff N, Inductible DNA, demethylation mediated by the maize Suppressor-mutator transposon-encoded TnpA protein, *Plant Cell*, 2002, 14 (11), 2883-99

Czyz A, Stillmock K, Hazuda D, Reznikoff WS, Dissecting Tn*5* transposition using HIV-1 integrase diketoacid inhibitors, 2007, *Biochemistry*, 46, 10776–89

Dai J, Xie W, Brady T, Gao J, Voytas D, Phosphorylation regulates integration of the yeast Ty5 retrotransposon into heterochromatin, *Molecular Cell*, 2007, 27 (2), 289-99

Darling PJ, Holt JM, Ackers GK, Coupled energetics of lambda cro repressor self-assembly and site-specific DNA operator binding I: analysis of cro dimerization from nanomolar to micromolar concentrations, *Biochemistry*, 2000a, 39 (37), 11500-7

Darling PJ, Holt JM, Ackers GK, Coupled energetics of lambda cro repressor self-assembly and site-specific DNA operator binding II: cooperative interactions of cro dimers, *Journal of Molecular Biology*, 2000b, 303 (3), 625-38

Davies DR, Goryshin IY, Reznikoff WS, Rayment I, Three-dimensional structure of the Tn*5* synaptic complex transposition intermediate, 2000, *Science*, 289, 77–85

Dawson & Finnegan D, Excision of the *Drosophila mariner* transposon *Mos1*: comparison with bacterial transposition and V(D)J recombination, *Molecular Cell*, 2003, 11, 225-35

de la Cruz NB, Weinreich MD, Wiegand TW, Krebs MP, Reznikoff WS, Excision Characterization of the Tn*5* transposase and inhibitor proteins: a model for the inhibition of transposition, *Journal of bacteriology*, 1993, 175 (21), 6932-38

Delelis O, Carayon K, Saïb A, Deprez E, Mouscadet J-F, Integrase and integration: biochemical activities of HIV-1 integrase, *Retrovirology*, 2008, 5 (114), 1-13

de Soultrait VR, Lozach PY, Altmeyer R, Tarrago-Litvak L., Litvak S, Andréola ML, DNA aptamers derived from HIV-1 RNase H inhibitors and strong anti-integrase agents, *Journal of Molecular Biology*, 2002, 324 (2), 195-203

DeMarco R, Venancio T, Verjovski-Almeida S, *SmTRC1*, a novel *Schistosoma mansoni* DNA transposon, discloses new families of animal and fungi transposons belonging to the CACTA superfamily, 2006, *BMC Evolutionary Biology*, 6 (89)

Derval P & Edelson B, Recognition of the DNA minor groove by pyrrole-imidazaole polyamides, 2003, *Current Opinion in Structural Biology*, 13, 284-99

Doak TG, Doerder FP, Jahn CL, Herrick G, A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif, *PNAS*, 1994, 91 (3), 942-46

Dooner H & Weil C, Give-and-take: interactions between DNA transposons and their host plant genomes, *Current Opinion in Genetics and Development*, 2007, 17, 486-92

Dyda F, Hickman A, Jenkins T, Engelman A, Craigie R, Davies DR, Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases, *Science*, 1994, 266 (5193), 1981-6

Eisen JA, Benito MI, Walbot V, Sequence similarity of putative transposases links the maize *Mutator* autonomous element and a group of bacterial insertion sequences, *Nucleic Acids Research*, 1994, 22, 2634–36

Elick TA, Bauser CA, Fraser MJ, Excision of the piggyBac transposable element in vitro is a precise event that is enhanced by the expression of its encoded transposase, *Genetica*, 1996, 98 (1), 33-41

Emmons SW, Klass MR, Hirsh D, Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*, *PNAS*, 1979, 76 (3), 1333-37

Emmons SW, Yesner L, Ruan KS, Katzenberg D, Evidence for a transposon in *Caenorhabditis elegans*, *Cell*, 1983, 32 (1), 55-65

Evgen'ev MB, Zelentsova H, Shostak N, Kozitsina M, Barskyi V, Lankenau DH, Corces VG, Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in Drosophila virilis, *PNAS*, 1997, 94 (1), 196-201

Evgen'ev MB & Arkhipova IR, Penelope-like elements-a new class of retroelements: distribution, function and possible evolutionary significance, *Cytogenetic and Genome Research*, 2005, 110 (1-4), 510-21

Farkash EA & Prak ET, DNA damage and L1 retrotransposition, *Journal of biomedicine and biotechnology*, 2006, 2006 (1), 37285

Faure A, Calmels C, Desjobert C, Castroviejo M, Caumont-Sarcos A, Tarrago-Litvak L, Litvak S, Parissi V, HIV-1 integrase crosslinked oligomers are active in vitro, *Nucleic Acids Research*, 2005, 33, 977-86

Feldmann T, Alex R, Suckow J, Dildrop R, Kisters-Woike B, Muller-Hill B, Single exchanges of amino acids in the basic region change the specificity of N-Myc, *Nucleic Acids Research*, 1993, **21** (22)**,** 5050-58

Feldmar S & Kunze R, The ORFa protein, the putative transposase of maize transposable element *Ac*, has a basic DNA binding domain, *EMBO Journal*, 1991, 10 (13), 4003-10

Feng X & Colloms S, *In vitro* transposition of *ISY100*, a bacterial insertion sequence belonging to the *Tc1/mariner* family, *Molecular Microbiology*, 2007, 65 (6), 1432-43

Feschotte C, *Merlin*, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS*1016* Insertion Sequences, *Molecular Biology and Evolution*, 2004, 21 (9), 1769-80

Feschotte C & Pritham E, Non-mammalian c-integrases are encoded by giant transposale elements, *Trends in Genetics*, 2005, 21 (10), 551-52

Feschotte C, The *piggyBac* transposon holds promise for human gene therapy, *PNAS*, 2006, 103 (41), 15008-13

Feschotte C, Transposable elements and the evolution of regulatory networks, *Nature Review Genetics*, 2008, 9 (5), 397-405

Finnegan D, Eukaryotic transposable elements and genome evolution, *Trends in Genetics*, 1989, 5 (4), 103-107

Fischer SE, van Luenen HG, Plasterk RH, Cis requirements for transposition of *Tc1*-like transposons in *C. elegans*, *Molecular and General Genetics*, 1999, 262 (2), 268-74

Foster TJ, Davis MA, Roberts, DE, Takeshita K, Kleckner N, Genetic organization of transposon Tn*10*, 1981, *Cell*, 23, 201–13

Fraser MJ, Brusca JS, Smith GE, Summers MD, Transposon-mediated mutagenesis of a baculovirus, *Virology*, 1985, 145 (2), 356-61

Fraser MJ, Ciszczon T, Elick T, Bauser C, Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (*IFP2*) and *tagalong* (*TFP3*) from the baculovirus genome in cell lines from two species of Lepidoptera, *Insect Molecular Biology*, 1996, 5, 141–151

Franz G, Loukeris TG, Dialektaki G, Thompson CR, Savakis C, Mobile *Minos* elements from Drosophila hydei encode a two-exon transposase with similarity to the paired DNA-binding domain, *PNAS*, 1994, 91 (11), 4746-4750

Frey M, Reinecke J, Grant S, Saedler H, Gierl A, Excision of the *En/Spm* transposable element of *Zea mays* requires two element-encoded proteins, *EMBO Journal*, 1990, 9 (12), 4037-44

Garrison B, Yant S, Mikkelsen JG, Kay MA, Postintegrative gene silencing within the *Sleeping Beauty* transposition system, *Molecular and Cellular Biology*, 2007, 27, 8824-33

Germon S, Bouchet N, Casteret S, Carpentier G, Adet J, Bigot Y, Augé-Gouillou C, *Mariner Mos1* optimization by rational mutagenesis, *Genetica*, 2009, publié en ligne

Geurts AM, Yang Y, Clark KJ, *et al*, Gene transfer into genomes of human cells by the sleeping beauty transposon system, *Molecular Therapy*, 2003, 8 (1), 108-17

Goodwin TJ, Butler MI, Poulter RT, Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi, *Microbiology*, 2003, 149, 3099-109

Goodwin TJ & Poulter RT, A new group of tyrosine recombinase-encoding retrotransposons, *Molecular Biology and Evolution*, 2004, 21 (4), 746-59

Gueguen E, Rousseau P, Duval-Valentin G, Chandler M, The transpososome: control of transposition at the level of catalysis, *Trends in Microbiology*, 2005, 13, 543-49

Grivna ST, Pyhtila B, Lin H, MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis, *PNAS*, 2006, 103, 13415-20

Guiot E, Carayon K, Delelis O, Simon F, Tauc P, Zubin E, Gottikh M, Mouscadet JF, Brochon JC, Deprez E, Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity, *Journal of Biological Chemistry*, 2006, 281, 22707-19

Hackett CS, Geurts AM, Hackett PB, Predicting preferential DNA vector insertion sites: implications for functional genomics and gene therapy, *Genome Biology*, 2007, 8, Suppl 1-S12

Hagemann S & Pinsker W, Drosophila P transposons in human genome?, *Molecular Biology and Evolution*, 2001, 18 (10), 1979-82

Haniford DB, Transpososome dynamics and regulation in Tn*10* transposition, *Critical Reviews in Biochemistry and Molecular Biology*, 2006, 41, 407-24

Haren L, Ton-Hoang B, Chanlder M, Integrating DNA: transposases and retroviral intégrases, *Annual Reviex of Microbiology*, 1999, 53, 245-81

Hazuda D, Iwamoto M, Wenning L, Emerging pharmacology: inhibitors of human immunodeficiency virus integration, *Annual Reviex of Pharmacology and Toxicology*, 2009, 49, 377-94

Heierhorst J, Lederis K, Richter D, Presence of a member of the Tc*1*-like transposon family from nematodes and *Drosophila* within the vasotocin gene of a primitive vertebrate, the Pacific hagfish *Eptatretus stouti*, *PNAS,* 1992, 89 (15), 6798-802

Hemmrich G, Miller DJ, Bosch TC, The evolution of immunity: a low-life perspective, *Trends in Immunology*, 2007, 28 (10), 449-54

Hickman AB, Perez ZN, Zhou L, Musingarimi P, Ghirlando R, Hinshaw JE, Craig NL, Dyda F, Molecular architecture of a eukaryotic DNA transposase, *Nature Structural and Molecular Biology*, 2005, 12 (8), 715-21

Hiom K & Gellert M, Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination, *Molecular Cell*, 1998, 1 (7), 1011-19

Hromas R, Wray J, Lee S-H, *et al*, The human set and transposase domain protein Metnase interacts with DNA Ligase IV and enhances the efficiency and accuracy of non-homologous end-joining, *DNA Repair*, 2008, 7, 1927-37

Islam MD, Yoshimoto M, Yahara S, Okuno S, Ishiguro H, Yamakawa O, Identification and characterization of foliar polyphenolic composition in sweetpotato genotype, 2002, Journal of Agriculture and Food Chemistry, 50, 3718-22

Ivics Z, Hackett PB, Plasterk RH, Izsvák Z, Molecular reconstruction of *Sleeping Beauty*, a Tc*1*-like transposon from fish, and its transposition in human cells, *Cell,* 1997, 91 (4), 501-510

Izsvák Z, Ivics Z, Plasterk RH, *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates, *Journal of Molecular Biology*, 2000, 302 (1), 93-102

Izsvák Z , Khare D, Behlke J, Heinemann U, Plasterk RH , Ivics Z, Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in *Sleeping Beauty* transposition, *Journal of Biological Chemistry*, 2002, 277 (37), 34581-8

Izsvák Z, Stüwe EE, Fiedler D, Katzer A, Jeggo PA, Ivics Z, Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells, *Molecular Cell*, 2004, 13 (2), 279-90

Jacobson JW & Hartl DL, Coupled instability of two X-linked genes in *Drosophila mauritiana*: germinal and somatic mutability, *Genetics*, 1985, 111 (1), 57-65

Jing N, De Clercq E, Rando RF, Pallansch L, Lackman-Smith C, Lee S, Hogan ME, Stability-activity relationships of a family of G-tetrad forming oligonucleotides as potent HIV inhibitors, 2000a, *Journal of Biological Chemistry*, 275, 3421-30

Jing N, Marchand C, Liu J, Mitra R, Hogan ME, Pommier Y, Mechanism of inhibition of HIV-1 integrase by G-tetrad-forming oligonucleotides *in vitro*, 2000b, *Journal of Biological Chemistry*, 275, 21460-67

Jurka J & Kapitonov V, PIFs meet *Tourists* and *Harbingers*: A superfamily reunion, *PNAS*, 2001, 98 (22), 12315-16

Kapitonov V & Jurka J, Molecular paleontology of transposable elements from *Arabidopsis thaliana*, *Genetica*, 1999, 107 (1-3), 27-37

Kapitonov V & Jurka J, Rolling-circle transposons in eukaryotes, *PNAS*, 2001, 98, 8714-19

Kapitonov V & Jurka J, Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome, *PNAS*, 2003, 100 (11), 6569-74

Kapitonov V & Jurka J, *Harbinger* transposons and an ancient HARBI1 gene derived from a transposase, *DNA cell and biology*, 2004, 23 (5), 311-24

Kapitonov V & Jurka J, RAG1 core and V(D)J recombination signal sequences were derived from *Transib* transposons, *PLOS Biology*, 2005, 3 (6), 998-1011

Kapitonov V & Jurka J, Self-synthesizing DNA transposons in eukaryotes, *PNAS*, 2006, 103 (12), 4540-45

Kapitonov V & Jurka J, *Helitrons* on a roll: eukaryotic rolling-circle transposons, *Trends in Genetics*, 2007, 23 (10), 521-29

Kapitonov V & Jurka J, A universal classification of eukaryotic transposable elements implanted in Repbase, *Nature Review Genetics*, 2008, 9, 411-12

Katayanagi K, Miyagawa M, *et al*, Three-dimensional structure of ribonuclease H from *E. coli*, *Nature*, 1990, 347, 306-9

Kaufman PD & Rio DC, *P* element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor, *Cell*, 1992, 69 (1), 27-39

Kawakami K & Shima A, Identification of the *Tol2* transposase of the medaka fish *Oryzias latipes* that catalyzes excision of a nonautonomous *Tol2* element in zebrafish *Danio rerio*, 1999, *Gene*, 240, 239-44

Kawakami K, Imanaka K, Itoh M, Taira M, Excision of the *Tol2* transposable element of the medaka fish *Oryzias latipes* in *Xenopus laevis* and *Xenopus tropicalis*, *Gene*, 2004, 338, 93-98

Kennelly P, A Protein kinases and protein phosphatases in prokaryotes: a genomic perspective, *FEMS Microbiology Letters*, 2002, 206 (1), 1-8

Keravala A, Liu D, Lechman E, Wolfe D, Nash J, Lampe D, Robbins P, Hyperactive *Himar1* transposase mediates transposition in cell culture and enhances gene expression *in vivo*, *Human Gene Therapy*, 2006, 17 (10), 1006–18

Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH, Mut-7 of C. elegans, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD, *Cell*, 1999, 99, 133-41

Klenchin VA, Czyz A, Goryshin IY, Gradman R, Lovell S, Rayment I, Reznikoff WS, Phosphate coordination and movement of DNA in the Tn*5* synaptic complex: role of the (R)YREK motif, *Nucleic Acids Research*, 2008, 36 (18), 5855-62

Klink T, Woycechowsky K, Taylor K, Raines R, Contribution of disulfide bonds to the conformational stability and catalytic activity of ribonuclease, *European Journal of Biochemistry*, 2000, 267 (2), 566–72

Koga A, Suzuki M, Inagaki H, Behsso Y, Hori H, Transposable element in fish, *Nature*, 1996, 383, 30

Koga A, Shimada A, Shima A, Sakaizumi M, Tachida H, Hori H, Evidence for recent invasion of the medaka fish genome by the *Tol2* transposable element, *Genetics*, 2000, 155, 273-81

Koga A, Higashide I, Hori H, Wakamatsu Y, Kyono-Hamaguchi Y, Hamaguchi S, The *Tol1* element of medaka fish is transposed with only terminal regions and can deliver large DNA fragments into the chromosomes, *Journal of Human Genetics*, 2007, 52 (12), 1026-30

Kramerov DA & Vassetzky NS, Short retroposons in eukaryotic genomes, *International Review of Cytology*, 2005, 247, 165-221

Kunze R, Behrens U, Courage-Franzkowiak U, Feldmar S, Kühn S, Lütticke R, Dominant transposition-deficient mutants of maize Activator (Ac) transposase, *PNAS*, 1993, 90 (15), 7094-8

Lachner M & Jenuwein T, The many faces of histone lysine methylation, *Current Opinion in Cellular Biology*, 2002, 14 (3), 286-98

Lam WL, Seo P, Robison K, Virk S, Gilbert W, Discovery of amphibian Tc1-like transposon families, *Journal of Molecular Biology*, 1996, 257 (2), 359-66

Lampe DJ, Churchill ME, Robertson HM, A purified *mariner* transposase is sufficient to mediate transposition *in vitro*, *EMBO Journal*, 1996, 15 (19), 5470-79

Lampe DJ, Grant TE, Robertson HM, Factors affecting transposition of the *Himar1 mariner* transposon *in vitro*, *Genetics*, 1998, 149 (1), 179-87

Lampe DJ, Akerley BJ, Rubin E, Mekalanos J, Robertson HM, Hyperactive transposase mutants of the *Himar1 mariner* transposon, 1999, *PNAS*, 96 (20), 11428–33

Langdon T, Jenkins G, Hasterok R, Neil Jones R, King I, A high-copy-number CACTA family transposon in temperate grasses and cereals, 2003, *Genetics*, 163, 1097-108

Lander ES, Linton LM, Birren B *et al,* Initial sequencing and analysis of the human genome, *Nature*, 2001, 409 (6822), 860-921

Laski FA, Rio DC, Rubin GM, issue specificity of Drosophila *P* element ransposition is regulated at the level of mRNA splicing, *Cell*, 1986, 44, 7-9

Lee J, Cordaux R, Han K, Wang J, Hedges DJ, Liang P, Batzer MA, Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons, *Gene*, 2007, 390 (1-2), 18-27

Lee S-H, Oshige M, Durant S *et al,* The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair, *PNAS*, 2005, 102 (50), 18075-18080

Leroy H, Castagnone-Sereno P, Renault S, Augé-Gouillou C, Bigot Y, Abad P, Characterization of Mcmar1, a *mariner*-like element with large inverted terminal repeats (ITRs) from the phytoparasitic nematode Meloidogyne chitwoodi, *Gene*, 2003, 304, 35-41

Levy AA, Walbot V, Regulation of the timing of transposable element excision during maize development, *Science*, 1990, 248, 1534-7

Li F, Papworth M, Minczuk M, Rohde C, Zhang Y, Ragozin S, Jeltsch A, Chimeric, DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes, 2007, *Nucleic Acids Research*, 35 (1)**,** 100-12

Li Y & Dooner H, Excision of *Helitron* transposons in maize, 2009, *Genetics*, 182 (1)**,** 399-402

Lipkow K, Buisine N, Chalmers R, Promiscuous target interactions in the *mariner* transposon *Himar1*, *Journal of Biological Chemistry*, 2004a, 279 (47), 48569-75

Lipkow K, Buisine N, Lampe DJ, Chalmers R, Early intermediates of *mariner* transposition: catalysis without synapsis of the transposon ends suggests a novel architecture of the synaptic complex, *Molecular and Cellular Biology*, 2004b, 24 (18), 8301-11

Lisch D, Chomet P, Freeling M, Genetic characterization of the *Mutator* system in maize: behavior and regulation of *Mu* transposons in a minimal line, 1995, *Genetics*, 139 (4), 1777–96

Lisch D, *Mutator* transposons, 2002, *Trends in Plant Science*, 7 (11)**,** 498-504

Liu D, Bischerour J, Siddique A, Buisine N, Bigot Y, Chalmers R, The human SETMAR protein preserves most of the activities of the ancestral *Hsmar1* transposase, *Molecular and Cellular Biology*, 2007, 27 (3), 1125-32

Lohe AR & Hartl DL, Germline transformation of *Drosophila virilis* with the transposable element *mariner*, *Genetics*, 1996, 143(1), 365-74.

Lohe AR, De Aquiar D, Hartl DL, Mutations in the *mariner* transposase: the D,D(35)E consensus sequence is non-functional, *PNAS*, 1997, 94 (4), 1293-97

Lohe AR & Hartl DL, Efficient mobilization of *mariner* in vivo requires multiple internal sequences, *Genetics*, 2002, 160 (2), 519-26

Lubkowski J, Yang F, Alexandratos J *et al*, Structure of the catalytic domain of avian sarcoma virus intégrase with a bound HIV-1 integrase-targeted inhibitor, *PNAS*, 1998, 95, 4831-36

Luo G, Ivics Z, Izsvak Z, Bradley A, Chromosomal transposition of a *Tc1/mariner-like* element in mouse embryonic stem cells, *PNAS*, 1998, 95 (18), 10769-77

Lyon MF, X chromosome inactivation: a repeat hypothesis, *Cytogenetics and cell genetics*, 1998, 80 (1-4), 133-37

Lyon MF, LINE-1 elements and X chromosome inactivation: a function for 'junk' DNA?, *PNAS*, 2000, 97 (12), 6248-49

Lyon MF, Do LINEs have a role in X chromosome inactivation, *Journal of biomedicine and biotechnology*, 2006, 2006 (1), 59746

Marchand C, Krajewski K, Lee HF *et al*, Covalent binding of the natural antimicrobial peptide indolicidin to DNA abasic sites, *Nucleic Acids Research*, 2006, 34 (18), 5157-65

Maruyama K, Schoor KD, Hartl DL, Identification of nucleotide substitutions necessary for trans-activation of *mariner* transposable elements in Drosophila: analysis of naturally occurring elements, *Genetics*, 1991, 128 (4), 777-84

Masson P, Rutherford G, Banks JA, Fedoroff N, Essential large transcripts of the maize *Spm* transposable element are generated by alternative splicing, *Cell*, 1989, 58, 755-65

Mates L, Izsvak Z, Ivics Z, Technology transfer from worms and flies to vertebrates: transposition-based genome manipulations and their future perspectives, *Genome Biology*, 2007, 8, Suppl 1-S1

McClintock B, Mutable loci in maize, *Carnegie Inst. Washington Year Book*, 1948, 47, 155-69

McClintock B, The origin and behaviour of mutable loci in maize, *PNAS*, 1950, 36 (6), 344-55

McClintock B, Mutations in maize and chromosomal aberrations in neurospora, *Carnegie Inst. Washington Year Book*, 1954, 53, 254-60

McGinnis W, Shermoen AW, Beckendorf SK, A transposable element inserted just 5' to a Drosophila glue protein gene alters gene expression and chromatin structure, *Cell*, 1983, 34 (1), 75-84

Medhora MM, MacPeek AH, Hartl DL, Excision of the *Drosophila* transposable element *mariner*: identification and characterization of the *Mos* factor, *EMBO Journal*, 1988, 7 (7), 2185-89

Medhora M, Maruyama K, Hartl DL, Molecular and functional analysis of the *mariner* mutator element *Mos1* in *Drosophila*, *Genetics*, 1991, 128 (2), 311-18

Melek M, Jones JM, O'Dea MH, *et al*, Effects of HIV intégrase inhibitors on the RAG1/2 recombinase, *PNAS*, 2002, 99 (1), 134-7

Mills RE, Bennett EA, Iskow RC, Devine SE, Which transposable elements are active in the human genome?, *Trends in Genetics*, 2007, 23 (4), 183-91

Miskey C, Izsvak Z, Plasterk RH, Ivics Z, The *Frog Prince*: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells, *Nucleic Acids Research*, 2003, 31 (23), 6873-81

Miskey C, Izsvak Z, Kawakami K, Ivics Z, DNA transposons in vertebrate functional genomics, *Cellular and Molecular Life Sciences*, 2005, 62 (6), 629-41

Miskey C, Papp B, Mates L, Sinzelle L, Keller H, Izsvak Z, Ivics Z, The ancient *mariner* sails again: transposition of the human *Hsmar1* element by a reconstructed transposase and activities of the SETMAR protein on transposon ends, *Molecular and Cellular Biology*, 2007, 27 (12), 4589-4600

Misra S & Rio D, Cytotype control of Drosophila P element transposition: the 66 kd protein is a repressor of transposase activity, *Cell*, 1990, 62 (2), 269-84

Mitra R, Fain-Thornton J, Craig N, *piggyBac* can bypass DNA synthesis during cut and paste transposition, *EMBO Journal*, 2008, 27 (7), 1097-109

Mizuuchi K, Polynucleotidyl transfer reactions in transpositional DNA recombination, *Journal of Biological Chemistry*, 1992, 267 (30), 21273-6

Moerman DG & Waterston RH, Spontaneous unstable unc-22 IV mutations in *C. elegans* var. Bergerac, *Genetics*, 1984, 108 (4), 859-77

Moerman DG, Kiff JE, Waterston RH, Germline excision of the transposable element Tc1 in *C. elegans*, *Nucleic Acids Research*, 1991, 19 (20), 5669-72

Munoz-Lopez M, Siddique A, Bischerour J, Lorite P, Chalmers R, Palomeque T, Transposition of *Mboumar-9*: identification of a new naturally active *mariner*-family transposon, *Journal of Molecular Biology*, 2008, 382 (3), 567-72

Nakayama K, Yamashita A, Kurokawa K, *et al*, The Whole-genome sequencing of the obligate intracellular bacterium Orientia tsutsugamushi revealed massive gene amplification during reductive genome evolution, *DNA Research*, 2008, 15 (4), 185-99

Nowotny M, Retroviral integrase superfamily: the structural perspective, *EMBO Reports*, 2009, 10 (2), 144-51

Ohtsu K, Hirano HY, Tsutsumi N, Hirai A, Nakazono M, *Anaconda*, a new class of transposon belonging to the *Mu* superfamily, has diversified by acquiring host genes during rice evolution, *Molecular Genetics and Genomics*, 2005, 274, 606-15

Oosumi T, Belknap WR, Garlick B, *Mariner* transposons in human, *Nature*, 1995, 378 (6558), 672

Palomeque T, Antonio Carillo J, Munoz-Lopez M, Lorite P, Detection of a *mariner*-like element and a miniature inverted-repeat transposable element (MITE) associated with the heterochromatin from ants of the genus Messor and their possible involvement for satellite DNA evolution, *Gene*, 2006, 371 (2), 194-205

Pardue ML & DeBaryshe PG, Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres, *Annual review of genetics*, 2003, 37, 485-511

Peterson PA, A mutable pale green locus in maize, *Genetics*, 1953, 38, 682-83

Paulsen IT, Banerjei L, Myers GS *et al*, Role of mobile DNA in the evolution of vancomycin-resistant Enterococcus faecalis, *Science*, 2003, 299 (5615), 2071-74

Piao Z, Shibayama K, Mori S, Wachino J, Arakawa Y, A novel insertion sequence, IS*1642*, of *Mycobacterium avium*, which forms long direct repeats of variable length, *FEMS Microbiology Letters*, 2009, 291 (2), 216-21

Picault N, Chaparro C *et al*, Identification of an active LTR retrotransposon in rice, *The Plant Journal*, 2009, 58 (5), 754-65

Pietrokovsky S & Henikoff S, A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons, *Molecular & general genetics*, 1997, 254 (6), 689-95

Plasterk RH, Izsvak Z, Ivics Z, Resident aliens: the Tc*1/mariner* superfamily of transposable elements, *Trends in Genetics*, 1999, 15 (8), 326-32

Pledger DW, Fu YQ, Coates CJ, Analyses of cis-acting elements that affect the transposition of *Mos1 mariner* transposons in vivo, *Molecular Genetics and Genomics*, 2004, 272 (1), 67-75

Pledger DW & Coates CJ, Mutant *Mos1 mariner* transposons are hyperactive in *Aedes aegypti*, *Insect Biochemistry and Molecular Biology*, 2005, 35 (10), 1199-1207

Polard P, Prère MF, Chandler M, Fayet O, Programmed translational frameshifting and initiation at an AUU codon in gene expression of bacterial insertion sequence IS*911*, *Journal of Molecular Biology*, 1991, 222 (3), 465-77

Polard P & Chandler M, An *in vivo* transposase-catalyzed single-stranded DNA circularization reaction, *Genes & Development*, 1995, 9 (22), 2846-58

Polard P, Ton-Hoang B, Haren L, Bétermier M, Walczak R, Chandler M, IS*911*-mediated transpositional recombination *in vitro*, *Journal of Molecular Biology*, 1996, 264, 68-81

Poulter RT, Goodwin TJ, Butler MI, Vertebrate helentrons and other novel *Helitrons*, *Gene*, 2003, 313, 201-12

Prère MF, Chandler M, Fayet O, Transposition in *Shigella dysenteriae*: isolation and analysis of IS*911*, a new member of the IS*3* group of insertion sequences, *Journal of Bacteriology*, 1990, 172 (7), 4090-9

Pritham E, Feschotte C, Wessler S, Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans, *Molecular Biology and Evolution*, 2005, 22 (9), 1751-63

Pritham E, Putliwala T, Feschotte C, *Mavericks*, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses, *Gene*, 2007, 390, 3-17

Pritham E & Feschotte C, Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*, *PNAS*, 2007, 104 (6), 1895-900

Pritham E, Transposable elements and factors influencing their success in eukaryotes, *Journal of Heridity*, 2009, 100 (9), 648-55

Pyatkov KI, Arkhipova IR, Makhova NV, Finnegan DJ, Evgen'ev MB, Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements, *PNAS*, 2004, 101 (41), 1479-24

Raina R, Schläppi M, Karunanandaa B, Elhofy A, Fedoroff N, Concerted formation of macromolecular Suppressor-mutator transposition complexes, 1998, *PNAS*, 95 (15), 8526-31

Raizada MN & Walbot V, The late developmental pattern of *Mu* transposon excision is conferred by a cauliflower mosaic virus 35S-driven MURA cDNA in transgenic maize, 2000, *Plant Cell*, 12, 5–21

Reznikoff WS, Tn*5* transposition, In Craig N *et al*, *Mobile DNA II*, ASM Press, 2002, 403–21

Reznikoff WS, Tn*5* as a model for understanding DNA transposition, 2003, *Molecular Microbiology*, 47, 1199–206

Reznikoff WS, Tn*5* transposition: a molecular tool for studying protein structure-function, 2006, *Biochemical Society Transactions*, 34, 320–23

Reznikoff WS, Transposon Tn*5*, 2008, *Annual Review of Genetics*, 42, 269-86

Rice P & Baker T, Comparative architecture of transposase and integrase complexes, *Nature Structural Biology*, 2001, 8 (4), 302-07

Richardson JM, Dawson A, O'Hagan N, Taylor P, Finnegan DJ, Walkinshaw MD, Mechanism of *Mos1* transposition: insights from structural analysis, *EMBO Journal*, 2006, 25 (6), 1324-34

Rio D & Rubin G, Identification and purification of a Drosophila protein that binds to the terminal 31-base-pair inverted repeats of the *P* transposable element, *PNAS*, 1988, 85 (23), 8929-33

Rio D, P transposable elements in *Drosophila melanogaster*, In Craig N *et al*, *Mobile DNA II*, ASM Press, 2002, 484-518

Robert V & Bessereau J-L, Targeted engineering of the *Caenorhabditis elegans* genome following *Mos1*-triggered chromosomal breaks, *EMBO Journal*, 2007, 26, 170-83

Roberts AP, Chandler M, Courvalin P *et al*, Revised nomenclature for transposable genetic elements, *Plasmid*, 2008, 60 (3), 167-73

Robertson HM, The *mariner* transposable element is widespread in insects, *Nature*, 1993, 362, 241-45

Robertson HM & Lampe DJ, Distribution of transposable elements in arthropods, *Annual review of entomolgy*, 1995a, 40, 333-57

Robertson HM & Lampe DJ, Recent horizontal transfer of a *mariner* transposable element among and between Diptera and Neuroptera, *Molecular biology and evolution*, 1995b, 12 (5), 850-62

Robertson HM & Asplund ML, *Bmmar1*: a basal lineage of the *mariner* family of transposable elements in the silkworm moth, *Bombyx mori*, *Insect Biochemistry and Molecular Biology*, 1996, 26 (8-9), 945-54

Robertson HM & Martos R, Molecular evolution of the second ancient human *mariner* transposon, *Hsmar2*, illustrates patterns of neutral evolution in the human genome lineage, *Gene*, 1997, 205 (1-2), 219-28

Robertson HM & Zumpano KL, Molecular evolution of an ancient *mariner* transposon, *Hsmar1*, in the human genome, *Gene*, 1997, 205 (1-2), 203-17

Roman Y, Oshige M, Lee Y-J *et al*, Biochemical characterization of a SET and transposase fusion protein, Metnase: its DNA binding and DNA cleavage activity, *Biochemistry*, 2007, 46, 11369-76

Rose AM & Snutch TP, Isolation of the closed circular form of the transposable element *Tc1* in *Caenorhabditis elegans*, *Nature*, 1984, 311 (5985), 485-6

Ruan K & Emmons SW, Extrachromosomal copies of transposon *Tc1* in the nematode *Caenorhabditis elegans*, *PNAS*, 1984, 81 (13), 4018-22

Rubin E, Lithwick G, Levy AA, Structure and evolution of the *hAT* transposon superfamily, *Genetics*, 2001, 158 (3), 949-57

Ryabinin VA, Sinyakov AN, de Soultrait VR *et al*, Inhibition if HIV-1 integrase-catalysed reaction by new DNA minor groove: the oligo-1,3-thiazolecarboxamide derivatives, *European Journal of Medicinal Chemistry, 2000,* 35 (11), 989-1000

Sakai J, Chalmers R, Kleckner N, Identification and characterization of a pre-cleavage synaptic complex that is an early intermediate in Tn*10* transposition, *EMBO Journal*, 1995, 14, 4374-83

Sakai J, Kleckner N, Yang X, Guhathakurta A, Tn*10* transpososome assembly involves a folded intermediate that must be unfolded for target capture and strand transfer, *EMBO Journal*, 2000, 19, 776-85

Sarkar A, Sim C, Hong Y, Hogan J, Fraser MJ, Robertson HM, Collins F, Molecular evolutionary analysis of the widespread *piggyBac* transposon family and related ‘‘domesticated’’ sequences, *Molecular Genetics and Genomics*, 2003, 270, 173-80

Schläppi M, Raina R, Fedorrof N, Epigenetic regulation of the maize *Spm* transposable element: novel activation of a methylated promoter by TnpA, *Cell*, 1994, 77 (3), 427-37

Schultz WA, L1 retrotransposons in human cancer, *Journal of biomedicine and biotechnology*, 2006, 2006 (1), 83672

Sekine Y, Izumi K, Mizuno T, Ohtsubo E, Inhibition of transpositional recombination by OrfA and OrfB proteins encoded by insertion sequence IS3, *Genes to cells*, 1997, 2 (9), 547-57

Semenova EA, Marchand C, Pommier Y, HIV-1 integrase inhibitors: update and perspectives, *Advances in pharmacology*, 2008, 56, 199-228

Shao H & Tu Z, Expanding the diversity of the IS*630-Tc1-mariner* superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons, *Genetics*, 2001, 159 (3), 1103-15

Siguier P, Filée J, Chandler M, Insertion sequences in prokaryotic genomes, *Current opinion in microbiology*, 2006, 9 (5), 526-31

Silva JC, Bastida F, Bidwell SL, Johnson PJ, Carlton JM, A potentially functional *mariner* transposable element in the protist *Trichomonas vaginalis*, *Molecular Biology and Evolution*, 2005, 22 (1), 126-34

Simmons MJ & Bucholz LM, Transposase titration in *Drosophila melanogaster*: a model of cytotype in the P-M system of hybrid dysgenesis, PNAS, 1985, 82 (23), 8119-23

Singh R, Liburd J, Wardle S, Haniford D, The nucleoid binding protein H-NS acts as an anti-channeling factor to favour intermolecular Tn*10* transposition and dissemination, *Journal of Molecular Biology*, 2008, 376, 950-62

Sinzelle L, Pollet N, Bigot Y, Mazabraud A, Characterization of multiple lineages of Tc1-like elements within the genome of the amphibian *Xenopus tropicalis*, *Gene*, 2005, 349, 187-96

Sinzelle L, Jégot G, Brillet B, Rouleux-Bonnin F, Bigot Y, Augé-Gouillou C, Factors acting on *Mos1* transposition efficiency, *BMC Molecular Biology*, 2008, 9 (106)

Sinzelle L, Kapitonov V, Grzela D, Jursch T, Jurka J, Izsvak Z, Ivics Z, Transposition of a reconstructed Harbinger element in human cells and functional homology with two-transposon-derived cellular genes, *PNAS*, 2008, 105 (12), 4715-20

Sinzelle L, Izsvak Z, Ivics Z, Molecular domestication of transposable elements: from detrimental parasites to useful host genes, *Cellular and Molecular Life Sciences*, 2009, 66 (6), 1073-93

Slotkin RK, Freeling M, Lisch D, *Mu killer* causes the heritable inactivation of the *Mutator* family of transposable elements in *Zea mays*, *Genetics*, 2003, 165 (2), 781-97

Staveley BE, Heslip TR, Hodgetts RB, Bell JB, Protected *P*-element termini suggest a role for inverted-repeat-binding protein in transposase-induced gap repair in *Drosophila melanogaster*, *Genetics*, 1995, 139 (3), 1321-9

Steiniger M, Metzler J, Reznikoff WS, Mutation of Tn*5* transposase β-loop residues affects all steps of Tn*5* transpostion: the role of conformational changes in Tn*5* transposition, 2006, *Biochemistry*, 45, 15552–62

Sumitani M, Lee JM, Hatakeyama M, Oishi K, Cloning and characterization of Acmar1, a *mariner*-like element in the asiatic honey bee, Apis cerana japonica (Hymenoptera, Apocrita), *Archives of insect biochemistry and physiology*, 2002, 50 (4), 183-90

Tang M, Cecconi C, Kim H, Bustamante C, Rio D, Guanosine triphosphate acts as a cofactor to promote assembly of initial P-element transposase–DNA synaptic complexes, *Genes & Development*, 2005, 19, 1422-25

Tang M, Cecconi C, Bustamante C, Rio D, Analysis of *P* element transposase protein-DNA interactions during the early stages of transposition, *Journal of Biological Chemistry*, 2007, 282 (39), 29002-12

Tenzen T, Matsutani S, Ohtsubo E, Site specific transposition of insertion sequence IS630, *Journal of bacteriology*, 1990, 172 (7), 3830-36

Tosi LR & Beverley SM, *cis* and *trans* factors affecting *Mos1 mariner* evolution and transposition *in vitro*, and its potential for functional genomics, *Nucleic Acids Research*, 2000, 28 (3), 784-79

Toussaint A & Merlin C, Mobile elements as a combination of functional modules, *Plasmid*, 2002, 47 (1), 26-35

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD, A distinct small RNA pathway silences selfish genetic elements in the germline, *Science*, 2006, 313, 320-4

van Luenen HG, Colloms SD, Plasterk RH, Mobilization of quiet, endogenous *Tc3* transposons of *Caenorhabditis elegans* by forced expression of *Tc3* transposase, *EMBO Journal*, 1993, 12 (6), 2513-20

van Luenen HG, Colloms S, Plasterk RH, The mechanisme of transposition of *Tc3* in *C. elegans*, *Cell*, 1994, 79 (2), 293-301

van Luenen HG & Plasterk RH, Target site choice of the related transposable elements *Tc1* and *Tc3* of *Caenorhabditis elegans*, *Nucleic Acids Research*, 1994, 22, 262-9

van Maele B & Debyser Z, HIV-1 integration: an interplay between HIV-1 integrase, cellular and viral proteins, *AIDS Reviews*, 2005, 7 (1), 26-43

van Maele B, Busschots K, Vandekerckhove L, Christ F, Debyser Z, Cellular co-factors of HIV-1 integration, *Trends in Biochemical Sciences*, 2006, 31 (2), 98-105

Vastenhouw NL & Plasterk RH, RNAi protects the Caenorhabditis elegans germline against transposition, *Trends in Genetics*, 2004, 20 (7), 314-9

Vigdal TJ, Kaufman CD, Izsvak Z, Voytas DF, Ivics Z, Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other *Tc1/mariner* transposable elements, *Journal of Molecular Biology*, 2002, 323 (3), 441-52

Volff J-N, Turning junk into gold: domestication of transposable elements and the creation of new genes in eucaryotes, *BioEssays*, 2006, 28, 913-22

Vos JC & Plasterk RH, *Tc1* transposase of *Caenorhabditis elegans* is an endonuclease with a bipartite DNA binding domain, *EMBO Journal*, 1994, 13 (24), 6125-32

Vos JC, De Baere I, Plasterk RH, Transposase is the only nematode protein required for *in vitro* transposition of Tc*1*, *Genes & Development*, 1996, 10 (6), 755-761

Walbot V, Saturation mutagenesis using maize transposons, *Current Opinion in Plant Biology*, 2000, 3, 103–7

Walisko O, Izsvák Z, Szabó K, Kaufman CD, Herold S, Ivics Z, Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1, *PNAS*, 2006, 103 (11), 4062-7

Walker EL, Eggleston WB, Demopulos D, Kermicle J, Dellaporta SL, *Genetics*, 1997, 146, 681–93

Ward C, Wardle S, Singh R, Haniford D, The global regulator H-NS binds to two distinct classes of sites within the Tn*10* transpososome to promote transposition, *Molecular Microbiology*, 2007, 64 (4), 1000-13

Wardle SJ, O'Carroll M, Derbyshire KM, Haniford D, The global regulator H-NS acts directly on the transpososome to promote Tn*10* transposition, *Genes & Development*, 2005, 19, 2224-35

Watkins S, van Pouderoyen G, Sixma TK, Structural analysis of the bipartite DNA-binding domain of *Tc3* transposase bound to transposon DNA, *Nucleic Acids Research*, 2004, 32 (14), 4306-12

Wicker T, Guyot R, Yahiaoui N, Keller B, CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements, *Plant Physiology*, 2003, 132, 52-63

Wicker T, Sabot F, Hua-Van A *et al*, A unified classification system for eukaryotic transposable elements, *Nature Review Genetics*, 2007, 8, 973-82

Williamson E, Rasila KK, Corwin LK *et al*, The SET and transposase domain protein Metnase enhances chromosome decatenation: regulation by automethylation, *Nucleic Acids Research*, 2008, 36 (18), 5822-31

Wilson R, Orsetti J, Klocko AD, Aluvihare C, Peckham E, Atkinson PW, Lehane MJ, O'Brochta DA, Post-integration behavior of a *Mos1 mariner* gene vector in *Aedes aegypti*, *Insect biochemistry and molecular biology*, 2003, 33 (9), 853-863

Wilson MH, Coates CJ, George AL, *PiggyBac* transposon-mediated gene transfer in human cells, *Molecular Therapy*, 2003, 15 (1), 139-45

Yang W, Hendrickson WA, Crouch RJ, Satow Y, Structure of ribonuclease H phased at 2A resolution by MAD analysis of the selenomethionyl protein, *Science*, 1990, 249, 1398-405

Yang W, Lee JY, Nowotny M, Making and breaking nucleic acids: two-Mg^{2+}-ion catalysis and substrate specificity, *Molecular Cell*, 2006, 22, 5-13

Yoder JA, Walsh CP, Bestor TH, Cytosine methylation and the ecology of intragenomic parasites, *Trends in Genetics*,1997, 13, 335-40

Yoshiyama M, Tu Z, Kainoh Y, Honda H, Shono T, Kimura K, Possible horizontal transfer of a transposable element from host to parasitoid, *Molecular biology and evolution*, 2001, 18 (10), 1952-58

Yuan JY, Finney M, Tsung N, Horvitz HR, Tc4, a *Caenorhabditis elegans* transposable element with an unusual fold-back structure, *PNAS*, 1991, 88 (8), 3334-38

Zayed H, Izsvak Z, Khare D, Heinemann U, Ivics Z, The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition, *Nucleic Acids Research*, 2003, 31 (9), 2313-22

Zhang L, Dawson A, Finnegan DJ, DNA-binding activity and subunit interaction of the *mariner* transposase, *Nucleic Acids Research*, 2001, 29 (17), 3566-75

Zhao S, Gwyn L, De P, Rodgers K, A non-sequence-specific DNA binding mode of RAG1 is inhibited by RAG2, *Journal of Molecular Biology*, 2009, 387 (3), 744-58

Zhou L, Mitra R, Atkinson P, Hickman A, Dyda F, Craig N, Transposition of *hAT* elements links transposable elements and V(D)J recombination, *Nature*, 2004, 432, 995-1001

Zilberman D & Henikoff S, Silencing of transposons in plant genomes: kick them when they're down, 2004, Genome Biology 5 (249)

Annexes

1/ Publications et communications

Publications scientifiques :

- **Bouchet N, Bischerour J, Germon S, Guillard J, Dubernet M, Viaud-Massuard M-C, Delelis O, Ryabinin V, Bigot Y, Augé-Gouillou C**, First *mariner Mos1* transposase inhibitors, *Mini reviews in medicinal chemistry*, 2009, 9 (4), 431-39
- **Germon S, Bouchet N, Casteret S, Carpentier G, Adet J, Bigot Y, Augé-Gouillou C**, *Mariner Mos1* optimization by rational mutagenesis, *Genetica*, 2009, publié en ligne
- **Carpentier G, Adet J, Bouchet N, Demattei M, Renault S, Bigot Y, Augé-Gouillou C**, *Mariner* mobility requires a single helix-turn-helix in the N-terminal transposase domain, soumis à *EMBO Reports*

Communications orales

- **Bouchet N, Germon S, Bigot, Augé-Gouillou C**, Identification et caractérisation des premiers inhibiteurs de la transposase de l'élément *mariner Mos1*, **XVème Congrès des Eléments Transposables**, 3-5 juillet 2007, Rennes, France
- **Bouchet N**, Transposons World & Transposons Tools, **International Life Science Students' Conference**, 10-14 septembre 2008, Varsovie, Pologne
- **Bouchet N**, Identification and characterization of inhibitors of the eukaryotic *Mos1* transposase, **Experimental Sciences in 'Pablo de Ovalide' University**, 12-16 novembre 2008, Séville, Espagne

Communications affichées

- **Bouchet N, Bischerour J, Germon S, Viaud-Massuard M, Bigot Y, Augé-Gouillou C**, Identification and characterization of inhibitors of the eukaryotic *Mos1* transposase, **International Congress on Transposable Elements**, 20-23 avril 2008, Saint-Malo, France

- **Carpentier G, Bouchet N, Augé-Gouillou C**, **Bigot Y,** Phosphorylation pattern of the *mariner Mos1* transposase expressed with the baculovirus system, **International Congress on Transposable Elements**, 20-23 avril 2008, Saint-Malo, France

First *mariner Mos1* transposase inhibitors

Bouchet N, Bischerour J, Germon S, Guillard J, Dubernet M, Viaud-Massuard M-C, Delelis O, Ryabinin V, Bigot Y, Augé-Gouillou C,

Mini reviews in medicinal chemistry, 2009, 9 (4), 431-39

First *Mariner Mos1* Transposase Inhibitors

N. Bouchet[1,2], J. Bischerour[1,2], S. Germon[1], J. Guillard[1,2], M. Dubernet[1,2], M.C. Viaud-Massuard[1,2], O. Delelis[4], V. Ryabinin[5], Y. Bigot[1,2,3] and C. Augé-Gouillou[1,2,*]

[1]*Génétique, Immunothérapie, Chimie & Cancer CNRS UMR6239 - 37200 Tours, France;* [2]*University of Tours - 37200 Tours, France;* [3]*CHRU of Tours - 37200 Tours, France;* [4]*LBPA CNRS UMR8113 - Ecole Normale Supérieure - 94230 Cachan, France;* [5]*Institute of Chemical Biology and Fundamental Medicine - Siberian Division of Russian Academy of Sciences - 630090 Novosibirsk, Russia*

Abstract: We described chemical inhibitors of *Mos1* transposition. Some were already known to affect a related prokaryotic transposase (Tn*5*) or HIV-1 integrase, whereas the other were new compounds in this field. The new compounds were all organized around a bis-(heteroaryl)maleimides scaffold. Their mechanism of action depended on the chemical substitutions on the scaffold. The cross-activity, between HIV-1 integrase and *Mos1* transposase, of the new group of inhibitors showed that *Mos1* transposase could constitute an excellent surrogate HIV-1 inhibitor screen.

For Supplement material, please see the online version of the article.

Key Words: Heterocyclic chemistry, maleimides, transposon inhibitor, HIV-1 integrase inhibitor, mariner inhibitor.

INTRODUCTION

DNA transposons are Class II transposable elements that correspond to discrete DNA segments that are "naturally" able to move within genomes (for a review see [1]). In addition, some of them have been shown experimentally to have a wide-ranging capacity to move into species distantly related to the host from which they had initially been isolated. These properties have made it possible to develop transposon tools for insertional mutagenesis and germ-line transgenesis in model organisms, and potentially for gene therapy. *Mos1* [2], *Himar1* [3], *Minos* [4] *PiggyBac* [5], *Sleeping Beauty* [6], and *Tol2* [7] were selected as the main candidates for the development of such transposon tools.

Mos1 is a 1286-bp long element, which is terminated at both extremities by 28-bp inverted terminal repeats (ITRs) that are imperfectly conserved in sequence. *Mos1* contains a single open reading frame coding MOS1, a 345-amino acid transposase, and it is known to move within the genome of its hosts using a cut-and-paste mechanism. We hypothesize that this mechanism could well consist of four main steps: [1] MOS1 homo-dimerization, [2] assembly of the synaptic complex, [3] excision of *Mos1*, and [4] target recognition and insertion at a new locus (supplemental data, Fig. (**S1**)). Although the organization and composition of the *Mos1* synaptic complex are still debated [8,9], a complex consisting of a pair of *Mos1* ITRs and a transposase tetramer (*i.e.* Paired End Complex 2: PEC2) is probably the complex that carries the activities that allow the concerted integration of the transposon at the new locus [8]. These activities rely on a well-conserved triad of amino acids, DD34D, that together coordinate the cationic cofactors required for catalysis [10]. The organization of its catalytic core links MOS1 to a large family of transposases and integrases that share a similar catalytic core, known as the RNase H-like core [11]. Crystallographic studies have shown that despite the lack of sequence similarity, the structure of the catalytic core, including the location of the invariable DDE/D residues, is very similar in three transposases (MuA, Tn*5* and MOS1) [10,12], and two integrases (HIV-1 and RSV) [12-14]. Due to the presence of this highly conserved triad, these proteins have been designated "DDE-transposases". In fact the *mariner* transposases are an exception, and have a DDD triad [12,13]. DDE-transposases carry out phosphoryl transfer reactions using the same basic chemical steps and similar active sites [11]. Several bacterial transposons, such as *Mu*, simply nick and transfer the 3' ends of the transposon to the target [15]. Transposons of the IS*3* family perform an asymmetric nick at just one 3' end of the element, followed by a strand transfer immediately outside the other transposon end [16]. In both *Mu* and IS*3*, transposition requires DNA replication. Several bacterial transposons, such as Tn*5* and Tn*10*, nick and transfer the 3' ends of the transposon to the complementary strand, thus forming hairpins on the transposon ends [17,18]. In contrast, several eukaryotic transposons, such as Ac/Ds, nick the 5' ends of the transposon, and transfer the 3' ends of the flanking DNA to the complementary strand, thus forming hairpins on the flanking DNA [1]. Other eukaryotic transposons, such as *mariner*, perform a double strand DNA cleavage at the transposon ends without forming a hairpin, and transfer the 3' ends of the transposon to the target DNA [19]. Finally, the DDE-integrase of retro-transposons and retroviruses performs the integration steps using strand transfer reactions that look like those performed by the "cut and paste" DDE-transposases of the *mariner* family.

In order to improve our fundamental understanding of MOS1 activities, but also to provide a better way of controlling transposition in the context of transposon tools, it will be useful to have compounds that are able to inhibit transposase activities. To date, no such compounds have been avail-

*Address correspondence to this author at the GICC UMR6239, Université de Tours, UFR Sciences & Techniques, Parc Grandmont, Bat L, 37200 Tours, France; Tel: (33) 247 36 70 63; Fax: (33) 247 36 70 35;
E-mail: auge@univ-tours.fr

1389-5575/09 $55.00+.00 © 2009 Bentham Science Publishers Ltd.

able for eukaryotic transposases. The aim of the work reported here was to isolate and characterize the first chemical inhibitors of a *mariner* transposase. 159 chemical compounds were selected, and their ability to inhibit *Mos1* transposition *in vitro* was assessed. Since it had recently been demonstrated that Tn*5* transposase can be used to characterize inhibitors of HIV-1 integrase [20,21], we included inhibitors of these two proteins in our panel, and tested their activity as MOS1 inhibitors. The other compounds were selected because they shared structural and chemical properties with established HIV-1 integrase inhibitors. Among the ten chemicals identified as being the best *Mos1* inhibitors, six were already known to be inhibitors of Tn5 transposase and/or HIV-1 integrase. The other four constituted a new group of drugs, and they all contained a bis-(heteroaryl) maleimides scaffold. The mechanism of action of these new substances depends on the chemical substitutions on the scaffold. First, the compounds that are *N*-substituted inhibited protein-DNA complex assembly more efficiently than the compound that is not. Second, the compound that contains formyl groups (instead of the acid functions) more efficiently prevented integration (the last step in transposition) than the other compounds. We also provide evidence that bis-(heteroaryl) maleimide derivatives are useful inhibitors of HIV-1 integrase. This cross-reactivity (between HIV-1 integrase and *Mos1* transposase) of the new inhibitors once again confirms the close relationships that exist between the DDE-enzymes.

IDENTIFICATION OF MOS1 INHIBITORS

In order to identify inhibitors of *Mos1* transposase, we screened a collection of 159 molecules. The first part of the panel (29 compounds) that we tested consisted of inhibitors of Tn*5* transposase and/or HIV-1 integrase (for details see **S3**, experimental section in supplemental data). The other compounds consisted of substituted-heterocyclic derivatives. Although these compounds had never been tested as inhibitors of an integrase or transposase, they were selected because they shared structural and chemical properties with a previously characterized HIV-1 integrase inhibitor.

A transposition assay was performed using purified MOS1 protein fused to the MBP, and the pBC-3T3 plasmid carrying a pseudo-*Mos1* transposon to screen the collection of chemicals. The pseudo transposon was formed by two 3'ITRs and the tetracycline resistance gene, without promoter. The same plasmid was used both as the donor of pseudo-*Mos1*, and as the target for transposon integration: MOS1 excises a 3T3 from pCB-3T3, and then triggers the reinsertion of this pseudo-*Mos1* within the CAT gene of another pBC-3T3 molecule. Here, the *in-vitro* assay is carried out in the same way as had previously been found for the bacterial assay (*29*). Thus, in this assay, transposition events were revealed by promoter tagging, the tetracycline resistance being activated through the CAT gene promoter.

The concentration of compound used for the initial screen was 80 μM. 116 of the 159 compounds tested had an Inhibition Factor (IF) of less than 5, compared to the control experiment performed in the absence of inhibitor but in the presence of DMSO, and 33 had an IF of between 5 and 25 (80% < inhibition < 96%). Among the 29 compounds previously identified as HIV-1 integrase and/or Tn*5* transposase inhibitors, 25 were found in these groups (IF<25) (Fig. (**1a**)). The corresponding transposition frequencies are given as Supplemental data (Table (**S1**)).

The ten best compounds had an IF of greater than 25 (> 96% of inhibition). The structures of these inhibitors are shown in Fig. (**1b**). Compounds (**1**) to (**4**) were bis-(heteroaryl)maleimide derivatives, constituting hitherto-unknown inhibitors of DDE/D transposases. Compounds (**5**) and (**6**) were *N*-methylpyrrole polyamides, which are known to interact with DNA (review in [22]). Compound (**7**) was a biscoumarin derivative that was also shown to be an inhibitor of Tn*5* transposase and HIV1-integrase [20]. Compounds (**8**), (**9**) and (**10**) were a cinnamoyl derivative, a benzoic acid derivative and a thioxothiazolidin substituted by carboxylic acid, respectively. These three compounds had already been identified as Tn*5* transposase inhibitors [20]

Inhibitory Activity on Mos1 Excision

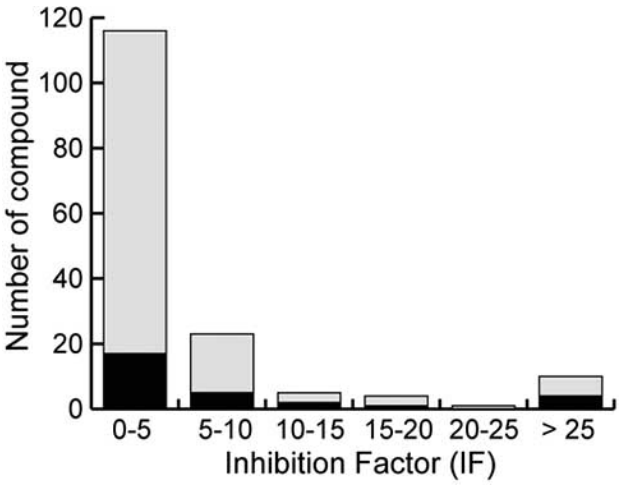
To further characterize the inhibitory activity of these ten compounds and their mechanisms of action on *Mos1* transposition, the ability of each chemical to inhibit transposon excision was tested *in vitro*, using pBC-3T3 as the transposon donor. MOS1 triggers the transposon excision from the plasmid, producing two linear DNA fragments, the transposon (T=1.2 kb) and the plasmid backbone (B=3.4 kb). After excision, the transposon can be reinserted into a target. The excision activity of the drug was measured by quantifying the plasmid backbone, because this is an end product of the reaction. As an example, the assay performed with compound (**9**) is shown in Fig. (**2a**). In the absence of drug, the transposase released the transposon from the donor plasmid (lane 8). At the highest concentration of drug (1 mM), excision of the transposon was completely abolished (lane 2). To determine the concentration of drug required to inhibit 50% of the reaction (IC50), the percentages of inhibition measured from two independent experiments were plotted against the concentration of drug. Similar experiments were performed with the other nine chemicals, and the corresponding IC50 values were estimated (Fig. (**2b**)). Compounds (**2**) to (**6**) were the most effective inhibitors of excision, with IC50 values around 10 μM (Table **1**). A second panel of drugs, with IC50 values ranging from 35 μM (compounds (**1**) and (**10**)) to 150 μM (compound (**9**)), included the Tn*5* inhibitors.

Our data indicated that all the inhibitors have a significant impact on the excision of the transposon. Because the Tn*5* transposase inhibitors (compounds (**7**) to (**10**)) blocked formation of the Tn*5* synaptic complex [20], and because compounds (**5**) and (**6**) were powerful DNA binding protein inhibitors (review in [22]), we decided to try to find out whether these six compounds act at the level of the formation of transposase-ITR complexes, a step that takes place before excision.

Inhibitory Effect on the Assembly of MOS-ITR Complexes

We first focused our attention on the ability of drugs 5 to 10 to inhibit the formation of complexes between MOS1 and the 3'ITR [8], using a gel shift mobility assay (EMSA). MOS1-ITR complexes were formed with a radiolabeled

a



b

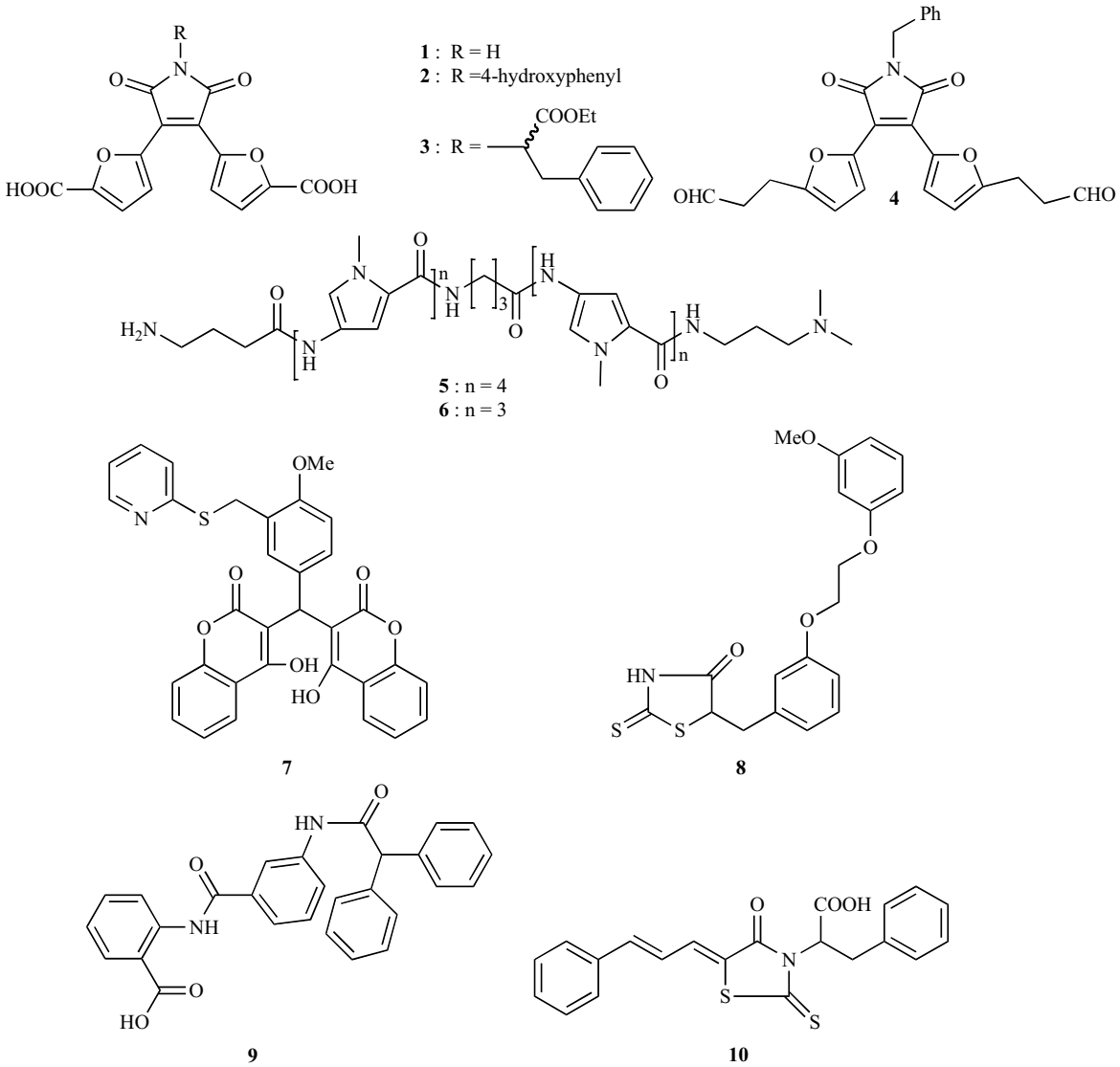


Fig. (1). Inhibitory activity of the 159 compounds. **(a)** Inhibition Factor of the 159 chemicals screened: each compound was tested at a concentration of 80 µM in a transposition assay. The inhibition factors (IFs), which correspond to the ratio between transposition frequency of the control reaction performed in the absence of the chemical, and the transposition frequency of the reaction in the presence of 80 µM of compound, are reported in the histogram. Chemicals previously identified as potent Tn*5* transposase or HIV-1 integrase inhibitors are shown in black. **(b)** Chemical structures of the ten compounds with IF > 25.

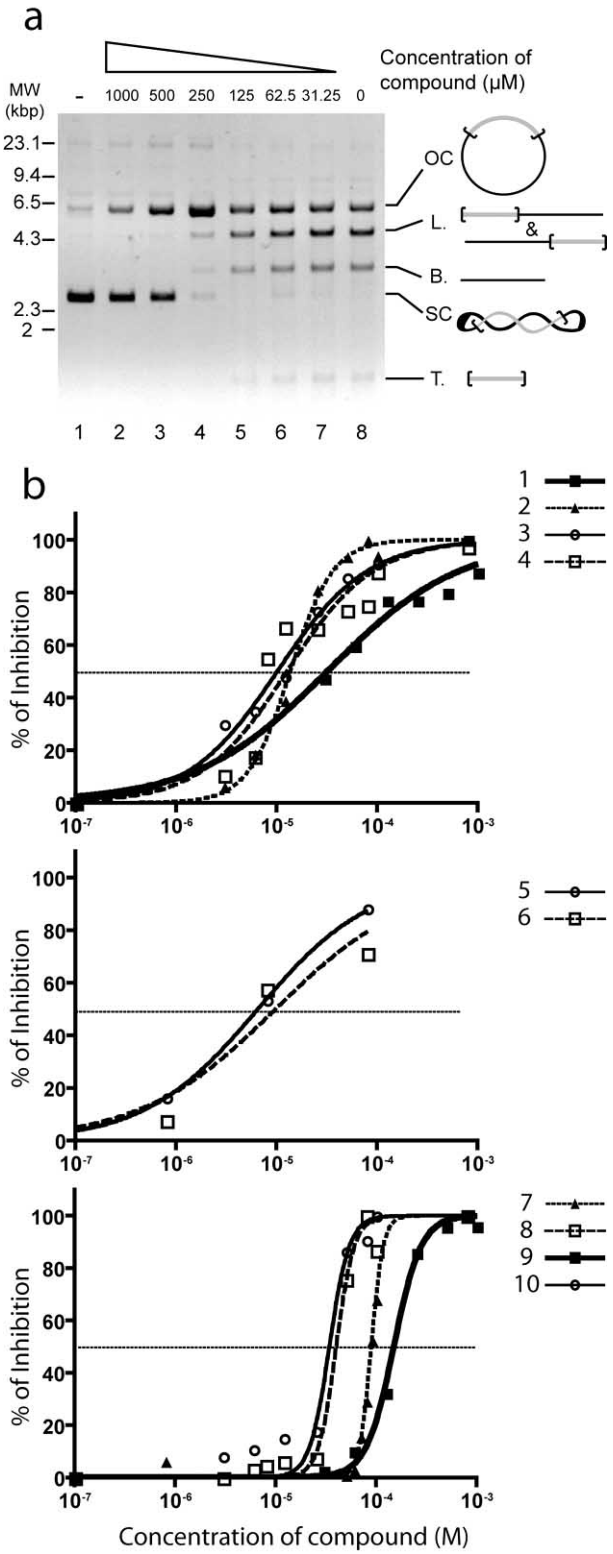


Fig. (2). Inhibition of transposon excision. **(a)** Inhibitory activity of compound (**9**) against transposon excision. The super-coiled pBC-3T3 plasmid (SC) is used as a transposon donor (lane 1). In the presence of transposase (lane 8), a first strand nicking reaction relaxed the donor plasmid (OC product). Second strand cleavage linearized the donor plasmid (L. product). Cleavage at both ITRs led to the excision of the transposon, releasing the transposon (T. product) and the plasmid backbone (B. product). The intermediate and final products of the reaction were visualized by electrophoresis on agarose gel and ethidium bromide staining. The reaction was performed in the absence of transposase (lane 1), without the inhibitor (lane 8) or with decreasing concentrations of compound (**9**) (lanes 2 to 7). The released backbone was quantified as an end product of the transposon excision reaction. **(b)** Inhibitory activity of the compounds. The activity of the ten compounds was plotted against compound concentration, and fitted on a sigmoid dose-response curve using GraphPad Prism software. First panel; compounds: (**1**) (black square), (**2**) (black triangle), (**3**) (open circle) and (**4**) (open square). Second panel; compounds: (**5**) (open circle) and (**6**) (open square). Third panel; compounds: (**7**) (dark triangle), (**8**) (open square), (**9**) (black square) and (**10**) (open circle).

Table 1. Effect of Each Inhibitor on *Mariner Mos1* Excision. IC50 Values for Excision (Exc.) are Reported in µM

Compound	IC50 Exc. (µM)	95% Confidence Intervals
1	33	18 < IC50<54
2	14	12 < IC50<15
3	10	7 < IC50<13
4	12	9 < IC50<18
5	6	5 < IC50<12
6	10	4 < IC50<22
7	89	86 < IC50<91
8	40	36< IC50<44
9	152	138 < IC50<166
10	34	29 < IC50<39

3'ITR (ITR70) in the presence of a concentration of drug equivalent to five-times the IC50 excision value, calculated for each compound. Our data (Fig. (**3a**)) indicated that the assembly of these complexes was almost totally prevented in the presence of compounds (**5**) to (**9**), and considerably inhibited (50% inhibition) in the presence of compound (**10**). The data observed here for compounds (**5**) and (**6**) are consistent with what is known about the mechanism of action of *N*-methylpyrrole polyamides [22]. These compounds are able to interact with DNA, and the slight modification in the migration pattern of the free probe (ITR70*, see Fig. (**3a**)) in the presence of these two compounds may account for this effect. Thus, compounds (**5**) and (**6**) targeted the DNA, and probably competed with MOS1 to bind to the ITR. Our data for compounds (**7**) to (**9**) are also consistent with the known mechanism of action of these drugs on the assembly of the Tn*5* transposase-ITR complex. Indeed, Ason *et al*. had previously shown (using EMSAs) that these compounds inhibit Tn*5* complex assembly, with IC50 values ranging from 5 to 45 µM [20]. Even though compound (**10**) efficiently inhibited Tn*5* transposase-ITR complex assembly, this compound seemed to have little impact on the assembly of MOS1-ITR complexes (see Fig. (**3a**)).

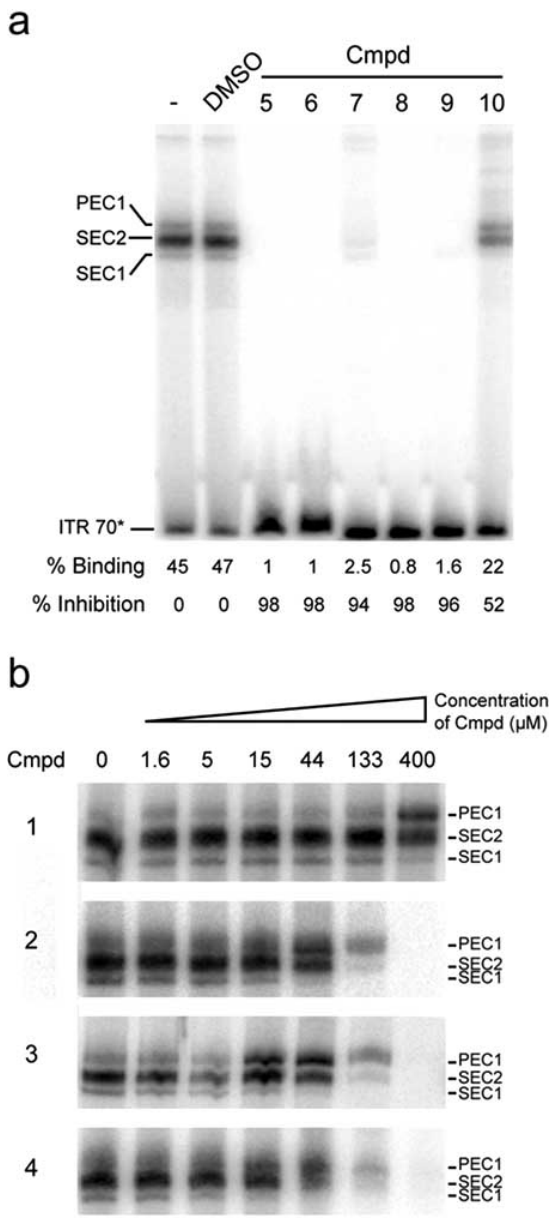


Fig. (3). Impact of the inhibitors on the assembly of MOS-ITR complexes. EMSAs were performed with 400 nM MBP-MOS1, 5 nM ITR70 as a probe and the 10 compounds (**1** to **10**) as indicated. **(a)** Compounds (**5**) to (**10**). (-): without compound; D: DMSO. The compounds are indicated at the top of the figure. They were used in concentrations corresponding to 5 times the IC50 reported in Table (**1**). Briefly, these concentrations were: 65 μM (compounds **(5)** and (**6**)), 440 μM (compound (**7**)), 200 μM (compound (**8**)), 750 μM (compound (**9**)) and 170 μM (compound (**10**)). The complexes observed and the unbound DNA (ITR70*) are indicated. SEC1 (Single end complex 1), SEC2 (Single end complex 2) and PEC1 (Paired end complex 1) were already described [8]. Percentage binding measured from the percentage of probe retained in the three complexes observed. Percentage inhibition measured as the loss of binding in the presence of the compounds. **(b)** Compounds (**1**) to (**4**). Accurate analyses of compounds (**1**) to (**4**)were performed using different concentrations of drugs ranging from 1.6 to 400 μM, as indicated at the top of the figure. The corresponding compounds are indicated on the left. Only complexes are shown and identified on the right (as in (**a**)).

As our findings for compounds (**5**) to (**10**) matched the published mechanism of action of these drugs, we focused our attention on the new drugs. Before doing any further investigations of the new transposase inhibitors (drugs (**1**) to (**4**)), we first checked to see whether these compounds are directed against MOS1, but not against the DNA (like compounds (**5**) and (**6**)). This was done either directly, testing the direct binding of the compounds to the ITR in electrophoretic mobility shift assay (EMSA) or indirectly using excision assays with large excess of competitor DNA (see supplemental data, Fig. (**S2**)). Our data showed that compounds (**1**) to (**4**) did not interact with the DNA.

We therefore assayed the ability of the new maleimide derivates to inhibit the formation of complexes between MOS1 and the 3'ITR. MOS1-ITR complexes were formed with a radiolabeled 3'ITR (ITR70) in the presence of drug concentrations ranging from 0 to 400 μM, and analyzed using EMSA (Fig. (**3b**)). Our data therefore indicated that compound (**1**) modified the ratio of MOS1-ITR complexes at a high drug concentration (400 μM, lane 7), leading to enrichment in PEC1 (paired end complex 1) to the detriment of SEC2 (single end complex 2). The other compounds acted similarly, modifying the ratio of MOS1-ITR complexes, but for drug concentrations ranging from 15 μM (compound (**3**), lane 4) to 44 μM (compounds (**2**) and (**4**), lane 5). At higher concentrations, compounds (**2**) to (**4**) were able not only to modify the ratio, but also to prevent the assembly of MOS1-ITR complexes (lanes 6 and 7). These data paralleled the IC50 excision values, which indicated that compounds (**2**), (**3**) and (**4**) were more efficient inhibitors of *Mos1* excision than compound (**1**).

As we think that MOS1 dimers are formed prior to ITR binding [23], we assayed drugs (**1**) to (**4**) to find out whether they were able to inhibit the formation of such dimers. Using a truncated form of MOS1 (Tnp[1-85]) that contains the MOS1 dimerization sub-domain, we measured MOS1 dimerization in the presence of compounds (**1**) to (**4**) by means of glutaraldehyde fixation experiments, as described [23]. None of the compounds inhibited the formation of Tnp[1-85] dimer at the 500-μM concentration (data not shown), which ruled out the possibility that these compounds could inhibit complex formation by destabilizing the N-terminal dimerization domain.

Taken together, these data indicated that these new inhibitors interacted with MOS1. This interaction could modify the organization of MOS1-ITR complexes, thus inhibiting *Mos1* excision, a subsequent step in the transposition reaction.

Effect of the New Inhibitors on Strand Transfer Reactions

Since excision and the strand transfer reaction are closely related reactions, and since strand transfer is a later step in the transposition reaction, we tested the new compounds to see whether they also inhibited the integration. Pre-cleaved ITRs were labeled on the 5' of the transfer strand, and used to form the integration complex. The inhibitor was added before the MOS1-ITR complexes had formed. ITR integration events in a target plasmid were detected after resolving

the reaction products on an agarose gel, as exemplified for compound (**1**), in Fig. (**4a**). One product, which co-migrated with linear forms of the target plasmid, was detected in the presence of both MOS1 and the catalytic cofactor Mg^{2+} (lane 7). The percentage inhibition produced by the new compounds was determined in four independent experiments, and plotted against the drug concentration (Fig. (**4b**)). Compounds (**1**), (**2**) and (**3**) inhibited the ITR integration, with IC50 values similar to those for excision (Table (**2**)), whereas compound (**4**) inhibited the strand transfer reaction 15-times more efficiently than excision. This indicated that compound (**4**) had some additional activity (when compared to compounds (**1**) to (**3**)) that might consist either of the inhibition of target capture or the direct inhibition of strand transfer catalysis, the two main steps of integration.

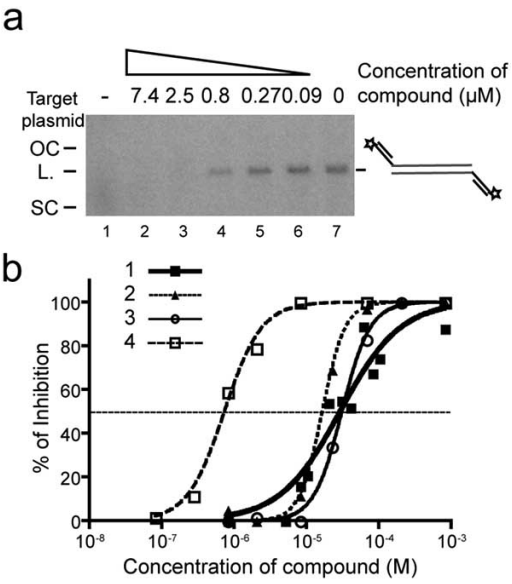


Fig. (4). Impact of the inhibitors on MOS1 strand transfer reactions. **(a)** *In-vitro* strand transfer inhibition activity of compound (**4**). Reactions were performed with pre-cleaved ITR substrate (radioactively labeled on the 5' extremity of the transferred strand). The pBC-KS vector was used as a super coiled plasmid target (SC). Reactions were stopped with EDTA, heated for 10 min at 75°C to disrupt the integration complex, and electrophoresed on a standard TBE-buffered 1% agarose gel. Concerted integration of transposon ends (two ITR) at the target site yielded a linear product (L.) as expected, whereas hemi-integration of transposon end (a single ITR) would have given a relaxed target plasmid (OC). After electrophoresis, the gel was stained for total DNA with ethidium bromide, dried and the radioactive products were recorded on a PhosphorImager. Lane 1; no transposase, lanes 2 to 6, decreasing concentrations of compound (**4**), lane 7; no inhibitor. **(b)** Inhibitory activity of compounds (**1**) to (**4**). The activity of compounds (**1**) to (**4**) (as measured in (**a**)) were plotted against the compound concentration and fitted on a sigmoid dose-response curve using GraphPad Prism software. Compounds: (**1**) (black square), (**2**) (black triangle), (**3**) (open circle) and (**4**) (open square).

Effect of the Inhibitors on HIV-1 Integrase 3' Processing and Strand Transfer Activities

We showed that Tn*5* and HIV-1 integrase inhibitors could also inhibit MOS1. To find out whether the new group of inhibitors were also able to block HIV-1 integrase, their activity was tested against the 3' processing and strand transfer activities of HIV-1 integrase. The assay was performed as previously described [24] using purified recombinant integrase and short radiolabeled oligonucleotide substrates, and the products were resolved on a denaturing urea gel. A typical assay, monitored using compound (**4**), is shown in (Fig. (**5a**)). Similar experiments were done using compounds (**1**) to (**3**), and the inhibition percentages found in three independent experiments were plotted against the drug concentrations (Fig. (**5b**) and (**5c**)). IC50 values for 3' processing, and strand transfer activities were calculated, and are reported in Table (**3**). Compounds (**2**) to (**4**) were efficient against both strand transfer and 3' processing. In contrast to compounds (**2**) and (**3**), compound (**4**) was significantly more efficient (5-times) against strand transfer activity (with an IC50 value of 5.5 μM) than against 3' processing (with an IC50 value of 29 μM). This data looks similar to what we observed using compound (**4**) against MOS1: integration was inhibited 15-times more efficiently than excision (see Tables (**1**) and (**2**)). Finally, compound (**1**) was the least effective against both HIV-1 integrase 3' processing and strand transfer activities. This effect, which had already been observed in MOS1 inhibition studies, was markedly enhanced for HIV-1 integrase.

Table 2. Effect of the New Inhibitors on *Mariner Mos1* Integration. IC50 Values for ITR Integration (Int.) are Reported in μM

Compound	IC50 Int. (μM)	95% Confidence Intervals
1	30	19 < IC50<45
2	16	15 < IC50<18
3	30	27 < IC50<34
4	0.7	0.6 < IC50<0.9

In conclusion, the bis(furanyl)*N*-maleimide derivatives that we have described as MOS1 inhibitors also have similar effects against HIV-1 integrase, demonstrating that this new group of inhibitors have cross activity against *Mos1* transposase and HIV1-integrase. As a consequence, MOS1 can be considered to be a new surrogate model for identifying HIV-1 integrase inhibitors.

CONCLUSIONS

The availability of specific inhibitors is very important for the analysis and control of elaborate mechanisms, such as transposition. For the first time, we have characterized chemicals that inhibit the transposition of *mariner*, the most widespread transposable element (TE) in eukaryotes. We used EMSA, excision and strand transfer assays to study in greater detail the mechanism of action of the ten most effective inhibitors identified in a panel of 159 compounds.

Tn5 and HIV-1 Integrase Inhibitors are Active Against MOS1

A first group of compounds consisted of compounds that had previously been identified as inhibitors of transposase

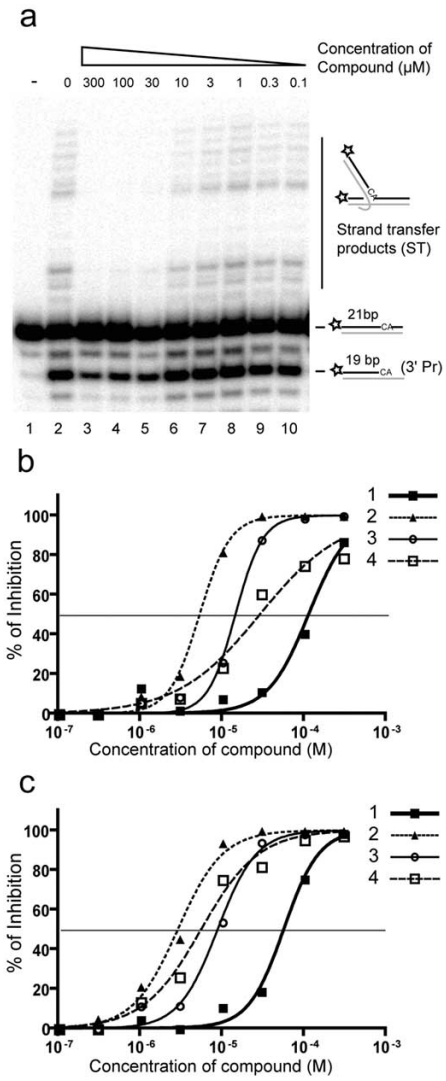


Fig. (5). Inhibitory activity of compound (4) on HIV-1 integrase 3' processing and strand transfer activities. **(a)** HIV-1 integrase inhibitory activity of compound (**4**): Assays were performed using purified recombinant HIV-1 integrase and a duplex oligonucleotide mimicking the U5 LTR extremity radioactively labeled on the 5' extremity of the processed strand. The reactions were stopped with EDTA, and electrophoresed on 18% acrylamide, 7M urea denaturing gel. 3' processing reaction yields a 19-mer product that is a substrate for the strand transfer reaction. Lane 1; no integrase, Lane 2; no inhibitor, lanes 3 to 10; decreasing concentration of compound (**4**). **(b)** 3' processing inhibitory activity of the compounds. The activity of the four compounds (**1** to **4**), from three independent experiments, was plotted against compound concentration, and fitted on a sigmoid dose-response curve using GraphPad Prism software. Compounds: (**1**) (black square), (**2**) (black triangle), (**3**) (open circle) and (**4**) (open square). **(c)** Strand transfer inhibitory activity of the compounds. The activity of the four compounds (**1** to **4**), from three independent experiments, was plotted against compound concentration, and fitted on a sigmoid dose-response curve using GraphPad Prism software. Compounds: (**1**) (black square), (**2**) (black triangle), (**3**) (open circle) and (**4**) (open square).

(Tn*5*) or integrase (HIV-1). Drugs 7 to 10 displayed a similar mechanism of action *versus* both MOS1 and Tn*5* transposase. They inhibited the formation of the complexes, probably by interfering with one of the DNA-recognition domains of the transposase. The Tn5 inhibitors were less effective against *Mos1* transposition, which may reflect differences between the mechanisms of DNA recognition and complex organization.

The target of the drugs allows us to distinguish between compounds (**5**) and (**6**) on the one hand, which bind DNA, and on the other, the other compounds, which are devoid of DNA-binding ability. Compounds (**5**) and (**6**) are hairpin polyamides, composed of a *N*-methylpyrrole polyamide connected by a γ-aminobutyric acid (γ-turn) linker, and bound to the minor groove of DNA [22]. They specifically interact with a short sequence of three or four A/T pairs for compound (**5**) and (**6**) respectively. These sequences are located at the tip, and also in the ITR substrate, which accounts for the effect of compounds (**5**) and (**6**) on *Mos1* transposition.

bis-(heteroaryl)maleimide Derivatives Constitute a New Scaffold of DDE-Enzyme Inhibitors

The panel of drugs that we tested as MOS1 inhibitors consisted mainly of hydroxyl-substituted heterocyclic molecules. Small molecules of this kind had already been identified as efficient inhibitors of HIV-1 integrase [25]. Nineteen of the compounds analyzed here were bis-(heteroaryl)maleimide derivatives. The most effective inhibitors we found were all organized around this scaffold (compounds (**1**) to (**4**)). However, scaffolds with no substituents were inefficient, suggesting that the bis-(heteroaryl)maleimide backbone itself is not the pharmacophore (data not shown). Maleimide *N*-substitution and heteroaryl substitution both had a major impact on both the mechanism of action and efficiency of the drugs. Here, we describe three examples. (i) Compound (**1**), which is not *N*-substituted, does not inhibit MOS1-ITR complex assembly, although differences in complexes ratio suggest that compound (**1**) may change the structure of the complexes, favouring PEC1 assembly. This compound has a similar effect against both the excision and strand transfer activities, and has slightly higher IC50 values than compounds (**2**), (**3**) and (**4**). (ii) Compounds (**2**) and (**3**), which are *N*-substituted, inhibit MOS1-ITR complex assembly more efficiently than compound (**1**). These compounds have a similar effect against both the excision and strand transfer activities. (iii) Compound (**4**), which is also *N*-substituted, but contains a formyl group (instead of the acid functions in compounds (**1**) to (**3**)) has similar activity against MOS1-ITR complex assembly and excision as compounds (**2**) and (**3**). Unlike compounds (**1**) to (**3**), compound (**4**) is 15-times more active against strand transfer activity than against *Mos1* excision.

Active bis-(heteroaryl)maleimide derivatives do not bind to DNA (this study and [26]), but they probably target the transposase or transposase-DNA complexes, with greater or lesser efficiency. There are (at least) two possible explanations for the differing efficiencies displayed by compound (**1**) and the other maleimide derivatives. First, the drug targets may differ as a result of the aromatic group substitution on the *N*-maleimide (compounds (**2**) to (**4**)). Second, all four compounds could have a similar target, but the bulky aromatic *N*-maleimide substitution may interfere with DNA

Table 3. Effect of the New Inhibitors on HIV-1 Integration. IC50 Values for 3' Processing (Proc.) and Strand Transfer (StT.) are Reported in μM

Compound	IC50 Proc. (μM)		IC50 StT. (μM)	
1	116	59 < IC50<228	58	35 < IC50<94
2	5.3	4.6 < IC50<6.1	3	2.3 < IC50<3.6
3	15	10 < IC50<19	9	6 < IC50<13
4	29	14 < IC50<60	5.5	4 < IC50<7.7

binding, whereas the non-substituted compound tolerates the presence of the DNA. Further studies are required to identify the exact target(s) and mechanism of action of this family of compounds.

Similarly, three situations were observed in the case of HIV-1 integrase inhibition. (i) Compound (**1**), which is not *N*-substituted, had higher IC50 values than compounds (**2**), (**3**) and (**4**). In addition, this compound yielded more variable data, with a wide 95% confidence interval: for instance the IC50 for 3' processing was (59-228) μM. This point was also confirmed (albeit to a lesser extent) against MOS1. This raises the question of the solubility of non-*N*-substituted compounds. (ii) Compounds (**2**) and (**3**), which are *N*-substituted, display similar activity against both the excision and the strand transfer activities, with very good 95% confidence intervals. (iii) Compound (**4**), which is also *N*-substituted but contains a formyl group, is 5 times more active against strand transfer activity than against 3' processing activity.

The effect of compound (**4**) on the final step in *Mos1* transposition and/or HIV-1 integration is very interesting. Indeed, the data presented here suggest that compound (**4**) prevents the integration step in both systems. During both transposition and retroviral integration, this step involves at least two phases: (i) capturing the DNA targeted for integration and (ii) the strand transfer reaction itself. Compound (**4**) could be the first chemical for studying this final step.

As a target, the DDD catalytic pocket is an excellent candidate for explaining the change in the mechanism of inhibition (drugs (**1**) to (**3**) *versus* (**4**)) and cross-activity of these drugs against HIV-1 integrase, because it carries the catalytic residues and also lies at the intersection of the two DNA substrates involved in the reaction. Compounds with a scaffold that is able to bind to this region might be capable of displaying several different inhibitory phenotypes, becoming specific inhibitors of complex formation, excision or strand transfer as a result of minor differences in their scaffold substitution.

On the other hand, differences in drug targets might also account for the different phenotypes of this family of compounds. Point mutations in the N-terminal domain of HIMAR1 (a MOS1 related *mariner* transposase) illustrate how changes in transposase-ITR interactions can have pleiotropic consequences. Some mutations completely destabilize the transposase-ITR complexes, whereas others allow the complexes to form, but inhibit the first or the second strand cleavage reaction [27]. This implies that compounds that target the N-terminal/ITR interaction could mirror the phenotype of N-terminal domain mutations. Identification of the target(s) of this family of compounds is already in progress, because it appears to be essential for further development, and for understanding the mechanism of action of these inhibitors in *Mos1* transposition and in the HIV-1 integration process.

Finally, we noted that bis-(heteroaryl)maleimide derivatives are new inhibitors of DDE-transposases, especially of *Mos1* transposase and HIV-1 integrase. Their identification may provide potent tools for studying the mechanism of *Mos1* transposition and/or for proposing new drugs that target HIV-1 integrase.

Cross Activity of DDE Enzyme Inhibitors

The ability of a drug to inhibit mechanistically-related enzymes (cross-activity) had already been illustrated by compounds that inhibit HIV-1 integrase, ASV and other related retroviral integrases [28]. Furthermore, diketo acid integrase inhibitors (5 CITEP and L-708,906) are also active against the distantly related RAG recombinases [29], although the activity of the drug is much lower (100 fold). More recently, screening for Tn*5* transposase inhibitors has been used to identify compounds that are active against HIV-1 integrase [20].

The work reported here constitutes a novel demonstration of cross-activity between Tn*5* and *Mos1* transposase inhibitors, and also, more interestingly, between MOS1 and HIV-1 integrase. The cross-activity of the new family of inhibitors we have identified indicates that MOS1 would provide an excellent surrogate model for an inhibitor screen. This surrogate model is very interesting for the following reasons. First, MOS1 transposition occurs in a eukaryotic environment, like the integrase, which makes it possible to envisage a test for inhibition that can be carried out in eukaryotic cells. Second, the structure of the catalytic domain of MOS1 is very similar to that of the retroviral RSV and HIV-1 integrases [9,10]. Third, the catalytic process of DNA cleavage looks similar to the processing reaction of retroviral integrase, the second strand cleavage reaction mimicking 3'-processing integrase activity (there is no inter-strand reaction implicating a hairpin intermediate in either of these reactions [19]). So far, the Tn*5* transpososome has been used as a model for the retroviral integrasome, but the *mariner* transpososome could constitute a new and probably better model of HIV-1 integrasome organization. Nevertheless, we should also note that these two models (*Mos1 versus* HIV-1) display

significant differences, such as the structure and arrangement of their DNA binding domains, the nature of the donor DNA substrate, and the DNA cleavage reaction (double strand versus single strand cleavage reaction) that mean that the two systems are not interchangeable. From this perspective, the use of cross-active drugs will provide an interesting tool for identifying similarities and differences between these two enzymes.

ABBREVIATIONS

IF = Inhibition factor

ITR = Inverted terminal repeat

LTR = Long terminal repeat

NTS = Non-transferred strand

RSV = Rous sarcoma virus

TS = Transferred strand

ACKNOWLEDGEMENTS

We would like to thank Dr Cécile Gueiffier (UFR Sciences Pharmaceutiques, Univ of Tours, Tours-FRANCE) who kindly gave us thirty-eight chemicals, Dr Marie-Line Andréola (UMR 5097 CNRS, Univ of Bordeaux 2, Bordeaux-FRANCE) who kindly gave us the G-quartets and Cédric Tissandier for his technical assistance. Dr M. Ghosh (Plougasnou, FRANCE) has revised the English text. Nicolas Bouchet holds a doctoral fellowship from the Ministère de l'Education Nationale, de la Recherche et de la Technologie. This work was funded by the University of Tours, the Centre National de la Recherche Scientifique; the Ministère de l'Education Nationale, de la Recherche et de la Technologie, and GDR CNRS n°2157. This work was also funded by grants from the European Commission (Project SyntheGeneDelivery, grant n°018716 to Y.B.), and the Association Française contre les Myopathies (grant n°11468 to C.A.G.).

REFERENCES

[1] Craig, N.L.; Craigie, R.; Gellert, M.; Lambowitz, A.M. *Mobile DNA II*. ASM Press: Washington DC, **2002**.

[2] Duverger, Y.; Belougne, J.; Scaglione, S.; Brandli, D.; Beclin, C.; Ewbank, J.J. A semi-automated high-throughput approach to the generation of transposon insertion mutants in the nematode Caenorhabditis elegans. *Nucleic Acids Res.,* **2007**, *35*:e11.

[3] Keravala, A.; Liu, D.; Lechman, E.R.; Wolfe, D.; Nash, J.A.; Lampe, D.J.; Robbins, P.D. Hyperactive Himar1 transposase mediates transposition in cell culture and enhances gene expression *in vivo*. *Hum. Gene Ther.,* **2006**, *17*, 1006-18.

[4] Kapetanaki, M.G.; Loukeris, T.G.; Livadaras, I.; Savakis, C. High frequencies of Minos transposon mobilization are obtained in insects by using *in vitro* synthesized mRNA as a source of transposase. *Nucleic Acids Res.,* **2002**, *30*, 3333-40.

[5] Kuwayama, H.; Yaginuma, T.; Yamashita, O.; Niimi, T. Germ-line transformation and RNAi of the ladybird beetle, Harmonia axyridis. *Insect Mol. Biol.,* **2006**, *15*, 507-12.

[6] Yant, S.R.; Huang, Y.; Akache, B.; Kay, M.A. Site-directed transposon integration in human cells. *Nucleic Acids Res.,* **2007**, *35*(7), e50.

[7] Koga, A.; Hori, H. The Tol2 transposable element of the medaka fish: an active DNA-based element naturally occurring in a vertebrate genome. *Genes Genet. Syst.,* **2001**, *76*, 1-8.

[8] Augé-Gouillou, C.; Brillet, B.; Hamelin, M.H.; Bigot, Y. Assembly of the mariner Mos1 synaptic complex. *Mol. Cell Biol.,* **2005**, *25*, 2861-70.

[9] Richardson, J.M.; Finnegan, D.J.; Walkinshaw, M.D. Crystallization of a Mos1 transposase-inverted-repeat DNA complex: biochemical and preliminary crystallographic analyses. *Acta. Crystallograph Sect. F. Struct. Biol. Cryst. Commun.* **2007**, *63,* 434-37.

[10] Richardson, J.M.; Dawson, A.; O'Hagan, N.; Taylor, P.; Finnegan, D.J.; Walkinshaw, M.D. Mechanism of Mos1 transposition: insights from structural analysis. *EMBO J.,* **2006**, *25*, 1324-34.

[11] Curcio, M.J.; Derbyshire. K.M. The outs and ins of transposition: from mu to kangaroo. *Nat. Rev. Mol. Cell Biol.,* **2003**, *4*, 865-77.

[12] Rice, P.A.; Baker, T.A. Comparative architecture of transposase and integrase complexes. *Nat. Struct. Biol.,* **2001**, *8*, 302-07.

[13] Dyda, F.; Hickman, A.B.; Jenkins, T.M.; Engelman, A.; Craigie, R.; Davies, D.R. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science,* **1994**, *266*, 1981-86.

[14] Davies, D.R.; Mahnke-Braam, L.; Reznikoff, W.S.; Rayment, I. The three-dimensional structure of a Tn5 transposase-related protein determined to 2.9-A resolution. *J. Biol. Chem.,* **1999**, *274,* 11904-13.

[15] Mizuuchi, K. Mechanism of transposition of bacteriophage Mu: polarity of the strand transfer reaction at the initiation of transposition. *Cell,* **1984**, *39*, 395-404.

[16] Duval-Valentin, G.; Marty-Cointin, B.; Chandler, M. Requirement of IS911 replication before integration defines a new bacterial transposition pathway. *EMBO J.,* **2004**, *23*, 3897-3906.

[17] Bhasin, A.; Goryshin, I.Y.; Reznikoff, W.S. Hairpin formation in Tn5 transposition. *J. Biol. Chem.,* **1999**, *274,* 37021-29.

[18] Kennedy, A.K.; Guhathakurta, A.; Kleckner, N.; Haniford, D.B. Tn10 transposition *via* a DNA hairpin intermediate. *Cell,* **1998**, *95,* 125-34.

[19] Dawson, A.; Finnegan, D.J. Excision of the Drosophila mariner transposon Mos1. Comparison with bacterial transposition and V(D)J recombination. *Mol. Cell,* **2003**, *11,* 225-35.

[20] Ason, B.; Knauss, D.J.; Balke, A.M.; Merkel, G.; Skalka, A.M.; Reznikoff, W.S. Targeting Tn5 transposase identifies human immunodeficiency virus type 1 inhibitors. *Antimicrob. Agents Chemother.,* **2005**, *49*, 2035-43.

[21] Barreca, M.L.; De Luca, L.; Iraci, N.; Chimirri, A. Binding mode prediction of strand transfer HIV-1 integrase inhibitors using Tn5 transposase as a plausible surrogate model for HIV-1 integrase. *J. Med. Chem.,* **2006**, *49*, 3994-97.

[22] Dervan, P.B.; Edelson, B.S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.,* **2003**, *13*, 284-99.

[23] Augé-Gouillou, C.; Brillet, B.; Germon, S.; Hamelin, M.H.; Bigot. Y. Mariner Mos1 transposase dimerizes prior to ITR binding. *J. Mol. Biol.,* **2005**, *351*, 117-30.

[24] Craigie, R.; Mizuuchi, K.; Bushman, F.D.; Engelman, A. A rapid *in vitro* assay for HIV DNA integration. *Nucleic Acids Res.,* **1991**, *19*, 2729-34.

[25] Dayam, R.; Neamati, N. Small-molecule HIV-1 integrase inhibitors: the 2001-2002 update. *Curr. Pharm. Des.,* **2003**, *9*, 1789-802.

[26] Dubernet, M. Synthèse d'hétérocycles azotés et oxygénés comme inhibiteurs de la chymase cardiaque humaine et à visée antidiabétique. *PhD thesis,* University François Rabelais of Tours, **2004**.

[27] Butler, M.G.; Chakraborty, S.A.; Lampe, D.J. The N-terminus of Himar1 mariner transposase mediates multiple activities during transposition. *Genetica,* **2006**, *127,* 351-66.

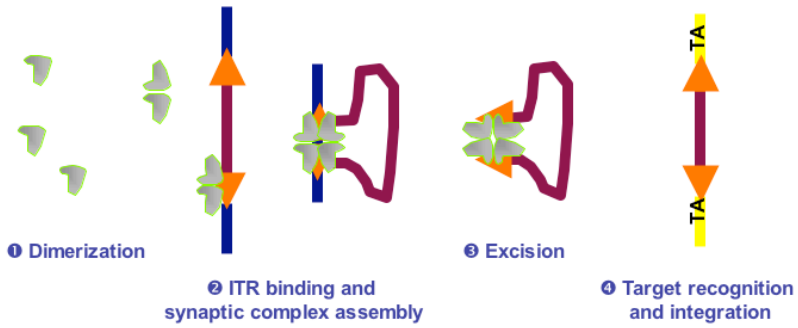
[28] Lubkowski, J.; Yang, F.; Alexandratos, J.; Wlodawer, A.; Zhao, H.; Burke, T.R.Jr.; Neamati, N.; Pommier, Y.; Merkel, G.; Skalka, A.M. Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proc. Natl. Acad. Sci. USA,* **1998**, *95*, 4831-36.

[29] Melek, M.; Jones, J.M.; O'Dea, M.H.; Pais, G.; Burke, T.R.Jr.; Pommier, Y.; Neamati, N.; Gellert, M. Effect of HIV integrase inhibitors on the RAG1/2 recombinase. *Proc. Natl. Acad. Sci. USA,* **2002**, *99*, 134-37.

Received: 07 November, 2008 Revised: 28 December, 2008 Accepted: 09 January, 2009

Bouchet et al, Supplementary data

Fig. S1. Schema for *Mos1* Transposition



According to the model we have previously proposed (Augé-Gouillou *et al*, J. Mol. Biol., 2005), *Mos1* transposition consists of four main steps: (**1**) dimerization of MOS1 (green forms) and (**2**) binding of one MOS1 dimer at each ITR (orange arrows) of the element (brown line), thus forming SEC2. Once bound, the dimers interact with each other, to form a paired ends complex (PEC2). (**3**) DNA strands are then cleaved between the end of the ITR and the flanking DNA. (**4**) Once the PEC2 has been excised, it can be inserted into a TA target dinucleotide by a trans-esterification reaction mediated by MOS1 tetramer.

Table S1. Effect of each inhibitor on *mariner Mos1* transposition

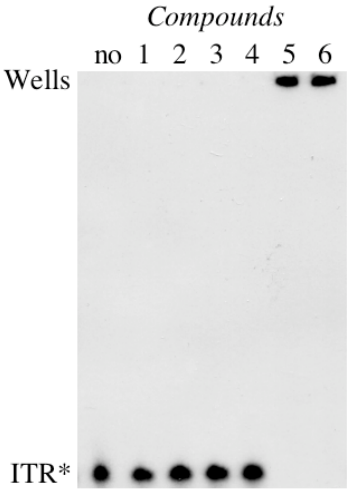
Compounds [80 µM]	Transposition frequency	Inhibition Factor
DMSO	$4x10^{-4}$	1
1	$1.09x10^{-6}$	367
2	$1.22x10^{-5}$	33
3	$4.01x10^{-6}$	100
4	$5.37x10^{-6}$	75
5	$8.38x10^{-6}$	48
6	$1.01x10^{-5}$	40
7	$4.85x10^{-6}$	82.5
8	$5.04x10^{-6}$	79
9	$1.64x10^{-5}$	25
10	$2.76x10^{-6}$	145

These results were obtained according to the procedure described in the main text as " *in vitro* transposition assays " (see materials and methods section). They are the mean for at least five independent assays.

It must be highlight that compare, and classify these inhibition factors (that are only a mean to rapidly evaluate several chemicals) with the IC50 values presented in the main text has no meaningful. Indeed, the transposition assays are not the sum of the independent biochemical assays performed to analyze each step of transposition (dimerization – formation of ITR/Tpase complexes – excision and strand transfer). In addition, *in vitro* transpositions were detected using promoter tagging (in the CAT gene), a method well known to under estimate the resulting transposition frequencies. From this standpoint, direct IC50 comparison, within each independent assay, is the most rigorous mean to analyze the data.

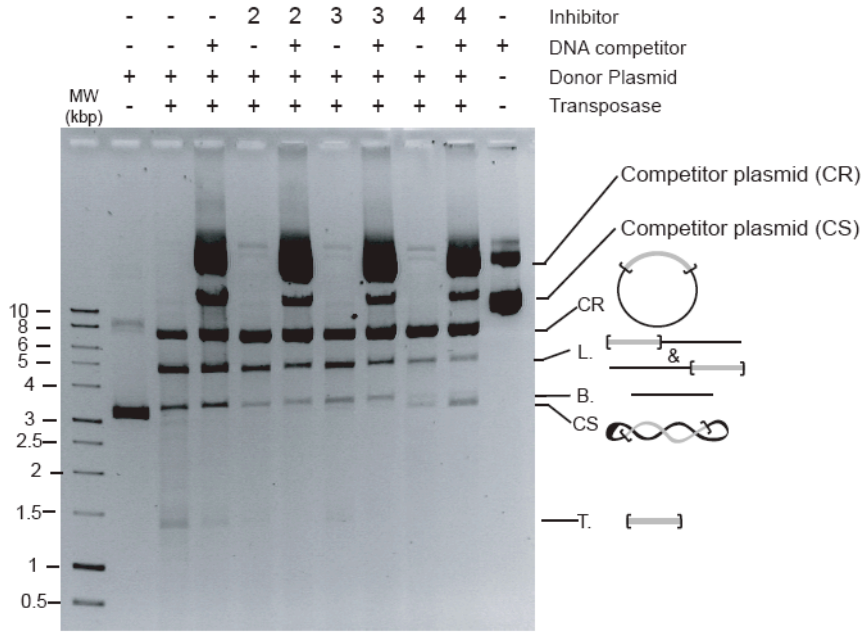
Fig. S2. Compounds 1 to 4 do not interact with the DNA

A. Direct demonstration: EMSA were performed using DMSO or large excess (400µM) of compounds 1 to 6, and 10 nM of ITR as a probe. No Tpase was added in the assay. Wells are indicated.



The data clearly shown that compounds 5 and 6 interact with the DNA, trapping the ITR in the well. This was not the case for compounds 1 to 4. Another line of evidence against the interaction of compounds 1 to 4 and the DNA is given in part B (see below).

B. Indirect demonstration: Excision assays were performed as described in Fig.2 (main text) using a 5-fold excess of DNA competitor as indicated (+), and in the presence of 40 µM of compounds 2 to 4. The super-coiled pBC-3T3 plasmid (CS) is used as a donor of transposon. In the presence of transposase, first strand nicking reaction relaxed the donor plasmid (CR product). Second strand cleavage linearized the donor plasmid (L. product). Cleavage at both ITR leads to the excision of the transposon releasing the transposon (T. product) and the plasmid backbone (B. product). The intermediate and final products of the reaction are visualized by electrophoresis on agarose gel and ethidium bromide staining. The released backbone is quantified as an end product of transposon excision reaction. The percentage of inhibition was measured as the decrease of the released backbone (B.) in the presence of compound when compared to the assay without compound.



As reported below the picture, the presence of DNA competitor does not modify the inhibitory effect (% of inhibition) of the bis-(heteroaryl)maleimide derivatives, thus indicating that these compounds do not target the DNA. The same data was obtained for compound 1 (not shown).

Mariner Mos1 optimization by rational mutagenesis

Germon S, Bouchet N, Casteret S, Carpentier G, Adet J, Bigot Y, Augé-Gouillou C,

Genetica, 2009, publié en ligne

Genetica
DOI 10.1007/s10709-009-9375-x

Mariner Mos1 transposase optimization by rational mutagenesis

Stéphanie Germon · Nicolas Bouchet · Sophie Casteret · Guillaume Carpentier · Jérémy Adet · Yves Bigot · Corinne Augé-Gouillou

Received: 5 February 2009 / Accepted: 25 May 2009
© Springer Science+Business Media B.V. 2009

Abstract *Mariner* transposons are probably the most widespread transposable element family in animal genomes. To date, they are believed not to require species-specific host factors for transposition. Despite this, *Mos1*, one of the most-studied *mariner* elements (with *Himar1*), has been shown to be active in insects, but inactive in mammalian genomes. To circumvent this problem, one strategy consists of both enhancing the activity of the *Mos1* transposase (MOS1), and making it insensitive to activity-altering post-translational modifications. Here, we report rational mutagenesis studies performed to obtain hyperactive and non-phosphorylable MOS1 variants. Transposition assays in bacteria have made it possible to isolate numerous hyperactive MOS1 variants. The best mutant combinations, named FETY and FET, are 60- and 800-fold more active than the wild-type MOS1 version, respectively. However, there are serious difficulties in using them, notably because they display severe cytotoxicity. On the other hand, three positions lying within the HTH motif, T88, S99, and S104 were found to be sensitive to phosphorylation. Our efforts to obtain active non-phosphorylable mutants at S99 and S104 positions were unsuccessful, as these residues, like the co-linear amino acids in their close vicinity, are critical for MOS1 activity. Even if host factors are not essential for transposition, our data demonstrate that the host machinery is essential in regulating MOS1 activity.

Keywords *Mos1* element · *Mariner* transposases · Hyperactive transposases · Phosphorylation · Rational mutagenesis · Transposon tools

Abbreviations

HTH	Helix-turn-helix
ITR	Inverted terminal repeat
MBP	Maltose binding protein
NHEJ	Non-homologous end joining

Introduction

Transposases are enzymes encoded by Class II transposable elements, and carry out most of the steps involved in transposon mobility, which occurs via a DNA intermediate that is precisely excised from a donor site and reinserted into an empty target site. In the IS*630-Tc1-mariner* family [*ITm*; (Shao and Tu 2001)], the transposases contain two domains: an N-terminal domain, which promotes binding to inverted terminal repeats (ITR, located at both ends of the transposon) and the assembly of the transposon-transposase complexes, and a C-terminal domain, which promotes catalytic functions (DNA strand cleavages and transfers). Regulation of transposition activity is essential to offset the adaptive advantages that transposable elements confer on their host genome, and at the same time to limit the damage

S. Germon · N. Bouchet · S. Casteret · G. Carpentier · J. Adet · C. Augé-Gouillou (✉)
GICC, Université François Rabelais de Tours, 37200 Tours, France
e-mail: auge@univ-tours.fr

N. Bouchet · S. Casteret · G. Carpentier · J. Adet · Y. Bigot · C. Augé-Gouillou
UMR 6239, CNRS, 37200 Tours, France

Y. Bigot
CHRU de Tours, UFR des Sciences & Techniques, Parc de Grandmont, 37200 Tours, France

S. Germon
Laboratoire d'Immunologie Parasitaire, Vaccinologie, UMR INRA 483, UFR de Pharmacie, 31 avenue Monge, 37200 Tours, France

Published online: 17 June 2009

they might cause by potentially lethal DNA rearrangements. Indeed, it has been suggested that "inactivity" could be an evolutionary strategy employed by many transposable elements (Reznikoff 2006). However, low transposition capacities make it difficult to use these elements and their transposase for genetic manipulations.

To get round this, a first strategy consists of obtaining hyperactive transposases. This has so far been successfully performed with several transposon systems, using approaches based on the screening of random point mutations, without taking the structure–function relationships of the proteins into account. The best-exemplified examples of this approach are those developed to improve the *Sleeping Beauty* (*SB*) transposase (SB10). They were achieved by screening for random point mutations occurring over the entire protein (Zayed et al. 2004), or concentrated within the N-terminal DNA-binding domain (Yant et al. 2004). An alternative approach is to screen mutant collections using transposition assays in bacteria. This was used to isolate hyperactive mutants of *Himar1* transposase with activities that have been increased 4- to 50-fold (Lampe et al. 1999). In this case, it is noticeable that the purified mutant transposases retain their hyperactivity in vitro, but to a lower extent. A third approach is to carry out rational mutagenesis. This has been done with SB10, on the one hand, using results based on phylogenetic-based analyses to target mutagenesis and, on the other, by optimizing the codon usage of the gene encoding the protein (Baus et al. 2005). Interestingly, whichever strategy is used, the results indicate that combining hyperactive mutations could result in synergistic enhancement of transposase activity, as well as a loss of activity.

In the past decade, a second strategy relying on directed mutagenesis has been devised to bypass the down-regulation mechanisms that limit the activity of certain transposases. This was undertaken with the Tn*5* transposase, and the stability of the transposase molecules was enhanced (Twining et al. 2001) by removing the proteolytic cleavage sites for *Escherichia coli* proteases by mutagenesis. However, it should be pointed out that the transposase mutants thus obtained were inactive both in vivo and in vitro, highlighting the difficulties of this approach. This issue has also been illustrated with the *P*-element, for which it has been proposed that the phosphorylation of several serine and threonine residues by the ATM kinase may be one of the pathways that regulates transposase activity. When these residues are replaced by alanine, *P*-transposase activity may be either increased or reduced (Beall et al. 2002).

In the past 5 years, several studies have demonstrated that the transposition of the *mariner Mos1* transposon is dramatically restricted in mammals, making it impossible to use this element to develop transposon based-vectors in human cells. In order to investigate this issue, we first developed rational methods to obtain a hyperactive version of MOS1. Next, taking into account the fact that the activity of MOS1 in human cells might be severely impaired by down-regulation mechanisms, we investigated possible rational designs for non-phosphorylable transposase mutants.

Results

Designing hyperactive MOS1

Based on data available in the literature on HIMAR1 and MOS1 (the *Himar1* and *Mos1* transposases), three rational approaches were developed to locate hyperactive single mutations in MOS1. The synergies and the antagonisms were then investigated in MOS1 prototypes combining these single hyperactive mutations.

Single MOS1 mutants containing mutations previously characterized in HIMAR1

Five single-point mutations allocating transposition hyperactivity to HIMAR1 have been reported in various studies (Lampe et al. 1999; Butler et al. 2006; Keravala et al. 2006) (Fig. 1). The first set of studies concerned three residues: Q131R, E137K and H267R. Sequence comparison revealed that these positions corresponded to residues R132, E137 and E264 in MOS1. The first residue is already an arginine (R) in MOS1. We first focused our attention on the E137K mutant, and using transposition assays in *E. coli* as previously described (Sinzelle et al. 2008) we found that it was 1.4-fold more active than the wild type MOS1 (MOS1[+]; Fig. 2a). Although this value is similar to that recovered by Pledger and Coates (2005), it does not revealed a significant difference to the wild type MOS1. The second set of studies consisted of two single-point hyperactive HIMAR1 mutants, E90A and E93A that corresponded to residues Q91 and E94, respectively in MOS1. Both residues are located in their respective transposase HTH motif, i.e., the ITR binding motif (Fig. 1). To verify the hyperactivity conferred by mutations at these residues in MOS1, we first used Q91R a mutant similar to Q91A, and previously isolated for other purposes (Zhang et al. 2001). We found that Q91R was ninefold more active in transposition than MOS1[+]. In contrast, the E94A MOS1 mutant was found to be inactive. Three other residues in the HIMAR HTH had previously been mutated, resulting either in no effect (I91A), or in proteins unable to carry out transposition (K99A and H104A). The corresponding MOS1 mutants, L92A and L92R at the first position, and Q100N, Q100E and Q100R at the second position, demonstrated no transposition activity, whereas the N105A

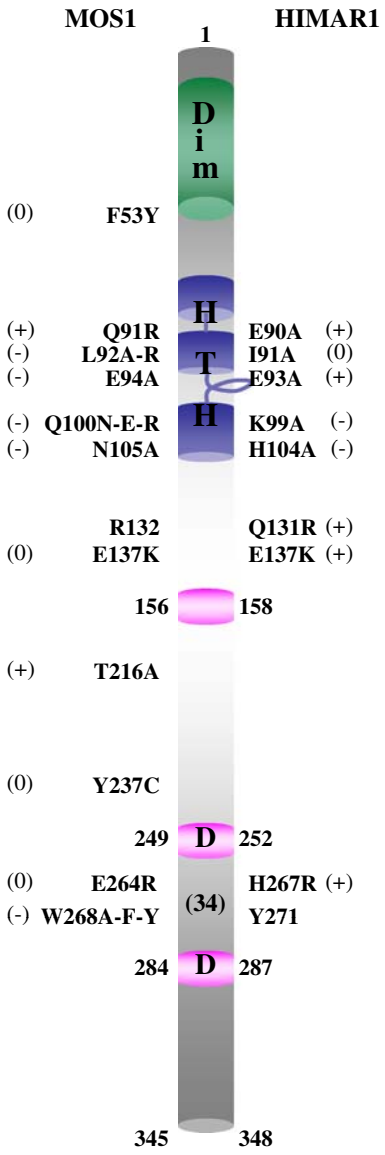


Fig. 1 *Himar1* and *Mos1* transposases. Functional domains are represented as follows: dimerization in *green* (Auge-Gouillou et al. 2005a, b; Butler et al. 2006), the HTH motif involved in ITR binding in *blue* (Brillet et al. 2007), and the catalytic domain (DD34D) in *pink* (Richardson et al. 2006). Single mutations used to obtain hyperactive MOS1 are indicated. The corresponding HIMAR1 positions are indicated. Mutant activity was described by Lampe et al. (1999), and Butler et al. (2006) for HIMAR1, Pledger and Coates (2005) and this study for MOS1

mutation at the third position had no impact on transposition efficiency.

Hyperactive single MOS1 mutants with altered protein–protein interactions

Richardson et al. (2004) has identified a single-point MOS1 mutant (T216A) from a yeast two-hybrid assay designed to identify transposase mutants that are able to interact more strongly with the wild-type protein. Recently, T216A was found to dimerize more in solution than MOS1[+] (Richardson et al. 2006). Since we had previously shown that MOS1 dimerization is required to initiate transposition (Auge-Gouillou et al. 2005a, b), we verified the transposition activity of the T216A MOS1 mutant, using transposition assays in *E. coli*. We found that it was 15-fold more active than MOS1[+] (Fig. 2a). This suggested that other MOS1 mutants affected at the level of their MOS1–MOS1 interactions might also be hyperactive. This was also supported by the fact that the Q91R MOS1 mutant, that we found here to be hyperactive, had also previously been shown to be a MOS1–MOS1 interaction mutant (Zhang et al. 2001). We therefore took advantage of 11 other MOS1 mutants displaying altered MOS1–MOS1 interactions, as described (Zhang et al. 2001): F53Y, L92H, A93S, S104P, V120L, L124S, W193R, I160T, Y237C, D279G and S302P. Two of them, F53Y and Y237C, revealed transposition efficiencies that were not higher but similar to that of MOS1[+] (Fig. 2a).

Single MOS1 mutants with altered aromatic residues in the D34D region

Numerous prokaryotic and eukaryotic transposases, and retroviral integrases have a catalytic domain that contains a DD[D/E] triad. Some of them, such as those related to the Tn*5* bacterial transposon, contain a well-conserved aromatic residue in the region spanning the two last aspartate residues. This residue, which is frequently a tryptophan, is involved in a "base flipping" process that occurs during excision (Bischerour and Chalmers 2007). Highly conserved aromatic residues located between the two last aspartate residues of the catalytic triad can also be found in some eukaryotic transposase families. For instance, the aromatic residue located at positions Y271 in HIMAR1 and W268 in MOS1 is conserved in every full-length *mariner* transposase so far available in a database. Three MOS1 mutants were therefore designed, one of which was expected to display either reduced transposition activity or none at all (W268A), and two with theoretically silent mutations, W268F and W268Y. Transposition assays in *E. coli* revealed that all three mutants were in fact completely inactive, thus indicating that this position was essential for MOS1 activity.

Transposition activity of MOS1 mutant combinations

Five mutants (F53Y, Q91R, E137K, T216A and Y237C) were selected for further analyses, due either to their own hyperactivity, or (when they were shown as efficient as the wild type MOS1) to their interaction properties. Before analyzing mutant combinations, we checked whether each

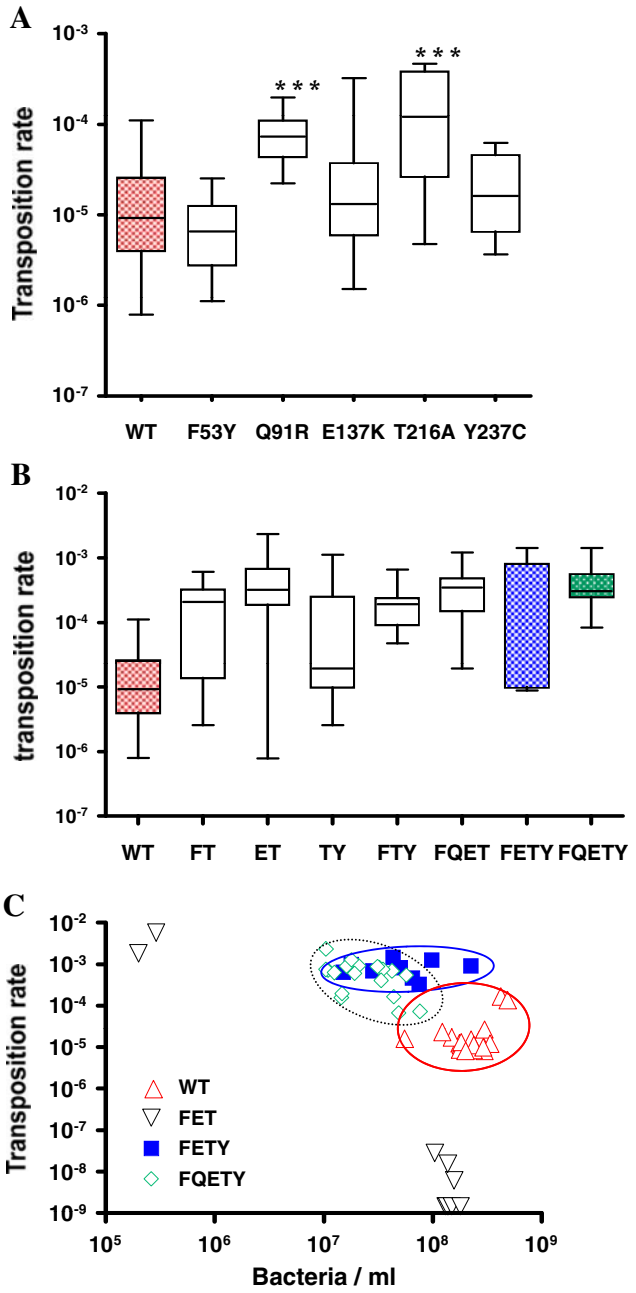


Fig. 2 Transposition rate of mutant transposases compared to that of the wild type MOS1 in transposition assays in bacteria. At least three independent clones from three independent transformations (giving nine values) have been recovered for each protein. Values are reported in a Tukey diagram with boxes-and-whiskers, and drawn using Prism 4 (GraphPad Software). The *middle line* in the box is the median, whereas the *top and bottom* of the box are the third and first quartiles, respectively. So-called "*whiskers*" indicate the maximum and minimum values obtained for each mutant. Nonparametric Mann–Whitney tests were performed (confidence interval: 95%) to compare the wild type MOS1 to each mutant assayed. **a** Single mutants. The wild type MOS1 is shown in *red*. *Stars* (***) indicated a significant difference ($P < 0.0005$) with the wild type MOS1. The other mutants present no difference with the wild type MOS1. **b** Multiple mutants. Combinations (double, triple or multiple) of mutations are as follows: F53Y-T216A (FT), E137K-T216A (ET), T216A-Y237C (TY), F53Y-E137K-T216A (FET), F53Y-T216A-Y237C (FTY), F53Y-Q91R-E137K-T216A (FQET), F53Y-E137K-T216A-Y237C (FETY), F53Y-Q91R-E137K-T216A-Y237C (FQETY). The wild type MOS1 is shown in *red*, FETY in *blue*, and FQETY in *green*. All mutants are significantly different ($P < 0.05$) to the wild type MOS1. **c** Cytotoxicity analyses. The concentration of bacteria at the end of the assays was plotted as a function of the transposition rate. Each point corresponded to an independent analysis. The wild type MOS1 is shown in *red*, FET in *black*, FETY in *blue*, and FQETY in *green*

of the five selected was able to carry out transposition events with intact transposon ends, and TA duplication at the insertion sites. For each of the MOS1 mutants, five independent and newly-inserted transposons were therefore cloned and sequenced (data not shown). Sequence analysis revealed that the ITRs were intact, and that all insertions had been into a TA dinucleotide that was duplicated upon insertion.

To find out whether combinations of these mutations have synergistic effects on MOS1 activity, double, triple, quadruple and quintuple mutant combinations were made and assayed for transposition into bacteria. Only the most interesting combinations are reported. We found that seven combinations, FT, ET, TY, FTY, FQET, FETY, and FQETY, all of which contained the T216A mutation, led to at least a 20-fold significant increase in activity compared to the activity of MOS1[+] (Fig. 2b). We also found that some combinations, such as QET, were completely inactive, demonstrating that some mutations can have antagonistic effects on one another. For one mutant, FET, we obtained results that were clone-dependent, and indicated either a loss of activity or a dramatic increase in transposition efficiency (200- to 800-fold; Fig. 2c). We postulated that these results might originate from the toxicity of the hyperactive FET transposase. We therefore determined the cell viability by calculating the bacterial concentration at the end of the assay, and searched for correlations between the bacterial concentration and the transposase(s) activity (Fig. 2c). Our data revealed that MOS1[+] transposition efficiency is not correlated to the bacterial concentration at the end of the assay. Similar data were obtained with inactive MOS1 versions, such as D(D249A)D and DD(D284N) (data not shown). This indicated that MOS1 has little or no cytotoxicity. In contrast, the results showed that the FET transposition efficiency was inversely correlated to the bacterial concentration at the end of the assay. Thus, when transposition occurred with FET, the bacterial concentration at the end of the assay was on average 5000- to 10000-fold lower than that obtained with MOS1[+] and DD(D284N). This indicates that FET expression (as detected in clones with high transposition rates) has a toxic impact on the bacteria. Consequently, bacteria that do not express an active FET might be selected during the

transposition assay (as indicated by clones with low or null transposition rates). In addition, our data also revealed that FQETY also has a cytotoxic effect when over-expressed in bacteria, although this was less pronounced than with FET, since on average it reduced the bacterial concentration at the end of the assay five- to tenfold (Fig. 2c).

Evaluating the transposition activities of the most hyperactive MOS1 prototypes in vitro

In an attempt to circumvent limitations due to toxicity, we tried to evaluate the activities of the FET and FETY using an in vitro approach. This required producing and purifying MOS1[+], FET and FETY using fusions with MBP (Maltose Binding Protein), as this made it possible to purify active *mariner* transposases without carrying out a denaturing step (Auge-Gouillou et al. 2005a, b; Butler et al. 2006). FET and FETY were therefore cloned in the pMal plasmid. Because we had observed toxic effects of the expression of hyperactive mutants of MOS1 on the bacteria survival, two methods for producing MBP-Tpases were tested. The first was used for MOS1[+] and FETY, as described (Auge-Gouillou et al. 2001). The second was used to produce proteins that are difficult to handle, such as FET. The latter has the specific feature of avoiding the induction of protein expression, limiting this to the level resulting from the pLac promoter escape. The three proteins were finally purified and quantified as described (Sinzelle et al. 2008). The in vitro transposition results are reported in Fig. 3. At all the transposase concentrations used (from 0 to 300 nM), we found that the MBP-MOS1 was more active than MBP-FETY and MBP-FET. These data are in agreement with those of Lampe et al. (1999), who found that hyperactive HIMAR1 mutants were less active in vitro than in vivo.

Designing non-phosphorylable MOS1

Previous studies on *P*-element transposase had suggested that eukaryotic transposases are sensitive to phosphorylation (Beall et al. 2002), and so we first investigated whether MOS1 was phosphorylated when produced in a eukaryotic context, and then checked for relationships between phosphorylation and activity.

MOS1 activity in eukaryotic cell extracts

Using an in vitro coupled transcription and translation eukaryotic system, we showed that MOS1 is labelled by ATP-γP^{32} as soon as it is synthesized, indicating that MOS1 is phosphorylated under these conditions (Fig. 4). When assayed in electrophoretic mobility shift assays (EMSAs), the transposase obtained under these conditions did not interact with *Mos1* ITR (not shown). In an attempt to further verify the phosphorylation status of a MOS1[+] expressed in a eukaryotic system, HeLa cells were transfected with pCS2+ plasmids, some of which expressed a MOS1[+] version tagged with the V5 antigen of HSV at its N-terminal end. The functionality of this fusion protein was confirmed using transposition assays in bacteria, and we found that this V5-MOS1[+] had the same properties as MOS1[+] (not shown). Crude protein extracts (containing MOS1 or not)

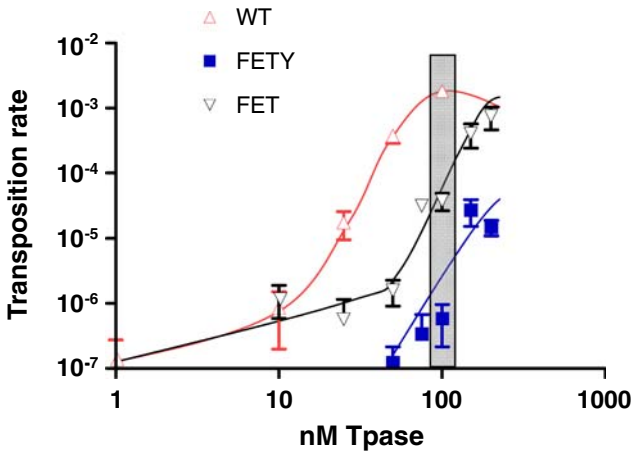


Fig. 3 Transposition rate as a function of Tpase concentration for three proteins: MBP-MOS1 (*red triangle*), MBP-FETY (*blue square*) and MBP-FET (*black triangle*). Medians and quartiles were calculated from at least five replicates (using Prism 4.0 GraphPad software). Values recovered for 100 nM Tpase are underlined in *gray*. At this concentration, the transposition activity of MBP-FETY and MBP-FET is significantly lower ($P < 0.05$) to that of the wild type MOS1

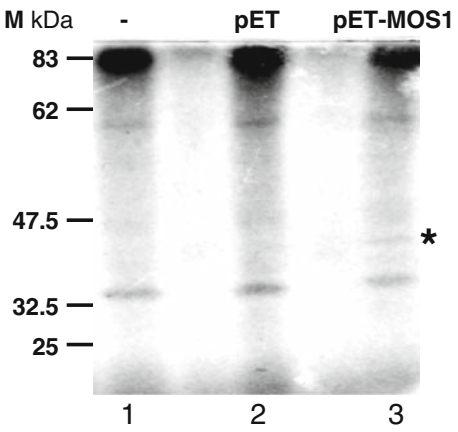


Fig. 4 MOS1 phosphorylation status using a eukaryotic in vitro expression system. SDS–PAGE analyses of samples from in vitro coupled transcription–translation reaction without plasmid (*lane 1*), with an empty pET26b+ (*lane 2*), and with pET26b+ coding MOS1 (*lane 3*). Protein syntheses were performed in the presence of ATP γ^{32}P. The resulting labelled MOS1 is indicated (*asterisk on the right margin*). *M* molecular weight markers. Data from B. Brillet's PhD (Brillet 2006)

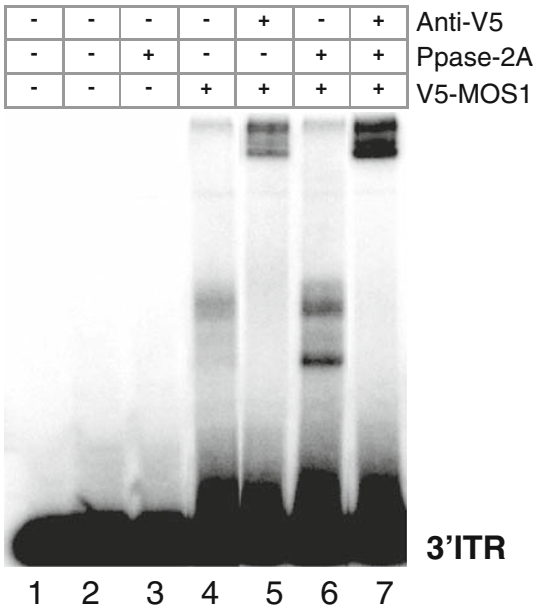


Fig. 5 EMSA for MOS1 activity analyses in HeLa cells crude extracts. EMSA were performed using the 3′ITR (ITR70) as a probe and HeLa crude extracts prepared from pCS2+ transfected cells (*lanes 2-3*) or pCS2 V5-MOS1 transfected cells (*lanes 4–7*). Extracts were either treated with Ppase-2A (*lanes 3, 6, 7*) and/or with anti-V5 antibody (*lanes 5–7*). *Lane 1*: probe alone

were prepared and subjected to EMSA together with *Mos1* 3′ITR as a probe (Fig. 5). Without treatment, weakly shifted complexes were detected, whatever the conditions used (lane 4). Because we used crude extracts, we confirmed that the shifted complexes were specific to the MOS1–ITR interaction by using large amount of competitor DNA, and achieving super-shift assays with a monoclonal anti-V5 antibody (lane 5). Our results revealed that all the complexes were shifted by this monoclonal antibody, thus demonstrating that they relied on specific interactions between MOS1 and its ITR. To find out whether phosphorylation affected the binding of the transposase to MOS1, the extracts were then treated with Ppase-2A, a phosphatase that specifically hydrolyses *O*-phosphate links in phosphorylated serine and theonine residues. Shifted complexes obtained with Ppase-2A treated extracts had higher signal intensities than the untreated samples (compared lanes 4 and 6). Once again, all these complexes were shifted by the monoclonal anti-V5 antibody (lane 7). Taken together, these data indicate that MOS1–ITR binding is sensitive either directly or indirectly to Ppase-2A treatment, thus supporting the hypothesis that phosphorylation may control MOS1 activity.

Non-phosphorylable MOS1 mutants

In-silico analyses of the MOS1 sequence were performed to identify the most phosphorylation-prone positions using four types of prediction software (Table 1). Fifteen residues were predicted to be phosphorylated by at least three methods. We analyzed some of these residues, which were located throughout the length of the entire MOS1 protein, and chosen as follows: three in the HTH sub-domain (T88, S99, and S104), one in the linker region, between the catalytic and ITR binding domains (T135), and two within the catalytic domain (S170, and S305). We also analyzed one residue in the N-terminal MOS1 dimerization sub-domain (S55), even though all the residues in this domain had been identified only twice as likely to be phosphorylated.

Impact of pseudo-phosphorylation on MOS1 activity

We first checked whether phosphorylation at these positions had any impact on MOS1 activity. To do this, pseudo-phosphorylated MOS1 were designed by replacing serine and threonine by negatively charged amino acids that mimic phosphorylation (i.e., aspartic acid and glutamic acid, respectively, (Dai et al. 2007)). The resulting MOS1 mutants were used in *E. coli* transposition assays. We hypothesized that positions insensitive to phosphorylation would lead to mutants with transposition activity similar to that of MOS1[+], whereas positions positively or negatively activated by phosphorylation would result in mutants with increased or reduced transposition activities. Two classes of mutant were obtained (Table 2): those with no (or very little) transposition activity, accounting for a putative down-regulation by pseudo-phosphorylation at these positions (T88E, S99D, S104D, T135E and S170D), and those showing little change as a result of the mutation (S55D and S305D), which were not analyzed any further.

Attempt to design non-phosphorylable MOS1

Our data therefore suggested that at least five positions are involved in the down-regulation of MOS1 activity by phosphorylation. Three of these (T88, S99 and S104) are part of the HTH motif, which allows specific recognition of *Mos1* ITR. Based on the data described above, the simplest hypothesis suggests that when expressed in mammalian cells, MOS1 undergoes phosphorylation at at least one of these positions, thus inhibiting its activity as a result of its reduced ability to bind to the ITR. This hypothesis is supported by the fact that HIMAR1 (which is active in human cells) had no phosphorylable residues at positions corresponding to two of these positions: T88 and S104 in MOS1, which correspond to K87 and G103, respectively in HIMAR1 (Butler et al. 2006). The third residue, S99, which corresponds to S98 in HIMAR1, is predicted to be non-phosphorylated. Consequently, replacing T88, S99 and S104 by non-phosphorylable residues might help to restore MOS1 activity in human cells. As has been done for the P

Table 1 Residues predicted to be phosphorylated in the MOS1 sequence

Residues	GPS	Kinase Phos	NetPhos 2.0	Net PhosK	MOS1 domain
S3				x	Dimerization
T11			x		
T13				x	
T24				x	
S28	x		x		
T42	x			x	
S55*			x	x	
Y72		x	x		HTH
T88*	**x**		**x**	**x**	
S99*	**x**	**x**		**x**	
S104*	**x**		**x**	**x**	
T135*	**x**	**x**	**x**	**x**	Linker
S141				x	Catalysis
S147	x		x	x	
T154		x	x	x	
S170*	**x**	**x**	**x**	**x**	
Y171	x	x	x		
T179	x				
S180	x		x	x	
T181	x		x	x	
S200				x	
Y219				x	
Y237	x	x	x	x	
S253	x				
T262				x	
S277	x		x	x	
Y283			x		
S302	x		x	x	
Y303	x		x	x	
S305*	**x**		**x**	**x**	
Y322			x		
Y345			x		

Four servers were used: GPS (http://bioinformatics.lcd-ustc.org/gps_web/index.php), KinasePhos (http://kinasephos.mbc.nctu.edu.tw), NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos), and NetPhosK (http://www.cbs.dtu.dk/services/NetPhosK). Residues that were predicted as being phosphorylated at least three times are in bold; residues that were analyzed in our studies are indicated by an asterisk

element (Beall et al. 2002), alanine substitution was first performed: T88A, S99A and S104A were assayed for transposition in bacteria and did not yield functional proteins (Table 3).

For each of the three positions, different substitutions were scored for HTH using NPS@ Helix-Turn-Helix motif prediction software (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html). All the scores suggest an almost 100% probability that the peptides (72–120) contained a helix-turn-helix motif, indicating that a single mutation within the HTH did not alter its structure, whereas the corresponding transposition efficiencies obtained using in vivo transposition assays were mainly near zero. Finally, we assayed two combinations of mutations (T88V/S104V, and T88C/S104V). The first was less active than the corresponding single mutants (Table 3), indicating that the effects of the mutations are synergistic.

Finally, three other positions (outside the HTH) had scores that appeared to be highly predictive of phosphorylation: T135, S170, and Y237. Transposition assays indicated that residues T135A and T135D both have activity similar to that of MOS1[+]. In contrast, S170 must be sensitive to modification, since S170D is non-functional whereas S170A has low activity (4.10^{-6}, i.e., 50-fold less than MOS1[+]). This was surprising, because S170 belongs to a loop region (residues 165–188) with no assigned function and no associated crystallographic structures (Richardson et al. 2006). Concerning the third position (Y237), our initial studies have revealed a mutation that is both active and a non-phosphorylable (Y237C).

Table 2 Transposition activity of MBP-MOS1 pseudo-phosphorylated mutants

Proteins[a] MOS1[+]	Transposition rate $10^{-4} \pm [5 \times 10^{-5}]$
S55D	$4 \times 10^{-5} \pm [3 \times 10^{-5}]$
T88E	0
S99D	0
S104D	0
T135E	$2 \times 10^{-6} \pm [2 \times 10^{-6}]$
S170D	0
S305D	$6 \times 10^{-5} \pm [2 \times 10^{-5}]$

Data are the median of at least five independent clones from two independent transformations

Standard deviations are indicated in brackets

[a] Data from B. Brillet's PhD (Brillet 2006)

Table 3 Transposition activity of helix-turn-helix (HTH) MBP-MOS1 mutants

Proteins	P-state at position	Transposition rate
MOS1[+]	WT	10^{-4}
T88A[a]	Non-P	$5 \times 10^{-7} \pm [4 \times 10^{-7}]$
T88C[a]	Non-P	$2 \times 10^{-6} \pm [2 \times 10^{-6}]$
T88E[a]	Pseudo-P	0
T88R[a]	Non-P	0
T88V[a]	Non-P	$10^{-5} \pm [6 \times 10^{-6}]$
S99A[a]	Non-P	0
S99C[a]	Non-P	$10^{-6} \pm [1.3 \times 10^{-6}]$
S99D[a]	Pseudo-P	0
S99V	Non-P	$10^{-7} \pm [5 \times 10^{-6}]$
S104A[a]	Non-P	$3 \times 10^{-7} \pm [10^{-6}]$
S104C[a]	Non-P	0
S104D[a]	Pseudo-P	0
S104E[a]	Pseudo-P	0
S104K	Non-P	$3 \times 10^{-6} \pm [4 \times 10^{-6}]$
S104N[a]	Non-P	0
S104P	Non-P	0
S104V	Non-P	$3 \times 10^{-6} \pm [4 \times 10^{-6}]$
S104Y[a]	Non-P	0
T88V/S104V	Non-P	0
T88C/S104V	Non-P	$2 \times 10^{-6} \pm [3 \times 10^{-6}]$

Data are the median of at least five independent clones from two independent transformations

Standard deviations are indicated in brackets

[a] Data from B. Brillet's PhD (Brillet 2006)

Discussion and conclusions

Our studies report rational mutagenesis performed to obtain hyperactive and non-phosphorylable MOS1 variants.

At least four hyperactive MOS1 prototypes, FET, FQET, FETY and FQETY were isolated, being from 30- to 800-fold more active than MOS1[+]. Our findings therefore suggest that these proteins might be highly toxic, probably leading us to under-estimate the transposition activities of these hyperactive MOS1. Moreover, cytotoxicity made it difficult to establish good and reproducible conditions for protein production, probably due to differences in bacterial growth during protein production. Indeed, bacteria that unfold over-expressed proteins probably have a very considerable advantage during bacterial growth, leading to the production of recombinant protein with very low specific activity. Altogether, our findings led us to the conclusion that in vitro transposition assays might be inaccurate for evaluating the activity of recombinant hyperactive transposases produced in bacteria.

Our data about non-phosphorylable MOS1 variants are relevant to the fact that the MOS1 HTH motif is essential for *Mos1* transposition. In consequence, it cannot be drastically modified if *Mos1* transposition is to be preserved. Further investigations will be required to confirm the HTH phosphorylation pattern. Nevertheless, phosphorylation of the serine and threonine contained within the HTH is a promising hypothesis to explain why *Mos1* has little activity in most invertebrates, and is unable to transpose in human cells. Indeed, at least two positions (T88 and S99) are putatively phosphorylated by ATM and/or DNA-PK kinases, which are activated when a mammalian cell suffers double strand damage to its genomic DNA. This is clearly the case during transposition.

This study is important since the findings reported here may facilitate the rational design of improved genetic tools for vertebrates. We conclude that hyperactive transposases are possible, even if the matrix is not a resurrected protein resulting from in-silico analyses from fossil transposon sequences. Nevertheless, as indicated by comparison between in vivo and in vitro analyses (performed for MOS1, (Pledger and Coates 2005) and this study, and for HIMAR1, (Lampe et al. 1999)), hyperactive transposases need to be established in the cell where they are used. Indeed, many factors may affect transposase activity, and engineering a hyperactive transposase in one system does not necessarily ensure that it will be successful in another system. Despite the commonly held opinion, this indicates that *Mos1*, and by extension *Himar1*, is probably sensitive to several host factors. One of these factors might be post-translational modification of MOS1, such as phosphorylations. The mass spectrometric definition of the phosphorylation patterns of MOS1 expressed in several different organisms will be required to enable us to reach any final conclusion about the role of these modifications in controlling the activity of this enzyme.

Our work raises several major questions about the value of developing hyperactive transposases for transposon tool technology, particularly in the field of gene transfers for therapeutic purposes. Indeed, we have demonstrated that the most hyperactive MOS1s are also those most toxic for the producer cell. We cannot rule out the hypothesis that what is true for MOS1 is also true for the other hyperactive transposases designed to date. For in vitro purposes, our data suggest that the purified hyperactive transposases never display as much specific activity as expected. This will be the case as long as in vitro producing systems, eventually based on the translation of a hyperactive transposase coding mRNA, are not used. Such systems already exist (RTS 500 Wheat Germ CECF kit—Roche Applied Science) and need only to be adapted to answer our questions. On the other hand, the use of hyperactive/cytotoxic transposases for in vivo purposes (gene therapy for instance) raises several problems. Firstly, cytotoxicity is never evaluated in DNA transfer assays mediated by transposases. Nevertheless, all of the studies that have been successfully performed to date have involved hyperactive transposases (Keravala et al. 2006; Wu et al. 2006). If these hyperactive transposases are toxic, this means that in any attempt to devise safe gene transfer processes, it becomes essential to evaluate their genotoxicity. Secondly, recent studies have revealed another source of poor performance associated with transposon based-vectors. Successful transgene transposition requires efficient transposition (where hyperactivity might be useful) followed by efficient transgene expression. However, hyperactivity does not provide a way of circumventing low transgene expression. This point was recently illustrated by Garrison et al. (2007): vectors based on the *Sleeping Beauty* (*SB*) transposon underwent a high degree (up to 94%) of post-integrative gene silencing, which varied with the promoter used for transgene expression. More recently, Dalsgaard et al. (2009) demonstrated that *SB*-based vectors, which are prone to transcriptional silencing, could be protected by insulators, resulting in long-term transgene expression. Taken together, these data indicate that producing a hyperactive transposase is no longer a possible solution, particularly in the light of their probable toxicity.

Finally, we now have at least three hypotheses that could account for the fact that *Mos1*-based vector activity is not detected in mammalian cells (including human cells). First, *Mos1*-based vectors might be completely silenced. This point has never been checked. Second, MOS1 is not sufficiently effective to promote a large-scale detectable level of transposition. Results obtained by Williams et al. (2005), and Robert and Bessereau (2007) using *Mos1* in *C. elegans* do not support this hypothesis. Third, MOS1 might be down-regulated by post-translational modification, as suggested by the present data. This point remains to be verified, but as illustrated here, it will be very difficult to design non-phosphorylable MOS1 for a possible use in human cells, due to the location of essential residues within the HTH that are possible targets for various cellular kinases. From this point of view, *Himar1* is actually the most efficient way to develop *mariner*-based genetic tools for mammals. On the other hand, *Mos1* remains a powerful tool for genome engineering in cold-blood vertebrates and some invertebrates, such as Nematodes.

Materials and methods

Site-directed mutagenesis

MOS1 transposase mutants were obtained with the Quickchange® Site-directed mutagenesis kit (Stratagene) according to the manufacturer's recommendations. The template used was the gene of MOS1 transposase cloned into the pGEM-T plasmid (Promega, Charbonnières, France). Briefly, 25 ng of template were amplified using complementary primers harboring the mutation to be introduced for 16 cycles (95°C 30 s, 55°C 1 min, and 68°C 8 min 30 s). The template was then degraded by DpnI (3u) treatment for 1 h at 37°C, and the PCR product was transformed into XL1-blue competent cells. Mutations were verified by sequencing (MWG biotech, Germany). After sequencing, MOS1 mutant genes were sub-cloned in pKK-233 plasmid (Clontech) for in vivo transposition assays, and in pMal-C2 plasmid (New England Biolabs) for in vitro analyses, and for in vivo transposition assays concerning the non-phosphorylable mutants. The abbreviations for the multiple mutants are as follows: F53Y-Q91R (FQ), F53Y-T216A (FT), F53Y-Y237C (FY), Q91R-Y237C (QY), E137K-T216A (ET), E137K-Y237C (EY), T216A-Y237C (TY), F53Y-E137K-T216A (FET), F53Y-T216A-Y237C (FTY), Q91R-E137K-T216A (QET), F53Y-Q91R-E137K-T216A (FQET), F53Y-E137K-T216A-Y237C (FETY), F53Y-Q91R-E137K-T216A-Y237C (FQETY).

Proteins production, purification and quantification

MBP-MOS1, MBP-FET, and MBP-FETY were cloned in a pMal-C2 plasmid (New England Biolabs) produced as fusion proteins linked to the maltose-binding protein (MBP), under the control of the pLac promoter. One colony (containing the corresponding pMal-Tpase plasmid) was taken from a fresh transformation (*E. coli*, strain DH5α), and grown at 37°C for 16 h whilst shaking in 2.5 ml LB-ampicillin (100 µg/ml).

Method 1. The overnight culture was then transferred into 500 ml of LB-ampicillin (100 µg/ml), and grown at

37°C whilst shaking until $OD_{600} = 1$. Protein production was induced by 1 mM IPTG for 30 min.

Method 2. The overnight culture was then transferred into 50 ml of LB-ampicillin (100 µg/ml), glucose (0.2%), and grown at 37°C for 8 h whilst shaking. This culture was then transferred into 1 l of LB-ampicillin (100 µg/ml), glucose (0.2%), and grown at 37°C for 16 h whilst shaking.

The cells were then harvested by centrifuging, suspended in 2.5 ml of buffer (20 mM Tris–HCl pH9, 200 mM NaCl, 1 mM DTT), and ruptured. Insoluble material was removed by centrifugation. The supernatant was the crude extract, containing the expected fusion proteins. The total protein concentration in the crude extract was about 3 mg/ml, as determined by a Bradford assay using bovine serum albumin as standard.

It must be noticed that, in contrast to what we currently obtained for the wild type MBP-MOS1, MBP-FET and MBP-FETY were difficult to produce with both methods, and productions were not reproducible, giving different results (in term of MBP-Tpases amounts) for a similar method of production. In addition, we failed to obtain MBP-FET using the first method, in amounts allowing in vitro studies. This was very probably due to the high toxicity of this protein. MBP-Tpases were then purified following the instruction of the manufacturer (New England Biolabs).

E. coli transposition assays

Transposition assays were performed as described (Sinzelle et al. BMC Mol Biol 2008). The genes coding MOS1 and the mutants were sub-cloned in the pKK-233 plasmid (Clontech), yielding plasmids designated pKK-Tnp. The pBC-3T3 was used as a pseudo-*Mos1* donor plasmid. It contained the pBR322 tetracycline resistance gene (without promoter) flanked by two *Mos1* 3′ITR. This plasmid was unable to confer resistance to *E. coli* cells in tetracycline concentrations over 10 µg/ml. Each pKK-Tnp and the pBC-3T3 were used to co-transform JM109 competent cells. We thus obtained several bacterial strains, each containing a pseudo-*Mos1* donor and a source of transposase (WT or variants). For each strain, five colonies were taken from a fresh co-transformation and grown at 37°C for 1 h, prior inducing transposition with 1 mM IPTG. The cells were grown for 1 h at 32°C. Transpositions were detected by plating the cells on LB medium containing tetracycline (12.5 µg/ml). Cells were titrated by plating an appropriate dilution of the cultured cells on LB medium. The transposition rate was calculated as the number of Tet^R colonies divided by the total number of cells. The transposition rate of the wild type transposase was taken as the reference value.

Similar assays were performed with the pMal-MOS1 vector (encoding for MBP-MOS1), and the pMal-MOS1-non-phosphorylable variants. Transposition rates recovered for MOS1 and MBP-MOS1 are rather similar.

In vitro transposition assays

In vitro transposition reactions were performed using pBC-3T3 as both the donor DNA and the target, consequently they are carried out in the same way as the previously used in the in vivo (bacterial) assay: transposition events were revealed by promoter tagging, the tetracycline resistance being activated through the *cat* gene promoter. Basic transposition reaction mixtures contained 1 mM Tris–HCl (pH 9), 5 mM NaCl, 20 mM $MgCl_2$, 0.5 mM EDTA, 5 ng/ml BSA, 600 ng of super-coiled pBC-3T3, and different concentrations of MBP-MOS1 (or MBP-FET, or MBP-FETY), in a volume of 20 µl. The reactions were allowed to proceed for 30 min at 30°C, before 10 µl of stop solution (0.4% SDS, 0.4 µg/µl proteinase K) was added. The reaction mixtures were then incubated at 37°C for a further 30 min, and then at 65°C for 10 min. The products were phenol-chloroform extracted, and ethanol precipitated with 1 µg of yeast tRNA using standard techniques. 10% of the reaction mixture was co-transformed with 0.01 ng of pBS-SK(−), used as positive transformation control (Stratagene) in electro-competent JM109 *E. coli*. Bacteria were grown at 37°C for 1 h in SOC media. Appropriate dilutions of each reaction mixture were plated on LB-ampicillin (100 µg/ml) agar to score for transformation efficiency, and on LB-tetracycline (12.5 µg/ml) agar and LB-chloramphenicol (80 µg/ml) agar to score for transposition rate. The transposition rate was the number of Tet^R colonies divided by the number of $Chloram^R$ plus Tet^R colonies. The efficiency of transformation was monitored from the number of Amp^R colonies. Each experiment was repeated five times.

In vitro coupled transcription/translation

The $T_NT^®$ T7 Quick Coupled Transcription/Translation System (Promega) was used to produced MOS1 in a eukaryotic (rabbit) system following the manufacturer instructions. Briefly, MOS1 open reading frame was cloned between *Nco*I and *Hind*III restriction sites of pET26b+ (Novagen), giving pET-MOS1. 0.5–1 µg of pET-MOS1 were incubated for 90 min at 30°C with 40 µl of T_NT mix and 2 µl (10 µCurie) of ATP $\gamma^{32}P$. Controls were performed using pET26b+ without the MOS1 sequence, and/or without plasmid. At the end of the incubation, 5 µl of samples were taken off and mixed with 5 µl of loading buffer. After boiling, samples were loaded on discontinuous 4–10% SDS–PAGE. Gels were run, dried and autoradiographed.

HeLa cells transfection and crude extracts preparation

HeLa cells were cultured in DMEM (PAA) supplemented with 10% foetal bovine serum. About 4.10^5 cells were seeded onto 6-wells plates, 1-day prior transfection. Cells were then transfected with TransPEI, according to producer's recommendation (Eurogentec). Briefly, pCS2+ (Invitrogen) or pCS2 V5-MOS1 (2 µg) and TransPEI (4 µl) were separately diluted into 50 µl of 150 mM NaCl solution, and gently mixed together. After 30 min at room temperature, the mixture was diluted into 1 ml OPTIMEM medium (PAA). Cells were then incubated with DNA–PEI complexes for 3 h. The transfection solution was then removed and replaced by fresh supplemented DMEM for 24 h at 37°C. Cells were scraped, washed with PBS and then disrupted in 20 mM Tris (pH9), 100 mM NaCl, 1 mM DTT and protease inhibitors (Roche Diagnostics) by making 5 freezing-thawing cycles. After centrifugation, proteins concentration in the supernatant (i.e., the crude extract) was determined using the Bradford assay (BioRad).

Electrophoretic mobility shift assays

Reactions were carried out in 20-µl containing 10 µg protein crude extract, 0.2 pmol ^{32}P-labeled ITR70 (Auge-Gouillou et al. 2005a, b), 25 mM EDTA, 5% glycerol, 10 µg herring sperm and EMSA buffer (100 mM Tris (pH 7.4), 100 mM NaCl, 1 mM DTT, protease inhibitors). When specified, 1.2 µg anti-V5 antibody (Invitrogen) and/or 0.5 U PPase-2A (Promega) were added to the mixture. Reactions were incubated for 2 h at 30°C and samples were separated by electrophoresis on native discontinuous 4–9% polyacrylamide gel containing 5% glycerol. Gel was visualized with a phosphorimager (Storm 860, Amersham Biosciences).

In-silico analyses

MOS1 serine, threonine and tyrosine phosphorylation predictions were performed using four different servers.

The Group-based Phosphorylation Scoring method or GPS (http://bioinformatics.lcd-ustc.org/gps_web/index.php) is based on a well-established rule to classify protein kinases into a hierarchical structure with four levels, and developing a simple approach to estimate the theoretically maximal false positive rate. GPS could predict kinase-specific phosphorylation sites for mammalian protein kinases in hierarchy, (Zhou et al. 2004; Xue et al. 2005).

KinasePhos (http://kinasephos.mbc.nctu.edu.tw) computationally predicts phosphorylation sites within given protein sequences. The known phosphorylation sites are categorized by substrate sequences and their corresponding protein kinase classes. Profile Hidden Markov Model is applied for learning to each group of sequences surrounding to the phosphorylation residues. By comparing to other approaches previously developed, this method has higher accuracy and provides not only the location of the phosphorylation sites, but also the corresponding catalytic protein kinases.

NetPhos 2.0 (http://www.cbs.dtu.dk/services/NetPhos) produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins (Blom et al. 1999).

NetPhosK 1.0 (http://www.cbs.dtu.dk/services/NetPhosK) produces neural network predictions of kinase specific eukaryotic protein phosphoylation sites. Currently NetPhosK covers the following kinases: PKA, PKC, PKG, CKII, Cdc2, CaM-II, ATM, DNA PK, Cdk5, p38 MAPK, GSK3, CKI, PKB, RSK, INSR, EGFR and Src (Blom et al. 2004).

Scores for HTH motifs prediction were obtained using the NPS@ web site server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html) (Dodd and Egan 1990).

Acknowledgments This work was funded by the University of Tours, the Centre National de la Recherche Scientifique; the Ministère de l'Education Nationale, de la Recherche et de la Technologie and GDR CNRS no. 2157. This work was also funded by grants from the European Commission (Project SyntheGeneDelivery, grant no. 018716 to YB), the Association Française contre les Myopathies (grant AFM no. 11468 to CAG), and the Agence Nationale de la Recherche (Project Magic Transpo, grant ANR-06-EMPB-014-01 to CAG).

References

Auge-Gouillou C, Hamelin MH et al (2001) The ITR binding domain of the mariner Mos-1 transposase. Mol Genet Genomics 265(1):58–65. doi:10.1007/s004380000386

Auge-Gouillou C, Brillet B et al (2005a) Mariner Mos1 transposase dimerizes prior to ITR binding. J Mol Biol 351(1):117–130. doi:10.1016/j.jmb.2005.05.019

Auge-Gouillou C, Brillet B et al (2005b) Assembly of the mariner Mos1 synaptic complex. Mol Cell Biol 25(7):2861–2870. doi:10.1128/MCB.25.7.2861-2870.2005

Baus J, Liu L et al (2005) Hyperactive transposase mutants of the sleeping beauty transposon. Mol Ther 12(6):1148–1156. doi:10.1016/j.ymthe.2005.06.484

Beall EL, Mahoney MB et al (2002) Identification and analysis of a hyperactive mutant form of *Drosophila* P-element transposase. Genetics 162(1):217–227

Bischerour J, Chalmers R (2007) Base-flipping dynamics in a DNA hairpin processing reaction. Nucleic Acids Res 35(8):2584–2595. doi:10.1093/nar/gkm186

Blom N, Gammeltoft S et al (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 294(5):1351–1362. doi:10.1006/jmbi.1999.3310

Blom N, Sicheritz-Ponten T et al (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 4(6):1633–1649. doi:10.1002/pmic.200300771

Brillet B (2006) Activité de la transposase de l'élément *mariner Mos1*: mise en place du complexe synaptique et régulation par phosphorylation. PhD of the University of Tours

Brillet B, Bigot Y et al (2007) Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. Genetica 130(2):105–120. doi:10.1007/s10709-006-0025-2

Butler MG, Chakraborty SA et al (2006) The N-terminus of Himar1 mariner transposase mediates multiple activities during transposition. Genetica 127(1–3):351–366. doi:10.1007/s10709-006-6250-x

Dai J, Xie W et al (2007) Phosphorylation regulates integration of the yeast Ty5 retrotransposon into heterochromatin. Mol Cell 27(2):289–299. doi:10.1016/j.molcel.2007.06.010

Dalsgaard T, Moldt B et al (2009) Shielding of sleeping beauty DNA transposon-delivered transgene cassettes by heterologous insulators in early embryonal cells. Mol Ther 17(1):121–130. doi:10.1038/mt.2008.224

Dodd IB, Egan JB (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. Nucleic Acids Res 18(17):5019–5026. doi:10.1093/nar/18.17.5019

Garrison BS, Yant SR et al (2007) Postintegrative gene silencing within the sleeping beauty transposition system. Mol Cell Biol 27(24):8824–8833. doi:10.1128/MCB.00498-07

Keravala A, Liu D et al (2006) Hyperactive Himar1 transposase mediates transposition in cell culture and enhances gene expression in vivo. Hum Gene Ther 17(10):1006–1018. doi:10.1089/hum.2006.17.1006

Lampe DJ, Akerley BJ et al (1999) Hyperactive transposase mutants of the Himar1 mariner transposon. Proc Natl Acad Sci USA 96(20):11428–11433. doi:10.1073/pnas.96.20.11428

Pledger DW, Coates CJ (2005) Mutant Mos1 mariner transposons are hyperactive in Aedes aegypti. Insect Biochem Mol Biol 35(10):1199–1207. doi:10.1016/j.ibmb.2005.06.002

Reznikoff WS (2006) Tn5 transposition: a molecular tool for studying protein structure-function. Biochem Soc Trans 34(Pt 2):320–323. doi:10.1042/BST20060320

Richardson JM, Zhang L et al (2004) Expression, purification and preliminary crystallographic studies of a single-point mutant of Mos1 mariner transposase. Acta Crystallogr D Biol Crystallogr 60(Pt 5):962–964. doi:10.1107/S0907444904003798

Richardson JM, Dawson A et al (2006) Mechanism of Mos1 transposition: insights from structural analysis. EMBO J 25(6):1324–1334. doi:10.1038/sj.emboj.7601018

Robert V, Bessereau JL (2007) Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks. EMBO J 26(1):170–183. doi:10.1038/sj.emboj.7601463

Shao H, Tu Z (2001) Expanding the diversity of the IS630-Tc1-mariner superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics 159(3):1103–1115

Sinzelle L, Jegot G et al (2008) Factors acting on Mos1 transposition efficiency. BMC Mol Biol 9(1):106. doi:10.1186/1471-2199-9-106

Twining SS, Goryshin IY et al (2001) Functional characterization of arginine 30, lysine 40, and arginine 62 in Tn5 transposase. J Biol Chem 276(25):23135–23143. doi:10.1074/jbc.M010748200

Williams DC, Boulin T et al (2005) Characterization of Mos1-mediated mutagenesis in Caenorhabditis elegans: a method for the rapid identification of mutated genes. Genetics 169(3):1779–1785. doi:10.1534/genetics.104.038265

Wu SC, Meir YJ et al (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. Proc Natl Acad Sci USA 103(41):15008–15013. doi:10.1073/pnas.0606979103

Xue Y, Zhou F et al (2005) GPS: a comprehensive www server for phosphorylation sites prediction. Nucleic Acids Res 33(Web Server issue):W184–187

Yant SR, Park J et al (2004) Mutational analysis of the N-terminal DNA-binding domain of sleeping beauty transposase: critical residues for DNA binding and hyperactivity in mammalian cells. Mol Cell Biol 24(20):9239–9247. doi:10.1128/MCB.24.20.9239-9247.2004

Zayed H, Izsvak Z et al (2004) Development of hyperactive sleeping beauty transposon vectors by mutational analysis. Mol Ther 9(2):292–304. doi:10.1016/j.ymthe.2003.11.024

Zhang L, Dawson A et al (2001) DNA-binding activity and subunit interaction of the mariner transposase. Nucleic Acids Res 29(17):3566–3575. doi:10.1093/nar/29.17.3566

Zhou FF, Xue Y et al (2004) GPS: a novel group-based phosphorylation predicting and scoring method. Biochem Biophys Res Commun 325(4):1443–1448. doi:10.1016/j.bbrc.2004.11.001

Mariner mobility requires a single helix-turn-helix in the N-terminal transposase domain

Carpentier G, Adet J, Bouchet N, Demattei M, Renault S, Bigot Y, Augé-Gouillou C,

soumis à *EMBO Reports*

Mariner mobility requires a single helix-turn-helix in the N-terminal transposase domain

Guillaume Carpentier[1,2], Jérémy Adet[1,2], Nicolas Bouchet[1,2], Marie-Véronique Demattei[1,2], Sylvaine Renault[1,2], Yves Bigot[2,3], and Corinne Augé-Gouillou[1,2]*

1) Université François Rabelais de Tours, GICC,

2) CNRS, UMR 6239,

3) CHRU de Tours,

UFR des Sciences & Techniques, Parc Grandmont, 37200 Tours – FRANCE,

*Corresponding author: Tel: +33 2 47 36 70 63, Fax: +33 2 47 36 70 42. E.mail: auge@univ-tours.fr

Exact character count: 23 607.

Abstract

Transposases are one of the main factors that drive the reshaping of eukaryotic genomes during evolution, but little is known about how they work. The structures involved in their DNA binding mechanism have not yet been fully identified. Here, we focus on the organization of the N-terminal domain the *mariner* transposase, MOS1, which is responsible for binding to the *Mos1* ends. Two putative organization models are currently envisaged: one is PAIRED-like and the other CRO-like. Both involve bipartite domains consisting either of two HTH sub-domains, or one dimerization sub-domain juxtaposed to the HTH. In an attempt to clarify this issue, we made two recombinant MOS1 in which the dimerization sub-domain had been replaced by a protein-protein interaction domain (from GAL4 or GCN4), and found that both chimeras were still able to perform transposition. This supports that the properties of the MOS1 ITR binding domain ressemble those of a CRO-like DNA-binding domain.

Keywords: Dimerization, CRO-like proteins, *MLE*, PAIRED-like proteins, *TLE*

Introduction

DNA transposons occur in all the biological kingdoms. To catalyze their mobility within genomes via a DNA intermediate, each transposon encodes and uses a specific transposase. These enzymes accomplish a series of tasks involving homo-oligomerization, binding to the inverted terminal repeats (ITRs) located at the DNA transposon ends, and catalysis of DNA strand cleavages and transfers, which allow transposition to occur (Craig et al, 2002). Based on analyses of their transposase and ITR sequences, nine super-families of eukaryotic transposons have been classified (Feschotte, 2008; Wicker et al, 2007), each encoding a transposase usually consisting of two main domains: an ITR-binding domain (IBD) and a catalytic domain. Interestingly, transposases encoded by members of six of these super-families are all composed of one IBD (of various origins) fused to a common catalytic domain that contains a DDE or a DDD triad. Transposase domain exchanges were found between DNA transposon sub-families (Auge-Gouillou et al, 2000), indicating that chimerization is an evolutionary mechanism that has impacted on every level of DNA transposon evolution: subclass, super-family, family and subfamily.

Here, we focus on the functional organization and the origins of transposase IBDs in two related transposon families, the *Mariner*-like elements (MLEs) and the *Tc1*-like elements (TLEs), both of which belong to the largest transposons super-family, the IS*630-Tc1-mariner* (*ITm*; (Shao & Tu, 2001)). The amino acid sequence of the catalytic domains of both families is closely aligned, whereas their IBD counterparts display no significant sequence similarities (Brillet et al, 2007). This can be interpreted as resulting either from more rapid and marked differentiation of IBDs than of the catalytic domain over the course of evolution, or from the presence of IBDs of different origins. One way to settle this issue is to confront functional and structural data. The sequences and structural analyses of the transposase IBD encoded by the TLE, *Tc3* (Watkins et al, 2004), revealed that it is composed of two helix-turn-helix (HTH) motifs involved in DNA binding and related to the DNA-binding domain (DBD) of the PAIRED-like proteins. PAIRED-like proteins contain two HTH motifs, both of which are involved in DNA binding: PAI is a specific recognition sub-domain, whereas RED simply stabilizes the interaction with the DNA, the two sub-domains being linked by an AT-hook (supplementary information I). The crystallographic data shows that PAIRED transposases interact with their 26-bp direct repeats as monomers (Watkins et al, 2004).

For the MLE, no IBD structure was solved, and the sequence features do not match those of a cardinal PAIRED-like domain. Two putative organization models are currently

envisaged. According to the first model, the TLE and MLE use the same ITR binding mode because they belong to the same super-family. This model therefore proposes that the MLE transposase ITR-binding domain is organized as a RED-PAI domain, i.e. similar, but in reverse orientation, to a PAIRED domain (Richardson et al, 2006). The second model proposes that MLE transposases contain an ITR-binding domain related to that of CRO-like proteins, i.e. a functional dimerization sub-domain linked to a single HTH involved in ITR binding (Auge-Gouillou et al, 2005b; Brillet et al, 2007; Feschotte, 2008; Pietrokovski & Henikoff, 1997). CRO-like proteins interact with their target DNA as dimers (Darling et al, 2000), and the target binding site contains a palindromic motif, as described for the lambda CRO operator (Albright & Matthews, 1998). The MLE ITR contains a palindromic sequence (Bigot et al, 2005), as expected for CRO-like proteins target DNA.

Transposition invariably necessitates the assembly of higher-order nucleoprotein complexes (known as synaptic complexes), and so we need to define both protein-protein and protein-DNA interactions if we are to understand the molecular mechanisms of transposition. With this in view, determining the organization of the MLE IBD is a key to establishing the stoichiometry of the synaptic complex: dimeric in the case of a PAIRED-like DBD, *versus* tetrameric for a CRO-like DBD. We used *Mos1* as a model for addressing the structure-function relationships of MLE IBD. *Mos1* is 1.3 kb in length, and has an imperfect 28-bp ITR. It encodes a single protein, MOS1, that consists of two large domains, a N-terminal IBD and a catalytic domain (Figure 1) (Richardson et al, 2006), and is both necessary and sufficient to trigger transposition. Like all MLE transposases, MOS1 IBD contains only one helix-turn-helix (HTH) motif spanning from residues 72 to 110 (Auge-Gouillou et al, 2001a; Pietrokovski & Henikoff, 1997; Zhang et al, 2001). The main difference between the two IBD organization models described above therefore resides in the role of the 50-first amino acids of MLE transposases: a non-specific DNA binding HTH motif in the case of the PAIRED model, and a dimerization motif in the case of the CRO model. In this study, we used chimeric MOS1 transposases to address the structure-function relationships of MOS1 N-terminal sub-domain. We made two chimeric transposases in which exogenous dimerization sub-domains (with no DNA binding activities) replaced the natural sub-domain. We show that these transposases with a single HTH are still able to promote true transposition. Our data therefore strongly support that the MOS1 N-terminal domain has functional properties similar to those of a CRO-like domain.

Results

Since the nature of the N-terminal sub-domain is a key aspect of both models, we investigated the ability to ensure transposition of MOS1 variants in which the N-terminal natural sub-domain had been replaced by an unrelated dimerization sub-domain. We selected two well-characterized dimerization domains known to have no DNA-binding activity: the yeast regulatory protein Gal4 [amino acids 41 to 100, (Li et al, 2007)], and the leucine zipper (LZ) motif of the yeast control general protein GCN4 (amino acids 253 to 278; (Feldmann et al, 1993)). These domains were substituted for the first 34 amino acids of MOS1 to yield the LZ[35-345], and Gal4[35-345] transposases, respectively (Figure 1). According to the PAIRED model, we would expect to obtain recombinant proteins unable to ensure transposition, whereas according to the CRO-like model this ability would be expected to be restored.

Transposition assays using chimeric transposases

The modified transposases were tested for their ability to trigger *Mos1* transposition in bacteria, the most sensitive test currently available. Reactions were performed using pBC-3T3 as both the donor and target plasmid, and promoter tagging was used to detect transposition events (Auge-Gouillou et al, 2001b); the sensitivity threshold was 2.10^{-10} events per assay (Figure 2A). We first checked that the first 34 amino acids of MOS1 were essential for MOS1 activity. No transposition events was detected using a *Mos1* transposase in which this region was either truncated (MOS1[35-345]) or had been mutated by punctual substitutions (V14P or S28P) that disrupt the local structure of this region (Table 1; (Auge-Gouillou et al, 2001b)a).

The results of transposition assays carried out using the LZ[35-345] and Gal4[35-345] transposases (Table 1) demonstrate that these proteins were indeed able to trigger *Mos1* transposition, albeit with reduced efficiency. Because synaptic complexes are highly organized structures, it is not so surprising to find that their transposition efficiency is altered by changing the way which the complexes are assembled (dimerization equilibrium, DBD accessibility to ITR, transposase-transposase interaction in the complex…).

Substituted transposases mediate *bona fide* transposition

The activity of the LZ[35-345] and Gal4[35-345] transposases having been confirmed in bacterial transposition assays, it was necessary to check whether what we had detected was true transposition. We confirmed that the events involved the mobility of a complete element

and also occurred at a TA dinucleotide that was duplicated upon integration. Five independent *in-vivo* transposition events for each recombinant transposase were analyzed by restriction mapping and sequencing. A single fragment of about 4600 bp was expected from donor pBC-3T3, which contained a single restriction site for *Hind*III (in the pseudo-transposon). When a pBC-3T3 molecule was used as the target, a pBC-3T3 with an additional 3T3 was recovered. These new plasmids in turn gave rise to two bands after *Hind*III digestion, the sizes of which depended on the location of the 3T3 insertion (Figure 2B).

Whichever substituted transposase was used, all insertions were due to complete elements, and were flanked by a TA duplication (Figure 2C), which are the expected landmarks for accurate transposition (Bryan et al, 1990). These data indicated that substituting the first 34 MOS1 amino acids by dimerization sub-domains does not abolish the transposition activity of MOS1. They therefore revealed that the HTH, but not the first 34 amino acids, is required to ensure specific DNA binding to the ITR and correct cleavages at the end of the ITR, to promote DNA strand transfer at the TA target dinucleotide, and to perform integration involving TA duplication at each end of the element.

Discussion

The data presented here indicate that the organization and functioning of the MLE IBD are similar to those of CRO-like domains. The chimeric transposases tested provide new evidence about the role of the first 34 MOS1-residues in the dimerization of the transposase. They demonstrate that a single HTH is needed for *Mos1* transposition, to ensure specific DNA binding to the ITR and correct cleavages at the end of the ITR, and to promote DNA-strand transfer at the TA target dinucleotide. Our molecular controls revealed that the transposition events triggered by both the chimeric transposases are as specific as those mediated by MOS1, since they involve complete elements with an intact ITR when integrated, TA dinucleotide duplication at the insertion sites, and retain their propensity to integrate into the *cat* gene (Crenes et al, 2009). Nevertheless, no information about the structure of the MOS1 dimerization sub-domain is yet available. Predictive data indicate that this sub-domain probably contains several alpha helices (Pietrokovski & Henikoff, 1997; Richardson et al, 2006), which may or may not be associated with beta sheets (Brillet et al, 2007). Further structural investigations will be required to elucidate this issue.

Our data also provide new insights into the assembly of the *Mos1* synaptic complex. One possibility is that MOS1 dimerization takes an HTH from each of two distinct MOS1 molecules in order to organize an efficient DNA-binding domain. Each HTH in the newly formed ITR-binding domain interacts with one half ITR (Bigot et al, 2005). This is similar to what is known to occur for CRO proteins (Albright & Matthews, 1998; Darling et al, 2000). At this point, MOS1 dimers are bound at both *Mos1* ITRs; it is thus tempting to propose that the *Mos1* synaptic complex is simply formed by the association of MOS1 dimers bound at both transposon ends. If this is the case, the *Mos1*-synaptic complex should contain a MOS1 tetramer, as previously proposed (Auge-Gouillou et al, 2005b). A similar hypothesis has been advanced for the *Himar1* element (Butler et al, 2006).

Finally, our results support the hypothesis that the differentiation between TLE and MLE families results from the association of the same catalytic domain with DBDs of different evolutionary origins.

Speculation

Our findings about the DNA-binding properties of MOS1 are consistent with a binding mode similar to that of CRO-like proteins. In this protein family, two binding pathways occur: one involving homo-dimerization of the protein before its interaction with the target DNA (Albright & Matthews, 1998; Darling et al, 2000), and the other the interaction of a monomer with the DNA target, followed rapidly by the capture of a second protein molecule to form a stable complex on the DNA target. In this latter case, the monomer is the binding species, and the second protein unit interacts cooperatively with the first protein to be bound. This pathway has been proposed for *Himar1* (Butler et al, 2006), for which evidence has been obtained for the binding of a monomer.

What about *Mos1*? No evidence for the direct binding of a monomer has ever been obtained. When monomers have been detected, they have always been associated with other complexes, suggesting the hypothesis that the monomer detected is in fact a dissociation product of higher complexes. In addition, gel filtration analyses (Richardson et al, 2006) indicate that dimers are the major MOS1 species in solution. Finally, the MOS1 T216A mutation results in a protein found only as a dimer in solution (Richardson et al, 2006), which is a hyperactive version of MOS1, both in bacteria and *in vitro* (Germon et al, 2009). The combination of these findings suggest that MOS1 dimerization is required before ITR binding can occur, as proposed (Auge-Gouillou et al, 2005a).

In view of our previous findings (Auge-Gouillou et al, 2005a), this suggests that differences in the binding modalities to the ITR among members of the MLE subfamilies might have emerged during evolution, by both full or partial domain exchanges with related counterparts, but also by accumulating mutations that can considerably modify both transposase interactions and ITR recognition.

Methods

DNA constructs

Genes encoding the yeast Gal4 dimerization domain (SPKTKRSPLTRAHLTEVESRLERLEQLFLLIFPREDLDMILKMDFLQDIKALLTGLFVQD; accession number: P04386), and the yeast GCN4 Leucine Zipper (ILPILEDKVEELLSKNYHLENEVASLKKLVGES; accession number: P03069) were synthesized by Entelechon (Germany). They are flanked by classic flexible linkers (MPGGGGSGGGGS/ /GGGGSGGGGSPG), and *Sma*I restriction sites. These domains were cloned in the pMal-c2 (New England Biolabs) vector at the *Sna*BI restriction site, giving pMal-Gal4 and pMal-LZ. These vectors were then used to insert the [35-345] part of MOS1 at the *Eco*RI and *Hind*III restriction sites. The final constructs were designated pMal-Gal4[35-345], and pMal-LZ[35-345], and they encode MBP-Gal4[35-345], and MBP-LZ[35-345] respectively.

Transposition assays in bacteria

Transposition reactions were performed in a bacterial transposition assay using pBC-3T3 as the donor DNA, and one of the pMal-Tpase expression vectors as the transposase source (Auge-Gouillou et al, 2001b). pBC-3T3 contained the pBR322 tetracycline-resistance gene (without promoter) flanked by two *Mos1* 3'ITRs. This plasmid was unable to confer resistance on *E. coli* in tetracycline concentrations of more than 8 µg/ml. Both pMal-Tpase and pBC-3T3 were used to co-transform JM109 competent cells, which were plated on agar containing LB-ampicillin (100 µg/ml) plus chloramphenicol (80 µg/ml). For each pMal-Tpase vector, five colonies were taken from a fresh co-transformation, and grown at 37°C for 1 hour in 1 ml LB, before inducing transposase expression with 1 mM IPTG. The cells were then grown for 5h at 32°C. Suitable dilutions of each reaction were plated with LB-tetracycline (12.5 µg/ml) agar to score for transposition, and with LB to titrate the bacteria. The transposition frequency was the number of tetracycline-resistant colonies divided by the total number of bacteria. Under these conditions, this method is able to detect transposition frequencies as low as 2.10^{-10}. Each experiment was repeated five times.

Insertion site analysis

Five tetracycline-resistant colonies from independent transposition assays were taken, and plasmids were extracted (Promega DNA purification system), diluted and used to transform DH5α-competent cells. This step was needed to purify the plasmids conferring the tetracycline-resistance phenotype. For each transformation, one tetracycline-resistant colony was taken, plasmids were extracted and analyzed using *Hind*III digest. Transposition signatures were verified by DNA sequencing, using primers anchored in the *cat*-coding gene of pBC (Crenes et al, 2009).

Acknowledgements

This work was supported by the University of Tours, le Centre National de la Recherche Scientifique, le Ministère de l'Education Nationale, de la Recherche et de la Technologie, and GDR CNRS n°2157. This work was also funded by grants from the European Commission (Project SyntheGeneDelivery, grant n°018716 to Y.B.), and the Association Française contre les Myopathies (grant n°11468 to C.A.G.). M. Ghosh (Plougasnou, FRANCE) revised the English text. The authors declare that they have no conflict of interest.

References

Albright RA, Matthews BW (1998) How Cro and lambda-repressor distinguish between operators: the structural basis underlying a genetic switch. *Proc Natl Acad Sci U S A* **95**(7)**:** 3431-3436

Auge-Gouillou C, Brillet B, Germon S, Hamelin MH, Bigot Y (2005a) Mariner Mos1 transposase dimerizes prior to ITR binding. *J Mol Biol* **351**(1)**:** 117-130

Auge-Gouillou C, Brillet B, Hamelin MH, Bigot Y (2005b) Assembly of the mariner Mos1 synaptic complex. *Mol Cell Biol* **25**(7)**:** 2861-2870

Auge-Gouillou C, Hamelin MH, Demattei MV, Periquet G, Bigot Y (2001a) The ITR binding domain of the Mariner Mos-1 transposase. *Mol Genet Genomics* **265**(1)**:** 58-65

Auge-Gouillou C, Hamelin MH, Demattei MV, Periquet M, Bigot Y (2001b) The wild-type conformation of the Mos-1 inverted terminal repeats is suboptimal for transposition in bacteria. *Mol Genet Genomics* **265**(1)**:** 51-57

Auge-Gouillou C, Notareschi-Leroy H, Abad P, Periquet G, Bigot Y (2000) Phylogenetic analysis of the functional domains of mariner-like element (MLE) transposases. *Mol Gen Genet* **264**(4)**:** 506-513

Bigot Y, Brillet B, Auge-Gouillou C (2005) Conservation of Palindromic and Mirror Motifs within Inverted Terminal Repeats of mariner-like Elements. *J Mol Biol* **351**(1)**:** 108-116

Brillet B, Bigot Y, Auge-Gouillou C (2007) Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica* **130**(2)**:** 105-120

Bryan G, Garza D, Hartl D (1990) Insertion and excision of the transposable element mariner in Drosophila. *Genetics* **125**(1)**:** 103-114

Butler MG, Chakraborty SA, Lampe DJ (2006) The N-terminus of Himar1 mariner transposase mediates multiple activities during transposition. *Genetica* **127**(1-3)**:** 351-366

Craig N, Craigie R, Gellert M, AM L (2002) *Mobile DNA II,* Washington DC. : ASM Press. .

Crenes G, Ivo D, Herisson J, Dion S, Renault S, Bigot Y, Petit A (2009) The bacterial Tn9 chloramphenicol resistance gene: an attractive DNA segment for Mos1 mariner insertions. *Mol Genet Genomics* **281**(3)**:** 315-328

Darling PJ, Holt JM, Ackers GK (2000) Coupled energetics of lambda cro repressor self-assembly and site-specific DNA operator binding I: analysis of cro dimerization from nanomolar to micromolar concentrations. *Biochemistry* **39**(37)**:** 11500-11507

Feldmann T, Alex R, Suckow J, Dildrop R, Kisters-Woike B, Muller-Hill B (1993) Single exchanges of amino acids in the basic region change the specificity of N-Myc. *Nucleic Acids Res* **21**(22)**:** 5050-5058

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* **9**(5)**:** 397-405

Germon S, Bouchet N, Casteret S, Carpentier G, Adet J, Bigot Y, Augé-Gouillou C (2009) Mariner Mos1 transposase optimization by rational mutagenesis. *Genetica* (revision)

Li F, Papworth M, Minczuk M, Rohde C, Zhang Y, Ragozin S, Jeltsch A (2007) Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. *Nucleic Acids Res* **35**(1)**:** 100-112

Pietrokovski S, Henikoff S (1997) A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. *Mol Gen Genet* **254**(6)**:** 689-695

Richardson JM, Dawson A, O'Hagan N, Taylor P, Finnegan DJ, Walkinshaw MD (2006) Mechanism of Mos1 transposition: insights from structural analysis. *Embo J* **25**(6)**:** 1324-1334

Shao H, Tu Z (2001) Expanding the diversity of the IS630-Tc1-mariner superfamily: discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. *Genetics* **159**(3)**:** 1103-1115

Watkins S, van Pouderoyen G, Sixma TK (2004) Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. *Nucleic Acids Res* **32**(14)**:** 4306-4312

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**(12)**:** 973-982

Zhang L, Dawson A, Finnegan DJ (2001) DNA-binding activity and subunit interaction of the mariner transposase. *Nucleic Acids Res* **29**(17)**:** 3566-3575

Legends to Figures

Figure 1. Proteins used in the study

Mos1 transposase is a 345-amino acid protein. Black bars represent truncated MOS1 (lacking amino acids 1 to 34), and colored boxes represent the fused dimerization domains: Gal4 (59 amino acids) in beige, and LZ (25 amino acids) in gray.

Figure 2. *In-vivo* transposition analyses

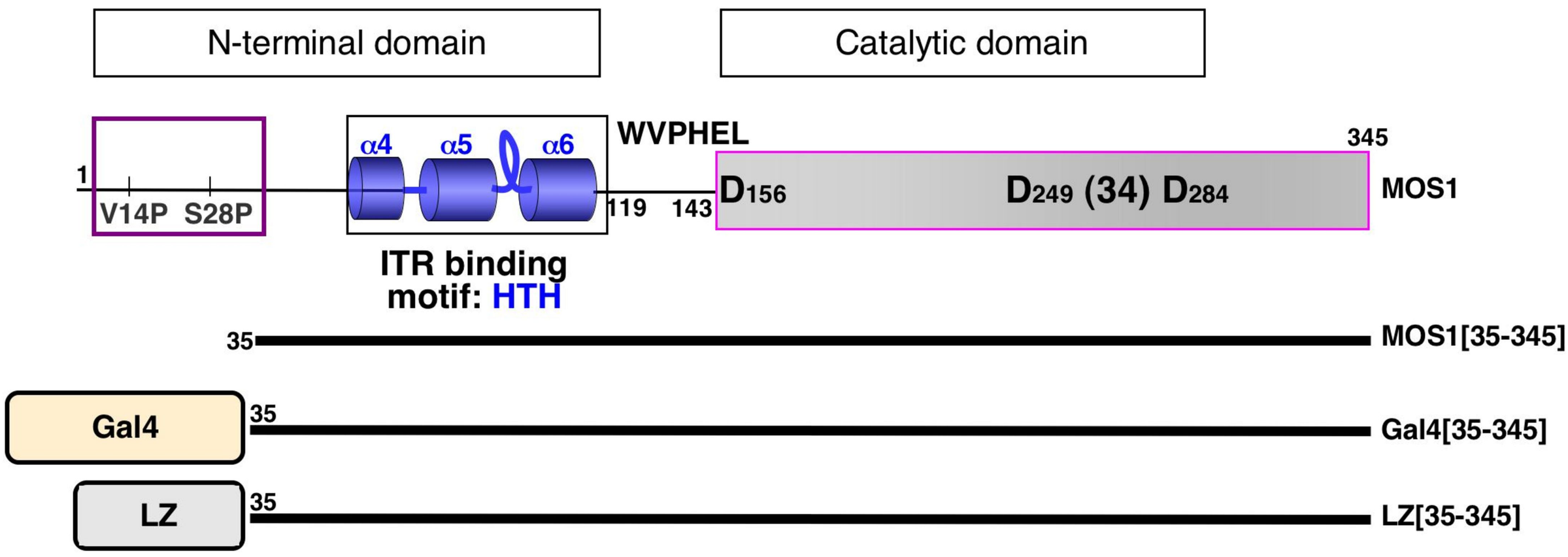
A. *In-vivo* transposition assays were performed using pBC-3T3 as the donor and target plasmid. Together with a plasmid encoding the transposase under investigation (pMal-Tpase), pBC-3T3 was co-transformed in bacteria. Transposition products were detected and recovered by promoter targeting. **B.** Analyses of Gal4[35-345] transposition products. After purification, plasmids were digested by *Hind*III, and analyzed on 1% agarose gel electrophoresis. MW: 1kb DNA steps ladder (Promega). T: pBC-3T3 (donor plasmid). Lanes 1 to 5: five independent clones from transposition assay. **C.** Insertion sites sequences. Target plasmids recovered in transposition assays were sequenced. Insertion sites are reported, with the duplicated TA (underlined), flanking the new inserted copy. All of them were located in the integration hot spot, the *cat* gene (Crenes et al, 2009). Only the four first bases of *Mos1* are indicated.

Tables

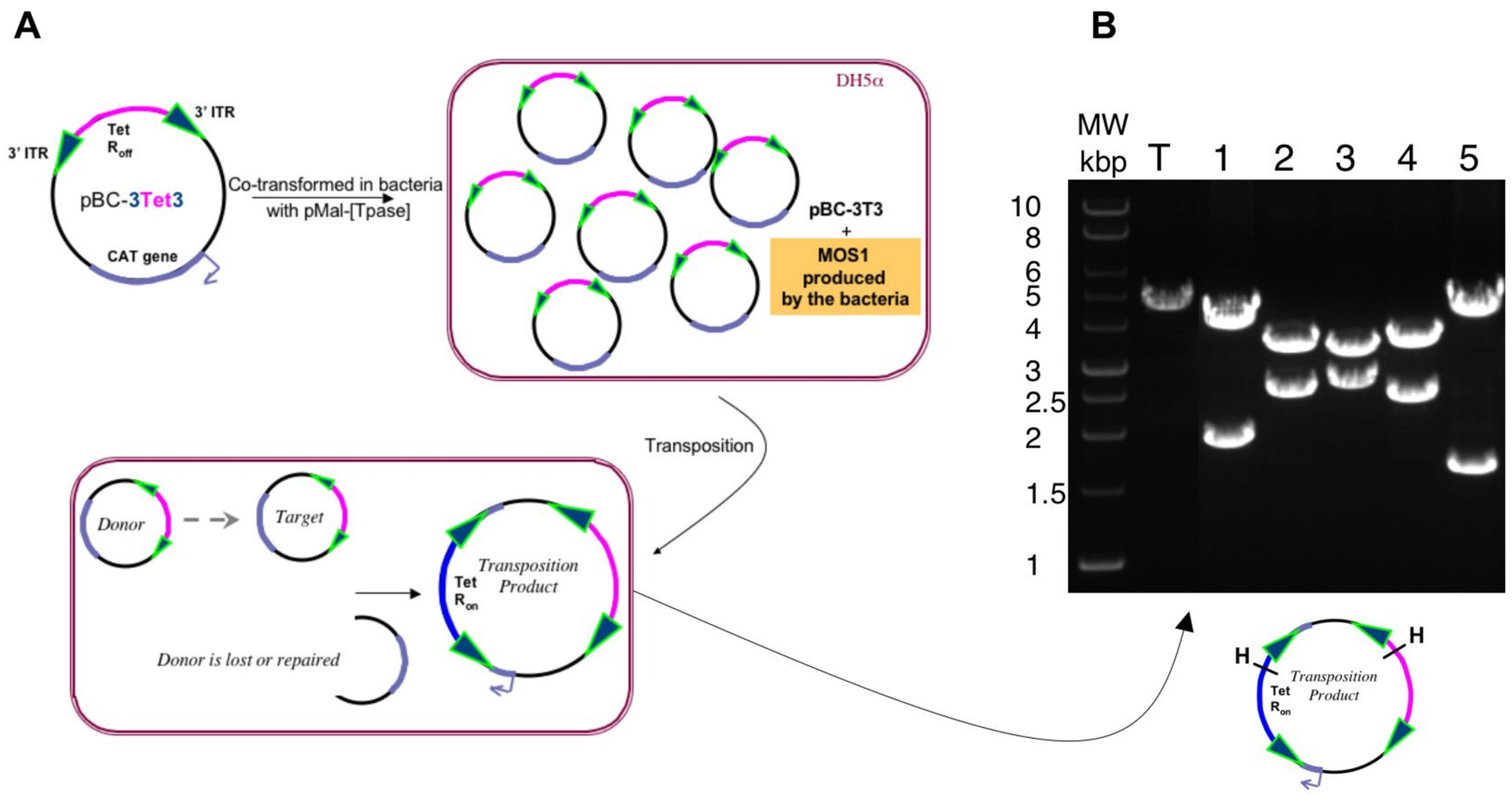
Table 1. Transposition frequencies (in bacteria) of the proteins used in the study are the median of at least five experiments. 1st quartiles are indicated in brackets.

Transposase	Transposition frequency
no	Undetected*
MOS1	10^{-4} [$3x10^{-5}$]
V14P	Undetected*
S28P	Undetected*
[35-345]	Undetected*
Gal4-[35-345]	$6.8x10^{-9}$ [$7x10^{-9}$]
LZ-[35-345]	$2x10^{-8}$ [$4x10^{-9}$]

*Sensitivity threshold: $2x10^{-10}$



Carpentier et al. Fig 1



C

3T3

pBC backbone Target ◀——▶ Target pBC backbone

with L[35-345]

TTTCTACACATA**TA**TCAGG-- // --CCTGA**TA**TGTGTAGAAACT

CTCTGGAGTGAA**TA**TCAGG-- // --CCTGA**TA**CCACGACGATTT

TACCACCGTTGA**TA**TCAGG-- // --CCTGA**TA**TATCCCAATGGC

GATGTGGCGTGT**TA**TCAGG-- // --CCTGA**TA**CGGTGAAAACCT

GAATTACAACAG**TA**TCAGG-- // --CCTGA**TA**CTGCGATGAGTG

with Gal4[35-345]

AGTTTCTACACA**TA**TCAGG-- // --CCTGA**TA**TATGTGTAGAAA

TGCTTAATGAAT**TA**TCAGG-- // --CCTGA**TA**CAACAGTACTGC

GTTGAATACTCA**TA**TCAGG-- // --CCTGA**TA**CTCTTCCCCCCC

TCCAGTTCGATG**TA**TCAGG-- // --CCTGA**TA**ACCCACTCGTGC

GTTGAATACTCA**TA**TCAGG-- // --CCTGA**TA**CTCTTCCTTTTT

Carpentier et al. Fig 2

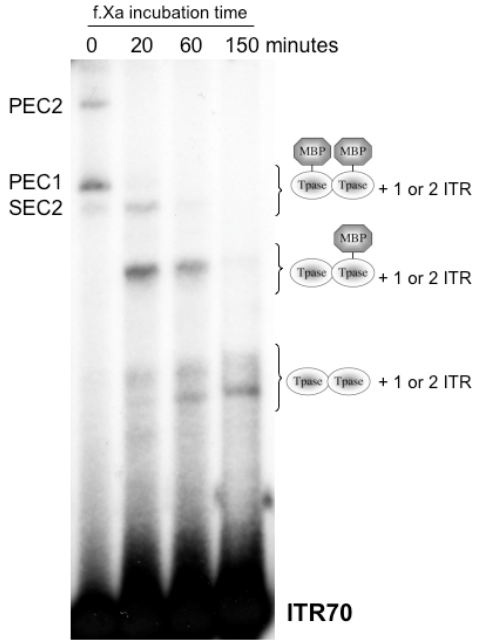
SUPPLEMENTAL INFORMATION

Carpentier et al: ***Mariner*** **mobility requires a single helix-turn-helix in the N-terminal transposase domain**

I. Features of PAIRED- and CRO-like domains, compared to those of the MOS1 ITR binding domain

	PAIRED domain	**CRO domain**	**MOS1**
Structure	2 HTH PAI + RED	1 HTH + dimerization structures	At least 1 HTH [72-110] + N-ter sub-domain
DNA binding domain	Bipartite PAI = (N-ter) specific. Able to bind DNA alone RED = (C-ter) unspecific. Unable to bind DNA alone	Monopartite HTH = specific. Unable to bind DNA alone	Monopartite HTH = specific. Unable to bind DNA alone
Dimerization sub-domain	No dimerization sub-domain identified	Dimerization sub-domain with α– and/or β– structures.	Identified sub-domain (aa: 1-34) of unknown structure
DNA target	No internal symmetry	Imperfect palindrome	Imperfect palindrome
DNA mode of binding on a single DNA	Monomer	Obligate dimer	Dimer (Sec2)
References	(Epstein et al, 1994; van Pouderoyen et al, 1997; Watkins et al, 2004; Xu et al, 1995)	(Jana et al, 1997; LeFevre & Cordes, 2003; Marmorstein et al, 1992)	(Auge-Gouillou et al, 2005a; Auge-Gouillou et al, 2005b; Bigot et al, 2005; Brillet et al, 2007; Pietrokovski & Henikoff, 1997; Richardson et al, 2006; Zhang et al, 2001)

II. MOS1 dimer interacting with *Mos1* ITR.



EMSAs were performed using 10 nM 3'ITR (ITR70, (Auge-Gouillou et al, 2005b)) and 400 nM MBP-MOS1 as probes. Mixtures were prepared in 5 mM $MgCl_2$, 100 ng BSA, 1 µg sonicated herring sperm DNA, and 10 mM Tris pH 9. After an incubation time of 20 minutes at 4°C (allowing complexes to be formed), the time courses for the digestion of complexes by factor Xa were done according to the manufacturer's guidelines (New England Biolabs).

According to the data previously obtained ((Auge-Gouillou et al, 2005b) Figure 3B), MOS1 interacts with the 3'ITR to form several complexes designated SEC2, PEC1 and PEC2 respectively. It had previously been proposed that PEC1 and PEC2 (paired end complexes) probably contain 2 ITR, and that SEC2 (a single end complex) contains a single ITR. In addition, it was previously proposed that PEC1 and SEC2 might contain an MOS1 dimer, whereas PEC2 might contain an MOS1 tetramer.

The occurrence of the MOS1 dimer in SEC2 and PEC1 is assayed here. We have taken advantage of the fact that MBP-MOS1 fusion proteins contain a factor-Xa cleavage site between MBP and MOS1. If SEC2 and PEC1 complexes contain the MBP-MOS1 dimer, then a three-band pattern would be expected following the cleavage of factor Xa : bands containing uncleaved MBP-MOS1 (in native SEC2 and PEC1, as seen at T=0), bands containing one cleaved and one uncleaved MOS1 simultaneously in complexes, and bands containing both cleaved MOS1 in complexes. This is what we found. Native PEC1 and SEC2 almost completely disappeared after 20 and 60 minutes of treatment respectively. At the same time, two groups of bands appear, corresponding to partially cleaved and totally cleaved complexes respectively. At the end of the assay, totally cleaved complexes accumulated, being the main species detected. Finally, PEC2 quickly disappears, but this disappearance does not result in additional bands, suggesting that PEC2 is sensitive to and dissociates in response to treatment with factor Xa. These data sustain that both SEC2 and PEC1 contain the MOS1 dimer.

III. References for supplemental information

Auge-Gouillou C, Brillet B, Germon S, Hamelin MH, Bigot Y (2005a) Mariner Mos1 transposase dimerizes prior to ITR binding. *J Mol Biol* **351**(1)**:** 117-130

Auge-Gouillou C, Brillet B, Hamelin MH, Bigot Y (2005b) Assembly of the mariner Mos1 synaptic complex. *Mol Cell Biol* **25**(7)**:** 2861-2870

Bigot Y, Brillet B, Auge-Gouillou C (2005) Conservation of Palindromic and Mirror Motifs within Inverted Terminal Repeats of mariner-like Elements. *J Mol Biol* **351**(1)**:** 108-116

Brillet B, Bigot Y, Auge-Gouillou C (2007) Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains. *Genetica* **130**(2)**:** 105-120

Epstein J, Cai J, Glaser T, Jepeal L, Maas R (1994) Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J Biol Chem* **269**(11)**:** 8355-8361

Jana R, Hazbun TR, Mollah AK, Mossing MC (1997) A folded monomeric intermediate in the formation of lambda Cro dimer-DNA complexes. *J Mol Biol* **273**(2)**:** 402-416

LeFevre KR, Cordes MH (2003) Retroevolution of lambda Cro toward a stable monomer. *Proc Natl Acad Sci U S A* **100**(5)**:** 2345-2350

Marmorstein R, Carey M, Ptashne M, Harrison SC (1992) DNA recognition by GAL4: structure of a protein-DNA complex. *Nature* **356**(6368)**:** 408-414

Pietrokovski S, Henikoff S (1997) A helix-turn-helix DNA-binding motif predicted for transposases of DNA transposons. *Mol Gen Genet* **254**(6)**:** 689-695

Richardson JM, Dawson A, O'Hagan N, Taylor P, Finnegan DJ, Walkinshaw MD (2006) Mechanism of Mos1 transposition: insights from structural analysis. *Embo J* **25**(6)**:** 1324-1334

van Pouderoyen G, Ketting RF, Perrakis A, Plasterk RH, Sixma TK (1997) Crystal structure of the specific DNA-binding domain of Tc3 transposase of C.elegans in complex with transposon DNA. *Embo J* **16**(19)**:** 6044-6054

Watkins S, van Pouderoyen G, Sixma TK (2004) Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. *Nucleic Acids Res* **32**(14)**:** 4306-4312

Xu W, Rould MA, Jun S, Desplan C, Pabo CO (1995) Crystal structure of a paired domain-DNA complex at 2.5 A resolution reveals structural basis for Pax developmental mutations. *Cell* **80**(4)**:** 639-650

Zhang L, Dawson A, Finnegan DJ (2001) DNA-binding activity and subunit interaction of the mariner transposase. *Nucleic Acids Res* **29**(17)**:** 3566-3575

2/ Matériel et Méthodes

2.1 Techniques courantes

2.1.1 Biologie moléculaire / ADN

La quasi-totalité des enzymes et kit utilisés pour la biologie moléculaire au cours de ma thèse ont été achetés chez *Promega* et *New England Biolabs*. Les techniques de biologie moléculaire classiques (préparation de plasmides, purification d'ADN, *etc*…) ont été réalisées selon les protocoles des fournisseurs.

Les électrophorèses d'ADN sont menées sur gel d'agarose 0,8%, 1X TAE (Tris-Acétate 0,04 M, EDTA 1 mM pH8), 1 µg/ml de bromure d'éthidium. La migration s'effectue dans un tampon TAE 1X entre 90 et 130 Volts selon la taille du gel. Le marqueur de taille classiquement utilisé est le phage λ digéré par HindIII. Les mesures de taille et de concentration d'ADN sont réalisées à l'aide d'un appareil automatique (*Syngene*) et les donnes traitées à l'aide des logiciels *Genesnap* et *Genetools*.

Les plasmides nécessitant une vérification de leur séquence ont été séquencés par *MWG Biotech*. Les séquences obtenues ont été analysées par les programmes *DNA Strider* et *ApE-A plasmid Editor*.

2.1.2 Bactéries et milieux de culture

Les **souches bactériennes** utilisées au long de ma thèse sont des *Escherichia coli JM109*, utilisées pour les tests de transposition, et des *E.coli DH5α*, utilisées pour les clonages ; ces bactéries ont été rendues électro-compétentes pour permettre leur transformation par choc électrique.

Leur préparation implique d'abord 3 repiquages des bactéries sur milieu BCC-agar ; 5 colonies sont alors mises en culture dans 2,5 ml de BCC, pendant 1 heure à 37° sous agitation (250 rpm). Ces 2,5 ml sont transvasés dans 25 ml de BCC, et incubés à 37°C sous agitation jusqu'à ce que la culture bactérienne atteigne une DO_{600nm}= 0,3. Les 25 ml sont alors

transvasés dans 250 ml de BCC, et mis en culture jusqu'à ce que DO_{600nm}= 0,3. La culture est alors mise dans la glace pendant 20 minutes.
A partir de là, tout se fait sous une température de 4°C avec du matériel préalablement refroidi. La culture bactérienne est alors centrifugée 10 min à 4000g à 4°C; le culot est resuspendu dans 250 ml de glycérol 10% froid, et le mélange est à nouveau centrifugé 10 min à 4000g à 4°C. L'opération est répétée avec 125 ml puis 10 ml de glycérol 10%. Le culot est alors repris dans 1 ml de glycérol 10%, et les bactéries sont aliquotées dans des tubes stériles par fractions de 45 µl. Les bactéries compétentes sont conservées à -80°C.
L'efficacité de transformation des bactéries, obtenue par une transformation par du pBS (plasmide porteur de la résistance à l'ampicilline) suivie d'un étalement sur LB-ampicilline, varie de 10^9 à 10^{10} transformants par µg d'ADN.

Les **milieux de culture** utilisés au laboratoire sont les suivants : LB (tryptone 1%, glucose 20 mM, extrait de levure 0,5%, NaCl 176 mM, NaOH 10 mM) ; LB/agar (LB et agar 1,5%), SOC (tryptone 2%, extrait de levure 0,5%, $MgCl_2$ 10 mM, $MgSO_4$ 10 mM, NaCl 10 mM, KCl 2,5 mM et glucose 20 mM) ; BCC (milieu cerveau-cœur (AES) 37 g/l).

Les antibiotiques sont utilisés aux concentrations suivantes : ampicilline (Sigma) 100 µg/ml, chloramphénicol (Boehringer) 80 µg/ml et tétracycline (Boehringer) 12,5 µg/ml.

2.1.3 Vecteurs

Le vecteur **pBC3T3** est un plasmide donneur de pseudo-transposon *mariner Mos1*, qui porte un gène de résistance à la tétracycline sans son promoteur, bordé par deux ITR 3' du transposon *Mos1* (Augé-Gouillou *et al*, 2001). Ce plasmide porte également un gène de résistance au chloramphénicol. L'ADN a été préparé en grande quantité selon le protocole de maxipréparation de Qiagen ; il a ensuite été traité à la protéinase K (SDS 0,13% final, protéinase K 0,13 µg/ml final), purifié au phénol-chloroforme, et précipité par de l'éthanol absolu.

Le **pBS** (pBlueScript II SK- de Stratagène) porte un gène de résistance à l'ampicilline.

Le **pBS (3')**, dérivé du plasmide pBS, porte l'ITR 3' du transposon *Mos1*. Un oligonucléotide ITR 30 double-brin avec des extrémités BamHI a été cloné au site BamHI du plasmide pBS.

Le **pMal-Tnp** est un vecteur qui exprime fortement la transposase de l'élément *Mos1*, via une induction à l'IPTG. Ce plasmide porte un gène de résistance à l'ampicilline.

Le **pVL** est le vecteur utilisé pour la production de protéines en système baculovirus/cellules d'insectes avec le kit BaculoGold (BS Sciences). La Tnp sauvage et les mutants de la Tnp produits en système eucaryote ont tous été clonés dans ce vecteur.

2.1.4 Protéines

La transposase (Tnp) de l'élément *Mos1* est fusionnée à la MBP (*Maltose Binding Protein*) ; cette protéine fusion est utilisée en routine au laboratoire. Elle est purifiée sur colonne d'amylose grâce à la partie MBP. L'activité de cette protéine fusion est identique à celle de la Tnp seule (Corinne Augé-Gouillou et Yves Bigot, communication personnelle).

2.1.5 Mutagenèse

Les mutants de la transposase ont été construits par mutagenèse des vecteurs d'expression pGemT-MBP-Mos1 ou pMal-MBP-Mos1. Une PCR est effectuée avec le vecteur d'expression contenant la transposase à muter et deux primers contenant la mutation voulue et qui s'annealent sur les brins opposés de la même séquence. Ces primers ont une taille comprise entre 25 et 45 bases, et le T_m des primers doivent être supérieur ou égal à 78°C. Le T_m est calculé par la formule suivante : $T_m = 81.5 + 0.41(\%GC) – 675/N - \%mismatch$, où le %GC correspond au pourcentage de nucléotides G et C dans la séquence du primer, N le nombre de nucléotides, et le %*mismatch* au pourcentage de bases différentes sur la séquence du primer par rapport à la séquence du vecteur à muter. La PCR est effectuée avec 10 ng du vecteur à muter, 125 ng de chaque primer, 1 mM de dNTP, 5 µl de *Pfu DNA Pol buffer*, 1 µl de *Pfu* polymérase, le tout dans un volume final de 50 µl. PCR : 30'' à 95°C, puis 16 cycles (30'' / 95°C + 1' à 55°C + 2'/kb du plasmide à 68°C). Le mélange de PCR est ensuite incubée 1h à 37°C avec l'enzyme DpnI, qui a la propriété de couper l'ADN méthylé, et va donc éliminer le plasmide de départ. 5 µl du mélange sont ensuite transformés dans des bactéries *E.coli DH5α* électro-compétentes et les mutants sont sélectionnés sur ampicilline, les vecteurs d'expression de la Tnp possédant le gène de résistance à l'ampicilline.

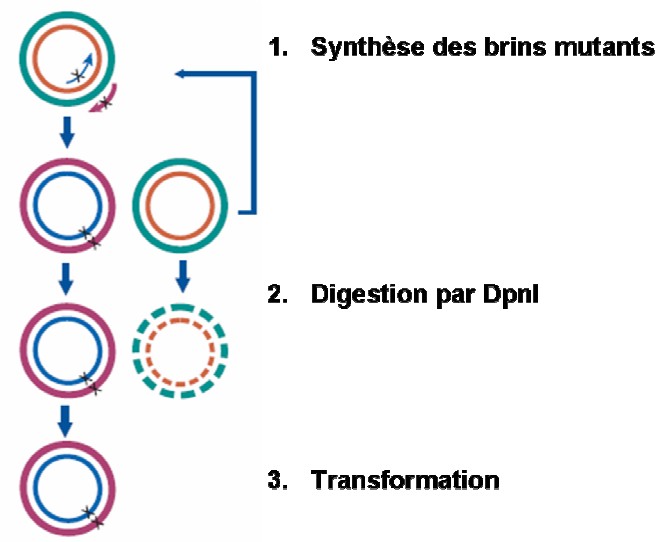


Figure A1. Schéma de la mutagenèse dirigée. (1) *Les brins mutants sont synthétisés par une PCR à la Pfu polymérase effectuée avec des oligonucléotides contenant la mutation désirée.* **(2)** *L'enzyme DpnI est ajoutée au mélange, et les plasmides d'origine, non porteurs de la mutation, sont digérés. Seuls restent dans le tube les produits de PCR, qui sont alors transformés dans des bactéries électro-compétentes* **(3)**.

2.2 Composés chimiques testés sur la transposase

Tous les composés ont été dissous dans le diméthyl-sulfoxide (DMSO), qui, aux doses utilisées, ne présente aucune activité inhibitrice sur la transposase MOS1, et aucune cytotoxicité sur les bactéries *E. coli*.

2.2.1 Composés de la Faculté de Pharmacie de Tours

111 composés testés proviennent des collections de produits de synthèse organique des laboratoires de chimie thérapeutique de la Faculté de Pharmacie de Tours :

- 73 sont issus du laboratoire du Pr Marie-Claude Viaud (GICC – UMR CNRS 6239, équipe 5) et sont numérotés MCV1 à MCV73. La majorité d'entre eux sont des dérivés maléimides

- 38 sont issus du laboratoire de Physicochimie des Matériaux et des Biomolécules (EA 4244) et nous ont été fournis par le Pr Cécile Enguehard-Gueiffier. Ils sont nommés CG1 à 38. Les 30 premiers sont des composés hétérocycliques issus de synthèses organiques effectuées dans le laboratoire, les 8 autres sont des composés naturels issus de plantes africaines.

2.2.2 Inhibiteurs de la transposase de Tn*5* et/ou de l'intégrase du VIH-1

Il s'agit de 27 composés décrits par Ason *et al* (2005), tous achetés chez Chembridge. Les composés R1 à R20 sont des inhibiteurs de la transposase de Tn*5*, six d'entre eux inhibant également l'intégrase du VIH-1 *in vitro*. Le composé R10 présente une activité *in vivo* sur l'intégrase du VIH-1, tout comme 7 composés dérivés du R10, nommés R10-A à R10-G.

Parmi ces composés aromatiques, certains sont liés structuralement : les dérivés de la coumarine (substance naturelle organique aromatique retrouvée chez les végétaux), les dérivés du cinnamoyle (composé naturel intervenant dans la voie de synthèse des polyphénols végétaux) et les dérivés de l'acide benzoïque (figure A2). Les autres apparaissent comme étant uniques.

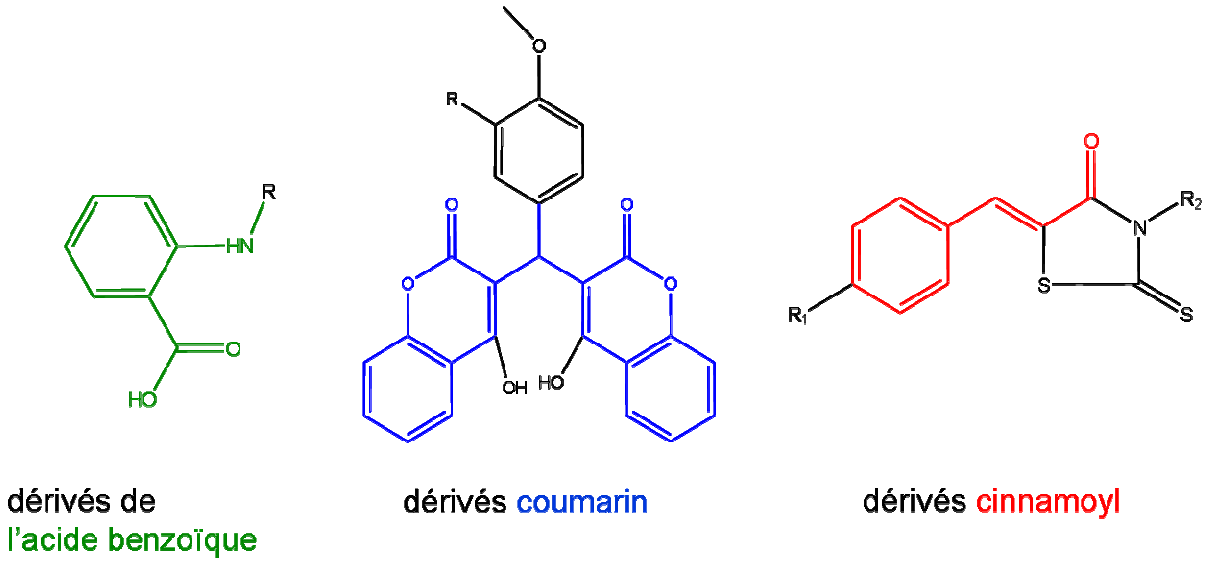


Figure A2. Structure des inhibiteurs de Tn*5*.

2.2.3 Composés divers

37 composés ont des origines diverses. Certains possèdent des propriétés inhibitrices de l'intégrase du VIH-1 ou présentent des caractéristiques structurales similaires à des inhibiteurs de l'intégrase ; d'autres n'ont aucune activité inhibitrice connue. Ils sont nommés de D1 à D37.

Parmi ces composés, il y a 16 thiazole-carboxamides, connus pour être des intercalants de l'ADN (Ryabinin *et al*, 2000), ainsi que 5 thiazolothiazépines présentant des similitudes avec des inhibiteurs de l'intégrase du VIH-1 (Aiello *et al*, 2004) (figure A3). 4 sont des

oligonucléotides riches en guanines formant des structures tétramériques appelées G-quartets (figure A3) et qui ont été décrits comme inhibiteurs de l'intégrase du VIH-1 (Jing *et al*, 2000a et 2000b ; Soultrait *et al*, 2002). L'acide 3,5-dicaféoylquinique, extrait de la patate douce (Islam *et al*, 2002), et l'*Integrase Inhibitor 3* de Merck font également parti de ces composés.

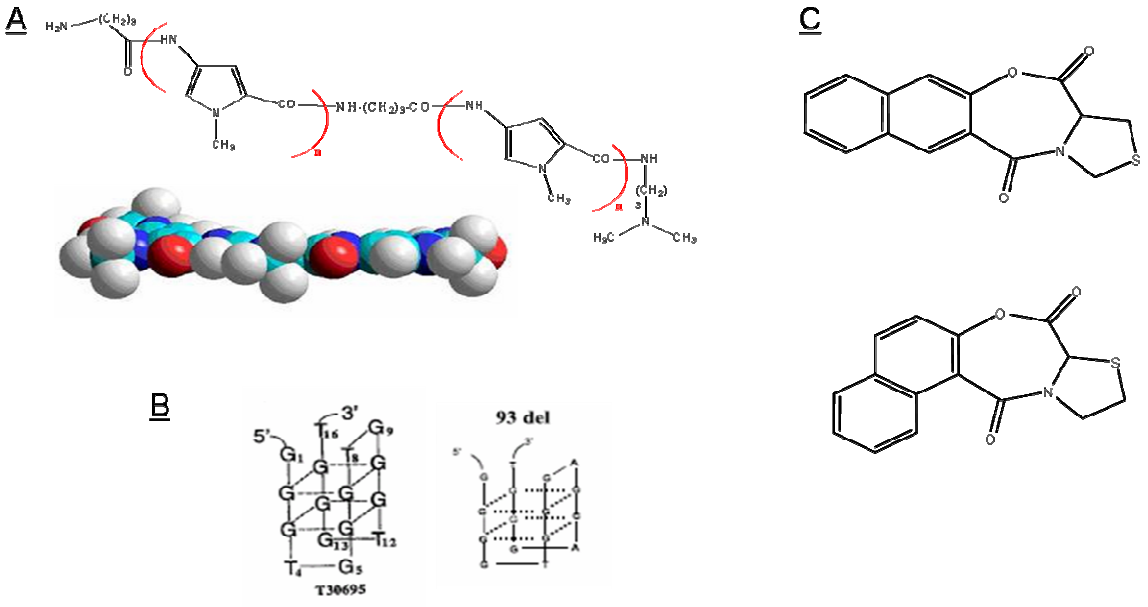


Figure A3. Structure des divers composés. (A) *Formule générale des thiazole-carboxamides et structure tridimensionnelle d'un carboxamide qui montre qu'il s'agit de composés de structure plane.* **(B)** *Représentation des oligonucléotides 93del et T30695 formant des G-Quartets (Jing et al, 2000a ; Soultrait et al, 2002)* **(C)** *Exemple de deux thiazolothiazépines, qui présentent des structures similaires à des inhibiteurs de l'intégrase du VIH-1.*

2.3 Analyses protéiques

2.3.1 Production de la transposase

La **transposase eucaryote** a été produite au laboratoire par le Dr Guillaume Carpentier, selon le système baculovirus / cellules d'insectes en utilisant le système BaculoGold (*BD Biosciences*) et selon les instructions du fournisseur. La présence du tag MBP permet la purification par une résine de maltose.

La **transposase procaryote** a été produite selon le protocole suivant. La MBP-Tnp (sauvage ou mutante) a été clonée dans le plasmide pMal-C2 (*New England Biolabs*). Une colonie issue d'une transformation dans des *E.coli DH5α* est mise en culture dans 2,5 ml de

LB-ampicilline (100 µg/ml) pendant 16 heures sous agitation à 37°C. Une dilution au 1/100 est ensuite inoculée dans 50 ml de LB-ampicilline (100 µg/ml) et incubée à 37°C sous agitation jusqu'à DO_{600} = 0,4. La production de MBP-Tnp est induite avec 1mM d'IPTG, et laissée en culture à 37°C sous agitation pendant 2h30. La culture est ensuite centrifugée 15 minutes à 6000 rpm à 4°C. Une fois le surnageant éliminé, le culot est congelé à -20°C sur la nuit, décongelé à température ambiante puis repris dans 5 ml de tampon (20 mM Tris pH 9, 100 mM NaCl, 1 mM DTT). La paroi bactérienne est éclatée par ajout de 0,5 mg/ml de lysozyme pendant 30 minutes à 4°C. Le mélange est centrifugé 20 minutes à 10000g. Le surnageant, qui correspond à l'extrait brut (EB), est récupéré ; 100µl d'EB sont dosés selon la méthode Bradford contre une gamme de BSA (0 à 10 µg).

La purification s'effectue ensuite grâce au tag MBP qui se trouve fusionné à la partie N-terminale de la transposase. Toutes les étapes s'effectuent à 4°C. 500 µl de résine de maltose (*New England Biolabs*) sont équilibrés avec 2 x 10 volumes de tampon (20 mM Tris pH 9, 100 mM NaCl, 1 mM DTT). La résine est culottée par décantation et le tampon est éliminé. Les 5 ml d'EB sont mélangés avec 150µl de résine de maltose puis mis à tourner sur une roue pendant 15 minutes. Le mélange est ensuite laissé décanter dans la glace. Le surnageant est récupéré (il correspond au flow). La résine est lavée à 5 reprises par 1 ml de tampon (20 mM Tris pH 9, 100 mM NaCl, 1 mM DTT) et une centrifugation de 10 secondes. Puis la protéine est éluée 3 fois : par ajout de 150 µl de tampon (20 mM Tris pH 9, 100 mM NaCl, 1 mM DTT, 10 mM maltose) puis décantation dans la glace. Trois surnageants sont donc récupérés et dosés selon la méthode Bradford.

2.3.2 Gel SDS-PAGE

Les électrophorèses sur gel de polyacrylamide (*concentration* : 4% acrylamide-bisacrylamide 30:0,83 (*Promega*), 0,4% SDS, 0,5 M Tris-HCl pH 6,8 ; *séparation* : 10% acrylamide/bisacrylamide 30:0,83, 0,4% SDS, 1,5 M Tris-HCl pH 8,8) ont été réalisées dans du tampon de migration SDS 1X (0,1% SDS, 25 mM Tris, 190 M glycine) sous un ampérage de 45 mA. Avant dépôt, les échantillons protéiques sont incubés 10 minutes à 100°C dans du tampon de dépôt (0,125 M Tris pH 6,8, 2% SDS, 20% glycérol, 0,005 % bleu de bromophénol, 1% β-mercaptoéthanol).

2.3.3 Coloration au bleu de Coomassie

Après migration, le gel de polyacrylamide est incubé dans une solution de bleu de Coomassie (50% méthanol, 10% acide acétique, 0,01% bleu de coomassie R250) pendant 10 minutes. Puis il est décoloré par des bains successifs d'une solution de décoloration (10% méthanol, 12% acide acétique).

2.3.4 Western Blot

Après migration, les protéines sur le gel de polyacrylamide sont transférées sur une membrane de nitrocellulose (Hybond, *Amersham Biosciences*). Les transferts sont effectués avec l'appareil Trans Blot SD (*BioRad*). La membrane est préalablement équilibrée 30 minutes dans du tampon de transfert (15% méthanol, 0,1% SDS, 48 mM Tris pH 8,8, 39 mM glycine), de même que le gel et 4 papiers Whatman. Le transfert des protéines est effectué après avoir déposé dans l'ordre sur l'anode deux papiers Whatman, la membrane, le gel et deux papiers Whatman. Le transfert a lieu pendant 30 minutes, à 10 volts. La membrane est saturée par 50 ml de PBS-lait (5% lait) pendant 1 heure. Elle est ensuite incubée 2 heures avec l'anticorps anti-MBP (1/50000) dans du PBS-tween-lait (0,05% tween, 5% lait), puis lavée par 25 ml PBS-tween (0,05% tween) trois fois pendant 5 minutes. La membrane est ensuite incubée avec l'anticorps anti-IgG de lapin couplé à la peroxydase (1/50000) pendant 1 heure, puis à nouveau lavée pendant 5 minutes trois fois par 25 ml de PBS-tween (0,05% tween). La révélation des protéines reconnues par l'anticorps anti-MBP est faite à l'aide du kit ECL Advanced (*GE Healthcare*) selon le protocole du fournisseur par l'imageur LAS-4000 (*Fuji*).

2.4 Tests de transposition

2.4.1 Test de transposition *in vitro*

Ce test permet de détecter les évènements de transposition par acquisition de la résistance à la tétracycline, le transposon *Mos1* ayant en effet été remplacé par un pseudo-transposon dans lequel le gène de résistance à la tétracycline remplace le gène codant la transposase, le tout flanqué de deux ITR 3'. Ce pseudo-transposon a été inséré dans le vecteur pBC, qui confère la résistance au chloramphénicol ; le gène *cat* de résistance au chloramphénicol est une cible préférentielle pour l'insertion de *Mos1* (Augé-Gouillou *et al*, 2001 ; Crénès *et al*, 2009). Ce nouveau plasmide, le pBC3T3 est donc à la fois le donneur et

le receveur ; la présence de transposase apportée en *trans* conduit à l'insertion du pseudo-transposon dans un pBC3T3 receveur, très majoritairement au niveau du gène *cat*. La transposase est apportée en *trans*, ce qui permet l'excision du pseudo-élément et sa réinsertion. Les évènements de transposition se font en majorité dans le gène *cat*, donc en aval d'un promoteur. Le mélange réactionnel est ensuite transformé dans des bactéries électro-compétentes *E.coli* JM109, l'efficacité de transformation étant vérifiée par l'ajout du plasmide pBS, porteur de la résistance à l'ampicilline. Les bactéries sont sélectionnées sur LB-ampicilline (l'efficacité de la transformation), LB-chloramphénicol (normalisation par rapport à la quantité d'ADN) et LB-tétracycline (détection des évènements de transposition).

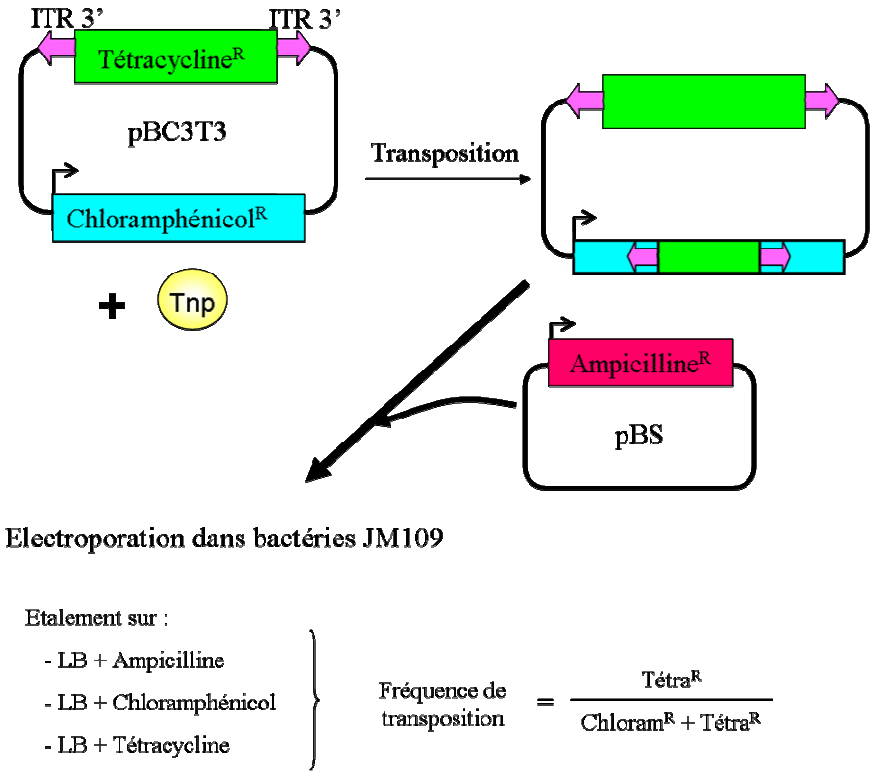


Figure A4. Principe du test de transposition *in vitro*. *Le plasmide pBC3T3 est porteur d'un pseudo-transposon composé de l'ORF, sans promoteur, du gène de résistance à la tétracycline bordé par les ITR 3' du transposon Mos1, et du gène cat de résistance au chloramphénicol. La Tnp purifiée ajoutée au mélange réactionnel va se fixer sur les ITR du pseudo-transposon, provoquant son excision et sa réinsertion préférentielle dans le gène cat. Les plasmides, ayant ou non subi un évènement de transposition, sont cotransformés avec le pBS par électroporation dans des bactéries E. coli JM109 électro-compétentes. Les bactéries sont sélectionnées sur les milieux LB-ampicilline, LB-chloramphénicol et LB-tétracycline.*

Le protocole est le suivant : le mélange réactionnel contient 1 mM Tris-HCl (pH 9), 5 mM NaCl, 20 mM $MgCl_2$, 0,5 mM EDTA, 5 ng/ml BSA, 600 ng de plasmide pBC3T3 super-enroulé et 80 nM de transposase dans un volume final de 20 µl. Le mélange est incubé 30

minutes à 30 °C. 10 µl d'une solution stop (0,4 % SDS, 0,4 µg/µl de protéinase K) est ajoutée pour stopper la réaction, puis le mélange est incubé 10 minutes à 65°C puis 30 minutes à 37°C. Les produits de réaction sont extraits par phénol-choloroforme, puis précipités avec de l'éthanol 100% avec 1 µg d'ARN de levure. 10 % du mélange réactionnel est cotransformé avec 0,01 ng de pBS (utilisé comme témoin positif de la transformation) dans des bactéries *E.coli JM109* électro-compétentes. Les bactéries sont mises en culture 1h à 37°C dans du milieu SOC. 25 µl sont étalés sur LB-ampicilline (100 µg/ml), 80 µl d'une dilution au 1/1000 sur LB-chloramphénicol (80 µg/ml) et le reste sur LB-tétracycline (12,5 µg/ml). L'efficacité de la transformation est vérifiée par le nombre de bactéries résistantes à l'ampicilline. Le taux de transposition correspond au nombre de bactéries résistantes à la tétracycline divisé par le nombre de bactéries totales (bactéries résistantes au chloramphénicol et à la tétracycline).

2.4.2 Test de transposition *in vivo*

Le test de transposition *in vivo* permet de détecter les évènements de transposition qui ont eu lieu en bactérie, par acquisition de la résistance à la tétracycline. Le principe du test est illustré dans la figure A5. Dans ce test, le rapport entre le nombre total de bactéries résistantes à la tétracycline et le nombre total de bactéries est utilisé pour calculer une fréquence relative de transposition.

Des bactéries compétentes *E. coli* JM 109 sont cotransformées par électroporation par les plasmides pMal-Tnp et pBC3T3. Elles sont étalées sur un milieu LB-chloramphénicol-ampicilline, et incubées la nuit à 37°C. Le lendemain, 5 clones sont inoculés dans 1 ml de LB. Les cultures bactériennes sont cultivées 1 heure à 37°C sous agitation (250 rpm). A T_0, les cultures sont mises en présence d'1mM d'IPTG pour induire la synthèse de la transposase. Elles sont ensuite incubées 5 heures à 32°C sous agitation (250 rpm). A T_0, 20 µl sont étalés sur milieu LB-tétracycline, et 20 µl d'une dilution au 1/1000 sur milieu LB. A T_5, 20 µl sont étalés sur milieu LB-tétracycline, et 50 µl d'une dilution au 1/50000 sont étalés sur milieu LB. Les boîtes sont ensuite incubées à l'étuve à 37°C sur la nuit, puis comptées le lendemain matin.

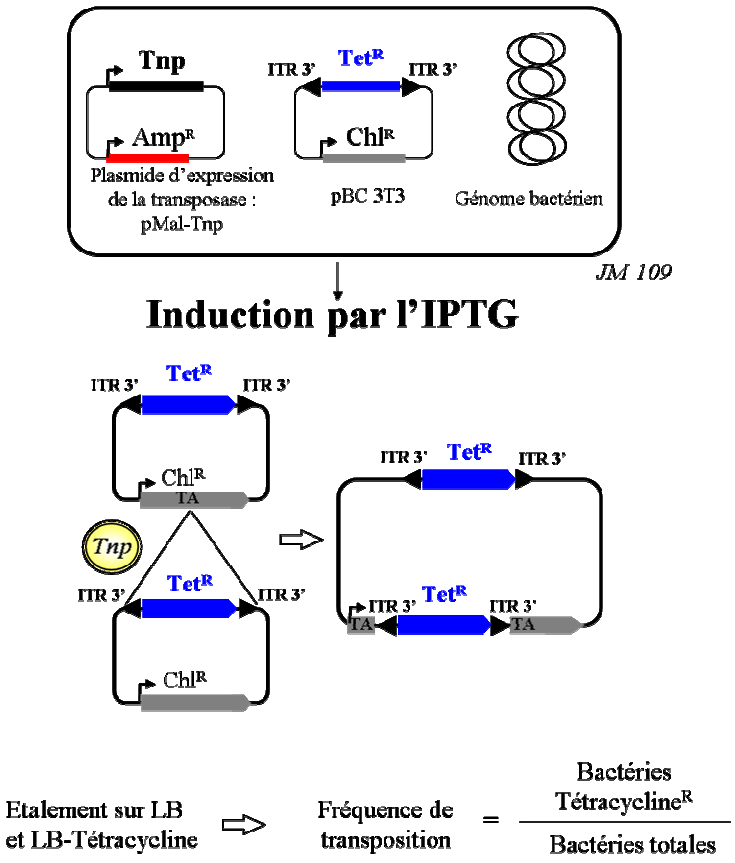


Figure A5. Principe du test de transposition in vivo. *Des bactéries E.coli JM109 sont cotransformées avec le plasmide pMal-Tnp et le pBC3T3, porteur du pseudo-transposon. Le pMal-Tnp porte un gène de résistance à l'ampicilline et le gène de la Tnp, placé sous le contrôle d'un promoteur inductible à l'IPTG. Après induction et synthèse, la Tnp va pouvoir se fixer sur les ITR du pseudo-transposon porteur d'un gène de résistance à la tétracycline, provoquant ainsi son excision et sa réinsertion. Les bactéries sont ensuite étalées sur LB et LB-tétracycline, et la fréquence de transposition est calculée par le nombre de bactéries résistantes à la tétracycline sur le nombre total de bactéries.*

2.5 Etudes des complexes ADN-protéine

2.5.1 Marquage radioactif des sondes ITR 3'

Pour le travail sur les inhibiteurs de la transposase MOS1, les expériences de retard sur gel ont été effectuées avec un ITR 3' long de 70 pb, excisé du plasmide pBS (3') par digestion EcoRI/XbaI dans un volume finale de 50 µl. Ces 50 µl de produit de digestion sont ensuite marqués en utilisant la DNA polymérase Klenow (*Promega*), de α^{32}P dATP, et du dCGT non marqué. La sonde radiomarquée est ensuite purifiée sur gel de polyacrylamide (6%) et resuspendu au final dans 50 µl de TE 1X.

Toutes les autres sondes ITR utilisées dans les travaux présentés dans ce mémoire ont été marquées avec du γ^{32}P avec de la T4 polynucléotide kinase (*Promega*). L'ITR classiquement utilisé est l'ITR UC (pour *UnCleaved*) long de 48 pb. La séquence de cet ITR ainsi que celle des ITR PC (pour *PreCleaved*) et PN (pour *PreNicked*) sont présentées dans la figure A6. L'ITR 260, utilisé lors des expériences d'ITR court/long, a été amplifié avec de la *Taq DNA polymérase* à partir du pBS (3') comme recommandé par le fournisseur (*Promega*)

et utilisant des primers universels. Le produit de PCR a été purifié sur gel d'agarose puis sa concentration a été estimée sur gel d'agarose.



Figure A6. Séquence des ITR UC, PN et PC. *La séquence interne à Mos1 est représentée en rouge. L'ITR UC (pour UnCleaved) correspond à l'ITR 3' de Mos1 avec de l'ADN flanquant à son extrémité. L'ITR PN correspond à un ITR UC ayant subi une première coupure sur le brin non-transféré. L'ITR PC correspond à l'ITR UC ayant été clivé sur ses deux brins par la transposase.*

Les ITR UC, PC et PN ont été marqués en utilisant de la T4 polynucléotide kinase (*Promega*) avec du γ^{32}P dans un volume final de 50 µl, comme recommandé par le fournisseur. Ils ont ensuite été purifiés à l'aide de colonnes MicroSpin G-25 (*GE Healthcare*).

2.5.2 Retard sur gel

Les réactions de liaison de la Tnp à l'ITR ont été menées avec 400nM de transposase et 20 nM de sonde ITR radiomarquée, dans 5 mM de $MgCl_2$, 50 nM NaCl, 10 mM Tris pH9, 5% de glycérol et 500 ng de BSA, dans un volume final de 20 µl. Le mélange est incubé à 4°C pendant 20 minutes. Les produits de réaction sont séparés sur un gel natif de 5% polyacrylamide (30 :0,93, acrylamide-bisacrylamide - *Promega*) - 0,5X TBE (22 nM Tris-22mM acide borique-0,6 mm EDTA). La migration s'effectue pendant 2h à 220 Volts, puis le gel est séché. Les gels sont ensuite révélés par un phospho-imager (Storm 860, *Amersham Biosciences*).

2.6 Etude de la phosphorylation de MOS1

2.6.1 ProQ Diamond

Cette technique permet de détecter si une protéine est phosphorylée ou non par coloration d'un gel de polyacrylamide à l'aide du colorant ProQ Diamond *(Molecular Probes).* Le protocole de coloration est celui décrit par le fournisseur.

Pour chaque protéine phosphorylée, l'intensité du signal est corrélée avec le nombre de groupement phosphates de manière linéaire. Le nombre de phosphates de l'ovalbumine, protéine phosphorylée du marqueur PeppermintStick (fourni avec le colorant), est connu (2 phosphates par molécule), de même que la concentration de la protéine dans le marqueur. Ceci permet de calculer approximativement le nombre de phosphates pour les molécules testées.

Complémentaire avec le colorant *ProQ Diamond*, le kit de relargage de phosphates *Serine/Threonine Phosphatase Assay System* de *Promega*, qui permet de vérifier l'activité des phosphatases sur un peptide phosphorylé.

2.6.2 Déphosphorylation de MOS1

Les réactions de déphosphorylation de MOS1 ont toutes été menées en suivant les indications des fournisseurs (*NEB* pour la PPase1, la λPPase et la LAR PPase ; *Promega* pour la PPase2A), en présence de 4 µg de transposase, dans un volume finale de 15 µl, 1h30 à 30°C. Ces conditions permettent d'obtenir des protéines entièrement déphosphorylées avec la PPase2A, la PPase1 et la λPPase.

3/ Catalogue des mutants de MOS1

Le catalogue des mutants de MOS1 présenté ici correspond aux mutants du laboratoire et aux mutants décrits dans la littérature. La catégorie ‘mutants simples’ regroupe des mutants ponctuels ainsi que des mutants doubles ou triples qui ont leurs mutations situés dans le même domaine de MOS1. La catégorie ‘mutants multiples’ regroupe quant à elle des protéines avec deux ou plusieurs mutations se trouvant dans des domaines différents de la protéine.

Les domaines sont présentés de différentes couleurs : le domaine de dimérisation est en bleu foncé, le domaine de liaison à l’ADN en bleu clair et le domaine catalytique en jaune. Les *linkers* et parties dont les fonctions ne sont pas clairement définies sont en blanc.

Les références correspondent soit à des membres du laboratoire (BB : Benjamin Brillet ; CAG : Corinne Augé-Gouillou, JB : Julien Bischerour ; NB : Nicolas Bouchet, SG : Stéphanie Germon) soit à des données publiées (Germon *et al*, 2009 ; Zhang *et al*, 2001).

Mutants simples	Transposition *in vivo* (% wt)	Référence	Observations
T11V	60-120	NB	variabilité des résultats de transposition *in vivo*
T11E	1	NB	
V14P	0	CAG	protéine protéolysée
S28D	0,3	NB	
S28P	0	CAG	protéine protéolysée
S28V	40	NB	disponible en version eucaryote
F53Y	50-60	SG ; Germon (2009) ; Zhang (2001)	mutant d'interaction Tnp-Tnp ; protéine protéolysée
S55C	30	BB	
S55D	80	JB	
S55V	40	NB	
D57A	25	NB	
D59A		BB	protéine protéolysée
D61A		BB	protéine protéolysée
D62A		BB	protéine protéolysée
L77A		BB	SEC2t et SEC2b observable en retard sur gel
L80A		BB	
L80E		JB	SEC2t et SEC2b observable en retard sur gel
L81A		BB	SEC2 très dur à visualiser en retard sur gel
L81H		JB	
D82A	200	BB	
E83A		BB	
D84A		BB	
D85A		BB	
T88A	1	BB	SEC2t et SEC2b observable en retard sur gel
T88C	4	BB	protéine protéolysée
T88E	0	BB	SEC2t et SEC2b observable en retard sur gel
T88R	0	BB	SEC2 très dur à visualiser en retard sur gel
T88V	20	BB	protéine protéolysée ; disponible en version eucaryote
T88C/S104A	1,4	NB	
T88C/S104V	2	NB	
T88V/S104V	0	NB	
Q91R	150	SG ; Germon (2009) ; Zhang (2001)	mutant d'interaction Tnp-Tnp ; hyperactif
L92A	0	SG ; Germon (2009)	protéine protéolysée
L92H	10	CAG	aucun complexe en retard sur gel ; protéine protéolysée
L92R	0	CAG ; Germon (2009)	protéine protéolysée
A93V	0,1	CAG	aucun complexe en retard sur gel ; protéine protéolysée

Mutants simples	Transposition *in vivo* (% wt)	Référence	Observations
S99A	0	BB	
S99C	2	BB	
S99D	0	BB	
S99V	0,2	NB	disponible en version eucaryote
S99V/Q100H/A102R		NB	
Q100E	0	SG ; Germon (2009)	aucun complexe visualisé en retard sur gel
Q100H	0	NB	
Q100N	0	SG ; Germon (2009)	
Q100R	0	SG ; Germon (2009)	
Q100R/Q101V	0	NB	réduit liaison à l'ADN, et interaction Mos1-Mos1
Q100H/Q114R		NB	
Q100R /Q101V/A102R		NB	
A102S		JB	
S104A	0,6	BB	
S104C	0	BB	
S104D	0	BB	aucun complexe en retard sur gel
S104E	0	BB	
S104K	15	NB	
S104N	0	BB	
S104P	0	CAG ; Zhang (2001)	aucun complexe en retard sur gel
S104V	15	NB	disponible en version eucaryote
S104Y	0	BB	
S104V/E109V		NB	
S104T/R106I		NB	
N105A	100	SG ; Germon (2009)	
R106A		Zhang (2001)	aucun complexe en retard sur gel
R106I	0	NB	
K112A		Zhang (2001)	profil en retard sur gel identique à la protéine sauvage
V120A	10-30	CAG ; Zhang (2001)	
L124S	0	CAG ; Zhang (2001)	insertions pas spécifiques au TA : résidu impliqué dans la capture de cible ?
T135A	85	BB	
T135E	12	BB	
C136S		NB	disponible en version eucaryote
E137K	230	SG ; Germon (2009)	
S147G	1,5	BB	aucun complexe visualisé en retard sur gel

Mutants simples	Transposition *in vivo* (% wt)	Référence	Observations
W159R	0	CAG ; Zhang (2001)	
L160T	0,1	CAG ; Zhang (2001)	
S170A	4,3	BB	
S170D	0	BB	profil en retard sur gel identique à la protéine sauvage
T216A	1600	SG ; Germon (2009) ; Richardson (2006)	majoritairement dimérique en solution utilisé pour la cristallographie de MOS1
Y237C	300	SG	
W268A	0	SG	
W268F	0	SG	
W268Y	0	SG	
D279G	0	CAG ; Zhang (2001)	
D284N		SG	
S302P	0,3	CAG ; Zhang (2001)	aucun complexe en retard sur gel
S305A	51	BB	
S305D	30	BB	
C336S		NB	disponible en version eucaryote

Mutants multiples	**Transposition *in vivo* (% wt)**	**Référence**
F53Y/Q91R	1400	SG
F53Y/T216A	4400	SG ; Germon (2009)
F53Y/Y237C	1200	SG
F53Y/E137K/T216A	très variable	SG ; Germon (2009)
F53Y/T216A/Y237C	2100	SG ; Germon (2009)
F53Y/Q91R/E137K/T216A	6100	SG ; Germon (2009)
F53Y/E137K/T216A/Y237C	6100	SG ; Germon (2009)
F53Y/Q91R/E137K/T216A/Y237C	3500	SG ; Germon (2009)
S55D/S147D		JB
Q91R/Y237C	1200	SG
Q91/E137K/T216A	0	SG
E137K/T216A	3500	SG ; Germon (2009)
E137K/Y237C	160	SG
T216A/Y237C	450	SG ; Germon (2009)

Nicolas BOUCHET

Mécanismes de transposition et de régulation de la transposase de l'élément *mariner Mos1*

Résumé

L'élément *mariner Mos1* est un élément transposable de classe II qui code une transposase, l'enzyme permettant aux transposons de se déplacer dans les génomes. Cette transposase possède un cœur catalytique DDE similaire à celui des intégrases rétrovirales.

Mon travail de thèse a consisté en l'étude de la transposase de *Mos1*, sous différents aspects. Mes résultats complètent le modèle de transposition précédemment établi au laboratoire et apportent une nouvelle vision de la formation du complexe synaptique qui permet l'excision de transposon du site donneur. Mes travaux ont également permis d'identifier le domaine de liaison à l'ADN de la transposase comme un domaine de type CRO. L'étude de la régulation de l'activité de la transposase, qui est phosphorylée, m'a permis d'élargir le modèle de transposition de *Mos1* au contexte cellulaire eucaryote, et des études d'ingénierie de la protéine posent la question de l'impact des facteurs d'hôtes lors de la transposition de *Mos1*.

Enfin, des inhibiteurs de la transposase de *Mos1* ont été identifiés et leur mode d'action caractérisé. Ces composés présentent également une activité sur l'intégrase du VIH-1 et une autre transposase à cœur catalytique DDE.

Mots clés : transposon, *Mos1*, enzymes à triade catalytique DDE, mécanisme de transposition, modifications post-traductionnelles, inhibiteurs de transposases/intégrases

---

Résumé en anglais

The *Mos1 mariner* element is a Class II transposable element, encoding a transposase, which is the enzyme allowing them to move in the genomes. This transposase has a DDE catalytic core like the retroviral integrases.

My work consisted in studying the *Mos1* transposase, under different aspects. My results complete the model of transposition previously established in the laboratory and bring a new vision of the formation of synaptic complex, which allows excision of the transposon donor site. The DNA-binding of *Mos1* transposase has also been identified as a CRO-like domain. Work on regulation of the activity of *Mos1* transposase, which is phosphorylated, allowed me to expand the model of *Mos1* transposition in a eukaryotic cell context. The engineering of the protein were also conducted and questions about the impact of host factors on *Mos1* transposition.

Inhibitors of *Mos1* transposase have been identified and characterized. These compounds also inhibit HIV-1 integrase and an other DDE transposase.

Keywords : transposon, *Mos1*, DDE enzymes, transposition mechanism, post-translationnal modifications, transposases/integrases inhibitors