



HAL
open science

Caractérisation biochimique, fonctionnelle et structurale de l'intégrase Pf-Int de plasmodium

Mehdi Ghorbal

► **To cite this version:**

Mehdi Ghorbal. Caractérisation biochimique, fonctionnelle et structurale de l'intégrase Pf-Int de plasmodium. Sciences agricoles. Université Paris Sud - Paris XI, 2012. Français. NNT : 2012PA114807 . tel-00685428

HAL Id: tel-00685428

<https://theses.hal.science/tel-00685428>

Submitted on 5 Apr 2012

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

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

UNIVERSITÉ PARIS-SUD 11

ECOLE DOCTORALE :

INNOVATION THÉRAPEUTIQUE : DU FONDAMENTAL A L'APPLIQUÉ

PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

DISCIPLINE :

STRUCTURE, FONCTION ET INGENIERIE DES PROTEINES

ANNÉE 2011 - 2012

SÉRIE DOCTORAT N° 1147

THÈSE DE DOCTORAT

soutenu le 28/02/2012

par

Mehdi GHORBAL

Caractérisation biochimique, structurale et
fonctionnelle de Pf-Int : une recombinaise à site
spécifique potentielle de *Plasmodium falciparum*

Directeur de thèse : Deshmukh GOPAUL Chef de Laboratoire, Institut Pasteur, Paris

Composition du jury :

Président du jury : Herman VAN TILBEURGH Professeur d'université, Paris-Sud 11, Orsay
Rapporteurs : Michael CHANDLER Professeur d'université, Paul-Sabatier, Toulouse
Mireille BETERMIER Directeur de recherche, CNRS, Gif-sur-yvette
Examineurs : Loïc RIVIERE Maître de Conférences Universitaire, Bordeaux Segalen
Jean François ALLEMAND Professeur, ENS, Paris

Caractérisation biochimique, structurale et fonctionnelle de Pf-Int : une recombinase à site spécifique potentielle de *Plasmodium falciparum*

RESUME

Plasmodium falciparum est un parasite protozoaire responsable de la forme la plus sévère de la malaria. Depuis quelques années, les cas de résistance aux antipaludiques sont devenus de plus en plus fréquents et de plus en plus répandus. En plus de sa résistance aux drogues actuellement disponibles, ce parasite reste jusqu'à aujourd'hui réfractaire aux vaccinations. L'identification de nouvelles approches basées sur l'inhibition spécifique de certaines de ses cibles moléculaires vitales est devenue une nécessité. La recombinase à site spécifique de *P. falciparum* (Pf-Int) est un enzyme qui a été récemment identifié dans le laboratoire à partir de PlasmoDB. Cette recombinase à site spécifique joue potentiellement un rôle clé dans le système de recombinaison nécessaire à la viabilité du parasite. Cette protéine de 490 acides aminés, soit ~57 kDa, contient une région C-terminale qui porte les résidus conservés du site catalytique des recombinases à tyrosine R-H-K-R-(H/W)-Y. La prédiction montre une région N-terminale qui ressemble à celle de l'intégrase du phage lambda avec un mélange de structures secondaires α et β .

Lors de ces travaux, nous avons d'abord montré par RT-PCR que le gène (MAL13P1.42) qui code pour PF-Int est transcrit pendant le cycle intra-érythrocytaire avec un maximum pendant la phase schizont. Nous avons ensuite essayé de montrer l'implication de Pf-Int dans le cycle parasitaire. Ceci a été réalisé grâce à un parasite (KO: *knock-out*) dont le gène *Pf-Int* a été invalidé. Ces analyses montrent que Pf-Int n'a aucun impact apparent sur le cycle de développement intra-érythrocytaire du parasite, en particulier sur la durée du cycle et le taux de croissance.

Au niveau moléculaire, nous avons également procédé à la production d'anticorps anti-Pf-Int en utilisant le fragment C-162 (Résidus 162-490). La comparaison des profils de marquage, par cet anticorps, des extraits protéiques du KO et du parasite sauvage par la technique de Western blot n'a pas permis d'identifier la protéine endogène dans le parasite sauvage. Dans le but de déterminer la localisation sub-cellulaire de Pf-Int, nous avons réalisé des essais de sur-expression de différentes protéines de fusion dans le parasite. Nous avons essayé de déterminer l'impact de trois codons d'initiation différents ainsi que l'impact de la présence de

la région N-terminale (1-190aa) de Pf-Int sur sa localisation subcellulaire en utilisant une chimère entre la partie N-terminale et la protéine GFP.

Lors de ces travaux, nous avons réussi à sur-exprimer différentes régions de Pf-Int sous forme recombinante dans *E. coli*. Nous l'avons d'abord caractérisé par des études biophysiques. Ainsi nous avons pu déterminer, par dichroïsme circulaire (CD), le contenu en structures secondaires de Pf-Int, qui est proche de celui des autres membres de la même famille. Nous avons également démontré sa stabilité par CD couplé à la dénaturation thermique. Le spectre RMN-1D a aussi pu être enregistré.

La troisième partie de nos travaux a concerné l'identification des cibles ADN de Pf-Int. Deux stratégies de recherche de cibles par affinité ont été utilisées au laboratoire en utilisant une première bibliothèque de séquences synthétisées chimiquement et une deuxième bibliothèque formée de fragments d'ADN génomique de *P. falciparum*. Ces deux approches ont permis l'identification de deux séries de cibles ADN. Grâce aux cibles ADN identifiées, nous avons pu démontrer l'interaction de différents fragments de Pf-Int avec ces cibles par des expériences de retard sur gel natif (EMSA). Nous avons aussi pu démontrer que les protéines recombinantes sont actives *in vitro*. En effet, ces dernières sont capables de former des complexes covalents en présence de l'ADN cible.

La conservation de la protéine, ainsi que son expression différentielle nous laisse à penser que son rôle est certes loin d'être élucidé, mais que Pf-Int reste une cible potentielle pour *P. falciparum*.

Mots clés : *Plasmodium falciparum*, recombinaison génétique à site spécifique, tyrosine recombinase, évolution.

Biochemical, functional and structural characterization of Pf-Int : a potential *Plasmodium falciparum* site specific recombinase

SUMMARY

Plasmodium falciparum is a protozoan parasite responsible for the most severe form of malaria. In recent years, cases of resistance to antimalarial drugs have become increasingly frequent and common. In addition to its resistance to drugs currently available, there is no vaccine available against this parasite till now. The identification of new approaches based on the specific inhibition of some of its molecular targets has become vital.

The identification of the Pf-Int site specific recombinase in *Plasmodium falciparum* by analysis of PlasmoDB is a new opportunity to study the role of genetic variation in this parasite as it needs to adapt to its hosts. This ~ 57 kDa protein contains a C-terminal domain carrying the putative tyrosine recombinase conserved active site residues R-H-K-R-(H/W)-Y, an N-terminus with a predicted alpha-helical bundle and a mixed alpha-beta domain resembling Lambda-Int.

Here, we show that the sequence is highly conserved among members of the Plasmodia. It is expressed differentially during distinct life stages as estimated by RT-PCR, namely with a peak in the schizont phase.

We then tried to show the involvement of Pf-Int in the parasitic cycle. We were able to create a parasite where the Pf-Int gene was knocked-out. The comparison test showed that Pf-Int has apparently no impact on the intraerythrocytic developmental cycle of the parasite, particularly in the cycle length and the growth rate.

At the molecular level, we produced two sets of anti-Pf-Int antibodies using the purified recombinant fragment C-162 (residues 162-490). Comparison of protein extracts from KO and wild parasite by Western blot technique using our antibody has failed to identify the endogenous protein in the wild type parasite.

We also tried to determine the subcellular localization of Pf-Int and the role of possible alternate initiation codons by over-expressing different constructs in the parasite *Plasmodium falciparum*. In order to determine the impact of the N-terminal region (1-190aa) of Pf-Int on its subcellular localization, we also created a chimeric protein using a fusion of Pf-Int(1-

190aa) with the GFP.

We successfully expressed a variety of the recombinant form of Pf-Int in *E. coli*. We have first determined its secondary structure content by circular dichroism (CD) and its solution stability by thermal denaturation-CD. An 1-D NMR spectrum was also recorded.

The third part of our work has involved the identification of the DNA targets of Pf-Int. Two search strategies conducted in the laboratory using a library of chemically synthesized sequences and a second library made of fragments of genomic DNA of *P. falciparum*. Both approaches have allowed the identification of two sets of target DNA. Secondly, electrophoretic mobility shift assays (EMSA) were used to show its affinity and specificity for DNA. The recombinant proteins were shown to be functional as they form a covalent complex with DNA. Thus Pf-Int could be a potential agent that binds to and alters DNA, either in a specific or in random fashion.

Its conservation and differential expression leads us to conclude that although its role is far from being understood, Pf-Int remains a key target for *P. falciparum*.

Key words: Genetic Site specific recombinase, *Plasmodium falciparum*, tyrosine recombinase, DNA binding protein, evolution.

Remerciements

Je tiens tout d'abord à adresser mes remerciements à mon directeur de thèse, Mr Deshmukh Gopaul, pour m'avoir accueilli au sein de son laboratoire. Je souhaite également remercier les directeurs des unités « Immunologie structurale » et « Plasticité des génomes », Mr Graham Bentley et Mr Didier Mazel. Je remercie également Mr Artur Scherf, directeur de l'unité « Biologie de l'Interaction Hôte Parasite » pour m'avoir permis d'accéder à son laboratoire.

J'adresse mes plus sincères remerciements aux deux rapporteurs Mme Mireille Bétermier et Mr Michael Chandler, pour leur disponibilité et l'intérêt qu'ils ont porté à mon travail de thèse.

Je remercie également Mr Herman Van Tilbeurg d'avoir accepté d'être président du jury ainsi que pour la formation qu'il m'a permis d'acquérir lors du master. J'associe à ces remerciements Mr Loïc Rivière et Mr Jean François Allemand, qui ont accepté d'être examinateurs du jury de thèse.

Je tiens à remercier chaleureusement Christophe Thomas et De Qiang Yao, ce fut un plaisir de partager ces trois années avec vous. Un énorme merci à toute l'équipe BIHP, en particulier Christine Benatar Scheidig et Jose Juan Rubio Lopez qui se sont tellement investis dans ce travail et qui n'ont jamais hésité à me faire profiter de leurs compétences et leur professionnalisme.

Merci à l'ensemble des personnes qui travaillent dans les plates-formes de l'Institut Pasteur, en particulier la plate-forme « Biophysique des macromolécules et de leurs interactions » et la plate-forme de « Cristallogénèse et Diffraction des rayons X ».

Enfin, j'adresse une pensée reconnaissante à ma famille et mes amis pour leurs soutiens et encouragements tout au long de ces longues années d'études.

Table des matières

Liste des figures	10
Liste des tableaux	13
Liste des abréviations	14
1 INTRODUCTION	16
1.1 Le paludisme	17
1.1.1 Données générales	17
1.1.2 Cycle de vie de <i>P. falciparum</i>	18
1.1.3 Les antigènes de surface et la variabilité antigénique chez <i>P. falciparum</i>	22
1.1.4 Variabilité antigénique chez d'autres protozoaires	29
1.2 La recombinaison à site spécifique	33
1.2.1 Les recombinases à tyrosine	34
1.2.2 Homologies structurales et fonctionnelles potentiels de Pf-Int	42
1.2.3 La variabilité antigénique générée par des recombinases à site spécifique	45
1.3 Délimitation du sujet	46
2 MATERIELS ET METHODES	50
2.1 Analyses bio-informatiques.....	51
2.2 Etude de la transcription du gène MAL13P1.42	51
2.3 Clonage.....	52
2.3.1 Constructions génétiques dans le but d'exprimer les protéines recombinantes dans <i>E. coli</i>	52
2.3.2 Constructions génétiques dans le but d'exprimer Pf-Int chez le parasite <i>P. falciparum</i>	52
2.4 Expression et purification de protéines recombinantes dans <i>E. coli</i>	54
2.4.1 Surexpression des protéines recombinantes	54
2.4.2 Purification des protéines recombinantes	55
2.5 Analyses biophysiques des protéines recombinantes Pf-Int	56
2.5.1 Analyse de l'homogénéité de la préparation par diffusion dynamique de la lumière..	56
2.5.2 Analyse de la structure secondaire en solution de Pf-Int par dichroïsme circulaire.....	57
2.5.3 Centrifugation analytique.....	57
2.6 Essais de détermination de la structure tertiaire de Pf-Int.....	58

2.6.1	Essais de cristallogénèse	58
2.6.2	Détermination de l'état de repliement de Pf-Int par résonance magnétique nucléaire 58	
2.7	Recherche des cibles ADN de Pf-Int	58
2.7.1	Recherche de cibles par la méthode « Systematic Evolution of Ligands by EXponential enrichment » (SELEX)	58
2.7.2	Recherche de cibles génomiques	60
2.8	Culture des parasites <i>P. falciparum</i>	61
2.8.1	Suivi des cultures de <i>P. falciparum</i>	61
2.8.2	Synchronisation des parasites	61
2.8.3	Transfection des constructions ADN dans les parasites <i>P. falciparum</i>	62
2.8.4	Vérification des transfections.....	63
2.8.5	Création du parasite 3D7-KO-Pf-Int.....	63
2.9	Production d'anticorps anti Pf-Int.....	64
2.9.1	Purification des anticorps anti-Pf-Int.....	64
2.9.2	Analyses des extraits parasitaires par Western blot	65
2.10	Purification des oligonucléotides ADN pour les mesures de fixation à Pf-Int.....	66
2.11	Amplification cibles ADN potentielles de Pf-Int à partir des bibliothèques de plasmides	66
2.12	Caractérisation de la fixation d'oligonucléotides par Pf-Int par retard sur gel (Electrophoretic mobility shift assay)	67
2.12.1	Détermination de l'état de l'ADN retenu au niveau des puits par Pf-Int par digestion à la protéinase k	67
2.12.2	Mesure de la spécificité de Pf-Int par des tests de compétition.....	68
3	RESULTATS	71
3.1	Étude du cadre de lecture ouvert identifié de Pf-Int.....	72
3.1.1	Etude de la conservation de Pf-Int	72
3.1.2	Prédiction de la structure tertiaire de Pf-Int	74
3.2	Etude de Pf-Int <i>in-vivo</i>	75
3.2.1	Analyse des extraits d'ARN messager	75
3.2.2	Vérification de l'inactivation du gène MAL13P1.42 (parasite 3D7-KO-Pf-Int)	76
3.2.3	Etude du taux d'expression protéique de Pf-Int	77
3.2.4	Implication de Pf-Int dans le cycle parasitaire : Comparaison de cultures des parasites K.O. vs 3D7 sauvage.....	77
3.2.5	Surexpression de Pf-Int dans les parasites	78
3.3	Purification des protéines Pf-Int recombinantes.....	81

3.4	Analyses physico-chimiques des protéines purifiées C-117, C-162, C-182 et C-192.....	83
3.4.1	Caractérisation de l'homogénéité de Pf-Int en solution par analyse par diffusion dynamique de la lumière.....	84
3.4.2	Analyse de l'état d'oligomérisation de Pf-Int par centrifugation analytique.....	84
3.4.3	Caractérisation des structures secondaires de Pf-Int : Analyse par dichroïsme circulaire	85
3.4.4	Essais de cristallogénèse et caractérisation préliminaire de la structure 3D par résonance magnétique nucléaire (RMN)	88
3.5	Identification des cibles ADN potentielles de Pf-Int.....	88
3.5.1	Résultats de la recherche de cibles par la méthode SELEX	89
3.5.2	Résultats de la recherche de cibles à partir de l'ADN génomique	89
3.5.3	Étude des cibles identifiées par la méthode Systematic Evolution of Ligands by EXponential enrichment (SELEX)	92
3.6	Détermination de l'état fonctionnel de Pf-int purifiée : Mise en évidence de la formation de complexes covalents.....	100
3.7	Étude de la spécificité de C-192 vis-à-vis de l'ADN.....	101
4	DISCUSSION	105
5	REFERENCES BIBLIOGRAPHIQUES.....	116
6	ANNEXES.....	125

Liste des figures

Figure	Page
Figure 1 : Distribution géographique de la malaria en 2010.	17
Figure 2 : Schématisation d'un mérozoite mettant en évidence les différentes organelles et les structures cellulaires.	18
Figure 3 : Cycle de vie de <i>Plasmodium falciparum</i>	19
Figure 4 : Représentation du cycle érythrocytaire détaillé de <i>P. falciparum</i>	20
Figure 5 : Représentation schématique des familles multigéniques codant pour les molécules de surface associées aux érythrocytes infectés par <i>P. falciparum</i>	22
Figure 6 : Représentation schématique d'un gène <i>var</i>	25
Figure 7 : Architecture et classification des protéines PfEMP1 du clone 3D7	25
Figure 8 : Organisation des gènes <i>var</i> , <i>rif</i> et <i>stevor</i> au niveau des chromosomes	26
Figure 9 : Etude de la similarité de séquence entre les répertoires de gènes <i>var</i> des trois souches 3D7, IT4 et HB3	28
Figure 10 : Les trois différentes issues de la recombinaison à site spécifique.	34
Figure 11 : Structure d'un protomère de Cre en complexe avec l'ADN.	35
Figure 12 : Comparaison des structures des tétramères de Cre, Flp et du « core binding domain » de λ -Int	35
Figure 13 : Organisation structurale de deux protomères de λ -Int en complexe avec l'ADN substrat (A) et de deux protomères de l'intégrase d'intégrons IntI4 (B).	36
Figure 14 : Structure tertiaire d'un protomère non actif de Flp en complexe avec l'ADN	37
Figure 15 : Comparaison des structures de (A) : la topoisomérase du virus de la variole (TopoIB, PDB: 3IGC) (B) : Cre (Y-SSR, PDB: 3C29).	37
Figure 16 : Organisation du site catalytique de Flp.	38
Figure 17 : Réaction de transfert d'un phosphoryl par une recombinase à tyrosine.	38
Figure 18 : Représentation schématique de la réaction de recombinaison catalysée par les Y-SSR.	39
Figure 19 : Organisation des substrats utilisés par λ -Int.	42
Figure 20 : Organisation structurale de λ -Int en complexe avec l'ADN substrat.	43
Figure 21 : Représentation schématique de l'organisation des intégrons.	44
Figure 22 : Comparaison des séquences protéiques prédites des homologues de Pf-Int.	72
Figure 23 : Alignement des séquences protéiques de Pf-Int avec ses homologues d'Apicomplexes, <i>Toxoplasma gondii</i> et <i>Neospora caninum</i>	73
Figure 24 : Représentation schématique de l'organisation des domaines de Pf-Int (490 aa)	74

Figure 25 : Prédiction de la structure tertiaire de l'intégrase de <i>P. falciparum</i> .	74
Figure 26 : Alignement de la structure primaire et la secondaire prédite de Pf-Int avec la structure primaire, secondaire prédite et secondaire connue de l'intégrase du phage lambda utilisé comme modèle.	74
Figure 27 : Modélisation de l'organisation du site catalytique de Pf-Int.	74
Figure 28 : Etude de l'expression de l'ARN messenger de MAL13P1.42 codant pour Pf-Int par RT-PCR.	75
Figure 29 : Stratégie utilisée pour l'invalidation du gène MAL13p1.42 de <i>P. falciparum</i> (Parasite 3D7-KO-Pf-Int).	76
Figure 30 : Vérification de l'invalidation du gène MAL13P1.42 par PCR.	76
Figure 31 : Étude comparative des extraits protéiques des parasites sauvage et KO.	77
Figure 32 : Suivi de la parasitémie des cultures 3D7 et 3D7-K.O. de-Pf-Int.	78
Figure 33 : A : Alignement des séquences ADN des régions 5' des gènes homologues de Pf-Int présents chez d'autres espèces de <i>Plasmodium</i> .	78
Figure 34 : Analyse de la surexpression de Pf-Int dans les parasites.	79
Figure 35 : Contrôle de la présence d'inserts de Pf-Int par PCR des parasites transfectés.	79
Figure 36 : Vérification sur gel SDS-PAGE de la purification du fragment C-162 de Pf-Int sur colonne d'affinité HisTrap.	81
Figure 37 : Vérification sur gel SDS-PAGE de la purification du fragment C-162 de Pf-Int sur colonne d'Héparine et contrôle du clivage de l'étiquette par la protéase TEV.	81
Figure 38 : Vérification sur gel SDS-PAGE de la purification du fragment C-162 de Pf-Int sur la deuxième colonne d'affinité HisTrap.	81
Figure 39 : Vérification sur gel SDS-PAGE de la purification du fragment C-162 de Pf-Int sur colonne d'exclusion stérique, HiLoad 16/60 Superdex 200pg.	82
Figure 40 : Vérification sur gel SDS-PAGE de la purification des fragments C-117, C-182 et C-192 après passage sur la colonne d'exclusion stérique.	82
Figure 41 : Analyse des protéines recombinantes par diffusion dynamique de la lumière.	84
Figure 42 : Détermination de l'état d'oligomérisation de la protéine C-162 par centrifugation analytique.	85
Figure 43 : Caractérisation des structures secondaires de C-117, C-162 et C-192 par dichroïsme circulaire.	85
Figure 44 : Étude de la stabilité thermique des protéines C-117 (A), C-162 (B) et C-192 (C) par dichroïsme circulaire-chauffé (thermal denaturing-CD).	86
Figure 45 : Analyse des protéines C-162 et C-117 par dichroïsme circulaire en UV proche.	87
Figure 46 : Etude préliminaire de l'organisation structurale de Pf-Int par RMN du proton.	88
Figure 47 : Vérification de la banque de cibles génomiques clonées par amplification PCR.	89

Figure 48 : Étude de l'interaction entre C-162 et un ADN cible génomique par des expériences de retard sur gel natif (EMSA).	91
Figure 49 : Étude de l'interaction entre C-162 et un ADN cible génomique par des expériences de retard sur gel natif couplé à la digestion par la protéinase k.	91
Figure 50 : Étude de l'interaction entre C-162 et un ADN cible génomique non marqué.	91
Figure 51 : Étude de l'interaction, par retard sur gel, entre les différents fragments purifiés de l'intégrase et l'ADN selex-8 (25 pb).	92
Figure 52 : Étude, par retard sur gel, de l'interaction entre les protéines 28-C192, C192, C182 et l'ADN selex8.	93
Figure 53 : Étude de l'effet de la nature du sel (KCl ou NaCl) sur l'interaction entre C-192 et l'ADN SELEX-8.	94
Figure 54 : Étude de l'effet de la concentration du sel NaCl sur l'interaction entre C-192 et l'ADN SELEX-8.	94
Figure 55 : Étude de l'effet de la concentration du KCl sur l'interaction entre C-192 et l'ADN SELEX-8.	94
Figure 56 : Étude de l'effet de la concentration de l'EDTA sur l'interaction entre C-192 et l'ADN SELEX-8.	95
Figure 57: Étude de l'effet de la concentration de MgCl ₂ sur l'interaction entre C-192 et l'ADN SELEX-8.	95
Figure 58 : Étude de l'effet de la concentration du Tween20 sur l'interaction entre C-192 et l'ADN SELEX-8.	96
Figure 59 : Étude de l'effet de la concentration du DTT sur l'interaction entre C-192 et l'ADN SELEX-8.	96
Figure 60: Étude de l'effet de la concentration d'ADN sur l'interaction entre C-192 et l'ADN SELEX-8.	97
Figure 61: Étude de l'effet de la taille de l'ADN sur l'interaction entre C-192 et selex-8.	98
Figure 62: Étude de l'effet de la concentration du compétiteur non spécifique poly (dI-dC) sur l'interaction entre C-192 et l'ADN SELEX-8.	98
Figure 63: Étude de l'interaction entre les fragments C-117, C-162 et C-192 et l'ADN selex-8 par retard sur gel en absence de poly (dI-dC).	99
Figure 64: Mise en évidence de la formation de complexes covalents entre les fragments C-117, C-162 et C-192 et l'ADN selex-8 par retard sur gel dénaturant SDS-PAGE (4-12%).	100
Figure 65: Étude par retard sur gel de l'interaction entre C-192 et l'ADN protect40 de la région promotrice du gène <i>phoP</i> de <i>M. tuberculosis</i> (40 pb).	102
Figure 66 : Étude de la spécificité de C-192 vis-à-vis de l'ADN.	103
Figure 67 : Étude de la spécificité de C-192 vis-à-vis de l'ADN.	103

Liste des tableaux

Tableau	Page
Tableau 1 : Amorces et oligos utilisés pour les clonages et les EMSA.	70-71
Tableau 2 : Comparaison de la composition en acides aminés de Pf-Int par rapport aux protéines homologues d'autres espèces de <i>Plasmodium</i> .	74
Tableau 3 : Contenu en structures secondaires des trois protéines purifiées C-117, C-162 et C-192 déterminé par dichroïsme circulaire.	87
Tableau 4 : Contenu en structures secondaires de différentes recombinaisons à tyrosine dont la structure tertiaire a été déterminée.	87
Tableau 5 : Séquences des cinq cibles ADN identifiées par la méthode SELEX	90
Tableau 6 : Séquences en acides aminés du peptide Met162-Leu191 et du peptide 28aa ajouté en N-terminal de C-192 (28aa-C192).	93

Liste des abréviations

λ -Int	: intégrase du phage lambda
°C	: degré Celsius
μ m	: micromètre
aa	: acide aminé
amu	: « atomic mass unit », unité de masse atomique
attB	: « bacterial attachment site »
attP	: « phage attachment site »
BIR	: « Break Induced Replication »
CD	: « circular dichroism », dichroïsme circulaire
CD 36	: « Cluster of Differentiation 36 »
CIDR	: « Cysteine-rich Interdomain Region »
CR1	: « Complement Receptor 1 »
CSA	: « chondroïtin sulfate A »
DBL	: « Duffy Binding like Domain »
DLS	: « Dynamic Light Scattering », diffusion dynamique de la lumière
DTT	: Dithiothreitol
EDTA	: acide éthylène diamine tétraacétique
ER	: réticulum endoplasmique
Fis	: « Factor for inversion stimulation », facteur de stimulation d'inversion
GR	: Globule Rouge
ICAM	: « Intercellular Adhesion Molecule-1 »
IHF	: « Integration Host Factor »
IntI4	: Intégrase d'intégrons de type 4
iRBC	: « infected Red Blood Cell », globule rouge infecté
kb	: kilo bases
MALDI-TOF	: « matrix-assisted laser desorption/ionization - time-of-flight », désorption-ionisation laser assistée par matrice – temps de vol
MHz	: Méga-hertz
MSP	: « Merozoite Surface Protein »
nm	: nanomètre
ORF	: « Open reading frame », phase ouverte de lecture
<i>P. falciparum</i>	: <i>Plasmodium falciparum</i>
PAGE	: électrophorèse sur gel de polyacrylamide
pb	: paires de bases
PBS	: « Phosphate buffer saline », tampon phosphate salin
PDB	: « Protein Data Bank »
PfEMP1	: « Plasmodium falciparum erythrocyte membrane protein 1 »
PlasmoDB	: « Plasmodium Database »
ppm	: partie par million
RMN	: résonance magnétique nucléaire
SDS	: « Sodium Dodecyl Sulfate », dodécylsulfate de sodium
smorf	: « Small Open Reading Frame »

S-SSR : « serine site-specific recombinase », recombinaison à site spécifique à sérine
TARE : « Telomere Associated Repeat Element »
TM : Transmembrane
VESA1 : « Variant Erythrocyte Surface Antigen 1 »
VSG : « Variable Surface Glycoprotein »
VSP : « Variable Surface Protein »
Xis : excisionase
Y-SSR : « tyrosine site-specific recombinase », recombinaison à site spécifique à tyrosine

1 INTRODUCTION



Figure 1 : Distribution géographique de la malaria en 2010. Bleu foncé : zones à haut risque de transmission. Bleu clair : zones à risque de transmission limité. 3,3 milliards de personnes vivent dans des zones à risque et 91% des décès ont lieu en Afrique [2].

1.1 Le paludisme

1.1.1 Données générales

Le paludisme, aussi appelé malaria, est une maladie infectieuse due à la présence d'un protozoaire du genre *Plasmodium* chez certains vertébrés comprenant des mammifères, des oiseaux et des reptiles. Toutes les espèces de *Plasmodium* ciblent et se multiplient à l'intérieur des érythrocytes de l'hôte vertébré [1]. Chez l'homme, la prolifération du parasite est à l'origine de manifestations cliniques différentes allant de l'infection asymptomatique jusqu'à l'infection sévère incluant des épisodes fébriles, des problèmes respiratoires, une anémie sévère et une obstruction de la circulation sanguine qui conduit à la mort en absence de traitement.

Quatre espèces de *Plasmodium* (*falciparum*, *vivax*, *malariae* et *ovale*) sont majoritairement à l'origine de la malaria humaine. Certains cas d'infection par l'espèce *P. knowlesi*, un parasite communément trouvé chez les singes, ont également été rapportés [2]. Ce parasite est propagé dans la nature par la piqûre infectante d'un insecte vecteur du genre *Anopheles* de l'ordre des *Diptera*, de la famille des *Culicidae*, sous-famille des *Anophelinae* [2]. Seule la femelle adulte hématophage joue le rôle de vecteur car le repas sanguin lui est nécessaire à la maturation des œufs [2]. Dans le cas du paludisme congénital, la malaria est transmise par voie placentaire de la mère infectée au fœtus [2].

Cette maladie constitue un véritable problème de santé publique. Son incidence annuelle mondiale, toutes formes confondues, est estimée par l'OMS à 247 millions de cas et à près d'un million de décès en 2008 (91% des décès en Afrique et 85% sont des enfants de moins de 5 ans) essentiellement imputés à l'espèce *falciparum* [2] avec au total 3,3 milliards de personnes vivant dans des zones à risque. La distribution géographique de cette maladie est limitée par celle de son vecteur. Elle affecte 108 pays principalement localisés dans la zone intertropicale, essentiellement les pays de l'Afrique sub-saharienne mais également d'Asie, d'Amérique latine, du Moyen Orient et d'Eurasie (Figure 1).

Le problème est d'autant plus préoccupant qu'il n'existe pas encore de vaccin contre cette maladie et que les traitements disponibles actuellement sont limités et loin d'être satisfaisants. Ainsi les drogues antipaludiques utilisées jusqu'à aujourd'hui tel que les chloroquines, sulfadoxine, et pyriméthamine administrés en mono ou en traitement combiné sont associés à des cas de plus en plus nombreux de résistance [3]. La seule substance à laquelle le parasite n'a pas encore développé de résistance est l'artémisine, dont l'OMS recommande l'utilisation

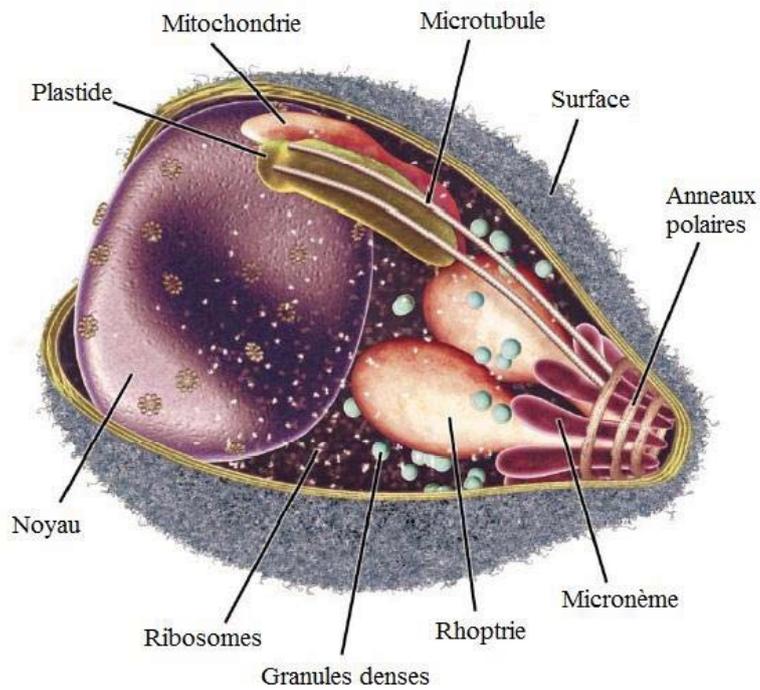


Figure 2 : Schématisation d'un mérozoïte mettant en évidence les différentes organelles et les structures cellulaires. Le mérozoïte est une cellule, d'environ 1 μm de longueur, spécialisée dans l'invasion des globules rouges. Le mérozoïte possède une extrémité apicale composée de différentes organelles : des micronèmes, des granules denses, des rhoptries et des anneaux polaires [159].

en traitement combiné (ACT, *Artemisinin-based Combination Treatment*) [2].

La recherche de nouvelles cibles thérapeutiques parmi les protéines qui participent aux différents processus métaboliques de *Plasmodium* est devenue de ce fait un impératif. Cependant dans ce type d'approche, il est important d'éviter les recouvrements possibles avec les cibles homologues présentes chez l'hôte humain. La complexité et la particularité des étapes du cycle de vie rendent ce type d'approche particulièrement intéressant car ce parasite utilise un arsenal unique pour le passage d'un hôte à l'autre et pour se développer à l'intérieur d'un même hôte. Ces changements majeurs du métabolisme dépendent des réorganisations du programme d'expression génétique. Afin de comprendre la complexité de la biologie du parasite, nous nous intéresserons dans les paragraphes suivants de l'introduction aux étapes clés du cycle parasitaire dans un premier temps, et nous développerons dans un deuxième temps la partie essentielle de ce travail, qui a pour objectif de mettre en évidence le rôle que pourrait jouer un enzyme de recombinaison dans le métabolisme de l'ADN chez *P. falciparum*.

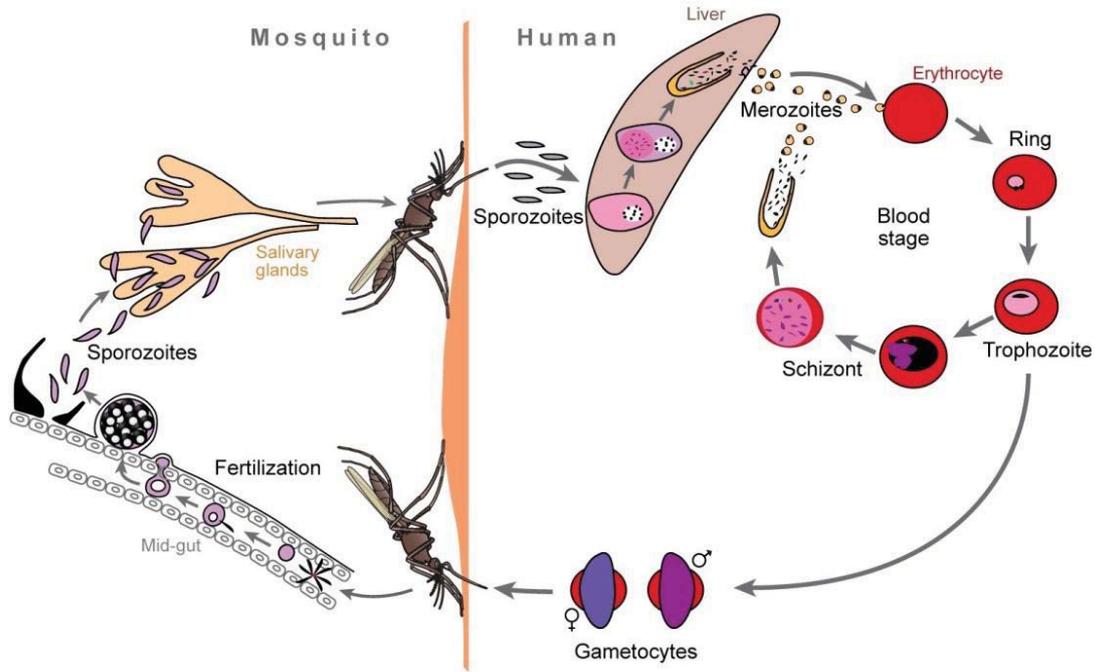
1.1.2 Cycle de vie de *P. falciparum*

Les parasites du genre *Plasmodium* appartiennent à l'embranchement des *Apicomplexa* qui inclut également les parasites *Toxoplasma gondii*, *Cryptosporidium parvum*, *Eimeria spp.* et *Theileria spp.* [4-6].

Ces parasites partagent un ensemble de caractéristiques morphologiques spécifiques de ce groupe, en particulier la présence d'un complexe apical à au moins un stade de leur vie (Figure 2). Le complexe apical consiste en un ensemble d'organelles spécialisées telles que les rhoptries, les micronèmes, les monorèmes, les granules denses et les anneaux polaires qui sont impliquées dans l'invasion des cellules de l'hôte.

En plus du complexe apical, le parasite possède d'autres organelles nécessaires à son développement tel qu'un noyau, un réticulum endoplasmique, un appareil de golgi, une mitochondrie et un apicoplaste. L'apicoplaste est une organelle issue de l'endosymbiose ancestrale d'une algue, homologue aux chloroplastes, présente chez tous les *Apicomplexa* sauf *Cryptosporidium* pour lequel celle-ci a été perdue au cours de l'évolution. Différentes voies métaboliques sont assurées par l'apicoplaste dont la biosynthèse d'acides gras, de terpénoïdes, de l'hème et des « clusters » fer-soufre. Cette organelle, essentielle à la survie de *P.*

A



B

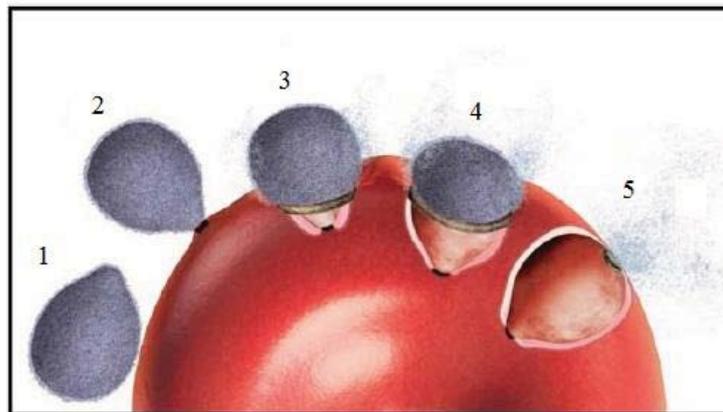


Figure 3 : A. Cycle de vie de *Plasmodium falciparum* [16]. Lors d'un repas sanguin, la femelle *Anopheles* injecte les parasites sous forme de sporozoïtes dans le derme de l'hôte sain. L'infection commence lorsque ces sporozoïtes envahissent le foie. Après plusieurs cycles de division cellulaire, les parasites sous forme mérozoïte sont relâchés dans la circulation sanguine où ils envahissent les globules rouges. Les mérozoïtes se développent successivement en stades anneau, trophozoïte, et schizont. Au bout de 48 heures, les cellules filles (mérozoïtes) sont relâchées et envahissent rapidement de nouveaux érythrocytes non infectés. Lorsque certaines formes parasitaires (gamétocytes) sont prélevées par la femelle *Anopheles*, un autre cycle de développement commence dans le moustique. Au bout de 10 jours, les parasites sous la forme de sporozoïtes se retrouvent dans les glandes salivaires du moustique prêts à être injectés à un autre individu. **B. Schématisation de l'invasion d'un globule rouge par un mérozoïte** [159]. Le contact initial entre le parasite et le globule rouge se fait de manière aléatoire (étape 1) ; Le parasite se réoriente ensuite pour pouvoir former la jonction entre l'extrémité apicale et la membrane du globule rouge (étape 2) ; Le processus d'internalisation commence alors par la formation d'une jonction autour du parasite (étape 3 et 4) et se termine par la formation d'une vacuole parasitaire (étape 5).

falciparum est caractérisée par la présence de quatre membranes. L'apicoplaste est une organelle semi-autonome qui possède son propre génome et sa machinerie d'expression mais qui utilise également plusieurs protéines codées par le génome nucléaire. Les gènes nucléaires qui codent pour les protéines exportées vers l'apicoplaste sont issus d'une relocalisation intracellulaire de gènes [7, 8].

P. falciparum possède trois génomes différents : Le génome nucléaire, qui a été séquencé en 2002 par *Gardner et al* (clone 3D7), a une taille de 23 mégabases et comprend 14 chromosomes qui portent près de 5300 gènes. Le génome plastidique a une taille de 35 kilobases (kb) et code pour une trentaine de protéines uniquement [9]. *P. falciparum* possède également un génome mitochondrial dont la taille est de 6 kb [10].

P. falciparum possède un cycle de vie complexe. Suivant l'hôte qu'il infecte et les différentes phases métaboliques, ce parasite alterne entre différentes formes morphologiques et états physiologiques. Le cycle parasitaire de *P. falciparum* est composé de deux phases : une phase de prolifération chez l'homme et une phase chez le moustique (Figure 3, A).

➤ La première phase de prolifération ayant lieu chez l'homme (Figure 3, A, partie de gauche) constitue la phase de multiplication asexuée appelée schizogonie et est composée de deux étapes. La première étape est initiée par la femelle *Anophele* infectée en inoculant des sporozoïtes dans l'hôte humain [11]. Ces sporozoïtes filiformes d'une dizaine de micromètres sont capables de se mouvoir grâce à l'utilisation d'un moteur actomyosine et de protéines adhésives libérées par les micronèmes. Les sporozoïtes vont rapidement (1-3 heures) migrer vers le foie via la circulation sanguine où ils infectent les cellules hépatiques et s'y multiplient [12]. Après plusieurs cycles de division, les cellules infectées relâchent dans la circulation sanguine des parasites sous forme de mérozoïtes et la deuxième étape de prolifération intra-érythrocytaire peut alors commencer. Cette étape commence lorsque les mérozoïtes libérés se répandent dans tout le système sanguin et infectent les érythrocytes disponibles pour s'y multiplier [6].

Le mérozoïte (~1 µm de long) a une forme ellipsoïdale avec une proéminence apicale à une de ses extrémités (Figure 2 et Figure 3, B). Il s'agit d'une cellule spécialisée dans l'invasion des globules rouges (GR). Le contact initial entre le mérozoïte et le globule rouge est réversible et peut avoir lieu de manière aléatoire sur toute la surface du mérozoïte (Figure 3, B, étape 1). Ces interactions ont lieu grâce à des protéines présentes à la surface des

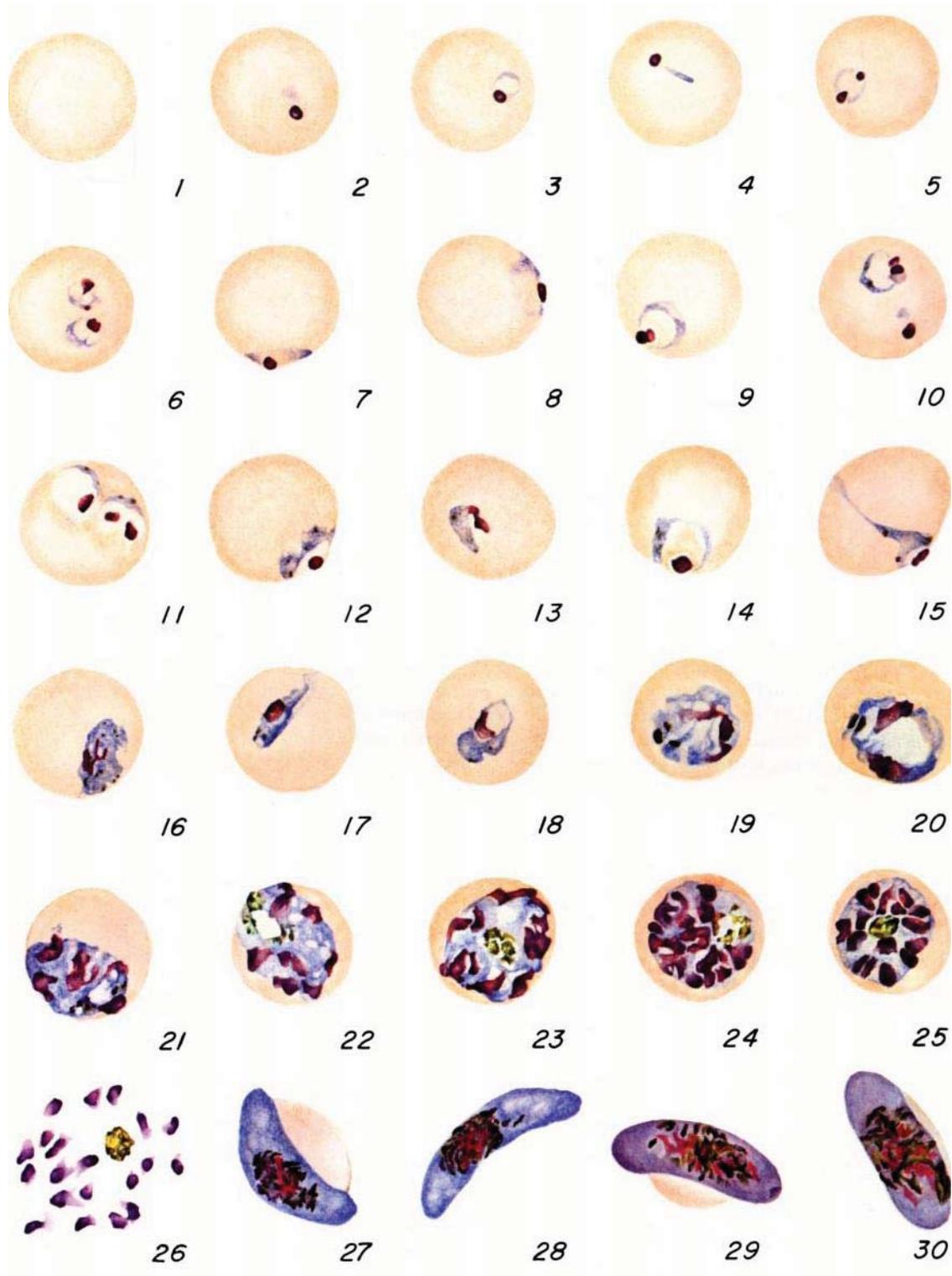


Figure 4 : Représentation du cycle érythrocytaire détaillé de *P. falciparum*. Les cellules sont observées au microscope après coloration Giemsa. 1 : Globule rouge sain ; 2-30 : Globule rouge infecté par un parasite *P. falciparum* ; 2-10 : stades anneau ; 11-18 : trophozoite ; 19-25 : schizont ; 26 : mérozoites libérés ; 27-30 : gamétocystes. La coloration (exemple : étape 9) montre que le parasite est constitué d'un noyau (rouge foncé), d'un cytoplasme (bleu) entourant une vacuole nutritive (clair) [160].

mérozoïtes (merozoïte surface proteins : MSPs-1,2,4,5,8 et 10) dont la plus étudiée MSP-1 qui est essentielle à l'invasion et à la survie des parasites. Une fois l'interaction établie, le parasite est réorienté de manière à ce que l'extrémité apicale soit au contact de la membrane érythrocytaire (Figure 3, B, étape 2). A la suite du repositionnement du mérozoïte, celui-ci relâche les protéines et les lipides contenus dans les rhoptries à l'interface du parasite et du globule rouge infecté (iRBC, « infected Red Blood Cell ») et il se forme une jonction irréversible. Une jonction de déplacement (Figure 3, B, étapes 3 et 4) est créée autour du mérozoïte pour former la vacuole parasitophore (Figure 3, B, étape 5) dont la membrane reste associée à la membrane plasmique du parasite durant le cycle intra-érythrocytaire [6, 13].

Lors du stade intra-érythrocytaire, les mérozoïtes se transforment d'abord en un disque fin avec une invagination au centre caractéristique du stade parasitaire anneau (Figure 4, étapes 2-10). Ces jeunes anneaux possèdent une région centrale fine et une région périphérique plus épaisse qui contient le noyau et le reste des organelles. Au stade anneau intermédiaire (10-15 h après l'invasion) le parasite a un diamètre de 2-3 μm et une épaisseur de 0,5 μm . A ce stade, il commence à internaliser le cytoplasme de la cellule hôte par un mécanisme de micro pinocytose. La digestion de l'hémoglobine de l'hôte constitue une source d'acides aminés nécessaire au développement du parasite. Ce processus produit aussi des molécules d'hème toxiques qui sont stockées sous forme cristalline connue sous le nom d'hémozoïne [6].

Une fois au stade trophozoïte (20-38 h après l'invasion), le parasite dont la taille est de 2-4 μm est à la phase la plus active de son métabolisme (Figure 4, étapes 11-18). Un grand nombre de ribosomes peuvent être observés prouvant que le niveau de synthèse de protéines est particulièrement élevé. La mitochondrie et l'apicoplaste ont augmenté de taille et leurs membranes deviennent visibles.

Lorsque le parasite commence sa maturation, il induit certaines modifications de l'iRBC grâce aux protéines parasitaires exportées vers le cytoplasme érythrocytaire.

La membrane cellulaire des iRBCs devient alors déformée et des protubérances (appelés « knob ») de 30-40 nm de hauteur et de 90 nm de diamètre apparaissent. Le rôle des « knobs » est d'exposer au maximum les protéines parasitaires exportées à la surface. Parmi celles-ci, on retrouve notamment la protéine PfEMP1 (*P. falciparum* « Erythrocyte Membrane Protein » 1) qui est la plus étudiée de toutes, et dont le rôle est de faciliter l'adhésion des iRBCs à l'endothélium des micro-vaisseaux ainsi que de permettre la formation des rosettes (adhésion de globules rouges parasités à des globules rouges non parasités). Une autre série de structures

appelées fissures de Maurer peuvent également être observées. Ces structures de 500 nm de diamètre et de 30 nm d'épaisseur sont issues de la membrane de la vacuole parasitophore et assurent le transport des protéines parasitaires à la surface de l'iRBC.

Le dernier stade du cycle intra-érythrocytaire (38-48 h après l'invasion, Figure 4, étapes 19-25) est établi par la présence des formes mûres schizont. Le parasite se prépare à la production des cellules filles mérozoïtes (16 à 20 cellules filles par schizont). 80 % de l'hémoglobine de la cellule hôte a été consommée et le parasite occupe presque la totalité du volume de l'érythrocyte. Son génome subit plusieurs divisions et les proéminences apicales ainsi que les précurseurs des rhoptries apparaissent dès que l'étape de quatre noyaux est atteinte. En accord avec ce changement métabolique, on observe que les protéines impliquées dans la réplication de l'ADN notamment DNA Pol α , DNA Pol δ et deux topoisomérases subissent une augmentation de leurs taux d'expression [14]. Une fois les seize noyaux apparus, les cellules filles se séparent par l'invagination de la membrane plasmique accompagnée de la formation des micronèmes et des granules denses. La mitochondrie et l'apicoplaste qui deviennent hautement ramifiés sont les derniers à se diviser. A la fin de ce stade (Figure 4, étape 25), les cellules filles mérozoïtes sont empaquetées dans le globule rouge avec un corps résiduel au centre de celui-ci contenant les cristaux d'hémozoïne [6]. Le schizont mûr provoque ensuite l'éclatement de l'érythrocyte (Figure 4, étape 26) en laissant se répandre les cellules filles permettant ainsi la dissémination de l'infection par une succession d'internalisation/relargage d'un érythrocyte à un autre.

➤ La deuxième phase du cycle parasitaire est la phase de différenciation sexuée (Figure 3, A, partie de droite). Cette phase commence approximativement 7 à 15 jours après l'apparition des parasites dans le sang des personnes infectées. Lors de cette phase, certains parasites subissent une différenciation vers une forme dite gamétocytaire qui est la seule forme capable d'assurer la transmission du parasite de l'hôte humain au moustique. Les gamétocytes mâles et femelles ainsi produits sont prélevés par le moustique lors d'un repas sanguin. En pénétrant dans l'intestin du moustique, les parasites subissent une baisse de température de près de 5°C, une élévation du pH ainsi que l'exposition à des molécules produites par le moustique tel que l'acide xanthurénique. Ces facteurs induisent la lyse des érythrocytes et la transformation des gamétocytes en macrogamètes femelles prêts à être fertilisés et l'activation des microgamétocytes mâles qui entrent en division cellulaire. Chaque microgamétocyte duplique son génome trois fois dans le but de produire huit microgamètes

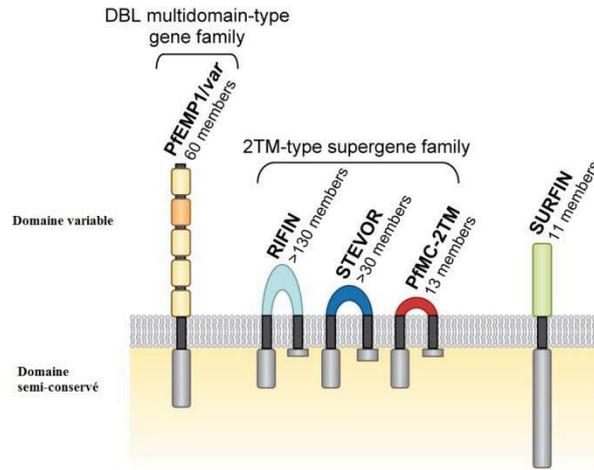


Figure 5 : Représentation schématique des familles multigéniques codant pour les molécules de surface associées aux érythrocytes infectés par *P. falciparum* [16]. Les molécules de surface de type PfEMP1 sont exprimées très tôt après l'invasion (18h après l'invasion) et sont impliquées dans l'adhésion des érythrocytes infectés aux micro-vaisseaux. Les protéines PfEMP1 sont les principaux antigènes reconnus par le système immunitaire. Les membres des autres familles multigéniques (*rif* (~130 gènes), *stevor* (~30 gènes), et *Pfmc-2TM* (~13 gènes), *surfIn* (~11 gènes)) sont exprimés dans les stades avancés du cycle érythrocytaire et sont également soumis à la variabilité antigénique mais leur rôle est encore à déterminer.

mobiles. La fécondation a lieu lorsqu'un microgamète et un macrogamète fusionnent pour donner un zygote qui subit une méiose et qui se différencie en ookinète. Cet ookinète est mobile et pourvu d'un complexe apical. Il traverse l'épithélium intestinal du moustique et se transforme alors en oocyste dont le noyau se divise pour donner de nouveaux sporozoïtes au niveau des glandes salivaires. Le cycle est bouclé lorsque ces sporozoïtes sont inoculés via la trompe de l'insecte à d'autres individus lors d'autres repas sanguins [15 , 16 , 17].

1.1.3 Les antigènes de surface et la variabilité antigénique chez *P. falciparum*

La persistance de *P. falciparum* durant la phase de prolifération intra-érythrocytaire dépend de l'expression de différentes molécules parasitaires à la surface des iRBCs. Les trophozoïtes commencent à produire des protéines qui sont exportées à la surface des GRs dès le début de l'infection. Ces antigènes de surface agissent comme molécules d'adhésion et sont à l'origine de la séquestration des globules rouges infectés dans le système vasculaire et de l'adhésion des iRBC à des GR non parasités (phénomène connu sous le nom de « rosetting »). Étant donné que ces protéines sont présentes à la surface des iRBCs, elles sont continuellement exposées à l'environnement extracellulaire où elles sont détectées par le système immunitaire de l'hôte [16]. Dans le but d'échapper à la pression du système immunitaire, le parasite a recours à la stratégie de variation phénotypique basée sur l'expression coordonnée de ses différents antigènes de surface.

1.1.3.1 Les antigènes de surface de *P. falciparum*

Les molécules exprimées à la surface des iRBC les plus étudiées proviennent des ensembles multigéniques (Figure 5), dont les protéines PfEMP1 codés par la famille des gènes *var* (60 gènes par génome haploïde), RIFIN codés par les gènes *rif* (repetitive interspersed family, 150 gènes), et les protéines STEVOR (35 gènes), SURFIN (11 gènes) et PfMC-2TM (*Plasmodium falciparum* « Maurer's Cleft 2 transmembrane domains protein », 13 gènes) [16].

Le rôle de ces antigènes de surface (essentiellement la protéine PfEMP1) est de permettre la cytoadhérence des iRBC contenant des parasites trophozoïtes et schizonts aux cellules endothéliales. Cette séquestration dans le système vasculaire permet aux parasites d'éviter l'élimination par la rate [16, 18]. Il résulte de cette cytoadhérence, une obstruction vasculaire à l'origine des symptômes observés lors de l'infection et une réponse immunitaire localisée [19]. Jusqu'à présent, douze récepteurs de l'hôte ont été identifiés permettant l'adhésion des iRBC [18]. Parmi ceux-ci, on note les récepteurs CD36 et L'ICAM-1 (« InterCellular Adhesion

Molécule 1 »). Le CD36 est le récepteur le plus commun et est associé au phénomène de « rosetting » [20]. L'ICAM-1 est le récepteur associé au paludisme cérébral. Quant aux autres récepteurs identifiés, ils sont soit rarement observés, soit observés dans des cas spécifiques. L'exemple le plus connu est celui de la molécule CSA (Chondroïtin Sulfate A) exprimée au niveau du placenta. L'adhérence des iRBC aux molécules de CSA placentaire *via* la protéine parasitaire VAR2CSA est responsable de la forme congénitale du paludisme [18].

1.1.3.2 La variabilité antigénique chez *P. falciparum*

La variabilité antigénique est un mécanisme utilisé par différents microorganismes dans le but d'éviter l'élimination par le système immunitaire de l'hôte. Deux mécanismes différents sont à la base de ce processus: un mécanisme génétique et un mécanisme épigénétique. Le mécanisme génétique est basé sur la mutation et/ou la recombinaison qui apportent une modification de la séquence ADN du gène qui code pour l'antigène. Il peut également impliquer la modification d'une séquence, modifiant ainsi le niveau de transcription ou de traduction du gène en question. Au contraire, le mécanisme épigénétique permet de réguler l'expression des antigènes uniquement en affectant l'expression d'un gène sans modifier la séquence d'ADN [21].

La variabilité antigénique générée par des mécanismes génétiques peut avoir lieu soit par des mutations aléatoires (dérive génétique) soit par des réactions programmées de recombinaison des gènes qui codent pour les protéines de surface. La variation antigénique générée par des mutations aléatoires est un mécanisme surtout utile pour les organismes dont le génome est répliqué d'une manière rapide et avec un faible degré de fidélité tel que le virus de l'immunodéficience humaine.

Un mécanisme plus sophistiqué basé sur l'utilisation des familles multigéniques est utilisé par les organismes à génome plus complexe. Une famille multigénique est un répertoire composé de plusieurs gènes (allant d'une dizaine jusqu'à un millier de gènes) qui codent pour autant de versions de la même protéine de surface. Ces versions sont suffisamment différentes pour pouvoir induire des réponses immunitaires distinctes mais gardent un certain degré d'homologie qui permet à la protéine d'assurer sa fonction. Le pathogène est capable de coordonner l'expression des différents gènes du répertoire de manière à ce qu'un seul gène soit exprimé à la fois.

Ce système, qui a été observé chez certains protozoaires, bactéries, virus ou champignons

permet au pathogène de se multiplier sans se faire éliminer, au moins de façon temporaire et d'établir une infection chronique causée par des vagues successives de parasitémie [16 , 22].

Chez les espèces *Plasmodium*, la variabilité antigénique a été, en premier lieu, observée par *Cox et al* [23] chez les souris infectées par *P. berghei*, suivi peu de temps après par l'observation de *Brown et Brown* [24] de la présence d'antigènes différents dans des érythrocytes infectés par *P. knowlesi* chez le macaque rhésus. Ces observations ont été suivies par d'autres chez les espèces *P. chabaudi* chez la souris et le *P. fragile* chez le macaque à toque [20]. La démonstration expérimentale de la variabilité antigénique chez *P. falciparum* a été faite en 1983 chez les singes *Saimiri* [16].

Chez *P. falciparum*, la variabilité antigénique dépend du contrôle de l'expression différentielle des différentes variantes des molécules parasitaires présentes à la surface des iRBC. En effet, ce processus encore mal compris chez *Plasmodium*, requiert l'expression mutuellement exclusive des gènes en exprimant à un temps donné un seul gène, et en maintenant tous les autres membres de la famille dans un état transcriptionnel silencieux. Ce mécanisme de commutation concerne aussi bien la famille des protéines PfEMP1 que les familles de protéines à deux domaines transmembranaires (RIFIN, STEVOR et PfMC-2TM) suggérant que ces dernières pourraient aussi être impliquées dans la variation antigénique [19]. Ainsi, en régulant la commutation entre les différents gènes, les parasites sont capables de retarder l'élimination par la réponse immunitaire et de maintenir une infection persistante.

Il a été récemment montré que la variation antigénique chez *Plasmodium* est régulée par des facteurs épigénétiques impliquant essentiellement des modifications au niveau des histones. La première démonstration a été faite par *Freitas-junior et al*, 2005 où les auteurs ont montré que les régions 5' en amont des gènes *var* actifs sont associées à une modification de type acétylation des histones H4 contrairement aux histones associées aux gènes *var* silencieux qui sont hypoacétylés [25].

Des études postérieures [26] ont permis d'affiner encore plus le modèle en montrant que les histones des régions 5' du gène transcrit pendant la phase anneau sont di et tri méthylés au niveau de la H3K4 (H3K4me2 et H3K4me3) et sont également acétylé au niveau de la H3K9 (H3K9ac) alors que les histones de région codante du gène sont modifiées en H3K9me3 et pas en H3K9ac. Au contraire, les gènes silencieux sont associés à une triméthylation de la H3K9 (H3K9me3) au niveau de la région 5' en amont et la région codante durant tout le cycle parasitaire.

En plus de la régulation par les modifications épigénétiques, d'autres niveaux de contrôle

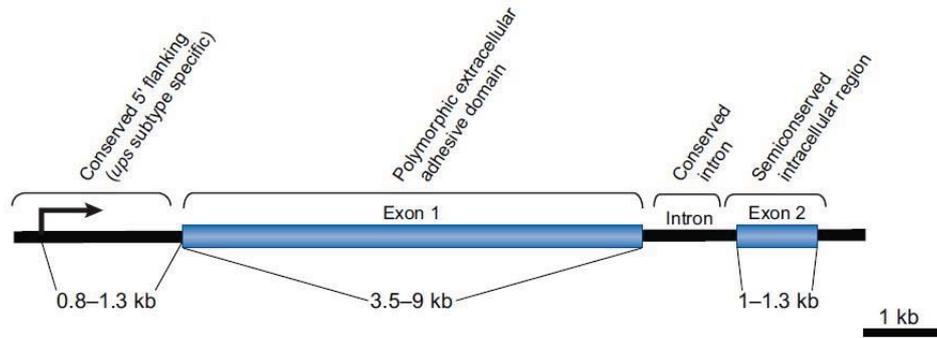
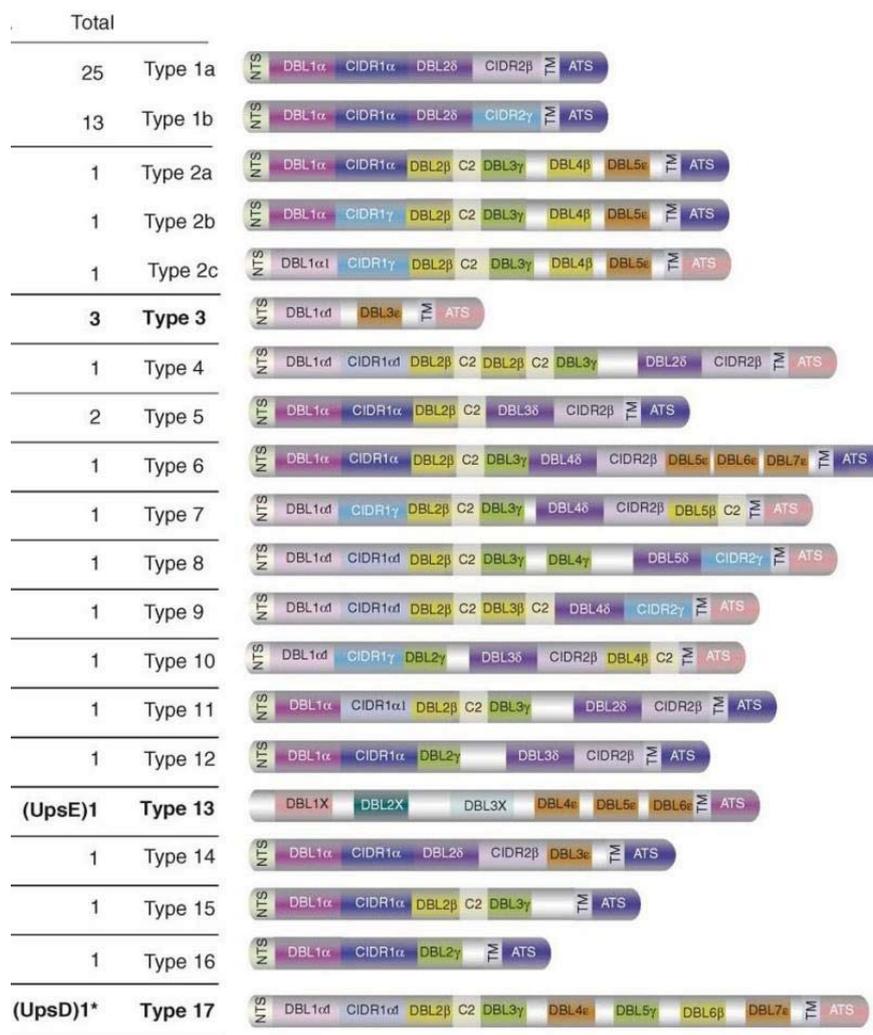


Figure 6 : Représentation schématique d'un gène *var* [16]. Les caractéristiques générales des gènes *var* sont représentées : exon 1 (3,5-9 kb) codant pour la partie extracellulaire variable, un exon 2 semi-conservé (1-1,3 kb) qui code pour la partie intracellulaire et un intron conservé.



59

Figure 7 : Architecture et classification des protéines PfEMP1 du clone 3D7 [30]. Les protéines PfEMP1 sont formées de plusieurs domaines : le domaine N-terminal (NTS), les domaines « Duffy Binding Like » (DBL), les domaines « Cysteine rich Interdomain Region » (CIDR), les domaines constants (C2), les domaines transmembranaires (TM) et un segment terminal acide (ATS).

pourraient exister. La présence de séquences d'ADN régulatrices en particulier l'intron qui possède une activité promotrice donnant des ARN non-codants, ainsi que la compartimentation de la périphérie nucléaire en régions associées à des gènes activés ou silencieux ont été démontrés [27].

Il a été également suggéré que la régulation de la transcription soit contrôlée en partie par des transcrits antisense non-codants (ncRNA) [28, 29], ou des petits ARN double brins (dsRNA) [16].

1.1.3.3 Organisation des gènes *var*, *rif* et *stevor*

Les gènes *var*, *rif* et *stevor* sont formés de deux exons. Le premier exon des gènes *var* a une longueur qui varie entre 3,5 et 9 kb, et code pour la région variable extra-cellulaire de la protéine ainsi que la région transmembranaire. Le deuxième exon a une longueur comprise entre 1 et 1,5 kb et code pour la partie relativement conservée cytosolique riche en acides aminés acides (ATS, « acidic terminal sequence ») (Figure 6). Le premier exon des gènes *rif* et *stevor* a une longueur comprise entre 50 et 75 paires de bases (pb) et code pour une hypothétique séquence signal. Le deuxième exon a une taille de près de 1 kb, avec l'exon des gènes *rif* étant plus grand que celui des gènes *stevor*.

La région variable des protéines PfEMP1, codées par les gènes *var*, est formée de trois domaines identifiés : les « Duffy Binding Like » (DBL ayant sept sous-groupes : α , $\alpha 1$, β , γ , δ , ϵ et x), les « Cysteine rich Interdomain Region » (CIDR α , $\alpha 1$, β , γ) et le domaine constant 2 (C2) [30]. En fonction de l'architecture des protéines PfEMP1, ces dernières ont pu être classées en 17 groupes différents [31] (Figure 7). Trois sous-domaines structuraux ont pu également être définis (S1-3) [32, 33].

Il a également été démontré que les domaines DBL ont des spécificités pour certains récepteurs. Ainsi, les DBL β ont une affinité pour l'ICAM-1, tandis que les DBL δ adhèrent aux PECAM-1 (« platelet-endothelial cell adhesion molecule 1 »), et enfin les DBL α sont associés à la fixation des héparanes sulfates (HS) et du CR1 (« complement receptor 1 »).

Les domaines CIDR se composent de trois régions : une région nommée M2 décrite comme étant la région minimale nécessaire pour la fixation au CD36, entourée par les régions moins conservées M1 et M3. Il a été également montré que le domaine CIDR α peut fixer l'immunoglobuline M et le PECAM-1 [34].

Une première étude, réalisée par *Smith et al* et qui a englobée 65 domaines DBL issus de 20

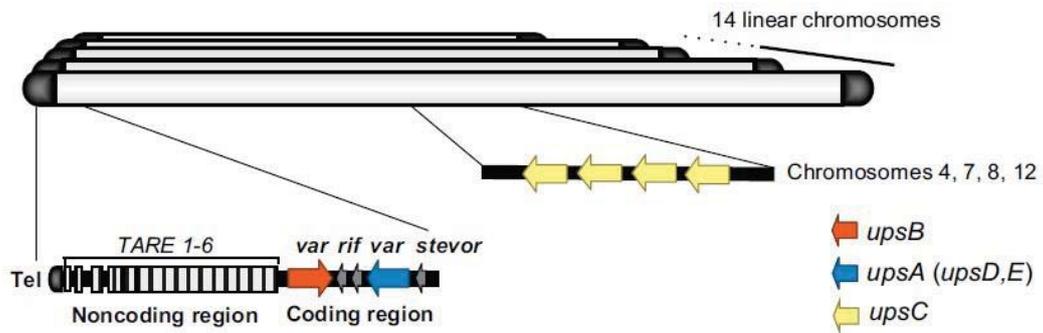


Figure 8 : Organisation des gènes *var*, *rif* et *stevor* au niveau des chromosomes [16]. Les chromosomes partagent une organisation globale semblable avec des régions télomériques non codantes (TARE1-6). Cette région est suivie par les gènes *var*, *rif* et *stevor* arrangés en tandem. Plusieurs tandems sont également trouvés au niveau des régions centrales des chromosomes.

PfEMP1 différents, avait permis d'identifier dix blocs homologues semi-conservés (HBa-j) d'une taille moyenne de 21 acides aminés, séparés par des régions hypervariables [35]. Une étude plus complète réalisée par *Rask et al.* a concerné l'analyse de 399 PfEMP1 provenant de sept génomes différents. Au cours de cette étude, la totalité des protéines et pas uniquement les domaines DBL ont été pris en compte. Ceci a permis d'identifier 628 blocs d'homologie (d'une taille moyenne de 18,6 acides aminés) couvrant en moyenne 83% d'un PfEMP1 donné, avec 4% uniquement d'auto chevauchement [34].

1.1.3.4 Classification des gènes *var*

Le séquençage complet du génome (clone 3D7) a permis de localiser les gènes *var*, *rif* et *stevor* dans les régions hautement polymorphiques des télomères des chromosomes alors que quelques-uns seulement existent près des centromères (Figure 8). Ainsi 60% des gènes *var* sont adjacents des régions sub-télomériques non codantes composées des TARE1-6 (« telomere associated repeat elements ») [9]. L'analyse génomique a révélé qu'un à trois gènes *var* peuvent exister dans chaque télomère. Ils peuvent être orientés de différentes manières : tête/tête, queue/queue (la plus fréquente) ou tête/queue. Quant aux gènes *var* de la zone centrale des chromosomes, ils sont orientés tête/queue. D'un point de vue évolutif, il est à noter que la synténie entre les différentes espèces de *plasmodium* est élevée pour les régions centrales des chromosomes, alors qu'elle est plus faible pour les régions sub-télomériques. Ceci suggère l'importance de ces régions pour la spéciation et l'adaptation à l'hôte [30].

Un premier traitement des séquences en amont des gènes *var* a révélé qu'il existe trois classes de séquences : upsA, upsB et upsC (11, 35 et 13 gènes respectivement). Ces trois classes d'ups sont associées préférentiellement à certains gènes *var*. Ainsi les upsB sont associés avec 22 des 24 gènes *var* télomériques et certains gènes centraux, les upsA sont associés aux gènes qui sont transcrits vers les télomères et les upsC sont associés aux *var* centraux. Une analyse plus affinée basée sur plus de critères dont la localisation chromosomale des gènes *var*, le sens de transcription, l'organisation des domaines des protéines qu'ils codent et la similarité de séquences des régions codantes et non codantes a permis de faire ressortir des groupes intermédiaires : B/A et B/C [36]. Les gènes du groupe B/A sont très similaires à ceux du groupe B de point de vue de la localisation et de l'orientation de la transcription et sont éloignés des télomères. Au contraire, les gènes du groupe B/C ont des séquences en amont upsB-like mais sont localisés dans les régions centrales des chromosomes [30]. D'un point de vue fonctionnel, il a été montré que l'adhésion au récepteur CD36 humain est limité aux

protéines PfEMP1 des groupes B et C. Les variantes du groupe A jouent un rôle différent et sont responsables du rosetting via l'interaction de leur DBL α avec les récepteurs CR1 présent à la surface des érythrocytes.

Le modèle actuellement admis est que tous les gènes *var* soient issus d'un *var* ancêtre-commun qui a été dupliqué en trois groupes (A, B et C) qui ont divergé plus tard. La comparaison des séquences codantes montre que le groupe A est très différent des groupes B et C. Ces deux derniers sont proches en séquence, même s'ils diffèrent par leur localisation au niveau des chromosomes. Cependant les régions de similarité sont essentiellement associées aux blocs d'homologie semi-conservés qui sont prédits pour former les éléments structuraux nécessaire à l'adhésion [30].

La comparaison des répertoires de gènes *var* de différents isolats a montré qu'il y a uniquement deux gènes conservés : *var1csa* et *var2csa* et un gène semi-conservé : *var type 3*. Les protéines codées par ces gènes sont identiques à plus de 75% sur plusieurs domaines alors que la plupart des autres protéines PfEMP1 (même les protéines qui partagent la même architecture des domaines) ont un taux d'identité de séquences en acides aminés inférieure à 50% entre les domaines homologues. De la même manière que le gène *var2csa* est nécessaire à l'établissement d'une infection congénitale via le récepteur chondroïtin sulfate A du placenta, les deux gènes *var1csa* et *var type3*, dont la fonction n'a pas encore été identifiée, pourraient jouer des fonctions importantes dans l'interaction hôte-parasite. Même s'ils ont des caractéristiques uniques, ces trois gènes se rapprochent des gènes du groupe A dans le sens où ils sont sub-télomériques, transcrits vers les télomères et ont des régions 5' en amont qui ressemblent à l'*upsA* [30].

1.1.3.5 Diversité des gènes *var* et conversion génique

Plusieurs travaux ont été consacrés à l'étude de la diversité des gènes *var* et à la manière dont cette dernière est générée. Le projet de séquençage du génome entier ainsi que les projets de séquençage d'autres isolats ont permis de constater que les séquences des gènes *var* d'un parasite donné sont extrêmement diverses, même si certaines familles de gènes peuvent être identifiées grâce à l'organisation relative des différents domaines des protéines PfEMP1 codées par ces gènes. La comparaison des génomes d'isolats provenant de zones géographiques distantes, ainsi que ceux de différentes souches de laboratoire, montrent que contrairement aux gènes de ménage qui sont hautement conservés, les gènes *var* sont très distants. Ceci suggère que ces derniers soient soumis à un mécanisme de génération de la

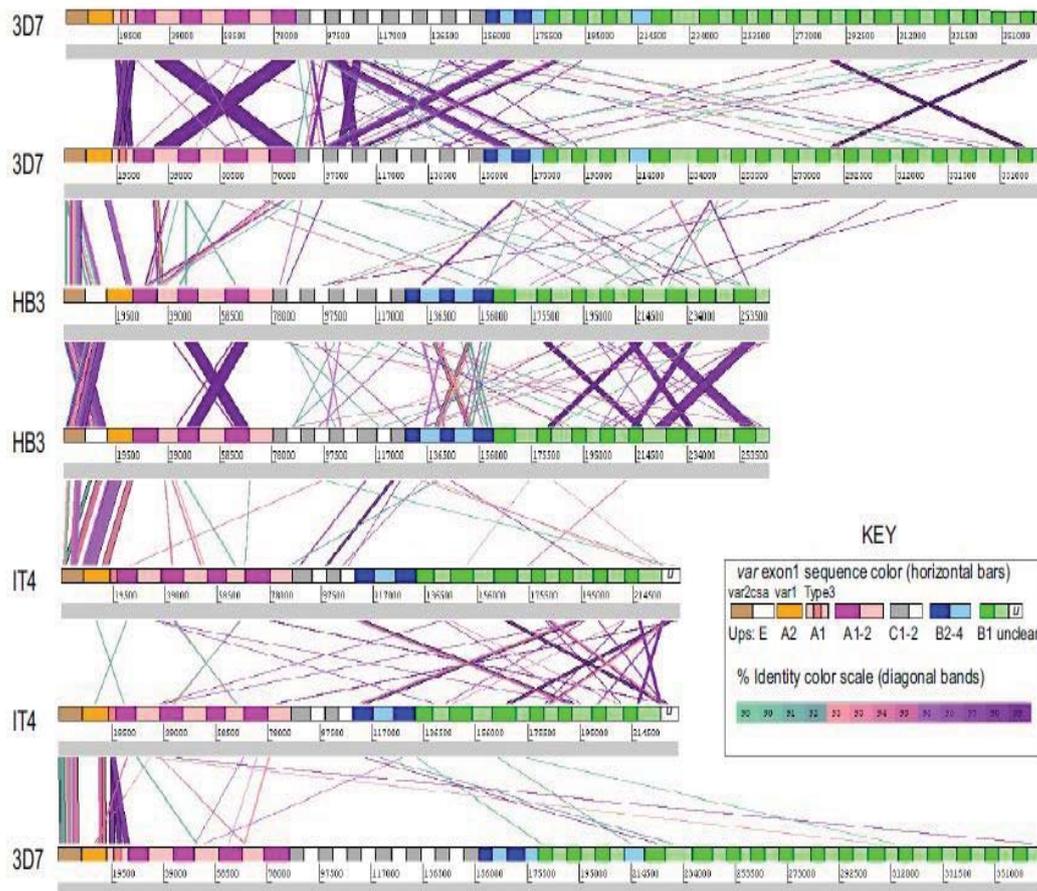


Figure 9 : Etude de la similarité de séquence entre les répertoires de gènes *var* des trois souches 3D7, IT4 et HB3 [30]. Les gènes *var* ont été arrangés en concatémères pour chaque souche (lignes horizontales) selon la région promotrice Ups. La similarité à plus de 90% par fragment de taille minimale de 90 nucléotides a été étudiée. Les barres diagonales représentent les correspondances d'identité obtenues entre les gènes *var* des différentes souches (3D7, IT4 et HB3). La coloration des barres représente le degré de similarité, bleu clair : 90%, violet : 99%. La comparaison des trois génomes les uns par rapport aux autres (barres diagonales entre 3D7 et HB3, HB3/IT4 et IT4/3D7) montre que les répertoires ne partagent qu'une faible identité de séquences suggérant que les gènes *var* sont divergents entre isolats. La comparaison de chaque génome par rapport à lui-même (barres diagonales 3D7/3D7 et HB3/HB3 et IT4/IT4) montre que plusieurs segments identiques (identité >90%) sont dupliqués dans chaque génome.

variabilité très accéléré qui résulte en une altération rapide des protéines codées [37].

Ainsi, l'analyse par *Frank et al* des gènes *var* du clone E5, qui est génétiquement différent mais proche du clone 3D7 (utilisé pour le projet de séquençage), montre qu'il ne partage que 50% de son répertoire avec 3D7. Une étude plus poussée des séquences codantes des gènes *var* spécifiques à E5 a montré qu'aucune d'entre elles, codant pour le DBL α , n'est présente en entier dans le génome du clone 3D7. Cependant, il a été observé qu'il existe plusieurs longs fragments dupliqués dans le génome d'E5, et qui sont identiques à des séquences présentes dans les gènes *var* du clone 3D7. La taille et l'identité élevée de ces séquences indique qu'il s'agit du résultat d'un événement récent [37].

Selon les observations de *kraemer et al*, 2007, où les auteurs ont analysé les séquences de l'exon 1 des gènes *var* de trois souches 3D7, IT4 et HB3, ces gènes auraient subi une duplication segmentale (Figure 9) mais qu'ils ont divergés extensivement plus tard entre les différentes souches [30].

Ces événements de transpositions duplicatives conduisent forcément à l'altération des séquences d'acides aminés menant à la production de nouvelles séquences hybrides qui sont probablement uniques à ce clone. De la même manière, l'analyse des gènes *rif* et *stevor* montre la présence de séquences dupliquées suggérant l'implication du même phénomène pour toutes les familles multigéniques [37].

La conversion de gènes est un mécanisme de génération de la diversité des antigènes qui a été observé chez différents agents infectieux notamment chez les parasites *Trypanosoma brucei* et *Babesia bovis*, ainsi que chez des bactéries tel que *Neisseria* et *Treponema*. Les mécanismes impliqués dans la conversion de gènes chez *P. falciparum* ne sont pas clairement identifiés mais les modèles actuellement proposés s'orientent vers l'implication de la voie de réparation homologue utilisée par les organismes pour réparer les cassures double brins de l'ADN chromosomique. Ce mécanisme étant basé sur l'utilisation de la similarité de séquence pour le choix de l'ADN matrice, il pourrait expliquer comment les trois types de gènes *var* ont divergé les uns des autres pour donner des sous-groupes [37].

Vu que, selon les cas, la durée de l'infection par *P. falciparum* peut être longue (des cas d'infection jusqu'à 480 jours ont été rapportés) et vu le nombre limité de gènes (~60 gènes *var* par parasite), il a donc été suggéré que le maintien de l'infection soit assuré *via* des vagues successives d'infection causées par des parasites génétiquement différents. Cette diversification du répertoire de gènes *var* dont disposent les parasites serait due à un phénomène de recombinaison mitotique qui permettrait au parasite d'évoluer continuellement

durant la phase érythrocytaire [37, 38].

Les études menées par *Freitas junior et al.* (2000) et *Taylor et al.* (2000) où les auteurs analysent la descendance du croisement de deux génotypes différents, montrent que celle-ci hérite d'un matériel génétique non-parental avec une fréquence très élevée. Les auteurs suggèrent que cette recombinaison qui a lieu pendant la méiose dans le moustique serait facilitée par l'architecture nucléaire permettant la localisation simultanée des gènes *var*, aussi bien sub-téломériques que centraux, dans des compartiments de la périphérie du noyau [39, 40].

1.1.4 Variabilité antigénique chez d'autres protozoaires

L'existence de familles multigéniques a été observée chez différents pathogènes de différents règnes. Chez les bactéries, les organismes les plus connus sont *Borrelia hermsii* (~30 gènes *vsp* et *vlp*) et *Neisseria meningitidis* (~19 gènes *pil* et 4-5 gènes *opa*) et *Neisseria gonorrhoeae* (~19 gènes *pil* et 11 gènes *opa*). Chez les champignons, *Pneumocystis carinii* possède ~85 gènes *msg*, ~85 gènes *msr* et ~85 gènes *pvt*. *Candida albicans* possède 9 gènes *als* et *Candida glabrata* possède 17-23 gènes *epa*. Chez les protozoaires, les familles multigéniques ont été observées chez trois parasites en plus de *Plasmodium*. Ainsi, le parasite *Trypanosoma brucei* possède ~1000 gènes *vsg*, *Babesia bovis* possède ~130-160 gènes *ves1a/ves1b* et *Giardia lamblia* possède ~150 gènes *vsp* [21].

Ces microorganismes pathogènes ont essentiellement recours à quatre mécanismes moléculaires pour assurer la variabilité antigénique. Le premier mécanisme est le contrôle de la transcription via des modifications épigénétiques, ce mécanisme est également connu sous le nom de commutation *in-situ*. Lors de ce processus, un seul gène du répertoire est transcrit alors que tous les autres sont dans un état silencieux. Le deuxième mécanisme est la conversion génique qui implique l'utilisation d'un seul site d'expression. La variabilité a lieu lorsque le gène présent au niveau du site d'expression est remplacé par un autre gène archivé dans un autre locus. Le troisième mécanisme est la conversion génique par segments. Lors de ce processus, les gènes et/ou les pseudogènes de la librairie transmettent un ou plusieurs fragments de leurs séquences codantes au gène transcrit. Le dernier mécanisme est le réarrangement, généralement par inversion d'ADN, des séquences promotrices d'un gène silencieux. Commun aux mécanismes impliquant la réorganisation topologique de l'ADN, est la nécessité d'avoir une protéine ou un complexe ayant les propriétés de pouvoir reconnaître, se fixer, couper et relier les brins d'ADN. Des solutions basées sur ce constat ont été testées et

mises en place chez plusieurs espèces. Quelques exemples sont décrits dans les paragraphes suivants.

1.1.4.1 La variation antigénique chez *Trypanosoma*

Trypanosoma brucei est un parasite unicellulaire flagellé responsable de la maladie du sommeil. La surface des trypanosomes est recouverte de la glycoprotéine VSG (Variable Surface Glycoprotein) qui est codée par plus de 1000 gènes et pseudogènes mais dont 7 % seulement sont des gènes fonctionnels. 80-90% des gènes *vsg* sont portés par les régions subtélomériques des chromosomes et une partie de la bibliothèque est portée par les mini-chromosomes qui sont présents dans la cellule [41].

Le cas de *T. brucei* est particulièrement intéressant car le parasite est capable d'échapper au système immunitaire en exprimant un seul de ses gènes grâce à différents techniques. Parmi celles-ci, le mécanisme de conversion génique est celui qui joue le rôle le plus important. Ce processus se manifeste par l'échange réciproque de gènes entre les 15 sites de transcription VSG-ES (VSG Expression Site) identifiés et localisés au niveau des télomères [42].

Lors de la conversion génique, un gène fonctionnel donné est copié d'un site d'expression silencieux vers un site actif. Dans ce cas, la région copiée englobe aussi bien la région codante que les régions qui la bordent. 90% des cassettes *vsg* variables sont bordées par deux régions homologues qui semblent jouer un rôle important dans la conversion génique : une région 3' conservée et des répétitions d'une séquence de 70 pb. Les études de mutation des protéines RAD51, BRCA2 et RAD51-3 montrent que la commutation a lieu en partie grâce à la recombinaison homologue médiée par RAD-51. Le modèle actuellement admis est celui dans lequel la réaction commence par une cassure double brin de l'ADN au niveau de la région des répétitions à 70 pb. Une fois la cassure introduite au niveau de la région de répétition d'un premier gène VSG présent dans un site d'expression, l'ADN est pris en charge par la machinerie de réparation des cassures doubles brins pour produire un ADN simple brin. Cet ADN simple brin issu du premier gène attaque par homologie le duplexe d'ADN d'un deuxième gène silencieux. Le complexe ainsi formé est résolu ensuite selon le modèle de réplication induite par cassure (BIR, Break-Induced Replication). La cause de l'introduction des cassures double brins n'a pas encore été identifiée mais les hypothèses actuellement proposées sont soit l'existence d'une endonucléase dédiée soit la présence de structures instables lors de la réplication des régions à répétitions. Des études antérieures avaient montré cependant que la présence/absence de la répétition n'a aucun effet sur la fréquence de commutation [42, 43].

Il existe également chez *Trypanosoma* un mécanisme de conversion génique par segments encore mal compris basé sur des réactions de recombinaison impliquant les régions codantes des gènes. Ce mécanisme complémentaire de la conversion génique joue un rôle important dans l'expansion de la librairie de gènes *vsg* en permettant la création de nouveaux gènes fonctionnels à partir de différentes portions des différents gènes et pseudogènes présents dans la librairie [41-44].

Le dernier mécanisme utilisé par *T. brucei* est basé sur des modifications épigénétiques qui sont assurées par l'histone déacétylase TbSir2 et l'histone méthyltransférase DOT1B qui permettent l'activation d'un seul des VSG-ES [45-47].

1.1.4.2 La variabilité antigénique chez *Giardia*

Giardia intestinalis (anciennement appelé *Giardia lamblia*) est le parasite responsable de la parasitose intestinale humaine giardiose. La surface de ce parasite est couverte par la protéine VSP (« variable surface proteins ») qui est codée par un répertoire constitué d'approximativement 200 gènes. Ces gènes sont dépourvus de séquences promotrices spécifiques et sont organisés en « cluster » de deux à neuf gènes.

Il a été montré que chez *Giardia*, la variabilité antigénique est soumise à une régulation à deux composantes : la première, découlant de l'épigénétique, assure la commutation *in-situ* entre les gènes. La deuxième composante agit au niveau post-transcriptionnelle par un mécanisme similaire aux RNAi. En effet, la présence de petits ARN nucléolaires (snoRNA) qui ont une complémentarité de séquence aux extrémités 3' de 22 gènes *vsp* a été démontrée [48]. On observe ainsi que plusieurs gènes *vsp* peuvent être transcrits mais que seul celui qui correspond à la protéine exprimée à la surface du parasite est accumulé dans le cytoplasme. Ces études montrent que le contrôle se fait par la protéine RdRP (« RNA-dependent RNA polymerase ») qui est activée lorsque plusieurs transcrits existent simultanément dans la même cellule. La protéine RdRP est une polymérase d'ARN qui produit des transcrits antisense pour donner des duplexes d'ARN (dsRNA) qui sont dégradés par la protéine DICER, une protéine à activité endonucléase dsRNA dépendante [49-51].

1.1.4.3 La variabilité antigénique chez *Babesia*

Babesia est un parasite protozoaire (plus de 100 espèces) transmis par la tique et qui est à l'origine de la babésiose. L'espèce la plus étudiée est *B. bovis* qui est l'agent responsable de la babésiose bovine. Dans le but d'échapper à la réponse immunitaire de l'hôte, *B. bovis* a

recours à la variabilité antigénique qui lui permet de modifier son antigène de surface majeur constitué de la protéine hétérodimérique VES1 (variant erythrocyte surface antigen-1). Les deux sous-unités de VES1 (VES1a et VES1b) sont codées par deux groupes de gènes (*ves1a* et *ves1b*) de la même famille multigénique *ves* composée de 130-160 gènes. Le séquençage du génome de *B. bovis* a montré que la majorité des gènes *ves1* sont organisés en paire de gènes (*ves1a/ves1b* ou *ves1a/ves1a*) orientés tête-tête et qui sont séparés par une région intergénique (<500 pb) hautement conservée. Les gènes qui ne sont pas compris dans ce type de structures sont majoritairement tronqués et leur fonction n'est pas encore connue. [52]. La transcription des *ves1* a lieu au niveau d'un site appelé LAT (Locus for active transcription) composé d'une paire (*ves1a/ves1b*) avec la région intergénique qui jouerait le rôle de promoteur bidirectionnel. Il a été suggéré que plusieurs sites LAT peuvent exister et que dans ce cas la transcription des gènes suivrait un mécanisme de transcription mutuellement exclusive avec un seul LAT transcrit à la fois [53].

B. bovis utilise un mécanisme de conversion génique par segments qui lui permet de changer graduellement la séquence du gène transcrit en y incorporant des segments de petites tailles issus des régions équivalentes des autres gènes *ves1*. La machinerie moléculaire n'a cependant pas encore été identifiée [54].

Le séquençage du génome de *B. bovis* a permis de montrer l'existence d'une deuxième famille multigénique de 44 membres : *smorf* (« small open reading frame »). Contrairement aux familles multigéniques de *P. falciparum*, les deux familles *ves* et *smorf* de *B. bovis* ne sont pas localisés dans les télomères mais sont distribués sur toute la longueur des chromosomes. Il a été montré que les gènes *smorf* ne sont pas soumis à une transcription mutuellement exclusive et que par conséquent différents transcrits peuvent exister dans la même cellule. Cependant, il n'a toujours pas été démontré s'il existe un contrôle post-transcriptionnel qui permet d'assurer la variabilité antigénique de la protéine codée par cette famille de gènes [55].

1.1.4.4 L'implication potentielle d'une recombinaison dans le phénomène de variabilité chez *P. falciparum*

Ainsi, les études montrent qu'il existe chez les protozoaires différents mécanismes qui permettent d'assurer la variabilité antigénique. Ces mécanismes sont basés sur des modifications épigénétiques, des modifications post-transcriptionnelles, des échanges de télomères ou des événements de conversion génique par gène entier ou par segments mais aucun mécanisme basé sur l'utilisation d'une recombinaison à site spécifique n'a été observé jusqu'à aujourd'hui.

Le séquençage du génome entier de *P. falciparum* et les analyses d'annotation ont permis d'identifier une protéine pouvant être un nouveau membre de la famille des recombinases à site spécifique de type tyrosine. Même si les études actuelles montrent que dans le cas de *P. falciparum* la variabilité antigénique des gènes *var*, *rifin* et *stevor* semble être soumise à un mécanisme de contrôle par modifications épigénétiques, l'hypothèse de la présence d'un deuxième mécanisme basé sur l'implication d'une recombinase n'est pas à écarter.

De plus, les études de *Kramer et al* [30], et de *Rask et al* [34] montrent que les gènes *var* sont organisés en blocs et qu'ils sont soumis à un mécanisme de recombinaison. Ce mécanisme implique l'échange de fragments de différentes tailles entre les gènes *var* et permet au parasite de diversifier sa librairie en générant continuellement de nouvelles variantes. *Kramer et al* [30] suggèrent que cette recombinaison soit assurée par un mécanisme de transposition duplicative mais l'existence d'une deuxième composante basée sur des réactions de recombinaison spécifique médiées par une recombinase dédiée n'est pas à écarter.

1.2 La recombinaison à site spécifique

Les recombinases à tyrosine (Y-SSR) font partie, avec les recombinases à sérine (S-SSR), de la famille des recombinases conservatrices. Elles sont dites conservatrices car elles ne nécessitent pas la synthèse *de novo* d'ADN. En effet, le nombre de liaisons phosphodiester présent au début de la réaction reste inchangé et aucune synthèse ou dégradation d'ADN n'a lieu. Classiquement, la recombinaison à site spécifique implique la participation de deux duplexes d'ADN et de quatre monomères de recombinases spécialisées. Durant la recombinaison conservatrice, les deux brins des deux partenaires sont clivés à des positions spécifiques, échangés puis reliés aux nouveaux partenaires. Contrairement à la recombinaison homologue qui nécessite une homologie de séquence étendue sur de grands fragments d'ADN, les recombinases à sites spécifiques catalysent des réactions de recombinaison entre des fragments dont l'identité de séquence est importante uniquement au niveau des régions de recombinaison qui bordent les ADN recombinés [56].

La recombinaison à site spécifique est impliquée dans de nombreux mécanismes biologiques tels que l'intégration/excision de phages (phage λ ou HP1), la résolution des co-intégrats résultants de la transposition reliés au transposon Tn3, l'amplification du plasmide 2 μ m de la levure, la résolution des concatémères pour le phage P1 ou encore dans l'inversion d'ADN tel qu'elle est réalisée par Hin chez *Salmonella* ou Gin du phage Mu [57, 58].

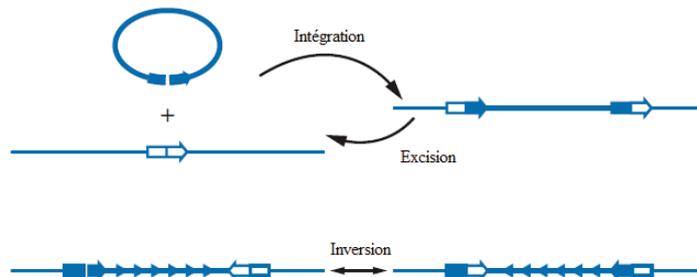


Figure 10 : Les trois différentes issues de la recombinaison à site spécifique. En fonction de l'organisation initiale des ADN substrats, l'issue peut être l'intégration si les deux sites sont présents sur deux ADN différentes, l'excision si les deux sites sont présents sur le même ADN et s'ils sont orientés dans la même direction ou l'inversion si les sites sont sur le même ADN et s'ils sont orientés dans deux directions opposés [57].

Chaque Y-SSR possède un site ADN spécifique de recombinaison. Chaque site de recombinaison (20 à 30 pb) est composé d'une paire de séquences répétées inversées appelés demi-sites disposées autour d'une région centrale de 6-8 pb appelée région cross-over. La recombinaison a généralement lieu entre deux sites d'ADN qui ont en commun cette région homologue centrale. La région centrale est asymétrique et donne de ce fait un sens au site de recombinaison [58].

En fonction de l'organisation initiale des ADN parentaux, la recombinaison à site spécifique peut avoir trois issues (Figure 10) : l'intégration, l'excision ou l'inversion. L'intégration résulte de la recombinaison de deux sites présents sur deux molécules d'ADN différentes dont l'une peut être sous forme circulaire. Pour les sites localisés sur la même molécule d'ADN, l'issue est déterminée par leur orientation l'un par rapport à l'autre; l'excision a lieu si les deux sites sont orientés dans la même direction alors que l'inversion a lieu lorsque les deux sites sont orientés dans deux directions opposées.

Toutes les recombinases à site spécifique identifiées à ce jour peuvent être classées dans deux groupes, en fonction du nucléophile qui catalyse la réaction de recombinaison : les Y-SSR et les S-SSR. Ces deux familles semblent avoir évolué séparément et sont éloignées aussi bien au niveau des séquences d'acides aminés qu'au niveau des mécanismes d'action. Les Y-SSR clivent et relient un seul brin d'ADN à la fois alors que les S-SSR clivent les deux brins en même temps. Les Y-SSR forment une liaison covalente phospho-tyrosine en 3' libérant ainsi un OH en 5', alors que les S-SSR forment une liaison covalente phospho-sérine en 5' libérant un OH en 3'. Les seules similarités qui existent entre ces deux familles sont qu'elles ont recours à la substitution nucléophile par une attaque en ligne du phosphate scissile et que, quelle que soit la fonction dans laquelle elles sont impliquées, elles reconnaissent des séquences d'ADN spécifiques et ne catalysent les réactions de réarrangement qu'entre ces dernières [59].

1.2.1 Les recombinases à tyrosine

Historiquement, la fonction des Y-SSR était associée à celle de l'intégrase du phage λ (λ -Int) impliquée dans l'intégration et l'excision de son génome. Cependant, il a été montré que les Y-SSR sont impliquées dans de nombreux processus biologiques.

Les exemples les plus étudiés autres que λ -Int sont les recombinases Flp de la levure, Cre du phage P1 et XerC et D bactériennes.

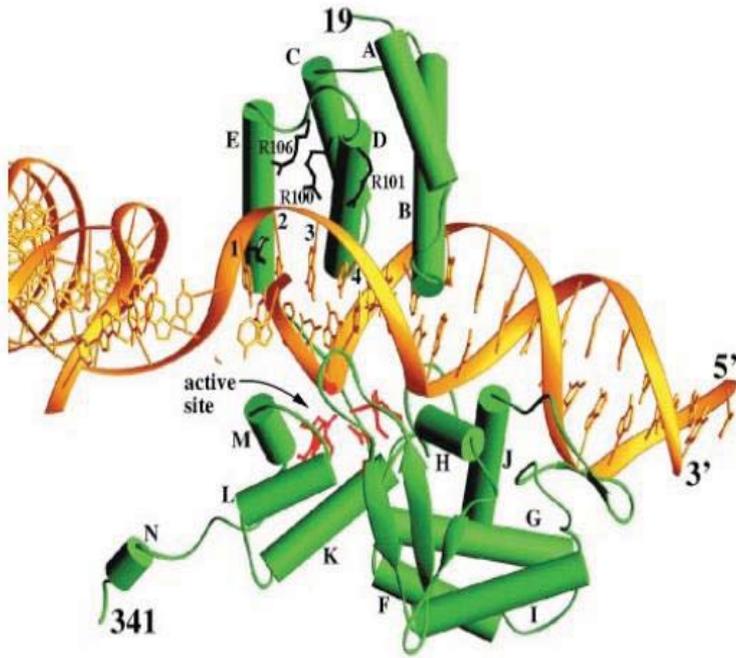


Figure 11 : Structure d'un protomère de Cre en complexe avec l'ADN. La structure montre que Cre (colorée en vert) est organisé en deux domaines N-terminal et C-terminal. Ces deux domaines s'organisent dans l'espace en pince (« C-shaped clamp ») qui se referme autour de l'ADN substrat (jaune). Les deux domaines sont formés de structures secondaires majoritairement en hélices- α (cylindres verts). Le domaine N-terminal : hélices A-E. Le domaine C-terminal : hélices F-N et d'un feuillet β unique (flèches vertes). Les acides aminés du site catalytique sont colorés en rouge [64].

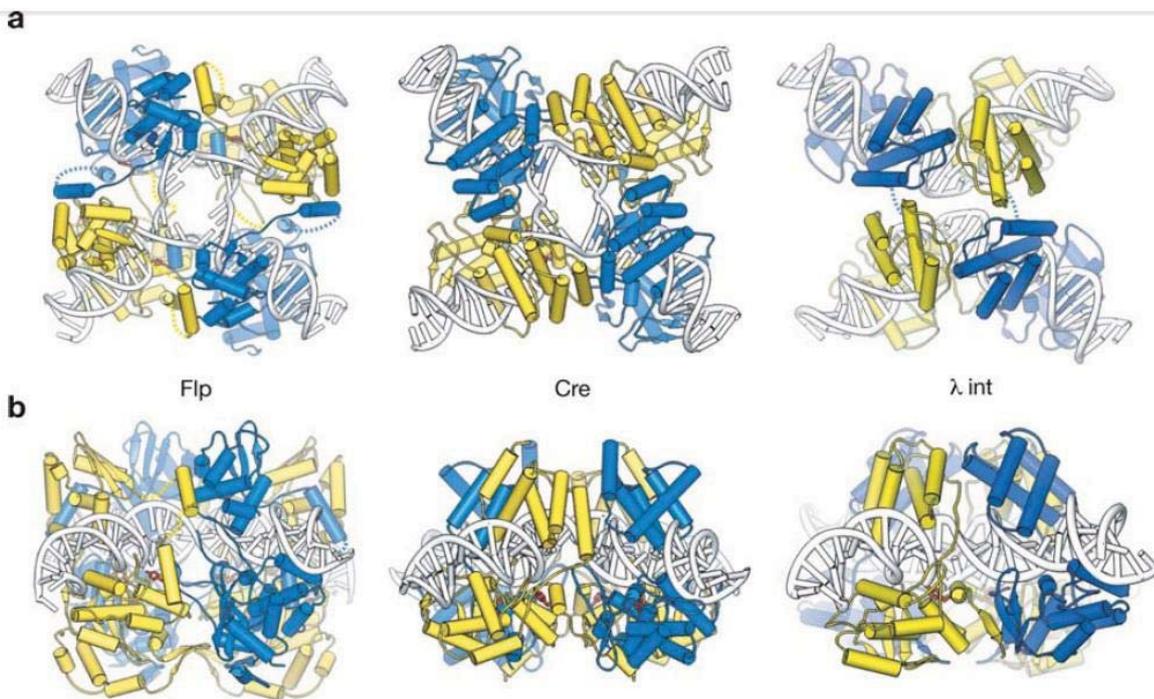


Figure 12 : Comparaison des structures des tétramères de Cre, Flp et du « core binding domain » de λ -Int [57]. (Panneaux de gauche : Flp; Panneaux du centre : Cre; Panneaux de droite : λ -Int). PDB IDs : Cre : 3Crx, Flp : 1M6X et λ -Int : 1Z19. (a) Vue avec les domaines N-terminaux en avant et les domaines C-terminaux catalytiques en arrière. (b) Vue avec une rotation de 90° C, l'ADN est horizontal au plan. Deux monomères sont représentés dans le premier plan et les deux autres dans l'arrière-plan. Les domaines N-terminaux sont au-dessus de l'ADN et les domaines catalytiques en-dessous. Pour chaque structure deux monomères sont colorés en jaune et les deux autres en bleu. Les structures montrent que les protéines partagent la même organisation tertiaire avec deux domaines N- et C-terminaux organisés majoritairement en hélices- α .

L'enzyme Flp sert à maintenir un nombre élevé de plasmides 2 μ m chez la levure, en inversant des séquences, par recombinaison entre les sites *frt* de 34 pb [60]. La recombinase Cre est une Y-SSR codée par le phage P1, chez qui elle sert à maintenir le génome sous forme monomérique en catalysant la résolution des dimères qui sont produits lors de la réplication de l'ADN [61]. Cette recombinaison a lieu entre les sites *loxP* de 34 pb. Les deux systèmes impliquant Flp et Cre sont des systèmes simples et ne nécessitent pas, *in vitro*, l'intervention de sites accessoires ou d'autres partenaires.

Le système XerC/D est un système conservé entre la plupart des bactéries et assure la ségrégation des chromosomes en agissant sur les sites *dif* pour convertir des dimères de chromosomes en monomères. XerC agit également sur les multimères des plasmides ColE1 et pSC101 en catalysant la recombinaison des sites *cer* et *psi* respectivement. L'utilisation de protéines accessoires dépend du contexte et des séquences d'ADN [62].

Les Y-SSR sont plus répandues chez les bactéries mais sont également trouvées chez certaines archées, et même chez des eucaryotes tels que des champignons ou des ciliés. Le nombre de protéines découvertes appartenant à cette famille ne cesse de croître et plus de 100 membres avaient été décrits en 1998 [56]. Selon *Buck et Flapan* une analyse par PSI-BLAST leur aurait permis d'identifier plus de 1000 séquences reliées à ce groupe [63].

1.2.1.1 Organisation structurale des recombinases à tyrosine

La structure du domaine catalytique de λ -Int a été rapportée en 1997 et depuis, plusieurs autres structures ont été décrites dont celles du domaine catalytique de l'intégrase HP1, de XerD, de l'intégrase d'intégrons de classe 4 (IntI4, de *Vibrio cholerae*) en présence d'ADN attC, de Flp et de Cre en présence d'ADN représentant les différentes étapes de la réaction [64-69]. Même si les séquences en acides aminés ne sont pas conservées, les études structurales montrent que les Y-SSR sont organisées le plus souvent en deux domaines N et C-terminal. Ces deux domaines forment une pince en C (« *C-shaped clamp* ») qui se referme sur l'ADN de part et d'autre (Figures 11, 12 et 13B). Le repliement du domaine catalytique C-terminal est très conservé (Figure 12) et présente des motifs reconnaissables. Certaines protéines de cette famille telle que λ -Int sont formées de trois domaines distincts (Figure 13, A). En effet, elles possèdent un domaine supplémentaire présent en amont des deux domaines rencontrés classiquement chez les autres Y-SSR. Ce domaine, appelé « *arm binding domain* », fixe des séquences ADN auxiliaires (« *arm sites* »), tandis que les deux autres domaines présents en aval fixent le site canonique de recombinaison (« *core sites* ») [70].

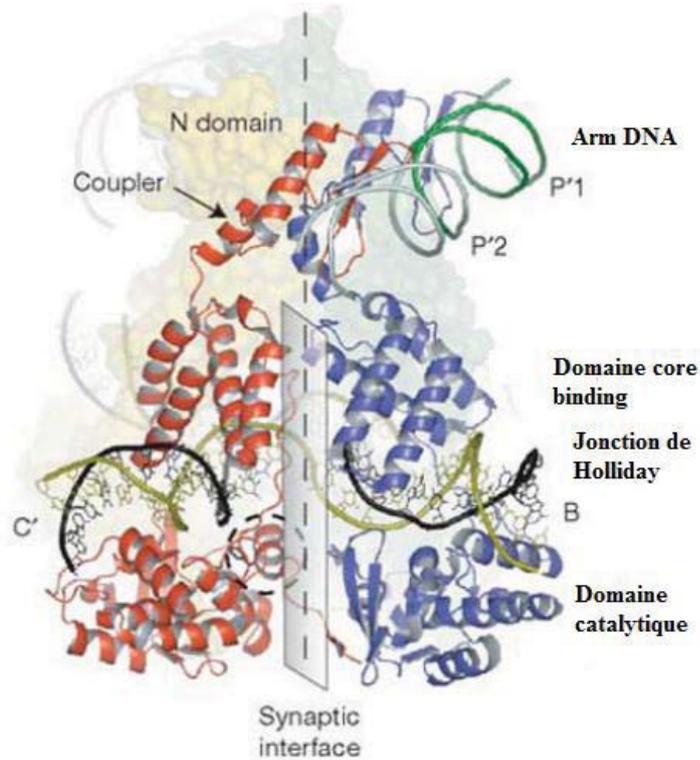
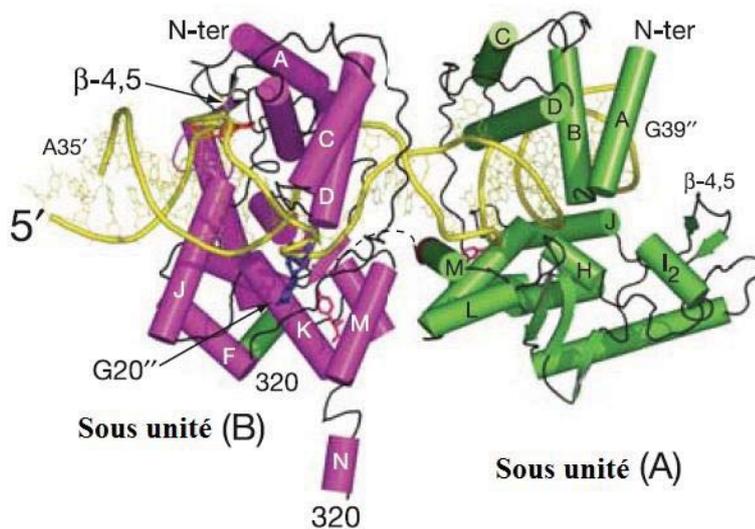
A**B**

Figure 13 : Organisation structurale de deux protomères de λ -Int en complexe avec l'ADN substrat (A) et de deux protomères de l'intégrase d'intégrons IntI4 (B). **A :** λ -Int est formée de trois domaines (du haut vers le bas) : le domaine « *arm binding* », le domaine « *core binding* » et le domaine catalytique. Le domaine « *arm-binding* » est connecté aux deux autres domaines par un connecteur (« *coupler* ») formé d'une hélice α . La figure montre l'interface entre un protomère actif (coloré en rouge à gauche) et un protomère non actif (coloré en bleu à droite). La région portant la tyrosine catalytique (cercle en pointillés) est repliée en hélice α dans le protomère actif (rouge) [67]. **B :** Deux protomères de IntI4 sont représentés : le protomère actif en vert (sous unité A) et le protomère non actif en magenta (sous unité B). L'intégrase d'intégrons IntI4 est formée de deux domaines : N-terminal (hélices A-C) et C-terminal (hélices F-N et d'un feuillet β) [66].

Les structures résolues de Cre, Flp, IntI4 et λ -Int en présence de leurs substrats montrent que les complexes adoptent des assemblages similaires. Les tétramères ont à la fois une symétrie rotationnelle d'ordre 2 et une pseudo-symétrie d'ordre 4. Le complexe adopte une conformation planaire des protéines avec l'ADN inclus dans le plan (Figure 12).

Ces mêmes structures montrent qu'il existe un contact très important entre les protéines et l'ADN. Il a ainsi été observé que le domaine N-terminal interagit avec le sillon majeur de l'ADN près de la région centrale. Le domaine C-terminal interagit avec les sillons mineur et majeur du substrat ADN [64-67].

Les structures montrent qu'en plus de l'interaction entre les protéines et l'ADN, chaque domaine catalytique d'un protomère donné interagit avec celui du protomère voisin en échangeant, d'une manière cyclique, une partie de son domaine C-terminal. Il s'agit du segment qui suit immédiatement l'hélice portant la tyrosine catalytique. Dans les cas de Cre et des intégrases d'intégrons, la partie échangée est une hélice (α -N pour Cre, Figure 11 ; α -N pour IntI4, Figure 13 B) [64, 66]. Le segment échangé par λ -Int forme un brin bêta lorsqu'il est tassé au sein du monomère adjacent et permet la structuration de la région portant la tyrosine catalytique (Figure 13A) [67]. Les protéines Flp et l'intégrase du virus SSV1 constituent des exceptions car c'est le domaine qui porte la tyrosine catalytique qui est échangée (Figure 14) [65, 71]. Le site actif est par conséquent assemblé en *trans*. Cependant, la structure de Flp montre que le reste du domaine C-terminal (hélices O-Q et un feuillet β) est replié du côté de son monomère d'origine (Figure 14) [65].

L'implication du domaine N-terminal dans la formation de la synapse varie selon les cas. Dans le cas de Cre, le domaine N-terminal participe fortement à la formation du complexe avec des surfaces étendues entre les monomères via des interactions hélice-hélice (α -A/ α -E) [64]. En revanche, dans le cas de Flp, les domaines N-terminaux des différents monomères échangent une deuxième hélice qui s'insère dans une poche hydrophobe (Figure 14) [65]. Dans le cas de λ -Int, les domaines fixant les sites accessoires interagissent fortement entre eux et avec les domaines centraux des monomères adjacents (Figure 13 A) [67].

Concernant la spécificité vis à vis de la séquence d'ADN de ces recombinaises, chaque domaine protéique entourant les régions centrales du site de recombinaison insère une hélice dans le sillon majeur de l'ADN. Cependant, les contacts entre chaînes latérales et bases sont rares et ne participent que faiblement à la spécificité de la réaction. Les contacts médiés par des molécules d'eau, les interactions avec le squelette phosphate de l'ADN, ainsi que la

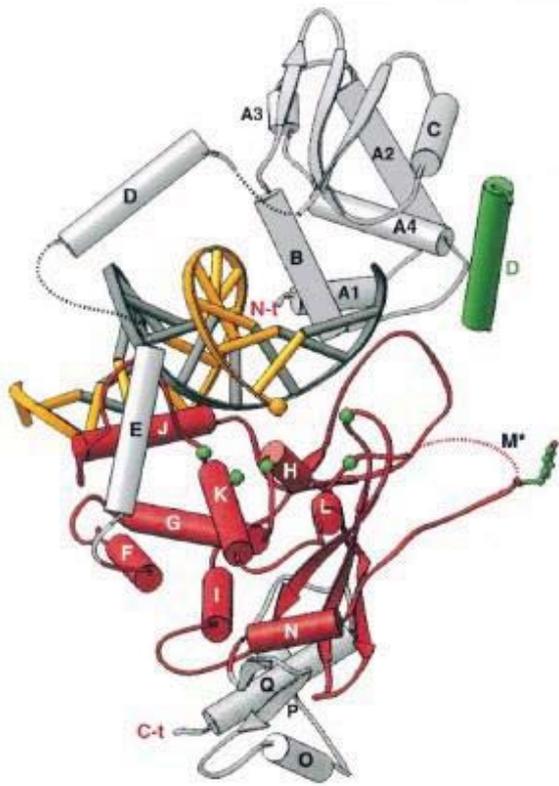
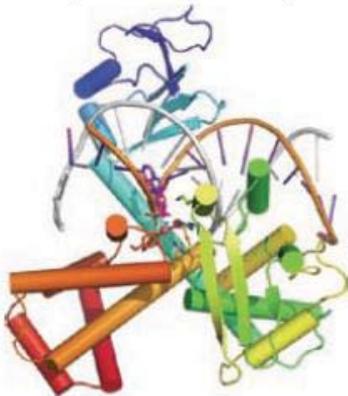


Figure 14 : Structure tertiaire d'un protomère non actif de Flp en complexe avec l'ADN [65]. Flp est organisé en deux domaines : un domaine N-terminal (hélices A1-C et un feuillet β) et un domaine C-terminal (hélices E-Q et deux feuillets β). Les deux domaines sont connectés par l'hélice D. L'hélice D est échangée par les différents protomères et participe à la formation de la synapse. La tyrosine catalytique est portée par la région échangée M* qui permet la formation d'un site catalytique en *trans*. Cette région est déstructurée dans les protomères non actifs et se structure en l'hélice M dans les protomères actifs. Le domaine homologue au domaine C-terminal de Cre est coloré en rouge.

A Topoisomerase IB (3IGC)



B Cre-LoxP (3C29)

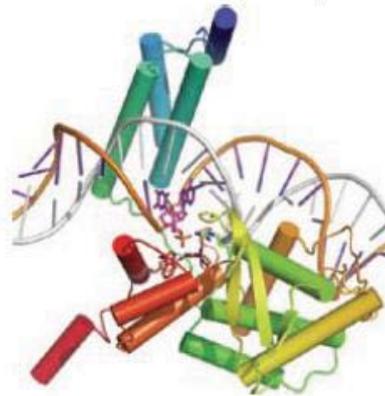


Figure 15 : Comparaison des structures de (A) : la topoisomérase du virus de la variole (TopoIB, PDB: 3IGC) (B) : Cre (Y-SSR, PDB: 3C29). La comparaison des structures de Cre (Y-SSR) et la topoisomérase du virus de la variole (TopoIB) montre que les deux familles ont une organisation tertiaire de leur domaine C-terminal catalytique similaire (coloré en vert, jaune et rouge) [59].

lecture indirecte des conformations dépendant de la séquence jouent probablement un rôle important. La spécificité peut également être renforcée à l'étape de catalyse par l'accessibilité au phosphate du site catalytique.

1.2.1.2 Les recombinaisons à tyrosine et les topoisomérases IB

Malgré les différences de séquences en acides aminés et de fonctions, les structures montrent que les Y-SSR et les topoisomérases IB possèdent de nombreuses similarités (Figure 15). En effet, les deux domaines catalytiques sont situés dans la partie C-terminale et partagent les mêmes éléments structuraux et la même topologie de repliement. En supplément du domaine C-terminal, les deux classes de protéines contiennent un domaine N-terminal qui participe aussi à la fixation de l'ADN. Bien que les éléments structuraux des domaines N-terminaux soient différents (hélices α pour les Y-SSR *versus* feuillet β central formé de 5 brins pour les topoisomérases IB), ces deux domaines se fixent à l'ADN via une interaction avec le sillon majeur situé du côté opposé au domaine catalytique [64, 72]. Ces deux classes de protéines partagent également le même mécanisme d'action impliquant la formation d'une liaison covalente entre la protéine et l'ADN via la tyrosine catalytique. La réaction ne nécessite pas la participation d'ions métalliques mais se base sur la charge positive des chaînes latérales des acides aminés du site actif pour activer le transfert du phosphoryle. Elle ne nécessite pas de ressources énergétiques (ATP) et utilise, *in vivo*, l'énergie stockée dans les super-enroulements pour assurer le relâchement de l'ADN [73]. Cependant, *in vitro* la réaction est aussi possible, et on imagine que les énergies de fixation et de changement de conformation lors de la constitution de la synapse (tétramérisation et torsion imposée sur l'ADN) sont autant de réserves mobilisables pour permettre à la réaction de se faire. Cependant, il existe clairement des différences. En particulier les Y-SSR et les S-SSR sont par définition spécifiques à une séquence donnée alors que les topoisomérases ne le sont généralement pas. Il a été montré que la topoisomérase du virus de la variole a une activité séquence-dépendante [74]. La deuxième différence majeure est que, durant la réaction de recombinaison, les SSR catalysent la coupure de deux doubles hélices (quatre brins au total) par cycle tandis que, dans le cas des topoisomérases, le clivage concerne un seul, voir deux brins par cycle [59].

Les rôles des résidus autres que la tyrosine nucléophile du site actif ont été mieux compris au fil des différentes études menées dans divers laboratoires, notamment les études des structures des Y-SSR précédemment citées et celles de certaines protéines de la famille des DNA topoisomérase IB [72, 75], ainsi que les études de mutagenèse [72] [76]. Ces études ont

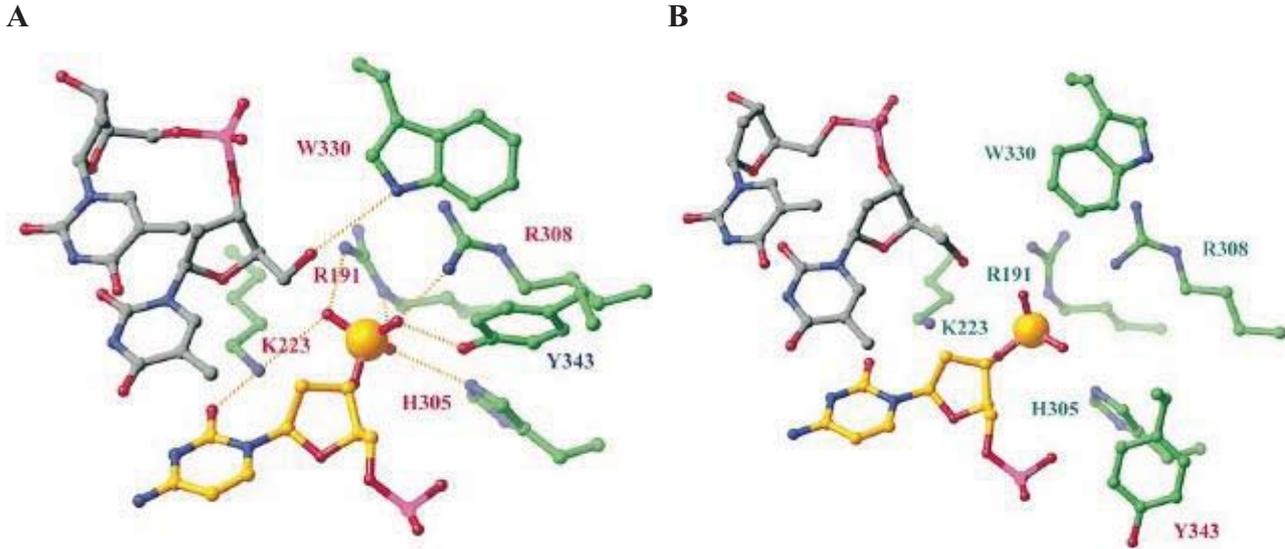


Figure 16 : Organisation du site catalytique de Flp. Le site catalytique est formé des acides aminés R-K-H-R-W-Y. Le phosphate scissile est coloré en jaune. La figure montre l'organisation des deux états actif et non actif du site catalytique. **A** : Site catalytique d'un monomère actif avec le phosphate scissile pentacoordonné grâce aux chaînes latérales des acides aminés R-K-H-R-W-Y. **B** : site catalytique d'un monomère non actif. Le phosphate n'est pas coordonné et la réaction ne peut pas avoir lieu [65].

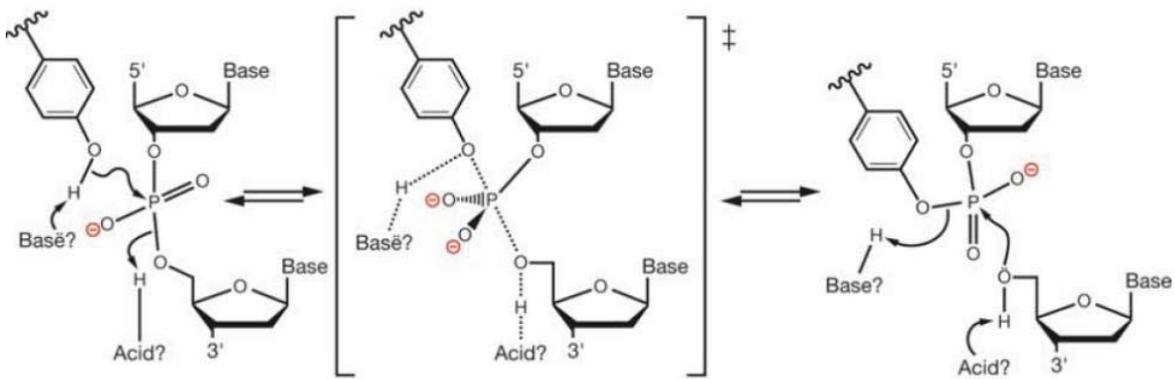


Figure 17 : Réaction de transfert d'un phosphoryl par une recombinase à tyrosine. La coupure de l'ADN se fait par une réaction en ligne de l'hydroxyle de la chaîne latérale de la tyrosine sur le phosphate de l'ADN. La réaction est facilitée par la présence d'une base qui accepte le proton de la chaîne latérale et d'un acide qui joue le rôle de donneur de proton à l'oxygène de l'ADN. (++) : état transitoire [57].

permis d'identifier les résidus hautement conservés (R-K-H-R-H/W-Y) qui caractérisent la famille des Y-SSR et qui s'assemblent dans l'espace pour former le site actif (Figure 16) [57, 65, 77, 78]. Il a été observé dans le cas des topoisomérases IB humaine et virale que le site actif est formé d'une tyrosine nucléophile et de la triade d'acides aminés positivement chargés K-R-H.

1.2.1.3 Mécanisme de recombinaison

1.2.1.3.1 Mécanisme catalytique

Les recombinases sont des protéines classées parmi les polynucléotidyltransférases qui agissent en transférant un nucléotide d'une chaîne polynucléotidique vers une autre. Dans ce type de réactions, un phosphate scissile peut être transféré à (i) une molécule d'eau menant à l'hydrolyse de la liaison phosphodiester, (ii) une chaîne latérale d'un acide aminé menant à une liaison covalente entre l'ADN et la protéine ou (iii) une deuxième chaîne polynucléotidique qui donne un ADN recombinant [79]. Le transfert du groupement phosphoryl est réalisé de la même manière que pour les autres classes de protéines phosphoryl transférases telles que les nucléases, phosphatases ou kinases [80, 81]. Toutes les réactions de phosphoryltransférases suivent quasiment le même mécanisme de substitution nucléophile SN₂ (Figure 17).

Contrairement aux autres protéines à activité phosphotransférase, les Y-SSR ne nécessitent pas d'ions métalliques pour la catalyse; au lieu de cela, le site catalytique est formé d'acides aminés hautement conservés, chargés positivement, dont les chaînes latérales qui constituent un ensemble de donneurs de liaisons hydrogènes qui jouent le rôle de stabilisateurs du phosphate scissile [59].

La formation d'un arrangement atomique contenant un oxygène en ligne avec le phosphate scissile pentacoordonné est l'état de transition typiquement retrouvé dans ce mécanisme (Figure 17, panneau central). Le départ de l'hydroxyle 5' (Figure 17, panneau de droite) est assuré ensuite par la protonation de celui-ci avec plusieurs candidats possibles, notamment deux Arginines ou deux Histidines du groupe d'acides aminés conservés R-K-H-R-H/W-Y tous portés par le domaine C-terminal de la protéine. La ligature en tant que réaction inverse de la réaction de coupure, est effectuée par l'hydroxyle de l'ADN, vraisemblablement tout seul, après la traversée de l'espace synaptique.

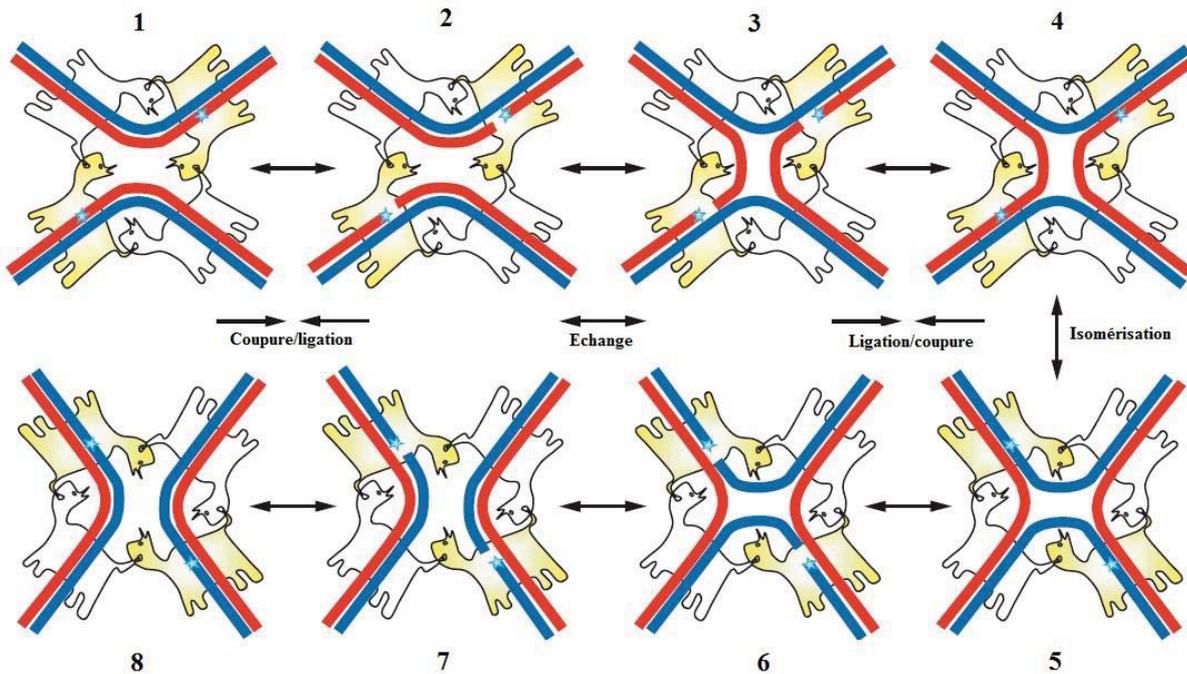


Figure 18 : Représentation schématique de la réaction de recombinaison catalysée par les Y-SSR [57]. La synapse est formée par deux duplexes d'ADN sur lesquels sont fixés quatre protomères de recombinases. Chaque protomère échange de manière cyclique une partie de sa région C-terminale. Les protomères actifs sont colorés en jaune et les protomères non actifs en blanc. Les étoiles représentent la position des phosphates scissiles. Une fois la synapse formée (étape 1), les deux protomères actifs clivent chacun un brin de l'ADN substrat et forment une liaison covalente 3'phospho-tyrosine avec l'ADN (étape 2) en libérant une extrémité 5'OH à l'extrémité des brins coupés. Les brins coupés sont alors échangés (étape 3) et le 5'OH attaque la liaison 3'phospho-tyrosine opposée (étape 4). La structure formée qui résulte de l'échange du premier brin est un intermédiaire formé de quatre branches d'ADN appelé « jonction de Holliday ». L'isomérisation de la synapse permet d'activer les deux protomères initialement inactifs et de désactiver les deux autres (étape 5). Les deux protomères activés réalisent une deuxième série de coupure/ligature (étapes 6 et 7) qui permet d'obtenir deux duplexes d'ADN recombinés (étape 8).

1.2.1.3.2 Formation de la synapse et schéma réactionnel

Les réactions de recombinaison sont réalisées en deux étapes de clivage et de ligature d'ADN (Figure 18). La réaction débute lorsque quatre monomères de recombinaison se fixent de manière spécifique chacun sur son demi-site induisant la formation d'un complexe synaptique [56]. Ce complexe synaptique présente une symétrie d'ordre 2 et une pseudo-symétrie P4. Cette asymétrie fait que seulement deux monomères seront actifs à un instant donné, de manière à ce que seulement un brin de chaque ADN est coupé, échangé et lié à la fois. Le clivage et la ligature se font à des positions spécifiques et identiques du côté 5' de la région centrale des deux ADN partenaires [59] et l'échange a lieu de manière à ce qu'aucune information génétique ne soit perdue ou gagnée et que les ADN produits n'aient aucune cassure [58]. Une fois la synapse formée (Figure 18, étape 1), les deux monomères actifs (protomères jaunes) catalysent le clivage de l'ADN par l'attaque de la chaîne latérale de la tyrosine catalytique sur le squelette phospho-diester de l'ADN (Figure 18, étape 2). Cette réaction de trans-estérification induit d'un côté la formation d'un complexe intermédiaire covalent 3'phospho-tyrosine entre la recombinaison et la molécule de l'ADN et de l'autre côté la libération d'un groupement 5'OH. Ce groupement hydroxyle libre en 5' peut ressouder la coupure et restaurer le substrat d'origine. Pour que la réaction de recombinaison ait lieu, les brins coupés doivent être échangés. L'échange de brin se fait par test d'homologie de la région centrale (Figure 18, étape 3). Le modèle actuellement admis est que 2-3 bases adjacentes au site de clivage soient échangées et suffisent pour que la réaction ait lieu [56, 82]. Une fois l'homologie assurée, l'hydroxyle 5' attaque la liaison 3'phospho-tyrosine du deuxième couple recombinaison/ADN induisant ainsi l'échange de brin (Figure 18, étape 4). La structure qui résulte de l'échange du premier brin est un intermédiaire appelé Jonction de Holliday (HJ) formé de quatre branches d'ADN et qui adopte une géométrie planaire facilitée par la pseudo-symétrie P4 [59]. L'isomérisation de cette synapse est déterminant pour l'issue de la réaction. En effet, l'isomérisation permet d'activer les deux molécules de recombinaison initialement inactives et de désactiver les deux autres (Figure 18, étape 5). Cette activation conduit à une deuxième série de coupure/ligature (Figure 18, étapes 6 et 7) qui permet la résolution de la jonction par clivage de la deuxième paire de brins d'ADN suivie par la libération des produits ADN recombinés (Figure 18, étape 8) [58]. Dans certains cas tel que chez les intégrases d'intégrons, la deuxième réaction de coupure n'a pas lieu et la jonction de Holliday n'est pas résolue par la deuxième paire de recombinaisons mais par des enzymes du processus cellulaire, probablement celles du complexe de réplication [83].

Pour les systèmes simples tel que Cre et Flp, le site formé de deux demi-sites et d'une région cross-over est suffisant à la recombinaison. Cependant, les sites de recombinaison peuvent être plus compliqués et contenir des sites auxiliaires nécessaires à la fixation d'autres molécules de recombinases ou d'autres cofacteurs qui vont jouer le rôle de régulateurs et permettre la structuration du site de recombinaison.

Dans la majorité des cas, les recombinases sont capables d'assurer à elles seules la recombinaison, mais l'utilisation de protéines accessoires semble être importante pour de nombreux systèmes. Ces protéines accessoires semblent jouer un rôle architectural qui permet l'assemblage du complexe par l'introduction de certaines contraintes au niveau de l'ADN. Ainsi, la résolution des dimères de ColE1 par XerC-XerD [84] nécessite un système plus complexe avec un site de recombinaison (appelé site *cer*) plus grand qui contient des sites accessoires sur près de 190 pb. Ces sites accessoires ont pour rôle la fixation des protéines arginine repressor (ArgR), leucine aminopeptidase (PepA) et *l'anaerobic growth regulator* (ArcA) sous sa forme phosphorylée pour pSC101 [85, 86]. Ces trois protéines sont nécessaires à la recombinaison qui a lieu entre les substrats super-enroulés contenant les sites *cer* répétés dans la même direction sur la même molécule. Ces exigences agissent comme un filtre qui permet uniquement la résolution et la production de monomères et non la réaction inverse.

La recombinaison dans le système XerC/D nécessite la présence du cofacteur FtsK qui possède une activité DNA translocase ATP-dépendante et qui reconnaît les sites *dif*. En l'absence de FtsK, XerC catalyse le clivage et l'échange de brins mais la jonction de Holliday qui en résulte ne peut aller plus loin sur le chemin réactionnel et sera résolue en substrats. Cependant en présence de FtsK, XerD est activé et peut alors catalyser la deuxième coupure [81]. La résolution de dimères de plasmides ColE1 (par son action sur les sites *cer*) et pSC1001 (par son action sur les sites *psi*) par le système XerC/D a été montrée en l'absence de FtsK [57]. Un exemple plus complexe est celui de λ -Int et sera développer dans le paragraphe 1.2.2.1.

1.2.1.4 Régulation de la réaction de recombinaison

Une question récurrente de l'étude des Y-SSR est celle de l'activation de la moitié du site uniquement, pour qu'ainsi, à un moment donné, seuls deux monomères soient actifs. Cette activation alternée des monomères permet d'éviter le clivage des deux brins au même moment. Ainsi, pour chaque ADN, le brin clivé doit être relié avant que son partenaire soit

coupé. Ceci permet d'éviter la formation de produits inintéressants pour la réaction de recombinaison tels que des linéarisations de chromosomes, des structures en tige boucle ou des jonctions à trois voies (*three-way junctions*) qui peuvent avoir lieu lorsque deux protomères adjacents sont activés en même temps [87-89]. L'activation différentielle de la moitié des sites est essentiellement définie par la géométrie de la synapse. La symétrie d'ordre 2 impose que deux protomères posés sur deux duplexes opposés adoptent des conformations similaires. La symétrie approximative quadruple facilite l'isomérisation et l'activation des autres monomères. La symétrie n'étant pas parfaite, deux types d'interfaces (I et II) séparant les protomères s'alternent dans la synapse. Cette différence au niveau des interfaces influe directement sur l'assemblage du site actif et régule ainsi l'activité des demi-sites. Cependant, la manière dont ces interfaces influencent la réaction varie d'un système à un autre.

Dans le cas de Flp, l'activation est contrôlée par la structuration de l'hélice échangée en *trans* à partir du protomère adjacent et qui porte la tyrosine catalytique. Cette hélice est formée uniquement dans les interfaces I alors que, dans les interfaces II, cette hélice est désordonnée et la tyrosine est éloignée du site catalytique [90].

Dans le cas de λ -Int, la régulation se fait aussi par le positionnement de la tyrosine. En effet, dans les protomères actifs, l'agencement du segment échangé en brin bêta permet de structurer l'hélice qui le précède et de positionner la tyrosine catalytique dans le site actif. Dans les unités non actives, la région échangée est déstabilisée, l'hélice est déstructurée et la tyrosine mal positionnée [67]. Dans le cas de Cre, la région charnière entre l'hélice portant la tyrosine et l'hélice qui la suit, qui est échangée en *trans*, adopte une conformation différente en fonction des protomères [57]. Dans au moins deux cas, la symétrie du système est brisée. Pour XerC/D, cette rupture est provoquée non pas par la conformation mais par la séquence des protomères qui sont différents. La différence est créée par la succession de deux protéines différentes qui fixent deux séquences différentes de chaque côté de la séquence centrale. Les deux protéines partenaires semblent échanger une partie de leur région C-terminale qui aurait un rôle dans la régulation de l'activité [91]. Dans le cas de l'intégrase d'intégrons IntI4 et de son substrat VCR_{bs}, le site attC est asymétrique, avec d'un côté un ADN double brin, au centre une bulle de 3 pb désappariées, et de l'autre côté une région contenant trois bases extra-hélicales G20'', T12'', et T10''. En effet, l'interaction de la protéine avec la base extra-hélicale T12'' fait déplacer la région C-terminale des protomères non actifs de près de 15°. Il en résulte le déplacement de l'hélice M et le repositionnement de la tyrosine loin du site actif (Figure 13, B) [66].

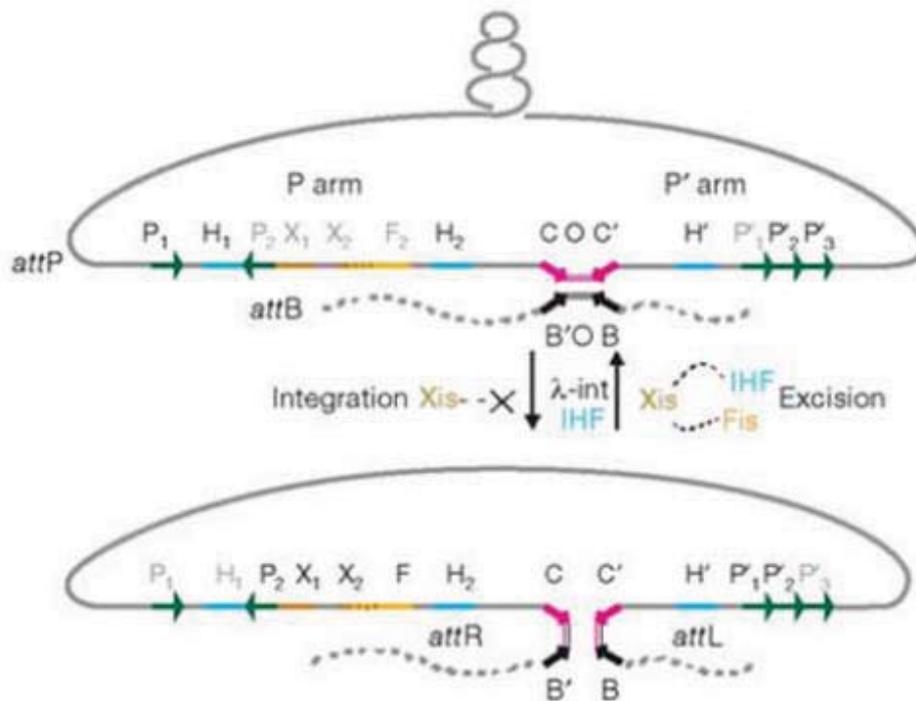


Figure 19 : Organisation des substrats utilisés par λ-Int. Les sites de recombinaison attP (« *attachment phage* ») et attB (« *attachment bacteria* ») sont représentés [67]. Le substrat attB est formé par le site de recombinaison BOB'. Le substrat attP a une taille de 240 pb et est formé par le site de recombinaison COC' où l'échange de brins a lieu, ainsi que les sites P1, H1, P2, X1, X2, F, H2 et H', P'1, P'2, P'3. Ces derniers constituent des sites de fixation du « *arm binding domain* » de l'intégrase et des protéines accessoires Xis (excisionase produite par la bactérie), Fis (« *Factor for inversion stimulation* » produit par le phage) et IHF (« *Integration host factor* »).

1.2.2 Homologies structurales et fonctionnelles potentiels de Pf-Int

Dans nos études de la recombinase à tyrosine de *P. falciparum* Pf-Int, que nous avons identifié, nous avons procédé à une étude de prédiction de la structure tertiaire de la protéine, ainsi qu'à la recherche de cibles potentielles ADN. Au vue des résultats de ces deux études développées en détail dans la suite du manuscrit, nous proposons l'hypothèse que Pf-Int ressemble à λ -Int de part son organisation structurale tertiaire et à l'intégrase des intégrons de part ses substrats ADN. Nous avons donc choisi de développer ces deux systèmes dans les paragraphes suivants de l'introduction.

1.2.2.1 L'intégrase du phage lambda

L'intégrase du phage lambda (λ -Int) est utilisée par ce dernier pour l'intégration et l'excision de son génome dans et à partir du génome de l'hôte en fonction des cycles lysogénique ou lytique par lesquels passe le phage [92]. En plus des deux domaines classiquement retrouvés chez les Y-SSR, λ -Int est formée d'un domaine auxiliaire supplémentaire présent du côté N-terminal de la protéine appelé « *arm binding domain* ». λ -Int catalyse l'échange ordonné par paire de quatre brins d'ADN entre deux substrats de recombinaison différents : attP (« *phage attachment site* ») et attB (« *bacterial attachment site* ») lors de l'intégration, attL et attR lors de l'excision [93, 94] (Figure 19). Le substrat bactérien attB est formé par le site de recombinaison BOB'. Le substrat attP a une taille de 240 pb et est formé par le site de recombinaison COC' où l'échange de brins a lieu, ainsi que les sites P1, H1, P2, X1, X2, F, H2 et H', P'1, P'2, P'3 (Figure 19). Ces sites constituent des sites de fixation du « *arm binding domain* » de l'intégrase et des protéines accessoires Xis (excisionase produite par la bactérie), Fis (« *Factor for inversion stimulation* » produit par le phage) et IHF (« *Integration host factor* » produit par la bactérie). Lors de l'intégration, les deux ADN attP et attB sont alignés et recombinaison par l'intégrase, aidée par le facteur IHF qui induit la courbure de l'ADN nécessaire à la formation du complexe d'intégration [67] [92]. La réaction d'intégration donne un prophage dont l'ADN est bordé par les sites attL et attR. Durant la phase lytique et en présence des facteurs Xis, Fis et IHF, le phage est excisé et les sites attP et attB sont régénérés [67]. L'interaction de l'ADN avec le domaine auxiliaire de λ -Int se fait *via* les sites P, celle de Xis *via* X, celle de Fis *via* F et celle d'IHF *via* H. Comme les deux réactions d'intégration et d'excision impliquent différents substrats et différents cofacteurs, elles ne sont pas réciproques contrairement aux réactions réalisées par les recombinases Flp et Cre [92].

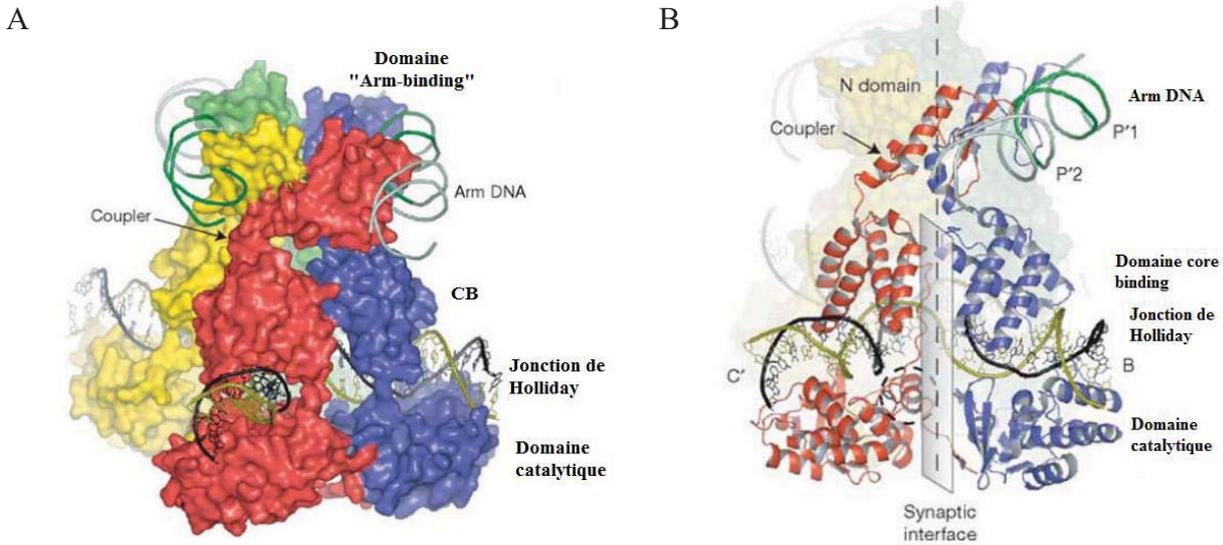


Figure 20 : Organisation structurale de λ -Int en complexe avec l'ADN substrat [67]. λ -Int est formée de trois domaines : le domaine « arm binding », le domaine « core binding » et le domaine catalytique. La figure montre l'organisation de la synapse formée par λ -Int en complexe avec l'ADN substrat. La synapse est formée par quatre protomères (rouge, bleu, jaune et vert), une jonction de Holliday et deux ADN « arm DNA ». Les trois domaines de l'intégrase sont organisés en trois couches superposées. La première couche est formée par les domaines « arm-binding » (en haut) de chaque protomère qui sont échangés de manière cyclique avec le protomère adjacent et qui fixent les ADN Arm-sites. Les deuxième et troisième couches sont formées par les domaines « core binding » et le domaine catalytique fixés à la jonction de Holliday. **A** : figure montrant l'enveloppe globale des protéines. **B** : figure montrant les structures secondaires des protéines et l'interface entre deux protomères.

Les domaines auxiliaires « arm binding domains » de l'intégrase jouent un rôle crucial dans la recombinaison. En effet la recombinaison ne peut pas avoir lieu en leur absence et leur interaction avec les sites P favorise le contact du cœur de la protéine avec les séquences COC' et BOB' [92]. L'implication des domaines auxiliaires, des séquences ADN accessoires et des facteurs auxiliaires (IHF, Xis, Fis) permet à λ -Int d'avoir un certain niveau de régulation qui n'est pas possible avec les autres Y-SSR plus simples [67, 94].

L'assemblage tétramérique, formé de quatre monomères d'intégrase entiers et de deux ADN, est l'un des plus complexes dont la structure a été observée pour cette famille de protéines (Figure 20) [67]. L'intérêt de ces structures réside dans l'organisation des domaines les uns par rapport aux autres. En effet, les structures montrent que le système est formé par un premier assemblage, composé d'un tétramère de domaines C-terminaux catalytiques et N-terminaux liés à l'ADN, dont l'arrangement ressemble à celui de Cre/LOX. Dans un deuxième niveau de complexité, ce tétramère est lié à un autre tétramère formé par les domaines auxiliaires qui eux-mêmes sont fixés à deux duplexes de séquences ADN *arm-sites*. Il est également intéressant de noter que chaque domaine auxiliaire est échangé d'une manière cyclique et vient se tasser sur le monomère adjacent [67].

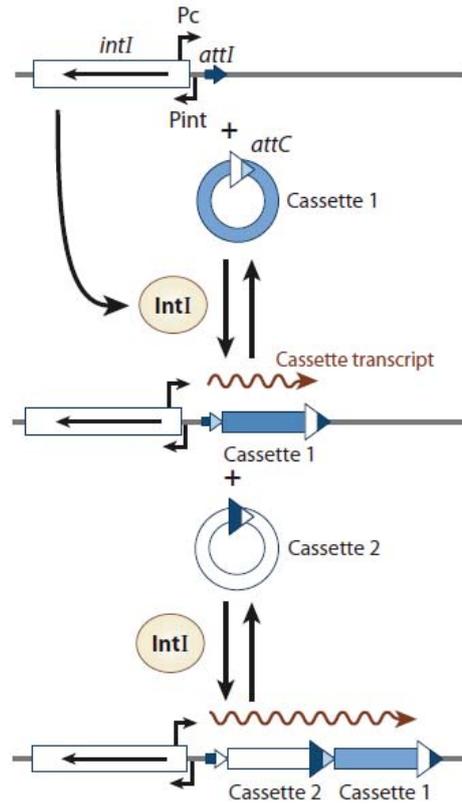
L'identification des cibles de Pf-Int homologues aux sites cœurs et aux sites auxiliaires de λ -int, n'a pas pu être faite pour diverses raisons, dont la principale est la richesse en A/T des séquences qui seront décrites plus bas. Cette difficulté, peut cependant refléter un autre choix d'organisation des sites. Un cas où les ADNs choisis sont apparemment aléatoires sauf pour un cœur GTTRRRY, est le cas des intégrases d'intégrons, qui utilisent une base structurelle plutôt que l'identité de séquence pour la sélection des sites de recombinaisons.

1.2.2.2 Les intégrases d'intégrons

Les études de recherche des cibles ADN de la recombinase à tyrosine de *P. falciparum* PF-Int et que nous développerons dans la partie résultats, suggèrent que les substrats de cette dernière soient formés d'ADN simple brin replié. Nous développerons donc dans le paragraphe suivant le système d'intégrase d'intégrons qui a recours à ce type de substrat.

Les intégrons sont des éléments génétiques capables d'acquérir et de réarranger des cadres de lectures ouverts (« Open Reading Frame », ORF) organisés en cassettes de gènes et de les convertir en gènes fonctionnels permettant ainsi leur expression [95]. Initialement identifiés comme étant un mécanisme utilisé par les bactéries Gram(-) pour l'acquisition de gènes de résistance, leur rôle a été élargi avec la découverte des intégrons chromosomiques présents

A



B

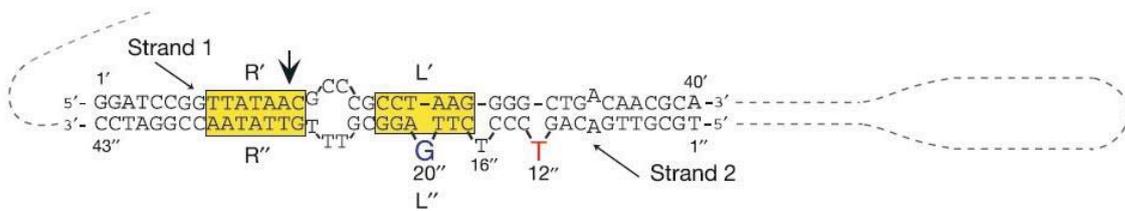


Figure 21 : A. Représentation schématique de l'organisation des intégrons [83]. Chaque intégron est formé d'un gène qui code pour l'intégrase (*intI*), un promoteur qui permet la transcription du gène *intI* (*Pint*), un promoteur qui permet la transcription des cassettes (*Pc*), un site d'insertion des cassettes (*attI*), un ensemble de cassettes et un site de recombinaison présent à l'extrémité de chaque cassette (*attC*). **B. Repliement d'un site *attC* [66].** Les sites *attC* sont constitués des régions R''-L'' et L'-R' séparées par une région centrale variable en taille et en séquence. Le repliement du site *attC* produit un site de recombinaison composé de deux duplexes L''-L' et R''-R' (colorés en jaune) séparés d'une région centrale en bulle et de trois bases extra-hélicales (G20'', T16'' et T12'') nécessaire à la fixation de la protéine. Le site de coupure est indiqué par une flèche noire.

dans des génomes de centaines d'espèces bactériennes. Ces structures appelés intégrons sédentaires ont une fonction autre que l'acquisition de phénotype de résistance [96]. Les intégrons sont formés de deux composantes majeures (Figure 21, A). La première est constituée d'éléments fonctionnels comprenant le gène codant pour l'intégrase et un promoteur (Pc) permettant la transcription des cassettes qui sont dépourvues de promoteurs. L'intégrase, qui est une Y-SSR, catalyse l'excision et l'intégration d'éléments génétiques (cassettes de gènes) entre deux sites de recombinaison différents *attI* et *attC* [97]. L'intégration des cassettes a lieu au site *attI*, adjacent au gène de l'intégrase [98]. La deuxième composante des intégrons est la collection de cassettes composées chacune d'un seul ORF suivi d'un site de recombinaison appelé *attC* [96].

L'analyse des différences entre les sites *attC* chez les intégrons mobiles montre que ces derniers sont très hétérogènes aussi bien du point de vue de leurs séquences que de leurs tailles. En revanche les cassettes sont homogènes en fonctionnalité et codent presque exclusivement pour des gènes de résistance aux antibiotiques [83, 99]. Les intégrases d'intégrons possèdent un système de recombinaison original qui implique l'utilisation d'un ADN double brin (*attI*) et d'un ADN simple brin replié en tige boucle provenant du brin bas du site *attC* [100]. Le site d'intégration *attI* est formé de deux demi-sites (L et R) de fixation des intégrases, le site L étant toujours dégénéré par rapport au site R. La région centrale est également différente entre les sites *attI*. Le point de recombinaison est localisé entre les bases G et TT dans le triplet conservé 5'-GTT-3' [101]. Pour les sites *attC*, chaque extrémité d'un site est formée par des séquences appelées R''-L'' et L'-R' séparées par une région centrale variable en taille et en séquence [83]. Il a été également montré que *in vivo* le brin du bas du site *attC* est 10^3 plus recombinant que le brin du haut et que les mutations qui déstabilisent la formation de la structure cruciforme abolissent la recombinaison [100]. Malgré la différence de séquence élevée qui existe entre les sites *attC*, ces derniers se replient de la même manière pour former des structures cruciformes (Figure 21, B) plus complexe que les sites *attI*. Une fois l'ADN simple brin replié (Figure 21, B), il forme un site de recombinaison constitué de deux duplexes L''-L'' et R''-R'' séparés par une région centrale en bulle et de trois bases extra-hélicales (G, T et T) qui imposent le choix du brin transféré en positionnant correctement les recombinases. La comparaison des sites *attC* montre que la conservation de séquence est limitée à deux triplets 5'-AAC-3' et 5'-GTT-3' inclus respectivement dans les régions R'' et R'. La recombinaison a lieu entre le G et TT portés par le duplexe R''/R''. Même si les triplets GTT et AAC sont entièrement conservés, il a été montré que seul le changement à la troisième

position du triplet GTT a un effet drastique sur la recombinaison [102]. De ce fait, la recombinaison dépend énormément de la reconnaissance de la structure de l'ADN indépendamment de la séquence. L'étude de *Bouvier et al* [103] montre que la spécificité est assurée par les bases extra-hélicale, d'une part G20" qui se retrouve logée dans une poche hydrophobe, d'autre part T12" stabilisé par deux histidines et une proline d'un feuillet β de la région supplémentaire caractéristique des IntI. Ceci permet l'identification du brin à recombinaison. Cette étude montre aussi que la recombinaison dépend de la structure de la région ADN non appariée centrale [66] [103]. La structure du complexe VChInt- attC obtenue par *MacDonald et al* [66] confirme qu'effectivement la majorité des contacts entre la protéine et l'ADN est de nature non spécifique et est formée par des interactions protéine-phosphate du squelette de l'ADN.

Il est à noter que la recombinaison peut se faire soit entre deux sites *attC* localisés sur le même ADN et engendrer l'excision d'une cassette circulaire, soit entre un site *attC* porté par la cassette circulaire et un site *attI*. Durant cette dernière, il y a formation d'un complexe composé de deux molécules de recombinases fixées sur l'ADN *attI* et de deux autres molécules fixées sur l'ADN *attC* replié. Un premier échange de brin a lieu suite à l'attaque de deux de ces molécules seulement, ce qui engendre la formation d'une jonction de Holliday qui sera résolue par une étape de réplication [95].

1.2.3 La variabilité antigénique générée par des recombinases à site spécifique

Chez certains micro-organismes, les réactions de recombinaison spécialisées impliquant l'inversion d'ADN ont évolué pour donner des mécanismes de génération de diversité génétique dans une population donnée. Ces réactions de réarrangements permettent de pré-adapter une partie de la population à un changement de l'environnement.

1.2.3.1 Exemple de réarrangements d'ADN uniques ou simples

Dans les cas les plus simples, le réarrangement de l'ADN consiste en une inversion d'un seul segment d'ADN qui génère un ou deux phénotypes. Les exemples incluent la variation de phase du flagelle chez *Salmonella enterica*, la variation de la piline chez *E. coli* ou *Moraxella*. Ainsi chez *Salmonella*, l'antigène variable H peut être exprimé à partir de deux *loci* différents H1 (appelé aussi *fliC*) et H2 (*fliB*) localisés dans des régions distinctes du génome. H2 contient un segment d'ADN de 996 pb qui peut subir une inversion réversible. Ce segment est bordé de deux séquences de recombinaison appelées *hixL* et *hixR* (26 pb) qui sont reconnus

spécifiquement par la recombinase à sérine HIN. L'inversion du segment induit le changement d'orientation du promoteur de *fljB* et *fljA* (répresseur de *fliC*). Lorsque *fljA* n'est pas transcrit, *fliC* est actif et H1 est synthétisé.

Chez *E. coli*, un système de switch « on-off » de la piline de type I est contrôlé par l'inversion spécifique d'un segment d'ADN (314 pb) qui contient le promoteur du gène *fimA* (protéine de la piline type I). Cette inversion est assurée par deux recombinases de type tyrosine FimB et FimE qui catalysent la recombinaison aux sites IRL et IRR de 9 pb localisés à l'extrémité du segment inversé [104, 105].

1.2.3.2 Exemple de cas de réarrangements multiples de segments d'ADN

Borrelia hermsii possède un système formé de 59 gènes *vmp* dont un seul est exprimé à la fois grâce à sa localisation au niveau du site d'expression. Ces gènes codent pour des lipoprotéines exprimées à la surface de cette bactérie, divisées en deux groupes selon leurs tailles (« *Variable large proteins* » : 37 à 40 kDa et « *Variable small proteins* » : 20 à 23 kDa), ou en quatre familles selon leurs séquences [108]. Il a été suggéré que la protéine ResT récemment identifiée pourrait jouer un rôle dans l'élasticité du génome qui caractérise cette bactérie [109]. ResT est une protéine qui semble avoir une activité « *hairpin telomere resolvase* » reliée d'une part aux transposases par le fait de posséder un motif qui lui permet de se fixer aux structures ADN tige-boucle et aux Y-SSR et topoisomérases IB par le fait qu'elle possède les résidus catalytiques conservés par ces familles [110, 111].

Plusieurs plasmides basés sur le système *shufflon* existent et utilisent des mécanismes d'inversions multiples. Par exemple, le plasmide R64 *shufflon* possède quatre segments d'ADN (A, B, C et D) inversibles, séparés et bordés par sept séquences *sfx* (19 pb). Ces segments peuvent être remaniés par une recombinase à tyrosine codée par le gène *rci*. Le remaniement de ces segments fait varier la région C-terminale de la protéine PilV du pili et permet ainsi de produire sept variantes [112, 113].

1.3 Délimitation du sujet

L'identification de nouvelles approches thérapeutiques antipaludiques basées sur l'inhibition spécifique de certaines des cibles moléculaires vitales pour ce parasite est devenue une nécessité dans la lutte contre ce fléau. En effet, en plus de sa résistance aux traitements actuellement disponibles, *Plasmodium* reste jusqu'à aujourd'hui réfractaire aux vaccinations. Grâce à un processus de variabilité de ses antigènes présents à la surface des érythrocytes qu'il infecte, ce parasite est capable d'échapper au système immunitaire de l'hôte et d'établir

une infection chronique. Différentes études suggèrent que les modifications épigénétiques constituent le facteur majeur impliqué dans la transcription mutuellement exclusive entre les différents gènes membres de la même famille multigénique. Cependant, ce mécanisme reste mal décrit et l'implication d'autres facteurs tel que des ARN non codant ou des ARNantisens n'a pas totalement été écartée. En plus de la transcription mutuellement exclusive, il a été montré que le parasite est capable de diversifier continuellement les gènes de sa librairie.

L'identification d'un nouveau mécanisme basé sur l'utilisation d'une recombinase à site spécifique pourrait être d'un intérêt capital pour la compréhension de la biologie de *Plasmodium* et nous proposons que la recombinase à tyrosine Pf-Int récemment identifiée dans notre laboratoire soit impliquée dans le mécanisme de variabilité antigénique. Cet enzyme peut être impliqué soit dans le mécanisme de régulation de l'expression des gènes soit dans le mécanisme de génération de la diversité.

Dans le but d'étayer cette hypothèse, nous avons procédé à la réalisation de différentes expérimentations ayant comme objectifs la définition de l'appartenance de la recombinase à la famille des Y-SSR, sa caractérisation biophysique comme une protéine globulaire stable, avec des indices de ses structures secondaire et tertiaire en solution, l'identification de cibles ADN potentiels, la détermination des interactions non covalentes avec celles-ci, la démonstration de la formation d'un complexe covalent et un suivi de son expression (ARNm, et protéique) *in vivo* pour mieux appréhender son rôle physiologique, et l'interférence avec celui-ci à travers la sur-expression de la protéine, et la génération d'un mutant K.O..

Pf-Int est une protéine de 490 aa codée par le gène MAL13P1.42 (locus Pf3D7_13 allant de la position 403,991 à la position 406,906). Cet enzyme a été identifié comme étant une hypothétique recombinase à tyrosine suite au projet de séquençage du génome entier de *P. falciparum*. En effet, les analyses automatisées de recherche de domaines, réalisées par le programme InterPro et accessibles dans la base de données PlasmoDB, ont permis d'identifier différents domaines dont deux qui sont homologues aux domaines N-terminal de λ -int (aa 192-300) et C-terminal de la famille des λ -int (aa 296-486). Le séquençage des génomes des autres espèces de *Plasmodium* a montré que ces deux domaines sont particulièrement conservés chez toutes les autres espèces. Même si les séquences en acides aminés des recombinases à tyrosine connues tel que Cre, Flp, λ -int et IntI sont divergentes, ces dernières ont des repliements tertiaires semblables. Les études préliminaires par prédiction de la structure tertiaire de Pf-Int ont permis de montrer qu'elle pourrait adopter un repliement semblable à celui des autres recombinases à tyrosine, en particulier celui de λ -int. Ces études de prédiction ainsi que les alignements de séquences ont permis de montrer que les résidus

catalytiques conservés chez tous les membres de cette famille et qui sont impliqués dans la réaction de recombinaison sont également présents chez Pf-Int. Pf-Int fait donc partie des rares recombinases à sites spécifiques identifiées chez les eucaryotes. Néanmoins, aucune étude n'avait été consacrée auparavant à cette protéine et le rôle ainsi que le mécanisme moléculaire qu'elle utilise restent inconnus.

Les expérimentations que nous avons réalisées lors de ces travaux de thèse se sont articulées en trois axes majeurs : l'identification des cibles potentielles, la caractérisation fonctionnelle et la caractérisation biochimique de la protéine.

Concernant la caractérisation biochimique, nous avons procédé à la production et purification de différentes régions de l'intégrase. Les différentes protéines ont été produites dans le système *E. coli* et les protéines recombinantes purifiées ont été caractérisées par des analyses biophysiques. Ces analyses ont concerné la détermination du contenu en structures secondaires par dichroïsme circulaire (CD), l'étude de la dénaturation thermique couplée au CD, l'analyse de l'état d'oligomérisation par centrifugation analytique et l'étude du repliement par RMN (résonance magnétique nucléaire) spectre 1-D. Nous avons également tenté, sans succès, de caractériser la structure tertiaire par des essais de cristallogénèse.

Le deuxième volet de la caractérisation biochimique a concerné l'étude de l'interaction des différentes protéines recombinantes purifiées avec des cibles potentielles ADN. Pour cela, nous nous sommes basés sur des résultats précédemment obtenus au laboratoire qui sont constitués de deux bibliothèques de cibles potentielles identifiées *in vitro* soit par SELEX soit par affinité à partir d'ADN génomique en utilisant comme appât la région C-162 de la protéine (allant de l'aa 162 à l'aa 490). Les analyses d'interactions entre les différents fragments purifiés de l'intégrase et les ADN cibles précédemment identifiés ont été faites par retard sur gel natif (EMSA, *Electrophoretic Mobility Shift Assay*). Nous avons entrepris de caractériser les interactions observées entre certains fragments et les ADN cibles ainsi que de démontrer que les protéines purifiées sont actives. Par ailleurs nous avons aussi étudié la spécificité des interactions observés.

Concernant la caractérisation fonctionnelle, nous avons essayé dans un premier temps de déterminer la localisation sub-cellulaire de Pf-Int au moyen d'anticorps anti-Pf-Int préparés à partir de la protéine recombinante purifiée (C-162) ou en utilisant différentes constructions génétiques dans le but de sur-exprimer l'enzyme dans le parasite. En effet, le parasite *Plasmodium falciparum* possède trois génomes avec lesquels pourrait interagir Pf-Int différents, ces génomes sont compartimentés dans trois organites différents (noyau,

mitochondrie et apicoplast), et déterminer la localisation sub-cellulaire de l'intégrase contribue à l'identification de ses cibles. Nous avons également essayé de déterminer l'impact de trois codons d'initiation possible sur l'adressage de la protéine. Dans un deuxième temps, nous avons essayé de déterminer l'impact de l'inactivation du gène codant pour Pf-Int sur le parasite en particulier la durée du cycle cellulaire et étude de la parasitémie. En effet, nous disposons d'une souche de parasite dont le gène codant Pf-int a été invalidé (souche 3D7-KO-Pf-Int créée par Christine Benatar, BIHP, Institut Pasteur). Les études futures de ce K.O. seront utiles pour pouvoir avancer dans la compréhension du rôle physiologique de Pf-Int.

2 MATERIELS ET METHODES

2.1 Analyses bio-informatiques

La prédiction de la structure secondaire de Pf-Int a été réalisée avec le programme PSIPRED V3.0 [114, 115] et le programme Discovery Studio (Accelrys). La recherche d'homologues structuraux de Pf-Int a été réalisée par les programmes PHYRE en mode « Normale » et « Intensive » [116], le programme GenTHREADER [117] et le programme pGenTHREADER [118]. La construction du modèle de Pf-Int selon la structure publiée de l'intégrase du phage lambda (PDB id : 1Z1B) a été réalisée grâce au programme Discovery Studio (Accelrys).

Les alignements des séquences de protéines et d'ADN ont été réalisés grâce au programme Clustal en mode automatique [119]. La recherche de motifs répétés dans les séquences d'ADN génomique a été réalisé avec les programmes IRF [120], MEME [121] et Einverted (EMBOSS) [122]. La prédiction des structures secondaires des cibles ADN génomiques identifiées a été réalisée au moyen du programme mFOLD à 37°C, 150 mM NaCl, 0 mM MgCl₂ et avec des valeurs différentes pour la variable « window parameter » [123]. La recherche de structures secondaires de l'ADN conservées a été réalisée par le programme LocARNA selon les modes « global » et « local » [124].

2.2 Etude de la transcription du gène MAL13P1.42

L'extraction de l'ARN a été réalisée à partir de cultures synchronisées de parasites. Les globules rouges infectés ont été récupérés par centrifugation à 1700 g pendant 5 minutes et le culot a été resuspendu dans 1 ml de PBS. Les cellules ont été centrifugées une deuxième fois et le culot resuspendu dans 1ml PBS contenant 0,05% saponine. Les parasites libérés ont été lavés avec 1ml PBS et conservés dans le trizol (Invitrogen) à -80°C. L'ADN a ensuite été éliminé par le kit DNA-free kit (Ambio, Applied Biosystems). L'ARN récupéré a été utilisé comme substrat pour la réaction RT-PCR en utilisant la transcriptase reverse SuperscriptII (Invitrogen) et des amorces oligo-dT (Invitrogen). La réaction de PCR en temps réel a été réalisée dans un volume de 20 µl comprenant 200 nM d'amorces (Pf-Int_B1 et Pf-Int_F1, Tableau 1) dans le mélange SYBR Green PCR master mix (Applied Biosystems) et en utilisant le thermocycler Realplex4 Eppgradients. Le gène seryl-tRNA synthetase (PF07_0073) a été utilisé comme contrôle interne. Les mesures ont été réalisées à la fin de l'étape d'élongation de chaque cycle et le niveau de transcription a été corrigé en fonction du taux d'expression du gène de ménage seryl-tRNA synthetase. L'expérimentation a été répétée à partir de deux préparations d'ARN.

2.3 Clonage

Les constructions génétiques ont été réalisées dans le but d'exprimer différentes versions de l'intégrase soit dans la bactérie *Escherichia coli* (*E. coli*) afin d'obtenir les quantités nécessaires pour réaliser les analyses biochimiques et biophysiques, soit dans *P. falciparum* afin de déterminer son rôle physiologique.

2.3.1 Constructions génétiques dans le but d'exprimer les protéines recombinantes dans *E. coli*

Concernant la surexpression de l'intégrase ou certains de ses domaines dans la bactérie *E. coli*, nous avons utilisé le plasmide pET43.1a(+) (Novagen) comme squelette de base dans lequel la région NusTag (834-2318) a été enlevée et remplacée par une séquence qui génère une nouvelle protéine de fusion avec une étiquette 6-histidines placée en N-terminal des différentes versions de Pf-Int. Un site de clivage à la protéase du TEV a également été inséré entre l'étiquette 6-histidines et la protéine d'intérêt. Pour les définitions suivantes, le codon d'initiation choisi comme référence est celui qui est employé par PlasmoDB dans son annotation actuelle. Ce codon d'initiation correspond à la position 406,906 du chromosome 13 et la traduction à partir de ce dernier permet d'obtenir un polypeptide de 490 acides aminés. Nous avons cloné cinq fragments du gène en utilisant les amorces présentées dans le tableau 1. La nomenclature utilisée pour la suite est en fonction de l'acide aminé (aa) de début et de fin (aa α - Ω) de la région de la protéine pour laquelle code le fragment cloné. Ainsi, ces cinq constructions permettent d'exprimer les régions suivantes de l'intégrase : aa 114-490, aa 162-490, aa 192-490, aa 303-490 et aa 114-180 appelées respectivement C114, C162, C192, C303 et V114-180. Il est à noter que l'expression des protéines est sous le contrôle du promoteur T7 et que le plasmide porte une cassette de résistance à l'ampicilline.

2.3.2 Constructions génétiques dans le but d'exprimer Pf-Int chez le parasite *P. falciparum*

Concernant la surexpression de la protéine dans le parasite, nous avons eu recours au plasmide pLN [125]. Ce plasmide porte deux cassettes de résistance à l'ampicilline et à la blasticidine et l'expression des protéines est sous le contrôle des éléments régulateurs 5'UTR du gène de la calmoduline et 3'UTR du gène hsp86. Quatre constructions ont été réalisées. La première version correspond à une fusion entre la région N-terminale de l'intégrase aa 1-190 avec la GFP en utilisant les amorces présentées dans le tableau 1. Trois autres versions du

gène ont été clonées en utilisant les amorces pLN-(+1)-AvrII, pLN-(-6)-AvrII, pLN-(+12)-AvrII et pLN-rev-XhoI. Ces constructions pLN 2, 3 et 4 correspondent à des protéines de fusion entre différentes versions de l'intégrase et un Flag-tag en C-terminal. Ces trois versions correspondent à trois ATG possibles dans la région du début du cadre ouvert de lecture de PlasmoDB. Elles reflètent l'incertitude qui existe quant au début de la traduction *in vivo* des mRNA de Pf-Int. Ainsi, l'ATG de la version du gène publiée dans plasmoDB est appelée (+1), celle de l'ATG conservé entre les différentes espèces de *Plasmodium* est appelée (+12), et enfin la troisième qui correspond à un ATG présent en amont de celui rapporté dans PlasmoDB est appelée (-6).

L'amplification des séquences codantes a été faite à partir du cDNA de *P. falciparum* (obtenu gracieusement de Christine Schedig-Benatar, du laboratoire BIHP, A. Scherf, Institut Pasteur) en utilisant le kit Expand High Fidelity PCR System (Roche). Les réactions PCR ont contenu 0,4 mM dNTP, 0,4 μ M de chaque amorce, 10 ng de cDNA, 1,5 mM MgCl₂, 2,5 U d'enzyme. Les amplifications ont été réalisées dans le thermocycleur Master cycler gradient (Eppendorf) en utilisant la programmation suivante : une étape de dénaturation à 94°C/2 minutes suivie de 25 cycles d'amplification (94 °C/15 secondes, hybridation pendant 30 secondes et 68 °C/2 minutes) suivi d'une élongation finale de 10 minutes à 72 °C.

Les produits d'amplification ont été digérés par les enzymes de restriction NcoI/BamHI pour le clonage dans pET et XhoI/AvrII (Fermentas) pour le clonage dans le pLN pendant une heure à 37 °C. Les produits de digestion ont été ensuite purifiés par le kit QIAquick PCR Purification Kit (Qiagen) selon les recommandations du fabricant. L'ADN élué a été dosé par spectrophotomètre nano-drop (Thermo Scientific).

Les plasmides ont été digérés par les enzymes de restriction correspondantes et l'ADN digéré a ensuite été chargé sur un gel d'agarose 0.5%-TBE(0.5X)/120 V en présence du marqueur de masses moléculaires (SmartLadder, Eurogentec). Après coloration au bromure d'éthidium et observation aux UV (Kodak EDAS 290), la bande correspondant au plasmide linéarisé a été coupée et l'ADN extrait en utilisant le kit QIAquick Gel Extraction Kit, (Qiagen). Le plasmide linéarisé a été ensuite déphosphorylé par traitement à l'enzyme Shrimp alkaline phosphatase (USB). 1 μ g de plasmide a été mélangé avec 1U d'enzyme dans un volume total de 20 μ l et le mélange a été incubé à 37 °C. La phosphatase a été désactivée au bout de 1h30 par chauffage à 65 °C pendant une heure. La ligation du plasmide et de l'insert a été réalisée par mélange de 0,1 pmol d'ADN plasmidique et 0,2 pmol d'insert avec 1,5 U de ligase T4 DNA Ligase

(BioLabs) dans un volume total de 20 μ l. Le mélange a été incubé à température ambiante pendant 10 heures et a ensuite été utilisé pour la transformation dans les bactéries *E. coli* DH5 α (Invitrogen) par choc thermique selon les recommandations du fournisseur.

10 colonies bactériennes issues de chaque ligature ont été repiquées chacune dans 5 ml de milieu LB additionné d'ampicilline 100 μ g/ml et incubées à 37 °C avec agitation pendant 12 heures. Ces cultures sont utilisées pour la préparation d'ADN plasmidique par le kit QIAprep Spin Miniprep Kit, Qiagen. Une vérification de la présence des fragments ADN d'intérêt est réalisée par digestion aux mêmes enzymes de restrictions utilisées pour l'étape de clonage. Les constructions positives sont soumises à la vérification de la séquence insérée par séquençage commercial (Beckman Coulter Genomics) à partir des sites T7prom et T7Term préexistantes sur le pET43.1a(+) et les sites pLN-HSP86 et pLN-CALMOD sur le plasmide pLN.

2.4 Expression et purification de protéines recombinantes dans *E. coli*

2.4.1 Surexpression des protéines recombinantes

La surexpression des protéines recombinantes a été faite dans *E. coli*, BL21-CodonPlus RIPL (Stratagene). Pour cela, 10 ng d'ADN plasmidique ont été mélangés avec 50 μ l de cellules compétentes. Le mélange est incubé pendant 30 minutes sur glace, soumis à un choc thermique de 42 °C pendant 20 secondes et incubé une deuxième fois sur la glace pendant 3 minutes. 300 μ l de milieu de culture liquide LB préchauffé à 42°C ont été ensuite rajoutés et le produit de la transformation a été incubé pendant une heure à 37 °C. Différents volumes (50 à 200 μ l) ont été étalés sur des boîtes de Pétri contenant du milieu de culture solide LB-agar supplémenté d'ampicilline 100 μ g/ml et de chloramphénicol 25 μ g/ml. Les boîtes ont été ensuite incubées pendant la nuit à 37 °C.

Une préculture a été préparée en repiquant une colonie dans 80 ml de milieu de culture LB liquide additionné des antibiotiques ampicilline à 100 μ g/ml et chloramphénicol à 25 μ g/ml, et incubé pendant la nuit à 37 °C à 210 tpm. Le lendemain, cette préculture a été centrifugée pendant 10 minutes à 900 x g et le culot a été resuspendu dans 20 ml de LB frais. Les bactéries lavées ont servi à ensemercer une culture de quatre litres de LB additionné des antibiotiques ampicilline et chloramphénicol aux mêmes concentrations décrites plus haut. La culture a été réalisée à 37 °C sous une agitation de 210 tpm. Une fois la valeur 0,7 d'absorbance à 600 nm (DO_{600}) atteinte, la surexpression de la protéine est induite par ajout

de 1 mM d'IPTG pendant 5 heures. A la fin de l'induction, les cellules sont collectées par centrifugation à 4 400 x g pendant 15 minutes et les culots sont resuspendus dans le tampon A (50 mM Tris-HCl pH 8, 1 M NaCl, 5 mM Imidazole, 10% glycérol, 0.1% triton X-100) et congelés à -80°C en aliquotes correspondant aux culots récupérés de 1 ou de 2 litres de culture.

2.4.2 Purification des protéines recombinantes

Les culots cellulaires ont été décongelés sous un filet d'eau froide. 17 µg/ml de PMSF et 0,25 µg/ml de leupeptine ont été ajoutés à la suspension et les cellules sont ensuite lysées par trois passages successifs à travers la cellule d'un Emulsiflex (Emulsiflex C-5, Avestin Europe GmbH) à 15000 psi à 4°C. Le lysat cellulaire est clarifié par centrifugation à 14000 tpm/1h/4 °C dans une centrifugeuse Sorvall RC5 en utilisant le rotor ss-34 et le surnageant est ensuite filtré à travers une membrane de seuil de coupure 0,45 µm (Minisart, Sartorius).

La protéine a été purifiée par chromatographie en phase liquide à l'aide d'un système AKTAprime plus (General Electric, Healthsciences) à 5 °C et en cinq étapes successives : une première colonne d'affinité aux ions Nickel (HisTrap Ni-NTA), une colonne d'affinité aux phosphates mimés par héparine, le clivage du tag, un deuxième passage sur la colonne d'affinité Ni-NTA et une dernière colonne de chromatographie d'exclusion stérique. Cinq tampons ont été utilisés pour la purification : Tampon A (50 mM Tris-HCl pH 8, 1 M NaCl, 5 mM Imidazole, 10% glycérol, 0.1% triton X-100); Tampon B : tampon A/250 mM imidazole/500 mM NaCl; Tampon C : 50 mM Tris-HCl pH 8, 0,5 M NaCl, 10% glycérol; Tampon D : tampon C/1 M NaCl. Tampon E (50 mM Tris-HCl pH 8, 0,5 M NaCl, 5% glycérol).

La colonne HisTrap HP 5 ml (Amersham biosciences, GE lifesciences) a été équilibrée avec le tampon A et l'échantillon a été chargé à un débit de 2 ml/min. La colonne a été ensuite lavée avec le tampon A jusqu'au retour de l'absorbance à 280 nm à la ligne de base de départ. Les protéines retenues sur la colonne ont été ensuite éluées par l'application d'un gradient de tampon B (tampon A, 500 mM NaCl, 250 mM Imidazole). Des fractions de 4 ml ont été prélevées tout au long de l'étape d'éluion. La présence d'une bande correspondant à la masse moléculaire attendue de la protéine sur gel de polyacrylamide-SDS (SDS-PAGE) a été utilisée pour suivre la progression de la purification.

Les fractions issues de l'étape précédente contenant la protéine d'intérêt ont été rassemblées et chargées à 2 ml/min sur une colonne HiTrap Heparin HP 5 ml (Amersham biosciences, GE lifesciences) pré-équilibrée avec le tampon C. La protéine a été ensuite éluée en augmentant la concentration en NaCl par application d'un gradient du tampon D.

Les fractions contenant la protéine d'intérêt ont été rassemblées et le tag polyhistidines a été coupé par incubation de la protéine avec la protéase du virus de la gravure du tabac (*Tobacco etch virus* ou TEV) à un ratio par masse de 1:10 de TEV:Pf-int en dialyse à 4°C contre un litre de tampon A pendant 15 heures en utilisant la membrane de dialyse Spectra/Por7, au seuil de coupure de 10 000 Da. La protéine dont le tag a été coupé a été séparée du reste du mélange par un deuxième passage sur la colonne HisTrap HP 5 ml. Les fractions contenant la protéine coupée ont été rassemblées et mises à dialyser contre le tampon E à 4°C/15 heures avec le changement du tampon au bout de deux heures.

La protéine dialysée a été ensuite passée sur une colonne de chromatographie d'exclusion stérique G75 Sephadex (Amersham biosciences, GE Lifesciences) pré-équilibrée avec le tampon E à un débit de 0,5 ml/minute.

Les fractions issues des différentes étapes d'éluion ont été contrôlées par SDS-PAGE. 10 µl de chaque fraction ont été mélangés avec 10 µl du tampon de *Laemmli* (2%-SDS). Le mélange a été dénaturé par chauffage à 95 °C pendant 10 minutes et chargé ensuite sur un gel de polyacrylamide Nupage Novex, 4-12% (Invitrogen) en présence du marqueur de masses moléculaire (PageRuler Plus Prestained Protein Ladder, Fermentas). La migration a été faite avec le tampon MOPS (Invitrogen) à 180 V et le gel a été coloré au bleu de Coomassie (SimplyBlue SafeStain, Invitrogen) selon les recommandations du fabricant.

2.5 Analyses biophysiques des protéines recombinantes Pf-Int

2.5.1 Analyse de l'homogénéité de la préparation par diffusion dynamique de la lumière

L'étude des protéines purifiées par diffusion dynamique de la lumière a été réalisée en utilisant un appareil Protein Solutions DynaPro (Wyatt Technology) à des températures allant de 4 °C à 40 °C. 100 µl de la solution protéique ont été utilisés pour chaque analyse. Chaque analyse comprenait 30 mesures de 5 secondes et le traitement des données a été réalisé au moyen du programme DynamicsV6 (Protein Solutions, Wyatt Technology).

2.5.2 Analyse de la structure secondaire en solution de Pf-Int par dichroïsme circulaire

Dans le but de déterminer la proportion des structures secondaires des constructions de Pf-Int purifiées, celles-ci ont été analysées par dichroïsme circulaire. Au préalable les protéines purifiées ont été dialysées pendant la nuit à 5 °C dans un litre du tampon (25 mM phosphate de sodium pH 7, 500 mM fluorure de sodium). Le tampon a été changé au bout de 2 et 4 heures. Les protéines dialysées ont été centrifugées à 26 000 x g/4°C/1h et les surnageants ont été filtrés par un filtre 0.22 µm (Millex, Millipore). Les spectres ont été enregistrés sur un appareil spectrophotomètre AVIV CD (modèle 215) en utilisant une cellule de 0,02 cm. Six mesures ont été réalisées en UV lointain et moyennées. Une mesure du tampon seul a été également enregistrée. Après soustraction du signal provenant du tampon et normalisation par rapport à la concentration de l'échantillon, le contenu en structures secondaires a été estimé soit par le programme CDPro soit sur le site internet Dichroweb (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) [126-130]. La dénaturation thermique a été faite à 220 nm en appliquant un gradient de température allant de 10 à 95 °C avec un incrément de 1 °C.

Des mesures en UV proche (250 à 350 nm) ont été également enregistrées en utilisant 2 ml de protéine à 1,2 mg/ml.

2.5.3 Centrifugation analytique

La protéine purifiée a été dialysée contre 1 l de tampon 50 mM Tris-HCl pH8, 500 mM NaCl pendant 15 heures à 4 °C, a travers une membrane Spectra/Por 10 000 Da, avec changement du tampon à 3h. Une fois dialysée, la protéine a été centrifugée à 16 000 x g/4 °C/1h et le surnageant filtré sur 0.22 µm (Millex, Millipore).

Concernant la protéine C-162, trois échantillons de 400 µl à trois concentrations différentes (0,77 mg/ml, 0,25 mg/ml et 0,086 mg/ml) ont été analysés. Les mesures des vitesses de sédimentation ont été faites au moyen d'une centrifugeuse ProteomeLab XL-1 (Beckman Coulter), rotor AN50-Ti à 42 000 tpm sous vide à 20 °C et à 276 nm. Les données ont été analysées par les programmes Sedfit et Sednterp [131, 132].

2.6 Essais de détermination de la structure tertiaire de Pf-Int

2.6.1 Essais de cristallogénèse

Les essais de cristallogénèse ont été réalisés par criblage robotique à la plate-forme de « Cristallogénèse et diffraction des rayons X » de l'Institut PASTEUR. Pour chaque protéine, trois concentrations différentes, allant de 5 mg/ml à 20 mg/ml, ont été testées. 670 conditions de cristallogénèse ont été testées en utilisant les kits commerciaux : HR2-110, HR2-112, HR2-122 et HR2-126 (Hampton Research), JBScreen Classic 1-8 (CS-101L – CS-108L, Jena Bioscience), MD1-01-02 (Molecular dimensions) et les kits Classic-130701, JCSG Core 1-130924, JCSG Core 2-130925, JCSG Core 3-130926 et JCSG Core 4-130927 (Qiagen). Les essais ont été réalisés à 18°C selon la méthode « goutte assise, 200 nl » en utilisant le robot Cartesian.

2.6.2 Détermination de l'état de repliement de Pf-Int par résonance magnétique nucléaire

La protéine C-162 a été dialysée dans du tampon composé de 20 mM Phosphate de sodium pH 7, 150 mM NaCl pendant 18 heures à 4 °C avec changement du tampon de dialyse à 2 et 4 heures. La protéine dialysée a été concentrée jusqu'à 1 mM, filtrée et centrifugée à 16 000 x g pendant 1 heure à 4 °C. L'analyse de la protéine par RMN du proton a été réalisée dans un spectromètre (Varian NMR system) 500 MHz à 25 °C.

2.7 Recherche des cibles ADN de Pf-Int

2.7.1 Recherche de cibles par la méthode « Systematic Evolution of Ligands by EXponential enrichment » (SELEX)

La recherche de cibles ADN par la méthode SELEX a été réalisée en plusieurs étapes. La première étape a consisté en la création d'une librairie de cibles synthétiques. Des fragments de 25 pb bordés de chaque côté par des sites d'hybridation d'amorces connues de 20 bp (5' ATG TCA TAT GGG ATT CGT CG (N)₂₅ GCA CCA TAT GCC ATG GAC TC 3') ont été synthétisés à une échelle de 10 DO(A_{260nm}) (Proligo, Sigma Proligo France) et purifiés par HPLC.

Le brin complémentaire a été synthétisé par un cycle d'amplification PCR en utilisant l'amorce complémentaire (5'GAG TCC ATG GCA TAT GGT GC 3') et la polymérase Pfu-turbo DNA polymérase (Stratagene, Agilent Technologies). La librairie a ensuite été amplifiée

par PCR (30 cycles, 94 °C, 50.4 °C et 72 °C dénaturation, hybridation et élongation respectivement). L'ADN a été ensuite précipité dans 100% éthanol, 0,3 M acétate de sodium pH 5,4 à -20°C pendant 1 heure et centrifugé à 13 200 rpm (Eppendorf 5415 R) pendant 30 minutes à 4°C et le culot obtenu a été lavé avec 1 ml d'éthanol à 70%. Le culot a été séché au speed-vac et resuspendu dans 50 mM Tris-HCl pH 7,8, 0,1 mM EDTA.

Afin de réaliser les étapes de sélection des ADN par fixation, élution et amplification des cibles SELEX, la protéine Pf-Int-C-162 pourvue d'une étiquette 6-His en C-terminal a été dialysée dans 1 L de tampon 50 mM Tris-HCl pH 8,0 , 150 mM NaCl et 10% glycérol pendant 15 heures. Les réactions de fixation ont été réalisées par mélange de 7,5 nmoles d'ADN SELEX et de 7,5 nmoles de Pf-Int-C-162 pendant 30 minutes à 25°C. 600 µl du mélange protéine/ADN ont été chargés sur une colonne pré-équilibrée (50 mM Tris-HCl pH 8,0 , 150 mM NaCl et 5% glycérol), Ni-NTA spin column (Qiagen). La colonne a été centrifugée à 2000 rpm pendant 2 minutes (Eppendorf 5415 R). La fraction non retenue a été ré-chargée dans la colonne et la centrifugation répétée. L'élution des ADN retenus a été réalisée par des séries de lavage contenant des concentrations croissantes de NaCl (0,250 à 1 M) dans 50 mM Tris-HCl pH 8 et 5% glycérol. Dans le but de récupérer les ADN retenus par la protéine et non élués à 1 M NaCl, une dernière élution a été réalisée avec 1 M NaCl, 200 mM imidazole, 50 mM Tris-HCl pH 8, 5 % glycérol. 5 µl de chaque fraction d'élution ont été utilisés pour une réaction d'amplification par PCR (34 cycles) en utilisant des amorces complémentaires aux régions connues de la librairie. La procédure mélange/élution a été répétée 5 fois. Pour chaque cycle, la fraction d'élution à la concentration la plus élevée en sel pour laquelle un produit d'amplification par PCR a été observé a été utilisée comme substrat. Les fractions utilisées pour les cycles sont 350 mM, 425 mM, 465 mM, 470 mM et 495 mM NaCl. Les ADN récupérés à la dernière élution réalisée à 750 mM NaCl ont été clonés dans le plasmide pCR2.1 vector (Invitrogen). Le produit de clonage a été utilisé pour la transformation de cellules compétentes et une sélection par coloration bleu/blanc en utilisant des boîtes de LB contenant 80 µg/ml de X -Gal et 100 µg/ml ampicilline a été réalisée. 60 colonies positives ont été récupérées, utilisées en midi-prep (Qiagen) et les inserts séquencés en utilisant les amorces M13 dont les sites sont présents sur le plasmide (Genome-express, Cogenics, Beckman-Coulter Genomics France).

2.7.2 Recherche de cibles génomiques

L'ADN génomique de *P. falciparum* wt-3D7 (110 µg/ml) a été coupé en fragments par sonication dans le but de produire des fragments de 200-300 pb (Bioruptor, Diagenode, Belgium). L'ADN soniqué a été chargé sur gel d'agarose et les bandes correspondant à 200-300 pb ont été récupérées et purifiées par le kit gel extraction kit (Qiagen). Les extrémités des fragments récupérés ont été étiquetées avec des adaptateurs par le kit « whole genome amplification Genoplex kit » (WGA2, Sigma-Aldrich) selon les recommandations du fabricant. Ces adaptateurs sont pourvus d'une partie connue qui a été utilisée pour amplifier le stock d'ADN par PCR.

La sélection des cibles a été réalisée en utilisant la protéine Pf-Int-C-162 pourvue d'une étiquette 6-His en N-terminal et fixée sur les billes Ni-NTA (Qiagen). Les billes ont été lavées avec le tampon 50 mM Tris-HCl pH 8, 50 mM NaCl et mélangées avec 1,1 µg/ml d'ADN et 1,3 µg/ml Pf-Int-C162. Les billes ont été lavées cinq fois jusqu'à disparition du signal à DO_{260nm} dans la solution de lavage (Nanodrop, Thermo Fisher). L'ADN a ensuite été élué en utilisant des concentrations croissantes allant de 250 mM à 750 mM NaCl dans le tampon Tris HCl pH8, avec des incréments de 50 mM. Dans le but de récupérer les ADN retenus par la protéine et non élués à 1 M NaCl, une dernière élution a été réalisée avec 750 mM NaCl, 250 mM imidazole et 50 mM Tris-HCl pH 8. Des contrôles utilisant l'ADN seul ou la protéine fumarate hydratase de *Mycobacterium tuberculosis* (protéine qui n'a pas d'affinité pour l'ADN) ont été réalisés afin de suivre la fixation non spécifique.

Les deux groupes d'ADN récupérés après élution à 750 mM NaCl et à 250 mM imidazole ont été amplifiés par PCR en utilisant les amorces du kit WGA2 et clonés dans les plasmides pCR-BluntII-topo (Invitrogen) avant transformation dans les bactéries compétentes Top10 cultivées dans le milieu LB additionné de 50 µg/ml kanamycine. Les colonies obtenues ont servi à la préparation d'ADN par le kit midi-prep (Qiagen) et au séquençage des inserts (Genome-express, Cogenics, Beckman-Coulter Genomics France). Les séquences obtenues ont été identifiées par BLAST en utilisant comme référence le génome de *P. falciparum*3D7.

La technique de clonage/séquençage étant limitée par le nombre de colonies utilisées, une deuxième approche a été utilisée pour l'identification des ADN génomiques retenus par affinité par Pf-Int-C-162. Dans cette approche, les deux groupes d'ADN obtenus après élution à 750 mM NaCl et à 250 mM imidazole ont été amplifiés par PCR afin de produire 4 µg d'ADN. Cet ADN a été utilisé comme produit d'hybridation sur une puce composée de fragments de 200 pb de l'ADN de *P. falciparum* (Roche Nimblegen Inc, Europe). L'analyse

statistique des sites d'hybridation identifiés a été réalisée par le fournisseur et les séquences correspondantes ont été identifiées par comparaison au génome de *P. falciparum* 3D7.

2.8 Culture des parasites *P. falciparum*

Les cultures de parasites ont été effectuées au laboratoire BIHP, Institut Pasteur (dirigé par Artur Scherf). Les parasites ont été cultivés dans des globules rouges humains dilués dans un milieu de culture RPMI 1640 (Invitrogen) supplémenté de 2X Hypoxanthine, 10% albumax, 50 µg/ml gentamicine à 37 °C sous une atmosphère contrôlée à 5% O₂, 3% CO₂ dans des flasques de culture de tissus (Falcon). Le milieu de culture a été renouvelé par aspiration/resuspension toutes les 48 heures et les parasites ont été dilués en fonction de la parasitémie avec les globules rouges frais.

2.8.1 Suivi des cultures de *P. falciparum*

Le suivi des cultures de *P. falciparum*, en particulier la détermination de la parasitémie, a été réalisé au moyen de frottis sanguins. 100 µl de culture ont été prélevés et centrifugés. Le culot a été resuspendu dans 40 µl et étalé sur une lame. Le frottis a été ensuite séché à l'air et fixé par immersion dans une solution de méthanol pendant 10 secondes. Une fois les cellules fixées, la lame a été plongée pendant 10 minutes dans une solution de Giemsa (Giemsa R solution, RAL Diagnostics) diluée à 10% dans le PBS. La lame a été ensuite lavée à l'eau et séchée. Les observations ont été réalisées au moyen d'un microscope optique Eclipse 50i (Nikon).

2.8.2 Synchronisation des parasites

L'obtention d'une culture de parasites synchrones enrichie en stade anneaux a été effectuée par lyse des formes matures au sorbitol. Après centrifugation de la culture asynchrone pendant 15 minutes à 800 x g, le culot de globules rouges infectés a été resuspendu dans une solution de sorbitol à 5% (5 volumes de sorbitol par volume de globules rouges infectés) puis le mélange a été incubé pendant 10 minutes à 37°C avec agitation régulière. Les globules rouges non lysés ont été récupérés par centrifugation 800 x g/10 minutes, lavés deux fois avec 20 ml de milieu de culture et remis en culture.

L'enrichissement de la culture en parasites de formes matures a été réalisé par un traitement au Plasmagel. La culture asynchrone a été centrifugée 5 minutes à 800 x g et le culot de

globules rouges infectés obtenu a été resuspendu dans la solution de Plasmagel et du milieu (1 volume de globules rouges + 2,4 volume Plasmagel + 1,4 volume de milieu de culture). Le mélange a été incubé à 37 °C et le surnageant récupéré au bout de 30 minutes. Le surnageant a été ensuite centrifugé pendant 10 minutes à 800 x g et le culot contenant les globules rouges infectés enrichis en formes matures a été remis en culture.

2.8.3 Transfection des constructions ADN dans les parasites *P. falciparum*

Deux méthodes de transfection des parasites ont été utilisées ; la première a été celle qui consiste à utiliser des globules rouges pré-chargés en ADN. Ainsi, 370 µl de globules rouges à 50% d'hématocrite ont été prélevés, lavés par resuspension dans 5 ml de milieu Cytomix et centrifugés. Le culot a été ensuite repris dans un volume total 370 µl de Cytomix et 30 µl de TE contenant 100 µg d'ADN plasmidique ont été ajoutés. Le mélange a été transféré dans une cuvette d'électroporation 0,2 cm (Gene Pulser/MicroPulser Cuvettes, Bio-Rad), placé à 4°C pendant 3 minutes et soumis ensuite à l'électroporation (0.31 kV; capacitance: 960 mFD, Bio-Rad Gene pulser). Les globules rouges transfectés ont été ensuite récupérés et la cuvette rincée avec 5ml de milieu. Le mélange récupéré a été centrifugé et le culot obtenu a été transféré dans une flasque de culture, inoculé avec 20µl de globules rouges infectés par des parasites de forme mature à haute parasitémie (8-10%) et incubé à 37°C. La culture a été suivie par frottis sanguin et coloration Giemsa. Le deuxième jour, le milieu a été changé au matin et au soir. Le troisième jour, la culture a été diluée par ajout de 900 µl de globules rouges frais et l'antibiotique de sélection a été ajouté (2,5 µg/ml blasticidin). Au bout de 14 jours, 0,350 ml de globules rouges frais ont été rajoutés.

La deuxième méthode de transfection consiste à transfecter directement les globules rouges infectés. 200 µl de globules rouges infectés par les parasites en forme anneaux et de parasitémie élevée entre 8 et 10% ont été resuspendus dans 400 µl de Cytomix et mélangés avec 100 µg d'ADN plasmidique. Le mélange a été transféré dans une cuvette d'électroporation de 0,2 cm et soumis à l'électroporation (0.31 kV; capacitance: 960 mFD, Bio-Rad Gene pulser). Les cellules ont été ensuite rincées de la cuvette avec 5 ml de milieu et centrifugées. Le culot a été transféré dans une flasque de culture et incubé à 37°C. L'antibiotique de sélection a été ajouté au bout de 6 heures (2,5 µg/ml blasticidin). La flasque a été changée toutes les semaines. Le sixième jour, 0,1 ml de globules rouges frais ont été

rajoutés à la culture et au bout de 14 jours, 70% de la culture ont été prélevés et additionnés de 0,2 ml de globules rouges frais.

2.8.4 Vérification des transfections

La présence des inserts d'intérêt dans les cultures issues des transfections a été vérifiée par PCR avec les amorces utilisées précédemment pour le clonage. L'ADN génomique a été extrait par le kit GenElute Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich) selon le protocole de purification des ADN à partir des globules blancs. Le kit Expand High Fidelity PCR System (Roche) a été utilisé pour la PCR et chaque réaction a contenu 0,4 mM dNTP, 0,4 µM de chaque amorce, 20 ng d'ADN, 1,5 mM MgCl₂, et 2,5 U d'enzyme. L'amplification a été faite en utilisant la programmation suivante : une étape de dénaturation à 94 °C/2 minutes suivie de 30 cycles d'amplification (94 °C/15 secondes, 50 °C/30 secondes et 68 °C/3,5 minutes) suivie d'une élongation finale de 10 minutes à 72 °C. Les produits d'amplification ont été contrôlés sur gel agarose à 0,5%-TBE 0,5X.

2.8.5 Création du parasite 3D7-KO-Pf-Int

Afin de créer une souche de parasite dans laquelle le gène MAL13P1.42 (2916 pb, 13 exons) a été invalidé (parasite 3D7-KO-Pf-Int), un plasmide contenant une version tronquée de ce gène a d'abord été cloné dans le plasmide pCC1 (Figure 29). Dans cette version tronquée, la partie centrale du gène (allant de la position 1000 à la position 2086) a été remplacée par une cassette de résistance à l'antibiotique WR99210 codant pour la protéine hDHFR. Cette construction a été réalisée avec les fragments du gène amplifiés par PCR en utilisant les amorces : 5' CC ccgcgg ATG TAT GTA TAT ATA TAT ATG TAT GC 3' et 5' GG actagt GGA CCC AAT TCG TGC AGT TC 3' pour le premier segment (box 1) et 5' CG gaattc TGG GAA ATA GAA AAT AAG CCC ATT 3' et 5' CGC cctagg TTA TAT TTT GAA GTT ATC AAG ACC A 3' pour le deuxième fragment (box 2). Le plasmide utilisé contient également une cassette de sélection négative codant pour la thymidine kinase. Les parasites ont été transfectés par électroporation avec 100 µg de plasmide et sélectionnés avec 10 nM d'antibiotique WR99210 (Jacobus Pharmaceutical Co. Inc.). Afin d'éliminer les parasites résistants au WR99210 mais pour lesquels la réaction de recombinaison par double « cross-over » n'a pas eu lieu, une sélection négative a été réalisée avec 4 µM ganciclovir (Roche). Des clones uniques ont par la suite été obtenus par dilution limite à partir des parasites

mutants et résistants au WR99210 et au ganciclovir.

L'invalidation du gène a été vérifiée par des réactions PCR en utilisant les combinaisons d'amorces suivantes (5' →3') : Chr13.IntF, Chr13.IntR, hDHFR 5', hDHFR 3', pCC1 ATG 5' et pCC1 TAA 3' (Tableau 1). La construction a été réalisée de manière à avoir un codon d'arrêt de la traduction immédiatement en aval de la « box 1 ». Les jonctions box1/cassette de résistance et cassette de résistance/box2 ont également été séquencées.

2.9 Production d'anticorps anti Pf-Int

Deux séries d'anticorps ont été produits par l'entreprise Eurogentec. Les deux séries ont été produites dans trois souris et trois rats dans deux procédures différentes à presque deux ans d'intervalle, suivant les protocoles du fabricant. Chacune de ces productions a nécessité 0,3 mg de la protéine purifiée C162 qui a été dialysée contre le tampon 50 mM Tris pH 8, 150 mM NaCl, 2% glycérol à 4°C pendant 15 heures, avec changement du tampon à 2 heures. Le protocole d'immunisation comprend quatre injections de 15 µg de protéine à t=0, t= 14, t=28 et t=56 jours. Pour chaque production, le choix des animaux a été fait suite à un test préliminaire par Western blot du sérum pré-immun. Les sérums finaux ont été récoltés au bout de trois mois, testés par Western blot contre la protéine recombinante purifiée, et congelés à -80 °C.

2.9.1 Purification des anticorps anti-Pf-Int

La purification des anticorps a été faite en utilisant la résine NHS-activated Sepharose 4 Fast Flow (GE-Healthcare) formée d'agarose couplé au N-hydroxysuccinimide. Un milligramme de protéine Pf-Int C162 purifiée a été dialysé dans un litre de tampon 20 mM Tris-HCl pH 8, 250 mM NaCl, 2% glycérol pendant 15 heures à 4 °C, avec changement du tampon à 3 heures. La protéine a été ensuite diluée à 2 mg/ml dans le tampon de couplage (0,2 M NaHCO₃, 0,5 M NaCl, pH 8,3). Parallèlement, un millilitre de résine a été coulé dans une colonne en verre (Sigma) et lavé avec 15 ml de 1 mM HCl. La protéine diluée a été passée sur la colonne et incubée sur la nuit à 4°C. Le lendemain, la résine a été lavée avec 2 ml du tampon de couplage et bloquée par 2 lavages successifs de 2 ml chacun du tampon 0,5 M éthanolamine, 0,5 M NaCl, pH 8,3. La résine a été ensuite lavée en alternant deux tampons de pH différents : 0,1 M Tris-HCl, pH 8 et 0,1 M acétate, 0,5 M NaCl pH 4.

Pour la purification des anticorps, la résine a été lavée avec successivement 10 ml de tampon

de charge (0,1 M de phosphate de sodium, pH 8) et 10 ml de tampon d'élution (0,1 M glycine, 0,5M NaCl, pH 3) avant d'être équilibrée avec 20 ml de tampon de charge.

50 µl d'anticorps anti-intégrase ont été dilués par ajout de 150 µl de tampon de charge puis centrifugés à 10 000 x g/4 °C pendant 15 minutes. Le surnageant qui a été récupéré a été injecté sur la colonne et le mélange incubé pendant la nuit à 4 °C. Le lendemain, la colonne a été lavée avec 20 ml du tampon de charge et les anticorps retenus ont été élués par 6 injections de 0,4 ml de tampon d'élution. L'éluât de chaque injection a été récupéré dans un tube contenant 0,1 ml de tampon 1 M Tris-HCl pH 8. Les fractions récupérées ont été concentrées sur filtre Microcon 0,5 ml au seuil de coupure de 10 000 (Amicon) et 0,1% BSA (Sigma), 150 mM NaCl, et 5 mM EDTA ont été ajoutés. Les anticorps ainsi purifiés ont été conservés à 4°C et ont été vérifiés par Western blot en utilisant la protéine Pf-Int C162 purifiée comme contrôle.

2.9.2 Analyses des extraits parasitaires par Western blot

Les globules rouges infectés ont été lysés par congélation/décongélation de -20°C à température ambiante. Les globules rouges lysés ont été resuspendus dans 20 ml de PBS et les débris cellulaires ont été éliminés par centrifugation pendant 20 minutes à 5 000 tpm à 4 °C. Les culots ont été ensuite lavés par resuspension dans 2 ml de PBS et centrifugés pendant 10 minutes à 16 000 tpm à 4 °C. L'opération de lavage a été recommencée jusqu'à ce que le surnageant soit limpide. Les parasites ont alors été resuspendus dans un volume total de 150 µl de 50 mM Tris-HCl pH 8, 0,1 mM EDTA et lysés aux ultra-sons par 3 applications de 15 secondes. Les extraits parasitaires ont ensuite été dénaturés par ajout d'un volume équivalent de tampon de *Laemmli* (2% SDS) et chauffage à 95°C pendant 15 minutes.

Les différents extraits ont été chargés sur un gel SDS-PAGE NuPage Novex 4-12% et séparés à 180V en tampon de migration MOPS. Après migration, les extraits ont été transférés sur une membrane nitrocellulose Hybond C extra (GE-Healthcare) pendant 200 minutes à 4°C en utilisant le NuPage Transfer Buffer (Invitrogen) additionné de 10% de méthanol. Une fois le transfert réalisé, la membrane a été saturée avec une solution de 3% de lait (Lait en poudre Régilait, France)-TBE 0,1% tween 20 pendant une heure à température ambiante. L'anticorps primaire a été ensuite ajouté à des dilutions différentes (allant de 1/200 à 1/1000) et la membrane a été incubée pendant 15 heures à 4 °C. La membrane a été ensuite lavée 5 fois avec 20 ml TBE-0,1% tween 20 et incubée en présence de l'anticorps secondaire dilué 1/1000

(Anticorps anti-souris ou anti-rat conjugué à la molécule AlexaFluor, Invitrogen) dans 3% lait-TBE-0,1% Tween 20 pendant 90 minutes à 4°C à l'abri de la lumière. La membrane a été ensuite lavée 5 fois avec 20 ml TBE-0,1% tween 20 et lue au scanner Odyssey (Li-COR) selon les recommandations du fabricant.

2.10 Purification des oligonucléotides ADN pour les mesures de fixation à Pf-Int

Les ADN simples brins marqués en 5' (Eurofins MWG) ou non marqués (Sigma-Aldrich) ont été purifiés par chromatographie en phase liquide sur un système binaire à pompes haute pression (HPLC) équipé de : pompes LC-ADVP, Auto-injecteur SIL-ADvP, détecteur de fluorescence RF-10 XL/Détecteur UV-Vis SPD-10AvvP (Shimadzu) et d'un four thermostaté pour colonne (Gecko 2000). Pour cela, l'ADN lyophilisé a été suspendu dans 500 µl de tampon A (25mM Tris-HCl pH 8, 0,5% acétonitrile), centrifugé pendant 20 minutes à 16 000 x g à 4°C. 480 µl ont été prélevés et injectés sur la colonne échangeuse d'anions TSK-GEL DEAE-5PW (Tosoh Biosciences) pré-équilibrée avec le même tampon. L'élution des ADN a été réalisée par l'application d'un gradient de tampon (25 mM Tris-HCl pH 8, 0,5% acétonitrile, 800 mM NH₄Cl). L'élution a été suivie en choisissant le détecteur approprié soit à 260 nm, soit à 700 nm. Les fractions contenant les ADN ont été séchées au concentrateur Speedvac (Savant) et les culots ont été resuspendus dans 200 µl d'eau. Le mélange est additionné de 20 µl d'une solution d'acétate de sodium 3 M pH 5,4 et de 1 ml d'éthanol à -20°C. Le mélange est ensuite incubé pendant une heure à -20°C et centrifugé pendant 30 minutes à 16 000 x g/4°C. Les culots d'ADN sont lavés deux fois avec 1 ml d'éthanol 70%, séchés au concentrateur Speedvac et resuspendus dans 50 µl de tampon 50 mM Tris-HCl pH 8, 0,1 mM EDTA et dosés au NanoDrop (Thermo Fisher).

Les ADN simples brins ainsi purifiés sont hybridés chacun avec leurs ADN complémentaires par l'application d'un programme (94°C/ 3minutes et en réduisant la température progressivement par application d'un gradient avec réduction de 0,7°C/min) en utilisant le thermocycleur Mastercycler 1 (Eppendorf).

2.11 Amplification cibles ADN potentielles de Pf-Int à partir des bibliothèques de plasmides

Les plasmides contenant les cibles génomiques potentielles de l'intégrase ont été utilisés comme matrice dans les réactions PCR pour amplifier les ADN avec les amorces M13 marquées à l'IRD-700 (MWG Biotech GmbH). Les produits ADN ont été purifiés soit

directement par le kit QIAquick PCR purification kit (Qiagen) soit après chargement sur des gels d'agarose. La purification à partir des gels d'agarose a été faite par deux méthodes différentes, soit par le kit Gel Extraction Kit (Qiagen) impliquant une étape de chauffage à 55 °C ou à froid par le kit DNA gel Extraction Kit (Millipore).

2.12 Caractérisation de la fixation d'oligonucléotides par Pf-Int par retard sur gel (Electrophoretic mobility shift assay)

Des expériences de retard sur gel ont été entreprises en utilisant les protéines purifiées et des ADN marqués en 5' par le marqueur fluorescent DY682 ou l'IRD700. Les ADN utilisés sont issus de deux séries différentes. La première série est constituée de fragments d'ADN clonés dans des plasmides, ces ADN sont amplifiés par PCR en utilisant les amorces M13forward et M13reverse dont une est marquée en 5'. La deuxième série est composée des séquences cibles identifiées précédemment par SELEX, ces séquences sont commandées en ADN simple brin dont un brin est marqué et sont purifiés et hybridés tel que décrit plus haut.

Les réactions standards de retard sur gel contiennent 10 nM d'ADN mélangés avec différentes quantités de protéine purifiée allant de 5 nM à 2 µM dans un volume de 20 µl de tampon 50 mM Tris-HCl pH 8, 150 mM NaCl, 0.1 mM EDTA, 5 mM DTT, 50 µg/ml BSA, 2% glycérol, 0.5% Tween 20 et 10 µg/ml poly (dI-dC) (Sigma Aldrich P4929). Le mélange est incubé à température ambiante pendant 20 minutes. Les réactions sont alors additionnées de 1 µl de glycérol à 60% et chargées sur un gel natif de polyacrylamide à 5%, 20 cm x 20 cm pré-équilibré dans le tampon TBE 0,5X (Biosolve) pendant une heure à 10 V/cm à 18°C. Les complexes sont séparés par application de 10 V/cm à 18°C pendant 90 minutes à l'abri de la lumière. Après électrophorèse, le gel est scanné et analysé sur le scanner Odyssey (Li-COR) selon les recommandations du fabricant.

2.12.1 Détermination de l'état de l'ADN retenu au niveau des puits par Pf-Int par digestion à la protéinase k

0,2 µM d'ADN marqué a été mélangé avec 40 µM de protéine Pfint-C162 dans 25 µl de tampon et incubés pendant 20 minutes à température ambiante. Des quantités croissantes (0,5 µg à 5 µg) de protéinase k ont été ajoutées et les mélanges ont été incubés à 37 °C pendant 30 minutes. Les réactions sont additionnées de 2 µl de tampon de charge et chargées sur gel polyacrylamide natif 5% -0,5X TBE.

2.12.2 Mesure de la spécificité de Pf-Int par des tests de compétition

Des tests de compétition ont été également réalisés en utilisant comme compétiteur différents ADN : un site de fixation modifié de l'intégrase d'integron attC, différents fragments de gènes de *plasmodium* (Hsp, Actin) ou une séquence aléatoire (Tableau 1). La protéine purifiée (0,3 μ M) est incubée avec 1 500 nM de compétiteur non marqué ou la sonde non marquée (Selex8-22bp) pendant 20 minutes à température ambiante dans 10 μ l du tampon de binding. 10 nM de la sonde marquée sont ensuite ajoutés et le mélange est incubé pendant 20 minutes supplémentaires et le mélange est séparé sur un gel polyacrylamide natif à 5%.

Un contrôle a été également réalisé en utilisant le domaine ApiAP2 de la protéine PFF_0200C produit en fusion avec la GST (donation de Rafael Miyazawa-Martins, BIHP, Institut Pasteur) et sa cible ADN (brin supérieur : 5' AAATAAGTGTGTGCACATGTAACCA 3') de la région 5'UTR du gène MAL7P1.119 et qui contient le motif reconnu par le domaine AP2 [133]. 10 nM de l'ADN Selex-8 ou de la sonde de AP2 sont mixés avec 1,25 μ M Pf-int-C192 ou 20 μ M AP2-domain dans un volume de 20 μ l et incubés pendant 20 minutes à 20 °C. Les complexes ADN-protéine sont ensuite séparés sur un gel polyacrylamide natif à 5%-TBE-0,5X.

Amorce	Sequence 5'--> 3'
pfint-C137-NcoI	GTTCCATGGACCTTTAAAGAAAATTAC
pfint-C182-NcoI	GGTCCATGGGTACCAACGTTTCTGTAG
pfint-C188-NcoI	GGTCCATGGAAAAAATCAAGTTGAAAAAATATATAGAAG
pfint-C192-NcoI	GGTCCATGGCGTTGAAAAAATATATAGAAG
pfint-C117-NcoI	GTTCCATGGACACCATAGTAG
pfint-C180-rev-BamHI	GAATTCGGATCCTTAGTATACACAAAACCTTACTC
pLN-(+1)-AvrII	CGCCCTAGGATGTATGCATATTTTTTTAGACG
pLN-(+12)-AvrII	CGCCCTAGGATGAATACATCCTTAAAG
pLN-(-6)-AvrII	CGCCCTAGGATGTATGTATATATATATATGTATGCATATTTTTTTAGACG
pLN-rev-XhoI	CCGCTCGAGTATTTTGAAGTTATCAAGACCATC
pfint-rev-BamHI	GAATTCGGATCCTTATATTTTGAAGTTATCAAGACCATC
m13-fw	GTAACGACGGCCAG
m13-rev	CAGGAAACAGCTATGAC
Ap2_bottom	TGGTTACATGTGCACACACTTATTT
Ap2_top	AAATAAGTGTGTGCACATGTAACCA
pLN-HSP86	TCATGTCTCAGTAAAGTC
pLN-CALMOD	GTATTATTCCAATGTGCATG
Selex8-22	CAAACACACTGTGCGCGCGTCG
Sonde-Ap2	AAATAAGTGTGTGCACATGTAACCA
Protect40	GACTGTTAGCAGACTACTGGCAACGAGCTTTCAGGAATTA
HSP70	GCTAGTGCAAAGGTTCAAACCAAATTTACCAGAATCCAAT
Actin II	TCTGAAGAAGCTGTTGCTTTAGTAGTTGATAATGGAAGTGGT
Aléatoire	GTACTIONCTAATCTAGTCGTAGTATGTATGCCTCAATGAA
pfint-C137-NcoI	GTTCCATGGACCTTTAAAGAAAATTAC
pfint-C182-NcoI	GGTCCATGGGTACCAACGTTTCTGTAG
pfint-C188-NcoI	GGTCCATGGAAAAAATCAAGTTGAAAAAATATATAGAAG
pfint-C192-NcoI	GGTCCATGGCGTTGAAAAAATATATAGAAG
pfint-C117-NcoI	GTTCCATGGACACCATAGTAG
pfint-C180-rev-BamHI	GAATTCGGATCCTTAGTATACACAAAACCTTACTC
pLN-(+1)-AvrII	CGCCCTAGGATGTATGCATATTTTTTTAGACG
pLN-(+12)-AvrII	CGCCCTAGGATGAATACATCCTTAAAG
pLN-(-6)-AvrII	CGCCCTAGGATGTATGTATATATATATATGTATGCATATTTTTTT

	TAGACG
pLN-rev-XhoI	CCGCTCGAGTATTTTGAAGTTATCAAGACCATC
pfint-rev-BamHI	GAATTCGGATCCTTATATTTTGAAGTTATCAAGACCATC
m13-fw	GTAAAACGACGGCCAG
m13-rev	CAGGAAACAGCTATGAC
Ap2_bottom	TGGTTACATGTGCACACACTTATTT
Ap2_top	AAATAAGTGTGTGCACATGTAACCA
pLN-HSP86	TCATGTCTCAGTAAAGTC
pLN-CALMOD	GTATTATTCCAATGTGCATG
Chr13.IntF (position sur le chromosome 407,151 à 407,127)	GAGAAGTTATTAATTTGGTGAAAG
Chr13.IntR (position 403,785 à 403,809)	CATATTAGA AACGTCCAAATA AAT
hDHFR 5'	GCATGGTTCGCTAAACTGCAT C
hDHFR 3'	TCATTCTTCTCATATACTTC
pCC1 ATG 5'	CTTTAAATTCATGCAAAAATTTAC
pCC1 TAA 3'	AACTATGCGGCATCAGAGCAG
Pf-Int_F1	ATTGGACTCAGACCATCG
Pf-Int_B1	AGTTTGTCATAGGAACGG

Tableau 1 : Amorces et oligos utilisés pour les clonages et les EMSA.

3 RESULTATS

L'identification du cadre ouvert de lecture correspondant à un nouveau membre de la famille des recombinases à tyrosine, Pf-Int, dans PlasmoDB nous a permis de procéder à la caractérisation bioinformatique, biophysique, et fonctionnelle de la protéine correspondante. Ces études constituent des étapes préliminaires à la poursuite des études de biologie structurale dans le but de déterminer l'organisation tertiaire de cette protéine. Par ailleurs, nous avons procédé au suivi de l'expression de la protéine *in vivo*, et avons essayé de déterminer son rôle dans la physiologie de *P. falciparum*. Nous avons ainsi procédé d'une part à l'inactivation du gène, et d'autre part à la mise en place d'un système de surexpression épisomale. Ces différentes étapes sont décrites dans les chapitres suivants.

3.1 Étude du cadre de lecture ouvert identifié de Pf-Int

Le premier volet de l'analyse bioinformatique de la protéine a concerné la caractérisation de sa structure primaire ainsi que sa conservation chez les autres espèces de *Plasmodia*, tout au moins au niveau des génomes séquencés jusqu'à présent. Le cadre de lecture ouvert identifié de la recombinase à tyrosine Pf-Int dans PlasmoDB (répertorié MAL13P1.42) consiste en un gène dont la taille est de 2916 pb. La région assignée à Pf-Int est bordée par les gènes PF13_0048 et PF13_0047 qui codent respectivement pour une nucléotide-hydrolase et une protéine conservée dont la fonction n'a pas encore été identifiée. Le locus MAL13P1.42 est composé de 13 exons et son épissage d'introns est prédit pour produire un ARN messager de 1,473 kb. La traduction de cet ARNm permettrait de produire une protéine de 490 acides aminés dont la masse moléculaire est estimée à 57 kDa.

3.1.1 Étude de la conservation de Pf-Int

Ce cadre de lecture est présent et conservé dans les génomes de *P. bergheii*, *P. chabaudii*, *P. knowlesi*, *P. yoelii*, et *P. vivax* avec une identité de séquence de la protéine correspondante allant entre 67% et 74%, (Tableau 2, Figure 22). La région partant du résidu 151 au résidu 490 est la plus conservée et présente entre 87% et 90% d'identité de séquence. La comparaison de synténie parmi les génomes de *P. falciparum*, *P. yoelii* et *P. vivax* montre que les régions codantes sont hautement conservées [134].

Espèce de <i>Plasmodium</i>	<i>P. bergheii</i> [442 aa]	<i>P. chabaudi</i> [447 aa]	<i>P. knowlesi</i> [472 aa]	<i>P. yoelii</i> [453 aa]	<i>P. vivax</i> [475 aa]
<i>P. falciparum</i> Pf-Int entière : [490 aa]	74% (1-442)	74% (1-447)	67% (1-472)	72% (1-453)	70% (1-475)
<i>P. falciparum</i> Pf-Int : (151-490)	88% (112-442)	87% (107-447)	90% (130-472)	88% (113-453)	90% (125-475)

Tableau 2 : Comparaison de la conservation des acides aminés de la recombinaise Pf-Int par rapport aux protéines homologues d'autres espèces de *Plasmodium*. Les tailles des protéines entières sont décrites par les valeurs entre les crochets []. Les pourcentages d'identité sont calculés par rapport aux positions précisées entre parenthèses (). L'analyse comparative montre que la protéine étudiée a une identité de séquence importante (~70%) et que la région la plus conservée est comprise entre les positions 151 et 490 (identité de séquence~88%).

Des homologues de Pf-Int ont également été identifiés chez d'autres *Apicomplexa* tels que *Toxoplasma gondii*, *Neospora caninum* et *Eimeria tenella*. L'alignement des séquences protéiques (Figure 23) de Pf-Int avec celles de *Toxoplasma gondii* et *Neospora caninum* est en accord avec celui des différentes espèces de *Plasmodium* (le génome d'*Eimeria tenella* n'étant pas encore finalisé, la séquence n'a pas été incluse dans l'analyse). En effet, la comparaison montre que la région C-terminale (à partir de la position ~175 de Pf-Int) est très conservée et que la région N-terminale est plus divergente.

Vu le degré élevé de conservation entre les espèces du genre *Plasmodia* ainsi que chez les autres *Apicomplexa*, on peut considérer que Pf-Int joue un rôle important dans la physiologie du parasite.

Il est donc légitime d'approfondir cette analyse et d'essayer de montrer si d'autres éléments de conservation peuvent être identifiés à d'autres niveaux, notamment l'organisation structurale de la protéine par rapport à d'autres membres de la famille des recombinaises à tyrosine, dont les fonctions ainsi que les structures sont établies. Ceci permettra d'extrapoler les rôles des différentes régions de la protéine, de confirmer les positions des résidus critiques à la catalyse et de poser une hypothèse sur le mécanisme de la réaction de recombinaison dans lequel Pf-Int pourrait intervenir.

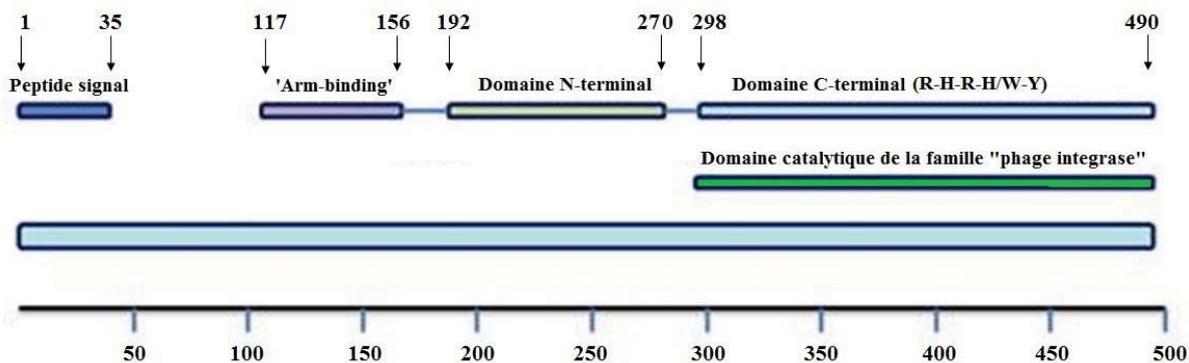


Figure 24 : Représentation schématique de l'organisation des domaines de Pf-Int (490 aa). Rectangle en bleu foncé : peptide signal en N-terminal (Position : 1 à 35 aa). Les annotations montrent que Pf-Int pourrait être organisée en trois domaines dont deux appartenant à la région hautement conservée entre les espèces de *Plasmodia* (~150-490aa) : le premier domaine s'étend de la position 298 à la position 490 (rectangle mince en bleu clair) et constitue le domaine C-terminal catalytique conservé par les Y-SSR (annoté (plasmoDB) comme étant : domaine catalytique de la famille « phage integrase »), le deuxième domaine s'étend de la position 192 à la position 270 (rectangle en vert clair) et est attribué (plasmoDB) par homologie au domaine N-terminal de λ -Int. La prédiction de la structure tertiaire de Pf-Int par le programme Phyre et l'utilisation de la structure de λ -Int (PDB : 1z1bA) comme modèle (programme Discovery Studio) ont permis de confirmer les annotations présentées par PlasmoDB, en particulier l'existence des deux domaines N- et C-terminaux et de montrer l'existence potentielle d'un troisième domaine (barre horizontale de couleur mauve, position 117 à 156) équivalent au « arm-binding domain » de λ -Int.

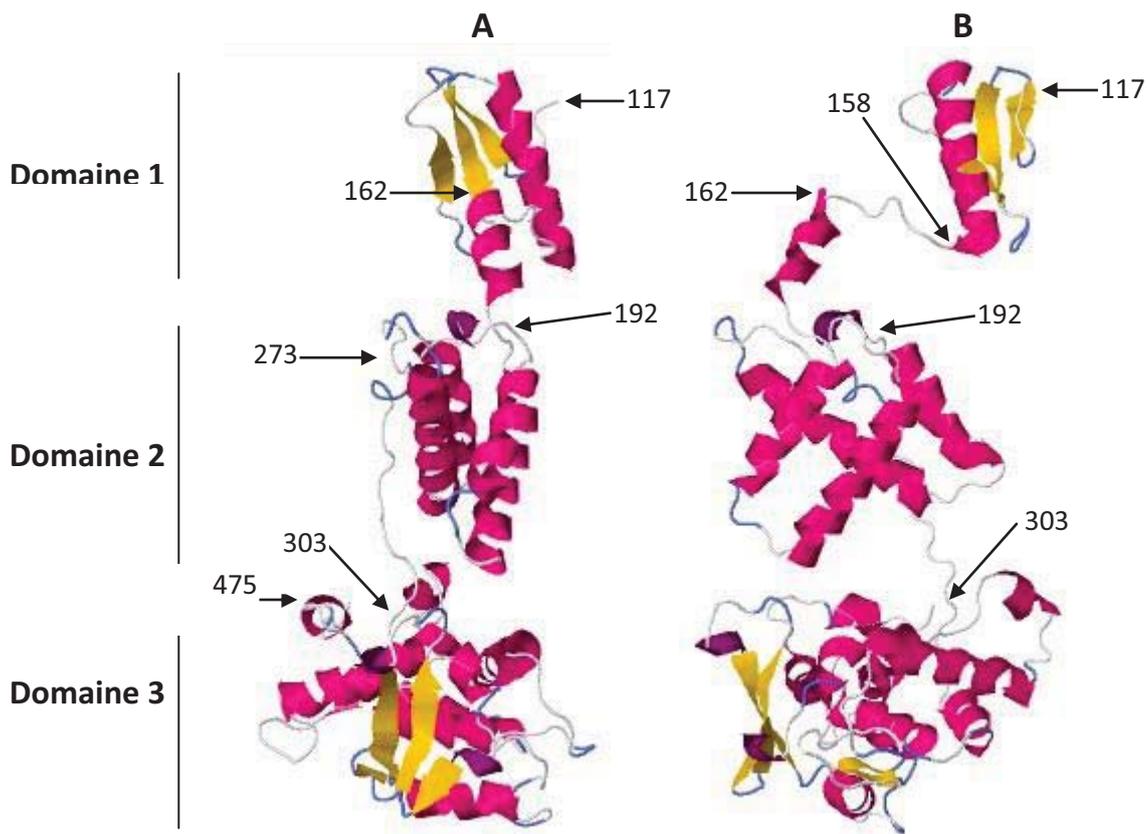


Figure 25 : Prédiction de la structure tertiaire de l'intégrase de *P. falciparum*.

La structure de l'intégrase du phage lambda (PDB-id c1z1bA) a été utilisée pour modéliser la structure de l'intégrase de *P. falciparum*. A, B sont des représentations à 90° du modèle de Pf-int. Les hélices α sont représentées en rubans rouges et les feuillets β en flèches jaunes. Les positions des acides aminés délimitant les domaines structuraux identifiés sont indiqués par des flèches noires. La protéine est organisée en trois domaines : le premier domaine (117-158 aa) est équivalent au domaine « arm-binding » de λ -Int, les deuxième (192-273 aa) et troisième domaine (303-475 aa) correspondent aux domaines N- et C-terminaux classiquement retrouvés chez les autres Y-SSR.

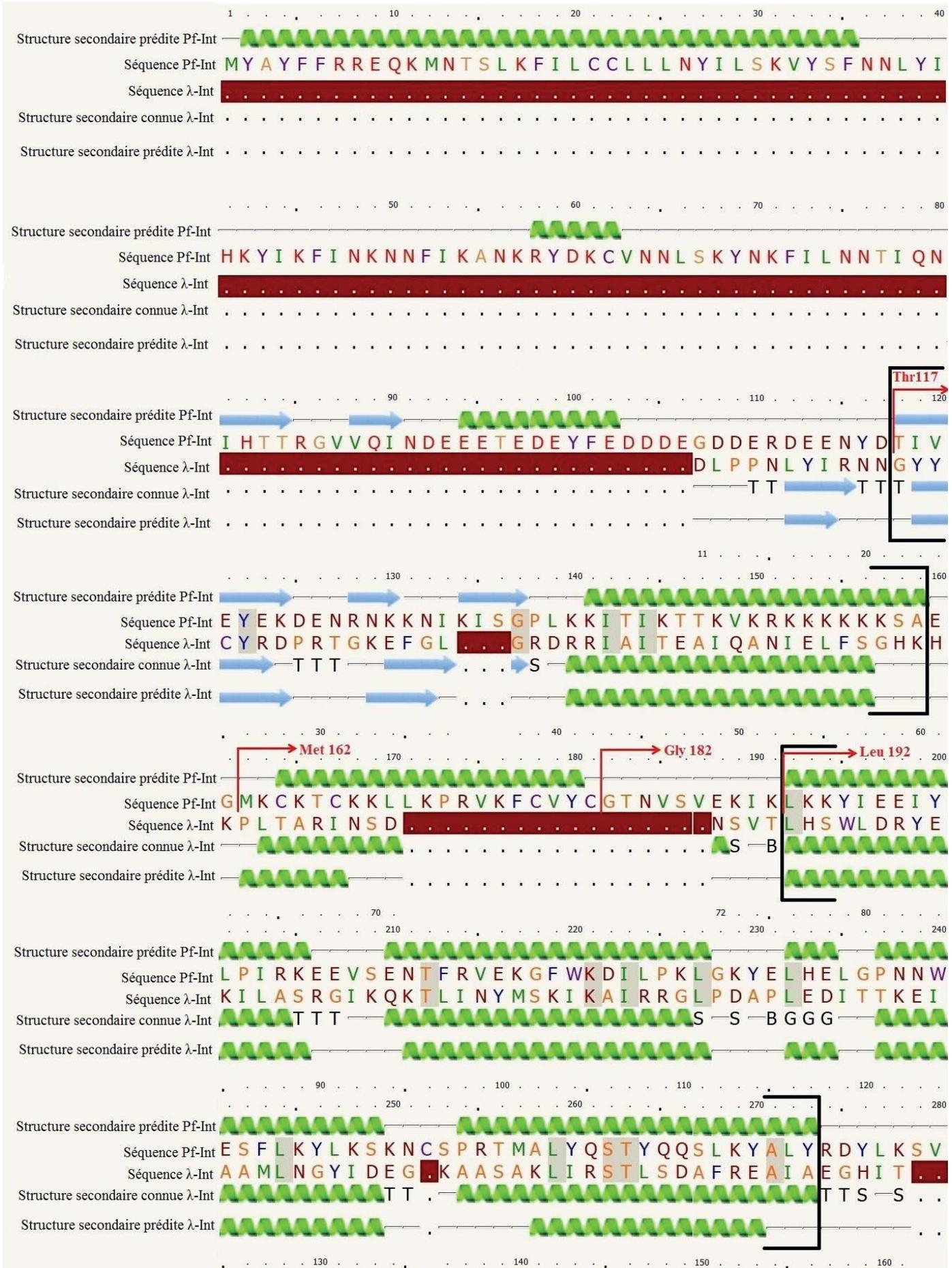
3.1.2 Prédiction de la structure tertiaire de Pf-Int

D'après les annotations automatisées réalisées par PlasmoDB à la suite du séquençage du génome entier de *P. falciparum*, Pf-Int serait composée de plusieurs domaines dont deux appartenant à la région hautement conservée aa 151-490 (Figure 24). Le premier domaine s'étend du résidu 298 au résidu 490 (Figure 24, barre en bleu clair) et est attribué au domaine catalytique conservé par la famille des recombinases de type intégrase de phage. Le deuxième domaine couvre la partie d'aa 192-270 (Figure 24, barre en vert) qui est attribuée par homologie au domaine N-terminal de l'intégrase du phage lambda. La présence d'un peptide signal et d'une hélice transmembranaire N-terminaux (position : 1-35 aa) est également prédite (Figure 24, barre en bleu foncé).

Malgré le faible pourcentage d'identité de séquences globales entre les différents membres, la famille des recombinases à tyrosine est caractérisée par plus de similitude de leurs C-terminus et une forte homologie des structures tertiaires [59].

Les essais de prédiction de la structure secondaire par le programme PSIPRED, de la structure tertiaire par le programme PHYRE et l'utilisation de la structure de l'intégrase du phage lambda (PDB : 1Z1B) comme modèle (Figures 25 et 26), nous ont permis de réaliser une prédiction de structure tertiaire de l'intégrase de *P. falciparum* (Figure 25). Seule la région allant du résidu 115 au résidu 490 de Pf-Int a pu être modélisée. Le pourcentage d'identité est de 13 % pour la région d'aa 115-490 et de 15 % pour la région d'aa 192-490. Cette prédiction montre que les principales parties de la structure tertiaire de l'intégrase de *Plasmodium* sont potentiellement présentes avec la même organisation que l'intégrase lambda-Int composée de trois domaines structuraux distincts (Figure 13, A). La figure 26 montre un alignement de la séquence primaire et des structures secondaires prédites à partir de Pf-Int avec la structure primaire et les structures secondaires prédites et connue de λ -Int utilisé comme modèle.

Chez Pf-Int (Figures 25 et 26), le premier domaine allant des aa 117 à 158 est composé d'un feuillet β à trois brins et d'une hélice α . Ce domaine serait l'équivalent du « arm-binding domain » présent dans λ -Int et qui sert à l'interaction avec les séquences auxiliaires (POP') présentes au niveau du site de fixation. Le deuxième domaine structural prédit s'étend de l'aa 192 à 273. Ce domaine est formé de quatre hélices α . Le troisième domaine composé de la région allant de l'aa 302 à 474 comprend un ensemble de 7 hélices α et de deux feuillets β de deux brins chacun. Ces trois domaines sont reliés entre eux par deux régions charnières : aa 159-191 et aa 274-301. La première région charnière est composée d'une hélice α allant de la



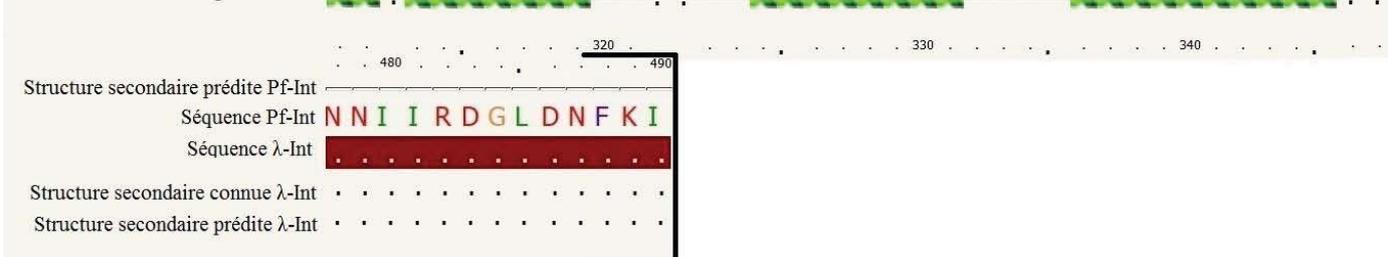
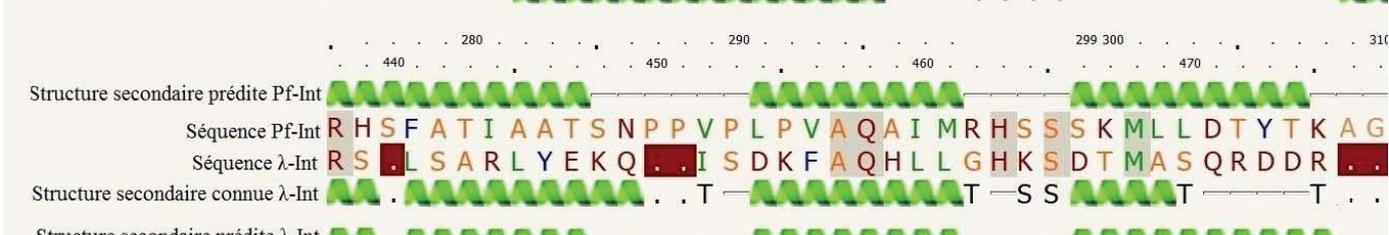
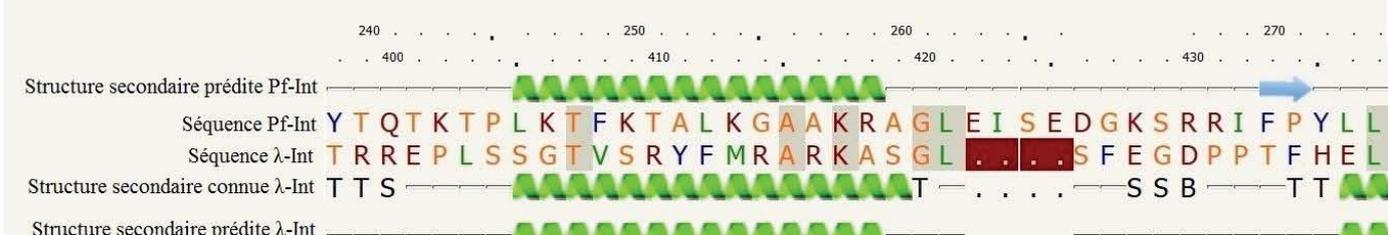
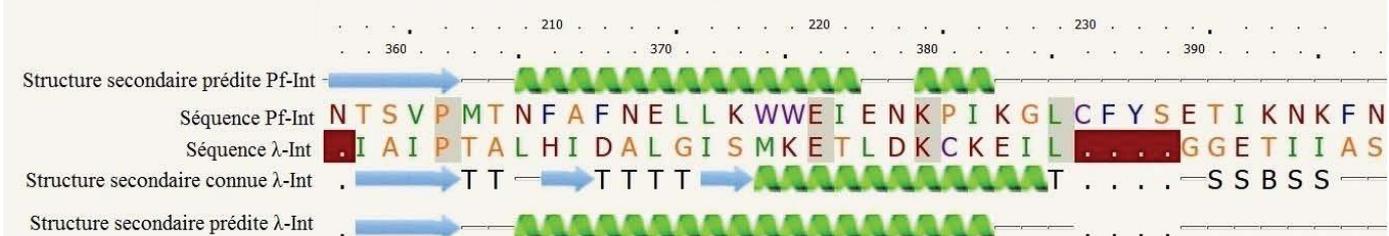
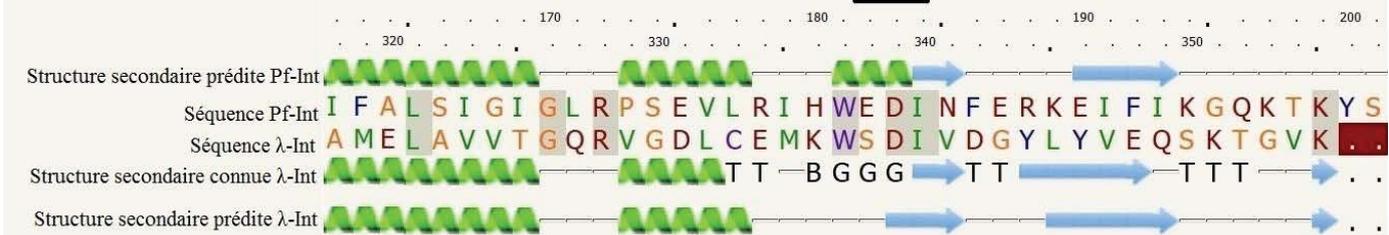
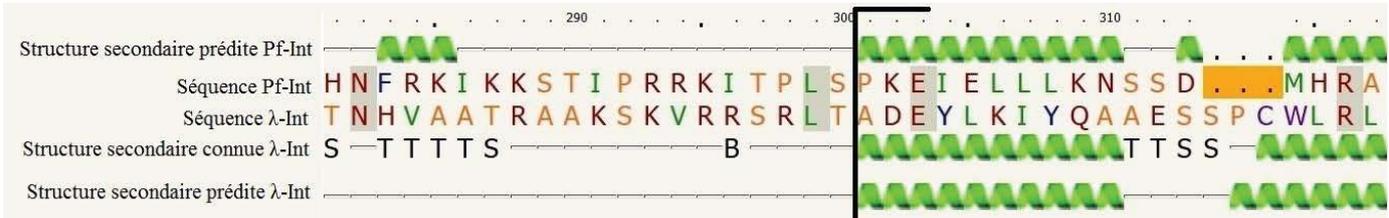
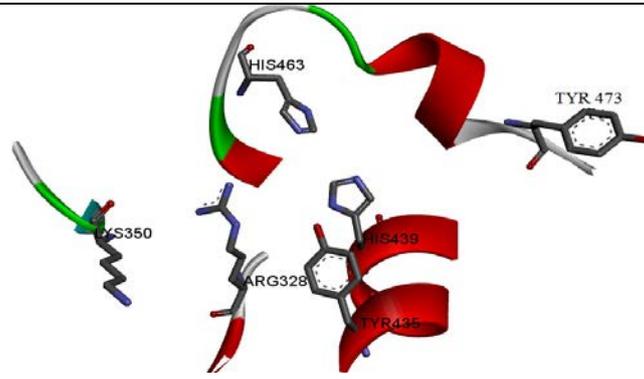
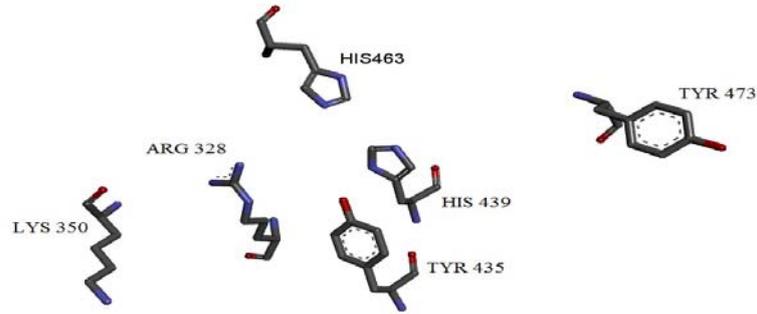


Figure 26 : Alignement de la structure primaire et de la structure secondaire prédite de Pf-Int avec la structure primaire, secondaire prédite et secondaire connue de l'intégrase du phage lambda utilisé comme modèle. Les acides aminés sont numérotés de 1 à 490 pour Pf-Int et de 11 à 356 pour λ -Int. Les hélices α sont symbolisées par des rubans verts, les brins β par des flèches bleues et les régions non structurées par des bâtons gris. Les insertions sont symbolisées par des points. Les insertions dans la séquence de λ -Int sont colorées en rouge foncé et les insertions dans la séquence de Pf-Int sont colorées en jaune. Les trois domaines prédits (aa117-158, aa192-273 et aa303-490) sont délimités par des crochets noirs []. Pour chacun des quatre fragments étudiés de la protéine (C-117, C162, C182 et C192), la position du premier acide aminé utilisé est indiquée par des flèches rouges (Thr 117, Met162, Gly 182 et Leu192). Attribution selon le dictionnaire de structure secondaire de protéines (DSSP) : G : hélice₃₁₀, T : coude « turn » impliquant des liaisons hydrogènes du résidu i à $i+n$ ($n=3,4,5$), B : résidu isolé dans un pont β , notamment formant des ponts hydrogènes décalés au lieu de parallèles sur deux séries de résidus ($i-1$ à $i+1$, $j-1$ à $j+1$), S : coude, angle de 70° entre les résidus allant des positions $i-2$ à $i+2$.

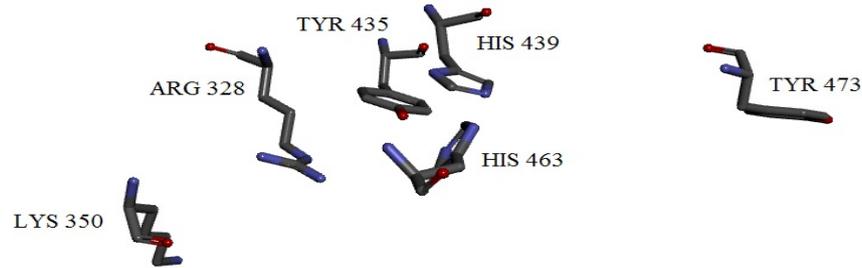
A



B



C



IntI1	119	PDEVVRIIGFLEG---EHLRFAQLLYGTGMI	ISEGLQLRVKDLD	FDHGTIIVREGKG-SKDRALM----	LPESLAPSLRE
IntI4	108	RDEIRRLLLEIVDP---KHQLPIKLLYSGSL	ELMECMRLRVQDID	FDYGAIRIWQGGK-GKNRTVT----	LAKELYPHLKE
XerD	121	LLQAPLIDQPLEL---RDKAMLEVL	YATGLRVSELVGLTMSDISLRQGVV	RV-IGKG-NKERLVP----	LGEEAVYWLET
XerC	121	RLLDIDINDPLAV---RDRAMLEVMYGAGL	ELSELVGLDIKHLDES	GEVWV-MGKG-SKERRLP----	IGRNAVAWIEH
PfInt	301	PKEIELLLKNSSD---MHR	IAFALSIGIGLRPSEVLR	IHWEDINFERKEIFI--	KG-QKTKYSNTSVPMTNFAFNELLK
LambdaInt	182	ADEYLKIYQAAESSPCWLR	LAMELAVVTGQ	RVGDLCEMKWSDIVDGYLYVEQ	SKT-G-VKIAIPTAL--HIDALGISMKE
Cre	146	RSLMENS	DRCQDI---RNLAFLGIAYNTLL	IAEIA	IRVVKDISRTDGGRLIHI-GRTKTLVSTAG--VEKALSLGVTK
		:	*	:	: . . .

IntI1	191	QLSRARAWLKDQAEGRSG-VALPD	ALERKYP	PRAGHSWPWFVFAQHTHST	DP	RS	GVVRRHHMYDQTFQRAFKR-AVEQA
IntI4	180	QIALAKRYYDRDLHQKNYGGVWL	PTALKEKYP	NAPYEFRWHYLFPSFQLSLD	PE	SD	VMRHHMNETVLQKAVRR-SAQEA
XerD	192	YLEHGRPWL-----		LNGV	S	I	DVLFPSQRAQQMTRQTFWHRIKH-YAVLA
XerC	192	WDLDR-----		LFG	S	E	DALFSLKLGKRISARNVQKRFAE-WGIKQ
PfInt	374	WWEIENKPIKG-----		LCFY	S	E	TIKNKFN-YTQTKTPLKTFKTALKG-AAKRA
LambdaInt	258	TLDKCKELGG-----		ETI	I	A	STRREPLSSGTVSRYFMR-ARKAS
Cre	220	LVERWISVSGVADDPNN-----		YLFC	R	V	RKNGVAAPSATSQLSTRALEGI
		:	:	:	:	:	FEATHRLIY

IntI1	269	GITK-----PATP	TLRHSFATALLRS-GYDIRTV-QDLLG	SDVSTM-I	Y	TH	VLVKGGAGVRSPLDALPPLTSER
IntI4	259	GIEK-----TVTCH	TLRHSFATHLLEV-GADIRTV-QEQLG	TDVKTQ-I	Y	TH	VLDRGASGVLSPLSRL-----
XerD	235	GIDSE-----KLS	PVLRHAFATHLLNH-GADLRV	V-QMLLG	S	D	LSTTQ-IYTHVATERLRQLHQHHPR
XerC	231	GLNN-----HVH	PKLRHSFATHMLESSG-DLRGV-QELLG	HANLSTTQ-I	Y	TH	LDFQHLASVYDAAHPRAKRGK--
PfInt	420	GLEISEDGKSRRIFPYLLRHS	FATIAATS	NPPVPLPVAQA	IMR	S	SSKMLLDYTKAGNNIIRDGLDNFKI-----
LambdaInt	297	GLSEFGD---PPTF	HELRS-LSARLYEKQISDKFA--QHLLG	HKS	D	T	MASQ-YRDDRGREWDKIEIK-----
Cre	274	GAKDDSGQRYLAWSGHS	ARVGAARDMARA-GVSI	PEIMQAG-G	T	N	VIMN-YIRNLDSETGAMVRLLEDGD-----
		*	:	*	:	:	:

Figure 27 : Modélisation de l'organisation du site catalytique de Pf-Int. A : Figure montrant les acides aminés catalytiques et une partie des structures secondaires dans lesquelles ceux ci sont impliqués. B : Organisation dans l'espace des chaînes latérales des acides aminés du site catalytique. C : Rotation de 90° de B. L'organisation dans l'espace des acides aminés Arg328, Lys350, Tyr435, His439, His463 et Tyr473, en particulier Arg328, Tyr435, His439 et His463 montre qu'ils sont capables de former le site catalytique de Pf-Int commun aux autres Y-SSR (R-H-R-(H/W)-Y) avec une spécificité due au remplacement de Arg438 par Tyr435. Les résidus Lys350 et Tyr473 adoptent des positions décalées par rapport aux autres acides aminés du site catalytique mais sont capables de se réorientés pour participer à la catalyse. D : Alignement de la séquence protéique du domaine catalytique de Pf-Int avec ceux d'autres Y-SSR. Vert : acides aminés catalytiques conservés, rouge : la tyrosine catalytique conservée : R328-K350-Y(H)435-R438-H463 and Y473

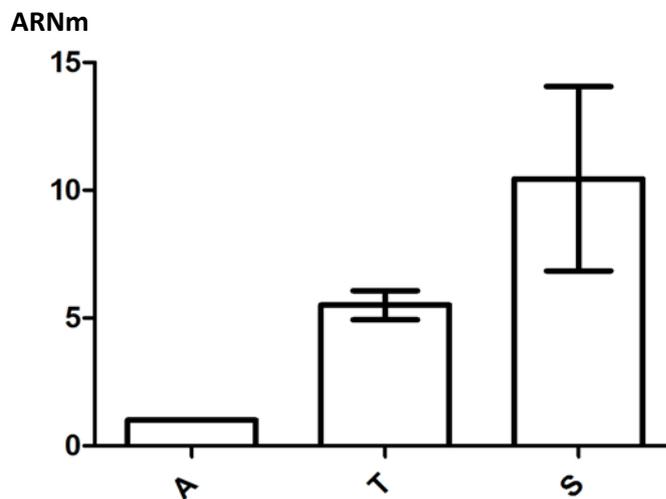


Figure 28 : Etude de l'expression de l'ARN messager de MAL13P1.42 codant pour Pf-Int par RT-PCR. L'analyse montre que le gène est transcrit de manière différente en fonction du stade parasitaire. Le niveau d'expression est 8 à 10 fois plus élevé pendant le stade schizont que pendant le stade anneau. Mesures réalisées en double (les histogrammes représentent les moyennes \pm écart-type). Ordonnée : taux relatif d'expression de l'ARN. Les expériences ont été réalisées en double. A : stade anneau. T : stade trophozoite. S : stade schizont.

position 164 à 180 alors que la deuxième est totalement désordonnée. Les deuxième et troisième domaines forment les éléments structuraux classiquement partagés par toutes les recombinases à site spécifique à tyrosine.

L'étude de la structure prédite a permis d'observer la co-localisation d'un certain nombre de résidus qui pourraient constituer le site catalytique : R₃₂₈ - K₃₅₀ - Y₄₃₅/H₄₃₉ - R₄₃₈ - H₄₆₃ - Y₄₇₃ (Figure 27, A, B et C). L'alignement de la séquence protéique du domaine catalytique de Pf-Int avec ceux d'autres recombinases à tyrosine a permis de montrer que ces résidus correspondent aux résidus conservés dans les Y-SSR (Figure 27, D).

3.2 Etude de Pf-Int *in-vivo*

La présence d'un cadre de lecture prédit, ainsi que la conservation de ce dernier ne présage ni le taux d'expression de la protéine correspondante, ni la présence de la fonction prédite dans la cellule. Une fois la recombinase identifiée et le cadre de lecture caractérisé, nous nous sommes d'abord intéressés à l'étude *in-vivo* de Pf-Int. Pour cela, nous avons procédé à la caractérisation des niveaux d'expression par les parasites de l'ARN messager et de la protéine. En prévision que le taux d'expression puisse être déterminant pour la fonction de Pf-Int, nous avons aussi créé une souche de parasite dont le gène codant Pf-Int a été invalidé (souche 3D7-KO-Pf-Int créée par Christine Benatar, BIHP, Institut Pasteur) et nous avons essayé d'observer l'apparition d'un nouveau phénotype.

3.2.1 Analyse des extraits d'ARN messager

L'étude du niveau d'expression de l'ARN messager de Pf-Int du parasite sauvage 3D7 par RT-PCR a permis de montrer que le gène est transcrit durant la phase intra-érythrocytaire de manière différente en fonction des stades parasitaires (Figure 28). Le niveau de transcription du gène est faible pendant le stade anneau et augmente progressivement pour atteindre un maximum au niveau du stade schizont. Le niveau d'expression à cette étape est 8 à 10 fois supérieur que celui du stade anneau. Cette observation est en concordance avec les observations de *Bozdech et al* [14] qui ont observé que le taux d'expression des protéines impliquées dans des mécanismes liés à l'ADN tel que la réplication DNA Pol α , DNA Pol δ et deux topoisomérases augmente durant ce stade.

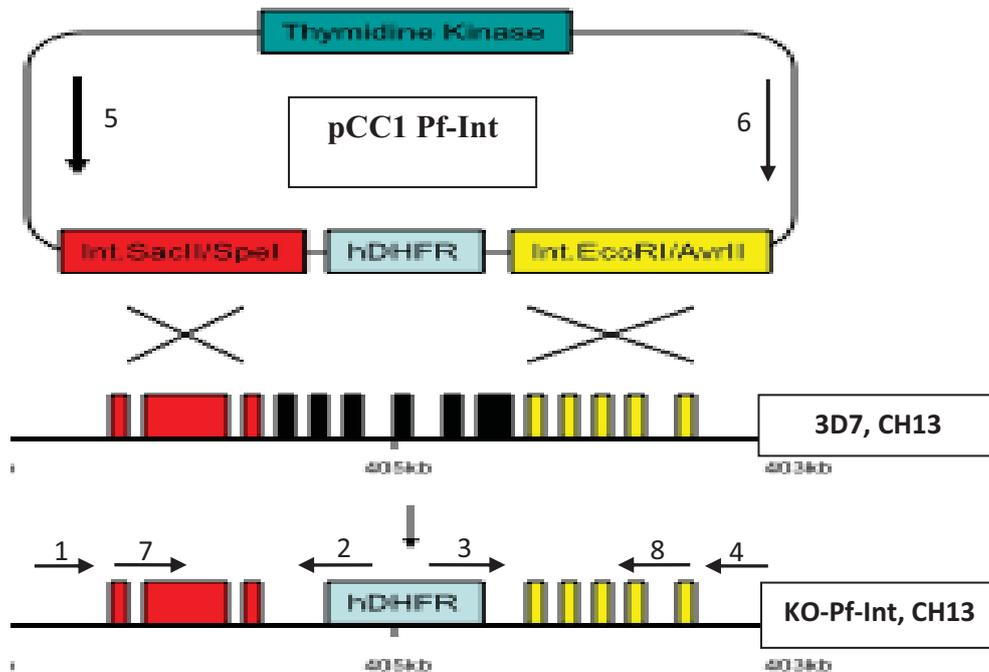


Figure 29 : Stratégie utilisée pour l'invalidation du gène MAL13p1.42 de *P. falciparum* (Parasite 3D7-KO-Pf-Int). La région centrale du gène (position 1000 à position 2086) du gène MAL13P1.42 (3.9 kb) a été remplacée par une cassette de résistance à l'antibiotique WR99210 codant pour la protéine hDHFR. Le plasmide utilisé porte également une cassette de sélection négative codant pour la thymidine kinase. Les parasites ont été sélectionnés par leur résistance aux antibiotiques WR99210 et ganciclovir. L'insertion a été vérifiée par amplification PCR en utilisant les couples d'amorces 1/2, 3/4, 5/6 et 7/8. Les amorces 1 et 4 sont spécifiques du chromosome 13, les amorces 2 et 3 sont spécifiques de la cassette de résistance hDHFR et les amorces 5 et 6 sont spécifiques du plasmide pCC1.

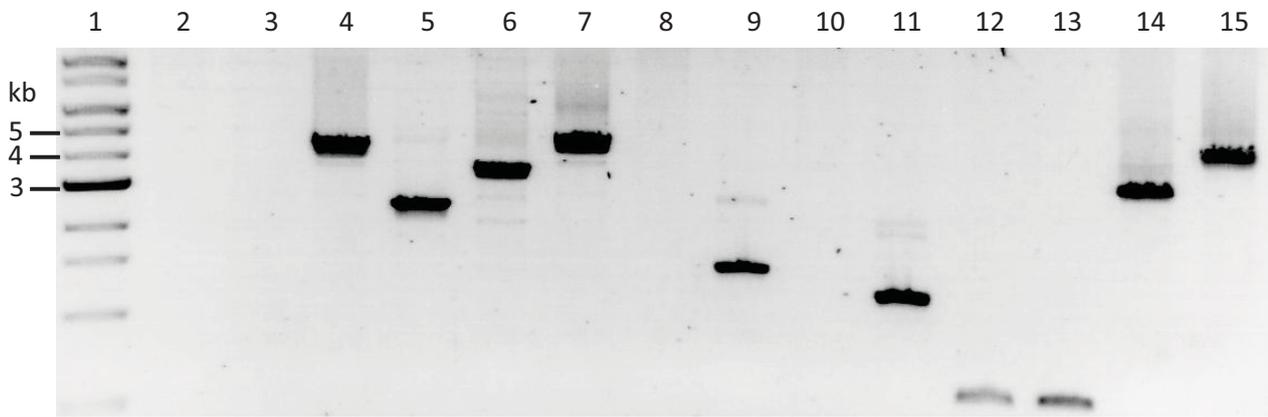


Figure 30 : Vérification de l'invalidation du gène MAL13P1.42 par PCR. Piste 1 : marqueurs de taille moléculaire (kb). Pistes 2, 6, 8, 10, 12 et 14 : Produits d'amplifications à partir des extraits d'acides nucléiques du parasite sauvage 3D7 ; Pistes 3, 7, 9, 11, 13 et 15 : Produits d'amplifications à partir des extraits d'acides nucléiques du parasite 3D7-KO-Pf-Int ; Piste 4 : Produit d'amplification à partir du plasmide pCC1-Pf-Int utilisé pour la transfection ; Piste 5 : Produit d'amplification à partir des plasmide pCC1 vide. Pistes 2, 3, 4 et 5 : amorces 5 et 6. Pistes 6 et 7 : amorces 1 et 4. Pistes 8 et 9 : amorces 1 et 2. Pistes 10 et 11 : amorces 3 et 4. Pistes 14 et 15 : amorces 7 et 8. Pistes 12 et 13 : contrôle de la qualité de l'ADN par les amorces spécifiques à un gène *var*.

3.2.2 Vérification de l'invalidation du gène MAL13P1.42 (parasite 3D7-KO-Pf-Int)

L'invalidation du gène qui code Pf-Int a été réalisée par remplacement de 1kb de la partie centrale du gène (allant de la position 1000 à la position 2086) par une cassette de résistance à l'antibiotique WR99210 codant pour la protéine hDHFR dont la taille est de 2 kb (Figure 30). Le contrôle de l'invalidation du gène MAL13P1.42 a été réalisé par des amplifications PCR en utilisant plusieurs couples d'amorces (Figure 29). Le premier couple d'amorces est constitué d'une amorce qui s'hybride sur la région chromosomale 3' du gène (amorce 1) et d'une deuxième amorce qui s'hybride sur la région chromosomale 5' du gène (amorce 4). La taille du fragment obtenu à partir des extraits du parasite sauvage 3D7 est de 3 kb (Figure 30, piste 6) alors que la taille obtenue pour le parasite 3D7-KO-Pf-Int (Figure 30, Piste 7) correspond bien à la taille attendue de 4 kb.

Pour la deuxième réaction d'amplification, le couple d'amorces utilisé est constitué de l'amorce 1 qui s'hybride sur le chromosome et d'une deuxième amorce (amorce 2) qui s'hybride sur la région 3' de la cassette de résistance. Pour la troisième réaction, le couple d'amorces utilisé est constitué de l'amorce 4 qui s'hybride sur le chromosome et d'une deuxième amorce (amorce 3) qui s'hybride sur la région 5' de la cassette de résistance. En accord avec ce qui attendu, l'amplification avec ces amorces à partir des extraits du parasite sauvage 3D7 n'ont pas donné d'amplification (Figure 30, Pistes 8 et 10 respectivement) alors que chez les parasites 3D7-KO-Pf-Int, une bande d'amplification a été observée pour chaque couple (Figure 30, Pistes 9 et 11 respectivement). Ceci prouve qu'il y a bien eu intégration de la cassette de la résistance au niveau des sites prévus du chromosome 13 et que cette intégration a eu lieu par un double cross-over. Dans l'hypothèse où une deuxième copie non invalidée du gène aurait pu être créée par erreur suite à ces essais d'invalidation, nous avons réalisé un contrôle avec un couple d'amorces (7 et 8) qui s'hybrident sur les régions « box1 » et « box2 ». Dans le cas où deux copies (l'une mutée et l'autre naturelle) existeraient, nous aurons deux bandes d'amplification de tailles 3 et 4 kb. Dans notre cas (Figure 30, Piste 15), une seule bande a été observée dont la taille est de 4 kb. Concernant le parasite sauvage, ce même contrôle a donné une seule bande à la taille attendue de 3 kb (Figure 30, Piste 14). Des contrôles additionnels ont été réalisés afin de confirmer l'absence du plasmide dans le parasite KO. Pour ces contrôles, nous avons utilisé les amorces spécifiques du plasmide 5 et 6. Dans l'hypothèse où une copie épisomale aurait pu être présente dans une partie des parasites, l'amplification par ces amorces aurait donné un produit de 4,5 kb (Figure 30, Piste 4). Cependant, dans notre cas, l'utilisation de ces amorces sur les extraits des parasites KO et

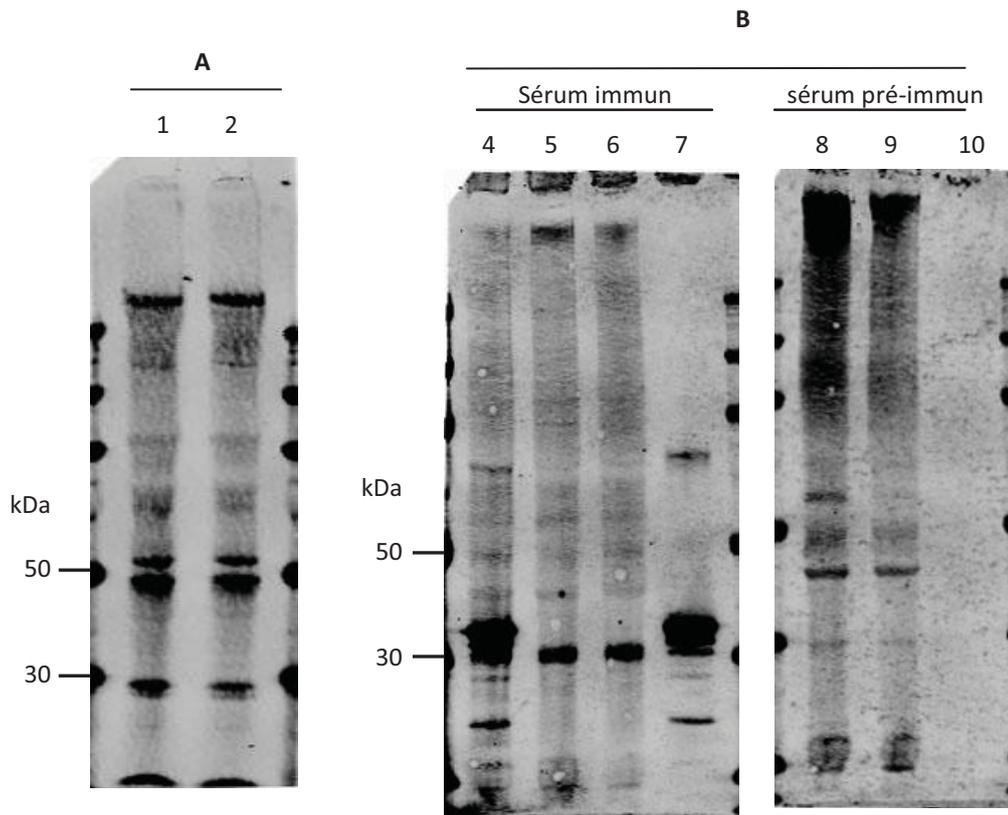


Figure 31 : Étude comparative des extraits protéiques des parasites sauvage et KO. L'analyse des extraits protéiques des deux souches 3D7 et KO par Western-blot en utilisant les anticorps anti-Pf-Int n'a pas permis de mettre en évidence une différence entre ces deux souches. En effet nous avons observé le même profil pour les deux souches et ceci en utilisant deux anticorps différents : Souris (Pistes 1 et 2) ou Rat (Pistes 5 et 6). Les deux bandes observées dans les extraits du parasite 3D7 (Piste 1) et dont les masses moléculaires (~45 et ~55 kDa) sont proches de la masse moléculaire théorique de Pf-Int (57kDa) ont également été observées dans les extraits du parasite KO (Piste 2). Les bandes ont été révélées par Odyssey à 700 nm.

Membrane A : anticorps produits dans la souris.

Piste 1 : Extraits parasitaires 3D7 sauvage.

Piste 2 : Extraits parasitaires K.O.

Membrane B : Pistes 4-7 : anticorps produits dans le rat. Pistes 8-10 : sérum pré-immun du rat.

Piste 4 : Extraits parasitaires 3D7 sauvage additionnés de 0,1 µg C-162 purifiée.

Pistes 5, 8 : Extraits parasitaires KO

Pistes 6, 9 : Extraits parasitaires 3D7 sauvage.

Pistes 7, 10 : 0,1 µg protéine C-162 purifiée.

Marqueur de masses moléculaires (50 et 30 kDa)

3D7 n'a pas donné d'amplification prouvant ainsi l'absence totale d'épisome. Un contrôle de la qualité de préparation de l'ADN a été également réalisé en utilisant un couple d'amorces spécifiques d'une région d'un gène *var* (Figure 30, Pistes 12 et 13).

Ces contrôles montrent donc que la région allant de la position 1000 à la position 2086 du gène MAL13P1.42 a bien été tronquée et remplacée par la cassette de résistance hDHFR et que par conséquent nous avons réussi à créer un parasite (3D7-KO-Pf-Int) déficient en Pf-Int.

3.2.3 Etude du taux d'expression protéique de Pf-Int

Une fois les niveaux de transcription du gène déterminés, nous nous sommes intéressés à l'étude de l'expression de la protéine Pf-Int produite par le parasite. L'observation de la protéine dans la cellule peut apporter deux types d'informations. Une première serait son taux d'expression, et deuxièmement son site d'action. L'étude de la localisation de Pf-int pourrait aussi permettre d'identifier les ADN cibles. Dans ce but, nous avons procédé à la production de plusieurs séries d'anticorps anti Pf-Int en utilisant un fragment purifié de la protéine (région allant de l'aa 162 à 490) (Anticorps d'origine Rat et Souris). Pour cet essai, nous avons également utilisé les extraits protéiques du parasite dont le gène qui code pour Pf-Int a été invalidé (souche 3D7-KO-Pf-Int).

Cependant, nos essais de Western blot (Figure 31), ne nous ont pas permis d'observer de différence entre les profils protéiques des extraits parasitaires de la souche sauvage et de la souche KO. L'absence de détection de Pf-Int par Western-blot peut être due soit à l'absence de traduction de l'ARN messager soit à sa traduction en faible quantité. Le contrôle de la transcription de l'ARN-m ainsi que le contrôle de la concentration de la protéine sont en accord avec la littérature concernant le taux d'expression régulé des protéines impliquées dans des mécanismes liés à l'ADN tel que des recombinaisons, gyrases ou des topoisomérases. Effectivement, l'expression des recombinaisons à un taux élevé, peut mener à des événements de recombinaisons aléatoires et causer des dommages à l'ADN [135].

3.2.4 Implication de Pf-Int dans le cycle parasitaire : Comparaison de cultures des parasites K.O. vs 3D7 sauvage

L'invalidation du gène codant Pf-Int peut potentiellement avoir un effet notable sur le comportement du parasite. Si Pf-Int est une protéine qui assure une fonction essentielle au cycle intra-érythrocytaire du parasite, son élimination devrait entraîner la mort du parasite. Si au contraire, sa fonction n'est pas vitale, on peut s'attendre soit à peu d'effet soit un

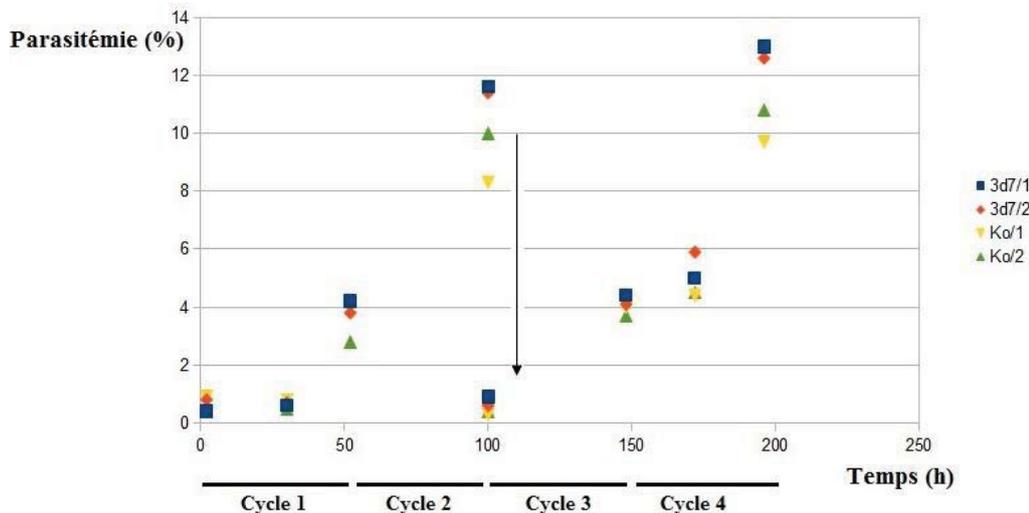


Figure 32 : Suivi de la parasitémie des cultures 3D7 et 3D7-K.O. de-Pf-Int. Le taux de la parasitémie pour les deux cultures a été mesuré au cours du temps. Les cultures ont été faites en double pour chaque souche et le suivi des cultures a été fait par observation au microscope optique après coloration Giemsa. Afin d'éviter la mort des parasites à haute parasitémie (~12%), une dilution (flèche noire) a été réalisée à la fin du 2e cycle. L'analyse montre que les deux souches ont le même temps de génération (~48 h/cycle) et des taux de ré-invasion relativement proches.

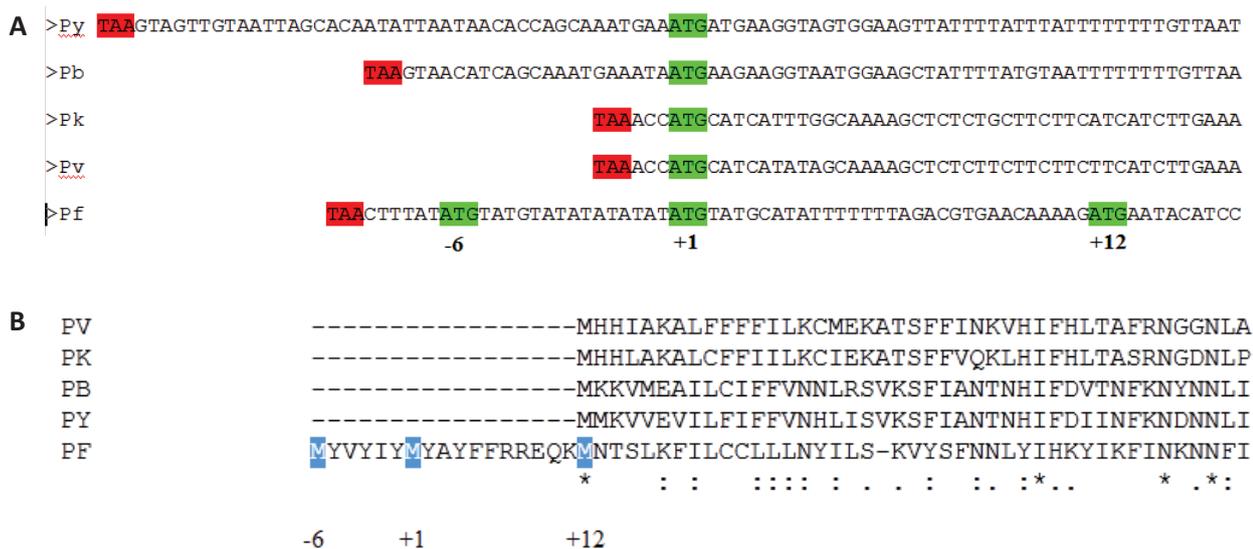


Figure 33 : A : Alignement des séquences ADN des régions 5' des gènes homologues de Pf-Int présents chez d'autres espèces de Plasmodium. Pour chaque séquence ADN, les codons d'initiation possibles sont colorés en vert et les triplets stop présents en amont en rouge. Pour l'espèce *P. falciparum* trois codons d'initiation nommés (-6), (+1) et (+12) existent dans le même cadre de lecture. Deux de ces ATG (-6 et +1) sont spécifiques à cette espèce. La traduction à partir de ces trois codons d'initiation produit des protéines dont les tailles sont 496, 490 et 478 acides aminés respectivement. B. Alignement de la région N-terminal des intégrases homologues des différentes espèces de Plasmodium. Py : *P. yoelli*, Pb : *P. berghei*, Pk : *P. knowlesi*, Pf : *P. falciparum*, Pv : *P. vivax*.

changement de phénotype visible. Pour cela, nous avons procédé à la comparaison de deux cultures synchrones des souches 3D7 et 3D7-KO-intégrase. Les cultures ont été faites en doubles et suivies par détermination du taux de parasitémie pendant quatre cycles cellulaires. Les cultures ont été diluées à la fin du deuxième cycle afin d'éviter la mort des parasites induite à haute parasitémie (~12%). L'analyse (Figure 32) montre que les deux souches ont le même temps de génération (~48 heures/cycle). Concernant le taux de multiplication déterminé par comparaison de la parasitémie des deux souches à la fin de chaque cycle (~48h, ~100h, ~150h et ~200h), les deux souches présentent des taux relativement proches avec un taux plus faible pour les parasites 3D7-KO-intégrase. Etant donné la difficulté de la ré-invasion à des taux aussi élevées de parasitémie (> 10 %) en culture classique, ces résultats doivent être confirmés par une deuxième étude. Pour cette étude, nous proposons de démarrer les cultures à des parasitémies plus faibles (~0.1 %) et d'utiliser un système de cultures en agitation qui permet d'avoir des cultures viables à des taux de parasitémies plus élevées. Dans le but de mieux caractériser ce mutant, il serait également intéressant de procéder à une analyse globale du transcriptome et d'étudier le taux de commutation des gènes impliqués dans la variabilité antigénique.

3.2.5 Surexpression de Pf-Int dans les parasites

En absence de détection de Pf-Int à partir d'extraits parasitaires natifs par nos anticorps (Résultats présentés dans la figure 31), et dans le but d'identifier les cibles ADN *in vivo*, nous avons essayé de sur-exprimer l'enzyme en fusion avec une étiquette (flag-tag) dans les parasites.

Il est important de noter que trois codons Met différents, nommés (-6), (+1) et (+12), existent dans le même cadre de lecture (Figure 33). Les premier (-6) et deuxième (+1) codons sont spécifiques à *P. falciparum* par rapport à d'autres espèces de *Plasmodia*. La traduction à partir du premier codon permettrait de produire une protéine de 496 acides aminés. Le deuxième codon (produisant potentiellement 490 acides aminés) est celui qui a été admis par PlasmoDB comme codon d'initiation. Le troisième (appelé codon (+12) et produisant potentiellement 478 acides aminés) est le codon conservé par toutes les espèces de *Plasmodium* analysées. L'intérêt de l'étude de ces trois codons d'initiation de la traduction est d'une part lié au fait qu'aucune étude n'a démontré lequel des trois est utilisé par *P. falciparum* et d'autre part que les acides aminés en N-ter ne sont pas connus car la protéine endogène n'a pas encore été purifiée. Une initiation alternative de la traduction pourrait exister et permettrait la synthèse

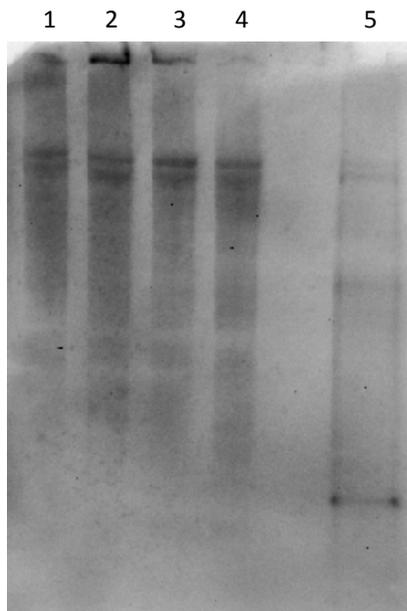


Figure 34 : Analyse de la surexpression de Pf-Int dans les parasites. Western blot en utilisant les anticorps anti-Flag. Les bandes ont été révélées par Odyssey à 700 nm.

1: Extraits protéiques du parasite sauvage.

2: Extraits protéiques du transfectant Flag -6

3: Extraits protéiques du transfectant Flag +1

4: Extraits protéiques du transfectant Flag +12

5: Contrôle positif (ALBA-Flag)

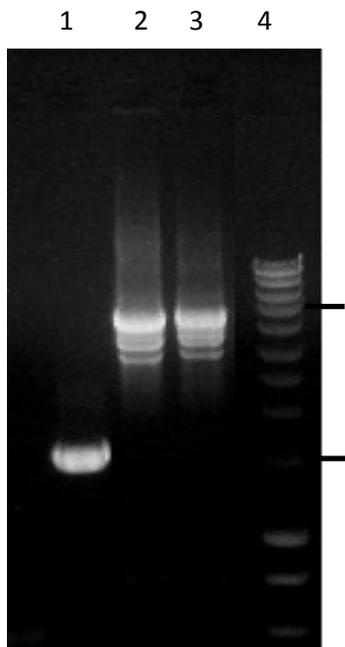


Figure 35 : Contrôle de la présence d'inserts de Pf-Int par PCR des parasites transfectés. Les extraits d'acides nucléiques des parasites transfectés ont été utilisés dans une réaction PCR. Deux couples d'amorces ont été utilisés : (i) amorces utilisées pour le clonage et (ii) amorces spécifiques au plasmide (pLN). Les produits d'amplification ont été observés aux UV après séparation sur gel d'agarose.

Piste 1 : produit d'amplification par les amorces utilisées pour le clonage, matrice : les extraits nucléaires du parasite transfecté +12. Piste 2 : Contrôle positif : produit d'amplification par les amorces pLN, matrice : le plasmide purifié utilisé pour la transfection. Piste 3 : produit d'amplification par les amorces pLN, matrice : les extraits nucléaires du parasite transfecté +12. Piste 4 : Marqueur de poids moléculaires (5 et 1,5 kilobases).

de trois formes protéiques différentes à partir d'un seul ARN messager. La différence entre ces trois protéines pourrait jouer un rôle important dans l'adressage de la protéine vers des compartiments cellulaires différents. Le décryptage du choix du codon d'initiation que fait *Plasmodium* pourrait expliquer l'absence de détection de la protéine par les analyses à partir des extraits parasitaires à cause de son implication dans un compartiment ou une structure non accessible par la technique Western blot.

Le mécanisme de double adressage a été observé chez différents eucaryotes y compris des protozoaires tel que *P. falciparum* ou *T. gondii* [136-138]. Chez *P. falciparum*, les études de Ragheb *et al.* [139] et Kehr *et al.* [137] concernant les protéines PfA-M1, une glutathion réductase et une thiorédoxine réductase de *P. falciparum* ont montré que selon le codon d'initiation de la traduction utilisé par le parasite, les protéines peuvent être adressées vers différents compartiments cellulaires.

Lors des essais de transfection, nous avons pu obtenir des parasites viables à partir des constructions de Pf-Int portant une étiquette FLAG pour les trois variantes: +1, +12 et -6. La détection dans des extraits protéiques de ces transfectants par Western blot (Figure 34) en utilisant l'anticorps anti-flag n'a pas permis d'observer les protéines surexprimées. Les bandes attendues aux masses moléculaires de 57 KDa approximativement n'ont pas pu être détectées sur le support d'analyse. Dans le but de vérifier la présence des inserts d'intérêt, nous avons procédé au contrôle des extraits des acides nucléiques de ces transfectants par PCR en utilisant comme amorces celles utilisées pour le clonage et des amorces spécifiques du plasmide utilisé (Figure 35). Ces contrôles montrent que les inserts sont présents à la taille attendue de 1,5 kb (Figure 35, piste 1), que les plasmides récupérés à partir des parasites ne sont pas tronqués (Figure 35, piste 3) et que les régions régulatrices sont présentes. En effet, la taille du produit d'amplification obtenu (4,5 kb) à partir de ces derniers (Figure 35, piste 3) correspond à la taille attendue et observée à partir de plasmide utilisé pour la transfection (Figure 35, piste 2). L'absence de protéine surexprimée peut donc être due à des mutations ponctuelles soit au niveau des régions régulatrices soit au niveau du gène lui-même provoquant un arrêt précoce de la traduction. La présence du transcrit et l'absence de mutation devront donc être vérifiés par Northern-blot et par séquençage respectivement. On peut aussi évoquer la possibilité de la dégradation de la protéine pour maintenir son niveau à un taux très faible.

Plasmodium possède trois génomes différents localisés dans trois compartiments distincts (noyau, apicoplaste et mitochondrie), par conséquent la détermination de la localisation sub-

cellulaire de Pf-Int est capitale pour la recherche de ses cibles et l'identification de son rôle. N'ayant pas pu déterminer la localisation sub-cellulaire de Pf-Int endogène avec nos anticorps et n'ayant pas pu la surexprimer en version étiquetée, nous avons procédé à la création d'une protéine chimère composée de la partie N-terminale (aa 1-190) de PF-Int fusionnée en N-terminal à la GFP. Malheureusement, cette construction s'est avérée létale et nous n'avons pas pu obtenir de parasite viable. La mort des parasites peut être due à plusieurs facteurs, notamment à l'interaction massive avec l'ADN *via* la partie N-terminale issue de Pf-Int, à la saturation des systèmes d'export des protéines ou à l'existence d'une fonction qui n'aurait pas été identifiée par les programmes de similitude et dont la surexpression conduirait à la mort des parasites.

Les études menées *in-vivo* présentées dans les paragraphes précédents ont permis de montrer que le gène identifié est fonctionnel. En effet, nous avons pu démontrer qu'il est transcrit selon un profil spécifique pendant le cycle intra-érythrocytaire, notamment au stade schizont. Malheureusement, les essais de détection de la protéine par des anticorps produits au laboratoire ne nous ont pas permis de déterminer le niveau d'expression la protéine Pf-Int. Ceci suggère soit que le niveau de traduction est maintenu très faible soit que Pf-Int est dégradée, probablement dans le but de minimiser l'impact que pourraient avoir des réactions de recombinaisons accidentelles sur la cellule si une molécule de ce type est trop abondante. En absence de détection de la protéine endogène et vu les essais de surexpression de la protéine étiquetée infructueux, nous ne pouvions pas poursuivre l'identification des cibles ADN *in-vivo* par une approche de sélection des régions retenues sur la chromatine (« Chromatine Immuno-precipitation » suivi d'hybridation sur puce, « ChIP on chip »).

Différents essais de détermination des cibles *in-vitro* ont été menés précédemment ou en parallèle utilisant la protéine recombinante produite dans le système *E. coli*. Ces étapes ont nécessité la mise au point de protocoles d'expression et de purification de la protéine, ainsi que la recherche de cibles à partir de librairies d'ADN génomiques ou synthétiques. Les différentes versions de Pf-int, identifiées à partir de l'organisation des domaines correspondant à des régions conservées, ont aussi été produites. L'ensemble de cette partie de chimie des protéines ainsi que la caractérisation physico-chimique sont décrites dans les chapitres suivantes.

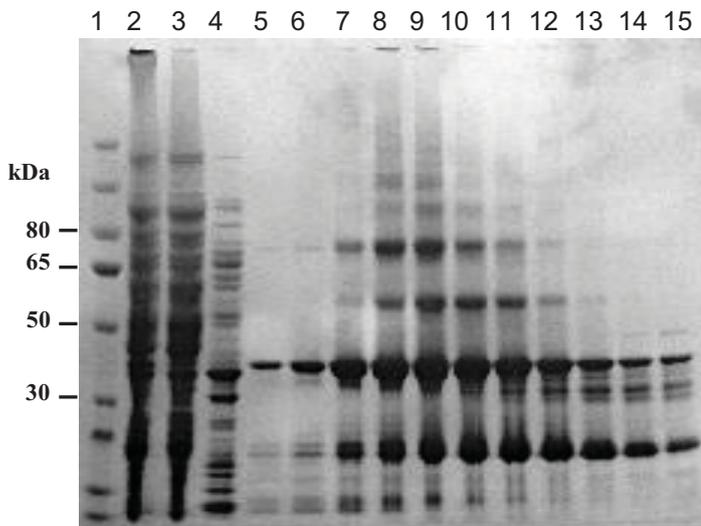


Figure 36 : Vérification sur gel SDS-PAGE de la purification du fragment C-162 de Pf-Int sur colonne d'affinité HisTrap. Piste 1: Marqueur de masses moléculaires (80, 65, 50 et 30 kDa). Piste 2: Lysat cellulaire chargé sur la colonne. Piste 3: Protéines non retenues. Piste 4: Lavage avec 25% tampon B. Pistes 5-15: Fractions d'élution. Le gel est coloré au bleu de Coomassie.

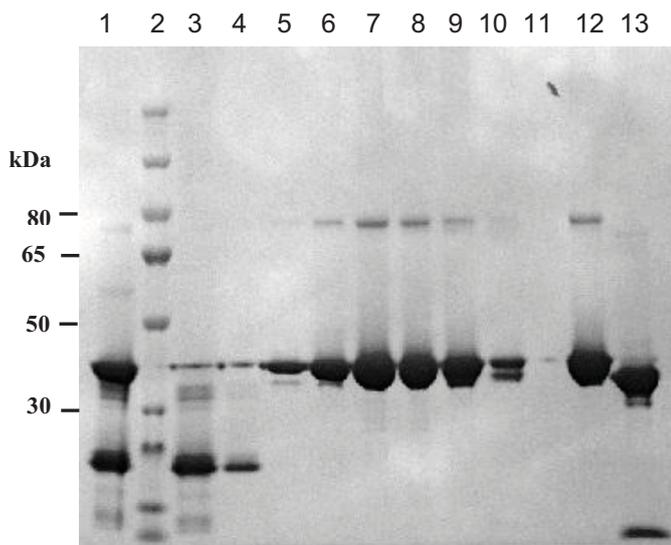


Figure 37 : Vérification sur gel SDS-PAGE de la purification du fragment C-162 de Pf-Int sur colonne d'Héparine et contrôle du clivage de l'étiquette par la protéase TEV. Piste 1: Extrait protéique chargé. Piste 2: Marqueur de masses moléculaires (80, 65, 50 et 30 kDa). Piste 3: Protéines non retenues. Piste 4: Lavage à 30% tampon B. Pistes 5-10: Fractions d'élution. Piste 12: Protéine avant coupure au TEV. Piste 13: Protéine après coupure au TEV.

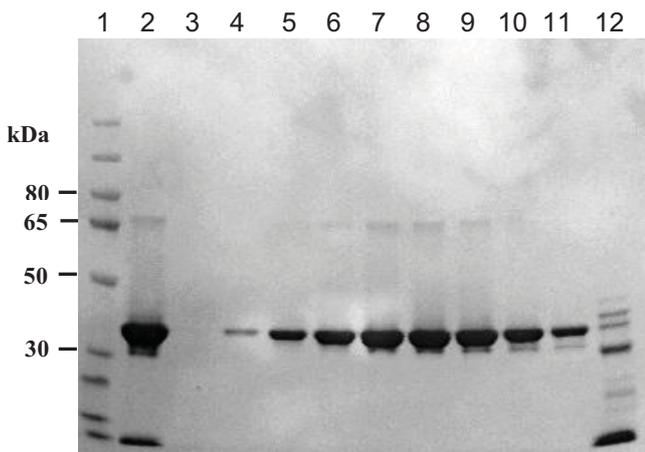


Figure 38 : Vérification sur gel SDS-PAGE de la purification du fragment C-162 de Pf-Int sur la deuxième colonne d'affinité HisTrap. Piste 1: Marqueur de masses moléculaires (80, 65, 50 et 30 kDa). Piste 2: Extrait protéique chargé. Pistes 3-12: Fractions d'élution.

3.3 Purification des protéines Pf-Int recombinantes

Dans le but d'étudier *in-vitro*, la recombinaise Pf-Int que nous avons identifié, nous avons procédé à la production de différentes formes recombinantes de la protéine. Nous avons également procédé à la recherche de cibles par deux méthodes différentes en utilisant comme appât la protéine purifiée C-162. La première étude a été réalisée par SELEX (*Systematic Evolution of Ligands by EXponential enrichment*) sur des fragments ADN synthétiques aléatoires de 25 pb. La deuxième étude a été réalisée sur des fragments d'ADN génomiques de *P. falciparum*.

Grâce aux résultats de prédiction de la structure de Pf-Int, nous avons réalisé différentes constructions pour l'expression de plusieurs fragments de l'intégrase (C-117 : allant de 117 à 490 aa, C-162 : de 162 à 490 aa, C-182 : de 182 à 490 aa et C-192 : de 192 à 490 aa). Les protéines produites dans *E. coli* BL21 DE3-RIPL sous la forme de fusion avec une étiquette 6-histidines sont solubles, et ont été purifiées par une succession d'étapes chromatographiques FPLC. Brièvement, la première étape de capture par chromatographie d'affinité sur colonne His-Trap a permis d'éliminer les débris cellulaires et les contaminants majeurs. Dans le cas de la protéine C-162, le contrôle sur gel SDS-PAGE (Figure 36) des différentes fractions issues de cette étape permet de voir que la protéine retenue sur la colonne migre à la masse moléculaire attendue (~40kDa).

Les fractions contenant les protéines de masse moléculaire attendue, issues de la première étape ont été rassemblées et utilisées dans une étape de purification intermédiaire sur colonne d'héparine sulfate. Comme on peut le constater lors du contrôle des fractions sur gel SDS-PAGE (Figure 37, pistes 1 à 10), cette étape a permis d'éliminer la quasi-totalité des contaminants car l'héparine mime le squelette phosphate de l'ADN. La protéine issue de cette étape de purification intermédiaire a été dosée (DO à 280 nm) et mélangée avec la protéase du TEV (portant également une étiquette 6-histidines) à un ratio protéine : protéase de 10:1. On peut voir que par rapport au contrôle (Figure 37, piste 12) plus de 90% de l'étiquette a été coupée (Figure 37, piste 13).

La protéine d'intérêt dont l'étiquette a été coupée, a été séparée du reste des produits issus de la réaction de coupure (protéase, l'étiquette coupée) et de la protéine de fusion non coupée, par un deuxième passage sur la colonne de chromatographie d'affinité (His-Trap), figure 38.

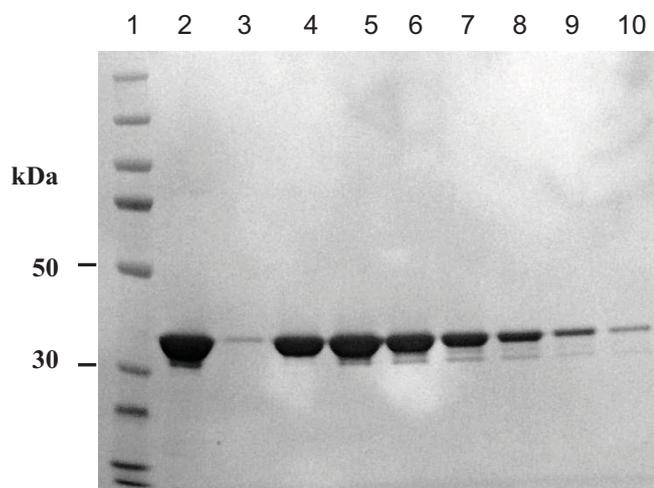


Figure 39 : Vérification sur gel SDS-PAGE de la purification du fragment C-162 de Pf-Int sur colonne d'exclusion stérique, HiLoad 16/60 Superdex 200pg. Piste 1: Marqueur de masses moléculaires (50 et 30 kDa). Piste 2: Extrait chargé. Pistes 3-10: Fractions d'élution.

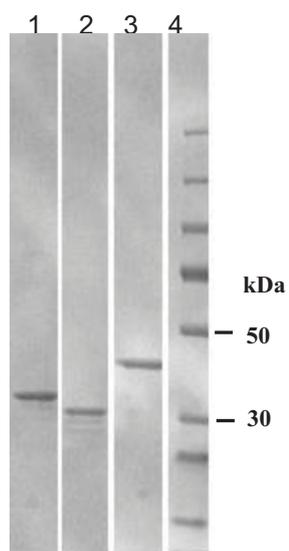


Figure 40 : Vérification sur gel SDS-PAGE de la purification des fragments C-117, C-182 et C-192 après passage sur la colonne d'exclusion stérique. Piste 1: C-182. Piste 2: C-192, Piste 3: C-117, Piste 4: Marqueur de masses moléculaires (50 et 30 kDa).

Une dernière étape de polissage a été réalisée par chromatographie sur colonne d'exclusion stérique. Cette étape a permis d'éliminer les produits résiduels et d'avoir la protéine purifiée dans le tampon de conservation (Figure 39).

La bande observée sur gel SDS-PAGE correspondant à la protéine purifiée a été découpée et le contenu a été analysé par digestion tryptique couplée à la spectrométrie de masse (MALDI-TOF). Ces fragments correspondent à plusieurs régions de la séquence primaire de la protéine Pf-int et nous ont ainsi permis de confirmer son identité. Par ailleurs, la mesure de la masse de la protéine purifiée entière (38568 amu) par MALDI-TOF nous a permis de vérifier qu'elle n'a pas subi de dégradation (masse théorique 38568.2).

Plusieurs bandes protéiques additionnelles de faible intensité ont été également observées sur les gels SDS-PAGE tout le long des étapes de purification. La bande dont la masse moléculaire est de ~70 kDa (Figure 38) a été analysée en spectrométrie de masse et la séquence identifiée comme étant celle de l'intégrase, ceci nous conduit à penser qu'il s'agit d'un dimère de l'intégrase. La présence d'un dimère observé par SDS-PAGE laisse à penser que les conditions de dénaturation pourraient avoir été insuffisantes. Quant aux autres bandes dont les masses moléculaires sont inférieures à celle de la protéine purifiée, elles semblent provenir de produits issus de la dégradation de cette dernière.

Les autres fragments de l'intégrase (C-117, C-182 et C-192) ont été purifiés selon le même protocole sauf que dans certains cas la coupure de l'étiquette 6-histidines a été omise. Dans le cas du 28aa-C192, l'extension de 28 acides aminés rajoutée dans la région N-terminale du fragment C-192 n'a pas été clivée.

La figure 40 montre que les fragments C-117, C-182 et C-192 ont été obtenus avec le même degré de pureté que le fragment C-162 et qu'ils migrent aux masses moléculaires attendues ~43, ~36 et ~34 kD. De la même manière que pour le fragment C-162, des bandes de faibles intensités et de masses moléculaires inférieures à celles des bandes majoritaires ont été observées. Ces bandes correspondent probablement à des produits de dégradation.

La masse moléculaire des différents fragments purifiés est estimée par chromatographie d'exclusion à 88, 70, 43 et 41 kDa pour C-117, C-162, C-182, C-192. Ces valeurs ne sont pas en concordance avec les masses moléculaires théoriques et observées par SDS-PAGE (43, 38, 36 et 34 kDa respectivement). Cette différence vient essentiellement du fait que la

chromatographie d'exclusion stérique permet de séparer les macromolécules en fonction de leur rayon hydrodynamique. La précision de cette technique est donc relative, dans la mesure où les formes des molécules utilisées pour la calibration et celles étudiées ne sont pas les mêmes mais sont supposées être identiques.

D'autre part, les courbes d'élution (Annexe 1) des fragments purifiés ne présentent pas une allure parfaitement en cloche (courbe de Gauss) mais on remarque qu'elles sont légèrement asymétriques. Ceci est probablement dû à l'interaction de la protéine avec la matrice de la colonne d'exclusion stérique. Cette interaction altère forcément les temps de rétention et par conséquent les estimations de masses moléculaires qui en découlent.

Il est cependant intéressant de noter que pour les deux fragments C-182 et C-192 les masses obtenues par exclusion stérique (43 et 41 kDa) sont relativement proches des valeurs théoriques (36 et 34 kDa) contrairement aux fragments C-117 et C-162 pour lesquels les masses obtenues par exclusion stérique (88 et 70 kDa) sont pratiquement le double des masses attendues (43 et 38 kDa). Cette différence importante entre les deux masses nous a conduits à penser que ces deux fragments pourraient être dimériques en solution. Dans le but d'étudier l'état d'oligomérisation du fragment C-162, nous avons procédé à son analyse par centrifugation analytique dont les résultats sont présentés dans la suite du texte. Par manque de temps, nous n'avons malheureusement pas pu analyser le fragment C-117.

Malgré la tendance de cette protéine à interagir avec la matrice de la colonne, l'étape de chromatographie par exclusion stérique reste une étape importante de purification et permet, en plus de l'élimination des contaminants résiduels, de changer efficacement le tampon de la protéine.

3.4 Analyses physico-chimiques des protéines purifiées C-117, C-162, C-182 et C-192

Une fois les différents fragments de Pf-Int purifiés, nous avons procédé à des analyses physico-chimiques dans le but de mieux les caractériser. Ces analyses ont englobé des études d'homogénéité des solutions protéiques, la détermination de l'état d'oligomérisation et une analyse initiale par résonance magnétique nucléaire (RMN) du fragment C-162. Nous avons également réalisé, par dichroïsme circulaire (CD), l'étude de la structure secondaire ainsi que de la stabilité thermique des fragments purifiés.

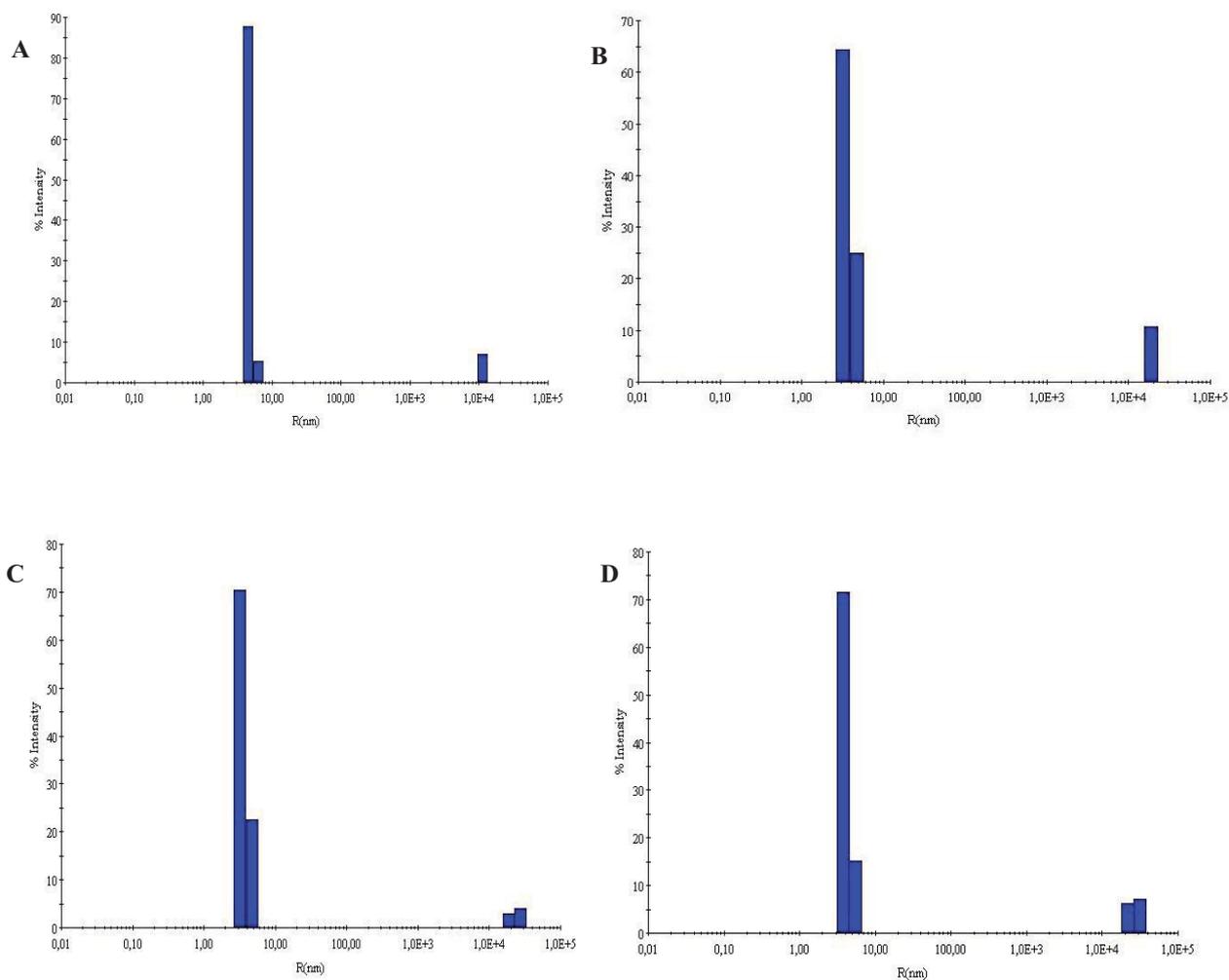


Figure 41 : Analyse de l'homogénéité protéines recombinantes par diffusion dynamique de la lumière.

A : C-117, B : C-162, C : C-192 et D : C-182. Les figures montrent l'intensité de la lumière diffusée (axe des ordonnées, % d'intensité) en fonction du rayon des molécules présentes dans la solution (axe des abscisses, R en nm). Les analyses montrent que les solutions protéiques sont majoritairement constituées d'une seule espèce ($R \sim 4$ nm) et qu'elles sont mono-disperses (<20% de polydispersité). Les analyses montrent également l'existence d'une faible quantité (<10 % de l'intensité) d'agrégats ($R > 10^4$ nm).

3.4.1 Caractérisation de l'homogénéité de Pf-Int en solution par analyse par diffusion dynamique de la lumière

La caractérisation de l'homogénéité d'une solution est une étape importante précédant d'autres études physico-chimiques. En addition de la détermination de la pureté de Pf-Int, nous avons procédé à son analyse par la diffusion dynamique de la lumière (DLS). Les expériences de diffusion de la lumière permettent en premier lieu, d'estimer le rayon hydrodynamique des espèces moléculaires présentes en solution et dans un deuxième temps de déterminer la proportion de ces espèces dans la solution étudiée. Cette analyse permet également d'estimer, en fonction de la largeur des pics obtenus, le degré de polydispersité des molécules. Ces mesures dites de polydispersités, permettent de déterminer la variété des espèces moléculaires en solution et de détecter la présence d'agrégats qui pourraient interférer avec la suite des études.

Les résultats présentés dans la figure 41 montrent que les quatre fragments purifiés ont approximativement le même profil avec deux espèces moléculaires observées. L'espèce moléculaire majoritaire (70 % d'intensité au minimum) correspond à des molécules dont les rayons hydrodynamiques ont été estimés à 4,6 , 3,6 , 3,7 et 4,1 nm pour C-117, C-162, C-182 et C-192 respectivement. Les pourcentages de polydispersité de ces espèces ont été estimés à 9,1 , 18,1 , 19,1 et 17,7 % respectivement. La deuxième espèce observée correspond à des agrégats de masses moléculaires élevées (rayon hydrodynamique $>10^4$ nm) dont l'intensité est inférieure à 10% et que nous n'avons pas pu éliminer malgré la colonne d'exclusion stérique, la centrifugation extensive et la filtration des échantillons. L'analyse par DLS des protéines purifiées a montré que les solutions protéiques sont suffisamment homogènes et présentent peu de polydispersité ($< 20\%$) et qu'elles sont donc d'une qualité raisonnable pour la suite des études.

3.4.2 Analyse de l'état d'oligomérisation de Pf-Int par centrifugation analytique

Devant la difficulté que nous avons eu pour observer des complexes du fragment C-162 avec certains ADN (résultats présentés dans les paragraphes suivants) et vu la différence de masse moléculaire observée par chromatographie d'exclusion stérique par rapport à la masse attendue, nous avons posé l'hypothèse que C-162 pourrait être un dimère en solution et que sa dimérisation gênerait la formation d'un complexe stable avec l'ADN.

Dans le but de déterminer l'état d'oligomérisation de C-162, nous l'avons analysé par

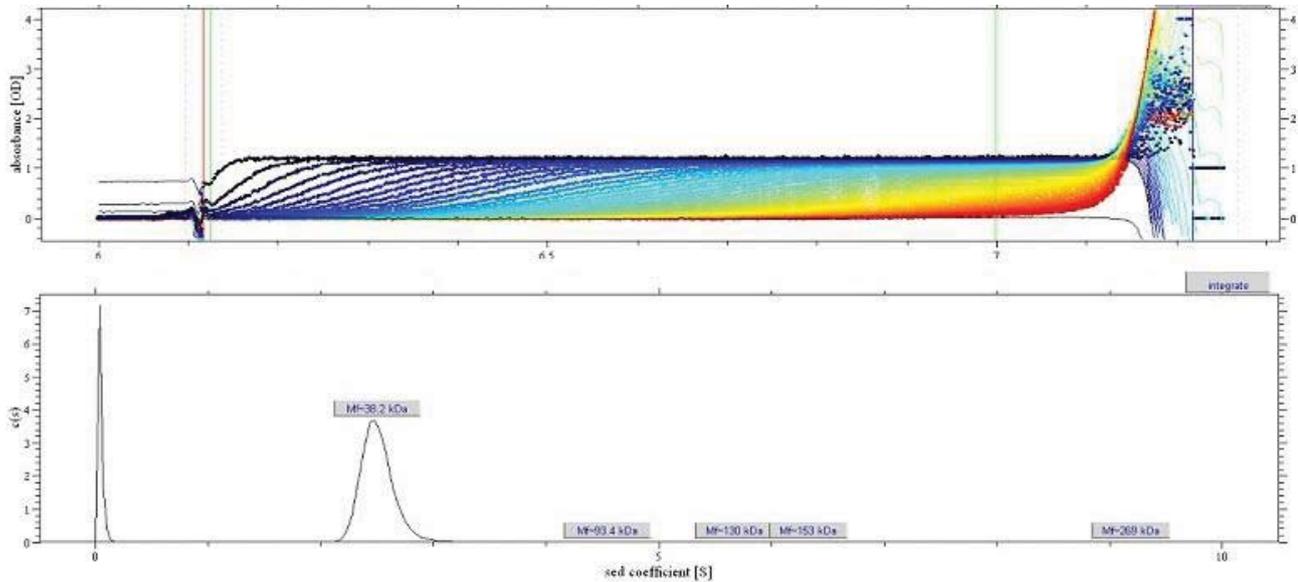


Figure 42 : Détermination de l'état d'oligomérisation de la protéine C-162 par centrifugation analytique. 1 mg/ml de protéine C-162 purifiée a été utilisé pour une analyse par centrifugation analytique en vitesse de sédimentation. A : données brutes et modélisation. B : Coefficients de sédimentation des espèces présentes en solution (Svedberg). L'analyse par centrifugation analytique montre la présence d'une seule espèce dont le coefficient de sédimentation est estimé à 2,7S. La masse moléculaire de C-162 est estimée à ~38 kDa et qu'elle est donc monomérique en solution à 1mg/ml.

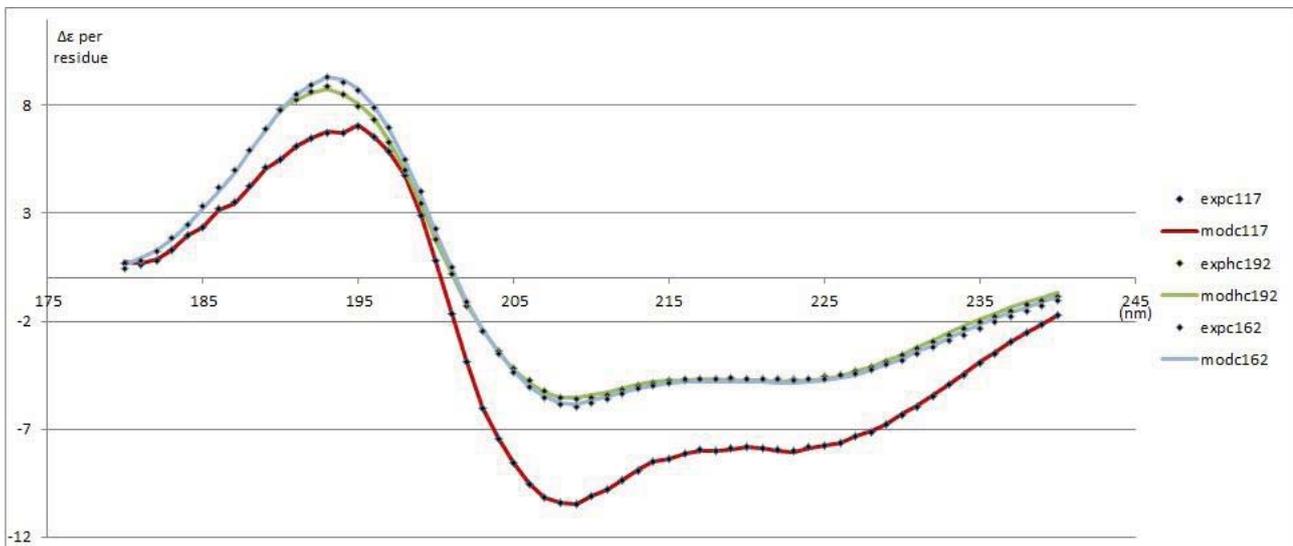


Figure 43 : Caractérisation expérimentale des structures secondaires de C-117, C-162 et C-192 par dichroïsme circulaire. Le contenu en structures secondaires est déterminé par mesure du dichroïsme en spectre UV lointain. Les données expérimentales sont représentées par des losanges. Les valeurs calculées selon le modèle proposé sont représentées par un trait plein rouge pour C-117, bleu pour C-162 et vert pour C-192. La comparaison des résultats obtenus par rapport aux bases de données par le programme CDSSTR ont permis de montrer que les trois fragments sont majoritairement composés de structures en hélice α (75%, 45% et 60% respectivement).

centrifugation analytique en vitesse de sédimentation à une concentration de 0,77 mg/ml. Cette étude a permis de montrer que C-162 est monomérique en solution (Figure 42). Par conséquent, les différences de masses observées en chromatographie sur gel d'exclusion stérique et les masses théoriques ne sont probablement pas dues à dimérisation de C-162. La présence du dimère observé sur gel SDS-PAGE au cours de la purification n'a pas été détectée.

3.4.3 Caractérisation des structures secondaires de Pf-Int : Analyse par dichroïsme circulaire

Les prédictions de structures secondaires de Pf-Int par comparaison de son contenu en acides aminés par rapport aux bases de données et d'autres recombinases de la famille Y-SSR ainsi que la prédiction de structure tertiaire (λ -int) ne valident pas pour autant son contenu en structures secondaires. Ces éléments contribuent à construire la carte d'identité de la protéine, et la probabilité qu'elle soit active en solution. Par ailleurs, la présence de structures typiques (Hélices α et feuillets β) conforte la probabilité que la protéine soit bien repliée. Dans le but d'estimer expérimentalement ce contenu des trois protéines recombinantes purifiées C-117, C-162 et C-192, ces dernières ont été soumises à des analyses de dichroïsme circulaire dans l'UV lointain (180-240 nm). Le signal de dichroïsme obtenu dans cette région du spectre dépend de l'organisation des liaisons peptidiques de la protéine. En fonction de la structure secondaire dans laquelle est impliquée la liaison peptidique (hélices α , brins β ...), on peut observer différents spectres caractéristiques. En comparant les spectres obtenus expérimentalement, normalisés par rapport au contenu en liaisons peptidiques, à des bases de données de protéines dont le contenu en structures secondaires est connu, nous pouvons avoir une estimation de la proportion en ces structures qui composent la protéine étudiée. Les résultats de ces analyses sont rassemblés dans la figure 43.

Pour chaque protéine, sont présentés les résultats expérimentaux (courbes en losanges) ainsi que les courbes de modélisations (« fitting ») théoriques générées par le programme CDSSTR représentées par un trait rouge, bleu et vert pour C-117, C-162 et C192 respectivement. Les proportions de structures secondaires résumées dans le tableau 3 montrent que les protéines adoptent une organisation plus riche en hélices α qu'en feuillets β . On remarque aussi que C-117 a un contenu en alpha significativement supérieur à celui présenté par les deux autres protéines.

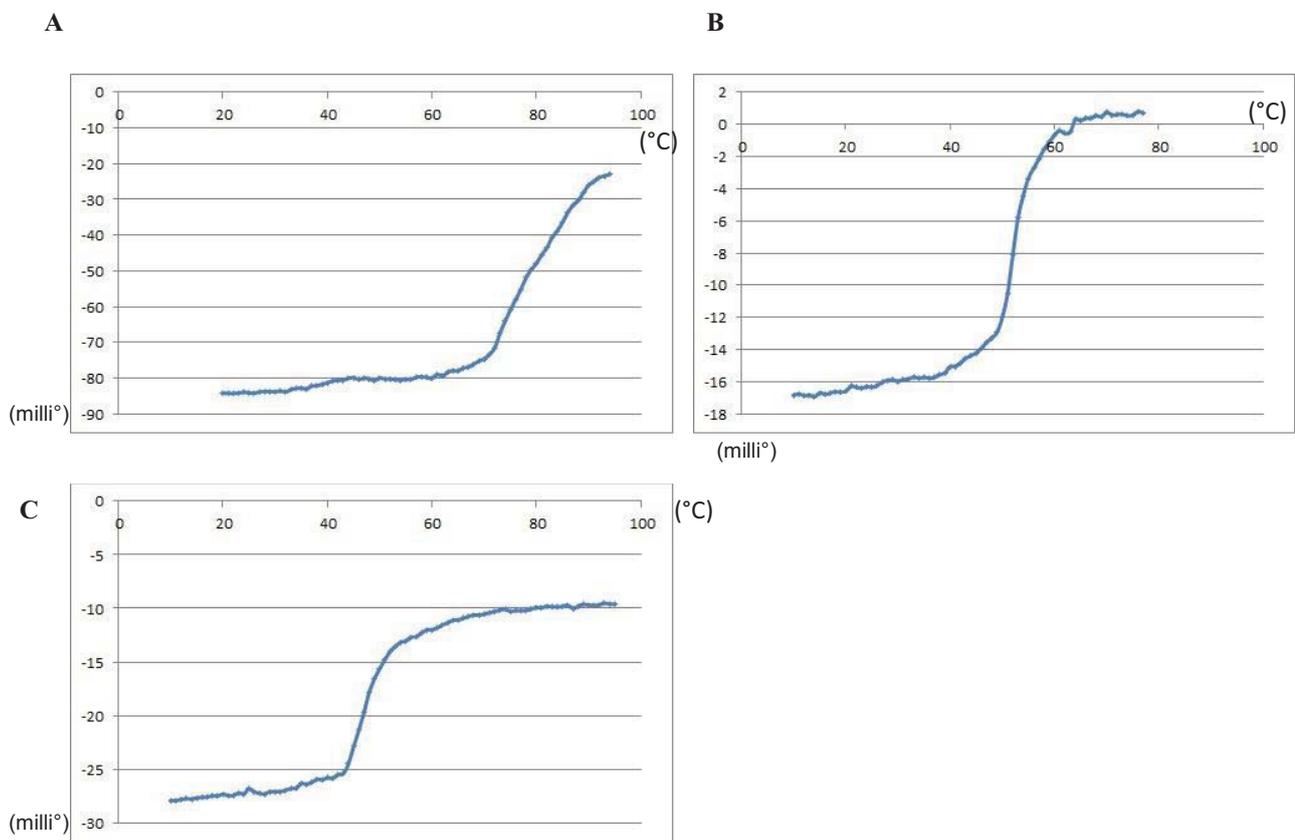


Figure 44 : Étude de la stabilité thermique des protéines C-117 (A), C-162 (B) et C-192 (C) par dichroïsme circulaire-chauffe (« thermal denaturing-CD »). Les courbes de dénaturation thermique montrent le dichroïsme à 220 nm (axe des ordonnées) en fonction de la température (axe des abscisses). Les analyses indiquent que les protéines C-162 et C-192 sont résistantes jusqu'à 50 et 40°C respectivement. Pour la protéine C-117, nous observons une réorganisation structurale de la protéine à 37°C et une dénaturation 70°C.

Le contenu riche en hélices a été également observé chez les autres membres de la famille des tyrosine-recombinases dont les structures 3D ont été déterminées (Tableau 4, lambda-Int, Cre, XerD, IntI).

Protéine \ Teneur en structures secondaires	Hélices α (%)	Feuillet β (%)	“Turn” (%)	Désordonné (%)	Total (%)
C-117	75	5	9	10	99
C-162	54	6	12	28	100
C-192	60	3	11	26	100

Tableau 3 : Contenus en structures secondaires des trois fragments protéiques de Pf-Int purifiées C-117, C-162 et C-192 déterminés par dichroïsme circulaire.

	Lambda-Int	Cre	Flp	Intégrase d'intégrons	XerD
PDB-id	1Z1B	2CRX	1FLO	2A3V	1A0P
Alpha (%)	47	53	47	50	55
Beta (%)	8	4	11	8	5

Tableau 4 : Contenus en structures secondaires de différentes recombinases à tyrosine dont les structures tertiaires ont été déterminées.

Un autre paramètre permettant d'attester de l'état natif d'une protéine et de sa stabilité est son point de fusion (T_m , « Melting temperature ») ainsi que son profil de transition vers et au delà du T_m . L'étude de la stabilité thermique par chauffage couplé au dichroïsme circulaire a également été réalisée pour les protéines C-162, C-192 et C-117. Les résultats (Figure 44, B, C) montrent que les fragments C-162 et C-192 se dénaturent à partir de 50 et 40°C respectivement. La dénaturation s'accompagnant de la précipitation de la protéine, la courbe résultante ne peut être résolue par application d'une équation de dénaturation simple à deux états. Par conséquent, la température de fusion (T_m) n'a pas pu être déterminée avec exactitude.

La même mesure pour le fragment C-117, montre que la courbe de dénaturation thermique peut être divisée en plusieurs phases. Une première transition est observée entre 30°C à 45°C.

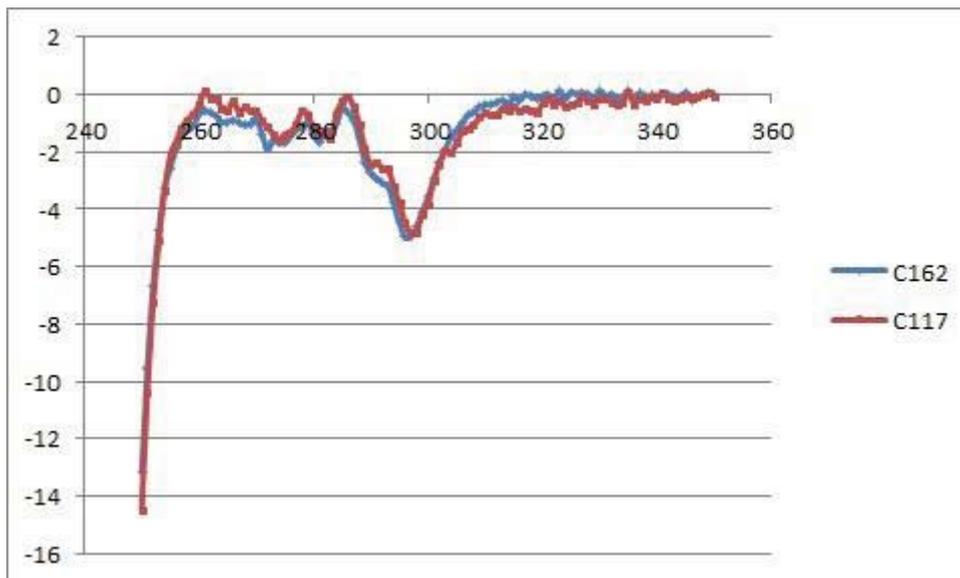


Figure 45 : Analyse des protéines C-162 (trait bleu) et C-117 (trait rouge) par dichroïsme circulaire en UV proche. L'observation d'un signal de dichroïsme dans la région UV proche indique que leurs acides aminés aromatiques des deux fragments analysés sont impliqués dans des structures tertiaires. Le signal observé est quasi-identique pour les deux fragments montrant qu'ils adoptent probablement le même repliement tertiaire.

Cette transition qui se traduit par une modification du signal global de dichroïsme révèle très certainement une réorganisation structurale de la protéine. La deuxième phase (70°C à 90°C) correspond à la phase de dénaturation de la protéine mais s'accompagne d'une précipitation importante. Ce phénomène observé également avec les autres fragments ne nous a pas permis de déterminer le T_m de la protéine C-117. On observe que la phase de dénaturation de C-117 intervient 20°C plus tard que celle de C-162. De ce fait, nous pouvons conclure que C-117 est relativement plus stable que C-162. Les trois valeurs de mesures de T_m pour C-117, C-162 et C-192 sont de l'ordre de ce qui est admis pour des protéines stables à température ambiante, 25°C. En procédant par étapes successives, nous apportons les éléments nécessaires pour estimer l'état fonctionnel de Pf-Int. Après les étapes de détermination de sa structure secondaire, nous nous sommes intéressés à l'étude de son état tridimensionnel, avec comme ultime résultat espéré, sa structure cristallographique.

L'implication des acides aminés aromatiques dans des structures tertiaires de la protéine engendre l'apparition d'un dichroïsme circulaire dans la région de l'UV proche (250-350 nm) du spectre. L'étude par spectroscopie de dichroïsme circulaire dans l'UV proche permet donc d'obtenir une information qualitative sur le repliement tertiaire de la protéine. L'observation d'un signal de dichroïsme en UV proche pour C-162 et C-117 (Figure 45) permet de montrer que les acides aminés aromatiques de ces protéines sont impliqués dans des structures tertiaires typiques de protéines repliées. Par ailleurs, les signaux très proches entre ces deux constructions de Pf-Int montrent que l'implication de ces aromatiques dans des structures tertiaires est probablement semblable pour les deux protéines. Malheureusement, les quantités très limitées de protéines purifiées obtenues pour les fragments C-192 et C-182 ne nous ont pas permis de réaliser la même étude par dichroïsme en UV proche.

Les résultats obtenus à partir des analyses biophysiques ont montré que les protéines purifiées sont obtenues sous forme homogène (<20% polydispersité), qu'une proportion importante est composée de régions repliées (60-80 % des régions adoptent une structure secondaire), que la structuration secondaire est très proche de celle adoptée par les autres membres de recombinases à tyrosine (majoritairement constituées de structures secondaires en hélices α), qu'elles sont stables et que les deux fragments C-162 et C-117 adoptent probablement des structures tertiaires très proches l'une de l'autre. Ces observations montrent que les protéines sont de qualité suffisante pour la suite des études en particulier la détermination des structures tertiaires par cristallogénèse.

3.4.4 Essais de cristallogénèse et caractérisation préliminaire de la structure 3D par résonance magnétique nucléaire (RMN)

Afin d'obtenir la structure tertiaire de Pf-Int, nous avons procédé à différents essais (différentes concentrations allant de 5 à 25 mg/ml) de cristallogénèse avec différents fragments purifiés, en particulier C-162 et C-192 et ceci en absence d'ADN. Malgré le recours à la plateforme robotisée de cristallogénèse de l'Institut Pasteur, ces essais n'ont pas été fructueux et n'ont pas permis d'avoir des cristaux de Pf-Int ou d'identifier des conditions pouvant mener à leur formation.

En complément des analyses menées plus haut et vu que les essais de cristallogénèse de différentes formes de Pf-Int n'ont pas été fructueux, nous avons entrepris d'établir une carte initiale de Pf-Int par des expériences 1D de Résonance magnétique nucléaire (RMN).

L'analyse de la protéine par RMN du proton a été réalisée à 1 mM avec différents spectres 1D enregistrés sur un spectromètre 500 MHz à 25 °C (Figure 46). La dispersion des pics de résonance et les valeurs des différents déplacements chimiques (δ) observés sur le spectre RMN-1D montrent que le fragment C-162 se comporte comme une protéine adoptant au moins partiellement un repliement tertiaire stable caractéristique des protéines globulaires. Effectivement, nous avons pu observer la présence de résonances de protons méthyles à des valeurs voisines de 0 ppm. Bien que plus larges, nous avons également identifié des résonances de protons amides à des valeurs supérieures à 8 ppm. Par ailleurs, plusieurs signaux α H à des valeurs supérieures à 5 ppm indiquent la présence de structures secondaires de type β précédemment déduits des expériences de dichroïsme circulaire menées dans l'UV lointain.

3.5 Identification des cibles ADN potentielles de Pf-Int

L'identification des ADN cibles est une étape importante pour l'analyse fonctionnelle de la recombinase Pf-Int. En effet, les tests fonctionnels nécessitent l'utilisation de ces ADN comme substrat. N'ayant pas pu utiliser la technique de « chromatine immuno-précipitation », deux approches de recherche de cibles par affinité ont été suivies pour identifier les cibles de Pf-Int. La première, réalisée par Mlle Claire Faltermeier au laboratoire, est basée sur la méthode SELEX (« Systematic Evolution of Ligands by EXponential enrichment »). Cette approche permet de s'affranchir de la limitation des séquences possibles en permettant toutes les combinaisons sur un court fragment d'ADN. Brièvement, le fragment

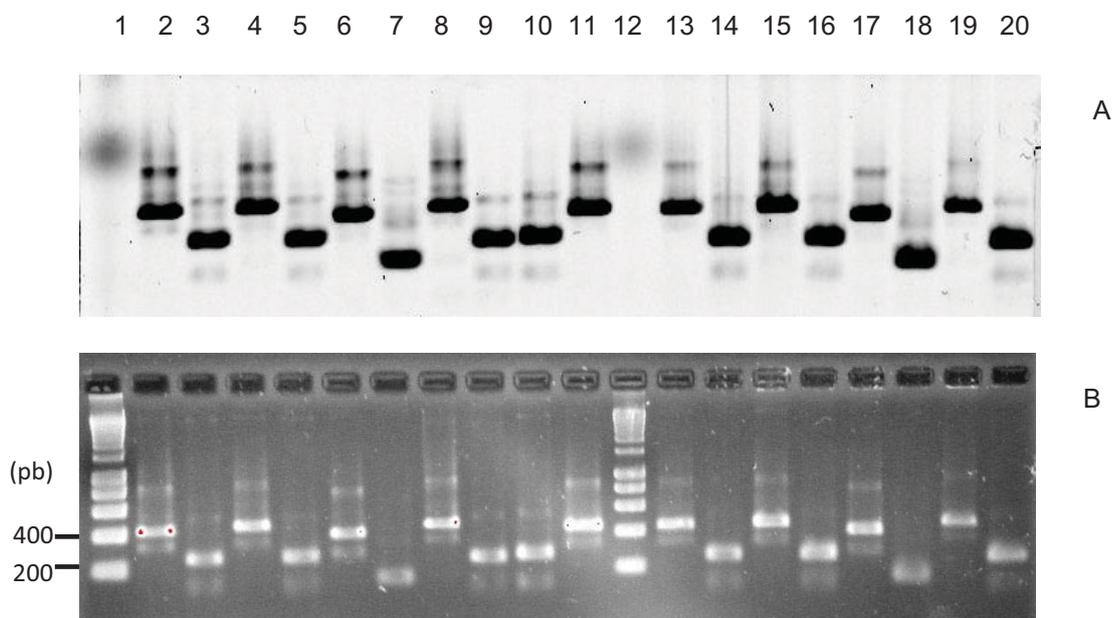


Figure 47 : Vérification de la banque de cibles génomiques clonées par amplification PCR. Les cibles génomiques clonées dans le plasmide pCR-BluntII-topo ont été amplifiées par PCR grâce aux amorces m13. Les amplicons ont été chargés sur gel d'agarose et le gel a été observé au scanner Licor Odyssey à 700 nm (Gel A) et aux UV après coloration au bromure d'éthidium (Gel B). Pistes 1 et 12 : Marqueur de tailles moléculaires (400 et 200 pb). Pistes 2-11 et 13-20 : produits d'amplification.

C-162 pourvu d'une étiquette 6-His en C-terminal a été immobilisé sur une matrice et une bibliothèque de fragments d'ADN de 25 pb a été utilisée pour l'identification de la cible ADN.

La deuxième approche (réalisée par Mlle Salma Bouizem) a permis d'identifier une deuxième série d'ADNs cibles en utilisant des fragments (200-400 pb) d'ADN génomique de *P. falciparum* comme substrat et la protéine C-162 pourvue d'un tag 6-His en N-terminal. La moitié des ADNs génomiques retenus a été utilisée pour une hybridation sur puce à ADN créée à partir de séquences du génome espacées de 200 pb (« Design » par le Laboratoire d'Artur Scherf) et l'autre moitié pour un clonage (Topo-blunt cloning) dans une librairie suivi du séquençage des inserts.

3.5.1 Résultats de la recherche de cibles par la méthode SELEX

La recherche d'ADN cible par la méthode SELEX a permis d'identifier un groupe de cinq séquences différentes de 25 pb (Tableau 5). L'alignement de ces séquences a permis d'identifier le motif conservé : (N)₆CAANC(A/C)(N)₂GT(N)₃CG(N)₇C(N)₄. La recherche de correspondance possible avec le génome de *P. falciparum* a permis de montrer qu'une partie (13 pb : CAAACACACTGTG) de la séquence Selex-8 est identique à une région d'un gène *rifin* (région 1600737..1600725 du chromosome 10).

Nom	Séquence ADN (5'→3')	Taille (pb)
Selex1	GCAACCCACAGTCCGGCCTGCAGCA	25
Selex6	GGCTCTCAACCCGGTTCGCCAGCGG	25
Selex8	ACCCAAACACACTGTGCGCGCGTCG	25
Selex14	CGCCAAACCAGGTCCGCTACGTGTG	25
Selex18	ACAACAAAGTCACGGGGCATCCAAC	25

Tableau 5 : Séquences des cinq cibles ADN identifiées par la méthode SELEX.

3.5.2 Résultats de la recherche de cibles à partir de l'ADN génomique

Une fois que les fragments d'ADN génomique retenus par affinité ont été récupérés, la moitié a servi pour la constitution d'une librairie de plasmides contenant les cibles potentielles. La présence d'inserts dans les plasmides de la librairie a été contrôlée par amplification PCR. La figure 47 montre le contrôle sur gel d'agarose des produits

d'amplification. On observe que dans certains cas, nous avons obtenus des amplicons dont la taille est de ~250 pb (Figure 47, pistes 3, 5, 9, 10, 14, 16 et 20). Ces amplicons sont issus de plasmides recircularisés et qui ne contiennent pas d'inserts. Dans certains cas, les amplicons ont des tailles inférieures à 200 pb (Pistes 7 et 18). Le séquençage de ces amplicons a montré que les plasmides ont été recircularisés sans insert et qu'une partie du plasmide lui-même a été perdue. Ceci est lié à de possibles réactions aléatoires de la topoisomérase présente en bout de chaîne du plasmide d'accueil utilisé et qui n'a pas de spécificité de séquence. Les amplicons dont la taille est supérieure à 250 pb ont été considérés comme positifs et les inserts ont été séquencés.

Au total, 58 séquences cibles potentielles ont été ainsi identifiées. L'addition de ces séquences à celles identifiées à partir de l'analyse par puce à ADN a permis de constituer un groupe de 149 séquences ADN qui constituent des cibles potentielles de Pf-Int (Annexe 2). Les séquences identifiées appartiennent à un groupe hétérogène de gènes et de régions génomiques. On retrouve ainsi des séquences issues des gènes fonctionnels et tronqués *var*, *stevor* et *rifin* ou issues des régions intergéniques proches de ces derniers (37 % des séquences). Parmi les 49 séquences issues des familles multigéniques, 46 correspondent à des régions des exons qui codent pour la partie variable extracellulaire des protéines correspondantes.

L'analyse de ces séquences par le programme MEME, le programme IRF, le programme eInverted (EMBOSS) et le programme Clustal n'a pas permis d'identifier de motif conservé qui pourrait correspondre à un site potentiel de recombinaison. Dans l'éventualité que les motifs puissent être éloignés les uns des autres, et pourraient se reconstituer en un site cible par repliement, comme c'est le cas pour les intégrases d'intégrons, l'analyse 2D des séquences identifiées a été réalisée. L'analyse du repliement des ADN par le programme mFOLD a permis de constater que ces ADN sont capables de former des structures secondaires de type tige-boucle grâce à la présence de plusieurs séquences répétées inversées de petite taille (Annexe 3). Cependant, aucun élément de structure secondaire ou repliement tige-boucle conservé n'a pu être identifié.

Une fois ces cibles ADN identifiées, nous avons procédé à des essais de caractérisation de l'interaction entre Pf-Int et ses cibles potentielles. Cette caractérisation peut se faire à deux niveaux. Le premier permet de montrer une interaction entre la protéine et l'ADN qui se traduit par un retard de migration sur un gel en conditions natives en présence de la protéine purifiée et le second permet de démontrer que cette interaction est spécifique.

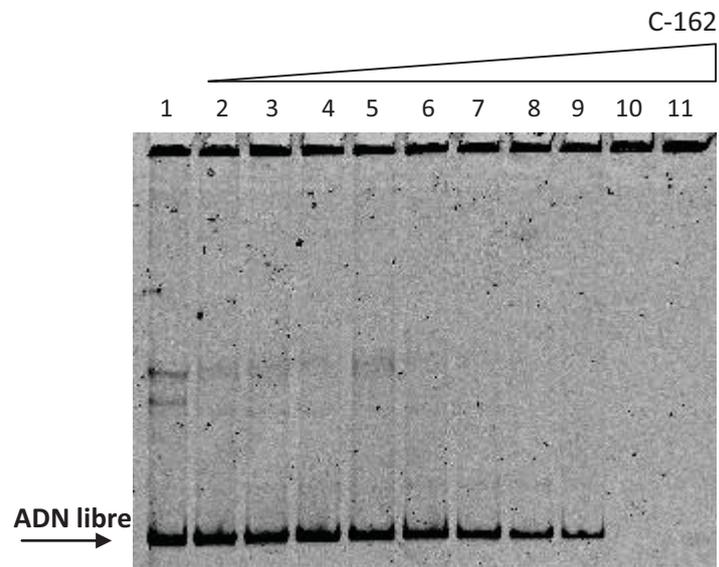


Figure 48 : Étude de l'interaction entre C-162 et un ADN cible génomique par des expériences de retard sur gel natif (EMSA). 10 nM d'ADN marqué à l'IRD 700 ont été mélangés avec des concentrations croissantes de protéine C-162. Les mélanges ont été séparés sur gel de polyacrylamide natif 5%. Piste 1: ADN libre. Piste 2: ratio = 50/1 (protéine/ADN), Piste 11: ratio =135/1. La figure montre que l'ajout progressif de C-162 purifié génère la disparition graduelle (Pistes 2 à 9) du signal de l'ADN sans permettre l'observation d'une bande d'ADN retardé. La disparition totale du signal de l'ADN a lieu à un ratio 125 :1 (Piste 10).

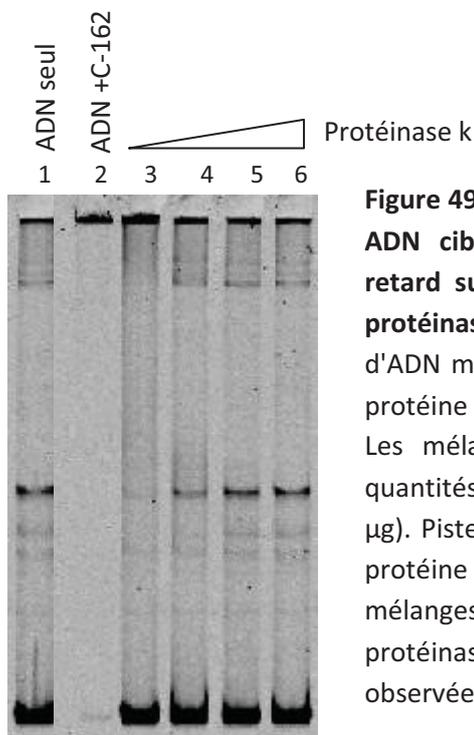


Figure 49 : Étude de l'interaction entre C-162 et un ADN cible génomique par des expériences de retard sur gel natif couplé à la digestion par la protéinase k. Gel polyacrylamide natif 5%. 10 nM d'ADN marqué (IRD 700) ont été mélangés avec la protéine C-162 (ratio : 200/1) pendant 20 minutes. Les mélanges ont été ensuite digérés par des quantités croissantes de protéinase k (0,5 µg à 2 µg). Piste 1 : 10 nM ADN. Piste 2-6 : 10 nM ADN + protéine C-162, ratio: 200/1. Piste 3-6: les mélanges protéine/ ADN ont été digérés par la protéinase K (0,5 µg à 2 µg). Les bandes ont été observées par Odyssey à 700 nm.

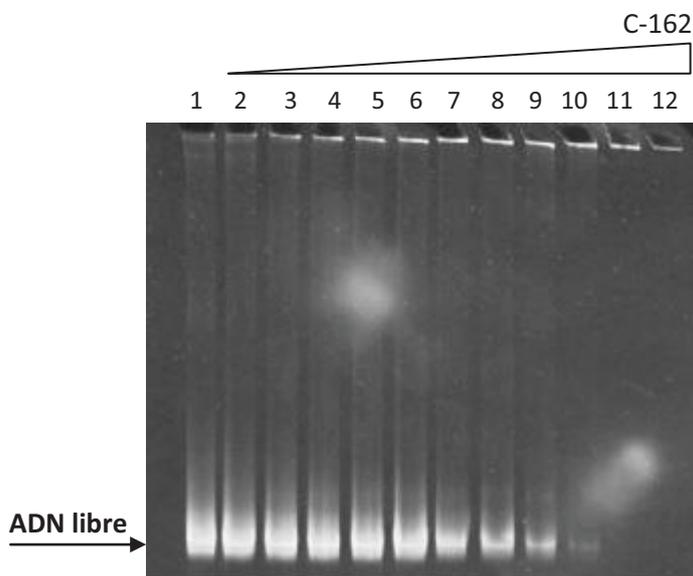


Figure 50 : Étude de l'interaction entre C-162 et un ADN cible génomique non marqué. Gel natif polyacrylamide 5%. 20 nM de produit PCR non marqué ont été mélangés avec des quantités croissantes de C-162. Le gel a été coloré au BET et observé aux UV. Piste 1 : 20 nM ADN. Pistes 2-12 : ADN additionné de quantités croissantes de protéine purifiée (Ratio protéine/ADN : de 2/1 à 50/1). L'analyse montre que la formation des complexes de masses moléculaire élevées retenus au niveau des puits du gel n'est pas due à la présence du marquage.

3.5.2.1 Étude de l'interaction de la protéine C-162/cibles génomiques par retard sur gel

Dans le but de caractériser, *in vitro*, l'interaction de la protéine C-162 avec les ADN génomiques sélectionnés, nous avons procédé à des essais de retard sur gel natif. L'ajout de quantités croissantes de protéine à une quantité fixe d'ADN provoque une disparition graduelle du signal émis par la sonde ADN libre (Figure 48, Pistes 2-9). La disparition complète du signal a lieu à un ratio protéine:ADN de $\sim 125:1$ (Figure 48, Piste 10) et nous n'avons pas pu observer de bande intermédiaire.

La disparition du signal n'est pas due à la dégradation de l'ADN ou la dégradation du marquage fluorescent puisqu'en digérant le complexe de haut poids moléculaire obtenu avec la protéinase K nous avons pu récupérer la totalité du signal de départ (Figure 49, Piste 6). La disparition du signal semble être due à la formation de complexes ADN-protéine de masses moléculaires élevées incapables de pénétrer dans le gel. La formation de ces espèces de masses moléculaires élevées n'est pas due à la présence du marquage IRD700 en extrémité des ADN puisqu'en utilisant des ADN non marqués et en colorant le gel au bromure d'éthidium (Figure 50), nous avons observé le même phénomène. Cependant, il est à noter que dans le but d'éviter le bruit de fond généré par la présence du poly (dI-dC), cette expérience a été réalisée en l'absence de ce dernier et que le ratio auquel nous avons observé une disparition complète du signal est de 40:1 (protéine:ADN).

Les essais de validation des cibles génomiques par les expériences présentées plus haut ont montré que contrairement à ce qui était attendu, nous n'avons pas pu isoler de complexes simples de faibles stœchiométries, capables de former des bandes d'ADN retardé bien définies. Cependant, ces expériences ont montré que C-162 est capable d'interagir avec l'ADN en formant des complexes de haut poids moléculaire. Ces complexes sont probablement formés grâce à des interactions aléatoires entre la protéine dont la charge est particulièrement élevée ($pI_{C-162} : 9,88$) et les ADNs utilisés.

Devant l'impossibilité d'isoler des complexes retardés avec les ADN génomiques, nous avons décidé d'utiliser des ADN plus petits (ADN SELEX), espérant ainsi réduire les interactions aléatoires.

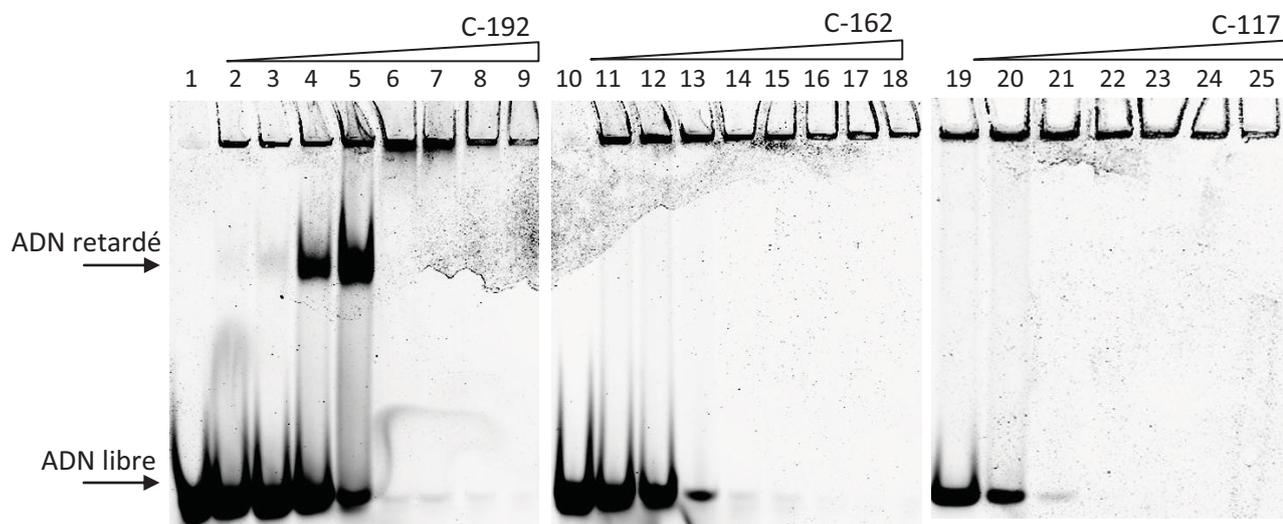


Figure 51 : Étude de l'interaction, par retard sur gel, entre les différents fragments purifiés de l'intégrase et l'ADN Selex-8 (25 pb). Gels natifs polyacrylamide 5%. 10 nM d'ADN marqué (Dy682) ont été mélangés avec des concentrations croissantes de la protéine (ratios protéine:ADN de 50:1 à 225:1). Les bandes ont été révélées par Odyssey à 700 nm. Pistes 1, 10 et 19 : ADN seul. Pistes 2-9 : fragment C-192. Pistes 11-18 : fragment C-162. Pistes 19-25 : fragment C-117.

3.5.3 Étude des cibles identifiées par la méthode Systematic Evolution of Ligands by EXponential enrichment (SELEX)

3.5.3.1 Étude de l'interaction intégrase/Séquence Selex par gel retard

La deuxième approche que nous avons utilisée, dans le but de caractériser l'interaction *in vitro* de l'intégrase avec l'ADN, a consisté en des expériences de retard sur gel natif utilisant différents fragments de la protéine et les ADN cibles sélectionnés par l'expérience SELEX. Les ADN ont été synthétisés en simple brin et hybridés au laboratoire. Ces expériences (Figure 51, pistes 11 à 18) montrent que de la même manière qu'avec les cibles génomiques, l'addition de quantités croissantes de la protéine C-162 à une quantité fixe d'ADN selex8 provoque la disparition graduelle du signal de l'ADN libre avec formation de complexes de haut poids moléculaire retenus au niveau des puits. Aucun complexe intermédiaire n'a été observé. La disparition totale du signal émis par l'ADN libre a été observée à un ratio ~175:1 (protéine:ADN).

Devant l'impossibilité d'observer des bandes d'ADN retardés, même en présence de fragments de petites tailles (25 pb), nous avons décidé d'utiliser différents fragments de la protéine : d'abord le fragment C-192 et ensuite les fragments C-117 et C-182. L'utilisation de C-192 a permis de mettre en évidence une interaction entre ce fragment de Pf-Int et l'ADN SELEX-8 (Figure 51, pistes 2 à 9). En effet, en présence de quantités croissantes de ce fragment, nous avons pu observer l'apparition graduelle d'un complexe ADN-protéine migrant à une vitesse plus faible que celle de l'ADN libre, sans s'agréger au niveau des puits. Cette interaction a été détectée pour des ratios allant de 50:1 à 125:1. Au-delà du ratio 125:1, nous avons observé la disparition complète du signal issu de l'ADN libre ainsi que celui issu de la bande d'ADN retardé avec une rétention de la totalité de l'ADN au niveau des puits.

Il semblerait donc, qu'on soit en présence de deux types d'interactions. Le premier type d'interaction est assuré par des interactions régulières, probablement *via* un site dédié à la fixation de l'ADN, menant à la formation d'une seule espèce stable à la stœchiométrie simple capable de former une bande retardée définie sur le gel natif. Le deuxième type d'interaction est le même que celui précédemment observé dans le cas du fragment C-162, cette interaction est de type aléatoire conduisant à la formation d'un agrégat composé potentiellement d'un mélange d'ADN seul, de protéine seule et de complexes ADN-protéine. Ces agrégats peuvent être produits par les charges positives localisées à la surface de la protéine et peuvent aussi être dus à l'exposition de surfaces hydrophobes de la protéine lors de son interaction avec de

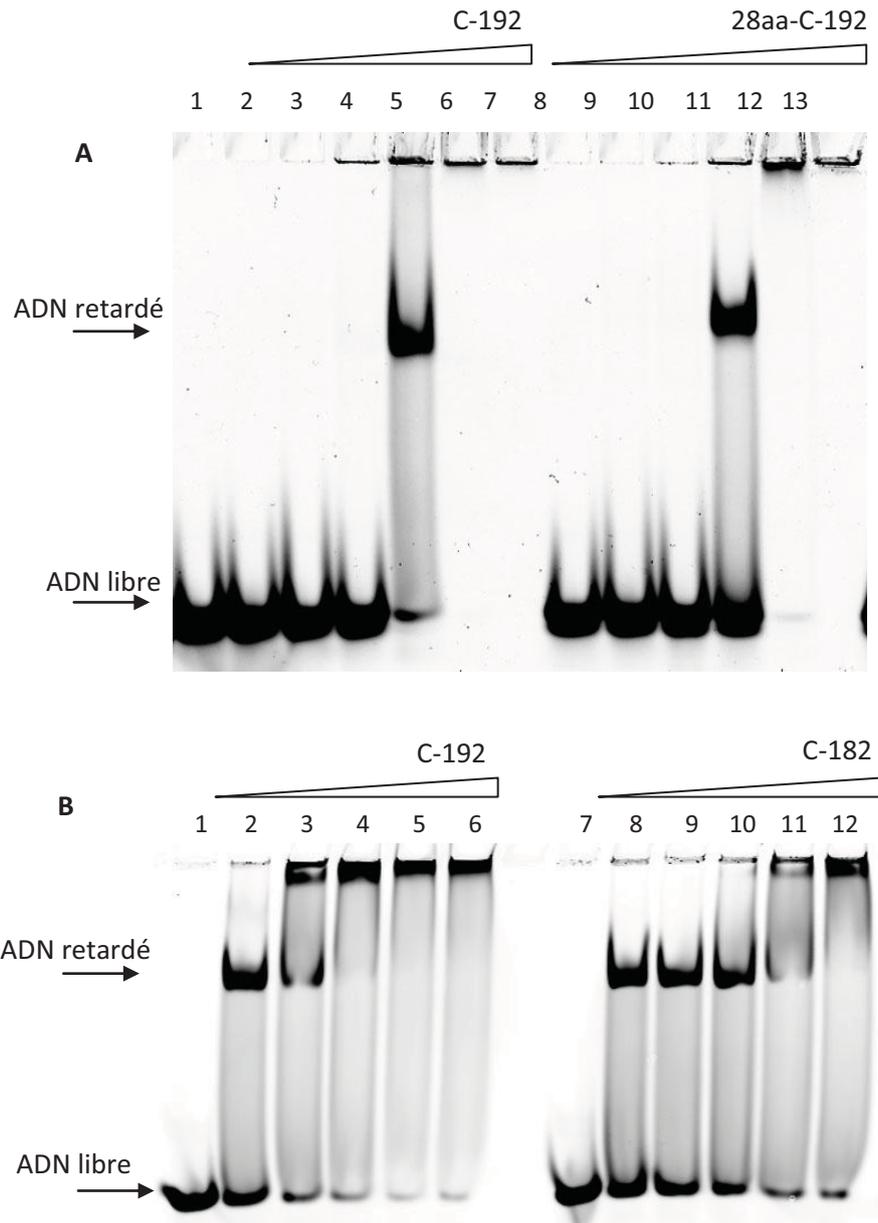


Figure 52 : Étude, par retard sur gel, de l'interaction entre les protéines 28aa-C192, C192, C182 et l'ADN Selex-8. Gel natif polyacrylamide 5%. 10 nM d'ADN marqué (Dy682) ont été mélangés avec des concentrations croissantes de la protéine. Les bandes ont été révélées par Odyssey à 700 nm. **Gel A :** ratios protéine/ADN : 2/1 à 500/1. Piste 1 : ADN seul. Pistes 2 à 7 : ADN additionné de C192. Pistes 8 à 13 : ADN additionné de 28aa-C192. **Gel B :** ratios protéine/ADN : 100/1 à 200/1. Piste 1 et 7 : ADN seul. Pistes 2-6 : fragment C-192. Pistes 8-12 : fragment C-182.

l'ADN.

Concernant le fragment C-117, ce dernier se comporte de la même manière que le fragment C-162 et aucune bande retardée bien définie n'a pu être observée dans les conditions classiques (Figure 51, pistes 19 à 25). Il est donc possible que la présence d'un peptide (Met162-Lys191 ou Asp117-Lys191) en addition à C-192 gênerait l'accès de l'ADN au site de fixation.

Dans le but de mieux caractériser l'implication de ces peptides dans la fixation de l'ADN par Pf-Int, nous avons décidé de créer une protéine de fusion 28aa-C192 équivalente en taille à C-162 mais pour laquelle la région Met162-Lys191 a été remplacée par une autre séquence protéique de 28 acides aminés différente de la séquence d'origine (Tableau 6).

	Séquence en acides aminés (N-terminal → C-terminal)
Met162-Lys191 (30 aa)	MKCKTCKKLLKPRVKFCVYCGTNVSVEKIK
Séquence des 28 aa utilisés en fusion 28aa-C192	MNKEITSGSGHHHHHSAGRSENLYFQG

Tableau 6 : Séquences en acides aminés du peptide Met162-Leu191 et du peptide 28aa ajouté en N-terminal de C-192 (28aa-C192).

Les analyses par gel retard (Figure 52, gel A) montrent que 28aa-C192 est capable de fixer les ADN selex8 de la même manière que C-192. Ceci montre que l'absence de bandes d'ADN retardé dans le cas de C-162 n'est pas simplement due à la présence physique du peptide Met162-Lys191 mais à sa composition.

Dans le but d'affiner l'impact qu'a le peptide Met162-Lys191 sur l'interaction des protéines avec l'ADN, nous avons décidé de produire le fragment intermédiaire C-182. Les expériences de retard sur gel montrent que ce dernier forme un complexe bien défini protéine-ADN que nous avons pu observer (Figure 52, gel B). Ce complexe s'établit pour un ratio allant de 100:1 à 150:1. Nous avons également remarqué que la formation de complexes retenus au niveau des puits n'a lieu qu'à partir du ratio 175:1 pour C-182 contrairement au C-192 en présence duquel nous observons la disparition complète d'ADN de la bande retardée à un ratio de 150:1. Le fragment C-182 a donc tendance à former des complexes plus stables et moins propices à la formation d'agrégats. D'après cet essai, il semblerait donc que c'est la région Met162-Cys181 qui inhibe la fixation de C-162 à l'ADN.

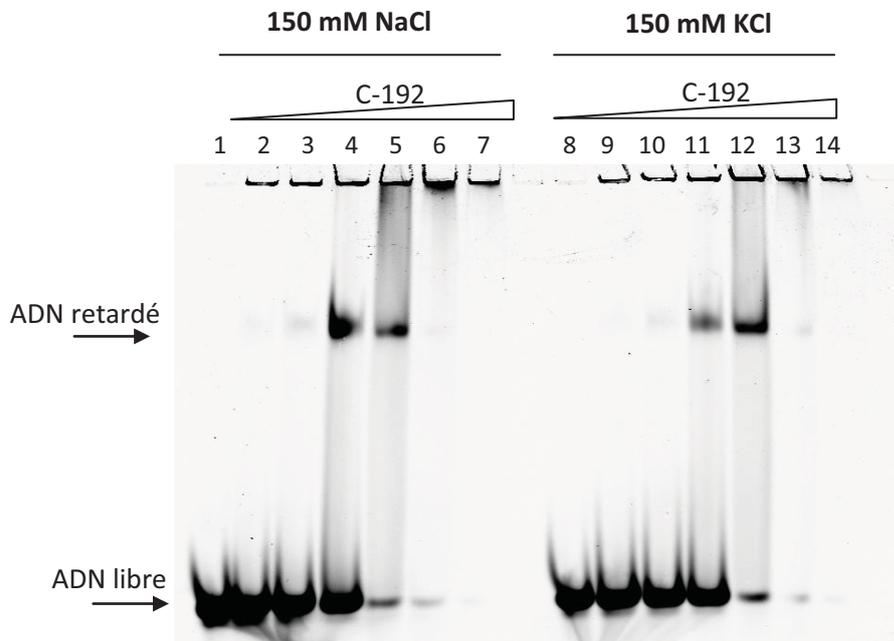


Figure 53 : Étude de l'effet de la nature du sel (KCl ou NaCl) sur l'interaction entre C-192 et l'ADN Selex-8. Gel natif 5% polyacrylamide. 10 nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de la protéine C-192 (ratio : 50/1 à 175/1) en présence de 150 mM NaCl (Pistes 1 à 7) ou en présence de 150 mM KCl (Pistes 8 à 14). Les bandes ont été révélées par Odyssey à 700 nm.

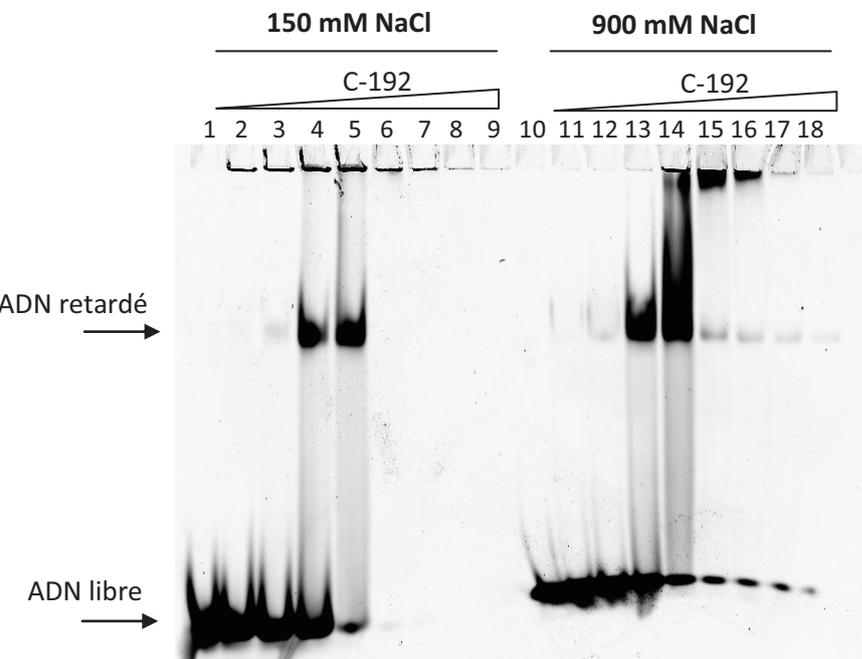


Figure 54 : Étude de l'effet de la concentration du sel NaCl sur l'interaction entre C-192 et l'ADN Selex-8. Gel natif polyacrylamide 5%. 10 nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de la protéine C-192 (ratio : 50/1 à 225/1) en présence de 150 mM NaCl (Pistes 1 à 9) ou en présence de 900 mM NaCl (Pistes 10 à 18). Les bandes ont été révélées par Odyssey à 700 nm.

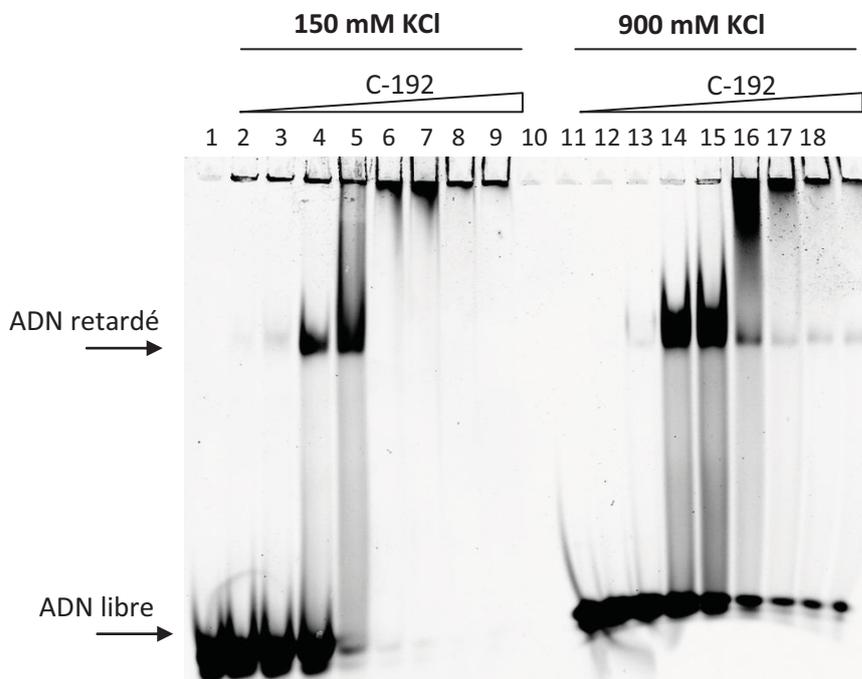


Figure 55 : Étude de l'effet de la concentration du KCl sur l'interaction entre C-192 et l'ADN Selex-8. Gel natif polyacrylamide 5%. 10 nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de la protéine C-192 (ratio : 50/1 à 225/1) en présence de 150 mM KCl (Pistes 1 à 9) ou en présence de 900 mM KCl (Pistes 10 à 18). Les bandes ont été révélées par Odyssey à 700 nm.

Ces premières analyses par retard sur gel natif nous ont permis de montrer, pour la première fois, la formation de complexes stables entre Pf-Int et un ADN cible (Selex-8). La formation de ces complexes dépend de deux phénomènes, le premier est dû à l'interaction aléatoire de la protéine avec l'ADN grâce aux charges réparties à sa surface ou à l'exposition de surfaces hydrophobes révélées lors du contact avec l'ADN alors que le deuxième est assuré par des interactions régulières probablement *via* un site dédié à la fixation de l'ADN. L'observation des complexes dépend particulièrement du fragment protéique utilisé, la région N-terminale de la protéine en amont de la position Cys181 semble jouer un rôle important dans la régulation de l'activité en inhibant l'accès de l'ADN au site réactionnel. Il est intéressant de considérer les modèles d'interactions potentielles découlant de ces observations. Une possibilité serait que la protéine fixe l'ADN dans une première étape de façon aléatoire, uniquement basée sur des interactions ioniques de surface. Cette étape serait suivie par une interaction plus spécifique via son site dédié. Il est aussi possible que des changements conformationnels de la protéine soient nécessaires afin qu'elle puisse se fixer à l'ADN

3.5.3.2 Étude de l'effet de la variation des conditions réactionnelles

De par leur nature de polymères chargés, on considère que les interactions protéine-ADN sont dépendantes de la force ionique qui les entoure. Par ailleurs, la reconnaissance finale de sites est potentiellement dépendante de changements conformationnels. L'interaction entre le fragment C-192 et l'ADN Selex-8 a ainsi été caractérisée, en particulier par l'analyse de l'impact que pourrait avoir la variation de différents facteurs tels que la force ionique, la présence de compétiteurs non-spécifique, la concentration d'ADN substrat, la présence d'agent réducteur et la concentration de détergent. Dans la suite des expérimentations, nous avons essayé de déterminer l'impact de ces facteurs sur le phénomène de formation d'agrégats et sur l'apparition de la bande d'ADN retardé.

➤ Effets de la force ionique :

Ainsi, nous avons observé que la formation du complexe ne dépend pas de la nature du sel mono valent utilisé dans la réaction (NaCl vs KCl, Figure 53), souvent présent à l'interface protéine ADN. Les bandes d'ADN retardé apparaissent à des ratios relativement identiques de C-192/Selex8.

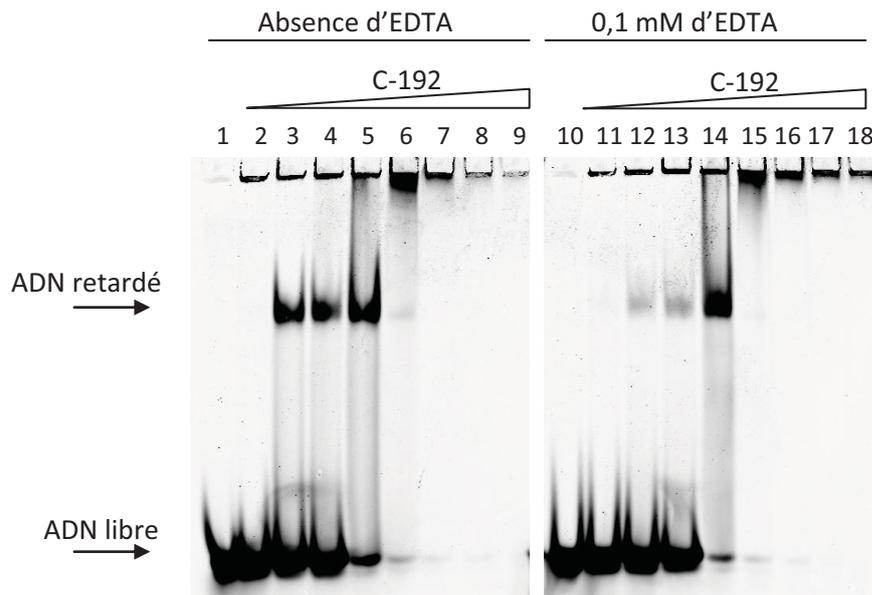


Figure 56 : Étude de l'effet de la concentration de l'EDTA sur l'interaction entre C-192 et l'ADN Selex-8. Gel natif polyacrylamide 5%-TE. 10 nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de la protéine C-192 (ratio :50/1 à 225/1) en absence d'EDTA (Pistes 1 à 9) ou en présence de 0,1 mM EDTA (Pistes 10 à 18). Les bandes ont été révélées par Odyssey à 700 nm.

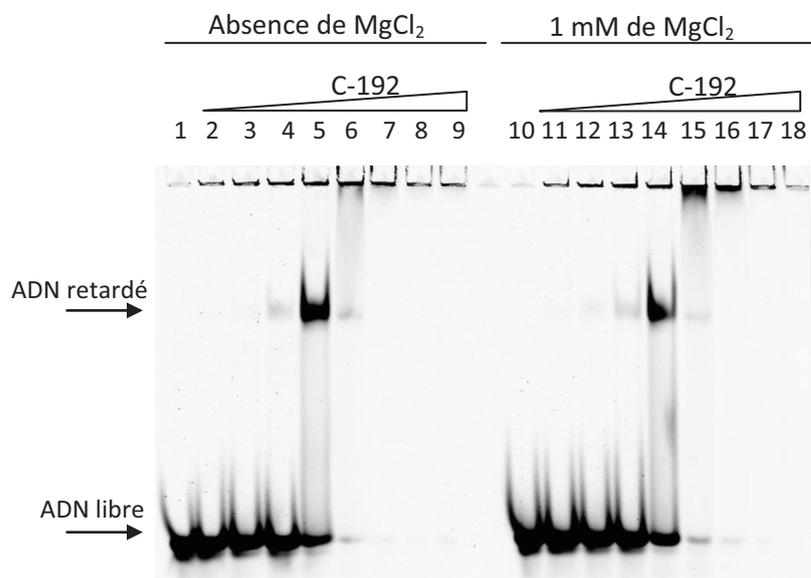


Figure 57: Étude de l'effet de la concentration de MgCl₂ sur l'interaction entre C-192 et l'ADN Selex-8. Gel natif polyacrylamide 5%-TE. 10nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de la protéine C-192 (ratio : 50/1 à 225/1) en absence de MgCl₂ (Pistes 1 à 9) ou en présence de 1 mM MgCl₂ (Pistes 10 à 18). Les bandes ont été révélées par Odyssey à 700 nm.

Nous avons également observé que le complexe peut être formé en présence d'une forte concentration en sel (900 mM NaCl, figure 54 et 900 mM KCl, figure 55). Ces concentrations élevées favorisent la non incorporation de l'ADN libre dans l'agrégat en haut des puits même à des ratios élevés (exemples : ratio protéine: ADN : 200/1 dans la piste 18 de la figure 54 et piste 18 de la figure 55).

Dans la majorité des cas, des concentrations aussi élevées en sel auraient pour effet l'inhibition de la fixation de l'ADN mais ce n'est pas le cas pour Pf-Int. Ceci montre que les interactions de type électrostatique ne sont que faiblement impliquées dans l'interface ADN/protéine au niveau du site de fixation.

Par ailleurs, en comparant l'intensité des bandes d'ADN retardé entre les pistes ayant un ratio ADN/protéine identique mais dont les concentrations en sel sont différentes (Figure 54, pistes 4, 5 et 13, 14 ; Figure 55, pistes 4, 5 et 14, 15), on remarque que l'intensité du signal est plus importante en présence de 900 mM NaCl ou KCl. Il semble donc que la présence d'une concentration élevée en sel favorise l'apparition d'une quantité plus importante de complexes retardés, cet effet est probablement dû à la neutralisation des charges électrostatiques présentes à la surface de la protéine et de l'ADN limitant ainsi la participation des complexes au phénomène responsable de la formation d'agrégats.

Nous avons également observé que l'absence d'EDTA favorise l'apparition de complexe retardé à des ratios plus faibles (75:1) alors qu'en présence de 0,1 mM EDTA, une bande d'intensité similaire n'est observée qu'à un ratio de 125:1 (Figure 56). Il semble donc que la fixation de l'ADN par la protéine soit sensible à la présence d'un ion divalent. Même si aucun ion divalent n'a été ajouté dans la solution pour la première série d'échantillons (Figure 56, pistes 1 à 9), la présence de faibles quantités issues des différents étapes de purification de la protéine et de l'ADN n'est pas à écarter et le simple fait de ne pas mettre l'EDTA dans la solution suffit à améliorer la formation des complexes. Plusieurs ions divalents pouvant être chélatés par l'EDTA sont connus pour leur implication dans la formation de complexes ADN/protéine dont les ions Mg^{2+} , Zn^{2+} , Ca^{2+} .

Dans le but d'identifier l'ion divalent impliqué dans la réaction, nous avons procédé dans un premier temps à des essais en présence de 1 mM $MgCl_2$, mais la présence du magnésium semble n'avoir aucun effet sur la formation des complexes (Figure 57). D'après les résultats des figures 54 et 55, il semble donc qu'un ion divalent, autre que le $MgCl_2$, soit impliqué dans la formation du complexe. Afin de tester l'implication du calcium, nous suggérons l'utilisation de l'EGTA, qui est spécifique pour cet ion.

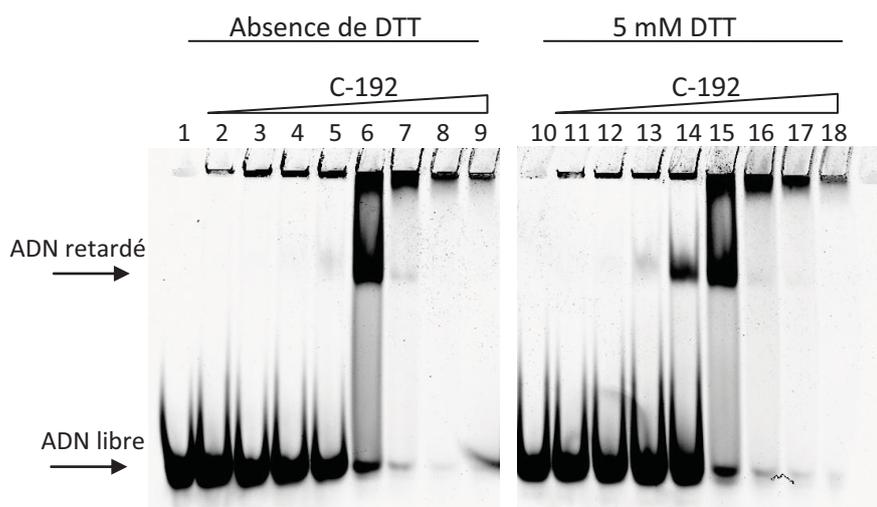


Figure 58 : Étude de l'effet de la concentration du DTT sur l'interaction entre C-192 et l'ADN Selex-8. Gel natif polyacrylamide 5%. 10 nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de la protéine C-192 (ratio : 50/1 à 225/1) en absence de DTT (Pistes 1 à 9) ou en présence de 5 mM DTT (Pistes 10 à 18). Les bandes ont été révélées par Odyssey à 700 nm.

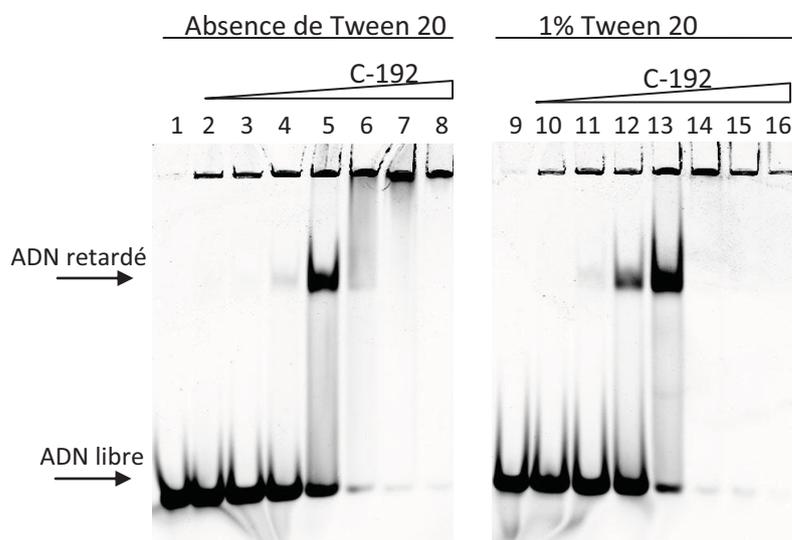


Figure 59 : Étude de l'effet de la concentration du détergent Tween 20 sur l'interaction entre C-192 et l'ADN SELEX-8. Gel natif polyacrylamide 5%. 10 nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de la protéine C-192 (ratio : 50/1 à 200/1) en absence de Tween 20 (Pistes 1 à 8) ou en présence de 1% Tween 20 (Pistes 9 à 16). Les bandes ont été révélées par Odyssey à 700 nm.

Les conditions de fixation idéales demandent beaucoup d'optimisations, et nous mettons en évidence dans la suite des expériences, d'autres paramètres qui pourraient influencer Pf-Int. Une partie de l'ensemble des conditions testées concernent surtout les agrégations observées lors des expériences d'EMSA. Ce phénomène souvent lié à la présence en excès de l'un des composants, ce qui engendre la formation de complexes multivalents, qui finissent par former un réseau supramoléculaire résultant en un précipité. Les interactions à la base de la formation de ce réseau peuvent être différentes. En premier, les interactions hydrophobes pourraient intervenir car la réorganisation nécessaire à la reconnaissance spécifique d'un site ADN requiert un changement d'entropie dû au nombre de charges masquées, n'étant plus en contact avec le solvant lors de la fixation de la protéine. Par ailleurs, il est possible que des ponts disulfures puissent se former entre des cystéines (en particulier les Cys252 et Cys386) qui n'étaient pas exposés jusqu'à lors, étant donné que Pf-Int semble être monomérique en solution (voir résultats d'ultracentrifugation analytique décrits plus haut). La présence d'agrégats massifs au niveau du puits pouvant donc être causés par les conditions redox de la solution ou par des interactions hydrophobes, nous avons mesuré le retard sur gel en présence de différentes concentrations de l'agent réducteur DTT et du détergent Tween 20.

➤ **Effets de la présence d'agent réducteur et de détergent sur la réaction de fixation**

Ainsi, nous avons observé que l'addition du DTT n'a qu'un impact limité sur la formation des agrégats et du complexe retardé. En effet, même en l'absence de DTT le complexe défini est retardé de façon identique à la condition standard (Figure 58).

La présence de 1% de Tween 20 dans la solution semble avoir un effet sur la formation de la bande d'ADN retardé (Figure 59). La concentration de 1% permet l'apparition des complexes retardés à un ratio plus faible (à 100:1) mais elle n'a aucun effet sur la formation d'agrégats qui apparaissent à partir d'un ratio identique dans les deux cas (Figure 59, pistes 6 et 14). Ce résultat est en accord avec les observations obtenues en présence de concentrations élevées en sel et confirme que les agrégats sont formés essentiellement, mais pas uniquement, *via* des interactions électrostatiques. La présence de Tween 20 semble favoriser une meilleure migration de l'ADN retardé, potentiellement grâce à son effet sur les extrémités marquées par le fluorescent Dy682. En effet ces molécules fluorescentes sont traditionnellement composées de structures moléculaires basées sur des cycles aromatiques, mais le Dy682 a été rendu partiellement chargé avec des groupements sulfates, limitant ainsi la formation d'agrégats.

L'action de fixer de l'ADN provoquerait de ce fait soit un changement de conformation qui

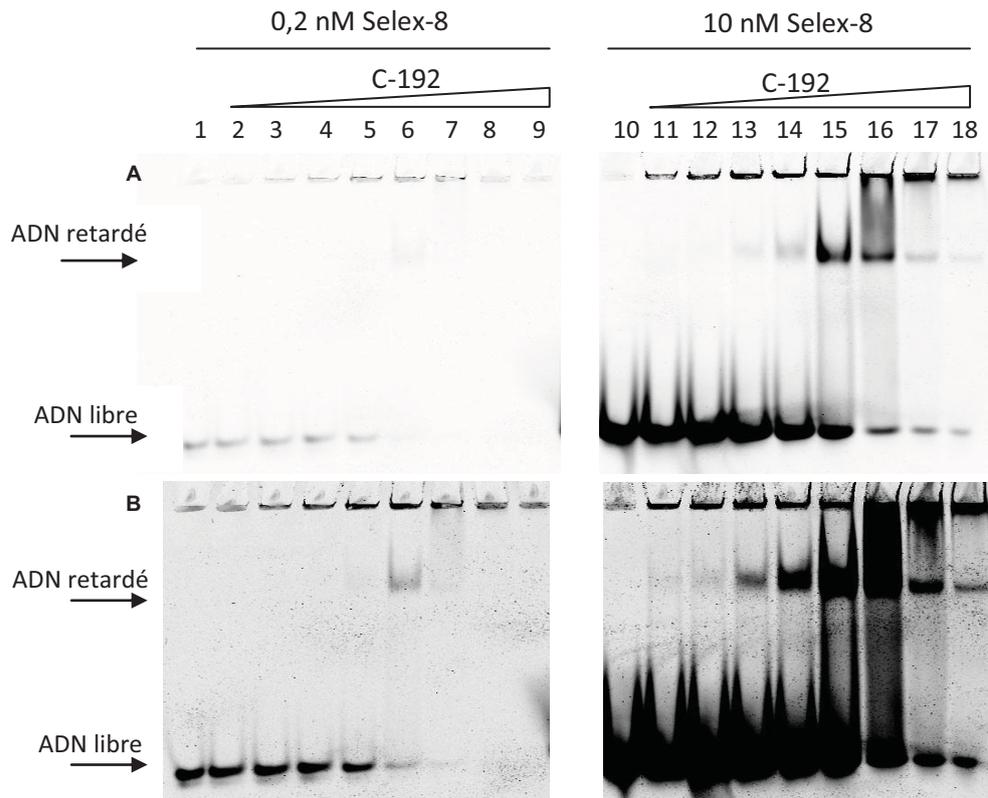


Figure 60: Étude de l'effet de la concentration d'ADN sur l'interaction entre C-192 et l'ADN SELEX-8. Gel natif polyacrylamide 5%. Pistes 1 à 9 : 0,2 nM d'ADN marqué (DY682). Pistes 10 à 18: 10 nM d'ADN marqué (DY682). Les bandes ont été révélées par Odyssey à 700 nm. Gel A : Intensité du laser à 5. Gel B : Intensité du laser à 7. Pistes 2-9, 11-18, l'ADN a été mélangé avec différentes concentrations de la protéine C-192 (0,5 à 2,25 μ M).

entraîne l'exposition de nouvelles surfaces soit hydrophobes ou ayant des charges complémentaires, ou formerait des compléments de surfaces bi-polaires (chargées +/-) ADN-protéines qui peuvent s'accumuler. La résistance de ce complexe à une très forte concentration de NaCl mène à penser que le composant hydrophobe de ces interactions n'est pas négligeable, et inversement, sa résistance au Tween 20, nous laisse supposer qu'il y a aussi un fort composant ionique. Une possibilité supplémentaire serait que ces interfaces ne soient pas ou peu accessibles au solvant, résistant ainsi à tous les types d'agents.

Si on considère que l'interaction de C-192 avec l'ADN Selex8 induit la formation d'un réseau dense, sa compacité devrait dépendre du taux de réticulation de la maille qui la compose qui lui-même dépend de la taille et de la concentration des éléments qui y sont impliqués. Nous avons donc testé ceci en étudiant l'effet de la longueur et de la concentration de l'ADN sur la fixation.

3.5.3.3 Effet de la concentration d'ADN

La figure 60 montre que la concentration d'ADN marqué (soit à 0,2 soit à 10 nM) dans la réaction de fixation a un effet sur les ratios d'observation des ADN retardés et de la formation d'agrégats. En présence de 0,2 nM de substrat ADN, la bande retardée a été observée à un ratio plus élevé (ratio = 7500 :1) que celui pour lequel on observe le maximum d'ADN retardé si la réaction est réalisée en présence de 10 nM (ratio 150:1).

Contrairement à ce qui serait attendu, notamment un retard complet au même ratio, indépendamment de la concentration de l'ADN, avec en particulier l'obtention du maximum de retard à des concentrations de protéines plus faibles pour la série à 0,2 nM d'ADN, nous avons observé que le complexe s'est formé à des concentrations élevées (Figure 60B, piste 6). La concentration de l'ADN marqué semble donc jouer un rôle sur le ratio à partir duquel nous observons l'apparition des complexes de masses moléculaires élevées retenus dans les puits, en limitant son apparition à faibles concentrations d'ADN. Par conséquent le phénomène d'agrégation serait partiellement un phénomène continu qui générerait au fil du temps des macro-complexes de plus en plus imposants, jusqu'à former un précipité au niveau des puits. La quantité d'ADN marqué présent au départ représente les difficultés de former le réseau d'agrégation à faible concentration d'ADN. Cependant, la présence de poly(dI-dC) pourrait contribuer à masquer la dépendance vis-à-vis de la concentration de l'ADN marqué. Il est à noter que, dans le cas de 0,2 nM ADN, nous avons été obligés d'utiliser 1,4 fois plus de puissance de laser pour pouvoir observer les bandes retardées.

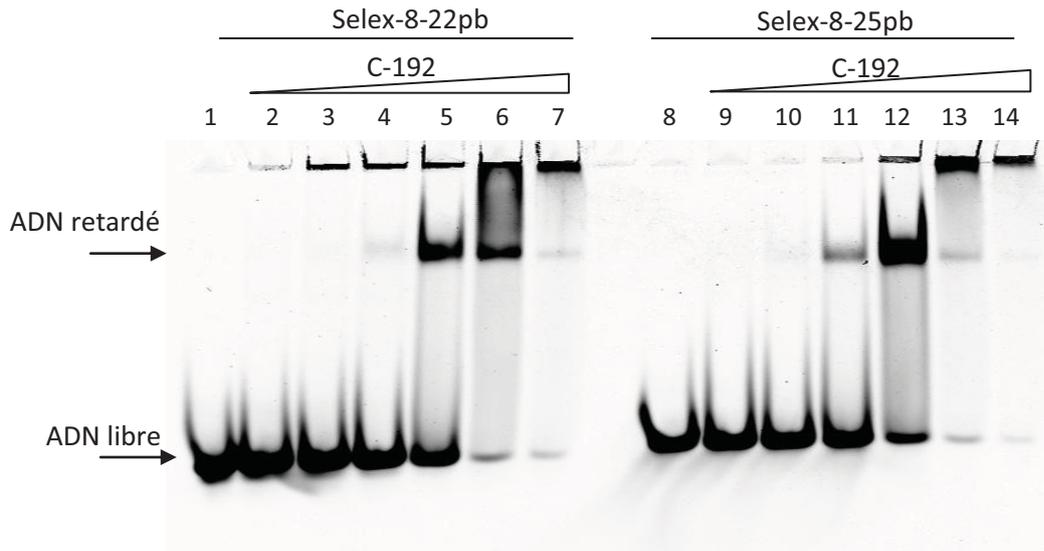


Figure 61: Étude de l'effet de la taille de l'ADN sur l'interaction entre C-192 et Selex-8. Gel natif polyacrylamide 5%. Pistes 1 à 7: ADN Selex-8-22pb. Piste 8 à 14: ADN Selex-8-25pb. 10 nM d'ADN marqué (DYE682) ont été mélangés avec différentes concentrations de la protéine C-192 (ratio : 50/1 à 175/1). Les bandes ont été révélées par Odyssey à 700 nm.

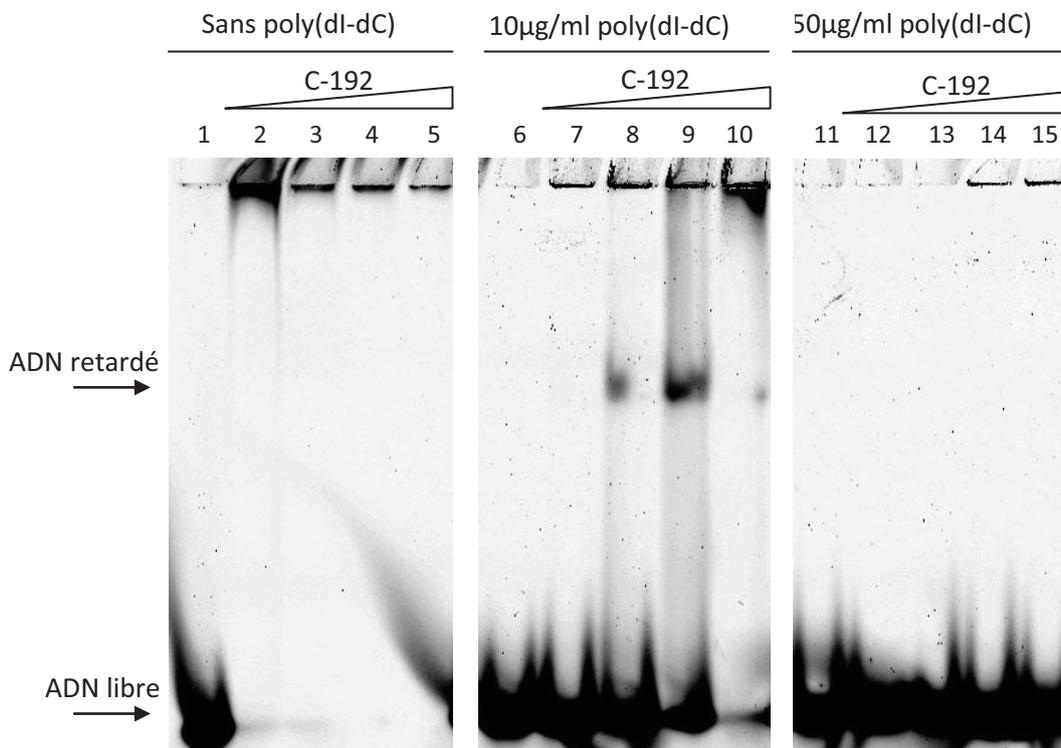


Figure 62: Étude de l'effet de la concentration du compétiteur non spécifique poly(dI-dC) sur l'interaction entre C-192 et l'ADN SELEX-8. Gel natif polyacrylamide 5%. 10 nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de protéine C-192 (ratios : 100/1 à 175/1) en présence de différentes concentrations de poly(dI-dC). Pistes 1 à 5: 0 µg /ml poly(dI-dC). Pistes 6 à 10 : 10 µg /ml poly(dI-dC). Pistes 11 à 15 : 50 µg /ml poly(dI-dC). Les bandes ont été révélées par Odyssey à 700 nm.

3.5.3.4 Effet de la taille de l'ADN sur les propriétés de fixation par Pf-int

Deux tailles d'ADN ont été testés (Figure 61), la protéine est capable de se fixer sur un ADN dont la taille est de 22 pb et de former des complexes à des ratios relativement identiques à ceux obtenus avec les ADN de 25 pb et ne semble avoir qu'un impact limité sur le phénomène de formation des agrégats retenus dans les puits. En effet, en présence de l'ADN plus court (selex-8-22pb) nous avons réussi à observer une bande d'ADN retardé au ratio 150:1 (Figure 61, piste 6) alors qu'au même ratio et en présence de l'ADN selex-8 de 25pb (Figure 61, piste 13), la majorité de l'ADN est retenu dans le puits.

Le changement de 3 pb ne semble pas empêcher l'agrégation et est peut-être une indication que l'interface créée au moment de la fixation sur l'ADN par Pf-Int n'inclut pas les extrémités des oligonucléotides. Ce phénomène d'agrégation par les extrémités a souvent lieu en cristallogénèse de complexes protéine-ADN, où ce sont les bases qui affleurent en 5' ou 3' qui engendrent un empilement favorable à la mise en bout-à-bout des molécules de la maille.

Finalement pour déterminer si l'addition de l'agent utilisé pour réduire les interactions non-spécifiques poly (dI-dC) pourrait interférer avec la réaction ou même y contribuer, nous avons testé l'effet de sa concentration.

3.5.3.5 Effet de la concentration du compétiteur non spécifique poly(dI-dC)

La figure 62 présente l'étude de l'effet de la concentration du polymère synthétique compétiteur non spécifique poly(dI-dC) en solution sur la formation des complexes protéine-ADN. En présence de 10 µg/ml de poly(dI-dC) (Figure 62, pistes 6 à 10), nous observons la formation d'une bande d'ADN retardé à un ratio de 125/1 et 150/1 (pistes 8 et 9) et la disparition du signal à 175/1 (piste 10). Par contre, en présence de 50 µg/ml de poly(dI-dC), nous n'avons observé ni une bande d'ADN retardé ni la formation d'agrégats même à un ratio de 175/1 (Pistes 11 à 15). En absence de poly(dI-dC), nous observons la disparition complète du signal de la sonde à partir d'un ratio 100/1 (piste 2). Ceci montre que la concentration du poly (dI-dC) intervient dans la formation du complexe retardé ainsi que dans celle des agrégats. Une des implications serait que le poly(dI-dC) peut également être fixé par la protéine et par conséquent participer à tous les phénomènes observés avec l'ADN cible, en particulier la formation d'agrégats. La contribution de petits ADNs non spécifiques à la stabilisation de complexes a été observée dans le cas de la protéine P53. Cette dernière agit sous forme tétramérique pour former un complexe avec son ADN spécifique et nécessite la présence de ces petits ADNs pour éviter l'agrégation [140].

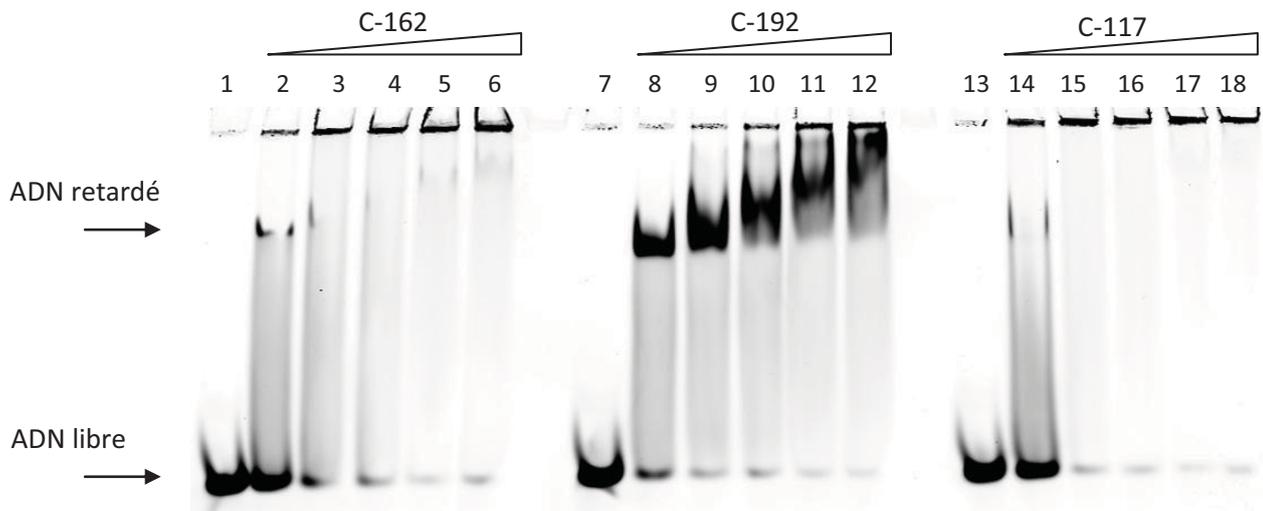


Figure 63: Étude de l'interaction entre les fragments C-117, C-162 et C-192 et l'ADN Selex-8 par retard sur gel en absence de poly (dl-dC). Gel natif polyacrylamide 5%. 10 nM d'ADN marqué (DY682) ont été mélangés avec différentes concentrations de protéine en absence de compétiteur poly (dl-dC). Pistes 1, 7 et 13 : ADN seul. Pistes 2-6 : fragment C-162 (ratios : 4,2/1 à 15,4/1). Pistes 8-12 : fragment C-192 (ratios : 4,2/1 à 15,4/1). Pistes 14-18 : fragment C-117 (ratios : 3,5/1 à 8,4/1). Les bandes ont été révélées par Odyssey à 700 nm.

Ainsi, pour la suite des expériences nous avons décidé de travailler à des concentrations usuelles décrites dans la littérature de 10 µg/ml. Cependant, dans le but de mieux caractériser le comportement des différents fragments de la protéine, nous avons réalisé des essais de retard sur gel en absence d'ADN compétiteur exogène.

En absence de compétiteur poly(dI-dC) nous avons pu observer (Figure 63) que les trois protéines interagissent avec l'ADN selex8 mais de manière différente les unes par rapport aux autres. En particulier, la présence de 42 nM du fragment C-192 (piste 8) induit le retardement de 90% des 10 nM de l'ADN utilisé. L'ajout de quantité plus importante de protéine (pistes 9-12) induit la formation de complexes de tailles plus importantes et dont la migration est encore plus retardée.

La présence des fragments C-162 (pistes 2 à 6) et C-117 (pistes 14 à 18) dans la même gamme de concentrations que celles utilisées pour C-192 induisent l'apparition d'une bande d'ADN retardé de faible intensité à 42 nM pour C-162 et 35 nM pour C-117. L'ajout de quantités plus importantes de protéines C-162 et C-117 induit la disparition totale du signal avec probablement la formation de complexes retenus au niveau des puits.

L'absence de poly(dI-dC) a donc permis d'observer de faibles quantités de complexes (en présence de C-162 et C-117) qu'on n'avait pas pu observer précédemment en présence du compétiteur. Ces complexes se forment probablement même en présence de poly(dI-dC) mais ne sont pas détectés car ils sont retenus dans les agrégats. Ce phénomène étant probablement amplifié lorsque on est en présence de quantités plus importantes de poly(dI-dC).

Il est par ailleurs intéressant de noter que pour les trois protéines (C-117, C-162 et C-192) les ratios protéine:ADN, pour lesquels les bandes retardées et les agrégats retenus sont observés dans les puits, sont différents en fonction de la concentration en poly(dI-dC). Ainsi, en absence de poly(dI-dC) les ratios protéine:ADN utilisés sont de l'ordre de 4:1 à 15:1 alors que pour les tests réalisés en présence de 10 µg/ml de ce compétiteur (Figure 51), les ratios sont supérieurs et sont de l'ordre de 50:1 à 200:1 (protéine:ADN).

Grâce aux différentes expériences de retard sur gel que nous avons réalisé, nous avons pu mettre en évidence, pour la première fois, l'interaction entre Pf-Int et les ADN cibles. Nous avons pu montrer que deux mécanismes interviennent dans cette interaction. Le premier, de type aléatoire, participe fortement dans la formation d'agrégats et est sensible en partie à la concentration de sel utilisée. Le deuxième mécanisme observé est basé sur l'utilisation d'un

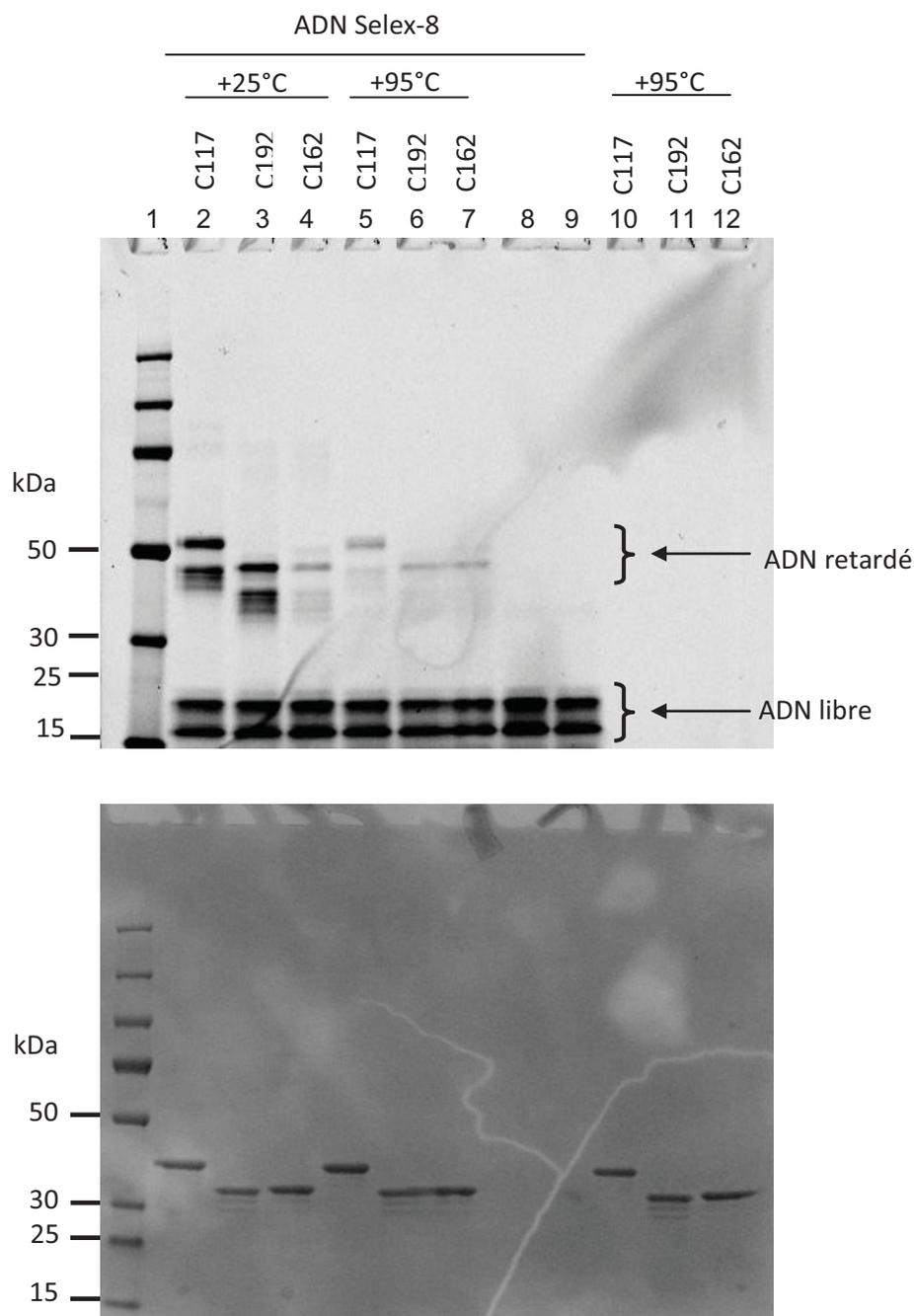


Figure 64: Mise en évidence de la formation de complexes covalents entre les fragments C-117, C-162 et C-192 et l'ADN Selex-8 par retard sur gel dénaturant SDS-PAGE (4-12%). 20 nM d'ADN Selex-8 marqué (DY682) ont été incubés avec 1,4 μ M de protéine en absence de poly (dI-dC) et de BSA. Les mélanges sont ensuite additionnés de 1% SDS, chauffés à 95 °C pendant 15 minutes et chargés sur gel SDS-PAGE. Dans un premier temps, les bandes d'ADN ont été révélées par Odyssey à 700 nm. Les bandes de protéines ont été révélées par coloration au SafeBlue Stain. Piste 1 : Marqueur de poids moléculaires (50, 30, 25 et 15 kDa). Pistes 2-4 : mélanges ADN/protéines additionnés de 1% SDS et incubés pendant 15 min à température ambiante. Pistes 5-7 : mélanges ADN/protéines additionnés de 1% SDS et incubés pendant 15 min à 95 °C. Piste 8 : ADN additionné de 1% SDS, 15 min à température ambiante. Piste 9 : ADN additionné de 1% SDS et incubé 15 min à 95 °C. Pistes 10-12 : 1,4 μ M de protéines dénaturées pendant 15 min à 95 °C. Pistes 2, 5 et 10 : fragment C-117. Pistes 3, 6 et 11: fragment C-192. Pistes 4, 7 et 12 : fragment C-162.

site dédié à la fixation de l'ADN, l'interaction de l'ADN avec la protéine *via* ce site permet de former des complexes plus homogènes capables de migrer en bande définie.

Nous avons également pu démontrer que la fixation de l'ADN par Pf-Int dépend du fragment protéique utilisé. En effet, il semble que la présence du peptide en amont de la position Gly182 limite l'accès de l'ADN au site de fixation. Ce n'est pas la présence physique d'un peptide qui est responsable de cette restriction mais c'est sa nature, puisque son remplacement par un autre peptide de longueur équivalente mais composé d'une autre séquence permet la formation de complexes. La région en amont de la position Gly182 pourrait donc agir comme régulateur de l'activité de l'enzyme. Nous avons également montré que l'interaction ne dépend que faiblement des facteurs du tampon, mis à part la concentration du poly(dI-dC) dont les proportions affectent considérablement le comportement de la protéine. L'implication d'un ion divalent dans la réaction de fixation de l'ADN a également été observée, l'ion impliqué n'a pas encore été identifié mais il ne s'agit vraisemblablement pas d'un ion magnésium.

En plus de leur capacité à former des complexes qu'on peut observer sur gel natif, les recombinauses à tyrosine ont la propriété de pouvoir former des complexes covalents avec l'ADN cible. En effet, lors de la réaction de recombinaison il se forme un complexe intermédiaire impliquant une liaison covalente phospho-tyrosine. La présence de complexes résistants aux conditions utilisées en électrophorèse sur gel polyacrylamide en conditions dénaturantes (SDS-PAGE) permettrait de démontrer que la protéine est purifiée sous forme active et qu'elle reste piégée sous forme d'un complexe covalent.

3.6 Détermination de l'état fonctionnel de Pf-int purifiée : Mise en évidence de la formation de complexes covalents

Dans le but de vérifier l'activité de Pf-Int, nous avons essayé de mettre en évidence la formation de complexes covalents entre les différents fragments purifiés et l'ADN. Les protéines purifiées ont été mélangées avec l'ADN marqué en absence de poly (dI-dC) et en absence de BSA. Deux séries de mélanges ont été réalisées : en présence et en absence de chauffage à 95°C.

La séparation des mélanges ADN/protéine par SDS-PAGE a permis d'observer la formation de complexes covalents (Figure 64, pistes 2 à 7) pour les trois protéines (C-117, C162, C192) montrant que ces dernières sont purifiées sous formes actives. Les complexes covalents ont

été observés aussi bien en présence (Figure 64, pistes 5-7) qu'en absence de chauffage (Figure 64, pistes 2-4). La quantité de complexes formés est cependant faible et les bandes retardées n'ont pu être observées qu'en utilisant le marquage de l'ADN (observation à l'Odyssey à 700 nm) et pas par coloration au bleu de Coomassie. Pour une raison que nous n'avons pas pu identifier, l'ADN libre migre sous la forme de deux bandes (entre et 15 et 25 kDa) (Pistes 8 et 9). La deuxième bande peut être due à la présence de l'ADN sous une conformation différente, le premier étant le duplexe d'ADN attendu après « annealing » et le deuxième un produit de dissociation partielle du duplexe, observé en présence de SDS. La présence de deux bandes peut également être due à la présence du même duplexe d'ADN qui réagissent différemment aux conditions de l'expérimentation, notamment à cause de l'interaction du Dy682 avec le SDS en formant un dimère.

On remarque que les deux formes d'ADN sont capables de former des complexes. Cependant, les quantités sont plus importantes pour l'une des deux formes d'ADN qui forme une bande plus intense.

Concernant les fragments C-162 et C-172 pour lesquels nous n'avons observé que des quantités limitées de complexes en conditions natives, nous avons réussi à observer des complexes covalents en conditions dénaturantes (Pistes 5 à 7). Ces protéines sont donc actives et par conséquent peuvent fixer l'ADN *via* le site catalytique. L'absence de complexes en conditions natives est donc probablement due à la faible stabilité des complexes non covalents formés en conditions natives.

Grâce à la mise en évidence de complexes covalents, nous avons pu démontrer l'appartenance de Pf-Int à la famille des recombinases à tyrosine. Cependant, dans le but de confirmer ces résultats, un contrôle impliquant une version de la protéine où la tyrosine catalytique Tyr473 a été mutée doit être réalisé. Par ailleurs pour s'affranchir des effets potentiels des conditions de réaction (impact du SDS sur le marquage), la réaction doit être répétée soit en changeant la position du fluorophore à une région interne de l'ADN soit utiliser d'autres types de marquage tel que la radioactivité.

3.7 Étude de la spécificité de C-192 vis-à-vis de l'ADN

En plus de pouvoir former des complexes covalents, les recombinases à tyrosine sont caractérisées par la capacité à se fixer spécifiquement aux sites de recombinaison. Dans le but d'étudier la spécificité du fragment C-192 vis-à-vis de l'ADN, nous avons d'abord utilisé un fragment non-spécifique d'ADN d'une taille de 40 pb (protect-40) de la région promotrice de

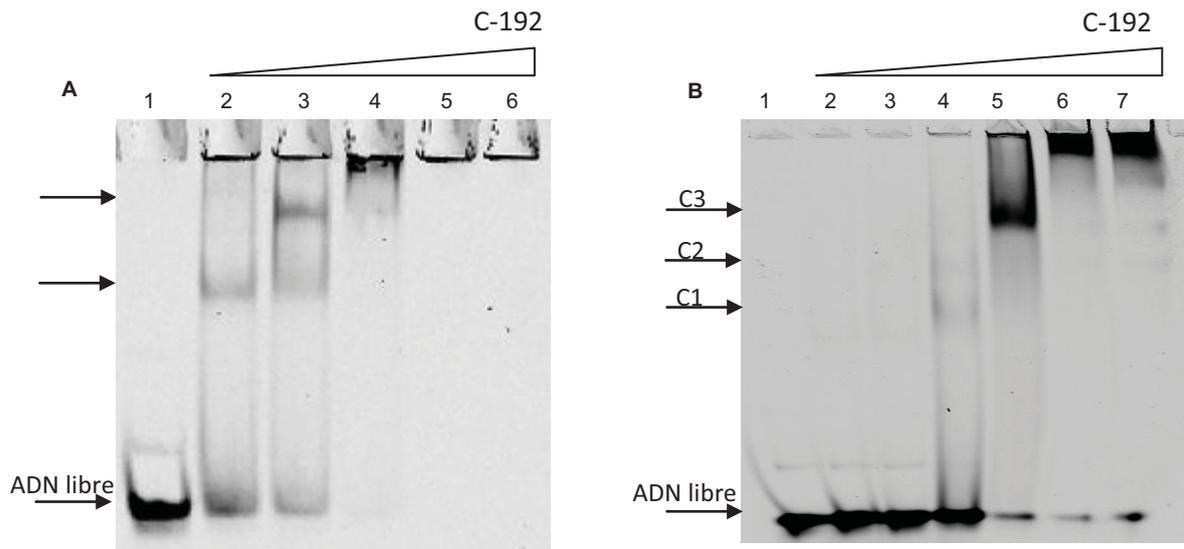


Figure 65: Étude par retard sur gel de l'interaction entre C-192 et l'ADN protect40 de la région promotrice du gène *phoP* de *M. tuberculosis* (40 pb). Gel natif polyacrylamide 5%. 10 nM d'ADN marqué (DY682) ont été mélangés avec des concentrations croissantes de C-192. **Gel A** : réactions réalisées en présence de 150 mM NaCl. Piste 1 : ADN seul. Pistes 2-6 : ADN additionné de C-192 (ratios protéine/ADN : 120/1 à 180/1). **Gel B** : réactions réalisées en présence de 900 mM NaCl. Piste 1 : ADN libre. Pistes 2 à 7 : ADN additionné de C-192, ratios protéine/ADN : 50/1 à 175/1. Les bandes ont été révélées par Odyssey à 700nm.

la protéine PhoP (*Mycobacterium tuberculosis*) disponible au laboratoire. Nous avons observé au cours de cette analyse (figure 65, gel A) un phénomène identique à celui précédemment observé en présence de l'ADN selex8. En effet, l'ajout de quantités croissantes de C-192 résulte en une disparition graduelle du signal de l'ADN libre avec apparition de complexes définis et d'autres de masses moléculaires élevées retenus au niveau des puits. La disparition complète du signal a lieu à un ratio 150:1 (protéine/ADN). Nous avons aussi pu observer la formation de deux espèces retardées intermédiaires. La deuxième partie de la figure (Figure 65, gel B) montre que ces complexes peuvent être formés en présence de 900 mM NaCl et que cette concentration élevée en sel favorise même la libération des complexes stables à partir des agrégats (Gel B, piste 5).

Les expériences réalisées avec l'ADN protect-40 confirment les suppositions que nous avons proposé suite à l'étude de l'impact du poly(dI-dC) et montrent que le fragment C-192 est capable de former plusieurs complexes stables avec un ADN autre que le selex8.

Ces expériences montrent également que le nombre de bandes retardées est dépendant de la taille de l'ADN. Il y a une différence de taille, de l'ordre de deux, entre protect-40 et selex8, et d'après nos observations cela entraîne la formation de deux complexes définis (Gel A, piste 3). On peut donc supposer que dans ce cas, deux monomères de C-192 se fixent sur le même duplexe d'ADN, avec le potentiel de former un dimère. Cette possibilité est d'autant plus plausible que dans le cas des recombinases à tyrosines dont les structures ont été élucidées, on peut observer que les sites de recombinaisons (~40 bp) sont occupés par des dimères d'enzymes. Ce phénomène est à la base de la formation de la synapse de recombinaison (deux dimères se faisant face, figure 12), et l'accumulation de ces unités peut éventuellement mener à l'agrégation des complexes.

Pendant, sur le deuxième gel (Figure 65, gel B), il est aussi possible de distinguer sur la piste 4 qu'une bande retenue intermédiaire de faible en intensité (C2) existe entre les bandes (C1) et (C3). Ceci pourrait affiner notre hypothèse de l'assignation de la bande C1 car il serait possible que cette bande soit formée d'un seul monomère de protéine et d'un duplexe d'ADN, que la bande intermédiaire C2 soit formée de deux molécules fixées sur un duplexe ADN, et que C3 soit formée d'un tétramère de molécules de C-192 fixées sur deux duplexes ADN. Ceci reste à être démontré de façon plus convaincante, par des expériences de « cross linking » et en déterminant la taille des complexes par spectrométrie de masse.

Dans le but de vérifier si la nature du système utilisé peut avoir un effet sur la spécificité des interactions protéine-ADN, nous avons procédé à une expérience d'interactions croisées

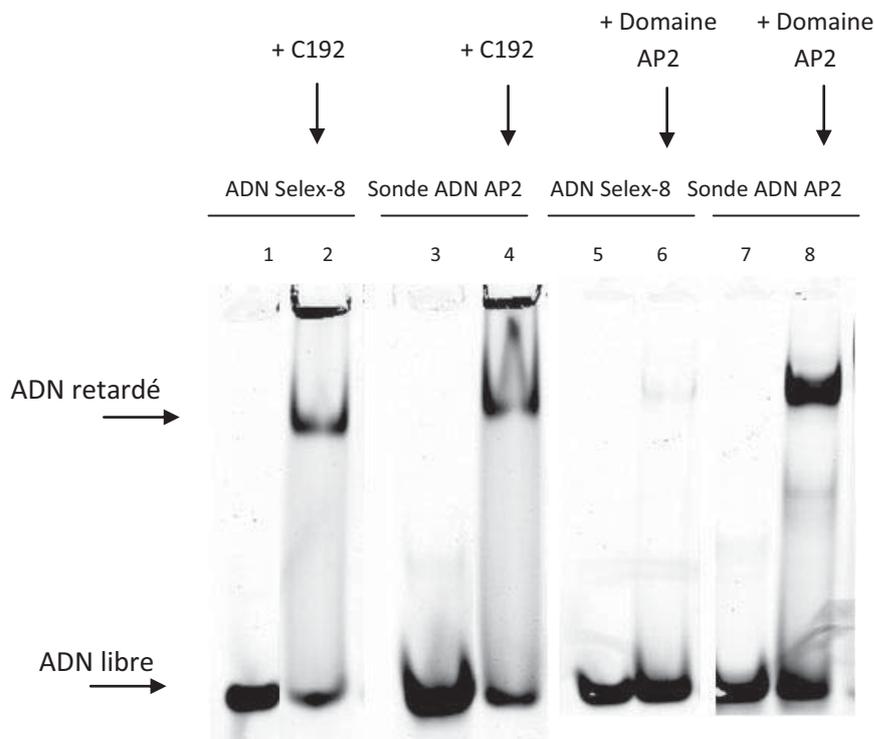


Figure 66 : Étude de la spécificité de C-192 vis-à-vis de l'ADN. Gel natif polyacrylamide 5%. Pistes 1, 2, 4 et 6 : ADN Selex-8. Pistes 3, 4, 7 et 8 : ADN sonde AP2. Pistes 1, 3, 5 et 7 : ADN libre. Pistes 2 et 4 : ADN additionné de 1,25 μ M C-192. Pistes 6 et 8 : ADN additionné de 20 μ M du domaine AP2. 10 nM d'ADN marqué (DYE682) ont été mélangés avec soit la protéine C-192 soit avec le domaine AP2. Les bandes ont été révélées par Odyssey à 700 nm.

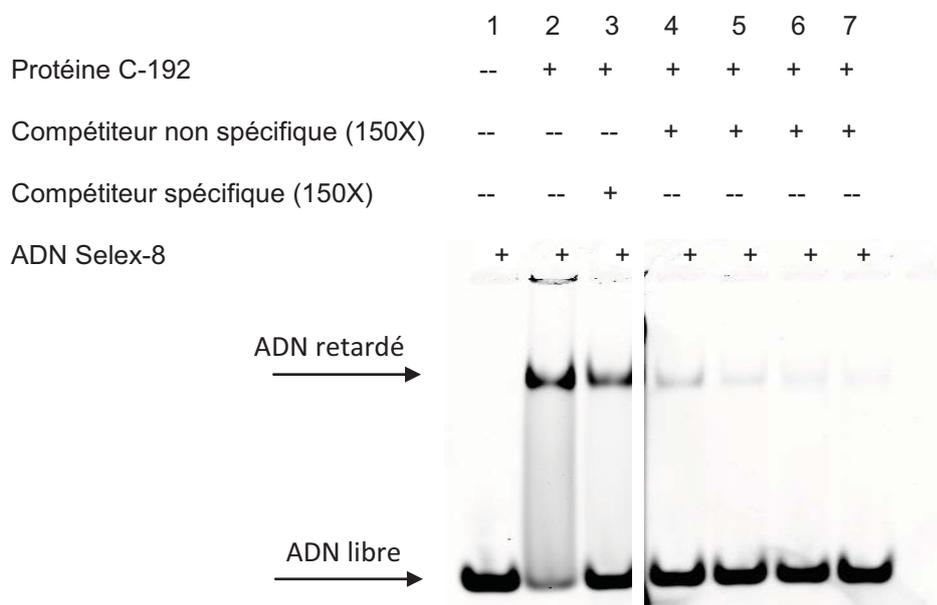


Figure 67 : Étude de la spécificité de C-192 vis-à-vis de l'ADN. Gel natif polyacrylamide 5%. Piste 1: 10 nM ADN Selex-8 marqué (DY682). Piste 2 : ADN Selex-8 marqué additionné de 0,3 μ M de la protéine C-192. Piste 3 : 0,3 μ M de la protéine C-192 ont été pré-incubés avec 150X (1500 nM) de la sonde Selex-8 non marquée (25 pb) avant d'être additionnés de 10 nM de l'ADN Selex-8 marqué (25 pb). Pistes 4-7 : 0,3 μ M de la protéine C-192 ont été pré-incubés avec 150X de différents ADN non marqués avant d'être additionné de 10 nM de l'ADN Selex-8 marqué. Les bandes ont été révélées par Odyssey à 700 nm. Piste 4 : Site de fixation de l'intégrase d'intégrons attC modifié (40 bp, sans structures internes). Piste 5 : Séquence du gène Hsp de *P. falciparum* (42 pb). Piste 6 : Séquence du gène de l'actine de *P. falciparum* (42 pb). Piste 7 : Séquence aléatoire (42 pb).

en utilisant un deuxième couple protéine/ADN. Le deuxième couple est formé d'un domaine issu d'un facteur de transcription ApiAP2 de *P. falciparum* et de sa séquence ADN spécifique (sonde AP2). Dans le contrôle (Figure 66), nous avons observé que le domaine ApiAP2 interagit spécifiquement avec sa sonde ADN. Ainsi, en présence de 20 μM d'ApiAP2, 50% de la sonde AP2 forme un complexe retardé (Piste 8). Par contre, seule une très faible partie de l'ADN selex8 est retardée par ApiAP2 (Piste 6).

Dans ces mêmes conditions, nous avons observé que C-192 interagit d'une manière identique avec la sonde AP2 (Piste 4) et selex8 (Piste 2). Cela nous permet de conclure que l'absence de spécificité précédemment constatée de C-192 n'est pas due au système utilisé mais à la protéine elle-même. Par ailleurs, nous observons aussi que C-192 forme un seul complexe défini avec la sonde AP2 ayant une taille de 25 pb.

3.8 Etudes de compétition (Pulse-Chase) pour déterminer la réversibilité de la fixation d'ADN et la spécificité de l'interaction

Dans le but de mieux affiner la spécificité de C-192 vis à vis de l'ADN selex-8, nous avons conduit des expériences de compétition en utilisant différents ADN comme compétiteurs (Figure 67). Dans la deuxième piste, l'ajout de 0,3 μM de C-192 induit la disparition de 90% du signal au niveau de l'ADN libre et l'apparition d'une bande d'ADN retardé. Le contrôle spécifique (Piste 3) qui a consisté en la pré-incubation de la protéine avec 1500 nM de selex8 non marqué, suivie de l'ajout de 10 nM de selex8 marqué permet l'observation de 90% du signal au niveau de la bande d'ADN libre et l'apparition d'une bande d'ADN retardé. Ceci confirme qu'il y a eu un échange de l'ADN non marqué par la sonde marquée et que cette dernière peut donc entrer en compétition avec la sonde non marquée.

La même expérimentation réalisée par la pré-incubation de la protéine avec 1500 nM de différents ADN compétiteurs non marqués de 40 pb (Pistes 4, 5 et 6) suivie de l'incubation avec 10 nM de la sonde selex8 marquée permet aussi la récupération de 90% du signal au niveau de la bande correspondant à l'ADN libre mais le signal provenant de l'ADN retardé est beaucoup plus faible que celui obtenu dans le contrôle (Piste 3). Il semble donc que l'échange est moins efficace en présence de ces ADN non spécifiques de 40 pb qui peuvent probablement fixer deux molécules C-192. Le complexe formé dans ce cas serait plus stable probablement grâce à la dimérisation des deux protomères fixés sur le même duplexe d'ADN. Il serait donc possible qu'on soit en présence d'un phénomène de coopérativité et qu'ainsi le recrutement du deuxième protomère permette la formation d'un complexe plus stable qui est

plus difficilement déplaçable que lorsqu'un seul protomère est présent. Il a été précédemment proposé qu'il existe une coopérativité dans la fixation dans les cas de lambda-int, Cre et Flp et [141 , 142].

Grâce aux différentes analyses de spécificité que nous avons réalisées pour caractériser l'interaction du fragment C-192 de Pf-Int avec l'ADN, nous avons pu montrer que ce dernier se fixe à l'ADN de manière non spécifique et que cette absence de spécificité n'est pas due au système que nous utilisons. Nous avons également pu montrer que le nombre de monomères fixés sur chaque duplexe d'ADN dépend de la taille de ce dernier. Malgré le fait que C-192 soit prédit comme étant le fragment de Pf-Int composé des domaines homologues aux domaines N- et C-terminaux classiquement retrouvés chez les autres membres recombinases à site spécifique à tyrosine, il semble que la région N-terminale en amont de la position 190 joue un rôle important dans la fonction de Pf-Int, en particulier sur la régulation de l'activité.

4 DISCUSSION

L'absence de vaccins ainsi que le nombre de plus en plus élevé des cas de résistance aux antipaludiques classiques conduisent à l'exploration de nouvelles approches. Dans le but d'identifier de nouvelles cibles, il devient nécessaire de comprendre les mécanismes moléculaires qui permettent à ce parasite d'interagir avec son environnement et de proliférer. Comme nous l'avons vu dans l'introduction, *P. falciparum* est un parasite intracellulaire capable de s'adapter à la réponse immunitaire de l'hôte grâce à un mécanisme de variabilité antigénique. Cette variabilité antigénique est assurée par l'expression coordonnée des antigènes parasitaires présents à la surface de l'érythrocyte infecté. Il est actuellement admis que l'expression mutuellement exclusive des antigènes soit associée à des modifications épigénétiques des histones. Cependant il n'est pas exclu que d'autres niveaux de régulation impliquant des modifications génétiques puissent exister [26, 40].

En plus de la régulation de la transcription des gènes, *P. falciparum* possède un mécanisme encore mal compris qui lui permet de générer de la diversité à l'intérieur même des familles multigéniques. Ce mécanisme serait basé sur l'échange de blocs, entre gènes, assuré par un mécanisme de transposition duplicative [30, 34, 37].

Au cours des travaux de thèse présentés dans ce manuscrit, nous avons tenté d'appréhender le rôle de la recombinaise à site spécifique Pf-Int dans la biologie du parasite *Plasmodium*. Cet enzyme a été récemment identifié au laboratoire et aucune étude précédente ne lui avait été consacrée dans la littérature. En effet, les recombinaises à site spécifiques sont très peu décrites chez les eucaryotes, et en particulier chez les *Plasmodia*. Leurs rôles et les mécanismes employés sont donc d'un intérêt particulier car peu connus, et pouvant potentiellement faire la liaison entre les recombinaises virales ou bactériennes, avec celles des cellules plus évoluées. Les objectifs majeurs de ces travaux ont concerné les études fonctionnelle, biochimique et structurale ainsi que la recherche des cibles potentielles de cette nouvelle recombinaise.

➤ **Les études fonctionnelles**

Les études *in-vivo* de Pf-Int ont permis de montrer que le gène identifié est effectivement transcrit et que l'ARN messager est épissé de manière conforme aux prédictions. L'étude des niveaux d'expression de l'ARN messager montre que la transcription du gène est régulée en fonction des stades parasitaires de la phase érythrocytaire.

Cependant, malgré la présence de l'ARN messager, les tentatives de détection de Pf-Int sous la forme protéique ne nous ont pas permis de vérifier sa traduction et de déterminer sa localisation subcellulaire. En effet, différents anticorps produits au laboratoire à partir de la

protéine recombinante (C-162) ont été utilisés en Western-blot et en immuno-fluorescence pour essayer de détecter la protéine sauvage du parasite sans donner de résultats. Il semblerait donc, qu'en plus de la régulation transcriptionnelle, l'expression de Pf-Int soit régulée à un niveau post-transcriptionnel. Il a été démontré que plusieurs protéines de *P. falciparum* sont soumises à ce type de contrôle [143, 144].

En absence de détection de Pf-Int endogène, nous avons tenté de le surexprimer (résidus 1-490) avec une étiquette de détection (FLAG-tag en C-terminal). Cette approche avait été utilisée auparavant au sein du laboratoire BIHP d'Artur Scherf pour déterminer la localisation des protéines ou dans le but d'identifier des cibles, *in vivo*, de certaines protéines par la technique « ChIP-chip ». Bien que nous ayons obtenu des parasites transfectés viables, nous n'avons pas pu détecter Pf-Int par Western blot et nous n'avons pas pu utiliser ces parasites. Ceci nous a conduits à poser la question sur le contrôle de son niveau d'expression *in vivo*. Dans ce but, nous avons étudié l'impact des deux autres codons d'initiation de traduction possibles présents chez *P. falciparum* en réalisant des essais de sur-expression de ces différentes versions de Pf-Int portant l'étiquette (FLAG-tag) en C-terminal. De la même manière que dans le premier essai, nous avons réussi à avoir des parasites viables mais nous n'avons pas pu détecter les protéines recombinantes. Il semblerait ainsi que l'expression de ces constructions de Pf-Int soient régulée de la même manière que la protéine endogène.

On peut donc considérer que la concentration de la protéine doit être régulée de manière à maintenir un niveau très faible, expliquant notre incapacité à la détecter dans les parasites. Plusieurs mécanismes peuvent être potentiellement impliqués dans cette régulation. Elle peut d'abord être assurée par un mécanisme post-transcriptionnel qui ne permettrait pas à l'ARN-m d'être traduit ou par un mécanisme post-traductionnel qui serait à l'origine de la dégradation rapide de la protéine. Nous suggérons que la présence du transcrit et l'absence de mutation soient vérifiés pour ces parasites.

L'analyse de la séquence primaire de Pf-Int montre qu'elle partage des caractéristiques communes avec des recombinases de levures (FLP du plasmide 2 μ m, et de ses équivalents portés sur les plasmides pKD1, pSB2, pSB3, pSM1 et pSR1) en particulier l'absence des motifs Thr-Gly and Leu-Gly-His précédemment décrits par *Esposito et Scocca* [145]. Plusieurs genres d'apicomplexes tel que *Plasmodium*, *Eimeria* ou *Toxoplasma* possèdent un plasmide de 35 Kb présent au niveau de l'apicoplaste. La majorité des protéines présentes au niveau de l'apicoplaste sont codées par le génome nucléaire et exportées vers l'organelle après traduction. L'export est réalisé via les voies de sécrétion (ER) au moyen de la région N-

terminal des protéines [146]. Même si aucun motif précis n'a été identifié, il a été montré que la région N-terminal de ces protéines est formée d'un peptide signal et d'un peptide de transit. Ce dernier a déjà été décrit chez les protéines exportées vers le chloroplaste chez les plantes. Le peptide de transit est composé majoritairement d'acides aminés hydrophiles et basiques, enrichi en lysine et asparagine [146]. Concernant Pf-Int, les prédictions (réalisées au moyen des programmes PATS et PlasmoAP [147-149]) montrent qu'effectivement cette protéine pourrait potentiellement être exportée dans l'apicoplaste. Dans le but de déterminer l'impact de la région N-terminale de la protéine sur la localisation subcellulaire de cette dernière, nous avons utilisé une chimère de Pf-Int (1-190, *start* défini par PlasmoDB) fusionnée à la GFP. Une telle stratégie avait été utilisée précédemment pour localiser certaines protéines exportées ayant les mêmes caractéristiques en N-terminal [150]. Cependant dans notre cas, elle n'a pas permis de résoudre le problème non-plus car cette construction s'est avérée létale, et nous n'avons pas pu obtenir des parasites. La mort des parasites exprimant la fusion 1-190-GFP implique que cette protéine chimère est traduite et qu'il n'y aurait plus de régulation. On pourrait donc conclure que c'est la partie 3' de l'ARN-m du gène qui jouerait un rôle de régulation négative en limitant sa propre traduction. A partir de cette expérience, nous pouvons conclure que la partie N-terminale de la protéine (aa 1 à 190) possède une certaine toxicité. Cette dernière pourrait être due à la fixation aléatoire et massive à l'ADN soit sous sa forme entière au moment où l'ADN nucléaire devient accessible ou sous sa forme maturée, par exemple après coupure d'une protéase à l'entrée de l'apicoplaste. Effectivement, nous avons précédemment montré par gel retard que le fragment C-117 génère plus d'agrégats que le fragment C-162 lors de la fixation avec l'ADN, qui lui-même génère plus d'agrégats que le C-192. Il paraît donc que l'ajout de séquences du côté N-terminal générerait de plus en plus d'agrégats. La toxicité de la région N-terminal (aa 1-190) pourrait aussi être due à la saturation des systèmes qui doivent prendre en charge son clivage [143]. Et enfin, elle pourrait être due à la présence d'une fonction qui n'aurait pas été détectée par les programmes de recherche de similitude et dont la sur-expression conduirait à la mort du parasite.

Le contrôle de la transcription de l'ARN-m ainsi que le contrôle de la concentration de la protéine sont en accord avec la littérature concernant le taux d'expression régulé des protéines impliquées dans des mécanismes liés à l'ADN tel que des recombinases, gyrases ou des topoisomérases. Effectivement, l'expression des recombinases à un taux élevé, peut mener à des événements de recombinaisons aléatoires, et causer des dommages à l'ADN [135].

En l'absence d'autres éléments d'explications, nous proposons que Pf-Int soit une protéine dont la fonction d'origine est liée à l'apicoplaste de part la présence de sa partie N-terminale

ayant potentiellement une fonction d'export, mais qui grâce à un mécanisme de trafic cellulaire alternatif pourrait être exportée vers le noyau où elle serait impliquée dans la variabilité antigénique. Plusieurs exemples de ces phénomènes de doubles localisations ont été documentés chez plusieurs espèces eucaryotes, tels que *Arabidopsis*, la levure et chez des *Apicomplexa* tel que *Toxoplasma* et *Plasmodium* [136-138]. Il est à noter que la présence d'une protéine à activité topoisomérase de type IB a précédemment été rapportée chez *P. falciparum* [151] et que si Pf-Int est effectivement exportée dans l'apicoplaste, elle aurait forcément une activité autre que topoisomérase. Une telle fonction serait en accord avec la nécessité d'assurer le maintien de cette organelle essentielle car le niveau d'expression maximal observé de Pf-Int coïncide avec le stade de division du parasite schizont, ceci en ayant par ailleurs un rôle dans la génération de diversité.

Dans le but d'étudier l'implication de Pf-Int dans la biologie du parasite, nous avons également créé une souche de parasite dont le gène codant Pf-Int a été invalidé (souche 3D7-KO-Pf-Int créée par Christine BENATAR, BIHP, Institut Pasteur). Les études comparatives de ces deux souches ont montré que Pf-Int n'est pas vitale pour le parasite pendant la phase érythrocytaire en culture. Ceci suggère que Pf-Int est nécessaire au parasite soit pendant la phase se déroulant chez le moustique soit lorsque le parasite est au contact du système immunitaire. En perspective, nous proposons de réaliser des cultures de longues durées et de suivre la commutation entre les gènes des familles multigéniques.

➤ **Organisation structurale de Pf-Int :**

Le cadre de lecture ouvert identifié par PlasmoDB consiste en un polypeptide de 490 acides aminés (Chromosome 13, allant de la position 403,991 à la position 406,906). Parmi les domaines prédits, deux semblent porter la partie fonctionnelle et englobent les régions des aa 192-300 et aa 296-486. Ces deux domaines sont annotés comme étant homologues au domaine N-terminal de l'intégrase du phage lambda (λ -Int) et au domaine catalytique de celui-ci. Dans les cas classiques des recombinases à tyrosine tel que Cre du Phage P1, ces deux domaines se replient en deux modules liés par un connecteur d'une dizaine de résidus et forment une pince (« *C-shaped clamp* »). Par conséquent, le site de fixation à l'ADN se trouve à l'interface entre les deux modules de la pince.

λ -Int, HP1-Int, l'intégrase du transposon Tn916 ou l'intégrase des HPI (High-Pathogenicity Island, Yersinia) sont des enzymes hétérobivalents et se fixent simultanément sur deux sites ADN : les sites de recombinaison et les sites accessoires [67, 152, 153]. L'utilisation de la structure de λ -Int (PDB-Id :1Z1B) comme modèle, montre qu'un troisième domaine « *arm-*

binding domain » pourrait exister dans la région des aa 117 à 156 de Pf-Int. Ce domaine N-terminal n'est pas à confondre avec le domaine N-terminal de Cre ou FLP, car il se positionne plus en amont. Chez les enzymes hétérobivalents, ce domaine interagit avec des sites accessoires présents sur le substrat ADN. Il a été montré qu'en fonction du contexte ce domaine agit comme un modulateur de l'activité λ -Int [70]. En effet, en absence des sites ADN accessoires (P, P'), sa présence inhibe la fixation de la protéine aux sites cœur de recombinaison (COC'). Cependant, la partie C-terminal isolée de λ -Int fonctionne parfaitement comme un topoisomérase. Plusieurs sites de fixation de protéines accessoires tel que IHF, Xis et Fis existent sur le substrat de λ -Int. Ces protéines qui courbent l'ADN permettent de ramener l'intégrase près des sites cœurs de recombinaison et de former un intasome fonctionnel.

Au niveau de l'assemblage quaternaire, la prédiction de la structure de Pf-Int permet aussi de voir que la tyrosine du site catalytique Y₄₇₃ est portée par la dernière hélice α de la région C-terminal. Cette hélice α étant probablement celle qui est échangée par les différents protomères pour former le tétramère réactionnel. Nous proposons donc que le site catalytique de Pf-Int puisse être formé soit en *cis* soit en *trans*, avec la tyrosine d'un monomère donné qui catalyserait la coupure de l'ADN dans le site catalytique du monomère local ou adjacent. L'échange de la tyrosine catalytique entre protomères avait déjà été montré pour la recombinase Flp de la levure *Saccharomyces cerevisiae* [65]. Dans le cas de Flp, la tyrosine catalytique n'est pas portée par la dernière hélice mais par une hélice intermédiaire (hélice M) qui est échangée. Les expériences de *Letzelter et al* montrent que le site catalytique de l'intégrase du bactériophage SSV1 est également assemblé en *trans* [36].

➤ **Identification des cibles de Pf-Int :**

Pour l'étude de Pf-Int nous nous sommes basés sur des résultats précédemment obtenus au laboratoire. En particulier deux études de recherche de cibles potentielles d'ADN cibles en utilisant C-162 comme appât. La première étude a été réalisée par SELEX (*Systematic Evolution of Ligands by EXponential enrichment*) sur des fragments ADN synthétiques aléatoires de 25 pb placés au centre de deux régions adaptatrices pour le clonage et l'amplification. La banque générée par cette étude a donné cinq séquences (SELEX 18, 8, 14, 1 et 6). Ces séquences ont en commun le motif : (N)₆CAANC(A/C)(N)₂GT(N)₃CG(N)₇C(N)₄. Il a été également observé que pour la séquence Selex8, 12 paires de bases correspondent à une séquence d'un gène *rif* (gène codant pour la protéine de surface RIFIN).

La deuxième étude a été réalisée sur des fragments d'ADN génomiques (clone 3D7) d'une

taille comprise entre 200 et 400 pb générés par sonication. 149 fragments retenus par affinité ont été identifiés soit par hybridation sur puce à ADN soit par clonage et séquençage. Concernant les séquences génomiques, il a été noté que 32% de ces séquences correspondent à des fragments issus des familles multigéniques *var*, *stevor* et *rif*. Sur les 49 séquences issues de ces familles multigéniques, 46 correspondent à des régions de l'exon codant pour la partie extracellulaire des protéines correspondantes. L'analyse de ces séquences n'a pas révélé de motif conservé. Cependant, si on considère ces ADNg sous la forme d'ADN simple brin, une majorité des séquences sont capables de former des structures secondaires de type tige-boucle car elles contiennent des régions palindromiques de petites tailles. L'ADN simple brin peut exister pendant les étapes de réplication ou de la réparation des cassures de l'ADN. Certaines intégrases, en particulier celles impliquées dans l'acquisition des cassettes par les intégrons sont d'avantage sensibles à la structure de leurs substrats et utilisent donc des simples brins repliés. Ces enzymes utilisent un premier site de recombinaison classique attI (site d'insertion) et un site de recombinaison attC (site présent sur les cassettes). Les sites attC sont hétérogènes en taille et se forment lors du repliement d'un ADN simple brin en structure tige-boucle [100]. Ce repliement n'est pas parfait sur toute la longueur de la séquence et engendre l'apparition de bases extrahélicales nécessaires à la reconnaissance de l'ADN par la protéine et à l'orientation du complexe [66]. Les ADN simples brins sont également utilisés par certains transposons (de la famille des transposons IS200/IS605) [154]. Ces derniers profitent de la présence de ce type de substrat au niveau du brin retardé lors de la réplication. Il a été montré que l'insertion et l'excision du transposon ont spécifiquement lieu au niveau du site 5'TTAC3' et dépendent de la formation de structures secondaires de l'ADN simple brin (LE, « left end » et RE, « right end ») [155].

Par conséquent, n'ayant pas clairement identifié, pour l'instant, ni des motifs linéaires ni des structures uniques en tiges boucles conservés comme substrats, nous ne pouvons pas proposer de cibles de Pf-Int de façon définitive. Il serait cependant intéressant de considérer, à la lumière de la localisation de la majorité des cibles identifiées au niveau des régions subtélomériques, qu'ils soient souvent dans des situations de remaniement d'ADN, et qu'ils pourraient facilement se structurer de part la teneur en séquences à symétrie palindromique qu'ils contiennent.

Dans le but de valider les cibles ADN identifiées, nous avons procédé à des expériences de retard sur gel natif (EMSA) en utilisant différents fragments de la protéine et différents ADNs. Les expériences d'EMSA ont permis d'étudier le mode de fixation d'ADN par Pf-Int, notamment s'il y a une différence d'interactivité des régions de la protéine, et la spécificité de

ces cibles potentielles. Ces analyses ont permis de montrer que les différentes constructions de la protéine réalisées interagissent de manières différentes avec l'ADN. Nous avons d'abord pu observer que tous les fragments testés avaient tendance à interagir de manière aléatoire avec l'ADN. En présence de quantités croissantes de protéine, ce type d'interaction engendre la formation de complexes hétérogènes conduisant à la formation d'agrégats. La formation de ces complexes hétérogènes est probablement due à l'interaction de la protéine via sa surface chargée à cause du point isoélectrique élevé ($pI > 10$) ou à des changements conformationnels qui sont induits au contact de l'ADN et qui révéleraient des surfaces hydrophobes. Il a été précédemment observé que pour l'enzyme Cre, l'ajout de l'ADN cible provoque des changements conformationnels qui causent la précipitation d'une partie de la protéine. Malgré la présence de ces interactions aléatoires ioniques ou hydrophobes, nous avons pu démontrer pour la première fois que certains fragments de Pf-Int sont capables de former des complexes ordonnés formant des bandes discrètes sur gel natif et stables en milieu à hautes forces ioniques (>900 mM NaCl). Ces complexes sont probablement formés par un mécanisme différent du premier et impliquent la participation d'un site de fixation dédié à l'ADN de la protéine, et dont l'interface serait protégée de l'impact ou de l'accessibilité du solvant. Malgré les différentes conditions que nous avons testées, nous n'avons pas pu nous affranchir du premier phénomène (interactions non-spécifiques) qui conduit à l'agrégation de la protéine, de l'ADN libre et/ou des complexes discrets formés.

Concernant les différences de comportement des différents fragments de la protéine vis-à-vis des ADNs, il est à noter que seuls les fragments C-182 et C-192 sont capables de former des complexes ordonnés et stables. Quant aux fragments C-117 et C-162, une faible quantité seulement de complexe a pu être formée et ce, en absence de compétiteur non spécifique. Par conséquent, il semblerait que la partie de la protéine qui permet la formation de la bande discrète avec de l'ADN soit comprise dans la région allant de aa 182 à 490. La partie N-terminal de la protéine semble apporter plusieurs contributions au fonctionnement de Pf-Int. De premier abord, elle semble empêcher l'accès de l'ADN à son site de fixation. Un changement conformationnel de cette partie N-terminal pourrait rendre le site disponible.

De ce fait, il serait important de reconsidérer les expérimentations d'identification des cibles ADN qui n'ont pas permis de mettre en évidence une séquence particulière et pour lesquelles nous avons utilisé le fragment C-162. En effet, si on admet que l'accès au site de fixation de l'ADN est partiellement condamné dans le fragment C-162, et sachant que les parties N-terminales des recombinase apportent la sélectivité de celles-ci, il est probable qu'on n'a pas encore pu identifier le motif nécessaire pour la réaction. Il est possible que, contrairement aux

fragments issus de l'ADN génomique de *P. falciparum* qui couvrent toutes les possibilités présentes dans la cellule, les ADN issus de la recherche par SELEX ne présentent qu'une partie seulement de la cible et pas la totalité du site reconnu par la recombinase.

La taille des fragments retenus (250-350 pb), et la richesse en A/T de ces séquences ne permettent pas de mettre en évidence facilement une courte séquence consensus, ou deux séquences qui seraient éloignées l'une de l'autre pour reconstituer un site. Deuxièmement, la partie N-terminale semble apporter à l'interaction avec l'ADN un peu plus de sensibilité. Il se comporte ainsi comme un élément de rapprochement initial à l'ADN de nature ionique, qui nécessiterait ensuite la participation de la partie C-terminale pour la formation de complexes plus stables et protégés du solvant.

Les complexes ADN/C-192 formés dans nos conditions de travail ont permis de montrer que ce fragment (aa 192-490) précédemment prédit comme étant la région homologue aux deux domaines minimaux des recombinases à tyrosine fixe l'ADN mais de manière non spécifique. L'absence de spécificité de ce fragment de la protéine pourrait être due à plusieurs raisons. Par exemple, la surexpression de la protéine étant réalisée dans le système hétérologue d'*E. coli*, ceci pourrait mener à l'absence de certaines modifications post-traductionnelles qui pourrait avoir un impact sur la spécificité.

D'autre part, nous notons la présence de deux motifs CXXC (C₁₆₄-X-X-C₁₆₇ et C₁₇₈-X-X-C₁₈₁) séparés de 10 acides aminés. Un module supplémentaire pourrait donc être formé par un doigt de zinc de type Cys₂Cys₂ et ainsi conférer la spécificité à la protéine. La présence hypothétique de ce doigt de zinc n'avait pas été prise en compte lors des expériences de recherche de cibles et la présence d'EDTA tout le long des étapes de purification de la protéine et de la préparation de l'ADN a pu déstabiliser ce domaine et interférer ainsi avec l'identification des ADN cibles [156]. Nous proposons donc que les expériences de recherche de cibles soient reprises en prenant en compte la présence hypothétique de ce doigt de zinc. Nous proposons également l'essai d'autres conditions réactionnelles dans le but de minimiser les interactions aléatoires entre l'ADN et la protéine. Dans le but d'éviter l'effet généré par les conditions d'expérimentation sur gel, nous proposons que des essais de fixation à l'ADN en solution soient entrepris. Des expériences de résonance plasmonique de surface permettraient de comparer l'affinité des différents fragments protéiques.

➤ **Activité catalytique de Pf-Int :**

Le mécanisme de la recombinaison par la famille des Y-SSR implique la formation d'un complexe covalent transitoire au sein de la synapse de recombinaison entre la tyrosine

catalytique et l'ADN substrat. Ce complexe covalent est résistant aux conditions dénaturantes (1% SDS et chauffage à 95°C) et a pu être précédemment observé pour d'autres recombinaisons [157]. Nos essais montrent que Pf-Int est purifiée sous forme active et qu'elle est effectivement capable de former un complexe covalent. Lors de ces expériences, nous avons réussi à observer des complexes covalents avec tous les fragments testés (C-117, C-162 et C-192). Contrairement aux observations obtenues sur gel natif, nous avons été capables d'observer des complexes avec les fragments C-117 et C-162. Il semble donc qu'une fraction de ces protéines soit capable de fixer l'ADN.

Cependant les quantités de complexes sont faibles. Cela pourrait être dû à plusieurs raisons, notamment l'utilisation d'un ADN substrat formé d'un demi-site uniquement (25 pb) et ne permettant pas à la protéine de se fixer dans une conformation active. Par ailleurs le taux de conversion catalytique (turnover) des enzymes de recombinaison est particulièrement faible de part la complexité même de la réaction. Ainsi l'intégrase FLP fonctionne avec un turnover d'approximativement 0.06 sites FRT min⁻¹ à partir de jonction de Holliday correspondant [158]. Le rendement d'accumulation de produits covalents sur gel de Pf-Int pourrait être amélioré soit par l'utilisation d'ADN substrat suicide, soit d'un ADN substrat plus grand contenant deux demi-sites et permettant la formation d'une moitié de synapse classique de recombinaison. Il est également important de vérifier la formation du complexe covalent via la tyrosine catalytique par la mutation de cette dernière. La présence de complexes en conditions natives et l'absence de complexes en conditions dénaturantes en présence de ce mutant permettra de démontrer que la réaction a effectivement lieu via le résidu muté. Il sera également important de déterminer l'identité du site ADN coupé. En effet, la coïncidence de plusieurs séquences retenues en liaison covalentes, identifiées soit par séquençage ou par spectrométrie de masse permettrait de convenir d'un point critique où se trouverait le phosphate scissile de la réaction.

D'une manière générale, les travaux de thèse présentés dans ce manuscrit ont permis de décrire, pour la première fois, la recombinaison à tyrosine Pf-Int produite par le parasite *P. falciparum*. Lors de ces travaux, nous avons essayé de démontrer l'implication de cette recombinaison dans le mécanisme de variabilité antigénique nécessaire à la survie du parasite. Nos travaux nous ont permis de montrer que cette protéine n'est pas nécessaire au parasite en culture pendant la phase intra-érythrocytaire. Nous avons également montré que le gène codant pour la recombinaison est transcrit mais que la traduction de l'ARN messager est régulée. De point de vue biochimique, nous avons pu montrer que les différents fragments

purifiés de la protéine réagissent de manière différente vis-à-vis de l'ADN, montrant que le mécanisme de fixation de l'ADN dépend de la maturation ou de la réorganisation structurale de la protéine qui permet à l'ADN d'accéder au site dédié. Nos essais montrent également que le fragment prédit comme étant homologue à la protéine Cre se fixe à l'ADN de manière non spécifique. La spécificité dépend donc soit de la région N-terminale de la protéine, soit de la présence de modifications post-traductionnelles ou de l'intervention d'un ou plusieurs autres facteurs. Les expérimentations exploratoires en conditions dénaturantes nous ont permis de montrer que les protéines purifiées sont actives et qu'elles sont capables de former des complexes covalents. La recherche des cibles potentielles de Pf-Int a permis d'identifier deux groupes de séquences mais dont l'étude n'a pas permis de déterminer le motif reconnu par Pf-Int. Ces travaux de thèse ont donc permis d'apporter les premiers éléments à la description de ce nouveau membre de la famille des recombinases à tyrosine dont l'implication dans la biologie du parasite *Plasmodium* reste à déterminer. La conservation de Pf-Int chez les autres membres de la famille des *Plasmodia* peut cependant indiquer que la fonction soit importante de part le mode de vie rigoureux que constitue le parasitisme et l'économie des ressources que cela implique.

5 REFERENCES BIBLIOGRAPHIQUES

1. Dzikowski, R., T.J. Templeton, and K. Deitsch, *Variant antigen gene expression in malaria*. Cell Microbiol, 2006. 8(9): p. 1371-81.
2. WHO, *World malaria report*. 2010.
3. White, N.J., *Antimalarial drug resistance*. J Clin Invest, 2004. 113(8): p. 1084-92.
4. Adl, S.M., et al., *The new higher level classification of eukaryotes with emphasis on the taxonomy of protists*. J Eukaryot Microbiol, 2005. 52(5): p. 399-451.
5. Morrissette, N.S. and L.D. Sibley, *Cytoskeleton of apicomplexan parasites*. Microbiol Mol Biol Rev, 2002. 66(1): p. 21-38; table of contents.
6. Tilley, L., M.W. Dixon, and K. Kirk, *The Plasmodium falciparum-infected red blood cell*. Int J Biochem Cell Biol, 2011. 43(6): p. 839-42.
7. Kalanon, M. and G.I. McFadden, *Malaria, Plasmodium falciparum and its apicoplast*. Biochem Soc Trans, 2010. 38(3): p. 775-82.
8. Lim, L. and G.I. McFadden, *The evolution, metabolism and functions of the apicoplast*. Philos Trans R Soc Lond B Biol Sci, 2010. 365(1541): p. 749-63.
9. Gardner, M.J., et al., *Genome sequence of the human malaria parasite Plasmodium falciparum*. Nature, 2002. 419(6906): p. 498-511.
10. Vaidya, A.B., R. Akella, and K. Suplick, *Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite*. Mol Biochem Parasitol, 1989. 35(2): p. 97-107.
11. Silvie, O., et al., *Malaria sporozoite: migrating for a living*. Trends Mol Med, 2004. 10(3): p. 97-100; discussion 100-1.
12. Ejigiri, I. and P. Sinnis, *Plasmodium sporozoite-host interactions from the dermis to the hepatocyte*. Curr Opin Microbiol, 2009. 12(4): p. 401-7.
13. Gaur, D., D.C. Mayer, and L.H. Miller, *Parasite ligand-host receptor interactions during invasion of erythrocytes by Plasmodium merozoites*. Int J Parasitol, 2004. 34(13-14): p. 1413-29.
14. Bozdech, Z., et al., *The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum*. PLoS Biol, 2003. 1(1): p. E5.
15. Talman, A.M., et al., *Gametocytogenesis: the puberty of Plasmodium falciparum*. Malar J, 2004. 3: p. 24.
16. Scherf, A., J.J. Lopez-Rubio, and L. Riviere, *Antigenic variation in Plasmodium falciparum*. Annu Rev Microbiol, 2008. 62: p. 445-70.
17. Kuehn, A. and G. Pradel, *The coming-out of malaria gametocytes*. J Biomed Biotechnol, 2010. 2010: p. 976827.
18. Janes, J.H., et al., *Investigating the host binding signature on the Plasmodium falciparum PfEMP1 protein family*. PLoS Pathog, 2011. 7(5): p. e1002032.
19. Howitt, C.A., et al., *Clonally variant gene families in Plasmodium falciparum share a common activation factor*. Mol Microbiol, 2009. 73(6): p. 1171-85.
20. Handunnetti, S.M., K.N. Mendis, and P.H. David, *Antigenic variation of cloned Plasmodium fragile in its natural host Macaca sinica. Sequential appearance of successive variant antigenic types*. J Exp Med, 1987. 165(5): p. 1269-83.
21. Deitsch, K.W., S.A. Lukehart, and J.R. Stringer, *Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens*. Nat Rev Microbiol, 2009. 7(7): p. 493-503.
22. Recker, M., et al., *Antigenic variation in Plasmodium falciparum malaria involves a highly structured switching pattern*. PLoS Pathog, 2011. 7(3): p. e1001306.
23. Cox, H.W., *A study of relapse Plasmodium berghei infections isolated from white mice*. J Immunol, 1959. 82(3): p. 209-14.
24. Brown, K.N. and I.N. Brown, *Immunity to malaria: antigenic variation in chronic*

- infections of Plasmodium knowlesi*. Nature, 1965. 208(5017): p. 1286-8.
25. Freitas-Junior, L.H., et al., *Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites*. Cell, 2005. 121(1): p. 25-36.
 26. Lopez-Rubio, J.J., et al., *5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites*. Mol Microbiol, 2007. 66(6): p. 1296-305.
 27. Swamy, L., B. Amulic, and K.W. Deitsch, *Plasmodium falciparum var gene silencing is determined by cis DNA elements that form stable and heritable interactions*. Eukaryot Cell, 2011. 10(4): p. 530-9.
 28. Chakrabarti, K., et al., *Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis*. Rna, 2007. 13(11): p. 1923-39.
 29. Li, F., et al., *Nuclear non-coding RNAs are transcribed from the centromeres of Plasmodium falciparum and are associated with centromeric chromatin*. J Biol Chem, 2008. 283(9): p. 5692-8.
 30. Kraemer, S.M., et al., *Patterns of gene recombination shape var gene repertoires in Plasmodium falciparum: comparisons of geographically diverse isolates*. BMC Genomics, 2007. 8: p. 45.
 31. Kraemer, S.M. and J.D. Smith, *A family affair: var genes, PfEMP1 binding, and malaria disease*. Curr Opin Microbiol, 2006. 9(4): p. 374-80.
 32. Juillerat, A., et al., *Structure of a Plasmodium falciparum PfEMP1 rosetting domain reveals a role for the N-terminal segment in heparin-mediated rosette inhibition*. Proc Natl Acad Sci U S A, 2011. 108(13): p. 5243-8.
 33. Higgins, M.K., *The structure of a chondroitin sulfate-binding domain important in placental malaria*. J Biol Chem, 2008. 283(32): p. 21842-6.
 34. Rask, T.S., et al., *Plasmodium falciparum erythrocyte membrane protein 1 diversity in seven genomes--divide and conquer*. PLoS Comput Biol, 2010. 6(9).
 35. Smith, J.D., et al., *Classification of adhesive domains in the Plasmodium falciparum erythrocyte membrane protein 1 family*. Mol Biochem Parasitol, 2000. 110(2): p. 293-310.
 36. Lavstsen, T., et al., *Sub-grouping of Plasmodium falciparum 3D7 var genes based on sequence analysis of coding and non-coding regions*. Malar J, 2003. 2: p. 27.
 37. Frank, M., et al., *Frequent recombination events generate diversity within the multi-copy variant antigen gene families of Plasmodium falciparum*. Int J Parasitol, 2008. 38(10): p. 1099-109.
 38. Duffy, M.F., et al., *Ectopic recombination of a malaria var gene during mitosis associated with an altered var switch rate*. J Mol Biol, 2009. 389(3): p. 453-69.
 39. Taylor, H.M., S.A. Kyes, and C.I. Newbold, *Var gene diversity in Plasmodium falciparum is generated by frequent recombination events*. Mol Biochem Parasitol, 2000. 110(2): p. 391-7.
 40. Freitas-Junior, L.H., et al., *Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of P. falciparum*. Nature, 2000. 407(6807): p. 1018-22.
 41. Rudenko, G., *African trypanosomes: the genome and adaptations for immune evasion*. Essays Biochem, 2011. 51: p. 47-62.
 42. Morrison, L.J., L. Marcello, and R. McCulloch, *Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity*. Cell Microbiol, 2009. 11(12): p. 1724-34.
 43. Boothroyd, C.E., et al., *A yeast-endonuclease-generated DNA break induces*

- antigenic switching in Trypanosoma brucei*. Nature, 2009. 459(7244): p. 278-81.
44. Barry, D. and R. McCulloch, *Molecular microbiology: A key event in survival*. Nature, 2009. 459(7244): p. 172-3.
 45. Alsford, S., et al., *A sirtuin in the African trypanosome is involved in both DNA repair and telomeric gene silencing but is not required for antigenic variation*. Mol Microbiol, 2007. 63(3): p. 724-36.
 46. Figueiredo, L.M., C.J. Janzen, and G.A. Cross, *A histone methyltransferase modulates antigenic variation in African trypanosomes*. PLoS Biol, 2008. 6(7): p. e161.
 47. Talbert, P.B. and S. Henikoff, *Chromatin-based transcriptional punctuation*. Genes Dev, 2009. 23(9): p. 1037-41.
 48. Saraiya, A.A. and C.C. Wang, *snoRNA, a novel precursor of microRNA in Giardia lamblia*. PLoS Pathog, 2008. 4(11): p. e1000224.
 49. Carranza, P.G. and H.D. Lujan, *New insights regarding the biology of Giardia lamblia*. Microbes Infect, 2010. 12(1): p. 71-80.
 50. Kulakova, L., et al., *Epigenetic mechanisms are involved in the control of Giardia lamblia antigenic variation*. Mol Microbiol, 2006. 61(6): p. 1533-42.
 51. Pucca, C.G. and H.D. Lujan, *Antigenic variation in Giardia lamblia*. Cell Microbiol, 2009. 11(12): p. 1706-15.
 52. Brayton, K.A., et al., *Genome sequence of Babesia bovis and comparative analysis of apicomplexan hemoprotozoa*. PLoS Pathog, 2007. 3(10): p. 1401-13.
 53. Zupanska, A.K., et al., *Universal primers suitable to assess population dynamics reveal apparent mutually exclusive transcription of the Babesia bovis ves1alpha gene*. Mol Biochem Parasitol, 2009. 166(1): p. 47-53.
 54. Al-Khedery, B. and D.R. Allred, *Antigenic variation in Babesia bovis occurs through segmental gene conversion of the ves multigene family, within a bidirectional locus of active transcription*. Mol Microbiol, 2006. 59(2): p. 402-14.
 55. Ferreri, L.M., et al., *Expression and strain variation of the novel "small open reading frame" (smorf) multigene family in Babesia bovis*. Int J Parasitol, 2011.
 56. Nunes-Duby, S.E., et al., *Similarities and differences among 105 members of the Int family of site-specific recombinases*. Nucleic Acids Res, 1998. 26(2): p. 391-406.
 57. Grindley, N.D., K.L. Whiteson, and P.A. Rice, *Mechanisms of site-specific recombination*. Annu Rev Biochem, 2006. 75: p. 567-605.
 58. Nancy L. Craig, R.C., Martin Gellert, Alan M. Lambowitz, *Mobile DNA II*. American Society for Microbiology, 2002.
 59. Yang, W., *Topoisomerases and site-specific recombinases: similarities in structure and mechanism*. Crit Rev Biochem Mol Biol, 2010. 45(6): p. 520-34.
 60. McLeod, M., S. Craft, and J.R. Broach, *Identification of the crossover site during FLP-mediated recombination in the Saccharomyces cerevisiae plasmid 2 microns circle*. Mol Cell Biol, 1986. 6(10): p. 3357-67.
 61. Austin, S., M. Ziese, and N. Sternberg, *A novel role for site-specific recombination in maintenance of bacterial replicons*. Cell, 1981. 25(3): p. 729-36.
 62. Rajeev, L., K. Malanowska, and J.F. Gardner, *Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions*. Microbiol Mol Biol Rev, 2009. 73(2): p. 300-9.
 63. Buck, D. and E. Flapan, *Predicting knot or catenane type of site-specific recombination products*. J Mol Biol, 2007. 374(5): p. 1186-99.
 64. Gopaul, D.N., F. Guo, and G.D. Van Duyne, *Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination*. Embo J, 1998. 17(14): p. 4175-87.

65. Chen, Y., et al., *Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping*. *Mol Cell*, 2000. 6(4): p. 885-97.
66. MacDonald, D., et al., *Structural basis for broad DNA-specificity in integron recombination*. *Nature*, 2006. 440(7088): p. 1157-62.
67. Biswas, T., et al., *A structural basis for allosteric control of DNA recombination by lambda integrase*. *Nature*, 2005. 435(7045): p. 1059-66.
68. Hickman, A.B., et al., *Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 Å resolution*. *Cell*, 1997. 89(2): p. 227-37.
69. Kwon, H.J., et al., *Flexibility in DNA recombination: structure of the lambda integrase catalytic core*. *Science*, 1997. 276(5309): p. 126-31.
70. Sarkar, D., M. Radman-Livaja, and A. Landy, *The small DNA binding domain of lambda integrase is a context-sensitive modulator of recombinase functions*. *Embo J*, 2001. 20(5): p. 1203-12.
71. Letzelter, C., M. Duguet, and M.C. Serre, *Mutational analysis of the archaeal tyrosine recombinase SSV1 integrase suggests a mechanism of DNA cleavage in trans*. *J Biol Chem*, 2004. 279(28): p. 28936-44.
72. Perry, K., et al., *Insights from the structure of a smallpox virus topoisomerase-DNA transition state mimic*. *Structure*, 2010. 18(1): p. 127-37.
73. Champoux, J.J., *DNA topoisomerases: structure, function, and mechanism*. *Annu Rev Biochem*, 2001. 70: p. 369-413.
74. Hwang, Y., A. Burgin, Jr., and F. Bushman, *DNA contacts stimulate catalysis by a poxvirus topoisomerase*. *J Biol Chem*, 1999. 274(14): p. 9160-8.
75. Patel, A., et al., *Crystal structure of a bacterial topoisomerase IB in complex with DNA reveals a secondary DNA binding site*. *Structure*, 2010. 18(6): p. 725-33.
76. Gibb, B., et al., *Requirements for catalysis in the Cre recombinase active site*. *Nucleic Acids Res*, 2010. 38(17): p. 5817-32.
77. Krogh, B.O. and S. Shuman, *Proton relay mechanism of general acid catalysis by DNA topoisomerase IB*. *J Biol Chem*, 2002. 277(8): p. 5711-4.
78. Nagarajan, R., et al., *Catalytic phosphoryl interactions of topoisomerase IB*. *Biochemistry*, 2005. 44(34): p. 11476-85.
79. Cousineau, B., et al., *Retrotransposition of a bacterial group II intron*. *Nature*, 2000. 404(6781): p. 1018-21.
80. Eckstein, F., *NUCLEOSIDE PHOSPHOROTHIOATES*. *Annual Review of Biochemistry*, 1985. 54: p. 367-402.
81. Knowles, J.R., *Enzyme-catalyzed phosphoryl transfer reactions*. *Annu Rev Biochem*, 1980. 49: p. 877-919.
82. Lee, J. and M. Jayaram, *Role of partner homology in DNA recombination. Complementary base pairing orients the 5'-hydroxyl for strand joining during Flp site-specific recombination*. *J Biol Chem*, 1995. 270(8): p. 4042-52.
83. Mazel, D., *Integrans: agents of bacterial evolution*. *Nat Rev Microbiol*, 2006. 4(8): p. 608-20.
84. Summers, D.K., *Derivatives of ColE1 *cer* show altered topological specificity in site-specific recombination*. *Embo J*, 1989. 8(1): p. 309-15.
85. Alen, C., D.J. Sherratt, and S.D. Colloms, *Direct interaction of aminopeptidase A with recombination site DNA in Xer site-specific recombination*. *Embo J*, 1997. 16(17): p. 5188-97.
86. Colloms, S.D., et al., *Xer-mediated site-specific recombination in vitro*. *Embo J*, 1996. 15(5): p. 1172-81.
87. Radman-Livaja, M., et al., *Arm sequences contribute to the architecture and*

- catalytic function of a lambda integrase-Holliday junction complex.* Mol Cell, 2003. 11(3): p. 783-94.
88. Lee, J. and M. Jayaram, *A tetramer of the Flp recombinase silences the trimers within it during resolution of a Holliday junction substrate.* Genes Dev, 1997. 11(18): p. 2438-47.
 89. Woods, K.C., et al., *Quasi-equivalence in site-specific recombinase structure and function: crystal structure and activity of trimeric Cre recombinase bound to a three-way Lox DNA junction.* J Mol Biol, 2001. 313(1): p. 49-69.
 90. Chen, Y. and P.A. Rice, *New insight into site-specific recombination from Flp recombinase-DNA structures.* Annu Rev Biophys Biomol Struct, 2003. 32: p. 135-59.
 91. Ferreira, H., et al., *Functional analysis of the C-terminal domains of the site-specific recombinases XerC and XerD.* J Mol Biol, 2003. 330(1): p. 15-27.
 92. Segall, A.M. and H.A. Nash, *Architectural flexibility in lambda site-specific recombination: three alternate conformations channel the attL site into three distinct pathways.* Genes Cells, 1996. 1(5): p. 453-63.
 93. Landy, A., *Dynamic, structural, and regulatory aspects of lambda site-specific recombination.* Annu Rev Biochem, 1989. 58: p. 913-49.
 94. Van Duyne, G.D., *Lambda integrase: armed for recombination.* Curr Biol, 2005. 15(17): p. R658-60.
 95. Cambray, G., A.M. Guerout, and D. Mazel, *Integrans.* Annu Rev Genet, 2010. 44: p. 141-66.
 96. Mazel, D., et al., *A distinctive class of integron in the Vibrio cholerae genome.* Science, 1998. 280(5363): p. 605-8.
 97. Stokes, H.W. and R.M. Hall, *A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrans.* Mol Microbiol, 1989. 3(12): p. 1669-83.
 98. Collis, C.M., et al., *Site-specific insertion of gene cassettes into integrans.* Mol Microbiol, 1993. 9(1): p. 41-52.
 99. Fluit, A.C. and F.J. Schmitz, *Resistance integrans and super-integrans.* Clin Microbiol Infect, 2004. 10(4): p. 272-88.
 100. Bouvier, M., G. Demarre, and D. Mazel, *Integron cassette insertion: a recombination process involving a folded single strand substrate.* Embo J, 2005. 24(24): p. 4356-67.
 101. Hall, R.M., D.E. Brookes, and H.W. Stokes, *Site-specific insertion of genes into integrans: role of the 59-base element and determination of the recombination cross-over point.* Mol Microbiol, 1991. 5(8): p. 1941-59.
 102. Frumerie, C., et al., *The relaxed requirements of the integron cleavage site allow predictable changes in integron target specificity.* Nucleic Acids Res, 2010. 38(2): p. 559-69.
 103. Bouvier, M., et al., *Structural features of single-stranded integron cassette attC sites and their role in strand selection.* PLoS Genet, 2009. 5(9): p. e1000632.
 104. Zieg, J. and M. Simon, *Analysis of the nucleotide sequence of an invertible controlling element.* Proc Natl Acad Sci U S A, 1980. 77(7): p. 4196-200.
 105. Gally, D.L., J. Leathart, and I.C. Blomfield, *Interaction of FimB and FimE with the fim switch that controls the phase variation of type 1 fimbriae in Escherichia coli K-12.* Mol Microbiol, 1996. 21(4): p. 725-38.
 106. Zhang, J.R., et al., *Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes.* Cell, 1997. 89(2): p. 275-85.
 107. Zhang, J.R. and S.J. Norris, *Genetic variation of the Borrelia burgdorferi gene*

- vlsE* involves cassette-specific, segmental gene conversion. *Infect Immun*, 1998. 66(8): p. 3698-704.
108. Hinnebusch, B.J., et al., *Population structure of the relapsing fever spirochete Borrelia hermsii as indicated by polymorphism of two multigene families that encode immunogenic outer surface lipoproteins*. *Infect Immun*, 1998. 66(2): p. 432-40.
 109. Chaconas, G., *Hairpin telomeres and genome plasticity in Borrelia: all mixed up in the end*. *Mol Microbiol*, 2005. 58(3): p. 625-35.
 110. Bankhead, T. and G. Chaconas, *Mixing active-site components: a recipe for the unique enzymatic activity of a telomere resolvase*. *Proc Natl Acad Sci U S A*, 2004. 101(38): p. 13768-73.
 111. Kobryn, K., A.B. Burgin, and G. Chaconas, *Uncoupling the chemical steps of telomere resolution by ResT*. *J Biol Chem*, 2005. 280(29): p. 26788-95.
 112. Komano, T., *Shufflons: multiple inversion systems and integrons*. *Annu Rev Genet*, 1999. 33: p. 171-91.
 113. Gyohda, A., et al., *Asymmetry of shufflon-specific recombination sites in plasmid R64 inhibits recombination between direct sfx sequences*. *J Biol Chem*, 2006. 281(30): p. 20772-9.
 114. Buchan, D.W., et al., *Protein annotation and modelling servers at University College London*. *Nucleic Acids Res*, 2010. 38(Web Server issue): p. W563-8.
 115. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices*. *J Mol Biol*, 1999. 292(2): p. 195-202.
 116. Kelley, L.A. and M.J. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server*. *Nat Protoc*, 2009. 4(3): p. 363-71.
 117. McGuffin, L.J. and D.T. Jones, *Improvement of the GenTHREADER method for genomic fold recognition*. *Bioinformatics*, 2003. 19(7): p. 874-81.
 118. Lobley, A., M.I. Sadowski, and D.T. Jones, *pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination*. *Bioinformatics*, 2009. 25(14): p. 1761-7.
 119. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0*. *Bioinformatics*, 2007. 23(21): p. 2947-8.
 120. Warburton, P.E., et al., *Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes*. *Genome Res*, 2004. 14(10A): p. 1861-9.
 121. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in biopolymers*. *Proc Int Conf Intell Syst Mol Biol*, 1994. 2: p. 28-36.
 122. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite*. *Trends Genet*, 2000. 16(6): p. 276-7.
 123. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction*. *Nucleic Acids Res*, 2003. 31(13): p. 3406-15.
 124. Smith, C., et al., *Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA*. *Nucleic Acids Res*, 2010. 38(Web Server issue): p. W373-7.
 125. Nkrumah, L.J., et al., *Efficient site-specific integration in Plasmodium falciparum chromosomes mediated by mycobacteriophage Bxb1 integrase*. *Nat Methods*, 2006. 3(8): p. 615-21.
 126. Sreerama, N. and R.W. Woody, *Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set*. *Anal Biochem*, 2000. 287(2): p. 252-60.

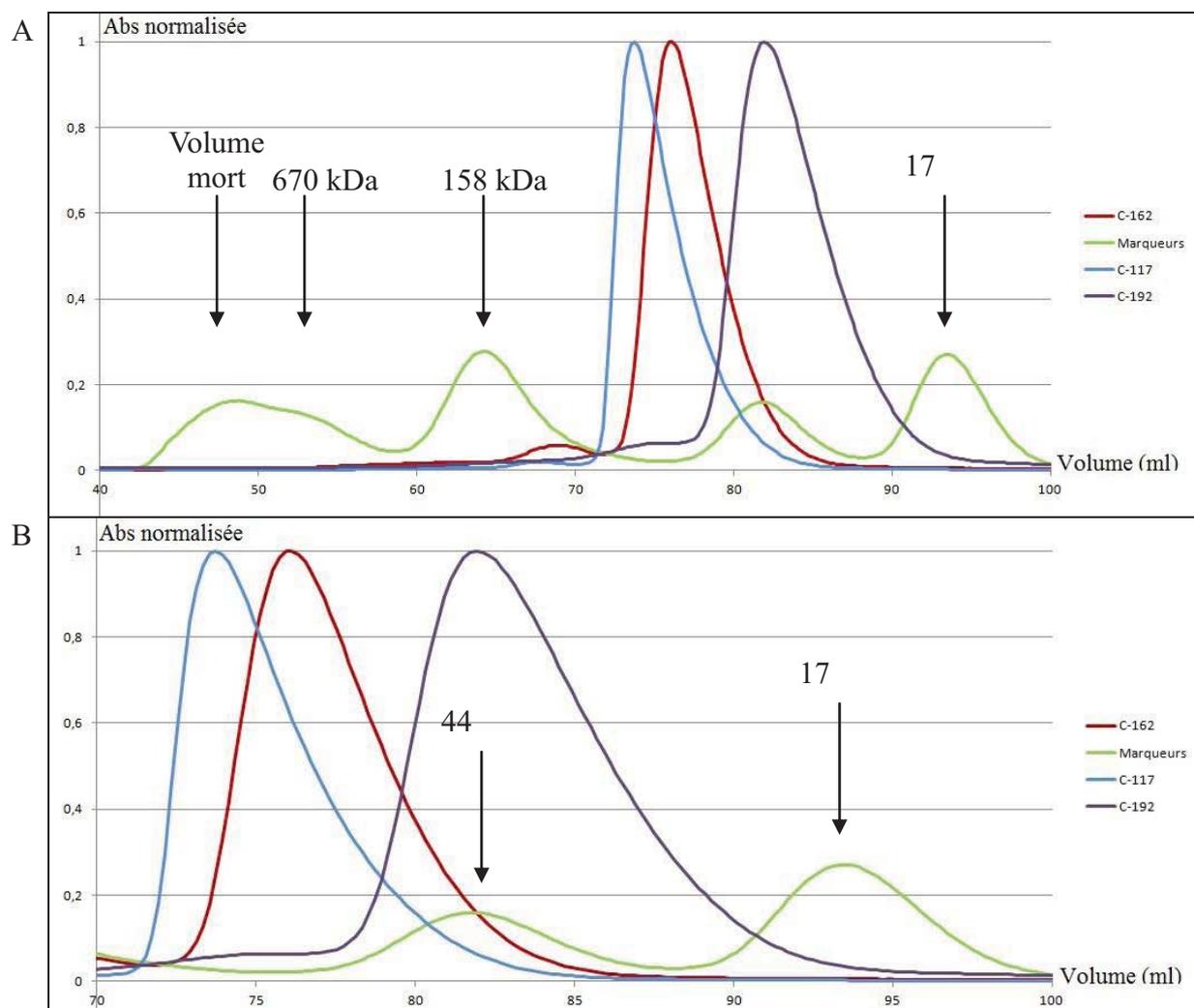
127. Whitmore, L. and B.A. Wallace, *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data*. *Nucleic Acids Res*, 2004. 32(Web Server issue): p. W668-73.
128. Whitmore, L. and B.A. Wallace, *Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases*. *Biopolymers*, 2008. 89(5): p. 392-400.
129. Johnson, W.C., *Analyzing protein circular dichroism spectra for accurate secondary structures*. *Proteins*, 1999. 35(3): p. 307-12.
130. Lees, J.G., et al., *A reference database for circular dichroism spectroscopy covering fold and secondary structure space*. *Bioinformatics*, 2006. 22(16): p. 1955-62.
131. Schuck, P., *Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling*. *Biophys J*. 78(3): p. 1606-19.
132. Lebowitz, J., M.S. Lewis, and P. Schuck, *Modern analytical ultracentrifugation in protein science: a tutorial review*. *Protein Sci*. 11(9): p. 2067-79.
133. De Silva, E.K., et al., *Specific DNA-binding by apicomplexan AP2 transcription factors*. *Proc Natl Acad Sci U S A*, 2008. 105(24): p. 8393-8.
134. Andrew P. Waters, C.J.J., *Malaria parasites, genomes and molecular biology*. Horizon Scientific Press, 2004.
135. Loonstra, A., et al., *Growth inhibition and DNA damage induced by Cre recombinase in mammalian cells*. *Proc Natl Acad Sci U S A*, 2001. 98(16): p. 9209-14.
136. Pino, P., et al., *Dual targeting of antioxidant and metabolic enzymes to the mitochondrion and the apicoplast of Toxoplasma gondii*. *PLoS Pathog*, 2007. 3(8): p. e115.
137. Kehr, S., et al., *Compartmentation of redox metabolism in malaria parasites*. *PLoS Pathog*, 2010. 6(12): p. e1001242.
138. Puyaubert, J., L. Denis, and C. Alban, *Dual targeting of Arabidopsis holocarboxylase synthetase1: a small upstream open reading frame regulates translation initiation and protein targeting*. *Plant Physiol*, 2008. 146(2): p. 478-91.
139. Ragheb, D., et al., *Distribution and biochemical properties of an MI-family aminopeptidase in Plasmodium falciparum indicate a role in vacuolar hemoglobin catabolism*. *J Biol Chem*, 2011. 286(31): p. 27255-65.
140. Anderson, M.E., et al., *Reciprocal interference between the sequence-specific core and nonspecific C-terminal DNA binding domains of p53: implications for regulation*. *Mol Cell Biol*, 1997. 17(11): p. 6255-64.
141. Ross, W. and A. Landy, *Patterns of lambda Int recognition in the regions of strand exchange*. *Cell*, 1983. 33(1): p. 261-72.
142. Ringrose, L., et al., *Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination*. *J Mol Biol*, 1998. 284(2): p. 363-84.
143. Le Roch, K.G., et al., *Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle*. *Genome Res*, 2004. 14(11): p. 2308-18.
144. Mair, G.R., et al., *Universal features of post-transcriptional gene regulation are critical for Plasmodium zygote development*. *PLoS Pathog*, 2010. 6(2): p. e1000767.
145. Esposito, D. and J.J. Scoocca, *The integrase family of tyrosine recombinases: evolution of a conserved active site domain*. *Nucleic Acids Res*, 1997. 25(18): p. 3605-14.
146. Tonkin, C.J., M. Kalanon, and G.I. McFadden, *Protein targeting to the malaria parasite plastid*. *Traffic*, 2008. 9(2): p. 166-75.
147. Foth, B.J., et al., *Dissecting apicoplast targeting in the malaria parasite*

- Plasmodium falciparum*. Science, 2003. 299(5607): p. 705-8.
148. Waller, R.F., et al., *Nuclear-encoded proteins target to the plastid in Toxoplasma gondii and Plasmodium falciparum*. Proc Natl Acad Sci U S A, 1998. 95(21): p. 12352-7.
 149. Zuegge, J., et al., *Deciphering apicoplast targeting signals--feature extraction from nuclear-encoded precursors of Plasmodium falciparum apicoplast proteins*. Gene, 2001. 280(1-2): p. 19-26.
 150. Ram, E.V., et al., *DNA organization by the apicoplast-targeted bacterial histone-like protein of Plasmodium falciparum*. Nucleic Acids Res, 2008. 36(15): p. 5061-73.
 151. Tosh, K., et al., *Plasmodium falciparum: stage-related expression of topoisomerase I*. Exp Parasitol, 1999. 91(2): p. 126-32.
 152. Wojciak, J.M., K.M. Connolly, and R.T. Clubb, *NMR structure of the Tn916 integrase-DNA complex*. Nat Struct Biol, 1999. 6(4): p. 366-73.
 153. Esposito, D., J.S. Thrower, and J.J. Scocca, *Protein and DNA requirements of the bacteriophage HP1 recombination system: a model for intasome formation*. Nucleic Acids Res, 2001. 29(19): p. 3955-64.
 154. Ton-Hoang, B., et al., *Single-stranded DNA transposition is coupled to host replication*. Cell, 2010. 142(3): p. 398-408.
 155. Ton-Hoang, B., et al., *Transposition of ISHp608, member of an unusual family of bacterial insertion sequences*. Embo J, 2005. 24(18): p. 3325-38.
 156. Nyborg, J.K. and O.B. Peersen, *That zincing feeling: the effects of EDTA on the behaviour of zinc-binding transcriptional regulators*. Biochem J, 2004. 381(Pt 3): p. e3-4.
 157. Serre, M.C., et al., *Cleavage properties of an archaeal site-specific recombinase, the SSV1 integrase*. J Biol Chem, 2002. 277(19): p. 16758-67.
 158. Meyer-Leon, L., R.B. Inman, and M.M. Cox, *Characterization of Holliday structures in FLP protein-promoted site-specific recombination*. Mol Cell Biol, 1990. 10(1): p. 235-42.
 159. Cowman, A.F. and B.S. Crabb, *Invasion of red blood cells by malaria parasites*. Cell, 2006. 124(4): p. 755-66.
 160. Coatney, G., W. Collins, M. Warren, and P. Contacos., *The Primate Malariae*. Bethesda: U.S. Department of Health, Education and Welfare., 1971.
 161. Gonnet, G.H., M.A. Cohen, and S.A. Benner, *Exhaustive matching of the entire protein sequence database*. Science, 1992. 256(5062): p. 1443-5.

6 ANNEXES

Annexe 1 :

Profils d'élutions de la colonne de chromatographie d'exclusion stérique de la protéine des protéines C-117, C-162 et C-192.



Profils d'élutions de la colonne de chromatographie d'exclusion stérique de la protéine des protéines C-117, C-162 et C-192. Les courbes représentent la variation de l'absorbance DO_{280nm} normalisée en fonction du volume d'élution (millilitre). **A** : Profils d'élution entre 40 et 100 ml. **B** : Grossissement de la partie allant de 70 à 100 ml. Courbe en vert : marqueurs de masse moléculaire (les pics d'élution sont indiqués par des flèches noires) : Thyroglobulin (670,000 kDa), γ -globulin (158,000 kDa), Ovalbumin (44,000 kDa) et Myoglobine (17,000 kDa). Rouge : C-162. Bleu : C-117. Mauve : C-192. Les profils d'élution (Partie B) montrent que les courbes sont asymétriques suggérant que les protéines interagissent avec la matrice de la colonne.

Annexe 2

Tableau récapitulatif des séquences cibles potentielles de Pf-Int. Ces séquences ont été identifiées par affinité à partir de fragments de l'ADN génomique de *P. falciparum* et de la protéine C-162 pourvue d'une étiquette 6-His en N-terminal.

Chromosome	Position sur le chromosome	Fonction selon PlasmoDB
chr\$1	51450 to 51690	Rifin
chr\$1	60690 to 61040	Rifin
chr\$1	350430 to 350730	Conserved protein unknown function
chr\$1	478050 to 478290	rrna encoding
chr\$1	556710 to 557130	Close to pfemP1
chr\$1	599620 to 599720	Close to Stevor
chr\$2	33560 to 33860	Rifin
chr\$2	54400 to 54740	Stevor
chr\$2	787000 to 787500	PHIST domain protein
chr\$2	889860 to 890280	Rifin
chr\$2	896800 to 897050	Rifin
chr\$2	902500 to 902560	Conserved protein
chr\$2	913650 to 914200	Rifin
chr\$3	58800 to 59200	Rifin
chr\$3	122460 to 122730	pfemP1 pseudo gene
chr\$3	191090 to 191170	IBR domain protein
chr\$3	232920 to 233070	Conserved protein unknown function
chr\$3	233670 to 233730	Conserved protein unknown function
chr\$3	233970 to 234030	Conserved protein unknown function
chr\$3	234270 to 234330	Conserved protein unknown function
chr\$3	1012980 to 1013580	Rifin
chr\$3	1015920 to 1016340	Rifin
chr\$3	1024920 to 1025160	Rifin
chr\$4	38100 to 38700	pfemP1
chr\$4	39300 to 39900	pfemP1
chr\$4	40200 to 40800	pfemP1
chr\$4	56700 to 57000	pfemP1
chr\$4	58200 to 58800	pfemP1
chr\$4	60600 to 60900	pfemP1
chr\$4	86300 to 86800	Rifin
chr\$4	88900 to 89500	Rifin
chr\$4	92540 to 92890	Rifin
chr\$4	95830 to 96460	Stevor
chr\$4	133560 to 133840	Surface asso. interspersed gene Surfin 4.1
chr\$4	133910 to 134190	Surface asso. interspersed gene Surfin 4.1
chr\$4	162500 to 163400	Rifin
chr\$4	165690 to 166250	Stevor
chr\$4	171900 to 172400	pfemP1 upstream
chr\$4	172900 to 173500	pfemP1
chr\$4	553140 to 553280	pfemP1
chr\$4	553490 to 553560	pfemP1
chr\$4	619530 to 619830	pfemP1 upstream

chr\$4	621000 to 621240	pfemP1 downstream
chr\$4	942550 to 942620	pfemP1
chr\$4	942830 to 942900	pfemP1
chr\$4	943180 to 943250	pfemP1
chr\$4	943460 to 943530	pfemP1
chr\$5	29900 to 30400	Rifin
chr\$5	545800 to 546120	Rab1b GTPase
chr\$5	773280 to 773600	Intergenic region
chr\$5	790400 to 790680	Mito. ribosomal prot L14
chr\$5	857300 to 857700	S-adenyl cystein lyase
chr\$5	1301280 to 1301760	Early transcribed Mb prot. 5 ETRAMP 5
chr\$5	1333400 to 1333520	pfemP1 Truncated
chr\$6	101800 to 102400	Elongation factor G
chr\$6	399800 to 399880	Mito. cardiolipin synthase
chr\$6	493110 to 493200	potential conserved protein
chr\$6	493700 to 493800	potential conserved protein
chr\$6	558440 to 558560	Intergenic region
chr\$6	623200 to 623500	enoyl acyl carrier reductase
chr\$6	731430 to 731970	pfemP1
chr\$6	735840 to 736380	Stevor
chr\$6	738360 to 739080	Stevor
chr\$6	741510 to 741690	Rifin
chr\$6	741960 to 742230	Rifin
chr\$6	880290 to 880390	Conserved Apicomplexan protein pfs77
chr\$6	880560 to 880700	Conserved Apicomplexan protein pfs77
chr\$6	1333080 to 1333350	Stevor
chr\$6	1333620 to 1333890	Stevor
chr\$6	1336300 to 1336700	Rifin
chr\$6	1336800 to 1337200	Rifin
chr\$1	60600 to 61150	Rifin
chr\$1	478000 to 478500	rrna encoding
chr\$2	204000 to 204500	5'-3' exonuclease
chr\$2	287800 to 288100	Intergenic region
chr\$3	232900 to 233100	conserved protein
chr\$3	366900 to 367500	HSP-90 homolog_1
chr\$3	798900 to 799100	Ubiquitin ligase
chr\$3	854400 to 854800	Intergenic region
chr\$3	855000 to 855400	Intergenic region
chr\$4	95800 to 96400	Stevor
chr\$4	157000 to 157500	Intergenic region
chr\$4	166000 to 166600	Stevor
chr\$4	171880 to 172400	Intergenic region
chr\$4	620900 to 621300	PfemP1
chr\$5	712700 to 712900	60S ribosomal protein L8
chr\$5	790390 to 790700	Mito. ribosomal protein L14
chr\$6	204500 to 204700	Glyoxalase 1
chr\$6	399790 to 399900	Mito. cardiolipin synthase
chr\$6	558300 to 558400	Intergenic region

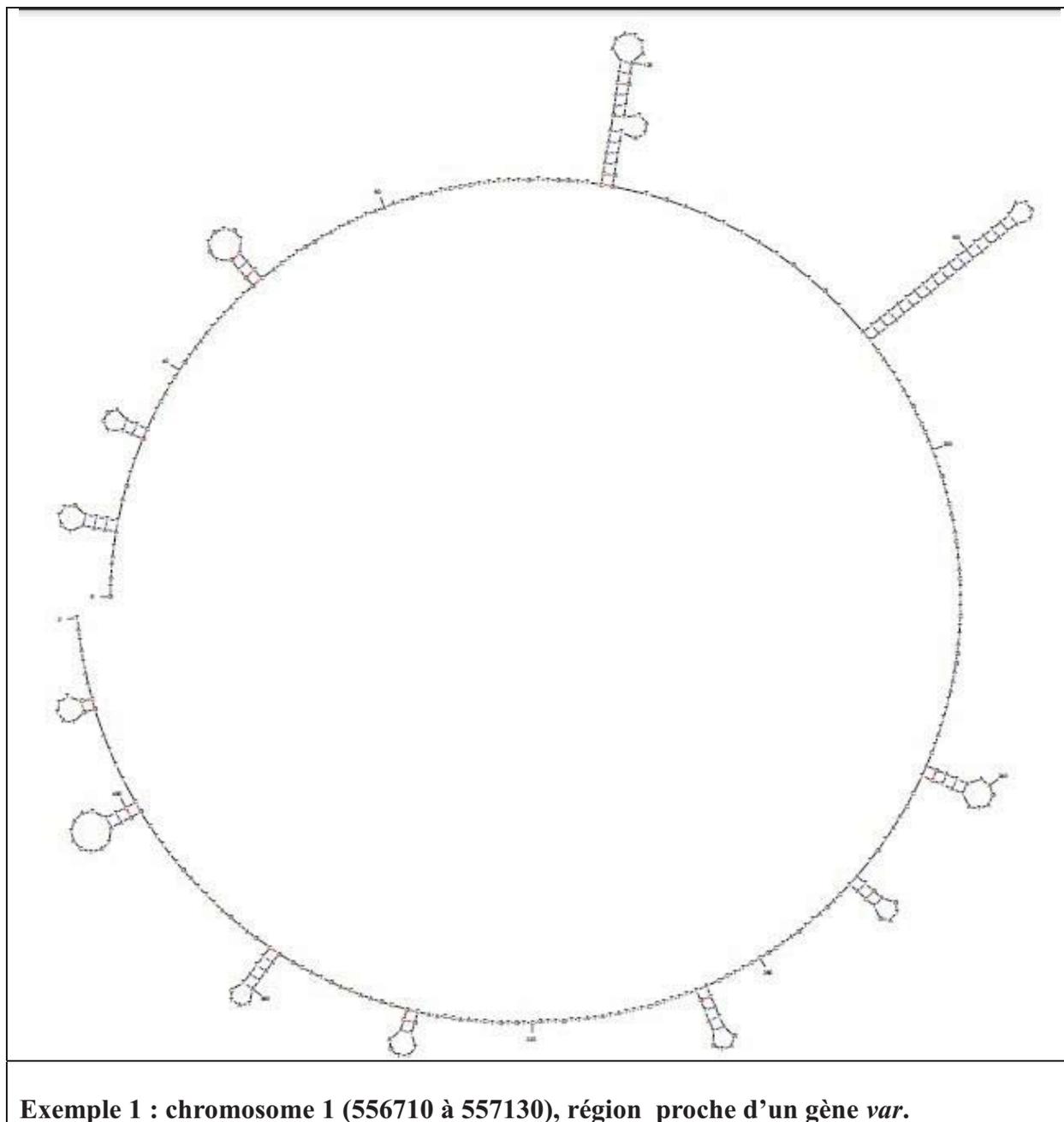
chr\$6	558420 to 558600	Intergenic region
chr\$6	723000 to 723200	PfemP1
chr\$6	912125 to 912650	RNA methyl transferase
chr\$6	1357200 to 1357700	pfemp1
chr\$7	428100 to 428450	DNA dep. RNA pol 2
chr\$7	574800 to 575150	pfemp1 region
chr\$7	653100 to 653400	pfemp1 region
chr\$7	742200 to 742500	cal-cam kinase
chr\$8	94100 to 94500	non protein coding
chr\$8	94900 to 95200	non protein coding
chr\$8	95200 to 95600	Intergenic region
chr\$8	95740 to 95980	non protein coding
chr\$8	320070 to 320300	Intergenic region
chr\$9	318220 to 318700	Arginase
chr\$9	440620 to 440980	FHA domain protein
chr\$9	880660 to 881100	conserved protein unknown function
chr\$9	1459320 to 1459780	PHIST domain protein
chr\$10	60700 to 61250	pfemp1 truncated
chr\$10	261850 to 262150	Intergenic region
chr\$10	266190 to 266330	conserved protein unknown function
chr\$10	266350 to 266550	conserved protein unknown function
chr\$10	1487270 to 1487500	Intergenic region
chr\$11	131570 to 131800	Early transcribed mb protein 11.2
chr\$11	584250 to 584500	Falcipain-3
chr\$12	57110 to 57350	pfemp1
chr\$12	764210 to 764380	close to pfemp1
chr\$12	1691180 to 1691600	close to pfemp1
chr\$12	1692610 to 1692850	close to pfemp1
chr\$12	1817360 to 1817600	Intergenic region
chr\$12	1983600 to 1983980	PF drosophila nmda1 protein
chr\$12	2190930 to 2191500	Rifin
chr\$13	525310 to 525540	Intergenic region
chr\$13	911830 to 912100	conserved protein unknown function
chr\$13	1044520 to 1044670	Intergenic region
chr\$13	1044740 to 1044850	PRPP synthase
chr\$13	1309750 to 1310130	Signal peptidase
chr\$13	1415230 to 1415550	MSP7 -like
chr\$13	2512810 to 2512960	MAL13P1
chr\$13	2659710 to 2659820	SRP receptor alpha SU
chr\$13	2833750 to 2833850	Rifin
chr\$13	2834230 to 2834420	Rifin-Stevor
chr\$14	39930 to 40210	Intergenic region
chr\$14	46990 to 47500	PEXEL
chr\$14	373010 to 373190	PF14_0093_rev
chr\$14	516070 to 516230	Intergenic region
chr\$14	516550 to 516650	N-myristoyl tfase
chr\$14	732000 to 732650	conserved protein unknown function
chr\$14	739350 to 739700	conserved protein unknown function

chr\$14	958250 to 958410	Intergenic region
chr\$14	1094240 to 1094500	Intergenic region
chr\$14	1673290 to 1673410	PF14_0390_rev
chr\$14	1681550 to 1681790	ser/thr kinase
chr\$14	2014590 to 2014790	conserved protein unknown function
chr\$14	2553570 to 2553810	cytochrome C1 precursor
chr\$14	2657350 to 2657590	PF14_0624_rev
chr\$14	2768330 to 2768570	Intergenic region
chr\$14	2899720 to 2899910	Intergenic region
chr\$14	3221030 to 3221510	PF14_0755_changed
chr\$14	3272650 to 3272910	Rifin

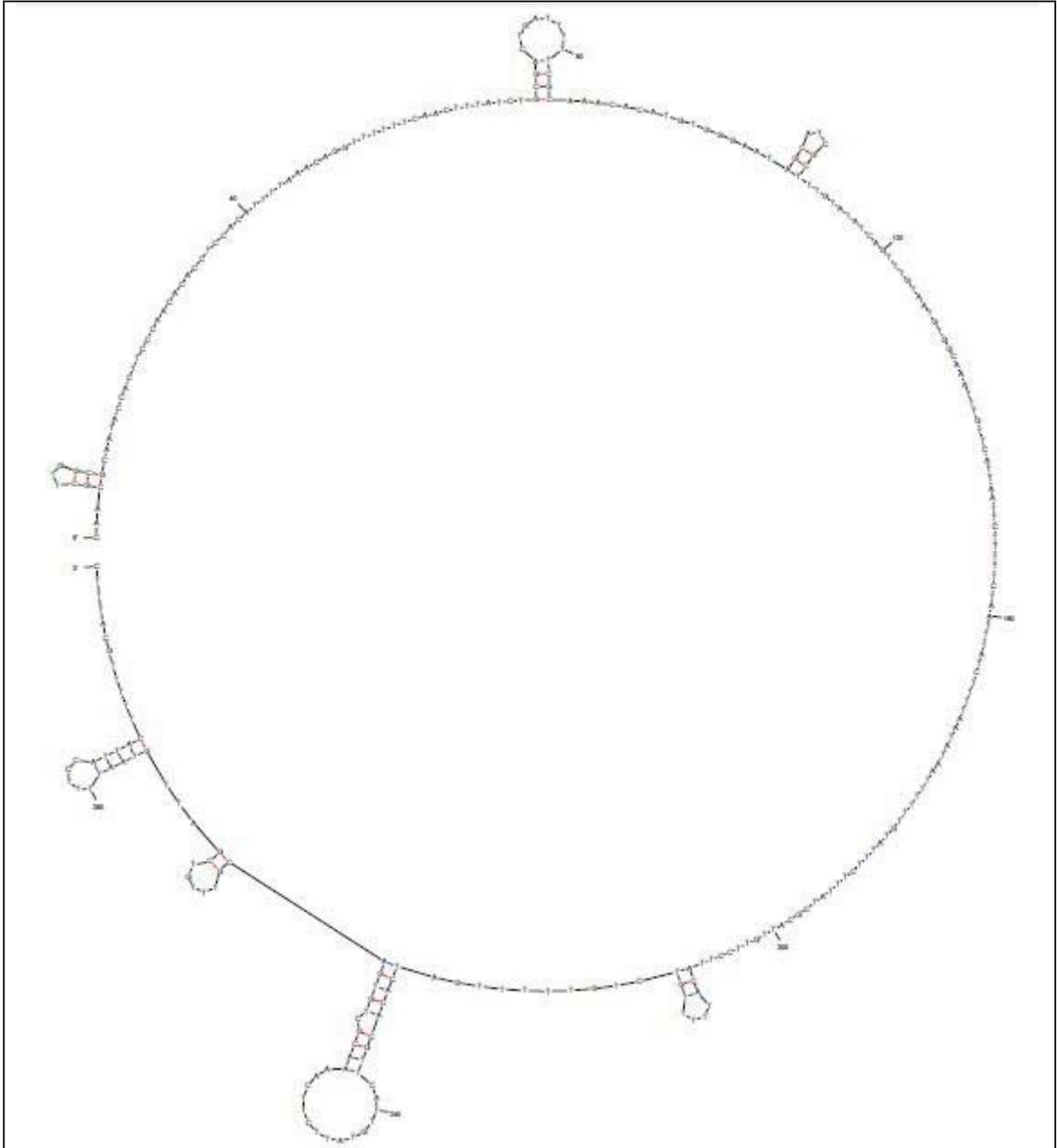
Annexe 2 : Tableau récapitulatif des séquences cibles potentielles de Pf-Int. Ces séquences ont été identifiées par affinité à partir de fragments de l'ADN génomique de *P. falciparum* et de la protéine C-162 pourvue d'une étiquette 6-His en C-terminal.

Annexe 3

Exemples de structures secondaires obtenues à partir de l'analyse par mFOLD des séquences d'ADN génomique.



B



Exemple 2 : chromosome 2 (33560 to 33860), gène *rif*. Pour cet ADN deux structures sont possibles (A et B).

Résumé de la thèse :

L'enzyme Pf-Int est un enzyme appartenant à la famille des recombinaases à site spécifique de type tyrosine, présent chez le parasite *P. falciparum*, qui a été récemment identifié dans le laboratoire. Cette recombinaase conservée chez les *Plasmodia* joue potentiellement un rôle clé dans le système de recombinaison nécessaire à la viabilité du parasite. Dans ces travaux de thèse, nous avons réalisé des études fonctionnelles (étude du niveau de transcription, étude d'un mutant K.O. et des essais de sur-expression de Pf-Int dans le parasite) ainsi que des études biochimiques et structurales (sur-expression hétérologue, purification de différents fragments de Pf-Int, caractérisation biophysique et structurale de ces fragments et des essais de détermination des cibles ADN de cette recombinaase.

Mots clés :

Plasmodium falciparum, recombinaison génétique à site spécifique, tyrosine recombinaase, évolution.

Laboratoires de rattachement :

Unité : Immunologie structurale, Institut Pasteur, CNRS URA 2185, 25-28 Rue du Dr. Roux 75724
Paris CEDEX 15

Unité : Plasticité du Génome Bactérien, Institut Pasteur, CNRS UMR 3525, 25-28 Rue du Dr. Roux
75724 Paris CEDEX 15

PÔLE : INGENIERIE DES PROTEINES ET CIBLES THERAPEUTIQUES

UNIVERSITÉ PARIS-SUD 11
UFR «FACULTÉ DE PHARMACIE DE CHATENAY-MALABRY »
5, rue Jean Baptiste Clément
92296 CHÂTENAY-MALABRY CEDEX