



HAL
open science

Mécanisme moléculaire de reconnaissance et de clivage du génome chez le bactériophage SPP1, un virus à ADN double-brin

Karima Djacem

► **To cite this version:**

Karima Djacem. Mécanisme moléculaire de reconnaissance et de clivage du génome chez le bactériophage SPP1, un virus à ADN double-brin. Biologie moléculaire. Université Paris-Saclay, 2016. Français. NNT : 2016SACLS462 . tel-01459676

HAL Id: tel-01459676

<https://theses.hal.science/tel-01459676>

Submitted on 7 Feb 2017

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.

NNT : 2016SACLS462

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE A
"L'UNIVERSITE PARIS-SUD"

ECOLE DOCTORALE N° (569)
Innovation thérapeutique du fondamental à l'appliquée
Spécialité de doctorat (Biochimie et biologie structurale)

Par

Mme Karima Djacem

Mécanisme moléculaire de reconnaissance et clivage du génome du bactériophage
SPP1, un virus à ADN double-brin

Thèse présentée et soutenue à « Gif-sur-Yvette », le « 08 Décembre 2016 » :

Composition du Jury :

Mr Michael Dubow	Professeur UPSud, Responsable, CNRS, I2BC, Orsay	Président du jury
Mme Mireille Ansaldi	Directrice de Recherche, CNRS, Aix-Marseille Université	Rapporteuse
Mr Eduardo Rocha	Directeur de Recherche, Institut Pasteur, Paris	Rapporteur
Mr David Bikard	Responsable d'équipe, Institut Pasteur, Paris	Examineur
Mme Leonor Oliveira	Chargée de Recherche, CNRS, I2BC, Gif-sur-Yvette	Directrice de thèse
Mr Paulo Tavares	Directeur de recherche, CNRS, I2BC, Gif-sur-Yvette	Directeur de thèse

Remerciements

En préambule de cette thèse, je tiens sincèrement à remercier toutes les personnes qui m'ont apportée leur soutien dans l'élaboration de ce travail de recherche. À l'issue de cette rédaction, je me rends compte que le doctorat n'est pas un travail de solitaire mais le résultat de l'intérêt et de l'implication de toutes les personnes qui m'ont aidée dans sa réalisation.

C'est donc naturellement que je commencerai par remercier les deux autres personnes à qui cette thèse appartient aussi.

En premier lieu, je tiens à remercier ma directrice de thèse, Leonor Oliveira, d'avoir suggéré ce projet et de me l'avoir confiée avec confiance pendant ces 4 dernières années. Je la remercie en particulier pour sa disponibilité et pour toutes ces longues discussions qui ont permis au projet d'évoluer et de prendre des chemins parfois imprévus, pour ses conseils, ses justes remarques et son soutien.

En second lieu, je voudrai également remercier mon « autre » directeur de thèse, Paulo Tavares, de m'avoir accueillie la première fois dans son équipe il y a quelques années de cela, de m'avoir encadrée en Master mais aussi d'avoir supervisé ce projet. Je le remercie de m'avoir toujours poussée à développer mon sens critique, à m'améliorer, à participer aux congrès, aux journées scientifiques, aux journées doctorales et j'en passe! Et surtout de m'avoir rapidement poussée du plongeur concernant la pratique de l'anglais.

Je suis fière d'avoir été formée par vous deux.

Je remercie le Dr Mireille Ansaldi ainsi que le Dr Eduardo Rocha d'avoir accepté de lire mon travail, de le juger et d'en être les rapporteurs. Je remercie également le Dr David Bikard et le Pr Michael Dubow d'avoir accepté de faire partie du jury.

Mes remerciements s'adressent également à l'équipe de la plateforme de séquençage haut débit de l'I2BC, sous la responsabilité du Dr Claude Thermes. Je remercie en particulier Yan Jaszczyszyn pour les discussions sur la méthodologie et la préparation des échantillons, Delphine Naquin et Yves d'Aubenton-Carafa pour les traitements des données et la préparation des figures. Je les remercie surtout pour leur disponibilité, les réunions de dernières minutes et les discussions autour de ces travaux.

Je remercie le Pr Catherine Berrier de l'institut des Neurosciences Paris-Saclay, qui m'a fournie le système « RTS 100 *E. coli* HY » pour la production *in vitro* des protéines et les essais de clivage *pac in vitro* ainsi que le vecteur piVex 2.3d. Je la remercie de m'avoir accueillie dans son laboratoire et de m'avoir expliquée la procédure.

Merci aux Drs Pascale Boulanger et Michael Chandler d'avoir accepté de juger mon travail de mi-thèse et d'avoir soulevé des questions qui m'ont permis de mieux appréhender certaines problématiques.

Je tiens à remercier Charlène Cornilleau de m'avoir formée à acquérir les bonnes pratiques de laboratoires à mon arrivée ainsi que Sandrine Brasilès qui a pris le relais.

Merci à la T4 ligase et à la pfu Turbo d'avoir toujours, à quelques exceptions près, répondu présentes,

grâce à vous, pleins de petits pKD existent. Merci à la région *pacL* qui, au bout d'une année a fini par coopérer. Je remercie moins la protéine gp1, avec qui les négociations ont été très difficiles.

Je remercie tous les collègues du département de Virologie, notamment Nolwenn, Julie, Yovan et Audrey, pour toutes les pauses café. Et les nouveaux arrivants, Lia et Andres à qui je souhaite d'aussi bonnes années que moi dans cette équipe.

À toute l'équipe « SPP1 » ; Audrey, Isabelle, Odile, Leonor, Paulo, Sandrine, ainsi que ceux qui sont partis; Virginija, Mohammed, Martha et Charlène, je vous remercie pour tous les bons moments, les fous rires, les pots dans la bibliothèque, les conseils, et surtout le soutien. Je suis contente d'avoir partagé avec vous, bureau, labo et pipettes ! Je remercie ceux d'entre vous qui m'ont encouragée pendant la rédaction. Vous avez été comme une famille en suivant toutes mes péripéties de paperasses et mes poisses, mais c'est aussi avec vous que j'ai fêté les bonnes nouvelles.

Je ne peux finir sans remercier toutes les personnes qui se sont retrouvées, de manière fortuite ou non, pour le meilleur ou le pire, entre le doctorat et moi. Ce sont certaines de ces personnes que j'aimerais remercier pour leur soutien moral, elles ont fait de mes années de thèses des années colorées. Je remercie la maison des chatons, tenue par Nina, qui m'a fourni l'ordinateur sur lequel j'ai écrit cette thèse. Ismael la Crevette polaire et Leandro, liés par la science et l'amitié. Julien Pepito l'arc en ciel, pour les tisanes et le soutien continu pendant que les champs brulaient. Sandrine la petite perruche pour sa présence, sa grande gentillesse et pour m'avoir accueillie dans son monde métalléux. A Nes & Adel.

A ma famille, qui m'a encouragée et soutenue de manière inconditionnelle malgré la distance. Sans vous je n'en serai pas là. Cette thèse vous est dédiée.

Sommaire :

<i>Chapitre 1 : Introduction</i>	1
1.1. Le compactage du génome	2
1.2. Assemblage de la particule virale et encapsidation du génome dans la lignée bactériophages caudés – herpèsvirus	4
1.2.1. Les bactériophages à queue	5
1.2.1.1. Voie d'assemblage de la capsidie chez les bactériophages à queue	5
1.2.1.2. Initiation de l'encapsidation chez les bactériophages à queue	8
1.2.1.2.1. Les bactériophages « <i>cos</i> »	9
1.2.1.2.2. Les bactériophages « <i>pac</i> »	10
1.2.1.2.3. Les bactériophages avec des extrémités redondantes fixes	12
1.2.2. Voie d'assemblage et encapsidation du génome dans les Herpèsvirus	13
1.3. Modèle d'étude : Le bactériophage SPP1 de <i>B. subtilis</i>	15
1.3.1. Voie d'assemblage chez SPP1	17
1.3.2. Initiation de l'encapsidation du génome chez le bactériophage SPP1	17
1.3.2.1. Organisation de la séquence <i>pac</i> chez SPP1	18
1.3.2.2. Petite sous-unité de la terminase contenant le domaine de liaison à l'ADN : gp1	19
1.3.2.3. Grande sous-unité de la terminase : gp2	21
1.4. Problématique	22
<i>Chapitre 2 : Matériels & Méthodes</i>	25
2.1. Souches bactériennes, plasmides et bactériophages	26
2.2. Construction des mutants	28
2.2.1. Préparation des bactéries compétentes	28

2.2.2. Mutagénèse dirigée.....	30
2.2.3. Clonage.....	32
2.2.3.1. Délétions <i>pacL</i>	32
2.2.3.2. Phages mutants SPP1 Mut6* et SPP1 Mut8*.....	33
2.2.4. Transformation et transfection.....	34
2.3. Séquençage des mutants construits.....	35
2.4. Test de complémentation.....	36
2.5. Test du clivage <i>pac in vivo</i>.....	37
2.6. Test de clivage <i>pac in vitro</i>, basé sur le système RTS 100 <i>E. coli</i> HY.....	37
2.7. Western Blot.....	38
2.8. Production et purification des phages.....	39
2.9. Purification des fragments <i>pac</i> pour séquençage à haut débit.....	40
2.10. PCR run-off.....	42
2.11. Séquençage haut débit, Next Generation Sequencing (NGS).....	42
2.12. Analyses bio-informatiques.....	44
 <i>Chapitre 3 : Résultats</i>	45
3.1. Etude de la région <i>pacC</i>, cible du clivage par la terminase	46
3.1.1. Effet des mutations dans <i>pacC</i> sur le clivage <i>pac</i>	46
3.1.2. Détermination du site de clivage <i>pac</i> chez SPP1 par PCR run-off et séquençage à haut débit.....	52
3.1.3. Détermination des sites de clivage <i>pac</i> des phages du groupe de SPP1	55
3.2. Etude des régions <i>pacL</i> et <i>pacR</i>, les sites de reconnaissance par la petite sous-unité de la terminase	59
3.2.1. Implication de la région <i>pacR</i> dans le clivage <i>pac</i>	59
3.2.2. Rôle de la région <i>pacL</i> , non encapsidée, dans le clivage <i>pac</i>	61

3.2.3. Test de clivage <i>pac in vitro</i> et <i>in vivo</i> dans <i>E.coli</i>	63
3.3. Etude de l'interaction gp1-ADN.....	65
<i>Chapitre 4 : Discussion</i>	73
4.1. Identification du site de clivage dans <i>pacC</i>	75
4.2. Les régions <i>pacL</i> et <i>pacR</i> sont nécessaires au clivage de <i>pac</i>	76
4.3. Modèle proposé.....	78
<i>Chapitre 5 : Conclusions et Perspectives</i>	83
<i>Bibliographie</i>	87

Liste des figures et tableaux :

Chapitre 1	
Figure 1.1 : Comparaison des stratégies de compactage d'ADN mises en place chez les eucaryotes, procaryotes et virus.	4
Figure 1.2 : Représentation schématique des différents groupes de phages à ADN double-brin (ADN db) de la famille des <i>Caudovirales</i> .	5
Figure 1.3 : Schéma général de la voie de morphogénèse chez les bactériophages à ADN db.	8
Figure 1.4 : Mécanisme d'encapsidation chez les phages « <i>cos</i> ».	10
Figure 1.5 : Mécanisme d'encapsidation de l'ADN par tête pleine (Headful).	12
Figure 1.6 : Voies générales de morphogénèse des virus à ADN db.	15
Figure 1.7 : Structure du bactériophage SPP1.	16
Figure 1.8 : Représentation schématique de la région <i>pac</i> de SPP1.	19
Figure 1.9 : Structure de la petite sous-unité de la terminase du phage SPP1-like SF6.	21
Chapitre 2	
Figure 2.1 : Représentation schématique du vecteur pKD33.	27
Figure 2.2 : Stratégie de purification des fragments <i>pac</i> .	41
Chapitre 3	
Figure 3.1 : Mutations introduites dans la séquence <i>pacC</i> .	48
Figure 3.2 : Test de clivage <i>pac in vivo</i>	50
Figure 3.3 : Effet des mutations dans <i>pacC</i> sur le clivage <i>pac</i> .	51
Figure 3.4 : Identification du site de clivage de <i>pac</i> de SPP1.	54
Figure 3.5 : Identification du site de clivage <i>pac</i> de SPP1 du Mut6* et Mut8*.	55
Figure 3.6 : Arbres phylogénétiques de la région d'ADN de <i>pac</i> et de <i>pac</i> -gène 1-gène 2 de SPP1 et ses phages proches.	56
Figure 3.7 : Alignement de séquence multiple de la séquence <i>pac</i> des phages des du groupe de SPP1.	57

Figure 3.8 : Identification du site de clivage <i>pac</i> des phages proches de SPP1, Lurz3 et SF6.	58
Figure 3.9 : Effet des mutations <i>pacR</i> sur le clivage <i>pac</i> .	60
Figure 3.10 : Effet des délétions <i>pacL</i> sur le clivage <i>pac</i> .	62
Figure 3.11 : Western Blot pour l'analyse de la production des protéines gp1 et gp2 dans le système RTS <i>E.coli</i> .	64
Figure 3.12 : Test de clivage <i>pac in vivo</i> dans <i>E. coli</i> .	64
Figure 3.13 : Prédiction de liaison à l'ADN de la protéine gp1 avec le logiciel BindN+.	66
Figure 3.14 : Alignement de séquences multiples des acides aminés de gp1 de SPP1 et ses phages proches en utilisant le logiciel ClustalW.	67
Figure 3.15 : Domaine de liaison à l'ADN (DBD) correspondant à la région N-terminal de la protéine gp1 de SF6.	67
Figure 3.16 : Représentation de la structure cristallographique de la protéine gp1 du phage SF6 (homologue de SPP1).	68
Figure 3.17 : Effet des mutants gp1 sur la reconnaissance et le clivage en <i>pac</i> .	69
Figure 3.18 : Test de complémentation <i>in vivo</i> des mutants gp1 avec SPP1 <i>sus70</i> .	70
Figure 3.19 : Phénotype des plages de lyses observé lors des tests de complémentation <i>in vivo</i> .	71
Chapitre 4	
Figure 4.1 : Modèle proposé pour la reconnaissance de <i>pac</i> par gp1.	79
<u>Tableaux</u>	
Tableau n°2.1 : Psi broth	29
Tableau n°2.2 : TfbI	29
Tableau n°2.3 : TfbII	29
Tableau n°2.4 : GM1/ GM2	29
Tableau n°2.5 : SP1	29
Tableau n°2.6 : Mutagenèse dirigée sur <i>pacC</i> et <i>pacR</i>	31
Tableau n°2.7 : Mutagenèse dirigée sur la protéine gp1	31
Tableau n°2.8 : Délétions dans <i>pacL</i>	33
Tableau n°3.1 : Mutants gp1 construits par mutagenèse dirigée	68

Liste des abréviations :

ADN	Acide Désoxyribonucléique
ATPase	Adénosine Tri Phosphatase
Cf	Concentration finale
cfu	« Colony forming units »
DO	Densité optique
DNase	Désoxyribonucléase
EDTA	Acide éthylène-diamine-tétraacétique
GFP	Protéine verte fluorescente
gp	« Gene product » (Protéine des phages)
HSV-1	Herpes simplex type 1
kDa	Kilo Dalton
Kb	Kilo paires de bases
LB	Milieu Luria-Bertani
NGS	« Next generation sequencing »
pb	Paire de base
PCR	« Polymerase chain reaction »
SDS	Sodium Dodécyl Phosphate
SPP1	<i>Subtilis</i> Phage Pavia 1
RNase	Ribonucléase
UL	Unit length (Unité de longueur)

Chapitre 1 : Introduction

1. Introduction

1.1. Le compactage du génome

Le compactage du génome est une stratégie observée dans toutes les cellules du domaine vivant. L'ADN, support de l'information génétique contenant selon les organismes des centaines de milliers à des milliards de paires de bases, doit impérativement être protégé pour assurer son intégrité et la transmission des gènes sans dommages. De plus, l'ADN doit être condensé pour être contenu à l'intérieur de la cellule. Chez *Escherichia coli* et *Bacillus subtilis*, deux des premiers organismes dont les génomes entiers ont été séquencés, la bactérie doit contenir un ADN de près de 4,6 (Blattner et al, 1997) ou 4,2 (Kunst et al, 1997) millions de paires de bases, respectivement. Ces tailles correspondent à une molécule d'ADN avec une longueur supérieur à 1 mm qui doit être contenue dans des cellules en forme de bâtonnet avec un diamètre autour de 1 μm et une longueur entre 2 et 5 μm . Chez les cellules eucaryotes, dont le génome confiné dans le noyau est souvent bien plus grand, cette étape est encore plus critique. En plus de cette contrainte d'occupation de l'espace physique dans la cellule, l'organisation du génome doit aussi le protéger des lésions ou mutations et également permettre le contrôle de l'expression des gènes (Becker, 2005).

Chez les eucaryotes, le compactage du génome est réalisé avec l'association de plusieurs protéines basiques appelées les histones qui forment un octamère sur lesquels s'entourent 146 pb d'ADN sans spécificité de séquence, formant ainsi un nucléosome. L'ensemble des nucléosomes compacté forment la chromatine retrouvée dans le noyau cellulaire (Cutter & Hayes, 2015). Chez les bactéries qui n'ont pas de noyau le chromosome linéaire ou, plus fréquemment, circulaire est condensé par super-enroulement de l'ADN catalysé par l'ADN gyrase et par interaction avec des protéines associées à l'ADN comme IHF ou HU (Thanbichler al, 2005 ; Badrinarayanan et al, 2015). Cette compaction de l'ADN conduit à la formation du nucleoïde bactérien qui se localise dans une région précise du cytoplasme

bactérien (Kleckner et al, 2014) (Hirano, 2015).

Chez les virus, on retrouve aussi cette notion de compactage équivalente à la condensation des chromosomes. Contrairement au génome cellulaire, la compaction de l'ADN viral se fait par son empaquetage dans une coque protéique qui le protège dans la particule virale infectieuse. Dans les cas des virus bactériens caudés (phages ou bactériophages) et les herpes virus qui ont une origine évolutive commune (Abrescia et al, 2012) leur ADN double-brin (ADN db) est encapsidé dans une structure (procapside) préformée (Prasad and Schmid, 2012) (Figure 1.1). Cette réaction effectuée par un puissant moteur moléculaire conduit à un compactage beaucoup plus important du génome que celui observé dans les nucléosomes eucaryotes ou dans le nucleoïde bactérien. Le confinement dans la capsid virale et des ions divalents jouent un rôle majeur dans la stabilisation de l'ADN viral dont la concentration atteint ~500 mg/ml (Roos et al, 2009). Cette stratégie d'empaquetage optimise donc la quantité d'information génétique transportée par la particule virale. Les mécanismes impliqués dans les étapes essentielles de reconnaissance spécifique du génome viral et de son encapsidation qui sont des cibles thérapeutiques potentielles lors de l'infection virale, restent méconnus.

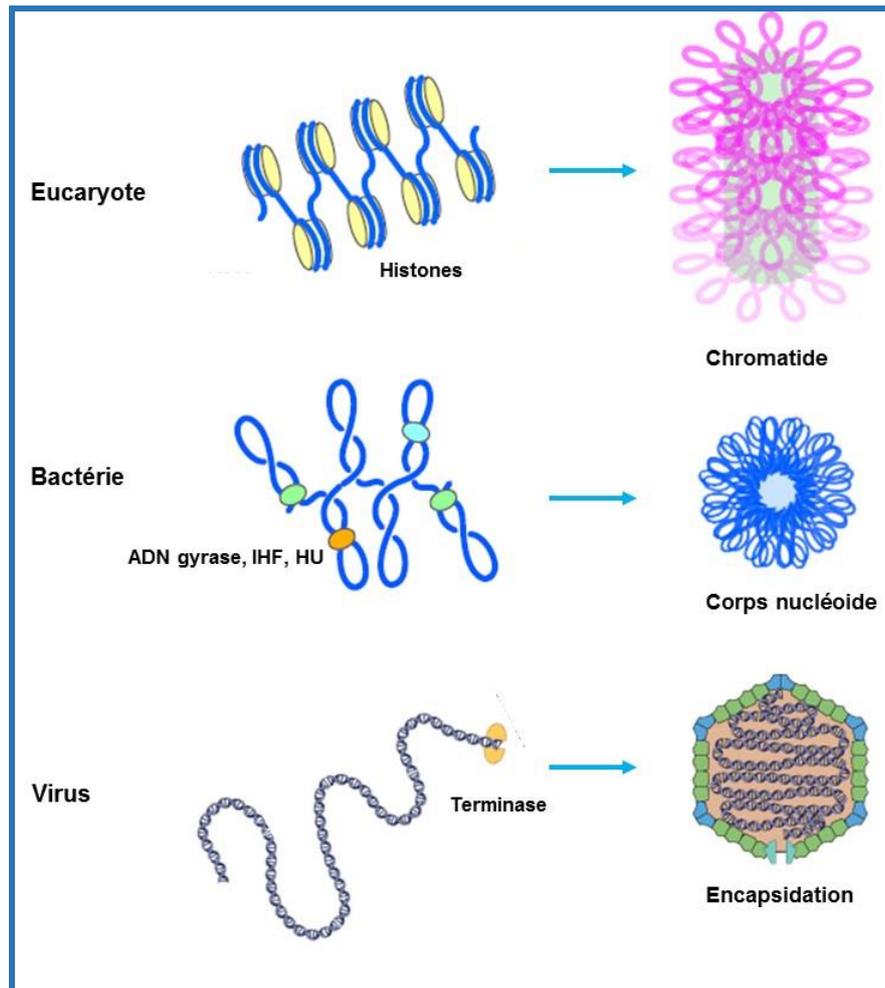


Figure 1.1. Comparaison des stratégies de compactage d'ADN mises en place chez les eucaryotes, les procaryotes et les virus. La condensation du chromosome est faite chez les eucaryotes et les bactéries par interaction avec des protéines spécifiques, les histones et des histones-like (HU, IHF...) respectivement, sur l'ensemble du génome. Chez les virus la condensation est le résultat du confinement du génome à l'intérieur de la capsid virale. Figure assemblée d'illustrations de Hirano (2015) (chromatine eucaryote et nucleoïde bactérien) et de <http://viralzone.expasy.org/virus>.

1.2. Assemblage de la particule virale et encapsidation du génome dans la lignée des bactériophages caudés-herpèsvirus

L'assemblage des particules virales est un processus complexe en plusieurs étapes dont l'objectif final est de produire des virus infectieux. Ils doivent être capables de détourner les machineries cellulaires, de répliquer leur génome en plusieurs copies, de synthétiser les différentes protéines et de sortir de la cellule hôte. L'encapsidation du génome virale est une des étapes les plus fascinantes du cycle viral. Elle a été largement étudiée chez les virus à

ADN db tel que les bactériophages à queue et les herpèsvirus.

1.2.1. Les bactériophages à queue

Les bactériophages sont les entités biologiques les plus présentes dans l'environnement et sont présents dans tous les écosystèmes (Ackermann, 2001). Les bactériophages sont classés selon le type de leur acide nucléique, la symétrie de la capsid et la présence ou non d'une enveloppe (Ackermann, 1998). Il a été observé que parmi les 5500 bactériophages étudiés au microscope électronique depuis 1956, 96% de ces phages sont des bactériophages à queue (ordre *Caudovirales*). Ils ont une symétrie binaire et sont caractérisés par une tête icosaédrique et une queue, longue ou courte, contractile ou non. Cet ordre est subdivisé en trois familles en fonction de la morphologie de la queue ; 1) *Myoviridae* (24.5%), longue queue contractile, 2) *Siphoviridae* (61%), longue queue non contractile, et 3) *Podoviridae* (14%), queue courte (Figure 1.2) (Murphy, 1995 ; Ackermann, 2007). Étant l'ordre de phages le plus important, et infectant la vaste majorité des espèces bactériennes (Ackermann, 2007), beaucoup de travaux ont été réalisés sur le groupe des **bactériophages à queue**.

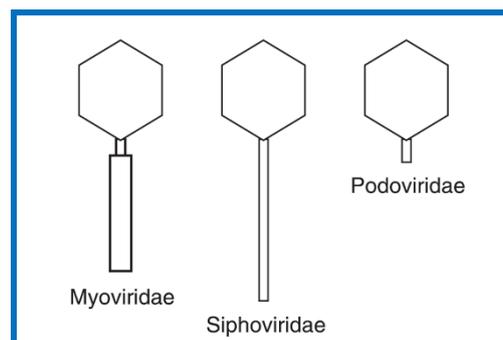


Figure 1.2. Représentation schématique des différents groupes de phages à ADN db de la famille des *Caudovirales* (Ackermann, 2007).

1.2.1.1. Voies d'assemblage de la capsid des bactériophages à queue

L'étude du cycle de vie des différents bactériophages à queue a apporté des

connaissances sur plusieurs mécanismes génétiques et biochimiques impliqués au cours de ce processus et a montré une voie de morphogénèse de la particule virale commune. En effet, les travaux sur certains des bactériophages les plus étudiés (phages λ , T4, T7, P22 et SPP1) ont montré les différentes étapes séquentielles de la voie d'assemblage des virions et les différentes interactions entre les protéines virales mises en jeu. Le processus commence par l'assemblage de la procapside, la première forme intermédiaire de la capsidie observée lors de la construction du virion. La procapside est un icosaèdre avec une forme sphérique d'une taille de 60 nm pour le phage λ (Lander et al, 2008), 60 nm chez T7 (Guo et al, 2014) et 61 nm chez P22 (Chang et al, 2007). Elle est constituée par de centaines de copies de la protéine majoritaire de la capsidie qui constituent sa coque extérieure dans laquelle les protéines d'échafaudage sont renfermées. Pendant l'assemblage de la procapside une protéine, nommée portale, forme un sommet spécialisé appelé aussi le « sommet portal » (Figure 1.3) (Moore et Prevelige, 2002). Celle-ci est un élément structural présent chez tous les bactériophages à queue. Malgré les différences de séquence d'acides aminés, les études par microscopie électronique ainsi que par diffraction de rayons X ont montré que les protéines portales partagent une structure cyclique homo-oligomérique (Orlova et al, 1999). Chez le bactériophage T7 ainsi que $\phi 29$, elle est formé par 12 sous-unités (T7 : Agirrezabala et al, 2005 ; $\phi 29$: Guasch et al, 2002). Chez SPP1, la protéine isolée est formée par 13 sous-unités (Dube et al, 1993) et par 12 sous-unités organisées autour d'un canal dans la particule phagique (Lebedev et al, 2007). La translocation de l'ADN à l'intérieur de la procapside à lieu à travers ce canal portal.

L'encapsulation de l'ADN commence par la reconnaissance et le clivage de l'ADN viral. Celui-ci est présent, le plus souvent, sous forme de concatémère composé par une suite de copies du génome phagique organisées en tête-queue (Black et al, 1989), résultant d'une réplication en cercle roulant (Takashi et al, 1975) ou par recombinaison (Mosig et al, 1998).

L'encapsidation du génome viral dans la procapside fait intervenir une protéine virale spécialisée, la terminase. Cette dernière est présente chez tous les bactériophages à queue et a un rôle essentiel dans le processus d'encapsidation du génome. La terminase est constituée chez la plupart des bactériophages par deux sous-unités, une petite sous-unité (TerS) qui varie selon les phages entre 16 et 21 kDa (T4: 18 kDa (Lin et al, 1997), P22 : 18,6 kDa (McNulty et al, 2001), SPP1 : 16 kDa (Oliveira et al, 2013)) , et une grande sous-unité (TerL) entre 46 et 72 kDa (T4 : 70 kDa (Lin et al, 1997), P22 : 57,6 kDa (McNulty et al, 2001), SPP1 : 48,8 kDa (Oliveira et al, 2013)).

L'encapsidation de l'ADN commence par la reconnaissance spécifique d'une séquence par TerS qui sera le point de départ de l'empaquetage du génome dans la particule virale (Figure 1.3). Cette reconnaissance est suivie par un clivage spécifique dans cette séquence par la TerL générant ainsi une extrémité libre prête à être encapsidée. Le complexe nucléoprotéique TerS-ADN-TerL va alors se fixer à la protéine portale par une interaction directe TerL-protéine portale, et démarrer la translocation de l'ADN vers l'intérieur de la procapside grâce à l'activité ATPasique de la TerL. Une fois la procapside remplie, la protéine TerL effectue un nouveau clivage endonucléolytique non spécifique ou spécifique, selon le phage, qui termine l'encapsidation. Des protéines viennent alors se fixer au sommet portal pour fermer le pore portal et former un complexe nommé connecteur, sur lequel viendra se fixer la queue pour finir la formation de la particule phagique. Le complexe terminase-ADN détaché de la capsid va interagir avec une autre procapside et démarrer un nouveau cycle d'encapsidation (Casjens, 2011 ; Oliveira et al, 2013).

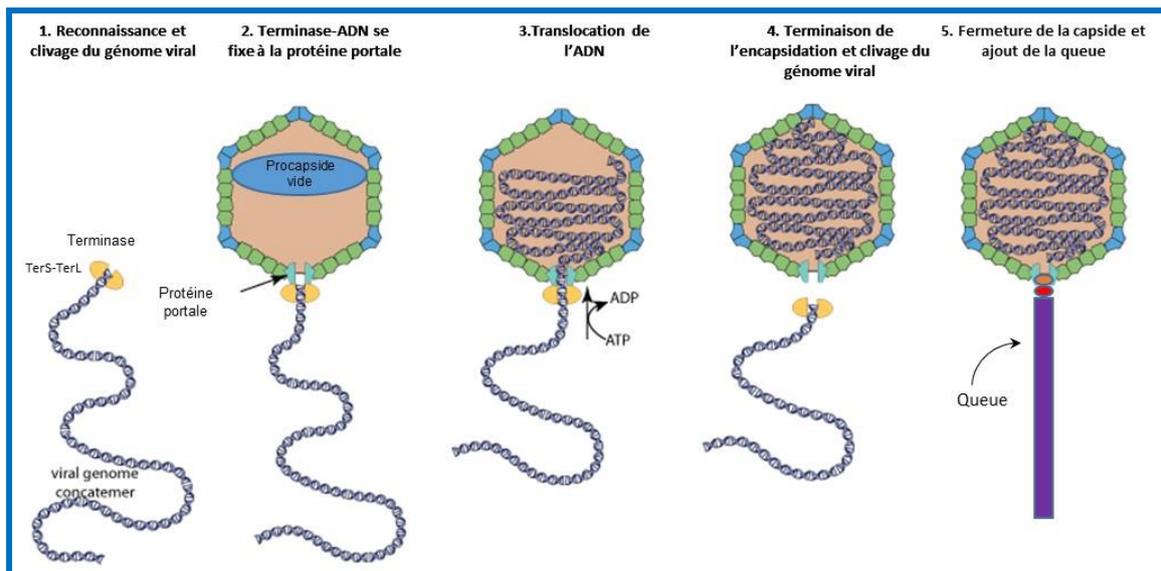


Figure 1.3. Schéma général de la voie de morphogénèse de la particule virale chez les bactériophages à ADN double-brin. Le concatémère viral est représenté en interaction avec les deux sous unités de la terminase TerS-TerL. Après reconnaissance et clivage, le complexe ADN-terminase se fixe à la protéine portale, par laquelle l'ADN est transloqué vers l'intérieur de la procapside grâce à l'hydrolyse de l'ATP par la protéine TerL. Une fois la tête pleine, une seconde coupure de l'ADN par la TerL a lieu finissant l'encapsidation du génome. Le complexe ADN-terminase se détache pour chercher une nouvelle procapside. Des protéines additionnelles viennent fermer le pore portal, suivi par la fixation de la queue. Figure adaptée de <http://viralzone.expasy.org/>.

1.2.1.2. Initiation de l'encapsidation chez les bactériophages à queue

Lors de l'infection des bactéries, la plupart des phages ne dégradent pas le génome de l'hôte et doivent donc reconnaître sélectivement le génome viral. L'encapsidation de l'ADN commence par la reconnaissance de régions spécifiques dans des concatémères produits par cercle roulant ou par recombinaison. On reconnaît trois stratégies d'encapsidation en fonction du site reconnu et des caractéristiques des molécules d'ADN encapsidées. C'est ainsi que trois groupes de phages se distinguent : i) Les phages dont l'encapsidation commence et se termine par une coupure dans la séquence « *cos* » et qui produit des extrémités cohésives, ii) Les phages dont l'encapsidation commence par la séquence « *pac* » et se termine dans une séquence différente, déterminée par le mécanisme par tête pleine avec la production d'extrémités redondantes et iii) Un troisième groupe qui inclut les phages T3 et T7, dont l'encapsidation commence et se termine sur un site spécifique, en produisant de longues

extrémités redondantes.

1.2.1.2.1. Les bactériophages « *cos* »

Les mécanismes d'encapsidation chez les phages « *cos* » ont été bien étudiés chez le bactériophage λ . L'assemblage de virions infectieux commence par la reconnaissance et le clivage d'une séquence spécifique du concatémère nommée « *cos* » pour *cohésif*, puisque l'encapsidation d'un monomère d'ADN dans la procapside conduit à la production d'un ADN linéaire avec des bouts cohésifs de 12 pb complémentaires (Catalano et al, 1995). Comme chez les autres bactériophages à queue, l'initiation de l'encapsidation est assurée par la terminase. La séquence *cos* du phage λ est une région de 200 pb avec plusieurs sites (Figure 1.4) : *cosN*, *cosB* et *cosQ* (Yang et al, 2008). La TerS de λ gpNu1 reconnaît spécifiquement 3 motifs R répétés de 16 pb dans la région *cosB* qui se trouve à droite du site de clivage par la terminase (Yang et al, 1997). L'interaction gpNu1- *cosB* va alors conduire la TerL de λ , gpA, sur le site *cosN* pour couper à une distance de 22 pb générant ainsi une extrémité libre prête à être encapsidée de manière unidirectionnelle dans la tête du phage. L'encapsidation de l'ADN commence alors grâce à l'activité ATPasique de la sous-unité gpA (Feiss & Catalano, 2005). Une fois la capsid pleine, gpA va terminer le cycle d'encapsidation avec un second clivage dans la séquence *cosN*. La région *cosQ* constituée de 7 pb et qui se trouve à 17 pb à gauche de *cosN* a été décrite comme nécessaire pour une terminaison de l'encapsidation sur le site approprié *cosN* (Wieczorek & Feiss, 2001).

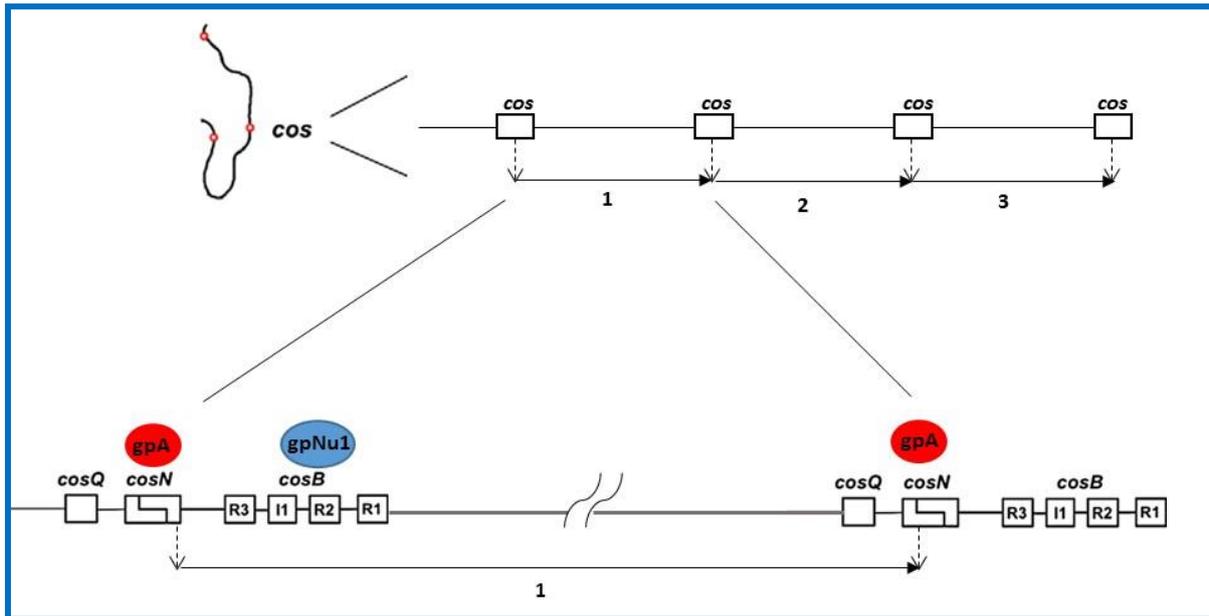


Figure 1.4. Mécanisme d'encapsidation chez les phages « *cos* ». L'ADN concatémérique du phage λ est représenté schématiquement dans la partie supérieure de la figure. Le polymère d'ADN est présenté avec les séquences *cos* entre chaque unité issue de la répllication en cercle roulant. Les flèches verticales représentent les sites de coupure dans *cos* et les flèches horizontales le nombre de cycles d'ADN encapsidé indiquant le sens de l'encapsidation du génome. Dans l'agrandissement du génome de λ en bas de la figure sont montrés les sites *cos* d'initiation et de terminaison de l'encapsidation avec les éléments de séquence nécessaires pour déterminer le clivage de *cos* (Feiss & Catalano, 2005 ; Wiczcerek & Feiss, 2001). Figure adaptée à partir de Yang et al, 2008.

1.2.1.2.2. Les bactériophages « *pac* »

Plusieurs bactériophages bien étudiés tels que les phages, P22, P1 et SPP1 font partie du groupe de phages qui encapsident leur génome à partir d'une séquence *pac*. Comme les phages *cos*, ce groupe encapside son ADN à partir d'un concatémère. La production des virus infectieux requiert l'encapsidation d'au moins une unité du génome qui passe par une première reconnaissance et clivage dans la séquence *pac*. En effet la petite sous-unité de la terminase reconnaît une séquence spécifique de l'ADN « *pac* » pour « packaging » démarrant ainsi le premier cycle d'une série processive et unidirectionnelle de cycles d'encapsidation. Le site de terminaison est cette fois-ci non spécifique, déterminé par la quantité d'ADN dans la capsid.

Chez le phage P22 qui infecte la bactérie *Salmonella typhimurium* (Zinder et Lederberg, 1952), l'initiation de l'encapsidation du génome commence par la reconnaissance

d'une séquence de 22 pb au sein de *pac* par la TerS, gp3 de P22 (Wu et al, 2002). Cette interaction est suivie par un clivage par la nucléase TerL (gp2) de P22 dans une région à gauche du site de reconnaissance à une distance de 120 pb. Chez P22 le clivage de *pac* est caractérisé par 6 coupures dans cette région, distribuées à 20 pb d'intervalle (Wu et al, 2002). Chez le phage P1 qui infecte une large gamme de bactéries Gram-négatives (Yarmolinsky et Sternberg, 1988), *pac* est une séquence de 162 pb. TerS, PacA dans P1, reconnaît 7 motifs répétés 5'-TGATCA/G-3' qui se trouvent dans 90 pb de la séquence *pac* de part et d'autre du site de clivage (Sternberg et Coulby, 1990). TerL, PacB dans P1, vient alors couper dans une séquence de 13 pb (Sternberg et Coulby, 1987). Le phage SPP1 qui infecte *B.subtilis* a une séquence *pac* de 268 pb dont l'organisation et position sont différentes par rapport à P22 et P1. Elle sera décrite en détail dans la section 1.3.2.1.

Selon les phages, la région *pac* varie en taille et en séquence mais tous ces phages adoptent la même stratégie de reconnaissance d'un site spécifique, suivie d'un clivage effectué par la terminase. Cette première reconnaissance est critique puisqu'elle représente le point de départ de plusieurs cycles d'encapsidation où les autres séquences *pac* ne sont pas coupées. En effet, une fois la capsid pleine d'ADN, TerL enclenche une nouvelle coupure cette fois non spécifique par le mécanisme de tête pleine déterminée par la quantité d'ADN à l'intérieur de la tête (Figure 1.5). La quantité d'ADN encapsidée est probablement déterminée par un capteur qui détecte quand la tête est pleine. Chez SPP1, des substitutions d'acides aminés dans la protéine portale gp6 (mutations *siz*) ont montré que l'ADN encapsidé était plus court que la gp6 sauvage. Une des mutations *siz* se trouve dans la partie de gp6 qui est à l'intérieur de la capsid, ce qui pourrait en constituer l'élément qui détecte la quantité d'ADN encapsidée en le bloquant par un changement de conformation. Une autre mutation dans gp6 qui induit l'encapsidation d'un ADN plus court, a aussi été associée à une réduction de l'activité ATPasique de gp2, ce qui basculerait l'activité de cette protéine vers la coupure

nucléasique qui termine le cycle d'encapsidation (Oliveira et al, 2013). Le mécanisme de terminaison par tête pleine conduit à ainsi à l'encapsidation d'une quantité d'ADN supérieure à une unité du génome viral, par exemple 104% chez SPP1 (Tavares et al, 1996). Le résultat de ce mécanisme est une population de molécules d'ADN encapsidées avec une permutation partielle de leur séquence et avec des extrémités redondantes (Casjens et Gilcrease, 2009). La redondance terminale permet la recirculation de l'ADN viral lors d'une infection afin de démarrer sa réplication dans le cytoplasme de la cellule hôte.

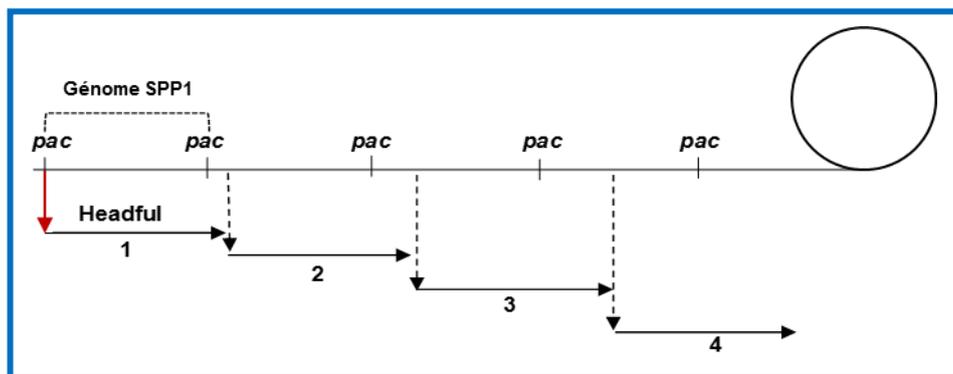


Figure 1.5. Mécanisme d'encapsidation de l'ADN par tête pleine « Headful ». Représentation schématique de l'ADN concatémère d'un phage *pac*, ici SPP1, issu du mode de réplication en cercle roulant. La première flèche verticale rouge (à gauche) indique le premier clivage spécifique dans *pac* qui génère une extrémité libre d'ADN qui est encapsidée lors du premier cycle d'encapsidation. Les flèches verticales suivantes représentent les clivages sur des séquences non spécifiques qui finissent chaque encapsidation, déterminées par le mécanisme d'encapsidation par tête pleine. Les flèches horizontales montrent la direction de l'encapsidation et indiquent les différents cycles d'encapsidation. La région entre deux sites *pac* correspond à une unité de génome de SPP1 (44.0 kb). La région entre deux flèches verticales correspond à la taille du chromosome mature de SPP1 encapsidé (45.5 kb) (Oliveira et al, 2013).

1.2.1.2.3. Les phages avec des extrémités redondantes fixes

Cette stratégie d'encapsidation est utilisée par les bactériophages T3 et T7 qui infectent *E. coli*. Comme les phages « *cos* » et « *pac* », la réplication de leur ADN produit un concatémère. L'initiation et la terminaison de l'encapsidation se fait par la reconnaissance et le clivage d'une séquence spécifique, tel que les phages « *cos* ». Cependant, ces phages sont caractérisés par la présence de longues extrémités à bouts francs de 160 pb pour T7 et 230 pb

pour T3. Les mécanismes impliqués dans ce processus sont peu connus mais ils requièrent probablement une synthèse de l'ADN des extrémités lors de l'encapsidation pour générer une répétition de leur séquence (Calendar, 1988).

1.2.2. Voie d'assemblage et encapsidation du génome dans les Herpèsvirus

Le processus utilisé par les bactériophages à queue pour l'assemblage de leur capsidite et l'encapsidation du génome viral dépasse la barrière eucaryote/procaryote. En effet, la même stratégie est observée chez les herpèsvirus (ordre *Herpesvirales*), des virus à ADN db qui infectent les animaux et les humains. La voie d'assemblage de la lignée des phages caudés et des virus d'Herpès suit un schéma général commun avec un ordre séquentiel et bien défini d'interactions protéine-protéine et protéine-acide nucléique. Comme chez les phages (Figure 1.6), le génome viral est protégé à l'intérieur d'une procapsidite préformée qui est le premier élément observé lors de la morphogénèse de ces virus eucaryotes. Celle-ci de forme sphérique est composée par la protéine majoritaire de la capsidite appelée UL19 qui renferme 1100 copies des protéines d'échafaudage UL26.5 et de la protéase virale UL26. La protéine portale UL6 est positionnée à un sommet unique de la procapsidite. La protéolyse des protéines d'échafaudage suivie par leur sortie et l'expansion de la procapsidite sont suivies par l'encapsidation de l'ADN (Baines, 2011).

L'encapsidation de l'ADN est initiée par la reconnaissance et le clivage d'une séquence spécifique du concatémère. Ce concatémère est constitué par la répétition de deux segments retrouvés dans le génome viral. Ils sont désignés par leur longueur : un segment long UL (unique long) et un court Us (unique short). Des boîtes répétées inversées sont présentes de part et d'autre de chaque segment. Des motifs riches en A/T, Pac1 et Pac2, en amont de UL, ont été décrits comme nécessaires pour le premier clivage du concatémère qui a lieu dans la séquence DR1 générant une extrémité 3' protubérante de sorte que Pac1 et Pac2 ne soient pas encapsidés. Comme chez les bactériophages à queue, cette étape est prise en

charge par les protéines de la terminase composé de 3 sous-unités ; **i**) la protéine UL28 qui a pour rôle la reconnaissance de l'ADN, **ii**) la protéine UL15 qui a une activité de nucléase et d'ATPase, et **iii**) la protéine UL33 dont le rôle définitif n'est pas complètement déterminé mais, est nécessaire pour l'encapsidation de l'ADN. L'extrémité libre de l'ADN générée sera ensuite transloquée à l'intérieur de la capsidie à travers un canal formé par 12 copies de la protéine portale UL6, grâce à l'énergie générée par la protéine UL15. Pour finir l'encapsidation, la terminase UL28/UL15 assure un second clivage dans la séquence DR1 dans la région Pac1 et Pac2. Ceci, génère la même coupure que lors du premier clivage conduisant à la production de molécules avec des extrémités identiques (Adelman et al, 2009 ; Catalano, 2005). S'ensuivent alors des étapes d'assemblages spécifiques aux herpèsvirus. La première est la fixation autour de la capsidie d'une couche de protéines, le tégument, au contact de la membrane interne du noyau et la seconde, l'enveloppe, dans le cytoplasme au contact des vésicules du réseau trans-Golgi (Mettenleiter et al, 2009).

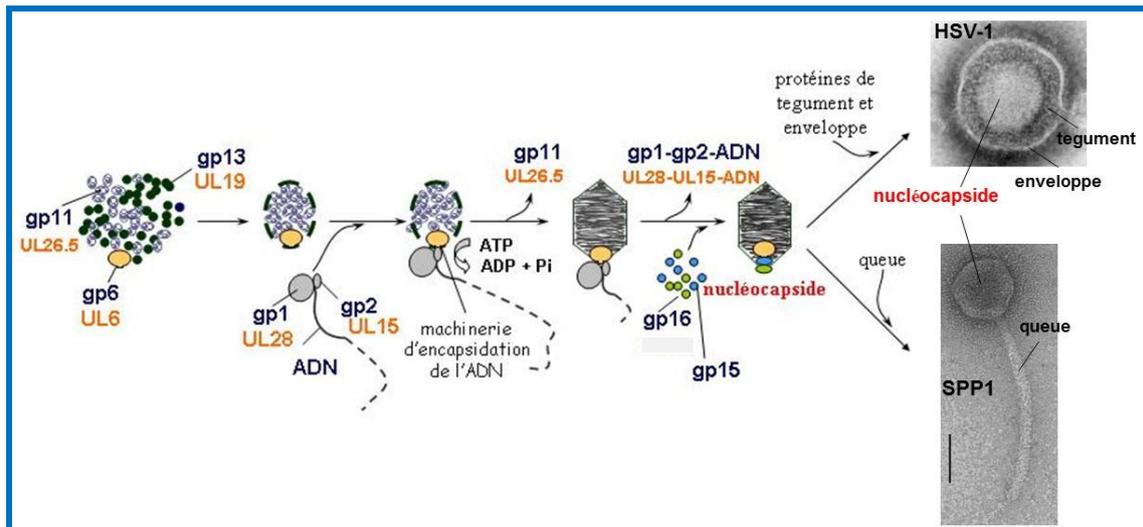


Figure 1.6. Voie générale de morphogénèse des virus à ADN db. Représentation schématique de l'assemblage des particules virales chez les herpèsvirus (ici HSV-1) en comparaison avec les bactériophages à queue (ici le phage SPP1). Dans les différentes étapes sont indiquées les protéines communes dans les deux groupes. Le cycle de vie commence par la formation de la procapside par les protéines UL19/gp13 (protéine majeure de la capsid), gp26.5/gp11 (protéine d'échafaudage) et UL6/gp6 (protéine portale). La petite et la grande sous-unité de la terminase, UL28/gp1 et UL15/gp2 respectivement, reconnaissent le génome viral concatémérique et introduisent un clivage spécifique sur l'ADN viral. Le complexe ADN-terminase interagit ensuite avec la portale sur un sommet unique de la procapside. Grâce à l'énergie de l'hydrolyse de l'ATP par UL15/gp2, l'ADN est transloqué à l'intérieur de la procapside à travers le canal de la protéine UL6/gp6. Une fois la tête pleine une deuxième coupure de l'ADN par la terminase termine l'encapsidation. La voie d'assemblage suit le même schéma général jusqu'à l'étape de terminaison de l'encapsidation, où les deux voies divergent avec des étapes d'ajout d'un téguement et d'une enveloppe pour les Herpèsvirus et d'ajout de la queue pour les bactériophages.

1.3. Modèle d'étude : le bactériophage SPP1 de *B. subtilis*

Comprendre les mécanismes d'encapsidation du génome viral dans la procapside est l'une des questions les plus intéressantes dans le cycle de vie des virus. Un apport significatif des connaissances actuelles sur cette problématique a été fourni en étudiant les bactériophages à queue comme le bactériophage de SPP1, étudié dans le laboratoire d'accueil. SPP1 est un phage bien caractérisé. Ses propriétés de transfection comme outils de biologie moléculaire (Trautner et Spatz, 1973), le séquençage de son génome entier (Alonso et al, 1997) ainsi que l'étude de sa voie morphogénétique (Oliveira et al, 2013) ont fait de lui un très bon modèle de recherche. De plus, les protéines impliquées dans les différentes étapes de son cycle de vie ont été identifiées et celles impliquées dans le moteur moléculaire responsable de l'encapsidation

de l'ADN ont pour la plupart des structures cristallographiques déterminées (Buttner et al, 2012 ; Smits et al, 2009 ; Lebedev et al, 2007).

Le bactériophage SPP1 est un phage lytique qui infecte la bactérie *B. subtilis*. Il a été isolé à Pavia (Riva et al, 1968). Les caractéristiques de sa particule virale font de lui un membre de la famille des *Siphoviridae*. Il est composé d'une capsidie icosaédrique d'un diamètre de 60 nm attachée à une queue hélicoïdale non contractile de 177 nm de long, avec à son bout une fibre responsable de son attachement sur la cellule hôte lors de l'infection (Fig 1.7) (White et al, 2012).

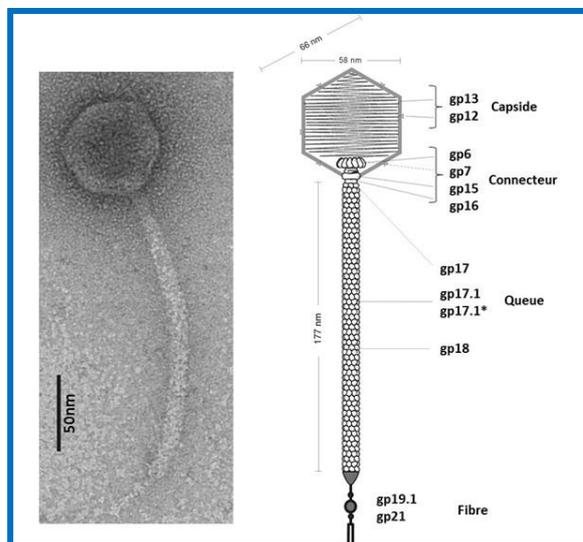


Figure 1.7. Structure du bactériophage SPP1. À gauche, observation de la particule virale de SPP1 au microscope électronique en coloration négative. À droite, organisation structurale du virion de SPP1. La localisation de ces différents composants protéiques est représentée. Adapté d'Alonso et al, 2006.

À l'intérieur de sa capsidie on retrouve un ADN linéaire double-brin d'une taille de 45,5 kb (Morelli et al, 1979; Tavares et al, 1992). Le chromosome mature SPP1 est une molécule qui présente des séquences terminales redondantes de 4% et une permutation circulaire partielle de l'ordre des gènes. Ces propriétés résultent du mécanisme d'encapsidation de l'ADN par tête pleine, en utilisant un substrat d'ADN concatémérique généré lors de la réplication du génome en cercle roulant. En effet, dans les phases précoces après l'infection, la molécule d'ADN est présente dans la cellule sous forme circulaire d'une taille de 44.007 pb (Alonso et al, 1997). C'est à partir de cette molécule circulaire que l'initiation de la réplication commence, générant d'abord de nouveaux cercles du génome par réplication thêta.

Un « switch » vers la réplication en cercle roulant (réplication sigma) s'ensuit générant des concatémères tête-queue du génome (Lo Piano et al, 2011)

1.3.1. Voie d'assemblage chez SPP1

Comme décrit dans la section 1.2.1.1, la voie de morphogénèse du phage SPP1 commence par l'assemblage d'une coque protéique icosaédrique formée par la protéine majeure de capsid gp13 (415 copies) qui entoure 100 à 180 de copies de la protéine d'échafaudage gp11. La procapside d'un diamètre de 55 nm est caractérisée par la présence de la protéine portale gp6 dans un sommet spécialisé de la capsid. Le bactériophage SPP1 fait partie du groupe de phages « *pac* ». L'encapsidation du génome commence donc par la reconnaissance de la séquence *pac* par TerS, gp1 dans SPP1, suivie par son clivage par TerL, gp2 dans SPP1 (Figure 1.6). Le complexe gp1-ADN-gp2 interagit par la suite avec gp6 dans la procapsid conduisant à l'assemblage du moteur moléculaire qui effectue la translocation du génome à travers le canal portal. Une fois la tête pleine d'ADN, une deuxième coupure non-spécifique a lieu par la nucléase gp2 finissant ainsi le premier cycle d'encapsidation. Les protéines gp15 et gp16 viennent ensuite fermer le pore portal sur lequel viendra se fixer la queue pour finir la morphogénèse du virus SPP1 (Oliveira et al, 2013).

1.3.2. Initiation de l'encapsidation du génome chez le bactériophage SPP1

Comme décrit chez les bactériophages à queue, en général, l'étape d'initiation de l'encapsidation requiert, chez SPP1, une reconnaissance et un clivage spécifique de l'ADN viral concatémérique par la terminase. Ici, nous nous intéresserons plus particulièrement à la région *pac* de SPP1, aux deux sous-unités de la terminase, gp1 et gp2, et aux connaissances actuelles concernant la reconnaissance et le clivage du site *pac*.

1.3.2.1. Organisation de la séquence *pac* chez SPP1

La séquence *pac* de SPP1 est une région de 268 pb qui chevauche la fin et le début de la séquence du génome du phage, coordonnées [43779-44010/1-37] de la séquence déposée sur Genbank X97918.2 (Figure 1.8). Des études par protection à la DNase avec de la gp1 purifiée et de l'ADN contenant la séquence *pac* ont permis de montrer que la protéine se lie à deux sites différents: à gauche la région *pacL* de 99 pb (non encapsidée) et à droite la région *pacR* de 33 pb (encapsidée) (Chai et al, 1995). Entre les deux régions, une région de 83 pb appelée *pacC* a été déterminée comme cible de clivage par gp2 (Bravo et al, 1990). La séquence *pac* chevauche en partie la séquence nucléotidique qui code pour la protéine gp1.

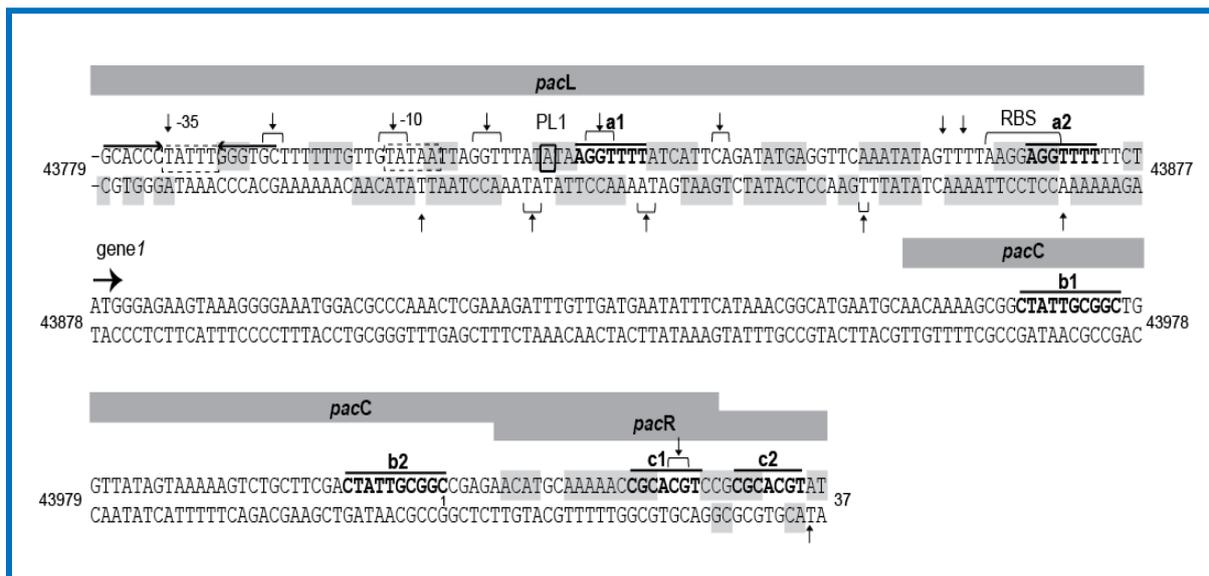


Figure 1.8. Représentation schématique de la région *pac* de SPP1. Les régions de reconnaissance par la protéine gp1 identifiées par un test de protection à la DNase, *pacL* et *pacR* (Chai et al, 1995), sont identifiées. Le profil de protection à la DNase I est représenté par des zones grises ombrées sur les deux brins de l'ADN. Les zones d'hypersensibilité sont représentées par des flèches verticales et des parenthèses. Les boîtes proposées comme étant des cibles de reconnaissance par la gp1 **a1**, **a2** dans *pacL* et **c1**, **c2** dans *pacR* sont identifiées (Chai et al, 1995). La région de clivage *pacC* est montrée dans la séquence avec ses deux boîtes répétées **b1** et **b2**. Le site de clivage se trouve dans la boîte **b2** proche de la région *pacR* encapsidée. Le début du gène *1* qui code pour gp1 et qui chevauche *pacC* et *pacR* est montré avec une flèche horizontale. En amont, le site de fixation du ribosome à l'ARN (RBS) est montré par une large parenthèse ainsi que le site de démarrage de la transcription PL1 (encadré) et ses promoteurs tardifs -10 et -35 (encadrés en pointillés) (Chai et al, 1992). La coordonnée 1 sous la boîte **b2** correspond au début du génome de SPP1 et au site de clivage dans *pac* comme définis par le séquençage de SPP1 (GenBank X97918.2).

1.3.2.2. Petite sous-unité de la terminase contenant le domaine de liaison à l'ADN : gp1

L'initiation de l'encapsidation de l'ADN requiert la reconnaissance spécifique de *pac* par gp1. Gp1 est une protéine de 147 acides aminés. La structure cristallographique de la protéine gp1 (TerS) de SF6, qui a une similarité de séquence en acides aminés de 85% avec la protéine gp1 de SPP1, a été déterminée (Figure 1.9). Elle est composée par une région N-terminale qui contient le domaine de liaison à l'ADN (DBD), un corps d'oligomérisation et par un domaine C-terminal. La protéine forme une structure en nonamère dans laquelle les domaines de liaison à l'ADN sont rangés autour du corps d'oligomérisation de manière flexible (Büttner et al, 2012). Il a été montré *in vivo* que la présence de la petite sous unité de

la terminase gp1 est essentielle pour l'initiation de l'encapsidation de l'ADN (Chai et al, 1992, 1994). Elle a pour rôle de reconnaître spécifiquement deux régions de *pac*: *pacL* (région non encapsidée ou extrémité gauche) et *pacR* (région qui est encapsidée ou extrémité droite) (Figure 1.8) (Chai et al, 1995). Il a été proposé qu'après interaction avec *pacL* et *pacR*, la gp1 rapprocherait ces deux sites pour former une structure d'ADN en boucle (Chai et al, 1995 ; Gual et al, 2000). Elle formerait ainsi une structure nucléoprotéique qui permettrait de positionner la grande sous-unité de la terminase, gp2, sur le sous-domaine *pacC*, qui se trouve entre *pacL* et *pacR*, de façon à ce que gp2 puisse couper l'ADN et démarrer son empaquetage dans la procapside (Alonso et al, 2006). Ce modèle reste cependant à démontrer. En plus de son rôle dans la reconnaissance de *pac*, la gp1 intervient dans la modulation des activités biologiques de la gp2. En effet il a été montré que la gp1 pouvait avoir un rôle répressif sur l'activité nucléase de la gp2 tout en augmentant son activité ATPasique permettant ainsi à cette dernière de « switcher » entre les deux activités (Gual et al, 2000 ; Camacho et al, 2003).

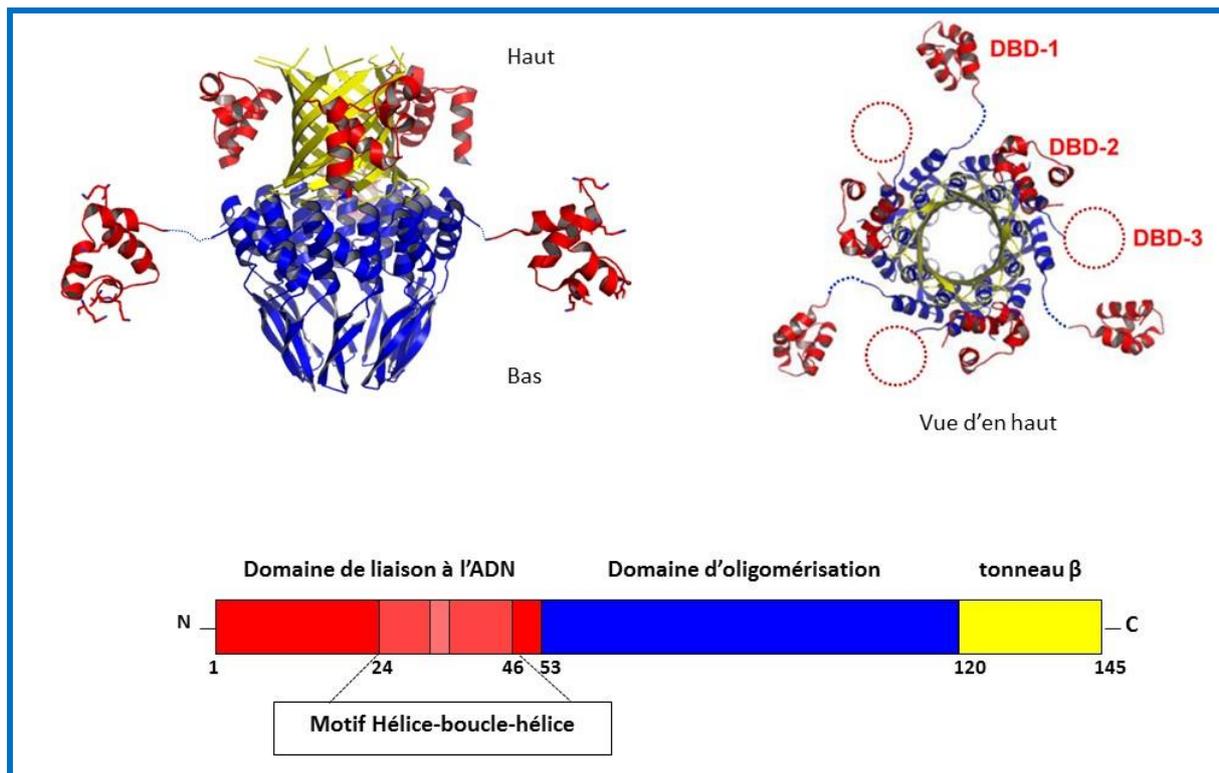


Figure 1.9. Structure de la petite sous-unité de la terminase du phage « SPP1-like SF6 ». Structure cristallographique de la gp1 en nonamère représentée sur deux angles d'observation. Deux rangées de domaines de liaison à l'ADN sont montrées autour du corps d'oligomérisation. La 3^{ème} rangée de DBD non observée est représentée en pointillé. En bas, représentation de la séquence linéaire de gp1 et de ses différents domaines. Code PDB : 3QZZ.

1.3.2.3. Grande sous-unité de la terminase gp2

La grande sous-unité de la terminase, gp2, est une protéine de 422 acides aminés, qui se trouve sous forme monomérique en solution. C'est une enzyme bi-fonctionnelle qui joue un rôle critique dans le mécanisme d'encapsidation de l'ADN. La gp2 possède en région N-terminale un domaine ATPase et un domaine nucléase en C-terminal (Smits et al, 2009). Son activité endonucléasique permet, d'une part, de générer une extrémité libre dans l'ADN concatémérique qui est essentielle pour le démarrage de l'encapsidation en coupant spécifiquement dans la séquence *pac* (Chai et al, 1997). D'autre part, cette activité nucléase permet aussi de terminer le cycle d'encapsidation en faisant une coupure cette fois non spécifique, lorsque la capsid est remplie d'ADN. Lors du processus d'encapsidation de l'ADN, gp2 est activement impliquée dans la translocation de l'ADN vers la procapsid,

puisque c'est son activité ATPasique qui fournit l'énergie nécessaire à cette étape (Oliveira et al, 2006). Lors de l'étape de reconnaissance et clivage de *pac*, la gp2 a été proposée d'introduire deux coupures décalées de 2 pb dans la séquence de *pacC* (83 pb) et plus spécifiquement dans des régions répétées de 10 pb appelées les boîtes **b1 et 2** (Fig. 1.8) (Gual et al, 2000).

1.4. Problématique

Nous avons décrit dans la section précédente, les éléments requis pour l'initiation de l'encapsidation de l'ADN dans la procapside de SPP1, en l'occurrence les sous unités de la terminase gp1 et gp2 et la séquence *pac* qui contient les sites de fixation de la gp1, *pacL* et *pacR* ainsi que le site de coupure par la gp2, *pacC*. La reconnaissance et le clivage de la séquence *pac* sont des étapes critiques puisque ceux-ci représentent le point de départ d'une série de cycles d'encapsidation le long d'un concatémère dans lequel les autres séquences *pac* restent intactes. Cette stratégie implique (i) une régulation très stricte du clivage sur *pac* et (ii) la formation d'un complexe nucléoprotéique des deux sous-unités de la terminase avec *pac* qui permet la coupure sur cette séquence et l'encapsidation unidirectionnelle qui s'en suit à partir de l'extrémité d'ADN qui a été créé. Les bases moléculaires de ces mécanismes restent très mal connues.

Comme décrit dans la section 1.3.2.1, des travaux de digestion à la DNase ont montré que la protéine gp1 purifiée interagit avec l'ADN au niveau des régions *pacL* et *pacR* contenues dans une région de 242 pb de la séquence *pac*. De plus, ces études associées à des analyses de séquence ont mis en évidence deux segments de 7 pb répétés directs mais avec des séquences différentes, appelés boîtes **a** et boîtes **c**, au niveau de *pacL* et *pacR*, respectivement (Chai et al, 1995). L'importance des boîtes **a** et **c** a été interprétée comme équivalente pour la reconnaissance de gp1. Des études *in vitro* avec des fragments de *pacL* altérés ont montré

d'autre part que la présence d'une seule boîte **a** dans la séquence de *pacL* ne changeait pas les profils de protection à la DNase. Il a ainsi été proposé qu'une seule boîte **a** au niveau de *pacL* serait suffisante pour permettre la liaison de gp1 sur l'ADN et la formation d'une structure en boucle de 204 pb, qui inclut le site de *pacC* (Chai et al, 1995). L'implication de ces séquences répétées dans *pacL* et *pacR* sur le mécanisme sous-jacent au clivage *pac in vivo* reste à démontrer.

D'autre part, des analyses *in vitro* ont montré que la gp2 a une activité endonucléolytique non spécifique, présentant cependant une certaine préférence pour la séquence 5'-RCGG↓CW-3' (Gual et al, 2000). La présence de distamycine, une molécule qui introduit des distorsions locales dans l'ADN et rentre en compétition avec la gp1 pour la formation du complexe gp1-*pac* (Chai et al, 1995), conduit à l'observation d'une coupure au niveau du site *pacC* de l'ADN par gp2 (Gual et al, 2000). L'ensemble de ces résultats suggère que la structure générée par le complexe gp1-ADN aiderait la protéine gp2 à se positionner sur la séquence *pacC* et subséquemment à la couper. Selon le modèle qui a été proposé, l'interaction avec 2 molécules de gp1 permettrait de positionner deux molécules de gp2 sur les deux boîtes **b**, mais la coupure n'aurait lieu que dans la boîte **b2** proche de la région *pacR* qui est encapsidée (Gual et al, 2000). La raison pour laquelle la coupure sur *pacC* semble avoir lieu au niveau des boîtes **b** et préférentiellement sur la boîte **b2** proche de *pacR* n'est pas encore connue. D'autant plus, que le clivage *in vitro* a montré que la gp2 est capable de couper les deux boîtes **b** (Gual et al, 2000).

L'ensemble des études *in vitro* a apporté des renseignements sur l'interaction terminase-*pac* mais ces travaux n'ont pas permis de reproduire *in vitro* la coupure de *pac* par gp1 et gp2 purifiées qui sont les deux composants du virus nécessaires et suffisants pour cliver *pac in vivo* (Cornilleau et al, 2013). Il a été aussi observé dans des expériences de mélange d'extraits de cellules infectées que l'ADN de SPP1 est encapsidé sélectivement à

partir de l'extrait donneur de la terminase (Dröge et Tavares, 2002). Ces résultats suggèrent que la régulation et coupure de *pac* dépend, ou est fortement favorisée, d'une association en *cis* entre la terminase et l'ADN qui la code. Une compréhension de ce mécanisme requiert donc la détermination des séquences de *pac* nécessaires pour la liaison de gp1 et la coupure par gp2 *in vivo*, ce qui constitue l'objectif majeur de mon projet de recherche. Plus précisément, la gp2 coupe-t-elle les boîtes **b** et en particulier celle proche de *pacR* par spécificité de séquence, ou à cause de sa position et de sa distance par rapport à la région *pacR* qui sera encapsidée? Les deux boîtes **b** ont-elles des rôles spécifiques *in vivo*? L'autre question concerne les rôles de *pacL* et *pacR* dans le processus de clivage de *pac in vivo*. Est-ce que la liaison de gp1 à *pacL* et *pacR* et particulièrement aux boîtes **a** et **c** est nécessaire au clivage de *pac*? L'identification des déterminants de séquence nécessaires pour la reconnaissance et clivage de *pac* apportera les bases pour l'étude, par la suite, des interactions terminase-*pac in vivo*. En effet, une fois la séquence *pac* étudiée, l'autre objectif est d'étudier l'interaction gp1-*pac* en se focalisant sur la protéine cette fois-ci. Nous essayerons d'identifier quels sont les résidus de la protéine gp1 impliqués dans l'interaction avec la séquence *pac*. L'ensemble de ces études permettra d'élaborer des stratégies pour la purification de ce complexe nucléoprotéique visant sa caractérisation biochimique et structurale.

Chapitre 2 : Matériels & Méthodes

2. Matériels & Méthodes

2.1. Souches bactériennes, plasmides et bactériophages

2.1.1. Souches bactériennes

- *E. coli* **XL1-Blue** a été utilisée pour amplification des plasmides contenant des mutations insérées par mutagenèse dirigée et pour les procédures de clonage.
- *B. subtilis* **YB886** (Yasbin et al, 1980) a été utilisée pour la multiplication des phages. Cette souche contenant le plasmide pBT163 (Chai et al, 1992) et ses dérivés (ce travail) a été aussi utilisée pour le test de clivage *pac in vivo* et pour les tests de complémentation.
- *B. subtilis* **HA101B** (Okubo and Yanagida, 1968), est la souche permissive qui a été utilisée pour l'amplification des différents phages contenant des mutations *sus*.

2.1.2. Plasmides

- Le plasmide **pBT163** (contenant la séquence *pac* ainsi que les gènes 1 et 2 de SPP1, codant respectivement pour la gp1 et la gp2) qui se réplique dans *E. coli* et dans *B. subtilis* (Chai et al, 1992) a été utilisé comme matrice pour l'introduction des mutations souhaitées dans la séquence *pac* et pour les sous-clonages de la région *pacL*. Il contient des gènes de résistance à l'érythromycine et au chloramphénicol. La sélection est faite avec 30 µg/ml érythromycine et 10 µg/ml chloramphénicol dans *E. coli* et avec 5 µg/ml chloramphénicol dans *B. subtilis*.
- Le vecteur pKD33 est un dérivé du vecteur **pIVEX2.3d** (Roche Biochemical Science) qui a été utilisé pour la surproduction des protéines gp1 et gp2. La protéine gp2 a une étiquette hexahistidine dans le terminal carboxylique pour faciliter la purification du complexe. pKD33 contient tous les éléments nécessaires pour l'expression *in vitro* des gènes 1 (codant gp1) et 2 (codant gp2) en présence de l'ARN polymérase T7 dans le système « RTS 100 E.

coli HY Kit Manual » (www.5prime.com) et *in vivo* dans la souche XL1-Blue (Figure 2.1). Il contient une résistance à l'ampicilline (100 µg/ml) et se réplique dans *E. coli*.

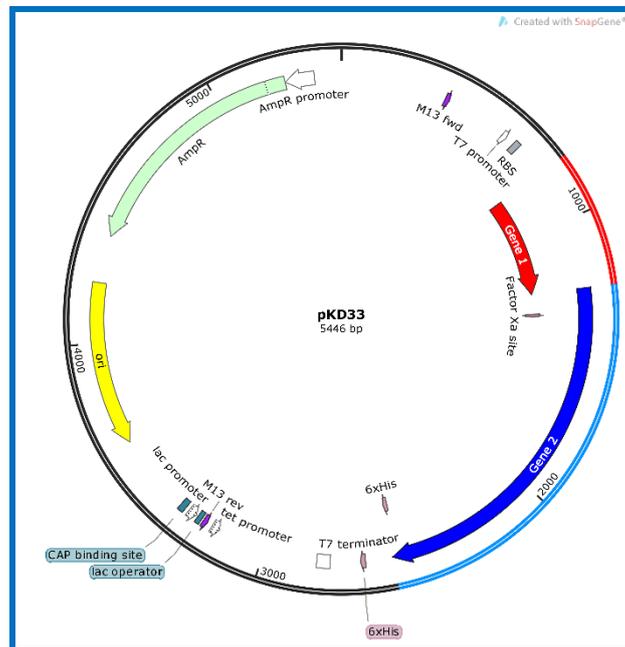


Figure 2.1. Représentation schématique du vecteur pKD33 dérivé du pIVEX3.2d, contenant les séquences du gène 1 et 2 sous le contrôle du promoteur T7 et avec la cassette hexahistidine en 3' du gène 2. Figure préparée avec Spnagene <http://www.snapgene.com/>.

- **pCC5** est un dérivé du plasmide pBT163 contenant les gènes 1 et 2 de SPP1. Le gène 2 code une gp2 avec une étiquette hexahistidine en terminal carboxylique. Il a été utilisé pour l'ingénierie de mutations dans la protéine gp1.

2.1.3. Bactériophages

- Les bactériophages SPP1 sauvage (Riva et al, 1968), les mutants SPP1Mut6* et SPP1Mut8* ainsi que les virus du groupe de SPP1, SF6 (Steensma and Block, 1979), 41c (Zsigray et al, 1973), Lurz2 et Lurz3 ont été séquencés et utilisés pour les alignements de séquences de la région *pac*-gène 1-gène 2 et/ou pour la cartographie des extrémités *pac*. Tous ont été titrés et amplifiés dans *B. subtilis* YB886. Les phages Lurz ont été isolés dans le sol du jardin du Dr Rudi Lurz (Tavares et Lurz, résultats non publiés). Les phages SPP1*sus70* et SPP1*sus19* qui contiennent un codon stop dans le gène 1 et 2, respectivement (Chai et al, 1992), ont été

utilisés pour vérifier l'activité biologique des protéines gp1 et gp2 par des tests de complémentation (section 2.4).

2.2. Construction des mutants

2.2.1. Préparation des bactéries compétentes et milieux de culture

- **Les bactéries XL1-Blue compétentes** ont été préparées par la méthode du chlorure de rubidium (Hanahan, 1985). Une culture de nuit de 3 ml a été inoculée dans 500 ml de milieu de culture Psi Broth (Tableau n°2.1) et soumise à une agitation à 37 °C jusqu'à une croissance qui atteint une DO = 0,48. Les cellules ont été gardées sur la glace pendant 15 minutes avant d'être culotées à 4°C pendant 5 minutes à une vitesse de 5000 g. Le culot a été resuspendu avec 0,4 du volume original du milieu TfbI (Tableau n°2.2) et gardé dans la glace pendant 15 minutes. Les bactéries ont été sédimentées par centrifugation à 4°C pendant 5 minutes à une vitesse de 5000 g. Le culot a été suspendu avec 0,04 du volume original du milieu TfbII (Tableau n°2.3). Les cellules ont été ensuite congelées à -80 °C.
- **Les bactéries *B. subtilis* YB886 compétentes** ont été préparées comme suit : une culture de la nuit a été diluée à 1:50 dans un volume final de 10 ml du milieu GM1 (Tableau n°2.4), puis incubée sous agitation à 37 °C. Quatre-vingt-dix minutes après la fin de la phase exponentielle de la courbe de croissance, celle-ci a été arrêtée avec l'ajout de 20% de glycérol. Les cellules pré-compétentes ont été ensuite conservées à -80 °C (Boylan et al, 1972).
- **Milieux de culture :** Les milieux de culture utilisés sont indiqués dans les tableaux de la page suivante.

Tableau n°2.1	
Psi broth (par litre) pH 7,6 avec hydroxyde de potassium. autoclavé	
Composant	Quantité
Bacto yeast extract	5 g
Bacto Tryptone	20 g
Sulfate de magnésium	5 g

Tableau n°2.2		
TfbI (par 200 mL) pH 5,8 avec acide acétique (200mM)		
Composant	Quantité	Concentration finale.
Acétate de potassium	0,588 g	30 mM
Chlorure de rubidium	2,42 g	100 mM
Chlorure de calcium	0,294 g	10 mM
Chlorure de manganèse	2,0 g	50 mM
Glycérol	30 ml	15% v/v

Tableau n°2.3		
TfbII (par 100 mL) pH 6,5 avec NaOH (200mM). Filtré.		
Composant	Quantité	Concentration finale.
MOPS	0,21 g	10 mM
Chlorure de calcium	1,1 g	75 mM
Chlorure de rubidium	0,121 g	10 mM
Glycérol	15 ml	15% v/v

Tableau n°2.4			
GM1		GM2	
Composant	Quantité	Composant	Quantité
SP1	100 ml	GM1	10 ml
Extrait de levure (10%)	1 ml	CaCl ₂ 0,05M	0,1 ml
Caséine (5%)	1 ml	MgCl ₂	0,25 ml

Tableau n°2.5	
SP1	
<i>Autoclave avant l'ajout du glucose.</i>	
<i>pH 7,2 avec NaOH, à 4 °C</i>	
Composant	Quantité (1 l)
K ₂ HPO ₄	14 g
KH ₂ PO ₄	6 g
Tri-sodium citrate.2H ₂ O	1 g
(NH ₂)SO ₄	2 g
MgSO ₄ .7H ₂ O	0,2 g
Glucose	5 g

2.2.2. Mutagenèse dirigée

L'introduction de mutations dans la région *pac* a été réalisée en suivant le protocole de Stratagène « Quick change Site-directed Mutagenesis » et en utilisant le plasmide pBT163 comme matrice. Les amorces ont été choisies de façon à introduire des mutations dans les séquences d'intérêt (*pacC* et *pacR*) sans changer la séquence en acides aminés du gène *I* dont la séquence codante chevauche *pac*. Dans un deuxième temps des mutations conduisant à des changements dans *gp1* ont été aussi introduites (marquées par un « * » dans le Tableau n°2.6). Des mutations ont été également insérées dans le domaine de liaison à l'ADN de *gp1* sur le plasmide pCC5 (Tableau n°2.7). Le manuel de Stratagène a été utilisé pour définir les stratégies de construction des amorces et celles-ci ont été commandées chez MWG (<http://www.eurofinsgenomics.eu/>). La réaction de mutagenèse a été faite dans un volume de 50 µl. Une concentration finale de 50 ng d'ADN a été utilisée, mélangée avec les oligonucléotides (125 ng), 2.5 U d'enzyme Pfu Turbo DNA polymérase, 5 µl de son tampon de réaction 10X et 5 µl de dNTPs 2 mM. L'amplification par PCR a été réalisée dans les conditions suivantes : un premier segment à 95 °C pendant 30 secondes, suivi par un second segment de 16 cycles à 95 °C pendant 30 secondes, 68 °C pendant 1 minute et 72 °C pendant 2 minutes/kb de la taille du plasmide. Après les réactions de PCR les produits obtenus ont été digérés avec 10 U de *DpnI*. Cette enzyme coupe exclusivement les résidus méthylés, présents uniquement dans le plasmide parental non muté. Les plasmides contenant les mutations ont été ensuite transformés dans *E. coli* XL1-blue (section 2.2.4).

Tableau n°2.6 : Mutagenèse dirigée sur *pacC* et *pacR*

	Nom	Oligonucléotides
Mutants <i>pacC</i>	Mutant 1	5' GCTTCGACTATTGCTGCTGAGAACATGCAAAAACC 3'
	Mutant 2	5' CAAAAGCGGCTATTGCTGCTGGTTATAGTAAAAAG 3'
	Mutant 4	5' GCAACAAAAGCTGCTATTGCTGCTGGTTATAG 3'
	Mutant 5*	5' CGGCAACAAAAGCTGCTATTAATGCTGGTTATAGTAAAAAGTCTGC 3'
	Mutant 6*	5' GCTTCGACTATTG GTAGT GAGAACATGCAAAAACCGCACGTCC 3'
	Mutants <i>pacR</i>	Mutant 9
Mutant 10		5' GCAAAAACCGCACGTCCG AGCT CGTATCGAGGAAAGATTGGC 3'
Mutant 11		5' ACATGCAAAAACCTCATGTCCG AGCT CGTAT 3'
Mutant 12		5' CAAAAACCTCATGT CAGAGCT CGTATCGAGG 3'
Mutant 13		5' GCCGAGAACATGCAGAAACCGCACGTCCGCGC 3'
Mutant 14		5' GCCGAGAACATGCAG GAAG CCGCACGTCCGCGC 3'
Mutant 15		5' GCCGAGAACATGCAG GAAGCCT CATGTCCGCGCACGTATCG 3'
Mutant 16		5' GC GAAGCCT CATGTCCG AGCT CGTATCGAGGAAAGATTGGC 3'

Les oligonucléotides utilisés pour la mutagenèse dirigée de la région *pacC* sont les mutants 1 à 6*. Ceux avec un astérisque (*) indiquent les constructions dont les mutations ont conduit à un changement en acide aminé. Le mutant 3 (Figure 3.1) a été construit avec l'oligonucléotide du mutant 2 en utilisant le plasmide du mutant 1 comme matrice. La même stratégie a été utilisée pour la construction des mutant 7* et mutant 8* (Figure 3.1) avec les oligonucléotides des mutants 4 et 6 à partir des mutants 5 et 7. Les mutants sur *pacR* sont les mutants 9 à 16. Les séquences en italiques sont les séquences homologues à SPP1.

Tableau n°2.7 : Mutagenèse dirigée sur la protéine gp1

	Nom	Oligonucléotides
gp1	K7E	5' GGAGAAGTAAAGGGG GAAT GGACGCCCAAAC 3'
	T9A	5' GAAGTAAAGGGGAAATGG GCAC CCCAAACCTCGAAAGATTTG 3'
	K11E	5' GGGGAAATGGACGCC CGAACT CGAAAGATTTGTTG 3'
	T27A	5' CGGCATGAATGCAGCAAAGCGGCTATTGCGGC 3'
	K28A	5' CATAAACGGCATGAATGCAACAG CAGC AGCGGCTATTGCGGCTGG 3'
	K28E	5' GGCATGAATGCAACAG GAAG CGGCTATTGCGG 3'
	T27AK28A	5' CATAAACGGCATGAATGCAG CAGCAGC AGCGGCTATTGCGGCTGG 3'
	K37E	5' GCTATTGCGGCTGGTTATAGT GAAAAGT CTGCTTCGACTATTGC 3'
	K38E	5' GCGGCTGGTTATAGTAAAG AGTCT GCTTCGACTATTG 3'
	K37EK38E	5'GGCTATTGCGGCTGGTTATAGT GAAGAGT CTGCTTCGACTATTGCG 3'
	K50E	5' CCGAGAACATGCAAG AACCG CACGTCCGC 3'
	R56E	5' CCGCACGTCCGCGCAG AAAT CGAGGAAAGATTGGCAC 3'

Les mutants de la petite sous-unité de la terminase (gp1) sont désignés par l'acide aminé muté, sa position dans la protéine ainsi que le nouvel acide aminé. Les caractères gras indiquent les nucléotides mutés. Ne sont présentés que les oligonucléotides du brin codant de la région *pac*-gène 1 - gène 2. Les séquences en italiques sont les séquences homologues à SPP1.

2.2.3. Clonage

2.2.3.1. Délétions dans *pacL*

Des réactions de PCR ont été effectuées pour amplifier un fragment contenant les gènes 1 et 2 de SPP1 allant jusqu'au site unique *Bam*HI du plasmide pBT163, avec des délétions dans la région *pacL* qui est située en amont du gène 1. Les amorces utilisées en 5' ont été choisies pour s'hybrider avec *pacL* et insérer en amont un site de restriction *Pvu*I. Un site *Eco*RI a aussi été ajouté pour produire des constructions contenant les mêmes sites de restriction que dans le plasmide original pBT163 (tableau 2.8). La deuxième amorce utilisée a été l'oligonucléotide pUC9 reverse 5' CTTTATGCTTCCGGCTCG 3' qui s'hybride avec une région de pBT163 située en aval du fragment cloné de SPP1 qui code pour *pac*-gène 1-gène 2. Après amplification, le fragment PCR a été purifié avec le kit QIAquick de QIAGEN puis digéré avec 5 U de *Pvu*I-*Bam*HI. Le fragment de PCR digéré a été cloné par ligation dans pBT163 coupé avec *Pvu*I-*Bam*HI, et puis transformé dans *E. coli* XL1-Blue. Ceci a permis d'obtenir des constructions où la région *pacL* a été progressivement délétée. Cette stratégie de sous-clonage a conduit à l'élimination d'une partie du vecteur en amont de *pacL*. Pour écarter un éventuel effet de cette région, nous avons construit un plasmide *pacL*-99 dans lequel la séquence sauvage de SPP1 a été amplifiée à partir de pBT163 et réintroduite dans le vecteur par digestion avec *Pvu*I-*Bam*HI, suivant la même stratégie décrite ci-dessus.

Tableau n°2.8 : Délétions dans *pacL*.

Nom	Amorces
<i>pacL</i> -15	5' CGGCCCGATCGCGGAATTCCCGGAAGGAGGTTTTTCTATGGGAGAAG 3'
<i>pacL</i> -36	5' CGGCCCGATCGCGGAATTCCCGGATGAGGTTCAAATATAGTTTAAAGG 3'
<i>pacL</i> -54	5' CGGCCCGATCGCGGAATTCCCGGAGGTTTTATCATTGATATGAGG 3'
<i>pacL</i> -58	5' GCCCGATCGCGGAATTCATAAGGTTTTATCATTGATATGAG 3'
<i>pacL</i> -68	5' GCCCGATCGCGGAATTCATTAGGTTTATATAAGGTTTTATCATT 3'
<i>pacL</i> -76	5' GCCCGATCGCGGAATTCGTTGTATAATTAGGTTTATATAAGG 3'
<i>pacL</i> -88	5' GCCCGATCGCGGAATTCGGGTGCTTTTTGTTGTATAATTAG 3'
<i>pacL</i> -93	5' GCCCGATCGCGGAATTCCTATTTGGGTGCTTTTTGTTGTATAATTAG 3'
<i>pacL</i> -96	5' GCCCGATCGCGGAATTCCTATTTGGGTGCTTTTTGTTGTATAATTAG 3'
<i>pacL</i> -99	5' GCCCGATCGCGGAATTCCTATTTGGGTGCTTTTTGTTGTATAATTAG 3'

Amorces directes utilisées pour la délétion successive de la région *pacL*. La région en italique indique la séquence homologue de SPP1 (*pacL*). Les parties soulignées sont les sites des enzymes de restriction ajoutées, *PvuI* et *BamHI* respectivement.

2.2.3.2. Construction des phages mutants SPP1Mut6* et SPP1Mut8*

L'amplification par PCR de la région *pac* a été réalisée utilisant comme matrice les ADN plasmidiques codant Mut 6* et Mut 8* (Tableau 2.6) avec les amorces SPP1-*pac99* 5' CCCGATCGCGGAATTCGGGACTAGTCTAAGCACCCCTATTTGGGTGCTTTTTGTTG 3' (en gras, les sites *PvuI*, *EcoRI* et *SpeI*, respectivement) et reverse *Eco9-1NCS* 5' CGCATTATGCGTACCCCC 3'. Les fragments PCR obtenus ont été purifiés avec le kit QIAquick (Qiagen). Ces derniers ainsi que l'ADN sauvage de SPP1 ont été digérés successivement par l'enzyme *SpeI* à 37 °C pendant la nuit suivi par son inactivation à 75 °C pendant 20 minutes et digestion avec *PspGI* à 75 °C pendant 1 heure et 30 minutes.

Les fragments de PCR digérés ont été purifiés avec le kit QIAquick. L'ADN de SPP1 sauvage digéré a été extrait deux fois avec du phénol et une fois avec du chloroforme pour inactiver les enzymes de restriction. L'ADN a été ensuite précipité par addition de 300 mM d'acétate de sodium (concentration finale) et de l'éthanol froid (-20 °C) à un ratio 2.5 éthanol : 1 solution d'ADN (v/v). Le mélange a été maintenu à -20 °C pendant 1 heure. L'ADN a été sédimenté par centrifugation pendant 15 minutes à 14.000 g, le surnageant a été

éliminé et le culot lavé avec 70% d'éthanol froid (-20 °C), puis re-centrifugé. Le culot d'ADN a été séché brièvement à 37 °C et resuspendu dans 40 µl de Tris 10 mM, pH 7,5. Les fragments de PCR *SpeI-PspGI* des Mut 6* et Mut 8* purifiés ont été introduits dans l'ADN de SPP1 digéré par *SpeI-PspGI* et purifié, par ligation avec la T4 ligase à 16 °C pendant la nuit. Une digestion des produits de ligation a été réalisée pendant 1 heure à 25 °C par *SwaI* afin d'éliminer les produits de re-ligation du génome parental de SPP1 sauvage puis transfectés dans *B. subtilis* (section 2.2.4). Pour obtenir des lysats provenant de clones purs, des plages de lyse isolées de chaque mutant ont été suspendues dans du tampon TBT (100 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl, pH 7,5) puis re-titrées par infection de *B. subtilis* pour réobtenir une plage de lyse isolée de chaque clone. Celles-ci ont été utilisées par la suite pour amplification de chaque clone comme décrit dans la section 2.8 et par Jakutyte et al. (2011). Cette stratégie a conduit à la construction de phages avec une délétion de 3500 pb entre le gène 47 et le début de la séquence *pac* de SPP1. La présence de cette délétion, des mutations dans *pacC* et l'absence de toutes autres mutations additionnelles dans *pac*-gène 1-gène 2 ont été vérifiées par séquençage (GATC Biotech) de produits PCR obtenus par amplification de cette région du génome de SPP1.

2.2.4. Transformation et transfection

2.2.4.1. Transformation d'*E. coli*

Les cellules *E. coli* compétentes (100 µl) (section 2.2.1) ont été incubées dans la glace pendant au moins 40 minutes en présence de 25 µl de chaque plasmide mutagénisé (section 2.2.2), du plasmide pBT163 (contrôle positif) et sans plasmide (contrôle négatif). Un choc thermique à 42 °C a été réalisé pendant 45 secondes. Un millilitre de milieu LB (1% de bacto-triptone, 0,5% d'extrait de levure, 0,5% de NaCl) a été ensuite rajouté aux différents échantillons et les bactéries incubées pendant 30 minutes à 37 °C. Les cellules bactériennes ont ensuite été centrifugées à 3340 g pendant 5 minutes. Le culot obtenu a été resuspendu

dans 100 µl de LB frais et étalé sur une boîte de LB agar contenant les antibiotiques adéquats. Les boîtes ont été mises en culture pendant la nuit à 37 °C.

2.2.4.2. Transformation et transfection de *B.subtilis*

Les bactéries pré-compétentes de *B. subtilis* YB886 préparées par la méthode de Boylan et al. (1972) (Section 2.2.1) ont été diluées 1:10 dans du milieu GM2 (Tableau n°2.4) et incubées avec agitation pendant une heure à 37 °C. La culture bactérienne a été supplémentée avec 20 mM de MgSO₄ et des aliquotes de 900 µl de culture ont été rajoutées à 1 µg d'ADN dans des tubes eppendorf suivie par une incubation statique de 30 minutes à 37 °C.

Dans le cas des transformations avec ADN plasmidique et les cellules ont été centrifugées pendant 5 minutes à une vitesse de 5900 g et le culot a été resuspendu dans 1 ml de LB préchauffé à 37 °C puis incubé pendant 30 minutes à 37 °C. Les bactéries sédimentées par centrifugation ont été resuspendues dans 100 µl de LB puis étalées sur boîte de LB supplémenté avec les antibiotiques adéquats et incubées à 37 °C pendant la nuit.

Dans le cas de la transfection, les bactéries incubées avec l'ADN de SPP1 ont été transférées dans une boîte de Pétri vide où on a rajouté par la suite ~100 µl d'une culture indicatrice de *B. subtilis* YB886 et 10 ml de LB semi-solide (0,7% agar) supplémenté avec 10 mM CaCl₂ (Jakutyte et al, 2011). Les bactéries ont été incubées à 37 °C pendant la nuit. Les clones de SPP1 obtenus ont été isolés et caractérisés comme décrit dans la section 2.8.

2.3. Séquençage des mutants

Après transformation d'*E. coli* avec les différentes constructions, des colonies ont été sélectionnées pour chaque mutant et mises en culture liquide dans du LB supplémenté avec les antibiotiques adéquats pendant la nuit à 37 °C, pour purification de l'ADN plasmidique. L'extraction des plasmides à partir d'*E. coli* a été réalisée avec le kit Nucleospin Plasmid

(Macheray Nagel). La concentration de l'ADN plasmidique a ensuite été mesurée par Nanodrop. Un séquençage systématique (GATC Biotech) des régions contenant les gènes *1* et *2* ainsi que la région *pac* a été réalisé pour vérifier la présence des mutations désirées, l'absence de mutations non désirées et la présence des sites *EcoRI* nécessaires pour les tests de clivage *pac* (section 2.5). Les oligonucléotides utilisés pour séquençage ont été Eco13-1CS 5'GAGTAAGTCGCCGCTCAGGC3' et Couple2-NCS 5'AGCTCCTCAATCCACATGCC'3. Les ADN de ces mêmes mutants purifiés à partir de *B. subtilis* ont également été séquencés.

2.4. Test de complémentation

L'activité biologique des protéines gp1 et gp2 des mutants étudiés dans ce travail a été quantifiée *in vivo* par un test de *trans*-complémentation. Des bactéries *B. subtilis* YB886 contenant les plasmides codant les différents mutants ont été mises en culture à 37°C dans du LB supplémenté avec les antibiotiques nécessaires, jusqu'à une densité optique de croissance équivalente à 10⁸ cfu/ml (DO = 0,8). Cent microlitres de dilutions sériées des phages SPP1*sus70* et SPP1*sus19* ont été mis sur boîte de Pétri avec 200 µl des différentes souches bactériennes puis mélangés avec 10 ml de LB semi-solide 0,7% agar supplémenté avec 10 mm de CaCl₂ et incubés à 37 °C toute la nuit (Santos et al, 1984). La complémentation a été estimée en comptant les plages de lyse de chaque dilution et comparée aux titres des bactéries infectées par la souche permissive HA101B, la souche non-permissive YB886 et le contrôle positif YB886 (pBT163). L'activité biologique des protéines a été aussi évaluée en comparant la taille des plages de lyse qui dépend du nombre de phages infectieux formés lors du processus infectieux.

2.5. Test du clivage *pac in vivo*

Nous avons utilisé la méthode décrite dans Cornilleau et al. (2013) pour ce test fonctionnel. Les souches de *B. subtilis* YB886 contenant les différents plasmides, en croissance exponentielle (DO = 0,8), ont été centrifugées et le culot bactérien a été resuspendu dans 1/10 du volume de culture initial dans une solution de lyse (100 mM Tris-HCl pH 7.5, 10 mM EDTA, 400 µg/ml RNase, 1 mg/ml lysozyme). Après 1 heure d'incubation à température ambiante, du SDS a été ajouté à une concentration finale de 0.5% et les lysats cellulaires ont été incubés pendant 10 minutes. De la pronase a été rajoutée et l'incubation continuée pendant 30 minutes à 37 °C. L'ADN total a été purifié par extraction avec du phénol/chloroforme suivie d'une dialyse pendant la nuit contre du tampon 10 mM Tris-HCl, 0,1 mM EDTA, pH 8,0 à 4 °C. La concentration et la pureté de l'ADN ont été évaluées par des mesures au Nanodrop. L'ADN a ensuite été digéré avec 100 U d' *EcoRI* (New England Biolabs) et les produits de digestion soumis à une électrophorèse sur un gel d'agarose 0,8%. L'ADN a été transféré sur membrane et analysé par la méthode de Southern Blot (Sambrook et al, 1989). Un fragment d'ADN provenant du plasmide pBT115 (Chai et al, 1992) contenant le site de clivage *pac* a été utilisé comme sonde. La membrane a été hybridée à 42 °C pendant 12 heures avec la sonde préparée et mélangée avec une solution de Dig Easy Hyb DIG (High Prime DNA Labeling and Detection Starter kit II, Roche Applied Bioscience) selon les instructions du kit. Un anticorps anti DIG/sonde a été utilisé pour l'immunodétection de l'ADN selon le protocole d'hybridation de Roche.

2.6. Test de clivage *pac in vitro*, basé sur le système RTS 100 *E.*

***coli* HY**

Un système de transcription-traduction couplées *in vitro* pour la synthèse de protéines (« Rapid Translation System RTS 100, *E. coli* HY Kit », Roche) a été utilisé dans le but de

développer un test de clivage *pac in vitro*. La transcription et la traduction se produisent simultanément dans la réaction : lorsque la T7 ARN polymérase transcrit le gène du plasmide matrice, les ribosomes fournis par le lysat d'*E. coli* commencent à traduire l'extrémité 5' de l'ARNm naissant. Cela rend le système très productif par rapport à l'utilisation de l'ARNm isolé. La protéine accumulée au cours de la réaction a été récoltée après 1-6 heures.

Les gènes d'intérêts gène 1 et 2 contenant la séquence *pac* ont été clonés sous le contrôle d'un promoteur T7 dans le vecteur pVex2.3d (section 2.1.2) et une étiquette hexahistidine a été ajoutée au domaine C-ter de la protéine gp2 aboutissant au plasmide pKD33 (Figure 2.1). Dans le compartiment de réaction, il a été mélangé 12 µl du lysat d'*E. coli*, 10 µl du mélange réactionnel fourni avec le RTS 100 *E. coli* HY kit, 12 µl d'acides aminés, 1 µl de méthionine et 5 µl de solution de reconstitution fournie avec le kit. Après le rajout de 10 µl d'ADN plasmidique pour obtenir une concentration finale de 0,5 µg, la réaction a été placée dans un tube Eppendorf Thermomixer Comfort à 30 °C dans la machine ProteoMaster (Roche). La réaction atteint 90% de production après 4 heures (www.5Prime.com). La réaction totale a été centrifugée pendant 10 minutes à 9200 g et le surnageant, le culot ainsi que l'extrait total ont été ensuite analysés par western blot pour estimer la production des protéines. Pour le test de clivage *pac*, l'extrait brut a été soumis à un traitement à la RNase (Cf= 500 µg/ml) pendant 30 minutes à 37 °C suivi de 50 µg/ml de Protéinase K et 10 mM d'EDTA pendant 1 heure à 65 °C. L'extrait a été directement digéré par *EcoRI* suivi de la détection du clivage en *pac* par Southern Blot tel que décrit dans la section 2.5.

2.7. Western blot

Les échantillons de protéines ont été dilués 1:2 avec le tampon de charge SDS 2x (0,125 M Tris-HCl pH 6,8, 0,2 M β-mercaptoéthanol, 4% SDS, 0,04% Bleu de Bromophénol, 25% de glycérol), et chauffés à 100 °C pendant 10 minutes. Les échantillons ont été ensuite soumis

à une migration électrophorétique sur gel de polyacrylamide (15%) pendant 2 heures, à 100 V/ 40 mA et transférés sur membrane de nitrocellulose pendant une heure, à 100 V dans un système de transfert liquide. La membrane a été incubée dans une solution de blocage (TBS (50 mM Tris-HCl pH 8, 150 mM NaCl)) avec 0,1% Tween et 5% de lait pendant 1h à température ambiante. La membrane a été ensuite incubée pendant 40 minutes avec les différents anticorps anti-gp1, anti-gp2 (dilution de 1:250 dans la solution de blocage) et anti-His pour la GFP (contrôle positif du système RTS, section 2.6) (dilution 1:1500). Après incubation, plusieurs lavages (4 à 5) de 10 minutes au TBS + 0,1% Tween ont été réalisés. La membrane a été ensuite incubée avec les anticorps secondaires anti-lapin (1:10000, pour les anticorps anti-gp1 et anti-gp2) ou anti-souris (1:12000), dans du tampon TBS + 0,1% Tween (+ 10% lait, pour l'anticorps anti-His) fusionnées à la « Horse Raddish Peroxidase » (HRP) pendant 1 heure à température ambiante. Après plusieurs cycles (6 à 8) de lavage de 10 minutes la membrane a été incubée avec une solution ECL (GE Healthcare) qui permet la détection par chimiluminescence des protéines immuno-marquées au contact d'un film d'autoradiographie. Des expositions de 30 secondes à une heure ont été effectuées.

2.8. Production et purification des phages

Les phages ont été produits dans la souche *B. subtilis* YB886 sauf dans le cas de SPP1*sus19* et SPP1*sus70* qui ont été multipliés dans la souche permissive HA101B. Les bactéries *B. subtilis* YB886 ont été inoculées par dilution 1:100 à partir d'une culture de nuit dans du milieu LB à 37 °C jusqu'à atteindre une densité optique correspondante à 10^8 cfu/ml (DO = 0,8). Les cellules ont été ensuite infectées avec les phages d'intérêt à une multiplicité de 3 phages/bactérie, en présence d'une concentration finale de 10 mM de CaCl₂. Les bactéries *B. subtilis* HA101B ont été infectées en suivant le même protocole avec l'exception que la culture a été supplémentée avec 100 mM CaCl₂. Après 2 heures d'infection, les débris

cellulaires ont été sédimentés par centrifugation à 4 °C pendant 20 minutes à une vitesse de 5900 g. Le surnageant a été centrifugé à la même vitesse toute la nuit, permettant ainsi de récupérer les particules virales. Les phages remis en suspension dans du TBT ont été purifiés dans un gradient de densité discontinu de 1,7, 1,5 et 1,45 g de CsCl₂ cm⁻³ préparé dans du TBT. La séparation a été réalisée par centrifugation dans un rotor SW41 (Beckman Coulter) à 20 °C pendant 5 heures à une vitesse de 35000 rpm. Les phages, apparaissant sous forme d'une bande bleue, ont été récupérés en perforant le tube avec une seringue. Après une dialyse contre du TBT pendant toute la nuit à 4 °C, ils ont été stockés à -4 °C (Sambrook et al, 1989).

2.9. Purification des fragments *pac* pour séquençage à haut débit

Après purification des phages sur gradient de chlorure de caesium (section 2.8), les ADN de SPP1 sauvage, 41c, SF6, Lurz2, Lurz3, SPP1Mut6* et SPP1Mut8* ont été extraits des particules virales par traitement à la RNase (500 µg/ml) pendant 30 minutes à 37 °C suivi de 10 mM d'EDTA à 55 °C pendant 30 minutes. Deux extractions au phénol et une au chloroforme ont été ensuite réalisées, suivies d'une dialyse pendant la nuit contre 10 mM Tris-HCl, pH 7,5 et 0,1 mM EDTA.

Afin de purifier les extrémités du génome viral issues du premier clivage du concatémère sur la séquence *pac*, 2 µg d'ADN de chaque phage ont été digérées par *Apa*LI (20U) à 37 °C pendant 4 heures (Figure 2.2). Les produits de digestion ont été ensuite soumis à une électrophorèse sur gel d'agarose de 0,8%. Les profils de digestion de SPP1 par l'enzyme *Apa*LI nous permet d'observer les bandes issues du clivage en *pac* par la terminase d'un côté et par *Apa*LI à l'autre extrémité. Ce fragment de restriction a une taille de 4600 pb. Il a été extrait et purifié du gel avec le kit NucleoSpin Gel and PCR Clean-up (Macheray-Nagel). La concentration de ce fragment a été mesurée avec l'appareil Qubit (Thermo Fisher

Scientific) et leur pureté vérifiée au Bioanalyzer (Agilent Technologies).

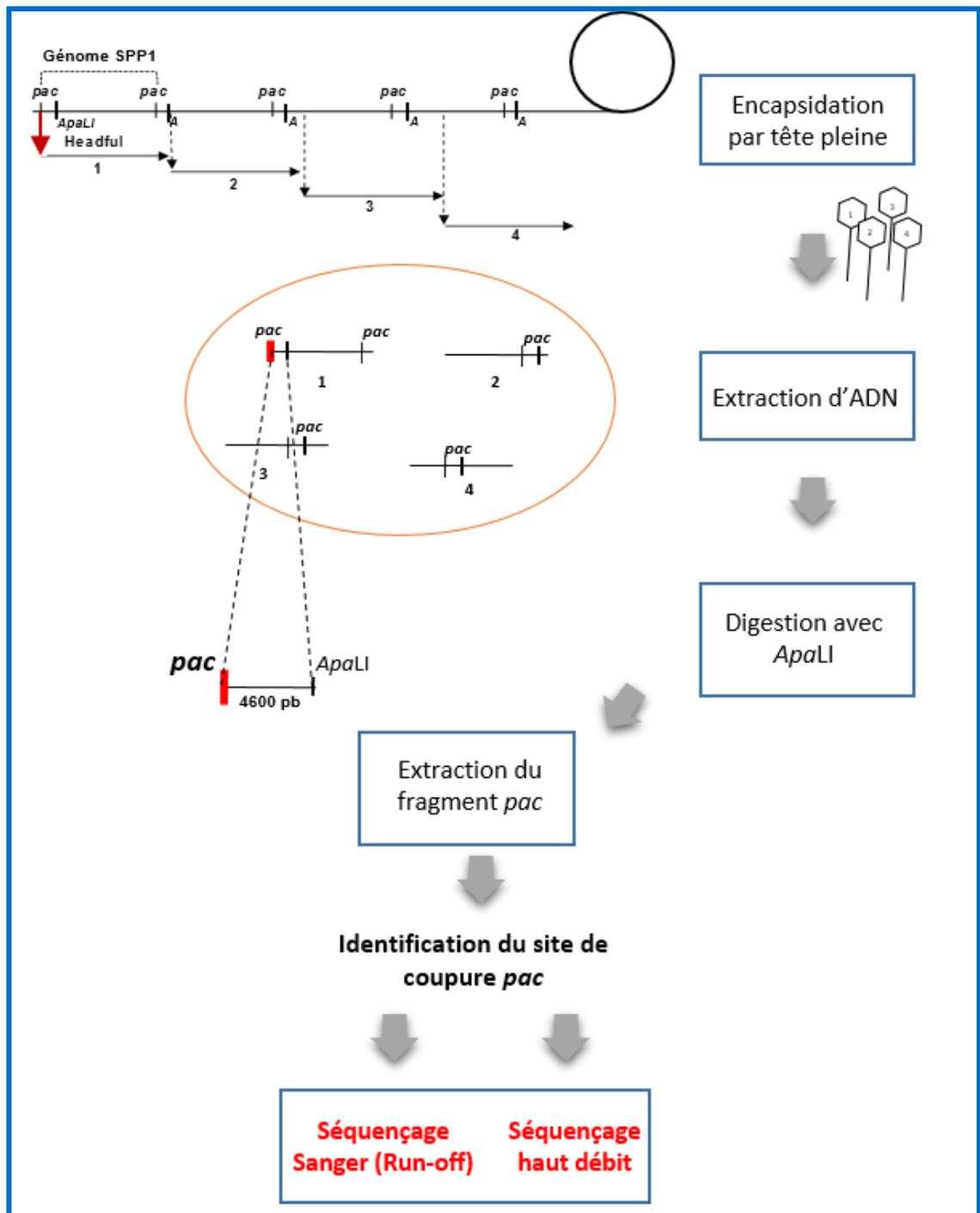


Figure 2.2. Stratégie de purification des fragments *pac*. L'encapsidation par mécanisme de tête pleine conduit à la production d'une population de molécules d'ADN encapsidés hétérogènes, dont une certaine proportion a été coupée sur *pac*. La digestion des génomes avec *ApaLI* nous permet d'identifier le fragment *pac* issu du clivage par la terminase et par l'enzyme de restriction. Ce fragment est ensuite analysé par séquençage run-off ou par séquençage à haut débit.

2.10. PCR Run-off

Un séquençage classique par la méthode Sanger a été réalisé sur les fragments *pac* purifiés des phages SPP1-*Apa*LI, SF6-*Apa*LI et Lurz3-*Apa*LI par GATC Biotech. Le principe de cette méthode consiste à initier la polymérisation de l'ADN à l'aide d'un oligonucléotide (amorce) complémentaire à une partie du fragment d'ADN à séquencer. Ici l'oligonucléotide Couple1-NCS 5'*TCCTTTGCTGATACTTCTACATGC*3' commun à tous les phages et qui s'hybride à une distance de 199 pb de la coupure *pac* a été utilisé pour faire un séquençage dans le sens 5' => 3'. L'élongation de l'amorce a été réalisée par une ADN polymérase. La synthèse du brin complémentaire a été réalisée par l'ajout des désoxyribonucléotides et d'une faible concentration de didésoxyribonucléotides modifiés avec un marqueur fluorescent. Ces derniers agissent comme terminateurs de chaîne et empêchent la poursuite de l'élongation car l'absence du groupement 3'OH n'autorise pas une nouvelle liaison phosphodiester (Sanger et al, 1977). Lorsque la polymérase atteint la fin d'un fragment aucun didésoxyribonucléotide complémentaire de la séquence référence n'est incorporé conduisant à une réduction du signal de séquence. Ceci permet d'identifier la fin de la séquence et donc du site de clivage de *pac*. La séparation des fragments d'ADN synthétisés a été effectuée par électrophorèse capillaire avec un laser et détecteur de fluorescence en aval pour identifier le didésoxyribonucléotide qui termine chaque fragment d'ADN. Les résultats de séquençage ont été analysés en utilisant le logiciel d'alignement BioEdit (Ibis Biosciences).

2.11. Séquençage haut débit, Next Generation Sequencing

(NGS)

Les fragments *pac-Apa*LI purifiés à partir de différents phages (Section 2.9) ont été analysés par la technologie de séquençage à haut débit Illumina. Cette méthode a l'avantage de produire des millions de séquences dans un seul « run » de séquençage. La préparation

d'une librairie, ici pour le fragment *pac* de chaque phage, a été faite avec le kit SPRIworks Fragment Library Kit I-10 (Beckman-Coulter). La procédure consiste à fragmenter de manière aléatoire l'ADN par sonication puis à réparer ses extrémités avec un mix contenant de la T4 polymérase, le fragment Klenow et la polynucléotide kinase. Les extrémités 3' de chaque brin d'ADN ont été ensuite adénylées pour permettre la ligation d'adaptateurs avec des extrémités dT à l'extrémité de chaque brin. Les fragments d'ADN contenant les adaptateurs ont été fixés sur une surface solide pour démarrer une amplification en « pont » grâce à des adaptateurs complémentaires pour effectuer la synthèse du brin complémentaire (Mardis, 2013). Après plusieurs cycles d'amplification, les brins ont été dénaturés créant un « cluster » de près de 1000 copies d'ADN simple-brin qui ont servi de matrice pour le séquençage fait par l'ajout de bases fluorescentes. Après chaque cycle d'incorporation d'une base, une image a été prise et l'émission de fluorescence de chaque « cluster » enregistrée. Cette procédure a été répétée pour obtenir des séquences qui ont été alignées par la suite avec le génome de référence (Mardis, 2013 ; Van Dijk et al, 2014). Le nombre de séquences qui démarrent à chaque position de l'extrémité *pac* des fragments séquencés a été utilisé pour déterminer la fréquence de coupure à ces positions par la terminase lors du premier cycle d'encapsidation de l'ADN de SPP1. En effet, sur un nombre total par exemple de 100 de séquences lues, les résultats NGS permettent de donner le nombre de séquences qui commencent avec la même base, par exemple, G= 48 séquences C=15 T=30 ou A=7. Ceci, permet de déduire la fréquence des fragments qui commencent à une base donnée. Les fréquences de coupure par *Apa*LI à l'extrémité opposée de *pac* ont été utilisées comme contrôle.

2.12. Analyses bio-informatiques

2.12.1 Alignement des séquences et arbres phylogénétiques

Les alignements de séquences ont été réalisés avec le programme CLUSTALW (<http://www.genome.jp/tools/clustalw/>). C'est un outil qui permet des alignements de séquences multiples et qui accepte différents formats d'entrée (FASTA, EMBL/SWISSPROT...) et de sortie (PHYLIP, GDE, FASTA) (Larkin et al, 2007). La visualisation des alignements a été mise en page en utilisant le serveur ESPript qui facilite la génération de figures mettant en évidence les identités de séquences, les mutations et les délétions (Robert and Gouet, 2014). Les arbres phylogénétiques des séquences d'ADN ont été construits en utilisant la plateforme Phylogeny.fr (www.phylogeny.fr) avec la méthode Neighbor joining du programme BioNJ (Gascuel, 1997) (Bootstrap : 1000). Les distances évolutives ont été calculées avec FastDist (Elias and Lagergren, 2007) et la mise en forme graphique des arbres réalisée avec TreeDyn (v198.3).

2.12.2 Analyses des données du séquençage haut débit

Les résultats de séquençage à haut débit ont été analysés en utilisant le logiciel Integrative Genomic Viewer (IGV). C'est un outil de visualisation pour l'exploration interactive de grands jeux de données génomiques. Les annotations de gènes de nombreux génomes sont incluses. Il ouvre notamment des fichiers BAM, qui contiennent les données d'alignement de séquences triés par position et indexés (Robinson et al, 2011).

Chapitre 3 : Résultats

3. Résultats

La réplication chez SPP1, comme chez d'autres bactériophages tels que P1 ou P22, aboutit à la formation de molécules d'ADN sous forme de concatémère. L'encapsidation du génome par un mécanisme de tête pleine, conduit quant à lui à la production de molécules d'ADN avec des terminaisons redondantes et une permutation circulaire (section 1.2.1.2.2). Pour expliquer ce mécanisme, un clivage unique dans le concatémère a lieu sur la séquence *pac* conduisant ainsi à l'encapsidation processive du concatémère. Puisque ce clivage directionnel n'a lieu qu'une seule fois par concatémère, ce mécanisme conduit à la production d'une population hétérogène de molécules d'ADN encapsidée (voir Figure 1.5 et Figure 2.2).

3.1. Etude de la région *pacC*, cible du clivage par la terminase

La digestion des molécules d'ADN d'un lysat de phages SPP1 avec une enzyme de restriction permet d'observer le fragment résultant du clivage en *pac* en quantité plus faible que les autres fragments de restriction (Ratcliff et al, 1979; Tavares et al, 1996). Afin d'apporter la lumière sur ce processus, la détermination du site de clivage a été importante et étudiée, notamment dans les travaux de Deichelbohrer et al. (1982). Leur analyse d'une région de 1000 pb autour de la séquence *pac* a montré que le clivage se passait préférentiellement au niveau du nucléotide CGGC↓CC avec une imprécision de 5 à 7 pb générant des extrémités 3' protubérantes. Dans cette section nous nous sommes intéressés à la région *pacC*, à la détermination de son implication dans la reconnaissance et le clivage de *pac*, ainsi qu'à la caractérisation du site de clivage par la nucléase.

3.1.1. Effet des mutations dans *pacC* sur le clivage *pac*

Afin de déterminer la spécificité du clivage *pac* nous avons étudié la séquence cible

pacC qui est clivée par la terminase. Des mutations dirigées ont été réalisées sur cette région connue comme étant préférentiellement clivée par la nucléase gp2 (Deichelbohrer et al, 1982) et plus particulièrement au niveau de la séquence 5'-GCGGC-3' qui se trouve dans la boîte **b2** 5'-CTATTGCGGC-3' (proche de la région *pacR* qui est encapsidée) (Figure 3.1). Par ailleurs, une deuxième répétition 5'-GCGGC-3' se trouve sur la même séquence *pacC* appelé boîte **b1**, ainsi qu'une troisième répétition en amont de la boîte **b1** (soulignées sur la Figure 3.1). Les deux motifs répétés 5'-CTATTGCGGC-3' dans *pacC* ont été identifiés comme cible de coupure par la nucléase gp2 *in vitro* en présence de distamycine A (Gual et al, 2000). Les mutations ont été introduites dans un premier temps dans chacun de ces motifs des boîtes **b1**, **b2** ainsi que dans le motif en amont de **b1** sans changer la séquence en acides aminés de la protéine gp1 dont le gène chevauche partiellement la séquence *pac* (Mut 1 à Mut 4, Figure 3.1). Dans un second temps, après inspection de la structure de la protéine gp1 du phage « SPP1-like » SF6 (Büttner et al, 2012 ; Greive et al, 2016), nous avons dégénéré plus les boîtes **b1** et **b2** (Mut 5*-Mut 8*, Figure 3.1). Ces changements conduisent à des substitutions de résidus au niveau du domaine de liaison à l'ADN (DBD pour « DNA-binding domain ») de gp1. Ils ont donc été réalisés en s'assurant qu'ils causent le moins possible de perturbations sur la structure de la protéine. L'alanine dans la boîte **b1** qui est exposé au solvant a été remplacée par un résidu polaire neutre (l'asparagine), la première alanine de la boîte **b2** qui se trouve dans le corps hydrophobe a été remplacée par une glycine (apolaire) et la deuxième alanine qui est à l'extérieur de l'hélice α_3 de gp1 a été remplacée par un résidu polaire (sérine). Des tests par *trans*-complémentation ont été effectués afin de confirmer que ces mutations n'ont pas d'effet sur la fonction de gp1 (résultats non montrés).



Figure 3.1. Mutations introduites dans la séquence *pacC*. La séquence et ses caractéristiques sont indiquées dans la partie supérieure de la figure. Les boîtes répétées **b** et **c** sont indiquées en gras et les régions protégées dans des essais de protection à la DNase et les sites d'hypersensibilité (Chai et al, 1995) sont indiqués comme dans la figure 1.8. La flèche noire dans la boîte **b2** indique le site de clivage préférentiel identifié au sein de *pacC* par Deichelbohrer et al. 1982. La séquence d'acides aminés de gp1 est indiquée en dessous des boîtes **b**. Les constructions portant les différentes mutations introduites au sein de *pacC* sont montrés en dessous de la séquence avec en gras rouge les nucléotides substitués. Les mutations dans ces zones qui conduisent aux changements des acides aminées sont marquées par un astérisque et les substitutions d'acides aminés sont identifiées dans le bas de la figure par des flèches.

L'effet des mutations sur *pacC* a été ensuite analysé par le test de clivage *pac in vivo*, basé sur un système de plasmide, pBT163 (Figure 3.2) (Chai et al, 1992 ; Cornilleau et al, 2013), qui contient les gènes *1* et *2* ainsi que la séquence *pac*. La présence de ces gènes, qui codent pour la gp1 et gp2, est suffisante et nécessaire pour mimer le clivage en *pac* par la nucléase dans *B. subtilis* qui a lieu lors d'une infection par SPP1 (Cornilleau et al, 2013). L'ADN total des souches de *B. subtilis* YB886, transformées avec les différents plasmides contenant *pacC* muté, a été digéré par *EcoRI* et les produits de digestions ont été séparés sur un gel d'agarose. Un Southern blot a permis ensuite l'hybridation de ces fragments d'ADN avec une sonde qui reconnaît la région *pac* (Figure 3.2). Comme on peut observer dans la Figure 3.2, une bande de ~700 pb est détectée en présence du plasmide pBT163 (témoin positif) en accord avec les tailles des fragments prédits suite au clivage *pac* par gp2 et digestion du plasmide par *EcoRI*. A titre de témoin négatif, le fragment de ~700 pb n'est pas observé avec le dérivé de pBT163 qui code la gp2 portant une mutation dans le site

catalytique (gp2D266N), qui bloque l'activité nucléase de l'enzyme et par conséquent le clivage *pac* (Cornilleau et al, 2013). D'autre part, et malgré la reproductibilité du test, le fragment de 242 pb prédit après clivage *pac*, n'a jamais été détecté. Ce résultat est en accord avec des travaux précédents (Cornilleau et al, 2013). Ceci peut être, très probablement, expliqué par la formation d'un complexe, pendant ou après le clivage, entre l'ADN viral et la terminase gp1/gp2 qui permet la protection de l'ADN qui va être encapsidé contre les attaques nucléolytiques des exonucléases cellulaires (Chai et al, 1992). Cependant, la nucléase responsable de la dégradation du fragment de 242 pb n'est toujours pas connue. Le fragment de 958 pb détecté dans le Southern blot correspond au fragment d'ADN *EcoRI* avec un site *pac* non clivé, alors que les bandes supérieures indiquent des fragments d'ADN partiellement digérés. La présence du fragment de 958 pb indique que le test de clivage *pac* utilisant le plasmide pBT163, reproduit bien le mécanisme du clivage *pac* observé lors d'une infection par le phage SPP1. Il mime fidèlement la coupure de *pac* et sa régulation, puisque celui-ci est un événement peu fréquent, et n'a lieu qu'une seule fois par série de cycles d'encapsidation à partir d'un concatémère.

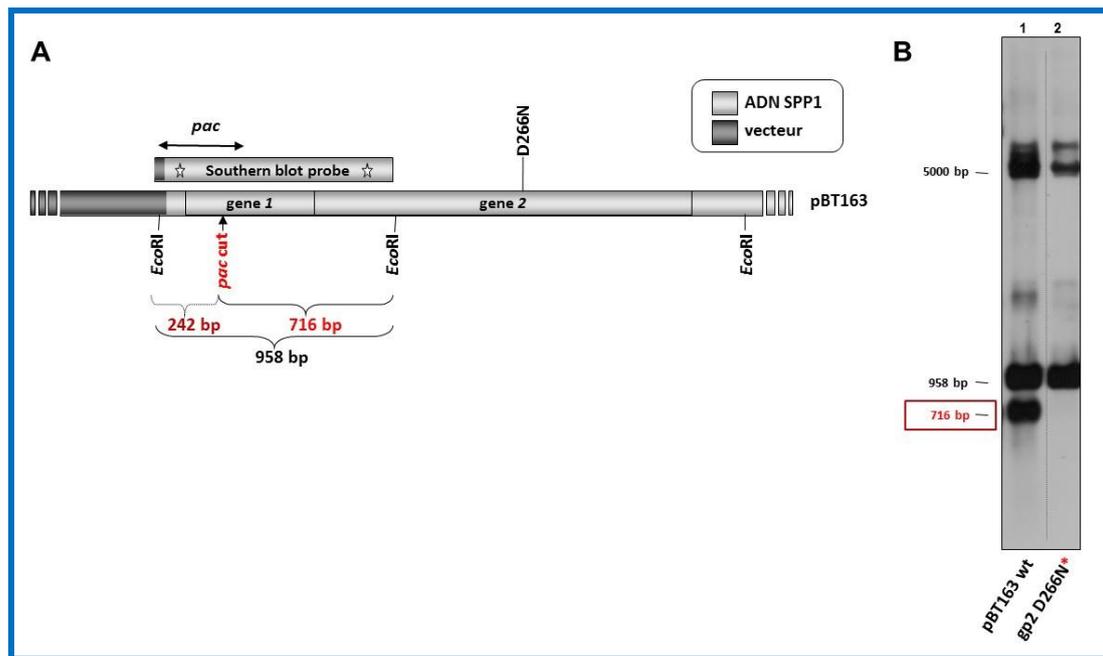


Figure 3.2 Test de clivage *pac* in vivo (A) Représentation schématique des gènes 1 et 2 dans pBT163 qui contient un fragment d'ADN de SPP1, comprenant les gènes 1 et 2 ainsi que la séquence *pac*, qui se trouve en partie à l'intérieur du gène 1. La flèche verticale indique le site de clivage *pac* situé dans le gène 1. Ce plasmide a été utilisé pour le test de clivage *pac* in vivo chez *B. subtilis*. Si le clivage *pac* a lieu, des fragments d'ADN de 716 et 242 pb sont attendus après digestion avec *EcoRI*. Ces fragments sont détectés après gel d'agarose suivi de Southern Blot. La région de l'ADN reconnue par la sonde d'hybridation utilisée pour le Southern Blot pour l'essai du clivage in vivo est indiquée (Cornilleau et al, 2013). (B) Southern blot de l'ADN digéré par *EcoRI* isolé à partir de la souche *B. subtilis* YB886 portant pBT163. La membrane a été hybridée avec la sonde représentée dans le panneau (A). puits 1) une bande de 716 pb a été détectée pour pBT163, en accord avec les tailles prédites des fragments après clivage de *pac* par gp2 et la digestion du plasmide avec *EcoRI*. En tant que témoin négatif (puits 2), le fragment de 716 pb n'a pas été observé avec un dérivé de pBT163 qui code pour la gp2 portant une mutation dans le site catalytique (gp2D266N; Cornilleau et al, 2013), qui bloque l'activité nucléasique de l'enzyme et donc le clivage de *pac*. En dépit de la reproductibilité de l'essai, le fragment de 242 pb prédit après clivage *pac*, n'a jamais été détecté, étant probablement dégradé par une nucléase après clivage sur *pac* (Cornilleau et al, 2013). Les bandes supérieures correspondent au plasmide super-enroulé et aux formes partiellement digérés.

Les résultats obtenus par Southern Blot montrent bien la présence du fragment de ~700 pb dans les mutations introduites dans la boîte **b2**, connue pour être le site préférentiel du clivage par la nucléase gp2, signe de que le clivage *pac* a bien eu lieu et que ces mutations n'ont pas d'effet détectable sur le clivage (Figure 3.3 Mut 1). Les mêmes résultats ont été observés pour les constructions avec des mutations dans les régions GCGGC qui se trouvent dans la boîte b1 et en amont de celle-ci (Mut 2-Mut 4). Les mutants construits afin de

dégénérer davantage la séquence *pacC* (Mut 5*-Mut 8*) n'ont montré aucun effet sur le clivage en *pacC* confirmant que la coupure par la nucléase n'a pas de spécificité de séquence.

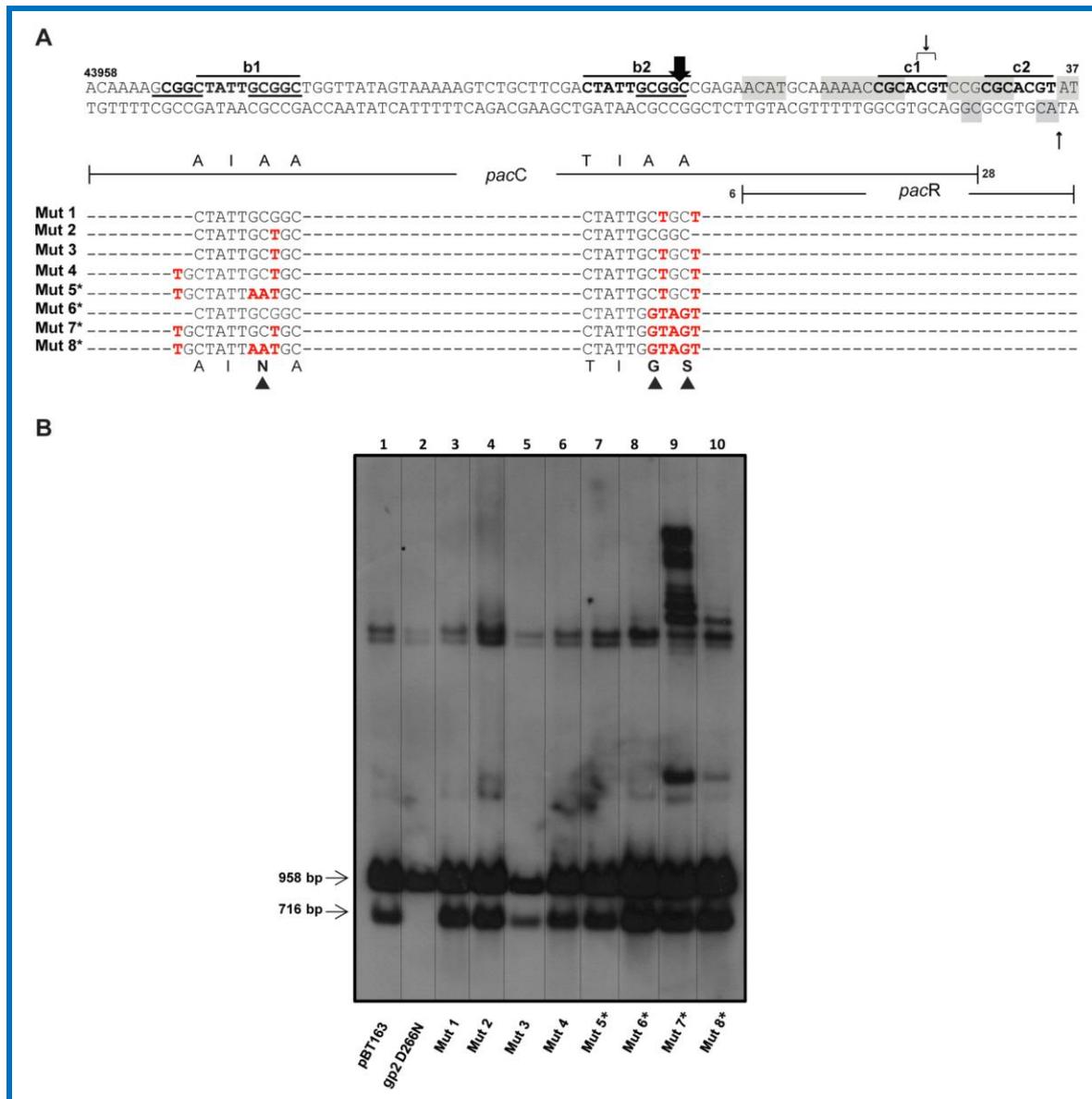


Figure 3.3. Effet des mutations dans *pacC* sur le clivage *pac*. (A) Mutations introduites dans la séquence *pacC* (Figure 3.1). (B) Southern Blot de l'ADN digéré par *EcoRI* extrait à partir de la souche *B. subtilis* YB886 portant les plasmides de contrôle ou des plasmides avec des mutations dans *pacC*. Les puits de gauche à droite sont: pBT163, contrôle positif; dérivé de pBT163 codant gp2 D266N contenant la mutation D266N dans le gène 2 qui bloque l'activité catalytique de la nucléase (Cornilleau et al, 2013), contrôle négatif; Mut 1-Mut 8*, mutants dans *pacC* comme affiché dans (A). L'expérience complète a été répétée trois fois indépendamment avec des résultats similaires.

3.1.2. Détermination du site de clivage *pac* chez SPP1 par PCR run-off et séquençage à haut débit

N'ayant pas observé de spécificité de séquence pour la coupure dans *pac*, du moins pas dans les boites **b1** et **b2** rapportés auparavant comme des cibles de la nucléase (Deichelbohrer et al, 1982 ; Gual et al, 2000), nous avons voulu identifier le site du clivage pour voir si celui-ci a lieu à une position définie malgré les mutations introduites dans les différentes boites ou s'il y a une certaine variabilité. Pour cela, nous avons entrepris une étude quantitative pour analyser les fragments dont les extrémités sont générées par le clivage sur *pac*. Cette analyse nous permettra non seulement d'identifier un site de coupure mais aussi d'observer sa fréquence.

La coupure du génome SPP1 par l'enzyme de digestion *ApaLI* produit 3 fragments de restriction dont un de 4646 pb résultant du clivage en *pac* par la terminase. Ce dernier comporte une extrémité naturelle due au clivage par la gp2 et une extrémité issue de la digestion par *ApaLI*. Le fragment *pac-ApaLI* a été purifié et soumis à un PCR run-off par séquençage de Sanger en utilisant un oligonucléotide non codant qui hybride à une distance de 199 pb de la position du site de clivage *pac*. L'analyse du chromatogramme (Figure 3.4A) a montré une réduction de la hauteur des pics à la position après le nucléotide G désigné par deux astérisques, indiquant la fin de nombreux fragments séquencés. Néanmoins, la lecture du chromatogramme se poursuit de 2 pb signalé par un pic à une hauteur plus faible, montrant une seconde population de fragments qui se termine à cette position (astérisque, Figure 3.4A). La séquence montre aussi la présence en aval de ces positions des nucléotides T, identifiés par des cercles points rouges, qui sont absents dans la séquence reverse complémentaire de *pac* de SPP1. Ceux-ci sont dus à l'addition de nucléotides A à l'extrémité de l'ADN, par la polymérase quand elle atteint l'extrémité d'un fragment d'ADN lors des réactions de séquençage (Clark et al, 1988). Cette analyse nous a permis d'identifier le site de clivage dans

la région *pac* en 3' de la boîte **b2** de *pacC* en mettant en évidence deux clivages préférentiels dans la position '5-GCGGCC↓GA-3' et un second site de clivage à 2 pb plus loin '5-GCGG↓CCGA-3' (représentés par des flèches noires creuses 3.4A). La présence d'un pic T non présent dans la séquence de SPP1 à 1 pb encore plus loin, indiquerait un potentiel autre site de coupure par gp2.

Dans l'objectif de quantifier la fréquence de clivage dans ces différentes positions et d'identifier une coupure principale, le fragment d'ADN *pac* a été également analysé par une technique de séquençage de nouvelle génération (NGS) Illumina, après réparation de l'extrémité *pac* et ligation des bouts francs à des adaptateurs. Les résultats présentés par un histogramme (Figure 3.4B), montrent un clivage principal à la position '5-GCGG↓CCGA-3' qui représente 50% des débuts des extrémités analysées. Des clivages à faible fréquence avant et après cette position sont observés avec une distribution presque symétrique. La majorité des clivages mise en évidence lors de cette étude est située au centre de l'hexa-nucléotide CGGCCG. Les résultats montrent que le clivage en *pac* produit des extrémités d'ADN *pac* hétérogènes, concordant avec les résultats observés par PCR run-off. Cette hétérogénéité n'est pas due à des contraintes expérimentales comme confirmé par les résultats obtenus avec l'extrémité *ApaLI*, utilisés comme contrôle, qui montre une coupure unique (100%) (Figure 3.4C). Ces résultats mettent en évidence un site de clivage en *pac* majoritaire dans '5-GCGG↓CCGA-3' à 1 pb en amont de la position précédemment décrite par Deichelbohrer et al. (1982).

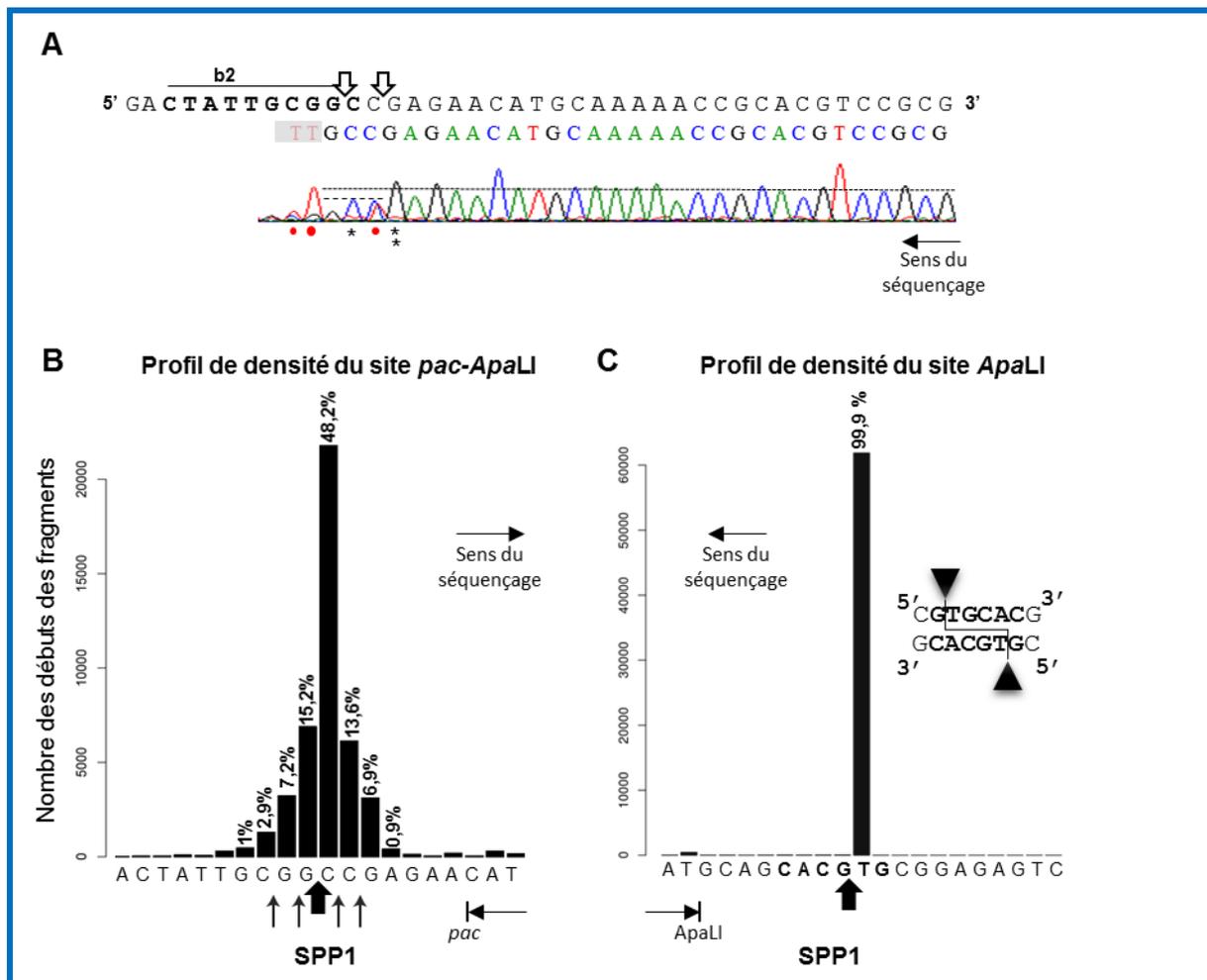


Figure 3.4. Identification du site de clivage de *pac* de SPP1. (A) Chromatogramme de séquençage du fragment *pac-ApaLI* obtenu par run-off. La séquence et le chromatogramme sont en brin inverse et complémentaire pour faire un alignement direct avec la séquence codante de la région *pac* où la boîte **b2** est identifiée. Le dernier nucléotide dont la hauteur du pic atteint l'intensité du signal de chromatogramme normal pour la même base est identifié par un double astérisque tandis que le nucléotide qui précède le signal au niveau de bruit de fond est identifié par un astérisque dans le chromatogramme. Les points rouges désignent les nucléotides T résultant de l'insertion artificielle d'un A à la fin de l'extrémité de l'ADN lors de la réaction de séquençage par la polymérase. Les sites de coupe *pac* identifiés sont indiqués par des flèches noires creuses. Des résultats identiques ont été obtenus dans trois expériences indépendantes. (B) Histogramme montrant la distribution des fragments *pac-ApaLI* dans la bibliothèque analysée par Illumina NGS qui commencent à la (aux) position(s) *pac* indiqué(s) en dessous. Le pic le plus élevé montre le nucléotide le plus fréquent au début des fragments *pac* de la bibliothèque, correspondant au site de clivage principal. Cette position est identifiée dans la séquence de SPP1 par une flèche noire pleine. Les clivages les moins fréquents, marqués par des flèches minces se trouvent dans les 8 paires de bases autour de la position de coupe principale *pac*. (C) Histogramme de la distribution des fragments d'ADN sur l'extrémité générée par la coupe *ApaLI* montrant que la coupe des endonucléases de restriction se produit à une position unique.

Nous avons montré que les mutations dans les boîtes **b** de *pacC* n'ont pas d'effet sur le clivage en *pac* détecté par le test de clivage *pac in vivo* (Figure 3.3). Nous avons alors voulu déterminer si ces changements ont un effet quantitatif sur la précision et/ou la position du

clivage. Pour cela, des phages qui contiennent les mutations qui dégénèrent le plus la région autour de la boîte **b2** (Mut 6*) ainsi que les deux boîtes **b** simultanément (Mut 8*) ont été construits. Les fragments *pac*-ApaLI de ces deux phages analysés par Illumina (NGS) (Figure 3.5) ont montré qu'il n'y avait pas de différence dans la position de coupure en comparaison avec le fragment *pac* sauvage (Figure 3.4B). On observe, cependant, une certaine augmentation de la fréquence de coupures autour de la position de clivage principale (Figure 3.5).

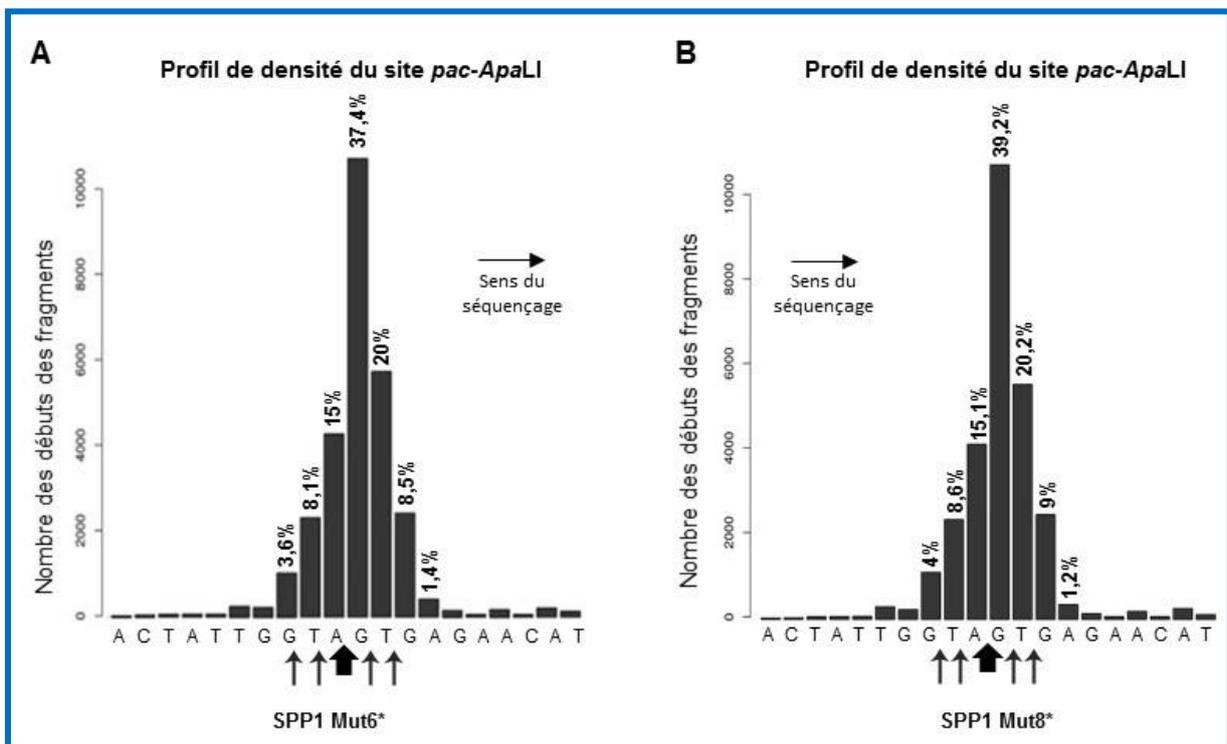


Figure 3.5. Identification du site de clivage *pac* de SPP1 Mut6* (A) et Mut8* (B). Histogramme montrant la distribution des nucléotides au début des fragments *pac*-ApaLI ainsi que leur pourcentage. L'approche expérimentale et la présentation des résultats sont identiques à ceux décrits dans la Figure 3.4B. Les flèches horizontales indiquent le sens du séquençage.

3.1.3. Détermination des sites de clivage dans *pac* des phages du groupe de SPP1

Au vu de ces résultats, nous avons voulu étendre l'analyse de la position du clivage en *pac* au groupe de phages proches de SPP1 afin de voir si les mécanismes impliqués dans ce

processus chez SPP1 sont conservés. Un séquençage des fragments *pac*-*Apa*LI a été réalisé sur les phages Lurz3 et SF6, comme précédemment décrit pour SPP1. Ces deux phages, ont été choisis en se basant sur l'arbre phylogénétique résultant de l'alignement multiple des séquences nucléotidiques de la région *pac* de SPP1 et ses phages proches (Figures 3.6 et 3.7).

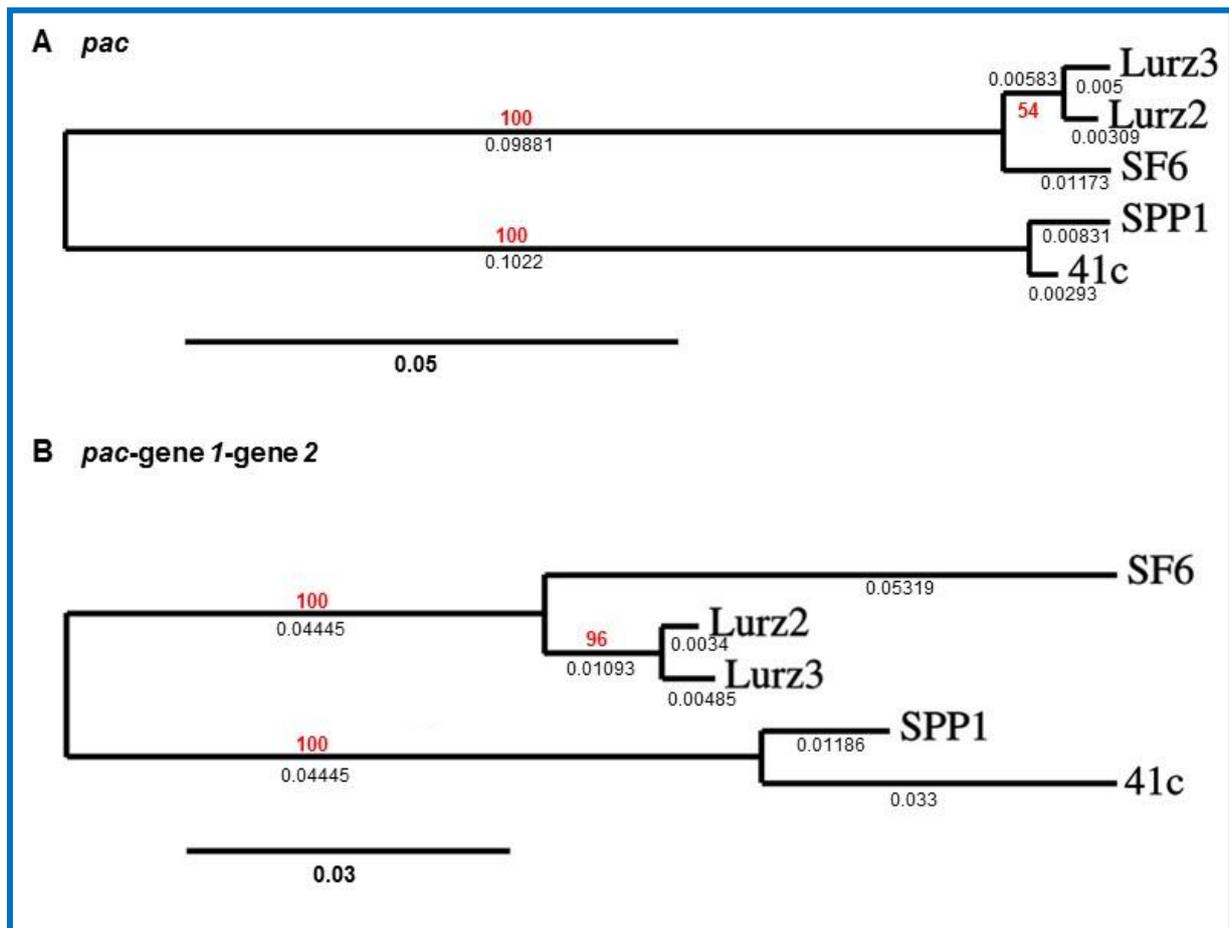


Figure 3.6. Arbres phylogénétiques de la région d'ADN de *pac* (A) et de *pac*-gène 1-gène 2 (B) de SPP1 et ses phages proches. Les arbres phylogénétiques ont été construits comme décrit dans la section 2.12.1 à partir d'un alignement de séquences avec ClustalW (<http://www.genome.jp/tools/clustalw/>). Les valeurs de confiance en pourcentage sont indiqués en rouge (Bootstraps : 1000). L'échelle montre la longueur des branches avec des valeurs indiquées en noir sous chaque branche. Ayant une similarité de séquence plus élevée, les valeurs de confiance du sous-groupe Lurz2, Lurz3 et SF6 augmentent pour la séquence complète de *pac*-gène 1-gène 2.

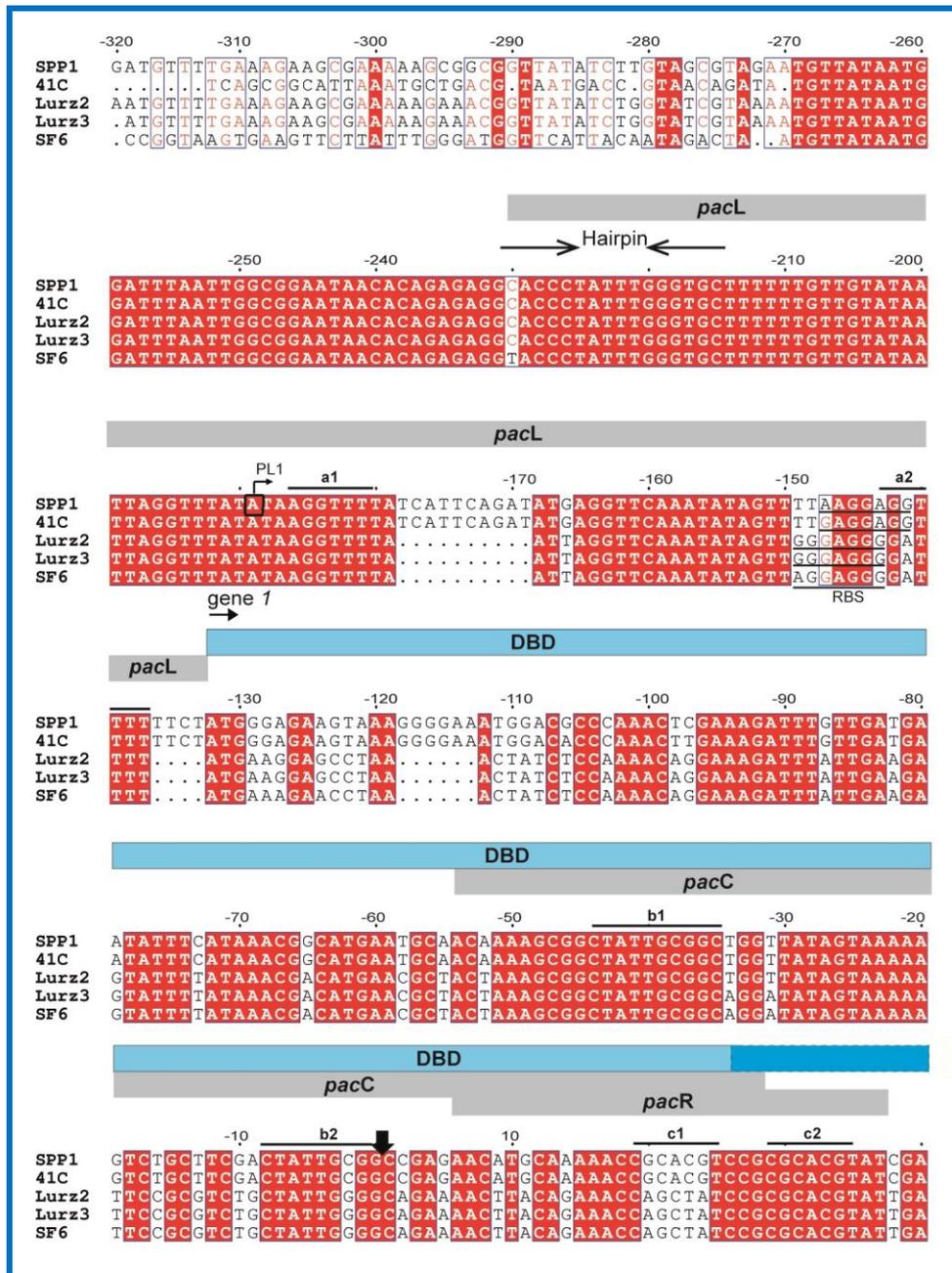


Figure 3.7. Alignement de séquence multiple de la séquence *pac* des phages du groupe de SPP1. Les cases rouges avec des caractères blancs indiquent une identité stricte, alors que les caractères rouges montrent les similarités au sein du groupe. Les sous régions de *pac* (*pacL*, *pacC* et *pacR*) sont indiquées au-dessus de l'alignement (barre grise) ainsi que les différentes boîtes. Le codon d'initiation du gène 1 et la région codante du domaine de liaison à l'ADN (DBD) de gp1 sont aussi indiqués (Büttner et al, 2012). La flèche noire verticale dans la séquence de SPP1 montre le site du clivage *pac* identifié par les expériences de NGS (Figure 3.4B). Les mêmes résultats ont été observés chez Lurz3 et SF6 (non montré). L'alignement multiple a été réalisé avec ClustalW (Larkin et al, 2007). La figure a été créée avec le logiciel ESPript (Robert et Gouet, 2004) puis adaptée.

Les résultats du séquençage run-off ont apporté les mêmes conclusions que chez SPP1 :

une première coupure suivie d'une deuxième coupure à une distance de 2 pb. Les résultats

obtenus par séquençage NGS ont permis l'identification d'un site de clivage principal à la position '5-GCGG↓CAGA-3' pour les deux phages ainsi que des coupures secondaires de part et d'autre (Figure 3.8). Cette étude montre que le clivage en *pac* Lurz3 et SF6 se passe à la même position que chez SPP1 mais avec des fréquences proportionnellement plus élevées aux deux positions adjacentes à celle du clivage principal.

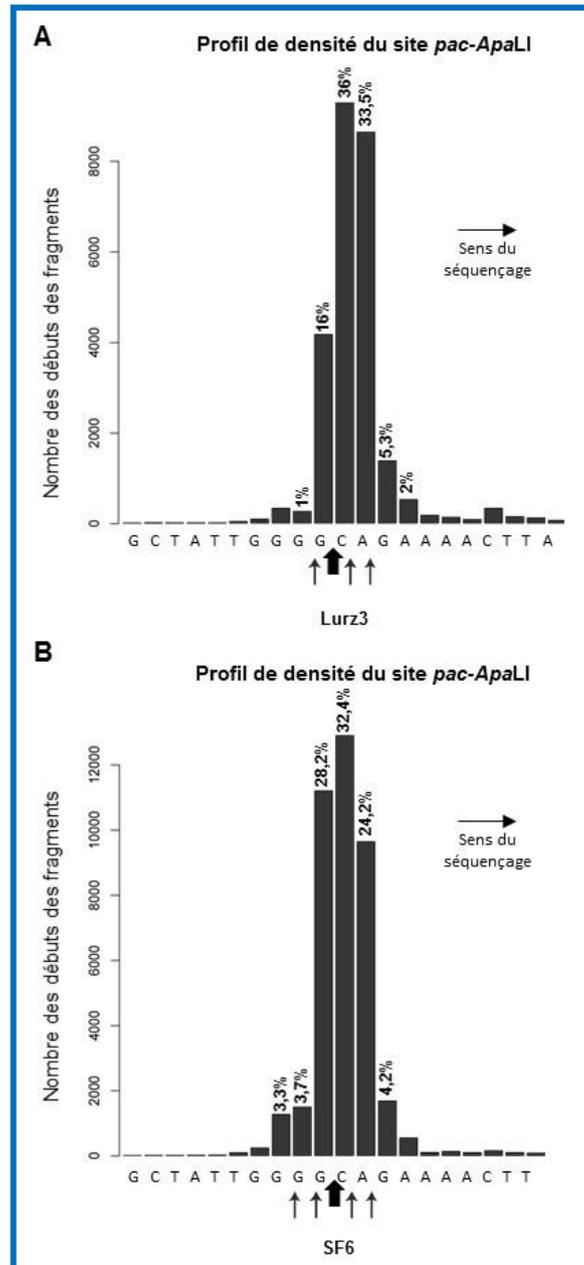


Figure 3.8. Identification du site de clivage *pac* des phages proches de SPP1, Lurz3 (A) et SF6 (B). Histogramme montrant la distribution des nucléotides des débuts des fragments ainsi que leur pourcentage. L'approche expérimentale et la présentation des résultats sont identiques à ceux décrits dans légende de la Figure 3.4B.

3.2. Etude des régions *pacL* et *pacR*, les sites d'interaction de la petite sous-unité de la terminase gp1

Une étude précédente de protection à la DNase des complexes gp1-*pac* a montré que la TerS de SPP1, gp1, se lie à deux sites différents de la séquence *pac*: *pacL* et *pacR*. En particulier, deux segments répétés de 7 pb appelés boîtes **a** et **c**, au sein de *pacL* et *pacR*, respectivement, ont été observés dans les zones protégées par la DNase (Chai et al, 1995). Cependant, la fonction de ces séquences répétées dans *pacL* et *pacR* et leur implication dans le mécanisme de clivage *pac in vivo* reste à démontrer.

3.2.1. Implication de la région *pacR* dans le clivage *pac*

Dans cette section nous présentons l'implication de la région *pacR* qui contient les boîtes **c1** et **c2** dans le clivage de *pac* par la terminase. Des mutations ont été introduites dans les boîtes **c1** et **c2** en premier lieu séparément, puis simultanément (Mut 9-Mut 11) (Figure 3.9). Des mutations ont été aussi introduites dans d'autres zones protégées par la protéine gp1. Une transversion C → A a été effectuée dans la zone protégée CCG entre les deux boîtes **c1** et **c2** (Mut 12) et une mutation en guanine a été introduite dans le motif AAAAA en une (Mut 13) ou deux positions (Mut 14) de la séquence poly-adénine qui se trouve en amont de la boîte **c1**. Ces deux dernières mutations ont aussi été combinées avec les mutations dans les deux boîtes **c1** et **c2** (Mut 15 et Mut 16). Les mutations dans les boîtes **c1** et **c2** n'ont montré aucun effet sur le clivage en *pac* (Figure 3.9) même lorsque la séquence est aussi changée entre les deux boîtes. Le clivage en *pac* a eu lieu également lorsque l'adénine en seconde position du motif AAAAA est changée en guanine (Mut 13). En revanche, la mutation additionnelle de l'adénine en dernière position bloque complètement le clivage en *pac* (Mut 14: AGAAG) révélant l'importance de cette région dans la reconnaissance et le clivage en *pac* par la

terminase.

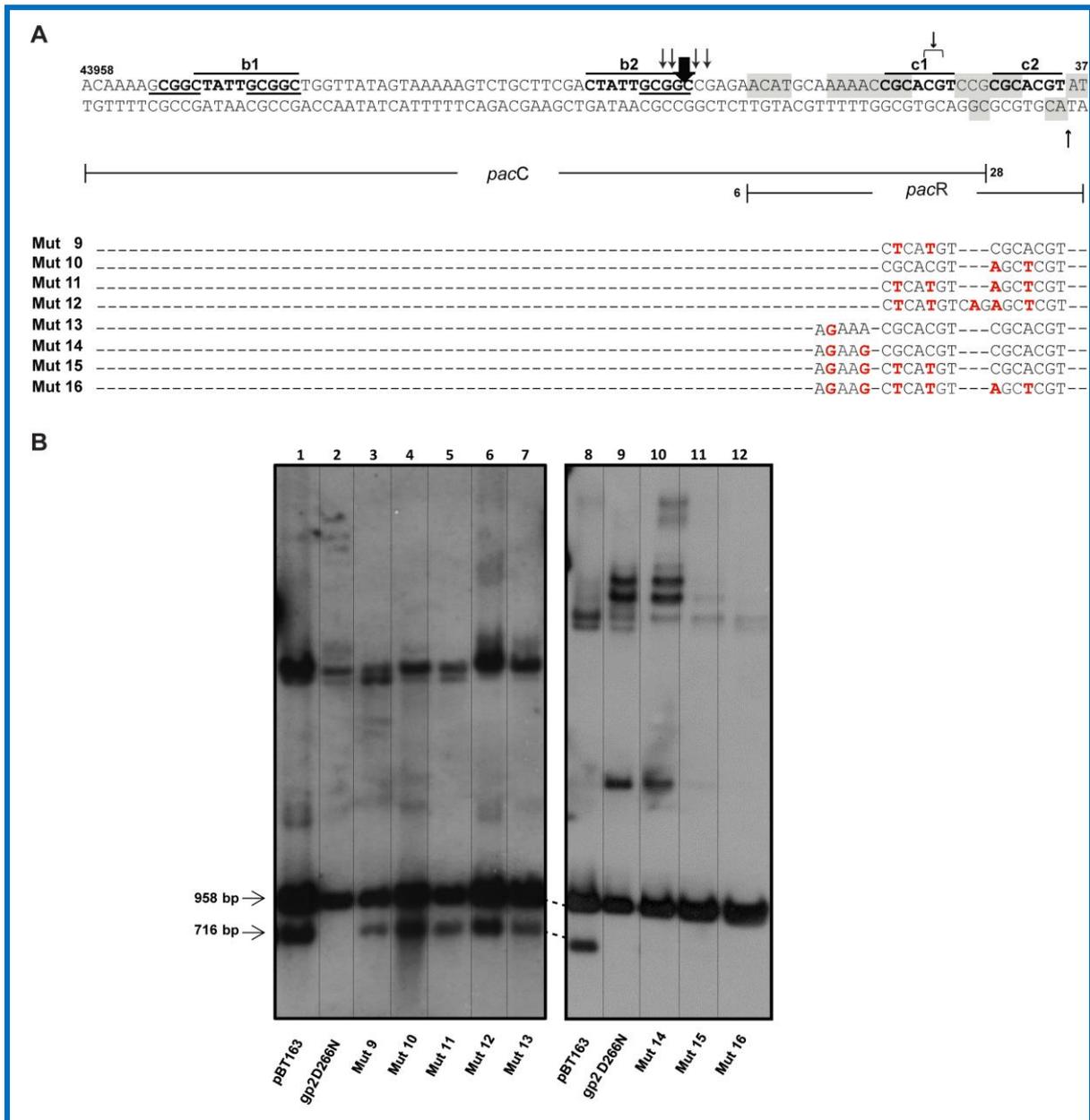


Figure 3.9. Effet des mutations dans *pacR* sur le clivage *pac*. (A) Mutations introduites dans la séquence *pacR*. La séquence et ses caractéristiques sont présentées dans la partie supérieure de la figure en incluant aussi la région *pacC* avec les boîtes **b** et les positions de clivage sur *pac* comme détaillé dans la Figure 3.3. Les boîtes répétées **c** de *pacR* sont indiquées en gras et les régions protégées dans des essais de protection à la DNase (Chai et al, 1995) sont indiquées par des zones grises, comme dans la figure 1.8. Les différents mutants construits sont indiqués en dessous avec les changements nucléotidiques introduits en rouge. (B) Southern Blot de l'ADN digéré par *EcoRI* extrait à partir de la souche *B. subtilis* YB886 portant le plasmide contrôle ou des plasmides contenant les mutations dans *pacR*. Les puits de gauche à droite sont: pBT163, contrôle positif; dérivé de pBT163 codant gp2 D266N contenant la mutation D266N dans le gène 2 qui bloque l'activité catalytique de la gp2 (Cornilleau et al, 2013), contrôle négatif; Mut 9-Mut 16, mutants *pacR* comme décrits dans (A). L'expérience complète a été répétée trois fois indépendamment.

3.2.2. Rôle de la région *pacL*, non encapsidée, dans le clivage *pac*

Après reconnaissance et clivage de *pac* par la terminase, le génome viral est encapsidé de façon unidirectionnelle. La région *pacR* est insérée à l'intérieur de la procapside, tandis que *pacL* est dégradée. *PacL* contient plusieurs zones protégées par la protéine gp1 lors de la digestion à la DNase (Figure 3.10) (Chai et al, 1995) ainsi que deux segments répétés appelés **a1** et **a2**, suggérant une implication de cette région dans la reconnaissance et le clivage en *pac*. Ceci reste à déterminer. La délétion de *pacL* à partir du plasmide pBT163 qui exprime les gènes 1 et 2 ainsi que la totalité de la séquence *pac* a été effectuée par sous-clonage tel que décrit dans la section 2.2.3. La région *pacL* est située en amont du gène 1, ce qui a permis d'envisager sa délétion afin de déterminer si elle a un rôle dans le clivage de *pac*. Ceci dit, il n'est pas possible de faire la délétion complète de *pacL* puisque celui-ci chevauche avec la séquence de liaison au ribosome (RBS) (Figure 3.10A). La délétion la plus longue testée a été celle allant jusqu'avant le site RBS, *pacL*-15 ne gardant que 15 pb de l'intégralité de *pacL* (Figure 3.10). On note aussi que malgré la délétion des promoteurs tardifs PL1 et PL2 de SPP1, les gènes 1 et 2 sont transcrits à partir du promoteur situé en amont du gène conférant la résistance au chloramphénicol dans pBT163. Le choix des délétions de la séquence *pacL* a été basée sur : i) la présence des boîtes **a1** et **a2**, ii) les zones de liaison à la gp1 mises en évidence par le test de protection à la DNase, iii) la présence des motifs répétés T, qui peuvent indiquer la présence de courbure dans la structure de la double hélice de l'ADN favorisant la liaison avec la terminase. Sur ces critères, onze constructions ont été réalisées (Figure 3.10). Le clivage sur *pac* a été observé pour les constructions *pacL*-96, -93, -88, -82 et -76 (Figure 3.10, puits 5 à 9). Le clivage en *pac* a été complètement bloqué pour toutes les constructions allant de *pacL*-15 jusqu'à *pacL*-58. La construction *pacL*-68 (puits 10) a aussi un effet sur le clivage en *pac* mais ne le bloque pas complètement. C'est à partir de la construction *pacL*-76

que le phénotype sauvage est rétabli. Ces résultats montrent qu'une région de 76 pb de *pacL* est nécessaire pour permettre la reconnaissance et la coupure endonucléolytique de *pac* par la terminase dans le système rapporteur pBT163.

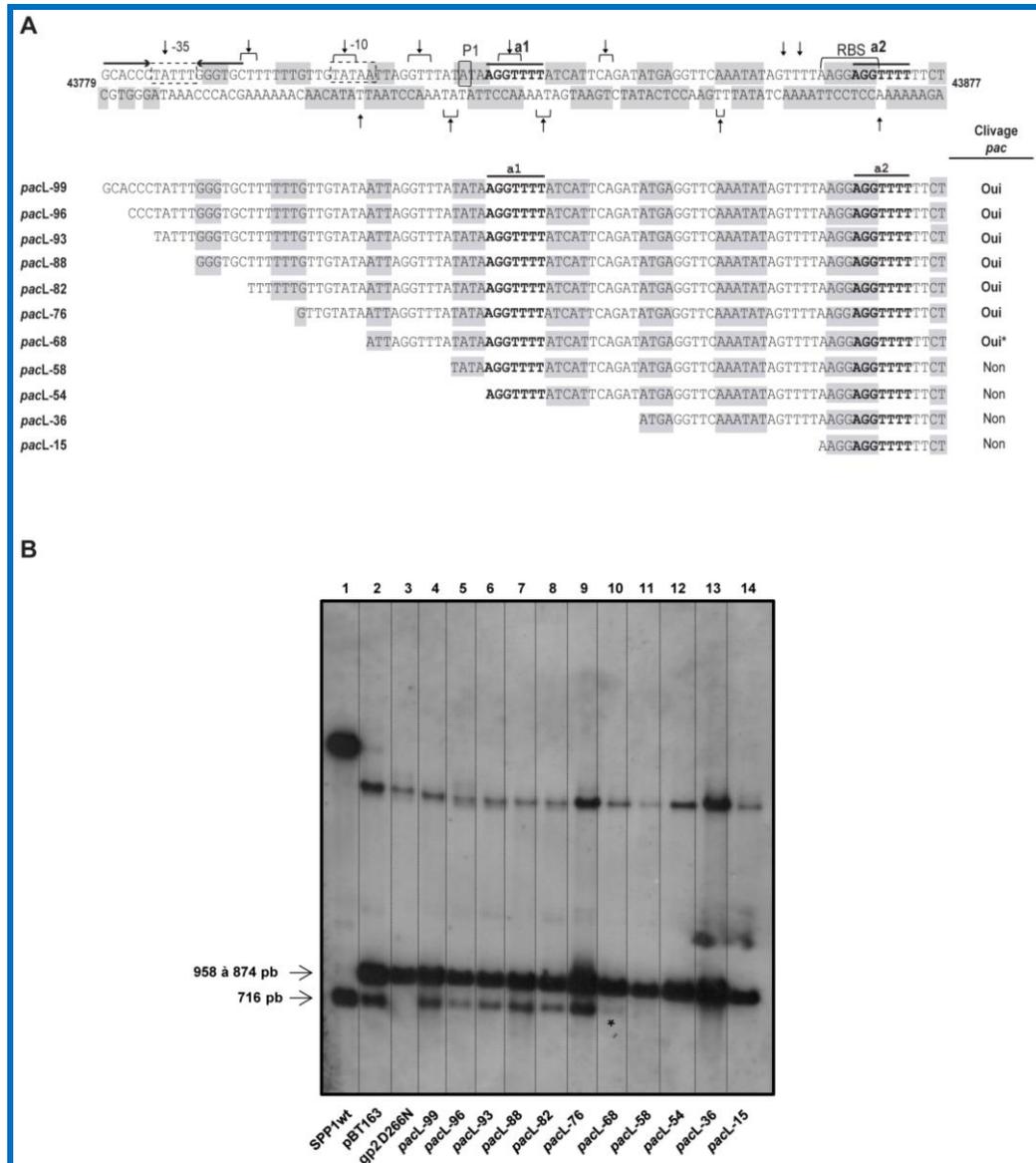


Figure 3.10. Effet des délétions *pacL* sur le clivage *pac*. (A) Représentation schématique de la région *pacL* et des constructions de délétions. La taille de la région initialement identifiée comme *pacL* a une longueur de 99 pb, correspondant aux coordonnées 43779 à 43877 de la séquence de SPP1. Elle se termine au nucléotide précédant le codon d'initiation du gène *I*. Les deux boîtes répétées **a** sont marquées. Les différentes constructions réalisées sont de 96-15 pb de la région *pacL* comme indiqué par leur nom à gauche de la figure. Les zones de la séquence marquées en gris indiquent la séquence protégée par gp1 dans les essais de protection à la DNase (Chai et al, 1995). (B) Southern Blot de l'ADN digéré par *EcoRI* isolé à partir de la souche *B. subtilis* YB886 portant les contrôles (puits 1 : SPP1wt : Contrôle positif avec l'ADN extrait des phages et qui contiennent l'extrémité issue du clivage sur *pac* ; puits 2 et 3 : contrôles positif et négatif comme dans la Figure 3.3 ; Puits 4 : construction codant *pacL*-99, similaire à celle dans le plasmide pBT163 et les différentes délétions de *pacL* (puits 5 à 14). L'astérisque de *pacL*-68 indique la présence d'un fragment de 716 pb en très faible intensité. Trois expériences indépendantes ont été répétées avec les mêmes résultats.

3.2.3. Test de clivage *pac* *in vitro* et *in vivo* dans *E. coli*

Après l'étude de la région *pac* et l'implication de ses différentes régions dans la reconnaissance et le clivage de *pac in vivo*, l'objectif suivant a été de décortiquer la mécanistique des différentes étapes de reconnaissance et de clivage dans *pac*.

Nous nous sommes basés sur un système de transcription-traduction *in vitro* pour la production de gp1 et gp2 en présence de *pac*. Le plasmide pKD33 contenant un fragment d'ADN de SPP1 incluant *pac*-gène 1-gène 2 sous le contrôle d'un promoteur T7 (Figure 2.1) a été utilisé dans une réaction de transcription-traduction *in vitro* avec le système RTS 100 *E. coli* HY. Un western blot a été réalisé (Figure 3.11) sur l'extrait brut de la réaction mais aussi à partir du culot et du surnageant afin de vérifier la production et la solubilité des protéines, gp1 et gp2. Comme contrôle positif pour confirmer que le système RTS fonctionne nous avons utilisé un plasmide codant la protéine GFP (puits 4, 7 et 10), et comme contrôle négatif le mix de réaction sans plasmide (puits 3, 6 et 9). Les protéines gp1 (16 kDa) et gp2 (48,8 kDa) pures ont été utilisées comme contrôles positifs (puits 1 et 2, respectivement). Les résultats obtenus montrent que la protéine gp1 est produite en grande quantité à partir du vecteur pKD33. Elle est présente dans l'extrait brut (puits 5) et se retrouve essentiellement dans le surnageant (puits 11). Quant à la protéine gp2, elle n'est observée que dans l'extrait brut de la réaction dans des quantités très faibles (astérisque dans le puits 5). Pour la suite de l'expérience, nous avons donc travaillé avec le vecteur PKD33 à partir de la réaction brute où les deux protéines sont présentes.

Un test de clivage *pac* sur l'ADN présent dans la réaction de transcription-traduction *in vitro* selon la méthode utilisée dans la section 2.5 n'a pas mis en évidence une coupure de la séquence *pac* présente dans pKD33. Nous avons émis l'hypothèse que ceci était due au contexte de l'expérience utilisant un système d'*E. coli*, sans pour autant exclure la possibilité que la gp2 ne soit pas présente en quantité suffisante dans le système, d'après les résultats du

western. Un test de clivage *pac* *in vivo* dans *E. coli* a alors été réalisé (Figure 3.12), montrant clairement qu'aucun clivage en *pac* n'est observé (Puits 4). Ceci indique qu'un facteur présent dans *B. subtilis* est fort probablement requis pour le clivage *pac* par la terminase et met en évidence une relation hôte-pathogène dans ce processus, élément qui n'a pas été décrit auparavant pour SPP1.

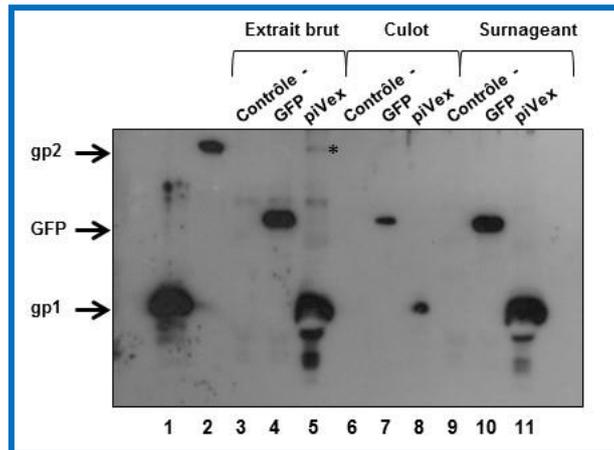


Figure 3.11 : Western Blot pour l'analyse de la production des protéines gp1 et gp2 dans le système RTS *E. coli*. puits 1 : gp1 pure, 16 kDa ; puits 2 : gp2 pure, 48,8 kDa ; puits 3, 6, 9 : contrôle négatif, réaction système RTS sans plasmide dans l'extrait brut, le culot et le surnageant respectivement. puits 4, 7, 10: plasmide avec GFP dans extrait brut, culot et surnageant ; puits 5, 8, 11: pKD33 codant gp1 et gp2 avec une étiquette hexa-histidine dans l'extrait brut, le culot et le surnageant, respectivement. La gp2 dans le puits 5 faiblement produite est indiquée par un *.

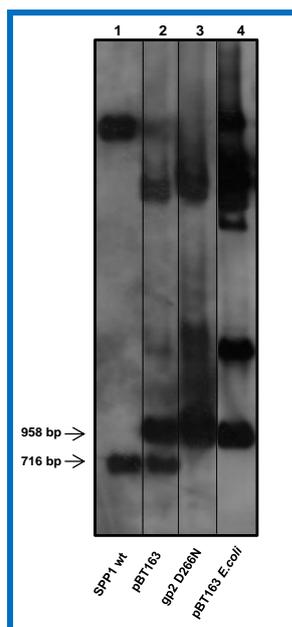


Figure 3.12. Test de clivage *pac* *in vivo* dans *E. coli*. Southern Blot de l'ADN digéré par *EcoRI* extrait à partir de particules purifiées de SPP1, de la souche *B. subtilis* YB886 (pBT163) ou de la souche *E. coli* (pBT163). La membrane a été hybridée avec une sonde homologue de l'ADN dans la région *pac* comme décrit dans Matériels et Méthodes. 1) SPP1 wt: ADN extrait à partir du phage sauvage, 2) pBT163 dans YB886, contrôle positif; 3) dérivé de pBT163 codant gp2 D266N dans YB886 contenant la mutation D266N dans le gène 2 qui bloque l'activité catalytique de la nucléase (Cornilleau et al, 2013), contrôle négatif; 4) pBT163 dans *E. coli*, contenant les gènes 1 et 2 ainsi que la séquence *pac*.

3.3. Etude de l'interaction gp1-ADN

L'étude de la séquence *pac* et des éléments nécessaires de cette région pour la reconnaissance et le clivage de *pac*, nous a permis de diriger les travaux vers l'étude des bases moléculaires impliquées dans la formation du complexe ADN-gp1 qui est l'étape clé pour l'initiation du processus d'encapsidation. En effet, malgré les connaissances que nous avons sur la reconnaissance et le clivage en *pac*, très peu d'information sur l'interaction gp1-ADN a été dévoilée. Puisque nous avons pu déterminer quels sont les sites nécessaires pour la reconnaissance de l'ADN par la gp1, la perspective immédiate a été d'étudier l'interaction gp1-*pac* avec pour objectif de déterminer la structure du complexe gp1-ADN.

Nous avons entamé cette étude. Dans la section suivante, nous présentons la démarche entreprise pour déterminer les résidus de la protéine gp1 impliqués dans l'interaction avec l'ADN. Nous avons étudié la structure cristallographique de la protéine gp1 du phage SF6 proche de SPP1 (Figure 1.9) partageant une identité de séquence de 84%. La protéine gp1 entière comprend une région N-terminale qui se lie à l'ADN (DBD pour « DNA-binding domain »), un corps oligomérique et une région C-terminale en tonneau de feuillets β . L'étude structurale a montré que ce polypeptide s'assemble en nonamère sous forme circulaire avec les DBDs organisés autour du corps oligomérique de manière hautement flexible (Figure 1.9, Büttner et al, 2012). Ces DBDs sont formés par 4 hélices α , α 1 [P10- F20], α 2 [A26- A30], α 3 [A40- N47], α 4 [P51- I57] (coordonnées chez SPP1) (Figures 3.13 - 3.15). L'identification des résidus potentiellement impliqués dans l'interaction avec *pac* a été basée dans un premier temps sur un logiciel de prédiction des interactions avec l'ADN, BindN+ (Wang et al, 2010). Ce logiciel a mis en évidence des résidus qui se trouvent essentiellement dans le DBD (Figure 3.13).

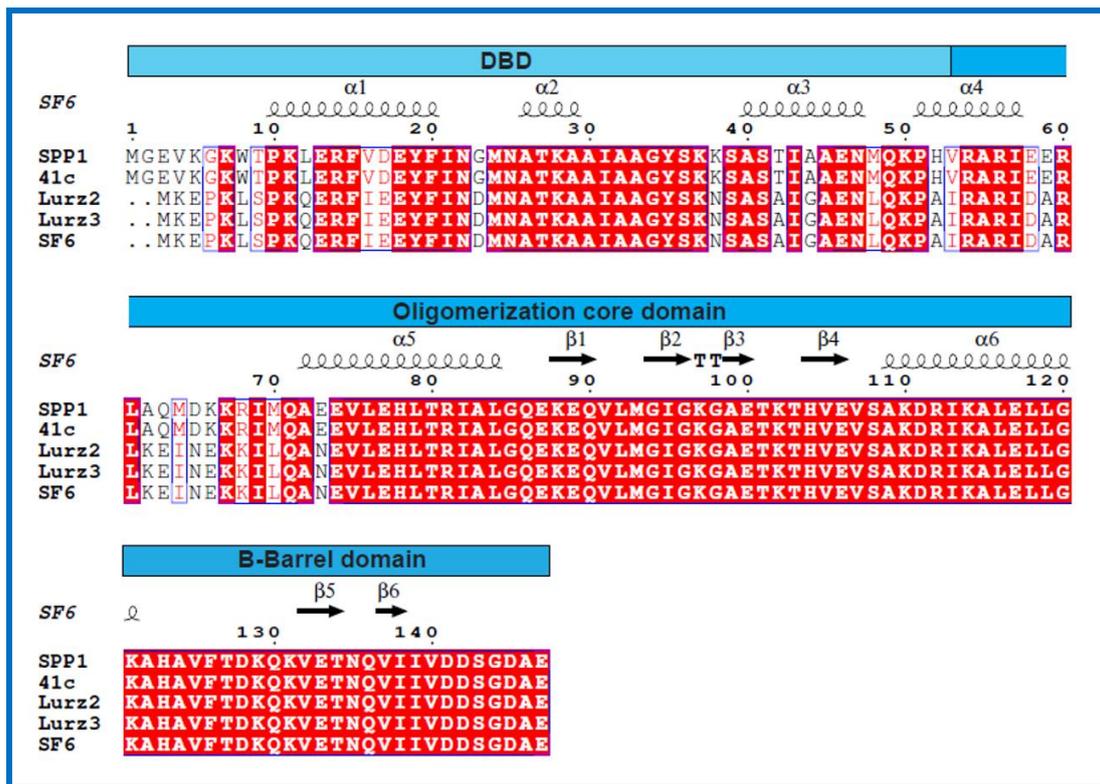


Figure 3.14. Aligement multiple de séquences des acides aminés de gp1 de SPP1 et ses phages proches en utilisant le logiciel ClustalW. La figure a été préparée avec ESPript. Les brins d'hélices alpha et bêta sont représentés au-dessus des alignements basés sur la structure cristallographique de SF6 (Büttner et al, 2012). Les cases rouges avec des caractères blancs indiquent une identité stricte, alors que les caractères rouges indiquent les similitudes de séquences.

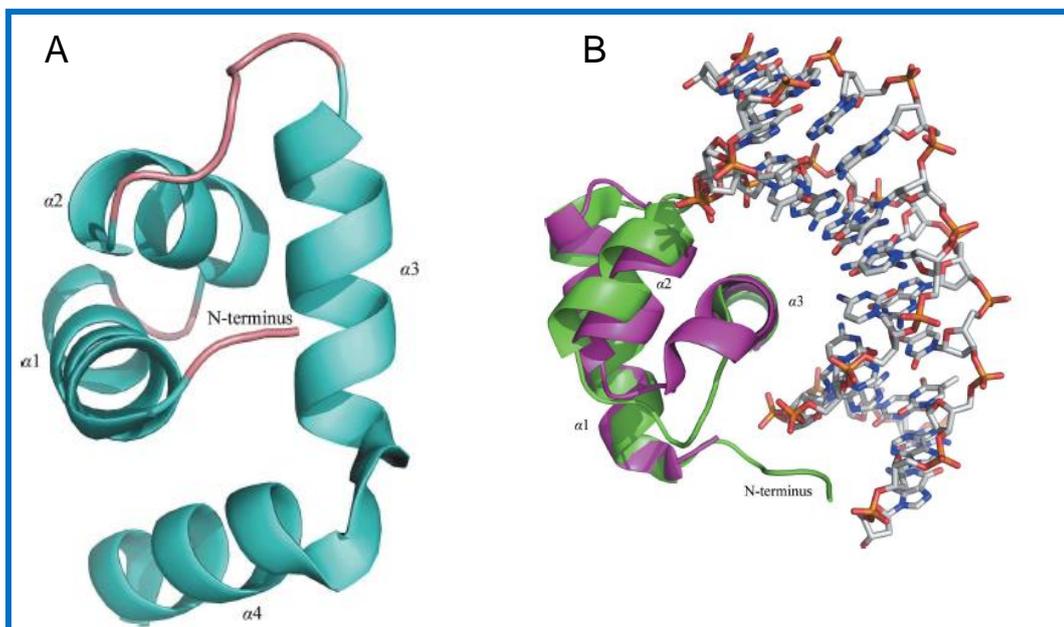


Figure 3.15. Domaine de liaison à l'ADN (DBD) correspondant à la région N-terminale de la protéine gp1 de SF6 (A) et superposition de la région N-ter de gp1 en vert avec le DBD de la protéine télomérique humaine TRF1 (en violet) en complexe avec l'ADN (B) (Benini et al, 2013).

Ensemble, ces analyses nous ont permis de mettre en évidence des résidus basiques qui se trouvent à la surface de la protéine et qui sont potentiellement impliqués dans l'interaction avec l'ADN. Même si le modèle de superposition dans la Figure 3.15B suggère une implication majeure de l'hélice α_2 , nous avons décidé de prendre en compte les résidus sur les autres hélices y compris l'hélice α_4 qui a probablement pour rôle la liaison du DBD au corps d'oligomérisation. Ainsi, neuf résidus ont été choisis et 12 mutants ont été construits par mutagenèse dirigée (Figure 3.16 et tableau 3.1) à partir du plasmide pCC5, dans l'objectif d'utiliser l'étiquette hexa-histidine dans des expériences de chromatographie par affinité ou de ChAP.

Tableau n°3.1: Mutants gp1 construits par mutagenèse dirigée.

Résidus cible	K7	T9	K11	T27	K28	K28	T27K28	K37	K38	K37 K38	K50	R56
Mutations	E	A	E	A	A	E	A	E	E	E	E	E

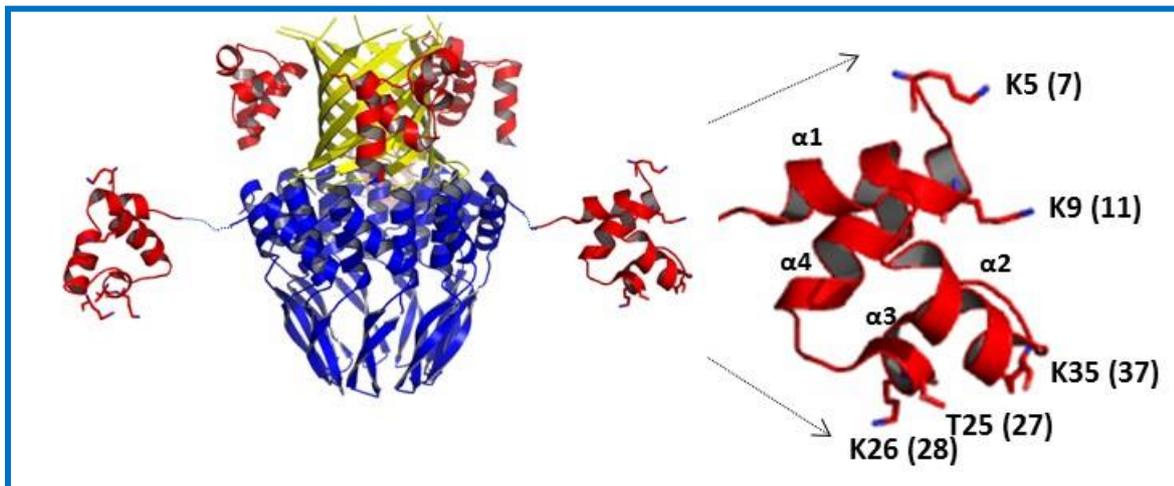


Figure 3.16. Représentation de la structure cristallographique de la protéine gp1 du phage SF6 (homologue de SPP1). Le corps d'oligomérisation est présenté en bleu, le tonneau en feuillet β qui est la région C-terminale de la protéine en jaune et les domaines de liaison à l'ADN (DBD) en rouge. L'agrandissement du DBD montre quelques résidus qui ont été mutagénisés dans cette étude pour tester leur rôle dans l'interaction du DBD avec l'ADN. L'équivalent de ces résidus chez la gp1 de SPP1 est indiqué entre parenthèses.

Ces mutants construits, nous avons ensuite analysé leur effet sur la reconnaissance et le clivage en *pac in vivo* dans *B. subtilis* par le test de clivage *pac*. Nous avons aussi

déterminé leur effet sur la production des phages par un test de complémentation *in vivo* en les infectant avec le phage *sus70* qui est déficient dans la production de gp1.

Les résultats obtenus avec le test de clivage *pac in vivo* ont montré que tous les mutants à l'exception de R56E affectent le clivage en *pac*. Pour la plupart des mutations l'effet est drastique, avec aucun clivage détecté, tandis qu'une réduction plus faible mais significative du clivage *pac* a été observée pour les mutants K28A et K37E (Figure 3.17).

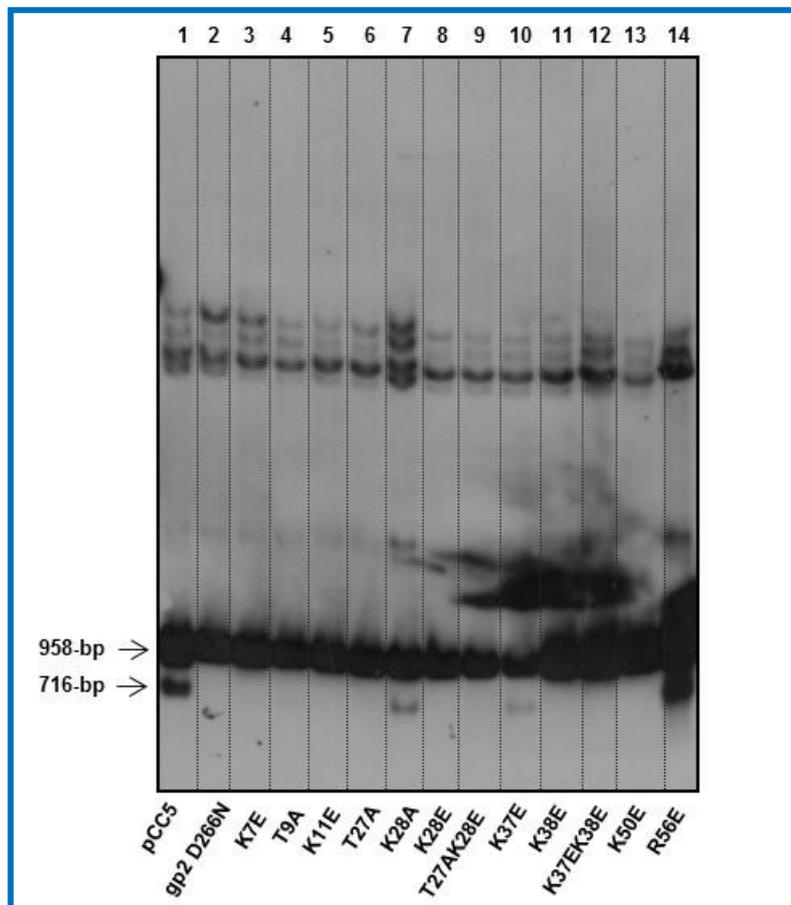


Figure 3.17. Effet des mutants gp1 sur le clivage en *pac*. Southern blot des ADN digérés par *EcoRI* des plasmides pCC5 et ses dérivés codant des gp1 mutants extraits à partir de *B. subtilis*. De gauche à droite; puits 1: pCC5: Contrôle positif ; puits 2: gp2 D266N : Contrôle négatif ; puits 3-14 mutants gp1.

Les résultats de complémentation avec SPP1*sus70* ont quant à eux conduit à des résultats différents. Tous les mutants gp1 ont un effet sur la production des phages, mais les niveaux de complémentation sont variables (Figure 3.18). Certaines mutations qui bloquent totalement le clivage *pac* ont un effet drastique sur la production de phages, cas du double

mutant K37K38 et du mutant K11. Á contrario, le mutant R56 qui n'a pas d'effet sur le clivage en *pac*, a un effet sur la fonction de gp1. De plus, tous les mutants, quelle que soit l'hélice du DBD mutée, ont présenté un phénotype avec une production de petites plages de lyse (Figure 3.19) indiquant une perturbation dans la multiplication du phage. Pour les mutants qui complètent partiellement la fonction de la gp1 tronquée produite par infection avec le phage SPP1_{sus70} (Figure 3.18) alors qu'ils bloquent le clivage en *pac* dans le test *in vivo*, nous suggérons que ces mutations permettent probablement la production d'une faible population de phages dans le contexte de l'infection par SPP1. L'initiation de l'encapsidation de leur génome peut avoir lieu soit à partir d'un clivage *pac* moins efficace et à un niveau qu'on ne peut observer avec le niveau de sensibilité du test de clivage *pac* avec le système plasmide utilisé dans cette étude, soit à partir d'une séquence alternative autre que la séquence *pac*. Pour le mutant R56E, ses phénotypes suggèrent que cette mutation affecte probablement une étape autre que la reconnaissance et le clivage en *pac*, permettant ainsi de déclencher le clivage mais pas de produire efficacement une population de virions infectieux.

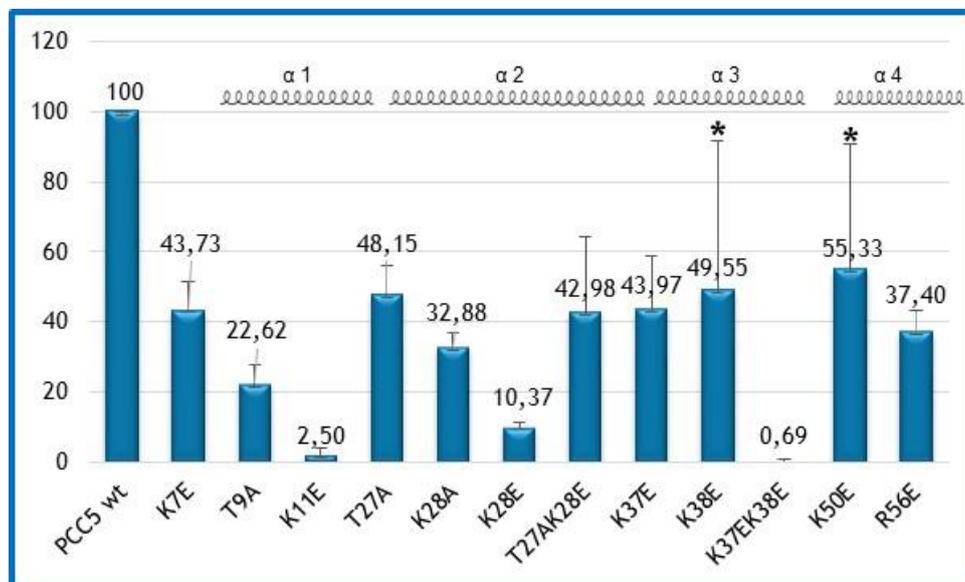


Figure 3.18. Test de complémentation *in vivo* des mutants gp1 avec SPP1_{sus70}. L'effet des mutations dans gp1 a été quantifié par titration et indiqué en pourcentage par rapport au plasmide contenant gp1 sauvage (pCC5) qui représente 100%. Les * indiquent les mutants dont le test a été fait deux fois.

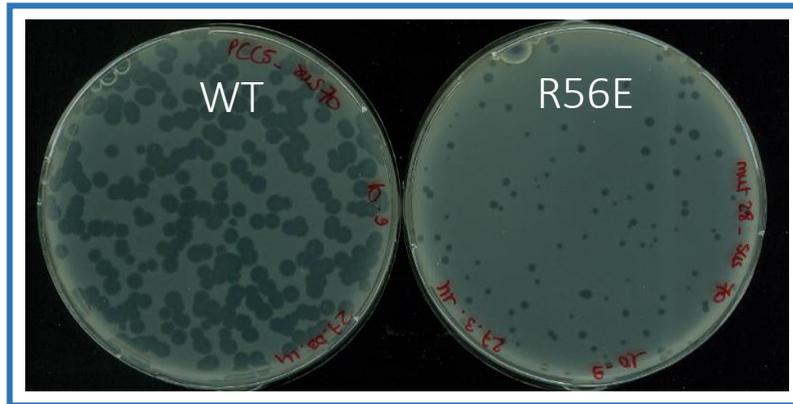


Figure 3.19. Phénotype des plages de lyse observé lors des tests de complémentation *in vivo*. A gauche sont présentées les plages de lyse (larges) produites en infectant des bactéries contenant gp1 (plasmide pCC5). A droite le mutant R56E qui présente des plages de lyse petites (comme chez tous les mutants sur gp1) et qui n'a pas d'effet sur le clivage en *pac* (Figure 3.17).

Chapitre 4 : Discussion

4. Discussion

L'encapsidation du génome viral dans la procapside est un processus biologique complexe qui est essentiel pour la multiplication virale. Il comprend plusieurs étapes séquentielles, la première étant la reconnaissance et le clivage d'une séquence spécifique de l'ADN viral concatémère. De nombreux phages utilisent une séquence appelée *pac*. La liaison à la séquence *pac* et le clivage qui suit, créant l'extrémité libre d'ADN qui sera insérée dans la procapside, est assurée par deux protéines non structurales, qui forment le complexe terminase: la petite sous-unité de la terminase, gp1 dans le bactériophage de SPP1, et la grande sous-unité de la terminase, gp2, dans SPP1. La séquence *pac* de SPP1 est quant à elle constituée de 3 sous-régions, *pacL* (région « left » non encapsidée), *pacR* (région « right » encapsidée) et *pacC*, région contenant le site de clivage. La complexité de ce système consiste dans le fait que les deux sous-unités de la terminase sont nécessaires pour la reconnaissance et coupure de l'ADN sur *pac*. Gp2 possède l'activité nucléase nécessaire au clivage et est également une ATPase. Gp1 qui, d'après des études *in vitro* de protection par la DNase, reconnaît et se lie aux régions *pacL* et *pacR* qui entourent le site de clivage *pacC* (Chai et al, 1992; Chai et al, 1995; Cornilleau et al, 2013). Une autre particularité réside dans le fait que la gp1 reconnaît des séquences dans sa propre séquence codante, puisque une partie de *pac* et notamment les domaines *pacC* et *pacR* se situent dans le gène qui code pour cette protéine. Enfin, le clivage *pac* est un événement rare et très bien régulé. Le présent travail avait pour but de contribuer à clarifier les mécanismes sous-jacents à la reconnaissance et coupure de *pac* par cette enzyme hétéro-oligomérique, qui ne sont compris pour aucun virus. Nous nous sommes intéressés au rôle des différents domaines de *pac* ainsi qu'à la question de la spécificité du clivage.

4.1. Identification du site de clivage dans *pacC*

Des études par NGS et run-off ainsi que la mutagénèse ciblée sur la région *pacC* ont permis de découvrir une propriété originale de cette interaction protéine-ADN complexe. Nous avons montré que la coupure endonucléolytique qui initie l'encapsulation de l'ADN *in vivo* a lieu au centre d'une séquence de 6 pb à une position précise de *pacC*, avec une variation de la coupure de part et d'autre de la position centrale, mais sans spécificité de séquence. Ceci a été démontré avec les expériences de mutagénèse dirigée sur la séquence *pacC* et en particulier sur ses boîtes **b**. Les mutations ont été introduites de façon à perturber la séquence en nucléotides, sans altérer la séquence d'acides aminés de la gp1 dont la région codante chevauche *pacC* et *pacR*. Ainsi, la fonction de la protéine gp1 n'est pas touchée, et si le clivage de *pacC* était spécifique de la séquence de la boîte **b** 5'-CTATTGCGGC-3', un phénotype serait attendu. Nos résultats sont en adéquation avec les données rapportées auparavant, indiquant la non spécificité de séquence de l'activité nucléasique de la gp2 *in vitro* (Gual et al, 2000 ; Smits et al, 2009 ; Cornilleau et al, 2013). Les travaux de Gual et al, 2000 ont montré que l'utilisation de la distamycine rentrait en compétition avec la protéine gp1. Cette drogue altère la conformation de l'ADN en se fixant sur les domaines riches en A/T et mime l'interaction gp1-*pac*. Ces expériences ont montré que l'introduction de courbure dans la région *pac* induisait des coupures dans *pacC* par la nucléase. Ils ont aussi montré qu'en présence d'un ADN super-enroulé contenant la séquence *pac*, la nucléase introduisait plusieurs petites coupures dans l'ADN sans montrer de spécificité de séquence. Nos résultats montrent que malgré ce manque de spécificité de la nucléase, le complexe terminase contrôle très efficacement son activité et la conditionne pour un clivage à une position précise de *pacC*.

4.2. Les régions *pacL* et *pacR* sont nécessaires au clivage de *pac*

Les travaux présentés nous ont aussi permis d'identifier les éléments nécessaires dans *pacL* et *pacR* pour le clivage sur *pac*.

D'une part, l'étude de la région *pacL* par délétions successives sans affecter la séquence codante de la *gp1* a été possible car la région *pacL* est située en amont du gène *I*, ce qui n'est pas le cas pour les régions *pacC* et *pacR*. La région *pacL* de 99 pb est caractérisée par la présence d'une région de courbures de l'ADN intrinsèques (Chai et al, 1995) et par plusieurs motifs protégés par la *gp1*. Ces motifs sont séparés par des sites d'hypersensibilité à la DNase éloignés de 10 pb ce qui est l'équivalent d'un tour d'hélice d'ADN, suggérant que la région *pacL* est enroulée par la *gp1* qui a une structure en tonneau. En effet, *gp1* de SPP1 qui est proche de SF6 (identité de séquence > 80%) est un nonamère cyclique avec un corps central entouré de neuf domaines de liaison à l'ADN (DBD) (Büttner et al, 2012) qui lui procure la possibilité de se lier à plusieurs sites d'ADN. Par ailleurs, les travaux de Greive et al. (2016) par résonance plasmon de surface de la région de reconnaissance de *pac* de SF6 ont montré que le domaine de liaison à l'ADN présente une faible interaction avec *pacL* (Greive et al, 2016). La manipulation de la région *pacL* a néanmoins montré qu'une séquence minimale de la région *pacL* est nécessaire pour la reconnaissance et le clivage en *pac* dans notre système de plasmide rapporteur. Ceci suggère que les différents DBDs de *gp1* se lient à plusieurs sites de la région *pacL* pour un assemblage efficace du complexe nucléoprotéique. En effet, la région contenant 58 pb de nucléotides en amont du codon de démarrage du gène *I*, qui inclut plusieurs sites protégés par la digestion à la DNase y compris les deux boîtes répétées **a**, n'est pas suffisante pour conduire à un clivage en *pac* par la nucléase (Figure 3.11). Une séquence plus longue de 68 pb permet un faible clivage tandis que le segment en amont de 76 pb est dispensable. Ces données associées aux résultats de protection à la digestion de DNase (Chai

et al, 1995) apportent les éléments nécessaires pour suggérer que malgré le fait que la région *pacL* ne soit pas encapsidée à l'intérieur de la procapside, la séquence minimale de 76 pb doit avoir un rôle central dans l'assemblage d'un complexe *pac*-terminase fonctionnel. La reconnaissance de la région *pacL* par les domaines de liaison de la protéine gp1 est probablement favorisée par la présence de plusieurs sites de courbure de *pacL* comme décrit précédemment par des études d'inhibition avec la distamycine, une drogue qui se lie préférentiellement aux séquences riches d'A/T du sillon mineur de l'ADN en introduisant des distorsions locales dans la double hélice (Chai et al, 1996). L'utilisation de drogues telles que la spermine ou Hoechst 33258, qui se lient au sillon mineur de l'ADN mais sans induction de distorsions locales, ne montre pas d'inhibition de la liaison à la région *pacL*.

D'autre part, des mutations dans la région *pacR* ont conduit à des effets différents sur le clivage *pac* selon les séquences perturbées. Les mutations introduites dans les deux boîtes répétées **c1** et **c2** ne montrent aucun effet sur la reconnaissance et le clivage en *pac*. Ceci indique que la manipulation de ses deux boîtes n'est pas suffisante pour affecter l'activité de la terminase. Cependant l'altération du motif poly adénine AAAAA en AGAAG a montré le rôle critique de cette séquence pour le clivage de *pac*. Cette région a été choisie car il s'agit d'une séquence protégée dans le profil de protection à la digestion par la DNase (Figure 3.9) mais aussi rapportée comme favorisant les liaisons à l'ADN (Bolchoy et al, 1991). Les résultats de clivage *pac in vivo* montrent que ces mutations bloquent complètement le clivage *pac*. Toutefois, en analysant la séquence *pac* dans un groupe de phages proche de SPP1 incluant SF6, Lurz2 et Lurz3, nous pouvons voir que dans le motif AAAAA, seuls 3 adénines sont strictement conservés (Figure 3.7) et que la boîte **c1** pouvait être encore dégénérée. En plus il a été rapporté que la gp1 de SPP1 peut être substituée par celle de SF6 et donner des phages chimériques viables (Chai et al, 1994). Nous pouvons aussi remarquer la présence d'un autre motif poly adénine AAAAA à 25 pb avant le début de la région *pacR*, mais une

mutation dans ce motif ne montre aucun effet sur le clivage en *pac* (résultats non montrés). Celui-ci n'étant pas protégé dans le profil de protection de la digestion de la DNase, et donc pas susceptible d'interagir avec gp1, écarte la possibilité que l'interaction gp1-*pacR* soit due à un contexte structurale créé par la présence des motifs poly-A. Ces résultats indiquent que la reconnaissance de *pacR* est plutôt séquence spécifique.

En analysant l'alignement de séquences nucléotidiques de *pac* de SPP1 et des différents phages proches (Figure 3.7) et les résultats de mutagénèses, on peut supposer que le clivage de *pac* est assez permissif à des altérations dans la région *pac*. En effet, la région *pac* accepte une variation de la séquence nucléotidique au niveau de la région *pacC* et des délétions dans la région *pacL*, tout en maintenant une position de clivage précise et inchangée. De plus, la construction de phages contenant une délétion dans la majorité de la séquence *pacL* (correspondant au Mutant *pacL*-15, Figure 3.10) ou des mutations dans la séquence poly-adénine de *pacR* (celles du Mut 14, Figure 3.9) ont conduit à la production d'une population de phages avec de petites plages de lyse (Tavares et al, non publié). On suppose que, contrairement au système rapporteur pBT163 dans lequel aucun clivage n'a été observé, dans les conditions physiologiques le phage arrive à survivre avec une séquence *pac* contenant ce type de mutations, suite à une pression sélective.

4.3. Modèle proposé

L'ensemble des données obtenues combiné avec les résultats de la littérature nous permettent de suggérer un modèle pour le processus de reconnaissance et de clivage de la séquence *pac*. L'étude des séquences *pacL*, *pacC* et *pacR* nous a permis de mettre en évidence les éléments nécessaires pour déclencher la coupure sur *pac* et l'initiation de l'encapsidation du génome de SPP1. Les résultats obtenus suggèrent que la reconnaissance de l'ADN viral ferait intervenir essentiellement la région *pacL*. La présence de plusieurs sites de protection à la DNase et d'hypersensibilité, indiquent que cette région s'enroule fort

probablement autour de la protéine gp1, ce qui est compatible avec sa structure en nonamère présentant plusieurs sites de liaison à l'ADN à sa surface. Cette interaction protéine-ADN par le biais de la région *pacL* représenterait le premier point d'ancrage de la gp1 pour ensuite se lier à une autre région plus éloignée, *pacR*, créant ainsi un deuxième point d'ancrage pour gp1 (Figure 4.1). La liaison à *pacL* et *pacR* permettrait à gp1 de créer un contexte structural cadencé de sorte qu'il puisse présenter la région de *pacC* au domaine nucléase de gp2 pour la positionner et couper sur une région de 5 pb en amont de *pacR* à une position précise. L'assemblage de ce complexe nucléoprotéique permettrait aussi de séquestrer gp2 et éviter ainsi une activité endonucléase non-spécifique qui attaquerait l'ADN dans la cellule. Finalement, son organisation asymétrique assurerait l'encapsidation unidirectionnelle de l'ADN lorsque la terminase interagit avec le sommet portal de la capsid. L'assemblage de ce moteur de translocation de l'ADN conduirait à un « switch » de la terminase, de son activité nucléase en ATPase.

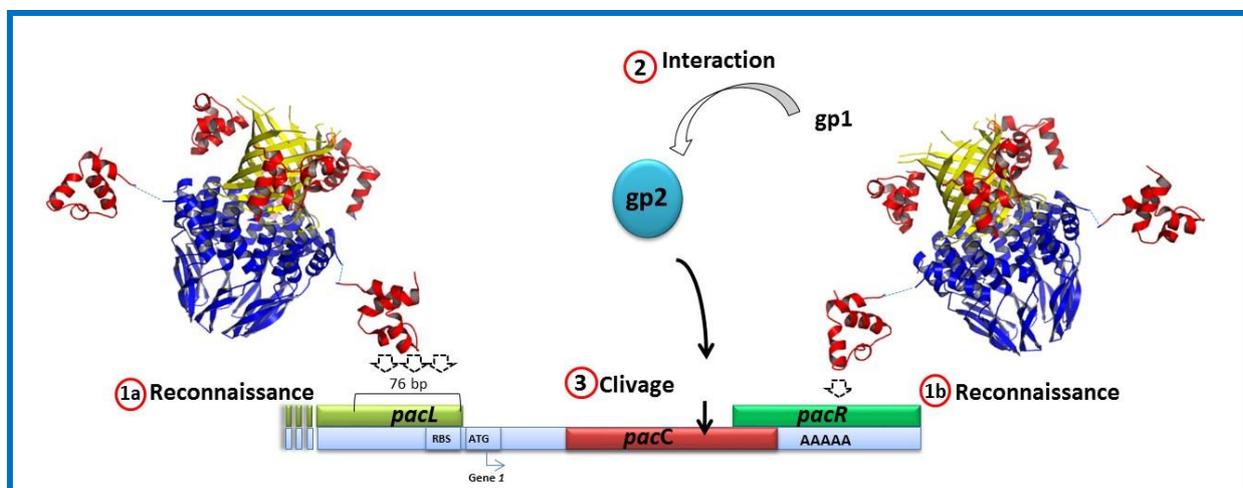


Figure 4.1: Modèle proposé pour la reconnaissance de *pac* par gp1. La reconnaissance des régions identifiées dans *pacL* et *pacR* par gp1 conduirait à une interaction avec gp2 dont le domaine nucléase se positionnerait pour insérer un clivage sur *pacC* à une position précise sans spécificité de séquence.

Ce type de mécanisme où le site de coupure est différent des régions de reconnaissance a déjà été décrit auparavant, autant chez les bactériophages qui encapsident l'ADN à partir de la séquence *pac* avec un système d'encapsidation par tête pleine, que chez les bactériophages

« *cos* » qui produisent des extrémités cohésives. Chez le bactériophage lambda par exemple, l'initiation de l'encapsidation du génome se fait au niveau d'un segment long de 200 pb appelé « *cos* ». Pour une initiation efficace de l'empaquetage. La protéine gpNu1, équivalente de la gp1 de SPP1, reconnaît 3 sites de 133 pb dans le segment *cosB*, puis coupe à une séquence spécifique appelée *cosN*, à gauche de *cosB* à une distance de 33 pb, dans laquelle l'introduction de deux « nicks » conduit à la production d'une coupure symétrique avec des bouts cohésifs de 12 pb (Catalano, 2005 ; Oh et al, 2016). Chez le bactériophage Mu l'encapsidation directionnelle est initiée par la reconnaissance d'une séquence *pac*. Celle-ci se trouve au début de l'extrémité gauche du génome entre les nucléotides 35-58 pb dans la région attL qui correspond aux régions de reconnaissance par les transposases. Il a été rapporté que le site de coupure a lieu à une distance de 100-250 pb en amont de la séquence *pac*, dans la séquence du génome de l'hôte *E. coli* ou Mu s'est inséré (Groenen et Van De Putte, 1985). Le phage P1 qui infecte une large gamme de bactéries Gram négatives, dont *E. coli*, démarre l'encapsidation de son génome en reconnaissant une séquence spécifique de 162 pb. L'enzyme pacase reconnaît 7 motifs répétés 5'-TGATCA/G-3' dans une région de 90 pb. L'interaction de l'enzyme avec ces différents motifs conduit à l'altération de la structure de l'ADN en formant une boucle qui présente un site de 13 pb à son centre accessible à la nucléase pour générer l'extrémité libre d'ADN (Sternberg et Coulby, 1987). Le phage de *S. typhimurium* P22, reconnaît une région asymétrique *pac* de 22 pb (Wu et al, 2002). Le clivage par la nucléase a lieu, cependant, à une distance de 120 pb sur plusieurs sites séparés de 20 pb (Casjens et al, 1992).

Les tests dans le système RTS *E. coli*, ont quant à eux mis en évidence une nouvelle donnée. Nous avons voulu développer un système de reconnaissance et de clivage en *pac in vitro* dans des conditions contrôlées. En effet, toutes les expériences pour reproduire le

processus *in vitro* en utilisant gp1 purifiée, gp2 et l'ADN portant la séquence *pac* ont échoué jusqu'à présent. Nous avons émis l'hypothèse que cela est dû à certaines exigences en matière de co-translation pour l'assemblage du complexe gp1-gp2-*pac*, notamment parce que le début de la séquence codante du gène *I* chevauche la plupart de la séquence *pac*. L'échec de la reproduction du clivage *pac in vitro* et dans la bactérie *E. coli* (section 3.2.3) nous a cependant permis de mettre la lumière sur un potentiel facteur de l'hôte de SPP1, *B. subtilis*, nécessaire dans ce processus. Ceci a déjà été décrit auparavant chez le phage lambda qui infecte *E. coli*. En effet un site *I2* entre la séquence de reconnaissance par gpNu1 (équivalente de gp1) et le site de clivage, a été rapporté comme étant une séquence de liaison à un facteur de l'hôte d'*E. coli*, IHF (Integration host factor). Des délétions dans cette séquence *I2* ont montré que l'efficacité de la coupure *in vitro* dans *cosN* ainsi que l'initiation de l'encapsidation étaient réduites. IHF est décrite comme une protéine qui introduit des courbures dans l'ADN et qui aiderait ainsi à mettre les autres sites de liaison en évidence pour la reconnaissance par la terminase (gpNu1). L'encapsidation de l'ADN *in vitro* a été observée comme réduite de 4 fois en absence de ce facteur bactérien (Catalano, 2005). Chez le bactériophage P1, cette caractéristique a aussi été observée. La reconnaissance et le clivage de *pac* se fait par les enzymes PacA et PacB, correspondant à la petite et grande sous-unité de la terminase, respectivement. Cependant un consensus de sites de liaisons à l'IHF a aussi été observé autour du site de clivage. Des expériences d'encapsidation *in vitro* chez P1 avec un extrait d'*E. coli* ont montré que ce facteur était nécessaire et que son absence réduisait drastiquement l'efficacité du clivage *pac in vitro* (Skorupski, 1994).

Chapitre 5 : Conclusions & Perspectives

5. Conclusions & Perspectives

La reconnaissance spécifique de l'ADN viral est essentielle pour l'empaquetage du génome. Dans ce travail, nous avons montré que :

- La spécificité du clivage dans la région *pac* de SPP1 est apportée par des régions bien définies dans *pacL* et *pacR* (délétions de *pacL* et mutagenèse de *pacR*).
- Le clivage sur *pac* par la nucléase de la terminase ne requiert pas de spécificité de séquence (Mutagenèse de la région *pacC*).
- Le site de clivage par la nucléase adopte toujours la même position malgré la dégénération de la région *pacC* (Séquençage haut débit des Mut 6* et Mut 8* de *pacC*).
- Des résidus du domaine de liaison à l'ADN (DBD) de gp1 ont un effet sur le clivage sur *pac* mais les résultats indiquent aussi des effets plus subtils sur le processus d'encapsidation.
- Les résultats de clivage *pac in vitro* associés au clivage *in vivo* dans *E. coli* suggèrent la présence d'un facteur bactérien apporté par l'hôte *B. subtilis*.

L'une des perspectives envisagée dans ce projet est de reproduire le système de clivage *pac in vitro*. La tentative de l'élaboration de cette expérience avec le système RTS pKD33 ayant échoué, nous pensons à la possibilité de rajouter un extrait de *B. subtilis* qui pourrait apporter le composant de l'hôte nécessaire pour ce clivage. Le fractionnement de l'extrait de *B. subtilis* pourrait, par la suite, permettre l'identification de ce composant. Le succès de ce système *in vitro*, nous permettra d'étudier le processus de reconnaissance et de clivage de l'ADN à des étapes distinctes. En effet, un plasmide avec seulement la protéine gp1 nous permettra d'étudier l'interaction gp1-*pac* ou l'interaction gp1-gp2 avec un plasmide contenant une nucléase qui ne coupe pas. De plus ce système permettrait également de vérifier

in vitro la séquence minimale nécessaire pour que la reconnaissance et le clivage de *pac* aient lieu.

L'identité du facteur de l'hôte qui est fort probablement nécessaire pour le clivage sur *pac* est une autre perspective importante. La présence d'une protéine associée à la chromatine équivalente à HU chez *E. coli*, a été observée chez *B. subtilis*. Cette protéine appelée Hbsu, se lie à l'ADN de manière non spécifique mais avec une préférence pour les zones de courbure de l'ADN. Hbsu interviendrait dans la réparation de l'ADN ainsi que dans des événements de recombinaison homologue (Fernandez et al, 1997). L'implication de cette protéine dans l'encapsidation du génome du phage, n'a pas encore été étudiée. Cette protéine représentera une première piste pour l'étude visant la recherche d'une interaction protéine (appartenant à l'hôte) et ADN du phage impliquée dans le processus de reconnaissance et clivage de *pac* chez SPP1. L'utilisation de la protéomique pourra aussi permettre de détecter les interactants de l'organisme hôte avec l'ADN ou les protéines du phage.

D'autre part, les mutants dans le domaine de liaison à l'ADN (DBD) de *gp1* construits dans ce travail ont tous montré un effet sur le clivage en *pac*. Les tests utilisés ici ne nous permettent pas de distinguer si ces résidus sont impliqués dans l'interaction *gp1-pac*, *gp1-gp2* ou dans le clivage sur *pacC*. C'est pourquoi, la prochaine étape est le développement d'une méthode d'analyse biochimique qui nous permette de distinguer entre ces différentes étapes. L'objectif étant de caractériser les résidus impliqués dans l'interaction directe avec l'ADN, nous avons pensé à l'utilisation de la méthode ChAP *in vivo* dans *B. subtilis*, qui est une variante de la méthode ChIP mais dont le complexe protéine-ADN formé est précipité par une purification par affinité d'une étiquette hexa-histidine. Nous avons commencé à élaborer cette expérience, mais les premières manipulations n'ont pas permis d'observer une interaction de la terminase avec une région précise de l'ADN. Des changements dans certaines étapes de purification ont apporté des améliorations mais l'expérience reste encore à perfectionner.

L'aboutissement de cette expérience nous permettrait d'utiliser les mutants gp1 décrits auparavant et de caractériser ceux qui sont impliqués dans la reconnaissance et dans l'interaction avec la région *pac*. Le succès de cette méthode permettra éventuellement de purifier le complexe ADN-gp1 pour des études structurales.

L'objectif majeur sera de caractériser le complexe gp1-gp2-ADN. En effet, on ne connaît pas la stœchiométrie des sous-unités gp1 et gp2, ni leur organisation structurale dans le complexe nucléoprotéique, ni comment de l'ADN est organisé au sein de ce complexe. Pour cela nous essayerons de produire et purifier le complexe ADN-terminase. Nous utiliserons les constructions que nous avons produites contenant des mutations dans la séquence *pacR*, *pacL* et dans gp1 qui sont susceptibles de bloquer les interactions gp1-*pac* et gp1-gp2 ou le clivage *pac* pour caractériser des états intermédiaires dans la voie d'assemblage de *pac*-gp1-gp2. Une fois ces complexes purifiés, et dans un objectif ultime, nous essayerons de faire une caractérisation structurale de ces complexes figés à différents états grâce à des études collaboratives de cryo-microscopie électronique et de cristallographie.

Bibliographie

- Abrescia et al. (2012) Structure Unifies the Viral Universe. *Annu Rev Biochem.* 81:795-822.
- Ackermann,W.H. (1998) Tailed bacteriophages: the order *Caudovirales*. *Advances in Virus Research.* vol. 51
- Ackermann,W.H. (2001) Frequency of morphological phage descriptions in the year 2000. *Arch Virol.* 146: 843-857
- Ackermann,W.H. (2007) 5500 Phages examined in the electron microscope. *Arch Virol.* 152: 227-243.
- Adelman,K., Salmon,B. and Baines,J.D. (2009) Herpes simplex virus DNA packaging sequences adopt novel structures that are specifically recognized by a component of the cleavage and packaging machinery. *PNAS.* 98(6): 3086–3091.doi10.1073ypnas.061555698.
- Alonso,J. C., Luder,G., Stiege,A.C., Chai,S., Weise,F. and Trautner,T.A. (1997) Analysis of the complete nucleotide sequence and functional organisation of *Bacillus subtilis* bacteriophage SPP1. *Gene.* 204:201-212.
- Alonso,J.C., Tavares,P., Lurz,R. and Trautner,T.A. (2006) *The Bacteriophages*, second edition.
- Badrinarayanan,A., Le, T.B.K., Laub, M.T. (2015) Bacterial Chromosome Organization and Segregation. *Annu. Rev. Cell Dev. Biol.* 31:171-99.
- Baines,D.J. (2011) Herpes simplex virus capsid assembly and DNA packaging: a present and future antiviral drug target. *Trends in Microbiology.* 19 (12): 606-6013.
- Becker,P.B. (2005) Nucleosome remodelers on track. *Nature Structural & Molecular Biology.* 12(9): 732- 733.
- Benini,S., Chechik,M., Lombardi,M.O., Polier,S., Leech,A., Shevtsov,M.B. and Alonso,J.C.(2013) The 1.58 Å resolution structure of the DNA-binding domain of

bacteriophage SF6 small terminase provides new hints on DNA binding. *Acta Cryst. F69*:376–381. doi:10.1107/S1744309113004399.

- Black,L.W. (1989) DNA packaging in dsDNA bacteriophages. *Annu. Rev. Microbiol.* 43: 267-292.
- Blattner et al. (1997) The Complete Genome Sequence of *Escherichia coli* K-12. *SCIENCE.* 277: 1453-1462.
- Bolshoy,A., McNamara,P., Harrington,R.E., and Trifonov,E.N. (1991) Curved DNA without A-A: experimental estimation of all 16 DNA wedge angles. *Proc Natl Acad Sci USA.* 88, 2312-2316.
- Boylan,R.J., Mendelson,N.H., Brooks,D. and Young,F.E. (1972). Regulation of the Bacterial Cell Wall: Analysis of a Mutant of *Bacillus subtilis* Defective in Biosynthesis of Teichoic Acid. *J of Bacteriol.* 111 (1): 281-290.
- Bravo,A., Alonso,J.C. and Trautner,T.A. (1990) Functional analysis of the *Bacillus subtilis* bacteriophage SPP1 pac site. *Nucleic Acids Res.* 18(10):2881-2886.
- Büttner,C.R., Chechik,M., Ortiz-Lombardía,M., Smits,C., Ebong,I.O., Chechik,V., Jeschke,G., Dykeman,E., Benini,S., Robinson,C.V., Alonso,J.C., Antson,A.A. (2012) Structural basis for DNA recognition and loading into a viral packaging motor. *Proc Natl Acad Sci USA.* 109: 811-816.
- Calendar,R. (1988) *The bacteriophages: Volume 2.* Plenum Press edition.
- Camacho,A.G., Gual,A., Lurz,R., Tavares,P., Alonso,J.C. (2003) *Bacillus subtilis* bacteriophage SPP1 DNA packaging motor requires terminase and portal proteins. *J Biol Chem.* 278: 23251–23259.
- Casjens,S. (2011) The DNA-packaging nanomotor of tailed bacteriophages. *Nature Reviews, Microbiology.* 9: 647.

- Casjens,S,R. and Gilcrease,E.B (2009) Determining DNA Packaging Strategy by Analysis of the Termini of the Chromosomes in Tailed-Bacteriophage Virions. *Methods Mol Biol.* 502: 91–111. doi:10.1007/978-1-60327-565-1_7.
- Casjens,S., Sampson,L., Randall,S., Eppler,K., Wu,H., Petri,J.B., Schmieger,H. (1992) Molecular genetic analysis of bacteriophage P22 gene 3 product, a protein involved in the initiation of headful DNA packaging. *J Mol Biol.* 227:1086-1099.
- Catalano,C.E., Cue,D., Feiss,M. (1995)Virus DNA packaging: the strategy used by phage lambda. *Molecular microbiology.* 16(6):1075.
- Catalano,C.E. (2005) viral genome packaging machines: Genetics, Structure, and Mechanisms. *Molecular Biology Intelligence Unit. Landes Bioscience.*
- Chai,S., Bravo,A., Luder,G., Nedlin,A., Trautner,T.A., Alonso,J.C. (1992) Molecular Analysis of the Bacillus subtilis Bacteriophage SPP1 region encompassing genes 1 to 6. The products of gene 1 and gene 2 are required for pac cleavage. *J. Mol. Biol.* 224: 87-102.
- Chai,S., Kruft,V., Alonso,J.C. (1994) Analysis of the Bacillus subtilis bacteriophages SPP1 and SF6 gene 1 product: a protein involved in the initiation of headful packaging. *Virology.* 202: 930–939.
- Chai,S., Lurz,R. and Alonso,J.C. (1995) The small subunit of the terminase enzyme of Bacillus subtilis bacteriophage SPP1 forms a specialized nucleoprotein complex with the packaging initiation region. *J Mol Biol.* 252: 386-398.
- Chai,S. and Alonso,J.C. (1996) Distamycin-induced inhibition of formation of a nucleoprotein complex between the terminase small subunit G1P and the non-encapsidated end (*pacL* site) of Bacillus subtilis bacteriophage SPP1. *Nucleic Acids Research.* 24 (2):282–288.
- Chai,S., Szepan,U., Alonso,J.C. (1997) Bacillus subtilis bacteriophage SPP1 terminase has a dual activity: it is required for the packaging initiation and represses its own synthesis. *Gene.* 197:251-256.

- Chang,J., Weigele,P., King,J., Chiu,W. and Jiang,W. (2007) Cryo-EM Asymmetric Reconstruction of Bacteriophage P22 Reveals Organization of its DNA Packaging and Infecting Machinery. *Structure*. 14:1073–1082.
- Clark,J.M. (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Research*. 16 (20).
- Cornilleau,C., Atmane,N., Jacquet,E., Callum,C., Alonso,J.C., Tavares,P. and Oliveira,L. (2013)The nuclease domain of the SPP1 packaging motor coordinates DNA cleavage and encapsidation. *Nucleic Acids Res*. 41:340-354.
- Cutter,A.R., Hayes,J.J. (2015) A brief review of nucleosome structure. *FEBS Letters*. 589:2914-2922.
- Deichelbohrer,I., Messer,W., Trautner,T.A. (1982) Genome of *Bacillus subtilis* bacteriophage SPP1: structure and nucleotide sequence of *pac*, the origin of DNA packaging. *J Virol*. 42:83-90.
- Droge,A. and Tavares P. (2002) In vitro packaging of DNA of the *Bacillus subtilis* bacteriophage SPP1. *J Mol Biol*. 296:1238-1243.
- Dube,P., Tavares,P., Lurz,R. and Heel,M.V. (1993) The portal protein of bacteriophage SPP1: a DNA pump with 13-fold symmetry. *The EMBO Journal*. 1 (2) (4):1303-1309.
- Elias,I. & Lagergren,J. (2007) Fast computation of distance estimators. *BMC Bioinformatics*. 8 (89).
- Feiss,M. and Catalano CE.(2005) Bacteriophage Lambda Terminase and the Mechanism of Viral DNA Packaging. In: Catalano CE, editor. *Viral Genome Packaging Machines: Genetics, Structure and Mechanism*. Landes Bioscience/Eurekah.com; Georgetown, TX. 5–39.

- Fernandez,S., Rojo,F. and Alonso,J.C. (1997) The *Bacillus subtilis* chromatin-associated protein Hbsu is involved in DNA repair and recombination. *Molecular Microbiology*. 23(6): 1169-1179.
- Gascuel,O. (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol*. 14(7), 685-695.
- Greive,S.G., Fung,H.K, Chechik,M., Jenkins,H.T., Weitzel,S.E., Aguiar,P.M., Brentnall,A.S., Glousieau,M., Gladyshev,G.V., Potts,J.R., Antson,A.A. (2016) DNA recognition for virus assembly through multiple sequence-independent interactions with a helix-turn-helix motif. *Nucleic Acids Res*. 44:776-789.
- Groenen,M.A.M, and Van De Putte,P. (1985) Mapping of a Site for Packaging of Bacteriophage Mu DNA. *Virology*. 144:520-522.
- Gual,A., Camacho,G.C and Alonso,J.C. (2000) Functional analysis of the terminase large subunit, G2P, of *Bacillus subtilis* Bacteriophage SPP1. *J Biol chem*. 275:35311
- Guasch,A., Pous,J., Ibarra,B., Gomis-Ruth,F.X., Valpuesta,J.M., Sousa,N., Carrascosa,J.L., and Miquel,C. (2002) Detailed Architecture of a DNA Translocating Machine: The High-resolution Structure of the Bacteriophage ϕ 29 Connector Particle. *J Mol Biol*. 315: 663-676.
- Guo,F. et al. (2014) Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. *PNAS*.4606-4614.
- Hanahan, D. (1985) Technique for transformation of *E. coli*. In: Gloever D.M. (Ed). *DNA cloning, a practical approach*, Washington DC. 109-139.
- Hirano,T. (2014) Condensins and the evolution of torsion-mediated genome organization. *Trends in Cell Biology*. 24(12).
- Jakutyte,L., Baptista,C., São-José,C., Daugelavičius,R., Carballido-López,R and Tavares,P. (2011) Bacteriophage Infection in Rod-Shaped Gram-Positive Bacteria: Evidence for a

Preferential Polar Route for PhageSPP1 Entry in *Bacillus subtilis*. *J Bacteriol.* 193(18): 4893-4903.

- Kleckner,N., Fisher,J.K., Stouf,M., White,M.A., Bates,D., Wit,G. (2014) The bacterial nucleoid: nature, dynamics and sister segregation. *Current Opinion in Microbiology.* 22:127-137.
- Kunst,F. et al. (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature.* 390: 249-256.
- Lander,J.C., Evilevitch,A., Jeembaeva,M., Potter,C.S., Carragher,B. and Johnson,J.E. (2008) Bacteriophage Lambda Stabilization by Auxiliary Protein gpD: Timing, Location, and Mechanism of Attachment Determined by Cryo-EM. *Structure.* 16: 1399-1406.
- Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., William,Mc., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R., Thompson,J.D., Gibson,T.J. and Higgins,D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics Application notes.* 23 (21): 2947–2948.
- Lebedev,A.A., Krause,M.H., Isidro,A.L., Vagin,A., Orlova,E.V., Turner,J., Dodson,E.J., Tavares,P., Antson,A.A. (2007) Structural framework for DNA translocation via the viral portal protein. *EMBO Journal.* 26: 1984-1994.
- Lin,H., Simon,M.N. and Black.L.W. (1997) Purification and characterization of the small subunit of phage T4 terminase, gp16, required for DNA packaging. *J Biol Chem.* 272(6):3495-501.
- Lo Piano,A., Martínez-Jiménez,M.I., Zecchi,L. and Ayora,A. (2011) Recombination-dependent concatemeric viral DNA replication. *Virus Research.* 160: 1-14.
- Mardis,E.R. (2013) Next-Generation Sequencing Platforms. *Annu Rev Anal Chem.* 6:287–303.

- McNulty,R., Lokareddy,R.K., Roy,A., Yang,Y., Lander,G.C., Heck,A.J., Johnson,J.E. and Cingolani,G. (2011) Architecture of the Complex Formed by Large and Small Terminase Subunits from Bacteriophage P22. *J Mol Biol.* 427(20):3285-99.
- Mettenleiter,T.C., Klupp,B.G., Granzow,H. (2009) Herpesvirus assembly: An update. *Virus Research.* 143: 222–234.
- Moore,S.D. and Prevelige,P.E.Jr. (2002) DNA Packaging: A New Class of Molecular Motors. *Current Biology.* 12: 96-98.
- Morelli,G., Fisseau,C, Behrens, B, Trautner,T.A., Luh,J., Ratcliff,S.W., Allison,D.P., and Ganesan,A.T. (1979) The genome of *B. subtilis* phage SPP1: The topology of DNA molecules. *Mol Gen Genet.* 10; 168(2):153-61.
- Mosig,G. (1998) Recombination and recombination-dependent DNA replication in bacteriophage T4. *Annu Rev Genet.* 32: 379-413.
- Murphy,F.A., Fauquet,C.M., Bishop,D.H.L., Ghabrial,S.A., Jarvis,A.W., Martelli,G.P., Mayo,M.A., and Summers,M.D. (1995). “Virus Taxonomy.” Sixth Report of the International Committee on Taxonomy of Viruses. *Arch. Virol. (Suppl. 10)*, Springer, Vienna.
- Oh,C.S., Sippy,J., Charbonneau,B., Hutchinson,J.C., Mejia-Romero,O.E., Barton,M., Patel,P., Sippy,R., Feiss,M. (2016) DNA Topology and the Initiation of Virus DNA Packaging. *PloS.One.*
- Okubo,S. and Yangida,T. (1968). Isolation of a Suppressor Mutant in *Bacillus subtilis*. *J of bacterial.* 95(3):1187-1188.
- Oliveira,L., Henriques A.O and Tavares P. (2006) Modulation of the viral ATPase activity by the portal protein correlates with DNA packaging efficiency. *J. Biol. Chem.* 281:21914-21923.
- Oliveira,L., Tavares,P. and Alonso, JC. (2013) Headful DNA packaging: bacteriophage SPP1 as a model system. *Virus Research.* 173: 247-259.

- Orlova,E.V., Dube,P., Zemlin,F., Lurz,R., Trautner,T.A., Tavares,P. and Beckmann,E. (1999) Structure of the 13-fold symmetric portal protein of bacteriophage SPP1. *Nature structural biology*. 6 (9): 842-846.
- Prasad,B.V.V. and Schmid,M.F. (2012) Principles of Virus Structural Organization. *Adv Exp Med Biol*. 726: 17-47.
- Ratcliff,S.W., Luh,J., Ganesan,A.T., Behrens,B., Thompson,R., Montenegro,M.A., Morelli,G. and Trautner,T.A. (1979) The genome of Bacillus subtilis phage SPP1: The arrangement of restriction endonuclease generated fragments. *Molecular and General Genetics MGG*. 168 (2): 165-172.
- Riva,S., Polsinelli,M. and Falaschi,A.(1968) A new phage of Bacillus subtilis with infectious DNA having separable strands. *J Mol Biol*.35:347-356.
- Robert,X. and Gouet,P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res*. 42(W1): W320-W324. doi: 10.1093/nar/gku316 ([freely accessible online](#))
- Roos,W.H. et al. (2009) Scaffold expulsion and genome packaging trigger stabilization of herpes simplex virus capsids. *PNAS*.106 (24): 9673-9678.
- Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G., Mesirov,J.P. (2011) Integrative Genomics Viewer. *Nat Biotech*. 29: 24 –26.
- Sambrook,J., Maniatis,T., Fritsch,E.F. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sanger,F., Nicklen,S. and Coulson,A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*. 74(12): 5463-5467.
- Santos,M.A., Herminia And Archer,L.J. (1984) Homology between Phages SPP1, 41c, 22a, p15 and SF6 of Bacillus subtilis. *J gen Virol*. 65: 2067-207

- Skorupski,K., Sauer,B., Sternberg,N. (1994) Faithful cleavage of the P1 packaging site (pac) requires two phage proteins, PacA and PacB, and two Escherichia coli proteins, IHF and HU. J Mol Biol. 243: 268-282.
- Smits,C., Chechik,M., Kovalevskiy,O.V, Shevtsov,M.B., Foster,A.W., Alonso,J.A. and Antson,A.A. (2009) Structural basis for the nuclease activity of a bacteriophage large terminase. EMBO Rep. 10: 592–598.
- Steensma,H.Y. and Blok,J. (1979) Effect of calcium ions on the infection of *Bacillus subtilis* by bacteriophage SF6. J Gen Viro. 42: 305-314.
- Sternberg,N. and Coulby,J. (1987) Recognition and Cleavage of the Bacteriophage P1 Packaging Site (pac) II. Functional limits of pac and Location of pac Cleavage Termini. J Mol Biol. 194: 469-479.
- Sternberg,N. and Coulby,J. (1990) Cleavage of the bacteriophage P1 packaging site (pac) is regulated by adenine methylation. Proc Nat Acad Sci USA. 87: 8070-8074.
- Tavares,P., Santos,M.A., Lurz,R., Morelli,G., de Lencastre,H. and Trautner.T.A (1992) Identification of a gene in *Bacillus subtilis* bacteriophage SPP1 determining the amount of packaged DNA. J Mol Biol.225:81-92.
- Tavares,P., Lurz,R., Stiege,A., Ruckert,B. and Trautner,T.A. (1996) Sequential Headful Packaging and Fate of the Cleaved DNA Ends in Bacteriophage SPP1. J Mol Biol. 264: 954-967.
- Takashi,S. (1975) The starting point and direction of rolling circle replicative intermediates of coliphage lambda DNA. Mol Gen Genet. 142:137-153.
- Thanbichler,M., Wang,S.C., Shapiro,L. (2005) The Bacterial Nucleoid: A Highly Organized and Dynamic Structure. Journal of Cellular Biochemistry. 96:506-521.
- Trautner.T.A., Spatz,H.C.(1973) Transfection on *B.subtilis*. Current Topics in Microbiology and Immunology. 62: 61-88.

- Van Dijk,E.L., Auger,H., Jaszczyszyn,Y., Thermes,C. (2014) Ten years of next-generation sequencing technology. Trends in Genetics. 30: 418-426.
- Wang,L., Huang,C., Yang,M.Q. and Yang,J.Q. (2010) BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. BMC Systems Biology. 4 (1):S3. DOI: 10.1186/1752-0509-4-S1-S3.
- Wieczorek,D.J. and Feiss,M. (2001) Defining cosQ, the Site Required for Termination of Bacteriophage λ DNA Packaging. Genetics. 158: 495–506.
- White,H.E., Sherman,M.B., Brasilès,S., Jacquet,E., Seavers,P., Tavares,P., Orlova,E.V. (2012) Capsid structure and its stability at the late stages of bacteriophage SPP1 assembly. J Virol. 86(12):6768-6877.
- Wu,H., Sampson,L., Parr,R., Casjens,S. (2002) The DNA site utilized by bacteriophage P22 for initiation of DNA packaging. Mol Microbiol. 45:1631-1646.
- Yarmolinsky,M. B., and Sternberg,N.L. (1988) Bacteriophage P1. In R. Calendar (ed.) The Bacteriophages. Plenum Press, NewYork. 291-438.
- Yang,Q., Hanagan,A., Catalano,C.E. (1997) Assembly of a nucleoprotein complex required for DNA packaging by bacteriophage lambda Biochemistry. 36. 2744–2752.
- Yang,Q., Maluf,N.K., Catalano,C.E. (2008) Packaging of a unit-length viral genome: the role of nucleotides and the gpD decoration protein in stable nucleocapsid assembly in bacteriophage lambda. J Mol Biol. 383(5):1037-1048.
- Yasbin, R.E., Mainon,V.C. and Young,F.E. (1976) Bacteriophage Resistance in *Bacillus subtilis* 168, W23, and Interstrain Transformants. J of Bacteriol. 125 (3):1120-1126.
- Zinder,N.D. and Lederberg,J. (1952) Genetic exchange in Salmonella. J. Bacteriol.64:679-699.
- Zsigray,R.M., Miss,A.L. and Landman,O.E. (1973) Penetration of a bacteriophage into *Bacillus subtilis*: blockage of infection by deoxyribonuclease. J Virol. 11: 69-77.

Titre : Mécanisme moléculaire de reconnaissance et clivage du génome chez le bactériophage SPP1, un virus à ADN double-brin

Mots clés : Bactériophage, encapsidation du génome, reconnaissance d'ADN, terminase, ADN double-brin

Résumé: La reconnaissance spécifique du génome viral et son encapsidation est une étape cruciale pour l'assemblage de particules virales. Chez SPP1, comme chez d'autres bactériophages à queue, le moteur moléculaire qui encapside le génome viral est composé de la terminase, une enzyme hétéro-oligomérique qui possède une activité ATPasique et nucléasique, et de la protéine portale, un oligomère cyclique par lequel l'ADN viral est transloqué. Dans un grand nombre de ces virus, l'encapsidation de l'ADN est initiée par la reconnaissance et le clivage d'une séquence spécifique nommée « *pac* ». C'est un événement qui se produit une seule fois au début d'une série de cycles d'encapsidation processive à partir d'un concatémère issu de la réplication du génome du phage. La région *pac* de SPP1 contient deux séquences (*pacL* et *pacR*) où TerS (gp1) se lie entourant la région (*pacC*) où TerL (gp2) coupe l'ADN de SPP1.

Ici, nous montrons qu'une région de la séquence *pacL* et qu'un motif polyadénine de *pacR* agissent ensemble pour promouvoir le clivage en *pacC*. La dégénération de la région *pacC* n'a pas montré d'effet sur que le clivage endonucléolytique qui a lieu à une position bien définie de *pacC* avec une précision de ~6 pb. Des études avec des phages proches de SPP1 ont montré une conservation dans la position du clivage, malgré des variations dans *pacC*, *pacR* ou dans la distance entre *pacL* et *pacC*. Les données sont compatibles avec un modèle dans lequel TerS interagit spécifiquement avec la région *pacL*, sur laquelle le multimère cyclique TerS doit s'enrouler, et le motif polyadénine de la région *pacR*. Le complexe nucléoprotéique résultant va créer un contexte structural qui permet de recruter et positionner le domaine nucléase de TerL pour une coupure très précise sur *pacC* sans spécificité de séquence.

Title : Molecular mechanisms of recognition and cleavage of the genome of bacteriophage SPP1, a double-stranded DNA virus

Keywords : Bacteriophage, genome packaging, DNA recognition, terminase, double-stranded DNA

Abstract: The specific recognition of the viral genome and its packaging is a critical step in viral particle assembly. In SPP1, as in many tailed bacteriophages, the macromolecular motor that encapsidates viral DNA is composed of terminase, a hetero-oligomeric enzyme possessing ATPase and nuclease activities, and of portal protein, a cyclic oligomer through which DNA is translocated. In a large number of these viruses, DNA packaging is initiated by recognition and cleavage of a specific sequence *pac*. This event occurs once at the beginning of a series of processive encapsidation events along a substrate concatemer of replicated phage genomes. The SPP1 *pac* region has two sequences where TerS (gp1) binds (*pacL* and *pacR*) flanking the segment where TerL (gp2) cleaves the

SPP1 DNA (*pacC*). Here we show that a sequence segment of *pacL* and a poly-adenine motif in *pacR* act together to promote cleavage at *pacC*. Extensive degeneration of *pacC* sequence has no detectable effect in *pac* cleavage. The endonucleolytic cut occurs at a defined position with a precision of ~6 bp. Studies with SPP1-related phages show conservation of the cut position, irrespectively of sequence variation in *pacC*, in *pacR* or changes in *pacL-pacC* distance. The data is compatible with a model in which TerS interacts specifically with a region of *pacL* that probably wraps around the TerS cyclical multimer, and a poly-A tract in *pacR*. The resulting nucleoprotein complex architecture positions TerL for accurate cleavage at *pacC* without specific sequence requirement.

