

MARIE-EVE LALIBERTÉ GAGNÉ

**IMPORTANCE DE L'EXTRÉMITÉ N-TERMINALE  
DE LA NUCLÉOCAPSIDE DANS L'ASSEMBLAGE  
DU VIRUS DE LA MOSAÏQUE DE LA PAPAYE**

Mémoire présenté  
à la Faculté des études supérieures de l'Université Laval  
dans le cadre du programme de maîtrise en biologie végétale  
pour l'obtention du grade de Maître ès sciences (M. Sc.)

DÉPARTEMENT DE PHYTOLOGIE  
FACULTÉ DES SCIENCES DE L'AGRICULTURE ET DE L'ALIMENTATION  
UNIVERSITÉ LAVAL  
QUÉBEC

2007

## Résumé

Le Virus de la Mosaïque de la Papaye est un virus végétal filamenteux flexible. Son unique protéine structurale, la protéine de capsid, est composée de 215 acides aminés. Le but de ce présent projet de recherche est de déterminer l'importance de l'extrémité N-terminale de la CP dans l'assemblage du PapMV. Des études antérieures à ce projet ont démontré que la délétion des cinq premiers acides aminés de la CP (CP<sub>6-215</sub>) n'affectait pas la capacité de la CP à se multimériser, alors qu'une délétion des vingt-six premiers acides aminés (CP<sub>27-215</sub>) conduisait à une incapacité de la CP à se multimériser. L'élaboration de mutants de délétion à l'extrémité N-terminale ainsi que leur expression dans la bactérie *E. coli* a permis de mettre en évidence l'importance de la phénylalanine retrouvée à la position 13 de la CP. Des mutations ponctuelles pour cet acide aminé ont permis de révéler le rôle que jouait l'hydrophobicité à l'extrémité N-terminale dans la multimérisation de la protéine de capsid.

## Avant-Propos

Les études au deuxième cycle ont représenté un beau défi pour moi, tant sur le plan intellectuel que personnel. C'est au cours des deux dernières années que je me suis épanouie et que j'ai découvert ma vraie passion : la recherche. Mon parcours n'a pas été sans accros, mais n'est-ce pas là la beauté de la chose que de réussir après avoir surmonté des obstacles? Bien sûr, plusieurs personnes ont contribué à cette réussite et je profite de cette occasion pour les remercier.

Tout d'abord, un merci tout spécial à mes parents, Diane et Jacques, qui m'ont toujours encouragée et supportée dans tout ce que j'entreprends. Les valeurs qu'ils m'ont inculquées ont été d'un grand secours tout au long de ma vie. Merci à ma grand-mère Gertrude Laliberté pour ses mille et une prières avant chaque événement important et pour ses toutes aussi nombreuses histoires drôles qui m'ont fait rire dans les pires moments. Merci à ma grande sœur Valérie sur qui j'ai toujours pu compter, à n'importe quelle heure du jour ou de la nuit et qui est toujours partante pour un petit souper et une bonne bouteille de vin. Merci à Philippe, mon rayon de soleil qui illumine tous mes jours nuageux. Merci à vous ma famille qui me supportez et me permettez de toujours avancer.

Merci à mes amis Julie, Marc-André, Sabrina, Frédérick, Claude et Catherine avec qui je peux me détendre et parler de tout et de rien. Un gros merci à tous mes collègues du centre de recherche en infectiologie et plus spécialement à mon équipe de recherche, car ils ont dû apprendre à me supporter dans les meilleurs moments comme dans les pires! Un merci spécial à Christine, Christian, Marilène et Jérôme pour toutes les discussions scientifiques ou non que nous avons partagées.

Je me dois de remercier Marie-Hélène Tremblay pour son incroyable travail sur le PapMV qui a conduit à l'élaboration de mon projet de recherche. Merci à Nathalie pour ses nombreux conseils et son expertise qui m'ont permis de sauver bien du temps et des soucis. Merci à Stéphane Gagné pour m'avoir accueillie dans son laboratoire au pavillon Marchand pour les nombreux tamisages moléculaires que j'avais à faire. Un immense merci à Katia

Lecours pour tout le temps qu'elle a consacré à m'aider et à m'expliquer ses mille et une découvertes. Merci aussi à Jean-Baptiste pour son aide.

Finalement, je tiens à remercier Denis Leclerc, mon directeur de recherche, qui m'a accueillie les bras ouverts au sein de son équipe et qui m'a donné l'opportunité de découvrir le monde fascinant du PapMV. Merci de m'avoir épaulée et encouragée à poursuivre tout au long de cette aventure.

Merci à vous tous encore une fois, car sans vous je n'aurais pu y arriver.

*Il n'y a point de génie sans un grain de folie. Aristote*

## Table des matières

Résumé.....	ii
Avant-Propos .....	iii
Table des matières .....	vi
Liste des abréviations.....	ix
Liste des tableaux.....	xi
Liste des figures .....	xii
1.0 Introduction.....	1
1.1 Les virus végétaux .....	3
1.1.1 Structures de capsides étudiées par la diffraction des rayons X .....	4
1.1.2 Structures de capsides étudiées par la résonance magnétique nucléaire .....	5
1.1.3 Les virus hélicoïdaux .....	5
1.1.4 L'assemblage viral.....	6
1.2 La famille des <i>Flexiviridea</i> .....	7
1.3 Les <i>Potyvirus</i> .....	8
1.4 La protéine de capside des <i>Potyvirus</i> .....	9
1.5 L'assemblage des <i>Potyvirus</i> .....	10
1.6 Les <i>Potexvirus</i> .....	12
1.6.1 Le Virus X de la Pomme de terre .....	12
1.6.2 Le Virus de la Mosaïque de la Papaye .....	12
1.7 La protéine de capside des <i>Potexvirus</i> .....	14
1.7.1 La protéine de capside de PVX.....	14
1.7.2 La protéine de capside du PapMV.....	15
1.8 L'assemblage des <i>Potexvirus</i> .....	21
1.8.1 L'assemblage du PVX.....	21
1.8.2 Assemblage du PapMV .....	23
1.9 Hypothèses et objectifs .....	27
2.0 Matériel et méthodes.....	28
2.1 Clonage .....	28
2.1.1 Réactions de polymérase en chaînes (PCR).....	28
2.1.2 Transformation de bactéries compétentes par choc thermique.....	31
2.1.3 Criblage des mutants.....	31
2.1.4 Les mutants de délétions.....	32
2.1.5 Les mutants ponctuels.....	33
2.1.6 Les mutants CP <sub>14-180</sub> et CP <sub>27-180</sub> .....	35
2.2 Production des protéines recombinantes dans <i>E.coli</i> .....	36
2.2.1 Transformation de bactéries compétentes par choc thermique.....	36
2.2.2 Induction de la production de protéines.....	36
2.2.3 Lyse des cellules bactériennes .....	37
2.3 Purification des protéines d'intérêts .....	37
2.4 Électrophorèse des protéines et transfert de type Western .....	38
2.5 Ultracentrifugation.....	38
2.6 Microscopie électronique.....	39
2.7 Spectroscopie en dichroïsme circulaire .....	39
2.8 EMSA .....	40

2.8.1	Préparation de la sonde marquée au <sup>32</sup> P	40
2.8.2	Réaction EMSA	41
2.9	Chromatographie par tamisage moléculaire sur FPLC	41
2.10	Traitement à l'acide acétique	42
2.11	Purification de PapMV à partir de plants infectés	42
2.12	Prédictions de structures secondaires	43
2.13	Échantillons RMN	44
2.12.1	Préparation des échantillons pour RMN	44
2.12.2	Enregistrement des spectres RMN	45
2.12.3	Analyse des spectres RMN	45
3.0	Élaboration d'un protocole standard pour la production des PapMV CP	46
3.1	Introduction	46
3.2	Résultats	47
3.2.1	La production de protéines	47
3.2.2	L'extraction des protéines (lyse bactérienne)	47
3.2.3	La purification des protéines	51
3.2.4	La dialyse des protéines	52
3.3	Discussion	53
<i>La production de protéines</i>		53
<i>La lyse bactérienne</i>		54
<i>La purification des protéines</i>		55
<i>La dialyse des protéines</i>		55
4.0	Implication de l'hélice Q18-S23 prédite dans l'assemblage du PapMV	57
4.1	Introduction	57
4.2	Résultats	61
4.2.1	Prédictions de structures secondaires des mutants de délétion CP <sub>15-215</sub> et CP <sub>21-215</sub>	61
4.2.2	Production et purification des mutants de délétion	62
4.2.3	Mutants ponctuels	63
4.2.4	Prédictions de structure secondaire pour E19K et E19P	63
4.2.5	Production et purification des mutants de délétion	64
4.2.6	Tamisage moléculaire de E19K et E19P	65
4.2.7	Microscopie électronique de E19K et E19P	65
4.3	Discussion	67
<i>Les mutants ponctuels</i>		67
5.0	Étude de l'implication de l'extrémité N-terminale dans la formation de NLPs	69
5.1	Introduction	69
5.2	Résultats	70
5.2.1	Mutant de délétion CP <sub>14-215</sub>	70
5.2.2	Mutant de délétion CP <sub>13-215</sub>	74
5.2.3	Mutant tronqué CP <sub>14-180</sub>	79
5.3	Discussion	83
<i>Les mutants de délétion CP<sub>13-215</sub> et CP<sub>14-215</sub></i>		83
<i>Les doubles mutants de délétion CP<sub>14-180</sub> et CP<sub>27-180</sub></i>		84
<i>Conclusion et perspectives</i>		85
6.0	Importance de l'hydrophobicité à l'extrémité N-terminale	87
6.1	Introduction	87

6.2 Résultats.....	88
6.2.1 Expression et purification des mutants F13 .....	88
6.2.2 Tamisage moléculaire .....	89
6.2.3 Microscopie électronique.....	91
6.2.4 Liaison à l'ARN.....	92
6.2.5 Microscopie électronique.....	94
6.2.6 Spectroscopie en dichroïsme circulaire .....	95
6.2.7 Rapport des absorbances à 280 et 260 nm pour les NLPs .....	95
6.2.8 Rapport des absorbances à 280 et 260 nm pour les disques .....	97
6.3 Discussion.....	98
7.0 Conclusion .....	102
Bibliographie .....	106
Annexe 1. Alignement des CPs des <i>Potexvirus</i> .....	109
Annexe 2. Milieux et tampons utilisés.....	112
Annexe 3. Séquences des clones.....	115
Annexe 4. Carte Vecteur d'expression pET-3d .....	131
Annexe 5. Prédications de structure secondaire .....	132



## Liste des abréviations

3D	Tridimensionnel
AIMV	" <i>Alfalfa Mosaic Virus</i> "
ADN *	Acide désoxyribonucléique
AltMV	" <i>Alternanthera mosaic virus</i> "
ARN *	Acide ribonucléique
BaMV	" <i>Bamboo mosaic virus</i> "
CD *	" <i>Circular dichroïsme</i> "
CP *	Protéine de capsid
CPMV	" <i>Cowpea Mosaic Virus</i> "
CsCMV	" <i>Cassava common mosaic virus</i> "
CVX	" <i>Cactus virus X</i> "
CymMV	" <i>Cymbidium mosaic virus</i> "
DO *	Densité optique
<i>E. coli</i> *	<i>Escherichia coli</i>
EMSA *	" <i>Electrophoretic mobility shift assay</i> "
FoMV	" <i>Foxtail mosaic virus</i> "
HCV	" <i>Virus de l'hépatite C</i> "
HMWF	" <i>High molecular weight forms</i> "
HVX	" <i>Hosta virus X</i> "
IPTG *	isopropyl-beta-D-thiogalactopyranoside
LMWF	" <i>Low molecular weight forms</i> "
LVX	" <i>Lily virus X</i> "
mAmp	milliampère
MRW	" <i>Mean residue weight</i> "
NLPs *	" <i>Nucleocapsid-like particules</i> "
NMV	" <i>Narcissus Mosaic Virus</i> "
ORFs *	" <i>Open reading Frames</i> "
PAMV	" <i>Potato aucuba mosaic virus</i> "
PapMV *	Virus de la Mosaïque de la Papaye
PapMV CP *	Protéine de capsid du virus du PapMV
PapMV CP WT *	Protéine de capsid du virus sauvage
PCR *	Réaction de polymérisation en chaîne
PepMV	" <i>Pepino mosaic virus</i> "
PhMV	" <i>Physalis mottle tymovirus</i> "
PIAMV	" <i>Plantago asiatica mosaic virus</i> "
poly(A)	Queue de polyadénines
PVA	" <i>Potato Virus A</i> "
PVBV	" <i>Pepper Vein Banding Virus</i> "
PVX	" <i>Potato Virus X</i> "

RMN *	Résonance magnétique nucléaire
RPM *	<i>“Revolution per minute”</i>
ScaVX	<i>“Scallion virus X”</i>
SMYEV	<i>“Strawberry mild yellow edge virus”</i>
TEV	<i>“Tobacco Etch Virus”</i>
TGB	<i>“Triple gene block”</i>
TMV	<i>“Tobacco mosaic Virus”</i>
TVX	<i>“Tulip virus X”</i>
UV *	Ultra Violet
VPg	<i>“viral protein genome-linked”</i>
WC1MV	<i>“White Clover Mosaic Virus”</i>
WT *	<i>“Wild type”</i>

\* Afin d’alléger le texte, ces abréviations seront utilisées tout au long du document.

## Liste des tableaux

<b>Tableau 1. Sommaire des caractéristiques pour les différents genres et virus faisant partie de la famille des Flexiviridea .....</b>	<b>8</b>
<b>Tableau 2. Tableau résumé des caractéristiques des protéines contenant des mutations à des acides aminés spécifiques.....</b>	<b>18</b>
<b>Tableau 3. Amorces utilisées pour les délétions du gène PapMV CP. ....</b>	<b>33</b>
<b>Tableau 4. Amorces utilisées pour les mutations ponctuelles dans le gène PapMV CP. ....</b>	<b>34</b>
<b>Tableau 5. Amorces utilisées pour les doubles délétions du gène PapMV CP. ....</b>	<b>35</b>
<b>Tableau 6. Amorces utilisées pour amplifier la séquence d'initiation d'assemblage du PapMV. ....</b>	<b>41</b>
<b>Tableau 7. Constitution des deux échantillons RMN de la PapMV CP.....</b>	<b>44</b>
<b>Tableau 8. Paramètres d'acquisition des 15N-HSQC .....</b>	<b>45</b>
<b>Tableau 9. Comparaison des rapports d'absorbance obtenu en divisant l'absorbance à 280 nm par celle à 260 nm pour les particules virales.....</b>	<b>97</b>
<b>Tableau 10. Comparaison des rapports d'absorbance obtenu en divisant l'absorbance à 280 nm par celle à 260 nm pour les disques traités à l'acide acétique.....</b>	<b>97</b>

## Liste des figures

Figure 1. Répartition des virus végétaux selon la composition de leur génome.....	3
Figure 2. Structure du dimère de la TMV CP.....	4
Figure 3. Représentation des trois étapes de l'assemblage chez le Virus de la Mosaïque du Tabac.....	7
Figure 4. Organisation du génome des potyvirus.....	9
Figure 5. Structure de la PVA CP proposé par Baratova <i>et al.</i> ....	10
Figure 6. Modèle d'assemblage du Pepper Vein Banding potyvirus.....	11
Figure 7. Caractérisation du PapMV.....	13
Figure 8. Organisation génomique du PapMV.....	14
Figure 9. Structure de la CP du PVX proposée par Baratova <i>et al.</i> .....	15
Figure 10. Comparaison du virus PapMV WT et des NLPs de la protéine recombinante CP <sub>6-215</sub> .....	16
Figure 11. Conservation des acides aminés de la partie centrale des CPs des potexvirus.....	17
Figure 12. Caractérisation de la protéine recombinante K97A.....	19
Figure 13. Caractérisation de la protéine recombinante E128A.....	20
Figure 14. Représentation schématique d'un segment du virion PVX comprenant environ 89 sous- unités pour 10 tours d'hélice.....	22
Figure 15. Représentation en microscopie électronique de l'assemblage à 25 °C de PapMV WT.....	23
Figure 16. Caractérisation de la protéine recombinante CP <sub>27-215</sub> .....	24
Figure 17. Prédiction de structure secondaire de la PapMV CP WT. Site Internet :.....	25
Figure 18. Spectroscopie de dichroïsme circulaire de CP <sub>27-215</sub> .....	26
Figure 19. Séquence des 30 acides aminés présents à l'extrémité N-terminale de la PapMV CP.....	27
Figure 20. Séquence en acides aminés de CP <sub>6-215</sub> .....	28
Figure 21. Schématisation des mutants de délétion de la PapMV CP.....	32
Figure 22. Schématisation des mutants ponctuels de la PapMV CP.....	34
Figure 23. Schématisation des doubles mutants de la PapMV CP.....	35
Figure 24. Profil d'expression protéique sur gel dénaturant SDS-PAGE 10 % de la CP <sub>6-215</sub> à différentes durées d'induction.....	47
Figure 25. Filtration sur gel de la protéine CP <sub>14-215</sub> sur la colonne Superdex <sup>TM</sup> 200 16/60.....	49
Figure 26. Filtration sur gel de la protéine CP <sub>6-215</sub> sur la colonne Superdex <sup>TM</sup> 200 16/60.....	50
Figure 27. Filtration sur gel de la protéine CP <sub>6-215</sub> sur la colonne Superdex <sup>TM</sup> 200 16/60.....	52
Figure 28. Filtration sur gel de la protéine E19P sur la colonne Superdex <sup>TM</sup> 200 16/60.....	53
Figure 29. Alignement de l'extrémité N-terminale des CP des Potexvirus.....	59
Figure 30. Prédiction de structure secondaire à l'extrémité N-terminale pour les PapMV CP <sub>6-215</sub> , CP <sub>15- 215</sub> et CP <sub>21-215</sub> .....	61
Figure 31. Profil de purification des protéines CP <sub>15-215</sub> et CP <sub>21-215</sub> .....	62
Figure 32. Comparaison de la protéine CP <sub>15-215</sub> avant et après la dialyse sur gel dénaturant SDS-PAGE.....	63
Figure 33. Prédiction de structure secondaire à l'extrémité N-terminale pour les PapMV CP <sub>6-215</sub> , E19K et E19P.....	64
Figure 34. Profil de purification des protéines E19K et E19P.....	65
Figure 35. Tamisage moléculaire sur colonne Superdex <sup>TM</sup> 200 16/60 des protéines E19K et E19P.....	66
Figure 36. Microscopie électronique des protéines CP <sub>6-215</sub> , E19K, E19P.....	66
Figure 37. Représentation schématique de la proline.....	68
Figure 38. Vérification de l'induction de la CP <sub>14-215</sub> sur gel SDS-PAGE.....	70
Figure 39. Gel dénaturant SDS-PAGE présentant la protéine CP <sub>14-215</sub> purifiée et dialysée.....	71
Figure 40. Tamisage moléculaire sur colonne Superdex <sup>TM</sup> 75 26/60 de la protéine CP <sub>14-215</sub> .....	72
Figure 41. EMSA des protéines CP <sub>14-215</sub> et CP <sub>27-215</sub> .....	73
Figure 42. 15N-HSQC de la PapMV CP <sub>27-215</sub> et CP <sub>14-215</sub> .....	74
Figure 43. Vérification de l'induction et de l'éluion de la CP <sub>14-215</sub> sur gel SDS-PAGE.....	75
Figure 44. Comparaison de la microscopie électronique des protéines CP <sub>6-215</sub> et CP <sub>12-215</sub> .....	76
Figure 45. Tamisage moléculaire sur colonne Superdex <sup>TM</sup> 200 16/60 de la protéine CP <sub>13-215</sub> .....	77
Figure 46. EMSA de la protéine CP <sub>13-215</sub> .....	78
Figure 47. Comparaison du spectre CD des NLPs de la protéine CP <sub>13-215</sub> à celui des NLPs de la protéine CP <sub>6-215</sub> .....	79

Figure 48. Expression et purification de la protéine CP <sub>14-180</sub> sur gel dénaturant SDS-PAGE..	80
Figure 49. Tamisage moléculaire sur colonne Superdex <sup>TM</sup> 75 26/60 de la protéine CP <sub>14-180</sub> .....	82
Figure 50. Alignement de l'extrémité N-terminale des CPs des <i>Potexvirus</i> .....	86
Figure 51. Schématisation des mutants ponctuels de la PapMV CP pour l'acide aminé F13.....	88
Figure 52. Vérification de l'induction des CPs F13A, F13G, F13L et F13Y sur gel SDS-PAGE.....	89
Figure 53. Gel dénaturant SDS-PAGE présentant les protéines F13A, F13G, F13L et F13Y purifiées et dialysées..	89
Figure 54. Tamisage moléculaire sur colonne Superdex <sup>TM</sup> 200 16/60 des mutants F13.....	90
Figure 55. Comparaison de la microscopie électronique des protéines CP <sub>6-215</sub> ; F13L; F13Y; F13A et F13G.....	92
Figure 56. EMSA des protéines CP <sub>27-215</sub> et F13A..	93
Figure 57. EMSA des protéines CP <sub>6-215</sub> , F13L et F13Y.....	94
Figure 58. Microscopie électronique des disques extraits à l'acide acétique des protéines CP <sub>6-215</sub> , F13L et F13Y.....	95
Figure 59. Spectre CD des NLPs traitées à l'acide acétique pour les protéines CP6-215; F13L et F13Y. ....	96
Figure 60. Représentation schématique de la Phénylalanine; la Tyrosine; la Leucine et l'Alanine.....	98
Figure 61. Représentation schématique du motif structural appelé «Leucine zipper». ....	100
Figure 62. Modèle d'assemblage du Pepper Vein Banding potyvirus.....	104

## 1.0 Introduction

Au cours des dernières décennies, l'intérêt de la communauté scientifique envers les virus d'origine végétale a grandement augmenté. En effet, la structure hautement répétitive des virus végétaux est reconnue efficacement par le système immunitaire, ce qui ouvre la porte au développement de plates-formes de vaccination versatiles dont le développement a été reconnu par l'organisation mondiale de la santé ([www.who.org](http://www.who.org)). Le but est de générer une plate-forme vaccinale universelle à laquelle seraient fusionnés des épitopes d'intérêt à la vaccination. La structure des virus végétaux possède des propriétés immunomodulatrices et peuvent être utilisés comme adjuvant capable de stimuler le système immunitaire en présentant à sa surface des milliers de copies d'un ou de plusieurs épitopes dérivés de pathogènes (Denis, Majeau *et al.* 2007). De plus, la rapidité de production des vaccins ainsi que les bons rendements obtenus à faibles coûts (Porta et Lomonosoff 1998) font de cette technologie une alternative aux vaccins traditionnels en permettant la vaccination à grande échelle dans les pays en voie de développement. Jusqu'à maintenant, plusieurs virus végétaux ont été étudiés à cette fin : le Tobacco Mosaic Virus (TMV), le potato Virus X (PVX), le Cowpea Mosaic Virus (CPMV) et le Alfalfa Mosaic Virus (AIMV) n'en sont que quelques exemples (Canizares, Nicholson *et al.* 2005).

C'est en utilisant le Virus de la Mosaïque de la Papaye (PapMV), un virus flexible à symétrie hélicoïdale, comme plate-forme vaccinale que notre équipe tente de mettre au point des vaccins contre, entre autres, l'hépatite C (HCV), la grippe (Influenza) et le cancer de la peau (mélanomes). Comme la structure tertiaire d'aucune protéine de capsid (CP) d'un virus flexible n'a pu être établie jusqu'à maintenant, la création de ces vaccins demeure complexe. Jusqu'ici, nous avons obtenu du succès à présenter des épitopes d'intérêt en vaccination en fusionnant les peptides à l'extrémité C-terminale de la CP du PapMV. Cependant, il serait utile de caractériser d'autres points de fusion sur la CP afin d'augmenter l'efficacité de la plateforme vaccinale.

Le présent projet de recherche, en collaboration avec le laboratoire du Dr S. Gagné (CREFSIP (Centre de recherche sur la fonction, la structure et l'ingénierie des protéines), U. Laval), un spécialiste de la résonance magnétique nucléaire (RMN), a pour but de

déterminer le rôle de l'extrémité N-terminale de la PapMV CP dans sa multimérisation, puisqu'une délétion des vingt-six premiers acides aminés à cette extrémité résulte à une incapacité pour la nucléocapside à se multimériser. Pour commencer, des informations générales sur les virus de plante seront abordées, suivies de comparaisons entre le PapMV et d'autres virus apparentés à propos de leur protéine de coque et de leur assemblage.

## 1.1 Les virus végétaux

Les virus sont des agents pathogènes ubiquitaires, qui infectent toutes les formes de vie sur terre. Les virus végétaux, pour leur part, sont composés d'un génome viral protégé, en général, par un seul type de protéine, la protéine de capsid, capable de se multimériser autour du génome. Le génome viral peut être constitué d'ADN (simple ou double brin) ou d'ARN (simple ou double brin, de polarité positive ou négative). Chez les virus végétaux, il y a une prédominance marquée pour les virus à ARN simple brin de polarité positive (Figure 1). Qu'ils soient constitués d'ADN ou d'ARN, les génomes viraux conservent tous les mêmes fonctions de base qui sont de coder pour les éléments essentiels à la réplication du virus, au mouvement de cellules en cellules et à l'assemblage en particules matures.

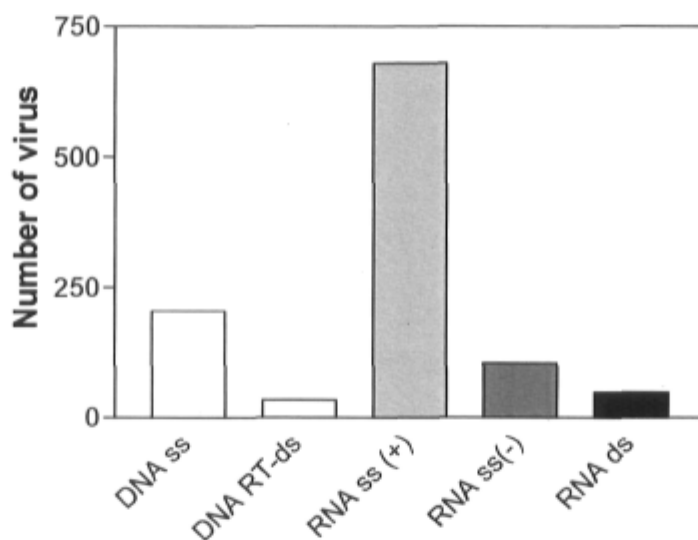


Figure 1. Répartition des virus végétaux selon la composition de leur génome. Légende : DNA ss : ADN simple brin; DNA RT-ds : ADN double brin rétro-transcrit; RNA ss : ARN simple brin (+) polarité positive, (-) polarité négative; RNA ds : ARN double brin. Adapté de : <http://www.dias.kvl.dk/plantvirology/taxonomy.html>.

Les virus végétaux sont classés en deux grands groupes selon la symétrie de leurs capsides : les virus à symétrie icosaédrique et les virus à symétrie hélicoïdale. Les virus icosaédriques ont été longuement étudiés et leur structure a été bien caractérisée. Leurs protéines de capsid sont reconnues pour leur repliement secondaire constitué majoritairement de



feuilletés- $\beta$ . Pour ce qui est des virus hélicoïdaux, les données structurales sont plus difficiles à obtenir. Jusqu'à maintenant, très peu d'études ont été publiées sur le sujet. Jusqu'ici, il a été impossible d'obtenir un échantillon homogène de virus flexible, ce qui a considérablement retardé les études structurales par cristallographie et par RMN.

### *1.1.1 Structures de capsides étudiées par la diffraction des rayons X.*

Pour la diffraction des rayons X, l'échantillon protéique à étudier doit se trouver de façon très ordonné, cristallin et uniforme. De plus, la pureté et la concentration de l'échantillon sont essentielles pour parvenir à de bons résultats. L'obtention de ce cristal est une étape laborieuse et limitante. Dans certain cas, comme pour les virus filamenteux, la cristallisation est impossible. Cependant, cette technique n'est pas complètement inutile pour les virus filamenteux puisqu'elle a pu être utilisée pour le virus de la mosaïque du tabac (TMV). En effet, ce virus peut former une solution gélosée très ordonnée où les hélices des virions sont alignées et qui peut diffracter les rayons X en produisant ainsi un patron de diffraction de fibre (Bawden 1936; Bernal 1941; Gregory 1965). Ainsi, la structure tertiaire de la CP du TMV a pu être déterminée (Figure 2) grâce à cette propriété de la protéine.



Figure 2. Structure du dimère de la TMV CP (PDB : 1EI7).

### 1.1.2 Structures de capsides étudiées par la résonance magnétique nucléaire

La limitation à l'utilisation de la RMN concerne la taille des molécules à tester. Pour un échantillon protéique, cette technique permet de déterminer la structure tridimensionnelle lorsque la protéine se trouve sous une forme monomérique (ou multimérique) inférieure à 30 kDa. Plus la molécule à tester est grosse, plus il risque d'y avoir chevauchement des pics enregistrés. Comme pour la diffraction aux rayons X, la concentration et la pureté de l'échantillon sont conditionnelles à l'obtention de bons résultats. De plus, les molécules doivent être stables tout au long de l'enregistrement des spectres. L'étude en RMN des virus filamenteux devient donc fastidieuse en raison de leur grande taille. Il faut donc trouver un intermédiaire de l'assemblage qui est stable, qui fait moins de 30 kDa et qui ne se multimérise pas en solution. Ces dernières exigences ne sont pas faciles à remplir lorsque l'on considère qu'une des caractéristiques essentielles des protéines de capside est leur grande capacité à se multimériser. C'est donc pour cette raison que très peu de structures de nucléocapside ont pu être déterminées grâce à cette méthode (Khorasanizadeh, Campos-Olivas *et al.* 1999; Campos-Olivas, Newman *et al.* 2000).

### 1.1.3 Les virus hélicoïdaux

Les virus hélicoïdaux peuvent être divisés en deux catégories selon la flexibilité de leurs capsides : on distingue les virus à bâtonnets rigides comprenant les genres *tobamo-*, *tobra-*, *hordei-* et *furovirus* et les virus filamenteux flexibles comprenant les genres *potex-*, *carla-*, *clostéro-*, *poty-* et *bymovirus*. C'est chez les virus à bâtonnets rigides que les données structurales recueillies sont les plus nombreuses et la structure au niveau moléculaire de quelques virus a été établie, comme pour celle du Virus de la Mosaïque du Tabac (Namba, Pattanayek *et al.* 1989; Bhyravbhatla, Watowich *et al.* 1998). Quant aux virus filamenteux flexibles, la structure tertiaire de leurs protéines de capside demeure toujours inconnue. Par contre, comme pour les virus à bâtonnets rigides, il a été établi que les repliements secondaires de leurs protéines de capside sont en grande partie constitués d'hélices alpha (Tremblay 2005).

#### *1.1.4 L'assemblage viral*

Afin de conserver leur pouvoir pathogène, les virus végétaux doivent être capables d'emballer leur génome viral. Cet auto-assemblage se déroule en trois étapes distinctes. Une représentation de cet assemblage chez TMV est illustrée à la Figure 3 (Culver 2002). La multimérisation de la protéine de capsid en dimères, trimères ou disques est la première étape essentielle de l'assemblage. Ensuite, survient l'étape de nucléation où les multimères formés s'associent au génome viral par leur domaine de liaison à l'ARN ou à l'ADN provoquant ainsi un changement de conformation des protéines. Finalement, les particules virales atteignent leur maturité lors de l'étape d'élongation où l'affinité entre les multimères augmente à la suite de changements conformationnels. Ce processus de coopérativité se poursuit jusqu'à la formation de la particule complète. Ces trois étapes sont d'une grande importance puisque chaque étape est limitante et conduit à la suivante. Plusieurs facteurs peuvent influencer les mécanismes d'assemblage des virus, dont les conditions de pH, de température, de concentrations en sel, en protéines et en acides nucléiques. La réaction à ces facteurs est propre à chaque famille virale et parfois même à chaque virus.

En plus de ces facteurs environnementaux, l'assemblage dépend aussi de la capacité des protéines de capsid à interagir entre elles (interactions protéines-protéines) et avec les acides nucléiques (interactions ADN-protéines ou ARN-protéines). Chez les virus icosédriques, l'équipe de Bink et Pleij a mis en évidence que les acides aminés les plus souvent impliqués dans la liaison à l'ARN sont la lysine et l'arginine, des résidus chargés positivement interagissant avec les charges négatives des groupements phosphate de l'ARN (Bink et Pleij 2002). Chez les virus en bâtonnets, les seuls domaines de liaison aux acides nucléiques connus sont retrouvés chez la famille des tobamovirus et sont constitués de trois arginines et d'un acide aspartique (Stubbs 1999).

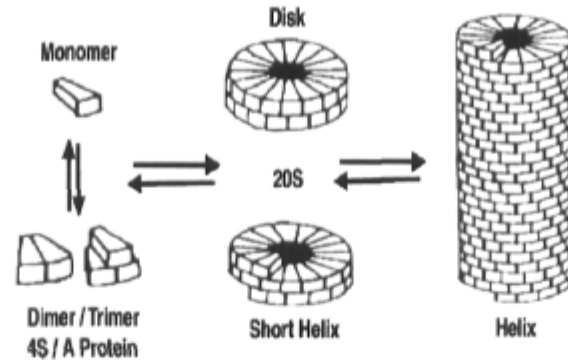


Figure 3. Représentation des trois étapes de l'assemblage chez le Virus de la Mosaïque du Tabac (Culver 2002).

## 1.2 La famille des *Flexiviridea*

Les virus végétaux faisant partie de la grande famille des *Flexiviridea* ont été regroupés ainsi car ils possèdent plusieurs caractéristiques communes (Adams, Antoniw *et al.* 2004). Comme le nom l'indique, ces virus sont flexibles et de longueurs variables (de 470 nm à plus de 1000 nm), mais avec un diamètre fixe de 12 ou 13 nm. Leur génome est constitué d'un ARN simple brin de polarité positive se terminant par une queue polyadénylée en 3'. Cet ARN code pour deux à six cadres de lecture ouverts (Open reading frames-ORFs). Cependant, la majorité des virus en possèdent cinq ou six.

Le premier ORF en 5' de l'ARN code pour une protéine de réplication de 150 à 250 kDa qui possède des motifs correspondants à une méthyl transférase, à une hélicase et à une ARN polymérase ARN-dépendante (Koonin et Dolja 1993). Les deuxièmes, troisièmes et quatrièmes ORFs codent pour les protéines impliquées dans le mouvement du virus de cellules en cellules dans la plante infectée. La plupart des virus possèdent le «*Triple gene block*» constitué des trois protéines de mouvement TGB1, TGB2 et TGB3, mais quelques virus parmi les *Flexiviridea* possèdent une seule protéine de mouvement appartenant à la superfamille des protéines '30K' (Melcher 2000). Ce sont les virus appartenant à cette catégorie qui ne possèdent que deux ou trois ORFs, en fonction de la position de la protéine de capsid sur le gène (s'il fait partie du premier ORF ou s'il se trouve sur un ORF indépendant). Pour les virus à cinq ou six ORFs, c'est sur le cinquième ORF que se trouve

l'information nécessaire à la transcription de la protéine de capsid de poids moléculaire variant de 22 à 44 kDa. Finalement, le sixième ORF présent chez quelques virus est soupçonné de coder pour une protéine ayant des propriétés de liaison aux acides nucléiques (Adams, Antoniw *et al.* 2004). Les *Flexiviridea* ont été divisés en plusieurs genres selon des critères plus précis. On y retrouve 8 genres : les *Allexivirus*, les *Capillovirus*, les *Carlavirus*, les *Foveavirus*, les *Potexvirus*, les *Trichovirus*, les *Vitivirus* et les *Mandarivirus*. De plus, six virus n'ont pas encore été assignés à un genre. Un sommaire des caractéristiques typiques pour chacun des genres ou virus non assimilés à un genre faisant partie de la famille des *Flexiviridea* est présenté au Tableau 1.

Genus	Virion length (nm)	ORFs	Movement proteins <sup>a</sup>	CP (kDa)
<i>Potexvirus</i>	470–580	5	TGB	22–27
<i>Mandarivirus</i>	650	6	TGB	34
<i>Allexivirus</i>	~800	6	TGB	26–29
<i>Carlavirus</i>	610–700	6	TGB	32–36
<i>Foveavirus</i>	800+	5	TGB	28–44
<i>Capillovirus</i>	640–700	2 or 3	30K	25–27
<i>Vitivirus</i>	725–785	5	30K	18–22
<i>Trichovirus</i>	640–760	3 or 4	30K	21–24
Viruses not assigned to a genus:				
<i>Banana mild mosaic virus</i>	580	5	TGB	27
<i>Cherry green ring mottle virus</i>	1000+	5	TGB	30
<i>Cherry necrotic rusty mottle virus</i>	1000+	5	TGB	30
<i>Citrus leaf blotch virus</i>	960	3	30K	41
<i>Potato virus T</i>	640	3	30K	24
<i>Sugarcane striate mosaic-associated virus</i>	950	5	TGB	23

Tableau 1. Sommaire des caractéristiques pour les différents genres et virus faisant partie de la famille des *Flexiviridea* (Adams, Antoniw *et al.* 2004).

### 1.3 Les *Potyvirus*

Les *Potexvirus* possèdent beaucoup de similarités avec d'autres virus filamenteux dont les *Potyvirus*. Le genre *Potyvirus*, genre majeur de la famille des *Potyviridae*, regroupe des virus filamenteux flexibles de 680 à 900 nm de long avec un diamètre de 11 à 15 nm (Urcuqui-Inchima, Haenni *et al.* 2001). Leur génome est constitué d'un ARN simple brin de polarité positive d'environ 10 kb entouré d'environ 2000 copies de la protéine de coque

(Urcuqui-Inchima, Haenni *et al.* 2001). À cause de la ressemblance dans l'organisation de leur génome, les *Potyvirus* ont été inclus dans le super-groupe des «picorna-like viruses». L'ARN viral possède une VPg (viral protein genome-linked) liée de façon covalente à l'extrémité 5' et une queue de poly(A) à l'extrémité 3'. Le génome code pour un seul et long cadre de lecture qui est traduit en une seule polyprotéine de 340 à 370 kDa qui sera clivée par la suite (Urcuqui-Inchima, Haenni *et al.* 2001) (Figure 4).

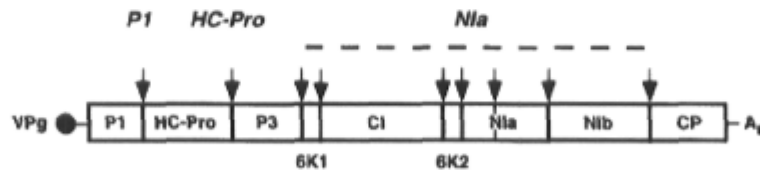


Figure 4. Organisation du génome des potyvirus. Tiré de Urcuqui-Inchima *et al.* (Urcuqui-Inchima, Haenni *et al.* 2001)

#### 1.4 La protéine de capsid des *Potyvirus*

Les *Potyvirus*, comme les *Potexvirus*, possèdent une seule protéine structurale. Les CPs des potyvirus sont de poids moléculaires variant entre 28 et 40 kDa (Urcuqui-Inchima, Haenni *et al.* 2001). Elles peuvent être divisées en trois domaines : l'extrémité N-terminale, la partie centrale et l'extrémité C-terminale. La partie la mieux conservée entre les membres de ce groupe est la partie centrale. Les deux extrémités possèdent des séquences en acides aminés variables et sont exposées à la surface des particules virales. Les extrémités ainsi exposées sont sensibles à la dégradation par la trypsine (Urcuqui-Inchima, Haenni *et al.* 2001). Quand les particules de potyvirus sont digérées par la trypsine, environ 50 acides aminés de l'extrémité N-terminale sont excisés (Shukla, Strike *et al.* 1988; Shukla et Ward 1989; Anindya et Savithri 2003).

Un modèle de la structure du Virus A de la Pomme de terre (PVA) a été développé à partir des résultats obtenus par la méthode de bombardement au tritium combiné aux résultats de prédiction de la structure secondaire de la PVA CP, proposant la composition suivante :

46% d'hélices-alpha et 16% de feuillets-bêta (Baratova, Efimov *et al.* 2001). Ce modèle est présenté à la Figure 5.

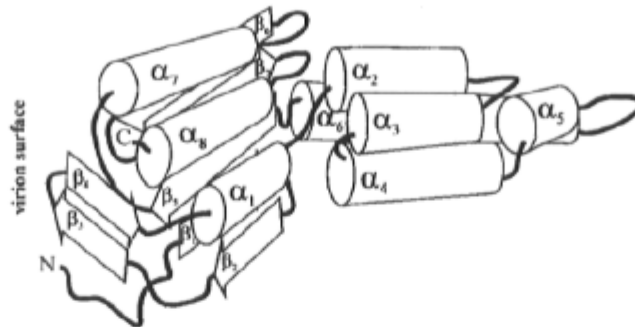


Figure 5. Structure de la PVA CP proposé par Baratova *et al.* (Baratova, Efimov *et al.* 2001).

### 1.5 L'assemblage des *Potyvirus*

Chez le potyvirus Johnsongrass Mosaic Virus, des mutations effectuées aux acides aminés possiblement impliqués dans un pont salin proposé par Dolja (Dolja et Koonin 1991) ont révélé que ces résidus conservés entre les différentes familles de virus hélicoïdaux végétaux sont essentiels à la formation de NLPs. Des mutations ont également été introduites dans le gène de la CP du potyvirus Tobacco Etch Virus (TEV). Les résidus S122, R154 et D198 du TEV sont importants dans la formation de virions (Dolja, Haldeman *et al.* 1994; Dolja, Haldeman-Cahill *et al.* 1995).

Anundya *et al.* ont récemment proposé un modèle d'assemblage pour le Pepper Vein Banding Virus (PVBV) à partir de protéines recombinantes contenant des délétions aux extrémités N- et C-terminales de la CP exprimées dans *E. coli* (Anindya et Savithri 2003). Ce modèle, présenté à la Figure 6, suggère l'implication des extrémités N et C terminales de la CP dans les premières étapes de l'assemblage. L'extrémité N-terminale d'une CP interagirait avec l'extrémité C-terminale d'une autre CP via des forces électrostatiques. Ce processus se poursuivrait jusqu'à l'obtention d'un intermédiaire de l'assemblage de coefficient de sédimentation de 16S. Ce «disque» serait indispensable à l'obtention de

particules virales complètes. Une fois le disque formé, les extrémités N- et C-terminales peuvent être enlevées sans interférer dans le processus d'assemblage.

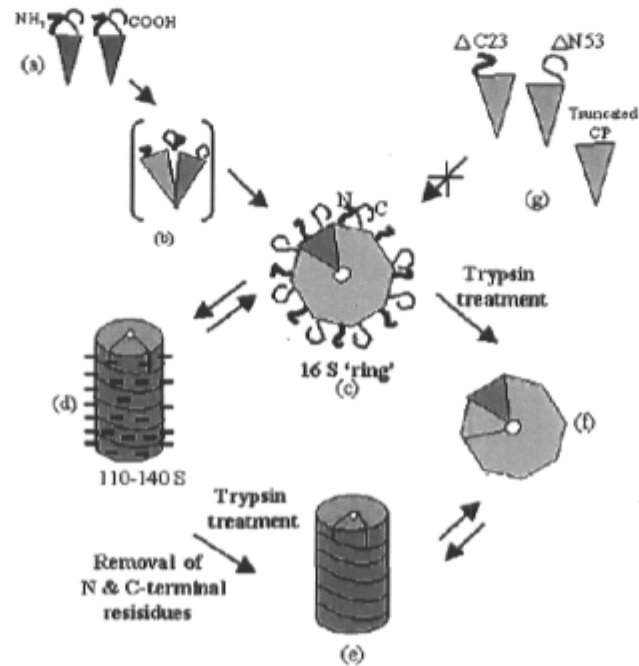


Figure 6. Modèle d'assemblage du Pepper Vein Banding potyvirus. (a) les extrémités N- et C-terminales interagissent ensemble (b) pour former un intermédiaire 16S en forme de disque (c) avec les extrémités exposées à la surface disponibles pour une digestion par la trypsine (f). Ces intermédiaires forment des NLPs à bas pH et à faible concentration ionique (d) et peuvent être reconvertis en intermédiaire 16S. Les extrémités N- et C-terminales exposées à la surface peuvent aussi être digérées par la trypsine à partir des NLPs (e). Ces particules trypsinisées peuvent également être dissociées en intermédiaire 16S (f). Les mutants CPAC23, CPAN53 (g) ne peuvent pas s'assembler en intermédiaire 16S. Tiré de (Anindya et Savithri 2003).



## 1.6 Les *Potexvirus*

Les virus du genre *Potexvirus* sont des virus filamenteux flexibles ayant des longueurs variant de 470 à 580 nm et un diamètre de 13 nm (Adams, Antoniw *et al.* 2004). Leur génome est constitué d'un ARN simple brin de polarité positive qui code pour cinq ORFs sur une longueur qui varie entre 5,9 et 7 kb. Les protéines de capsides codées par le cinquième ORF possèdent des poids moléculaires compris entre 18 et 27 kDa. Plus de vingt-cinq virus différents font partie du genre *Potexvirus* dont le virus modèle est le Virus X de la Pomme de terre (Potato Virus X (PVX)).

### 1.6.1 Le Virus X de la Pomme de terre

Le PVX est un virus flexible de 515 nm formé de 1300 copies de la protéine de capside enroulées autour de l'ARN simple brin de 6435 nucléotides (Huisman, Linthorst *et al.* 1988). Cet ARN possède une coiffe à l'extrémité 5', est polyadénylée en 3' (Skryabin, Kraev *et al.* 1988) et code pour cinq ORFs : l'ORF1 pour la réplicase, les ORF2, 3 et 4 pour les protéines de mouvement et l'ORF5 pour la protéine de capside (Huisman, Linthorst *et al.* 1988; Skryabin, Kraev *et al.* 1988).

### 1.6.2 Le Virus de la Mosaïque de la Papaye

Le Virus de la Mosaïque de la Papaye (PapMV) fait partie du grand groupe des virus filamenteux flexibles et est membre de la famille des *Flexiviridea* du genre *Potexvirus* (Adams, Antoniw *et al.* 2004). L'hôte principal du PapMV est la papaye (*Carica papaya* membre de la famille des Caricaceae), mais dix-sept autres espèces de plantes peuvent être également infectées (Purcifull et Heibert 1971). Les symptômes reliés à une infection au PapMV diffèrent selon l'hôte, mais se manifestent principalement par l'apparition d'une mosaïque (Figure 7A) ou de régions nécrosées sur les feuilles des plants infectés. Ce virus est principalement retrouvé au Pérou, au Venezuela, en Bolivie et aux États-Unis.

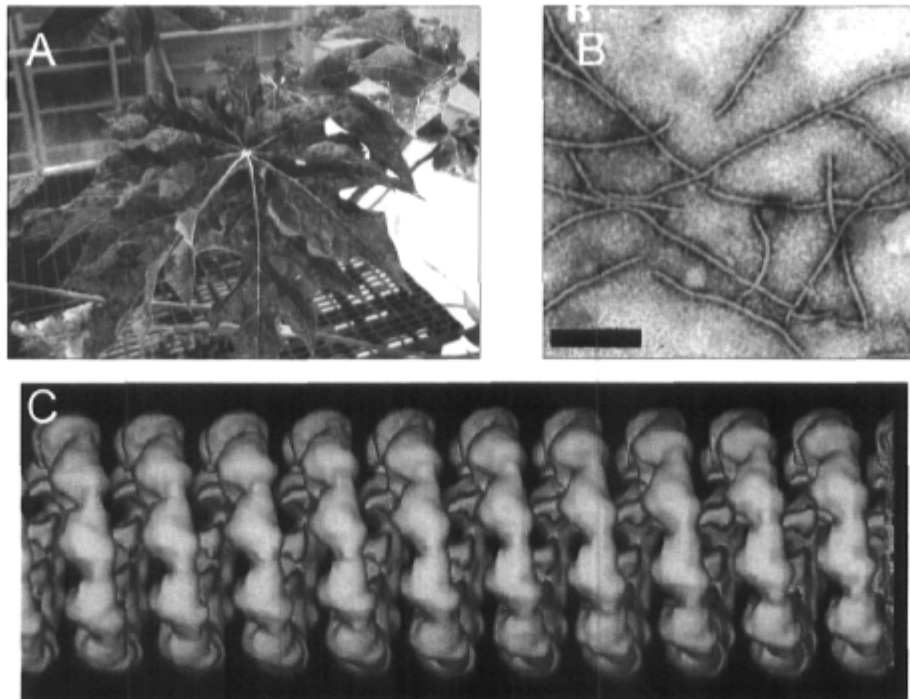


Figure 7. Caractérisation du PapMV. A) Mosaïque caractéristique à une infection par le PapMV sur des feuilles d'un plant de papaye cultivé en serre. B) Photo en microscopie électronique du PapMV. La barre représente 200 nm. (Photos prises par Mme Christine Paré). C) Représentation du PapMV en cryomicroscopie électronique. Collaboration avec le Dr Huilin Li, BNL, Road Island, NY.

Le génome du Virus de la Mosaïque de la Papaye est constitué de 6656 nucléotides (Sit, Abouhaidar *et al.* 1989). Comme pour le PVX, cet ARN simple brin de polarité positive code pour cinq protéines essentielles à la multiplication virale : une réplicase indispensable à la réplication du virus, trois protéines de mouvement TGB1, TGB2 et TGB3 (constituant le «*triple gene block*») impliquées dans le mouvement du virus de cellules en cellules dans la plante infectée et une protéine structurale, la protéine de capsid qui est responsable, entre autres, de la protection du génome viral (Figure 8). Comme chez tous les potexvirus, cet ARN comprend aussi deux régions non codantes aux extrémités 5' et 3', une coiffe à l'extrémité 5' ainsi qu'une queue de poly(A) à l'extrémité 3'. Les particules virales en forme de bâtonnets flexibles (Figure 7B) sont constituées de l'ARN viral entouré de 1400 copies de la protéine de capsid arrangées en hélice (Zhang, Todderud *et al.* 1993). La Figure 7C montre une représentation du PapMV en cryomicroscopie électronique. Chaque tour d'hélice est constitué de 8,75 sous-unités de la capsid et chaque protéine est associée à cinq nucléotides de l'ARN (Tollin, Bancroft *et al.* 1979).



Figure 8. Organisation génomique du PapMV. Légende : *Cap* : coiffe à l'extrémité 5', *TGB 1, 2 et 3* : protéines de mouvement du «triple gene block», *CP* : protéine de capsid, *Poly A* : queue polyadénylée à l'extrémité 3'. Tiré du mémoire de Marie-Hélène Tremblay (Tremblay 2005).

## 1.7 La protéine de capsid des *Potexvirus*

L'alignement des séquences des CPs des membres du genre *Potexvirus* permet de relever qu'il existe un bon nombre d'acides aminés conservés et de remplacements conservateurs (Abouhaidar et Lai 1989) dans la partie centrale de la protéine. Cet alignement nous révèle aussi que l'extrémité N-terminale est la partie la plus variable de la protéine. En effet, les différents potexvirus possèdent une extrémité N-terminale de longueur très variable (Annexe 1).

### 1.7.1 La protéine de capsid de PVX

La PVX CP isolée à partir des virus purifiés se retrouve sous 2 formes majeures d'agrégation en solution, une forme 3-5 S et une forme 10-15 S (Kaftanova, Kiselev *et al.* 1975). Lors de la préparation de virus purifiés à partir de plantes infectées, les particules virales du PVX sont sujettes à la dégradation par des protéases de la plante. Cette dégradation engendre le clivage de 19 à 21 acides aminés à l'extrémité N-terminale qui est présente à la surface des particules virales (Baratova, Grebenshchikov *et al.* 1992).

Les résultats obtenus lors du bombardement au tritium des formes non dégradées et dégradées du PVX combinés à ceux obtenus lors de la prédiction de structure secondaire de la CP, révélant que 45% du repliement de la protéine est sous forme d'hélices-alpha et qu'environ 5 % est sous forme de feuillets-bêta, ont été utilisés pour développer un modèle structural de la CP (Figure 9). Le bombardement au tritium des particules virales intactes révèle qu'environ 35 acides aminés à l'extrémité N-terminale sont situés à la surface des

virions. Sur les particules virales dégradées, l'extrémité C-terminale de la CP devient exposée au bombardement par le tritium suggérant que l'extrémité N-terminale masque l'extrémité C-terminale dans la protéine native (Baratova, Grebenschikov *et al.* 1992). Les études d'activité optique de Raman suggèrent toutefois que ce modèle contient une trop grande quantité de feuillets- $\beta$  bien définis aux extrémités N- et C-terminales (Blanch, Robinson *et al.* 2002).

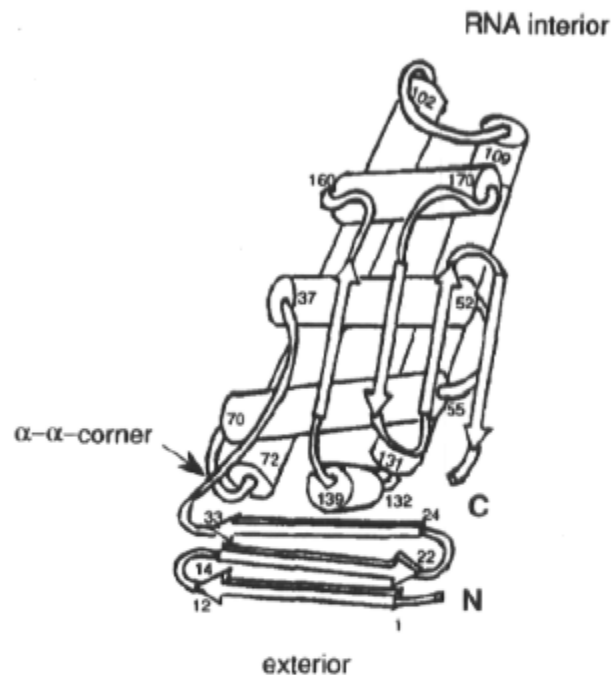


Figure 9. Structure de la CP du PVX proposée par Baratova *et al.* (Baratova, Grebenschikov *et al.* 1992).

### 1.7.2 La protéine de capsid du PapMV

La protéine de la capsid du PapMV est composée de 215 acides aminés et possède une masse moléculaire estimée à 23 kDa (Verde, Malorni *et al.* 1989). Selon une prédiction de la structure secondaire, près de 50 % de la protéine serait repliée en structure d'hélice alpha (Tremblay 2005). Une étude récente (Tremblay, Majeau *et al.* 2006) a démontré que la protéine de capsid portant une délétion sur les cinq premiers acides aminés de l'extrémité N-terminale (CP<sub>6-215</sub>) exprimée dans *Escherichia coli* était capable de s'assembler à l'intérieur de la bactérie. Lors de la purification, 80 % des protéines produites se trouvaient

sous la forme de disques de vingt sous-unités de la protéine de capside. Les 20 % restant était produits sous forme de pseudovirions. Ce résultat suggère donc que la protéine de capside du PapMV possède une affinité pour certains des ARNs présents dans le cytoplasme de la bactérie. La même étude a démontré que la protéine CP<sub>6-215</sub> et la protéine sauvage (PapMV CP WT) partageaient la même structure secondaire (Figure 10A). Par contre, une différence entre les particules virales formées par ces deux protéines est observable en microscopie électronique : les particules virales sauvages possèdent une longueur dix fois plus élevée que les pseudovirions de la protéine CP<sub>6-215</sub> avec des longueurs de 500 nm et de 50 nm respectivement (Figure 10B, C et D).

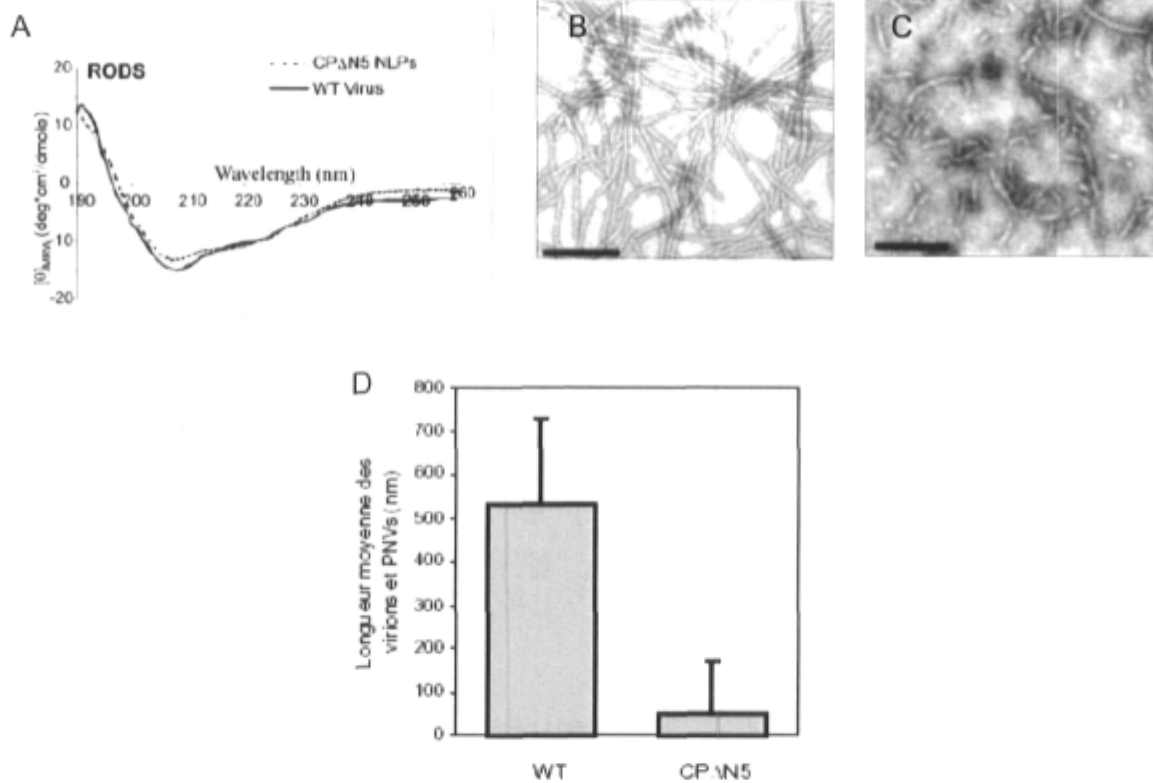


Figure 10. Comparaison du virus PapMV WT et des NLPs de la protéine recombinante CP<sub>6-215</sub>. A) Spectroscopie de dichroïsme circulaire de CP<sub>6-215</sub> NLPs en comparaison avec le virus PapMV WT. Photographie en microscopie électronique B) du virus sauvage (WT) C) de la protéine tronquée CP<sub>6-215</sub>. La barre représente 200 nm. D. Comparaison de la longueur des particules virales en microscopie électronique. ( $n = 150$ )

Une analyse de la PapMV CP réalisée par Marie-Hélène Tremblay (Tremblay 2005) démontre qu'elle contient quatre principaux domaines : un domaine hydrophobe compris entre les résidus 40 à 70, un domaine basique retrouvé entre les acides aminés 97 à 121 et deux domaines hydrophiles situés entre les acides aminés 110 à 140 et 155 à 200. Le domaine basique comprend 6 résidus chargés positivement soit trois lysines et trois arginines. De plus, la position relative de cinq de ces résidus est conservée dans les séquences de protéines de capsid de cinq potexvirus différents (Abouhaidar et Lai 1989). De plus, il a été suggéré par Abouhaidar et Lai en 1989 (Abouhaidar et Lai 1989) que les acides aminés R104, K133, K137, R161 et K198 pourraient être impliqués dans la liaison à l'ARN en interagissant avec les groupements phosphate chargés négativement de l'ARN.

PapMV	"QLASIVKASG TSLRKFCRYF APIIWNIRTD KTPPANWEA
Consensus	-LAG-VKVAN -TLRQFCRY- AKVVMNIRLD KNL PANWAK
PapMV	AGYKPNAKFA AFDFFDGVEN PAAMQPPTGL IIRSPTEERI
Consensus	AGYKEETKFA AFDFFDGVLN PAALEP-DGL IR-PTE-ER-

Figure 11. Conservation des acides aminés de la partie centrale des CPs des potexvirus (Tremblay, Majeau *et al.* 2006). D'après l'alignement des 20 CPs présentés à l'Annexe 1.

Une analyse des acides aminés chargés les plus conservés chez les potexvirus dans la région comprise entre les acides aminés 97 à 168 a été effectuée (Figure 11) et neuf protéines portant diverses mutations ont été produites. Leur capacité à être exprimées dans *E. coli*, la similitude de leur structure secondaire en spectroscopie en dichroïsme circulaire avec la protéine sauvage, leur capacité à s'assembler en particules virales et leur affinité pour l'ARN ont été observées. Les résultats obtenus pour six de ces protéines sont détaillés au Tableau 2.

	<b>Superposition du spectre CD avec CP<sub>6-215</sub></b>	<b>Formation de NLPs</b>	<b>Liaison de l'ARN EMSA</b>
K97A*	Oui	Non	Non
R104-K105-R108A*	Oui	Non	Oui
E128A	Oui	Oui	Oui
K133-K137A*	Oui	Non	Non
E148A	Oui	OUI	Oui
R161A*	Non	Non	N.D.

Tableau 2. Tableau résumé des caractéristiques des protéines contenant des mutations à des acides aminés spécifiques. Tiré du mémoire de Marie-Hélène Tremblay (Tremblay 2005).

Pour quatre des six protéines présentes dans le Tableau 2, la substitution d'un, deux ou trois acides aminés chargés par une ou des alanines a été fatale pour la formation de virions (protéines marquées d'un \*). Cependant, la filtration sur gel sur une colonne Superdex<sup>TM</sup> 75 a permis de déterminer que ces quatre protéines étaient surtout présentes sous forme multimérique plus grande que 70 kDa. Ces mutations ont pu affecter la protéine de capsid soit en modifiant sa structure secondaire l'empêchant ainsi de prendre la bonne conformation (R161A), soit en modifiant son affinité pour l'ARN (K97A et K133-K137A), soit en modifiant sa capacité à interagir avec les autres protéines et à s'assembler en NLPs (R104-K105-R108A). En effet, il a été suggéré par Marie-Hélène Tremblay (Tremblay 2005) que l'acide aminé R104 pourrait être impliqué dans la formation d'un pont salin avec la D142 puisque ces acides aminés sont conservés chez plusieurs famille de virus végétaux filamenteux (Dolja, Boyko *et al.* 1991). Toutes ces modifications dans les fonctions primaires de la protéine de capsid ont pour effet d'empêcher la formation de particules virales. La plupart de ces protéines étaient plutôt instables et difficiles à purifier à l'exception de la protéine K97A qui pouvait être purifiée en bonne quantité, demeurait relativement stable à 4°C et qui a été étudiée plus en détail.

Une analyse en filtration sur gel sur une colonne Superdex<sup>TM</sup> 200 de la protéine K97A a permis de démontrer la présence de molécules à faible poids moléculaire (correspondant probablement à la forme monomérique, dimérique et trimérique de la CP), de disques formés d'environ 20 sous-unités de la CP et, finalement, de molécules à haut poids

moléculaires (correspondant probablement à des agrégats non spécifiques de la CP) (Figure 12A) (Tremblay, Majeau *et al.* 2006). L'absence de NLPs a été confirmée en microscopie électronique (Figure 12B). De plus, l'incapacité de la protéine K97A à interagir avec l'ARN viral a été démontrée par des EMSA en utilisant une sonde marquée au  $\gamma$ - $^{32}\text{P}$  correspondant au 80 premiers nucléotides de l'ARN viral du PapMV (Figure 12C). L'utilisation croissante de protéines (passant de 0 ng à 1500 ng) avec une quantité fixe de sonde marquée (165 fmol) n'a permis d'observer aucune interaction entre l'ARN et la protéine sous forme de disque (flèche rouge). Le complexe ARN-protéine identifié sur le gel serait plutôt dû à une interaction entre la forme monomérique de la protéine et l'ARN marqué.

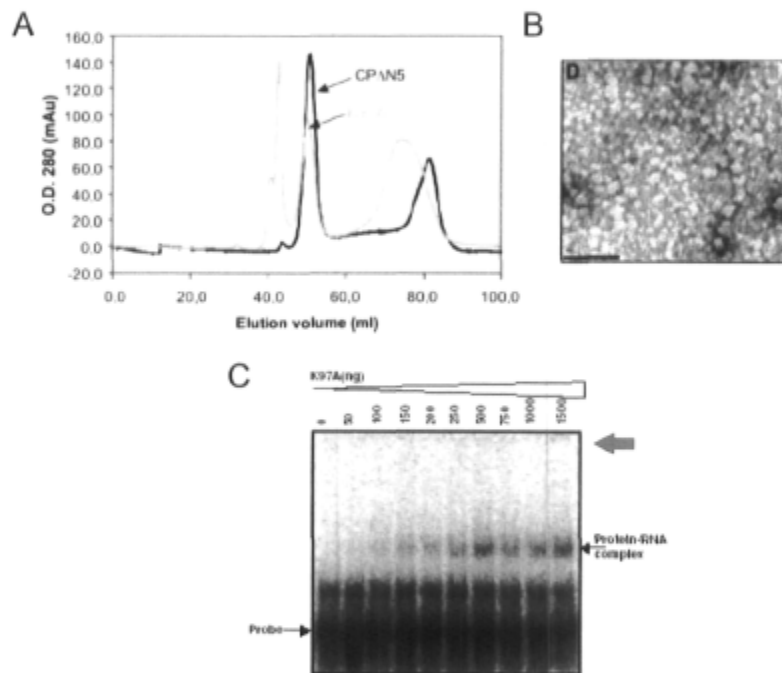


Figure 12. Caractérisation de la protéine recombinante K97A. A) Filtration sur gel de la protéine K97A sur la colonne Superdex<sup>TM</sup> 200 (ligne grise) en comparaison avec le profil obtenu avec le surnageant d'ultracentrifugation de la protéine CP<sub>6-215</sub> (ligne noire). B) Microscopie électronique de la protéine K97A purifiée. La barre représente 200 nm. C) EMSA de la protéine K97A. (Tremblay, Majeau *et al.* 2006)

Pour ce qui est des deux autres protéines présentes dans le Tableau 2, les mutations E128A et E148A ont plutôt eu un effet bénéfique sur la formation de particules. Cependant, les protéines portant la mutation E148A étaient instables et ont précipité lors de la dialyse



servant à enlever l'imidazole utilisé pour leur purification, limitant ainsi leur étude (Tremblay, Majeau *et al.* 2006).

Pour ce qui est de la protéine E128A, des analyses plus approfondies ont pu être effectuées grâce à la grande stabilité de la protéine. Une ultracentrifugation réalisée sur la protéine purifiée a permis d'établir que 100 % de la production se trouvait sous forme de NLPs (Figure 13A) (Tremblay, Majeau *et al.* 2006).

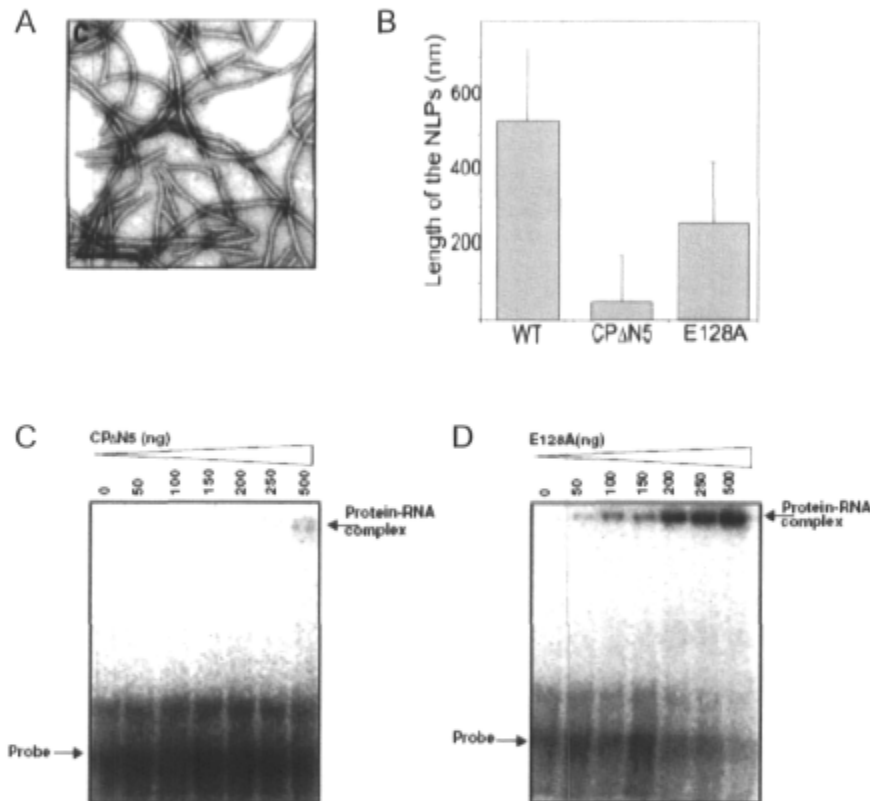


Figure 13. Caractérisation de la protéine recombinante E128A. A) Photographie en microscopie électronique du mutant E128A. La barre représente 200 nm. B) Moyenne des longueurs des particules virales du virus sauvage (WT), de CP $\Delta$ 6-215 (CP $\Delta$ N5) et de E128A. 150 NLPs ont servi à faire les calculs. EMSA de la protéine C) CP $\Delta$ 6-215. D) E128A (Tremblay, Majeau *et al.* 2006).

La structure secondaire des protéines formant les particules a été comparée à celle de la protéine sauvage et les spectres se superposaient. Cependant, la mutation E128A semble avoir modifié l'affinité de la protéine de capsid pour l'ARN viral, tel que le démontre l'EMSA présenté à la Figure 13C et D. Puisque la protéine E128A se retrouve uniquement

sous forme de NLPs, les disques ont du être extraits à l'acide acétique tel que décrit auparavant (Erickson, Bancroft *et al.* 1976). Alors que les disques de la protéine CP<sub>6-215</sub> commencent à avoir une affinité pour l'ARN lorsque 500 ng de protéines sont présentes, il ne faut que 50 ng de disques de la protéine E128A pour interagir avec l'ARN marqué, soit une efficacité dix fois supérieure. Finalement, la longueur des particules produites a été calculée pour le virus sauvage (WT), pour CP<sub>6-215</sub> et pour E128A et les valeurs ont été comparées (Figure 13B). Étonnamment, les particules virales formées dans *E. coli* pour E128A étaient cinq fois plus longues que celles de la protéine CP<sub>6-215</sub> avec une longueur moyenne de 250 nm. Ce résultat suggère donc que la protéine mutée E128A peut plus efficacement supporter l'initiation et l'élongation de la particule lors de l'assemblage.

## 1.8 L'assemblage des *Potexvirus*

Les mécanismes d'assemblage chez les membres du genre potexvirus n'ont pas fait l'objet de beaucoup d'études. C'est chez le PapMV qu'on détient le plus d'informations sur le processus.

### 1.8.1 L'assemblage du PVX

Parker *et al.* ont récemment étudié la surface du PVX en utilisant la diffraction de fibre (Parker, Kendall *et al.* 2002) et une modélisation d'un segment de virus contenant 89 sous-unités protéiques et 10 tours d'hélice a pu être obtenue (Figure 14).

L'impact de plusieurs mutations introduites à l'intérieur de la PVX CP sur la morphologie des virions et sur leur caractère infectieux dans les protoplastes a été étudié par Chapman *et al.* (Chapman, Hills *et al.* 1992). Les résultats obtenus pour des mutants contenant des délétions à l'extrémité N-terminale de la PVX CP supportent le modèle structural proposé par Baratova *et al.* (Baratova, Grebenschikov *et al.* 1992) qui prédit que cette extrémité serait exposée à la surface du virion. Cependant, des formes inhabituelles d'agrégation des virions dont les CP sont délétées à l'extrémité N-terminale ont été observées, ce qui suggère que cette extrémité participerait à des interactions intra- et intermoléculaires dans

la particule virale. Les différences de morphologie observées impliquent que cette partie de la CP facilite la production de virions et stabilise les particules virales.

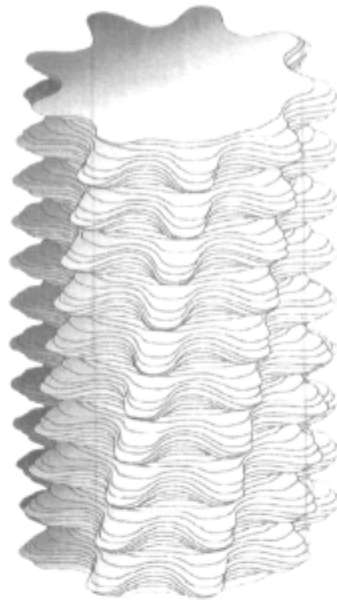


Figure 14. Représentation schématique d'un segment du virion PVX comprenant environ 89 sous-unités pour 10 tours d'hélice. Chaque protubérance représente une sous-unité (Parker, Kendall *et al.* 2002).

Une étude récente de Kozlovsky *et al.* (Kozlovsky, Karpova *et al.* 2003) portant sur la phosphorylation à l'extrémité N-terminale de la PVX CP confirme des résultats obtenus dix ans plus tôt. En effet, la mutation des résidus sérine et thréonine à l'extrémité N-terminale pour des résidus alanine et glycine non phosphorylables rend les particules virales moins stables et plus susceptibles à une digestion par la trypsine. Également, les sous-unités de la CP ainsi mutées se désassemblent spontanément aux extrémités des virions (Kozlovsky, Karpova *et al.* 2003). Kozlovsky *et al.* suggèrent que le domaine N-terminal de la CP est impliqué à une des étapes de l'assemblage. Ils suggèrent aussi que ces mutations engendrent une diminution de la force des interactions protéine:protéine et peut-être ARN:protéine et causent ainsi la perte de stabilité des particules conduisant au désassemblage spontané observé aux extrémités des virions.

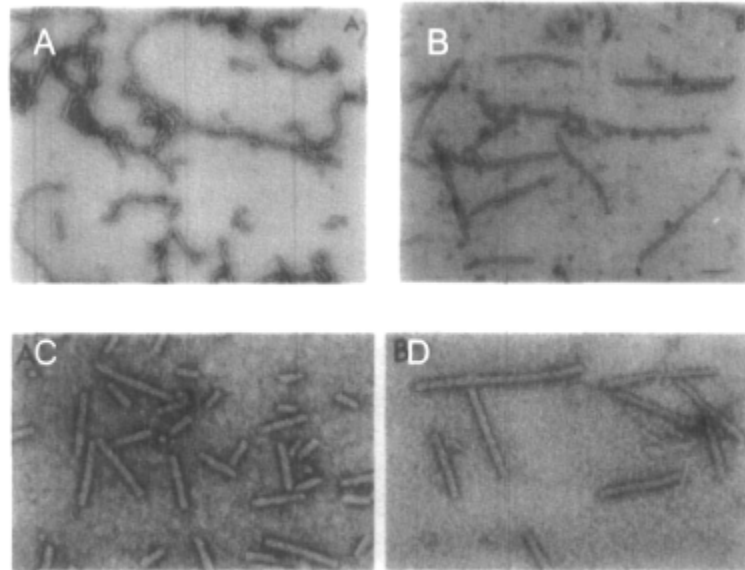


Figure 15. Représentation en microscopie électronique de l'assemblage à 25 °C de PapMV WT traité à l'acide acétique et de son ARN extrait du virus sauvage dans un tampon A) 10 mM MES pH 6.0; B) 10 mM Tris-HCl pH 8.0. La barre représente 100 nm. Adapté de (Erickson et Bancroft 1978). Représentation en microscopie électronique de l'assemblage à 25 °C de PapMV WT traité à l'acide acétique et de son ARN extrait du virus sauvage dans un tampon 10 mM Tris-HCl pH 8.0 dans un rapport protéine:ARN C) 1:4 et D) 1:20. Adapté de (Abouhaidar et Bancroft 1980).

### 1.8.2 Assemblage du PapMV

Les conditions nécessaires à l'assemblage du PapMV *in vitro* ont été longuement étudiées et des conditions optimales ont pu être déterminées : l'assemblage doit être effectué dans un tampon Tris-HCl 10 mM à un pH de 8.0 et à une température de 25°C avec un rapport ARN:protéine de 1:20 (Erickson et Bancroft 1978; Abouhaidar et Bancroft 1980) tel que montré à la Figure 15. Selon une étude publiée en 1978, l'initiation de l'assemblage se fait à l'extrémité 5' de l'ARN viral et ne nécessite que les quarante-sept premiers nucléotides du génome du PapMV (Abouhaidar et Bancroft 1978). La spécificité de l'assemblage est essentiellement observée lors de l'initiation de l'assemblage et est fonction du pH. À un pH de 8.0, la protéine de capsid du PapMV peut initier l'assemblage avec, outre son propre ARN, l'ARN viral de d'autres *Potexvirus* ainsi qu'avec des polyadénines et des polycytosines. Dans tous les cas, l'initiation se fait à l'extrémité 5' de ces ARN et nécessite une région riche en adénines et en cytosines, de telles régions étant considérées comme étant des régions sans structure secondaire. À un pH de 6.0, des assemblages non

spécifiques de la protéine de capsid ont été observés autour d'ARN et d'ADN hétérologues (Abouhaidar et Bancroft 1978).

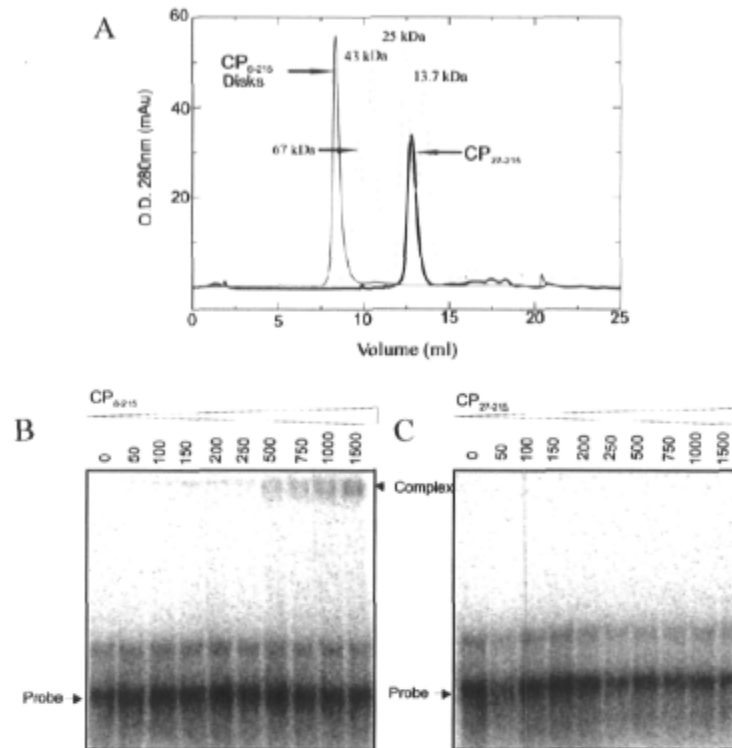


Figure 16. Caractérisation de la protéine recombinante CP<sub>27-215</sub>. A) Tamisage moléculaire de la CP<sub>27-215</sub> (courbe noire) en comparaison avec les disques de la CP<sub>6-215</sub> (courbe grise) sur une colonne Superdex<sup>TM</sup> 75 10/300. Le poids moléculaire du monomère a été estimé à 21,2 kDa à l'aide des marqueurs de poids moléculaire (courbe pointillée). EMSA de la protéine B) CP<sub>6-215</sub>. C) CP<sub>27-215</sub> (Lecours, Tremblay *et al.* 2006).

Les travaux de mutations de Marie-Hélène Tremblay ont mené à la découverte d'un mutant de délétion qui ne se trouvait que sous la forme monomérique lorsqu'elle était exprimée dans *E. coli* (Figure 16A). La protéine CP<sub>27-215</sub>, portant une délétion des vingt-six premiers acides aminés à l'extrémité N-terminale, a été longuement étudiée par Katia Lecours en collaboration avec Marie-Hélène Tremblay en vue de l'utiliser pour des études 3D en RMN. Il a été prouvé que la protéine tronquée était incapable de s'assembler *in vitro* en NLPs ni de former des disques de 20 sous-unités (Lecours, Tremblay *et al.* 2006). Un EMSA de la protéine CP<sub>27-215</sub> a démontré son incapacité à interagir avec les 80 premiers acides nucléiques de l'ARN du PapMV (Figure 16B).



incapacité pour la protéine d'interagir avec les autres protéines (Lecours, Tremblay *et al.* 2006). Cependant, cette incapacité à former des multimères n'est probablement pas causée par un mauvais repliement de la protéine, tel que le suggère les études en spectroscopie de dichroïsme circulaire (Figure 18) (Lecours, Tremblay *et al.* 2006).

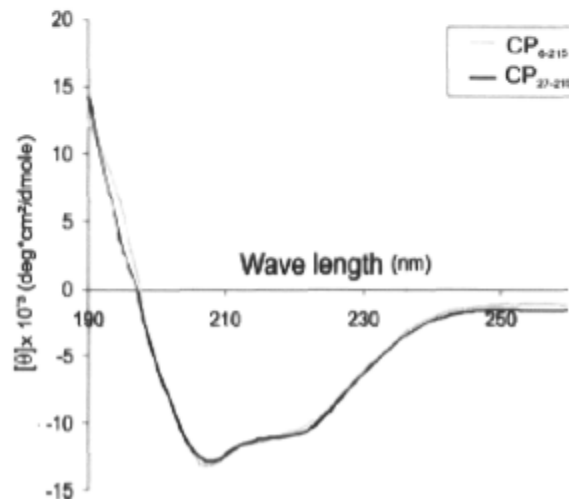


Figure 18. Spectroscopie de dichroïsme circulaire de CP<sub>27-215</sub> (monomère) en comparaison avec la protéine CP<sub>6-215</sub> (disques). Tiré de Lecours et al. (Lecours, Tremblay *et al.* 2006)

Malgré les informations obtenues avec ce mutant nous indiquant l'importance de l'extrémité N-terminale dans la multimérisation, les processus régissant l'assemblage du PapMV demeurent, encore aujourd'hui, méconnus. Nous supposons que les trois grandes étapes décrivant l'assemblage des virus de plantes présentés à la section 1.1.2 sont représentatives de ce qui se passe pour le PapMV, mais nous ignorons quels domaines de la protéine de capsid sont impliqués dans la multimérisation et à quel moment ils sont essentiels.

## 1.9 Hypothèses et objectifs

AA	1	10	20	↓	30
Sequence	<b>MSKS SMSTPNIAFPATTQEQMSSIKV</b> DPTS				
Consensus	○ ○				○ ○ ○ ○ ○ ○

Figure 19. Séquence des 30 acides aminés présents à l'extrémité N-terminale de la PapMV CP ainsi que la prédiction de structure secondaire associée à ces acides aminés. Légende : C : non structuré; H : hélice prédite; flèche rouge : portion déléetée du mutant CP<sub>27-215</sub>.

**L'objectif global de ce projet est de déterminer le rôle de l'extrémité N-terminale de la protéine de capsid e dans l'assemblage du virus de la Mosaïque de la Papaye.**

- a) Le premier objectif de ce travail était d'établir un protocole standard et reproductible de production et de purification des protéines de capsid e du PapMV.
- b) Le deuxième objectif visé est de vérifier l'importance de l'hélice Q18-S23 prédite à l'extrémité N-terminale de la PapMV CP (Figure 19) dans l'auto-assemblage de la CP.
- c) Le troisième but est d'isoler une forme monomérique de la CP propice aux études structurales 3D de la CP.
- d) Finalement, le dernier objectif de ce travail est de vérifier que le caractère hydrophobe de certains résidus de l'extrémité N-terminale de la CP (Figure 19) est important pour les étapes d'initiation de l'assemblage.



## 2.0 Matériel et méthodes

La composition des tampons, des milieux de culture et des réactifs pour les gels sont présentés à l'Annexe 2.

### 2.1 Clonage

Les formes mutantes du gène de la PapMV CP ont été clonées dans le vecteur d'expression pET-3d à partir du gène du mutant de délétion CP<sub>6-215</sub> effectué précédemment par Marie-Hélène Tremblay. La séquence de la protéine codée par ce gène est montrée à la Figure 20. Une alanine en position 2 (illustrée en rouge sur la Figure 20) a été ajoutée lors du clonage par l'utilisation du site de restriction NcoI qui code pour la méthionine initiale suivie d'une alanine. De plus, une queue de six histidines a été ajoutée à l'extrémité C-terminale de la protéine (illustré en bleu sur la Figure 20) afin de faciliter sa purification sur colonne en utilisant des billes de Ni<sup>+2</sup>. La carte du vecteur d'expression pET-3d a été placée à l'Annexe 4. Les oligonucléotides utilisés afin d'introduire les mutations dans le gène ont été synthétisés par le Service de synthèse d'ADN du Centre de Recherche du CHUL.

```

  MASTPN IAFPAITQEQ MSSIKVDPTS NLLPSQEQLK SVSTLMVAAK VPAASVTTVA
  LELVNFCYDN GSSAYTTVTG PSSIPEISLA QLASIVKASG TSLRKFCRYF APIIWNLRT
  KMAPANWEAS GYKPSAKFAA FDFFDGVENP AAMQPPSGLT RSPTQEERIA NATNKQVHLF
  QAAAQDNNFA SNSAFITKGQ ISGSTPTIQF LPPPE HHHHHH

```

Figure 20. Séquence en acides aminés de CP<sub>6-215</sub>. L'alanine additionnelle ajoutée par Marie-Hélène Tremblay est indiquée en rouge. La queue de six histidines en C-terminal de la protéine est indiquée en bleu.

#### 2.1.1 Réactions de polymérase en chaînes (PCR)

Trois stratégies d'amplification ont été utilisées afin de générer les différents mutants selon s'il fallait faire une délétion d'acides aminés ou s'il fallait introduire une mutation ponctuelle à l'intérieur du gène.

### *Stratégie # 1*

Pour les mutations ponctuelles, l'ADN polymérase *Pfu* native (Stratagene, catalogue # 600140) fut utilisée selon les conseils du fabricant. Le principe était d'utiliser deux oligonucléotides avec des extrémités franches qui débutaient soit par l'acide aminé à muter, soit par celui qui le précédait et qui pointaient vers des directions opposées afin d'amplifier la totalité du vecteur pET-3d. La longueur du produit d'amplification était vérifiée par migration sur gel d'agarose 0,8 % et comparée à une échelle d'ADN (New England Biolabs, catalogue # N3232, N3231). Lorsque la longueur du fragment correspondait à ~ 5,2 kb, l'ADN était purifié à l'aide du kit de purification des produits PCR (QIAGEN, catalogue # 28106). Puisque les produits PCR ainsi obtenus ont des extrémités franches, une étape de phosphorylation de ces extrémités à l'aide de la polynucléotide kinase T4 (New England Biolabs, catalogue # M0201) était nécessaire avant de circulariser le plasmide à l'aide de l'ADN ligase T4 (New England Biolabs, catalogue # M0202). La procédure est faite selon les directives du fabricant.

### *Stratégie # 2*

Pour les mutants de délétions, l'ADN polymérase *Long Template* (Roche, catalogue # 1759060) fut utilisée lors de la réaction de polymérase en chaîne. La PCR a été réalisée selon les directives du fabricant. Cette fois-ci, la stratégie était d'utiliser un premier oligonucléotide comprenant un site de restriction qui possède la méthionine initiatrice de la traduction (NdeI), suivi des nucléotides codant pour les premiers acides aminés situés après la délétion désirée. Le deuxième oligonucléotide utilisé possédait aussi le site de restriction utilisé pour le premier oligonucléotide ainsi que 18 nucléotides codant pour les six acides aminés du vecteur pET-3d précédant le gène de la PapMV CP. Les deux oligonucléotides utilisés pointaient vers des directions opposées et le produit de la réaction comprenait le gène partiellement délété en N-terminal de la PapMV CP ainsi que la totalité du vecteur pET-3d. La longueur du produit d'amplification était vérifiée par migration sur gel d'agarose 0,8 % et comparée à une échelle d'ADN (New England Biolabs, catalogue # N3232, N3231). Lorsque la longueur du fragment correspondait à ~ 5,2 kb, l'ADN était purifié à l'aide du kit de purification des produits PCR (QIAGEN, catalogue # 28106). Les extrémités du produit PCR devaient d'abord être digérées à l'aide de l'enzyme de

restriction NdeI (New England Biolabs, catalogue # R0111) selon les directives du fabricant. Une fois la réaction complétée, l'enzyme de restriction était inactivée 20 minutes à 65 °C. Une étape de purification du produit PCR digéré à l'aide du kit de purification des produits PCR (QIAGEN, catalogue # 28106) était nécessaire afin d'enlever l'enzyme inactivé ainsi que les nucléotides digérés afin d'éviter qu'ils interfèrent lors de la recircularisation du plasmide. C'est l'ADN ligase T4 (New England Biolabs, catalogue # M0202) qui fut utilisée pour faire la ligation selon les directives de la compagnie.

### *Stratégie #3*

C'est l'ADN polymérase *Expand High Fidelity* (Roche, catalogue # 1759078) qui fut utilisée pour les mutants CP<sub>14-180</sub> et CP<sub>27-180</sub>. Le premier oligonucléotide utilisé était complémentaire aux six ou sept acides aminés suivant la délétion des treize ou vingt-six premiers acides aminés à l'extrémité N-terminale de la PapMV CP respectivement. Un site de restriction NdeI était aussi présent au début de l'oligonucléotide. Le deuxième oligonucléotide utilisé en sens inverse comprenait les nucléotides codant pour les acides aminés 176 à 180, six histidines, un codon stop et le site de restriction BamHI. Les fragments obtenus à la suite de la réaction de polymérase en chaîne étaient vérifiés par migration sur gel d'agarose 0,8 % et comparés à une échelle d'ADN (New England Biolabs, catalogue # N3232, N3231). Lorsque la longueur des fragments correspondaient à ~ 500 pb pour CP<sub>14-180</sub> et ~ 460 pb pour CP<sub>27-180</sub>, l'ADN était purifié à l'aide du kit de purification des produits PCR (QIAGEN, catalogue # 28106). Les extrémités des produits PCR devaient d'abord être digérés à l'aide des enzymes de restriction NdeI (New England Biolabs, catalogue # R0111) et BamHI (New England Biolabs, catalogue # R0136) selon les directives du fabricant. Une étape de purification des produits PCR digérés à l'aide du kit de purification des produits PCR (QIAGEN, catalogue # 28106) était nécessaire afin d'enlever les enzymes inactivées ainsi que les nucléotides digérés afin d'éviter qu'ils interfèrent lors de la recircularisation du plasmide. C'est l'ADN ligase T4 (New England Biolabs, catalogue # M0202) qui fut utilisé pour faire la ligation entre les fragments purifiés et le vecteur pET-3d digéré avec NdeI et BamHI selon les directives de la compagnie.

### 2.1.2 Transformation de bactéries compétentes par choc thermique

Afin d'amplifier les gènes mutés de la PapMV CP, des bactéries compétentes *Escherichia coli* DH5- $\alpha$  ont été utilisées. La transformation des bactéries compétentes s'effectue en mettant sur glace 30 minutes 200  $\mu$ l de bactéries en présence d'environ 50 ng de plasmides, dans des microtubes de 1,5 ml (Ultident, catalogue # 24-SCT-150-SS-CS). Le choc thermique facilitant l'entrée du plasmide à l'intérieur de la bactérie est fait à 42°C pendant 90 secondes. Les bactéries sont ensuite mises sur glace pour cinq minutes. On ajoute ensuite 800  $\mu$ l de milieu 2xYT (Annexe 2) et on incube les bactéries à 37 °C pendant 45 minutes. Une fois l'incubation terminée, une brève centrifugation de 30 secondes à 15 800 x g est effectuée afin de culotter les bactéries. L'excédent de milieu est jeté ( 800  $\mu$ l) et les bactéries sont resuspendues dans le 200  $\mu$ l de milieu restant avant d'être étalées sur une gélose de 2xYT contenant de l'ampicilline 50 mg/L. Le gène de résistance à l'ampicilline contenu dans le vecteur d'expression pET-3d permet de sélectionner uniquement les bactéries ayant incorporé adéquatement le vecteur d'expression. Les géloses sont ensuite incubées 16 heures à 37 °C.

### 2.1.3 Criblage des mutants

À partir des colonies obtenues à la suite de la transformation des bactéries compétentes *E. coli* DH5 $\alpha$ , l'extraction et la purification de l'ADN des clones sont faites en utilisant le kit *QIAprep Spin Miniprep* (QIAGEN, catalogue # 27106). Les longueurs des ADNs purifiés sont vérifiées sur un gel d'agarose 0,8 % et comparées à une échelle d'ADN (New England Biolabs, catalogue # N3232, N3231). Lorsque les longueurs des fragments correspondent à ~ 5,2 kb, les ADNs sont soumis à une digestion enzymatique afin de vérifier la présence du gène de la PapMV CP. Cette digestion est effectuée à l'aide de l'enzyme de restriction NdeI (New England Biolabs, catalogue # R0111) situé en 5' du gène et de l'enzyme de restriction BamHI (New England Biolabs, catalogue # R0136) situé en 3' du gène. La longueur des fragments digérés est vérifiée sur un gel d'agarose 0,8 % en comparaison avec les marqueurs d'ADN (New England Biolabs, catalogue # N3232, N3231). Finalement, les clones correspondants à ces critères sont séquencés à l'aide d'une amorce complémentaire au promoteur T7 qui se trouve en amont du site de clonage multiple du

vecteur d'expression pET-3d. Les réactions de séquençages sont effectuées par le service de séquençage du centre de recherche du CHUL.

#### 2.1.4 Les mutants de délétions

Les mutants de délétion qui ont été réalisés sont illustrés à la Figure 21. Tous les clones possèdent une séquence codant pour la queue de six histidines en 3' du gène de la PapMV CP. Aucun autre acide aminé n'a été ajouté dans les séquences comme ce fût le cas pour CP<sub>6-215</sub> et CP<sub>27-215</sub> avec les alanines situées après les méthionines initiales.

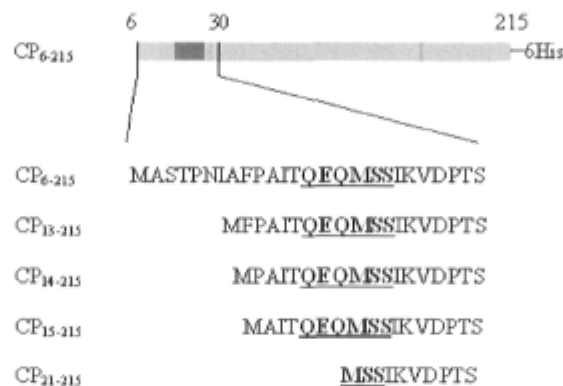


Figure 21. Schématisation des mutants de délétion de la PapMV CP. Les lettres soulignées représentent les acides aminés impliqués dans l'hélice  $\alpha$  prédite selon des logiciels de prédiction de structures secondaires (Annexe 5). Légende : 6His : queue de six histidines.

Le mutant CP<sub>12-215</sub> a été amplifié avec un site de restriction NcoI à l'extrémité 5' puisque cette fois-ci le site de restriction NcoI n'ajoutait pas une alanine supplémentaire au gène parce que l'acide aminé en position douze du gène de la PapMV CP est déjà une alanine. Les autres mutants portant des délétions à l'extrémité N-terminale ont été amplifiés avec un site de restriction NdeI à l'extrémité 5' afin de ne pas ajouter d'alanine supplémentaire. Tous les mutants de délétions ont été amplifiés avec le site de restriction BamHI en 3' de la séquence. Les amorces utilisées pour les PCR sont présentées au Tableau 3.

Nom		Séquence
CP <sub>13-215</sub>	Forward	5' ACGT <u>CATATG</u> TTCCCCGCCATCACCCAG 3'
	Reverse NdeI	3' GAAATTCTTCCT CTATAT <u>GTATACTGC</u> A
CP <sub>14-215</sub>	Forward	5' ACGT <u>CATATG</u> CCCCGCCATCACCCAGGAA 3'
	Reverse NdeI	3' GAAATTCTTCCT CTATAT <u>GTATACTGC</u> A
CP <sub>15-215</sub>	Forward	5' ACGT <u>CATATG</u> GCCCATCACCCAGGAACAAATG 3'
	Reverse NdeI	3' GAAATTCTTCCT CTATAT <u>GTATACTGC</u> A
CP <sub>21-215</sub>	Forward	5' ACGT <u>CATATG</u> AGCTCGATTAAGGTCGATCCA 3'
	Reverse NdeI	3' GAAATTCTTCCT CTATAT <u>GTATACTGC</u> A

Tableau 3. Amorces utilisées pour les délétions du gène PapMV CP. Les acides aminés correspondants aux sites de restrictions utilisés sont soulignés (NdeI : catatg). Légende : Forward : brin sens; reverse : brin complémentaire.

### 2.1.5 Les mutants ponctuels

Les mutants ponctuels réalisés à partir du gène de la CP<sub>6-215</sub> sont illustrés à la Figure 22. Tous les clones possèdent une séquence codant pour la queue de six histidines en 3' du gène de la PapMV CP. De plus, puisque les mutants ont été réalisés à partir du gène de la CP<sub>6-215</sub>, tous les clones possèdent une alanine supplémentaire située immédiatement après la méthionine initiale. Tous les mutants ponctuels ont été clonés avec les sites de restriction NcoI et BamHI. Les amorces utilisées pour les PCR sont présentées au Tableau 4.

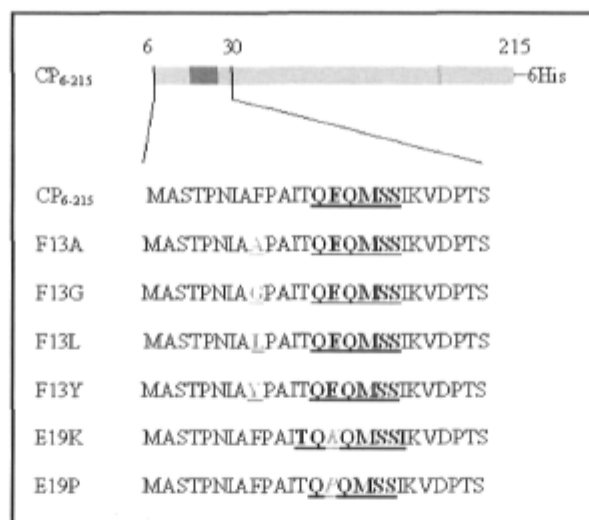


Figure 22. Schématisation des mutants ponctuels de la PapMV CP. Les lettres soulignées en noir représentent les acides aminés impliqués dans l'hélice  $\alpha$  prédite selon des logiciels de prédiction de structures secondaires (Annexe 5). Les lettres colorées et soulignées en gris représentent les acides aminés mutés. Légende : 6His : queue de six histidines.

Nom		Séquence
F13A	Forward	5' GCGCCCGCCATCACCCAGGAACAA 3'
	Reverse	3' CGTAGGTGTGGGTTGTATCGG 5'
F13L	Forward	5' CTGCCCGCCATCACCCAGGAACAA 3'
	Reverse	3' CGTAGGTGTGGGTTGTATCGG 5'
F13G	Forward	5' GGTCCCGCCATCACCCAGGAACAA 3'
	Reverse	3' CGTAGGTGTGGGTTGTATCGG 5'
F13Y	Forward	5' TATCCCGCCATCACCCAGGAACAA 3'
	Reverse	3' CGTAGGTGTGGGTTGTATCGG 5'
E19P	Forward	5' CCGCAAATGAGCTCGATTAAGGTC 3'
	Reverse	3' CGGAAGGGGCGGTAGTGGGTC 5'
E19K	Forward	5' AAACAAATGAGCTCGATTAAGGTC 3'
	Reverse	3' CGGAAGGGGCGGTAGTGGGTC 5'

Tableau 4. Amorces utilisées pour les mutations ponctuelles dans le gène PapMV CP. Les nucléotides correspondants aux acides aminés mutés sont écrits en Italique. Légende : Forward : brin sens; reverse : brin complémentaire.

### 2.1.6 Les mutants CP<sub>14-180</sub> et CP<sub>27-180</sub>

Les mutants CP<sub>14-180</sub> et CP<sub>27-180</sub> sont illustrés à la Figure 23. Des délétions de 13 ou 26 acides aminés ont été réalisées à l'extrémité N-terminale de la PapMV CP en plus de la suppression des 35 derniers acides aminés à l'extrémité C-terminale. Ces deux clones possèdent quand même une séquence codant pour la queue de six histidines en 3' du gène de la PapMV CP. Aucun autre acide aminé n'a été ajouté dans les séquences comme ce fût le cas pour CP<sub>6-215</sub> et CP<sub>27-215</sub> avec les alanines situées après les méthionines initiales.

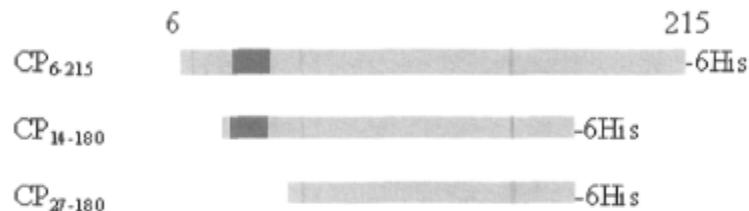


Figure 23. Schématisation des doubles mutants de la PapMV CP. La zone plus foncée à l'intérieure des bâtonnets représente les acides aminés impliqués dans l'hélice  $\alpha$  prédite selon des logiciels de prédiction de structures secondaires (Annexe 5). Légende : 6His : queue de six histidines.

Les mutants CP<sub>14-180</sub> et CP<sub>27-180</sub> ont été amplifiés avec un site de restriction NdeI à l'extrémité 5' afin d'éviter l'ajout d'une alanine supplémentaire et avec le site de restriction BamHI en 3' de la séquence. Les amorces utilisées pour les PCR sont présentées au Tableau 5.

Nom	Séquence
CP <sub>14-180</sub>	Forward 5' ACGT <u>CATATG</u> CCCGCCATCACCCAGGAA 3' Reverse 3' GTCCACGTAGAGAAGGTGGTAGTGGTAGTGGTAATCATT <u>CCTAGGCGTA</u> 5'
CP <sub>27-180</sub>	Forward 5' ACGT <u>CATATG</u> GATCCAACGTCCAATCTTCTG 3' Reverse 3' GTCCACGTAGAGAAGGTGGTAGTGGTAGTGGTAATCATT <u>CCTAGGCGTA</u> 5'

Tableau 5. Amorces utilisées pour les doubles délétions du gène PapMV CP. Les acides aminés correspondants aux sites de restrictions utilisés sont soulignés (NdeI : catatg; Bam HI : ggatcc). Légende : Forward : brin sens; reverse : brin complémentaire.



## **2.2 Production des protéines recombinantes dans *E.coli***

### 2.2.1 Transformation de bactéries compétentes par choc thermique

Afin d'amplifier les gènes mutés de la PapMV CP, des bactéries compétentes *Escherichia coli* BI21 (DE3) (Invitrogen, catalogue # C6000-03) ont été utilisées. La transformation des bactéries compétentes s'effectue en mettant sur glace 30 minutes 200 µl de bactéries en présence d'environ 50 ng de plasmides, dans des microtubes de 1,5 ml (Ultident, catalogue # 24-SCT-150-SS-CS). Le choc thermique facilitant l'entrée du plasmide à l'intérieur de la bactérie est fait à 42°C pendant 90 secondes. Les bactéries sont ensuite mises sur glace pour cinq minutes. On ajoute ensuite 800 µl de milieu 2xYT (Annexe 2) et on incube les bactéries à 37 °C pendant 45 minutes. Une fois l'incubation terminée, une brève centrifugation de 30 secondes à 15 800 x g est effectuée afin de culotter les bactéries. L'excédent de milieu est jeté ( 800 µl) et les bactéries sont resuspendues dans le 200 µl de milieu restant avant d'être étalées sur une gélose de 2xYT contenant de l'ampicilline 50 mg/L. Le gène de résistance à l'ampicilline contenu dans le vecteur d'expression pET-3d permet de sélectionner uniquement les bactéries ayant incorporé adéquatement le vecteur d'expression. Les géloses sont ensuite incubées 16 heures à 37 °C.

### 2.2.2 Induction de la production de protéines

Préparer une préculture contenant 20 ml de milieu 2xYT et 20 µl d'ampicilline 50 mg/ml (50 µg/ml) et inoculer avec 3-4 colonies de bactéries provenant des transformations réalisées à la section 2.2.1. Incuber à 37 °C avec une agitation de 250 RPM pendant 4 heures. Préparer une culture de 500 ml de milieu 2xYT dans un erlenmayer de 2 litres avec 500 µl d'ampicilline (50 µg/ml) et ajouter 1 ml de la préculture. Incuber à 37 °C avec une agitation de 250 RPM jusqu'à ce que la DO à 600 nm soit d'environ 0,6. Prélever 500 µl de la culture avant de l'induire avec 500 µl (1 mM) d'isopropyl-β-D-thiogalactopyranoside 1M (IPTG) et d'ajouter 250 µl d'ampicilline (25 µg/ml). Incuber à 22 °C avec une agitation de 225 RPM pour 16 heures. Prélever 500 µl de la culture après l'induction.

Centrifuger les 500 µl de culture prélevés plus tôt 10 minutes 15 800 x g. Resuspendre les culots ainsi obtenus dans 100 µl de tampon d'électrophorèse (Annexe 2) et ajouter 50 µl de tampon d'échantillon dénaturant 3X. Faire chauffer à 95 °C pour 10 minutes. Ces

échantillons représentent l'ensemble des protéines produites par les bactéries avant et après l'induction de la production de la protéine d'intérêt. L'efficacité de l'induction est vérifiée sur un gel tris-tricine.

### 2.2.3 Lyse des cellules bactériennes

Centrifuger la culture 15 minutes à  $8\,983 \times g$  à  $4\text{ }^{\circ}\text{C}$ . Jeter les surnageants et resuspendre le culot dans 20 ml de tampon de lyse froid (50 mM  $\text{NaH}_2\text{PO}_4$ , 300 mM NaCl et 10 mM imidazole, pH 8.0). Transférer le culot resuspendu dans un Falcon de 50 ml (Sarstedt, catalogue # 62.547.205). Ajouter 40  $\mu\text{l}$  de PMSF 20 mM (40  $\mu\text{M}$ ) et 40  $\mu\text{l}$  de lysosyme 100 mg/ml (0,2 mg/ml). Incuber à  $-80\text{ }^{\circ}\text{C}$  pour 30 minutes. Les culots peuvent être ainsi conservés plusieurs mois. Dégeler les culots sur glace. Une fois dégelés, lyser les bactéries en utilisant la Presse de French en faisant un seul passage à une pression de 750 PSIG. Faire un traitement DNase en ajoutant 900  $\mu\text{l}$  de DNase 10 000 U/ml (Sigma, catalogue # D5025) et 1,5 ml de  $\text{MgCl}_2$  1M filtré. Laisser agiter 15 minutes à  $22\text{ }^{\circ}\text{C}$  sur une plaque agitatrice. Centrifuger 30 minutes à  $20\,442 \times g$  à  $4\text{ }^{\circ}\text{C}$ . Centrifuger à nouveau 30 minutes à  $20\,442 \times g$  à  $4\text{ }^{\circ}\text{C}$ . Laver 3 ml de billes Ni-NTA (QIAGEN, catalogue # 30230) pour 500 ml de culture en ajoutant 10 ml de tampon de lyse. Dans centrifugeuse de table, centrifuger 2 minutes à  $913 \times g$  et jeter le surnageant. Faire 3 lavages avec le tampon de lyse. Incuber le surnageant de centrifugation contenant les protéines et les billes de nickel lavées à  $4\text{ }^{\circ}\text{C}$  sur une plaque agitatrice pour 16 heures. Les protéines possédant une queue d'histidine vont se lier aux billes de nickel pendant ce temps.

## **2.3 Purification des protéines d'intérêts**

Les protéines sont purifiées à l'aide d'une colonne de polypropylène (Econo-Pac Columns de Bio-Rad laboratories, catalogue # 732-6025). Le surnageant de la lyse contenant les billes de nickel est mis sur la colonne. Les protéines couplées aux billes sont lavées en utilisant 20 ml de tampon de lavage 1 (50 mM  $\text{NaH}_2\text{PO}_4$ , 300 mM NaCl et 20 mM imidazole, pH 8.0), 20 ml de tampon de lavage 2 (50 mM  $\text{NaH}_2\text{PO}_4$ , 300 mM NaCl et 50 mM imidazole, pH 8.0) et 20 ml de tampon de lavage 3 (10 mM Tris-HCl et 50 mM imidazole, pH 8.0). Les billes sont incubées pour une période de 3 heures dans 2 ml de

tampon d'éluion (10 mM Tris-HCl et 1 M imidazole, pH 8.0) avant d'être éluées par la gravité. Les protéines sont immédiatement dialysées une heure contre 500 ml de tampon 10 mM tris-HCl pH 8.0 froid, suivi d'une deuxième heure contre 1000 ml du même tampon. Finalement, une dialyse de 16 heures contre 4 litres du même tampon vient compléter le processus d'élimination de l'imidazole. Les protéines dialysées doivent être ensuite conservées à 4°C.

## 2.4 Électrophorèse des protéines et transfert de type Western

L'efficacité de la purification des protéines est vérifiée sur gel de protéines tris-tricine SDS-Page 10% coloré à l'aide du *GelCode Blue Stain Reagent* (Pierce, catalogue # 24590). Un marqueur de poids moléculaire, le *Prestained protein marker* (New England Biolabs, catalogue # P7708S), est utilisé afin d'estimer le poids moléculaire des protéines purifiées. Le transfert de protéine de type western sur une membrane de nitrocellulose est fait dans un appareil à transfert pendant 45 minutes à 200 mAmp. La membrane de nitrocellulose est bloquée pendant 30 minutes à 22 °C dans 25 ml de tampon de blocage (5 % de lait en poudre resuspendu dans du TBS). La membrane est ensuite incubée avec l'anticorps primaire (anticorps polyclonaux générés contre le PapMV) pour une dilution de 1/25000 2 heures à température pièce dans le même tampon. La membrane est lavée sous agitation successivement avec 25 ml de TBS pendant 10 minutes, 25 ml de TBS-Tween 0,05% pendant 3 minutes et 25 ml de TBS pendant 5 minutes. La membrane est incubée dans l'anticorps secondaire (goat anti-rabbit IgG H&L-AP) couplé à l'alkaline phosphatase (GBiosciences, catalogue # 786-10AP-R44) à une dilution 1/25000 pendant 1 heure à température pièce. Les mêmes étapes de lavage sont reprises avant de révéler la membrane avec le substrat de l'alkaline phosphatase (NBT/BCIP) (Pierce, catalogue # 34042).

## 2.5 Ultracentrifugation

Afin de séparer les protéines produites en deux populations, une ultracentrifugation est nécessaire. Les protéines dialysées (2 ml) sont diluées à la ½ en ajoutant 2 ml de tampon 10 mM tris-HCl pH 8.0. L'ultracentrifugation est réalisée à 100 000 g pour 90 minutes dans

un rotor Beckman SW60ti. Le surnageant contenant les disques et les formes de faibles poids moléculaires (< 450kDa) est prélevé et conservé à 4 °C. Le culot de l'ultracentrifugation contenant les particules virales est resuspendu dans 300 µl de tampon 10 mM tris-HCl pH 8.0 et conservé à 4 °C.

## 2.6 Microscopie électronique

La présence de pseudo-nucléocapsides virales est vérifiée par microscopie électronique. Les protéines purifiées sont diluées à une concentration de 50 ng/µl et déposées sur une grille de microscopie 400 mesh de formvar, recouverte d'une couche de carbon (Canemco, catalogue # V0400) pendant 6 minutes. Les grilles sont ensuite lavées deux fois avec 8 µl d'eau, puis colorées avec 8 µl d'acétate uranyle 2% (w/v). Les grilles sont séchées 2 heures dans un dessiccateur avant d'être observée au microscope électronique (JEOL-1010) avec un voltage d'accélération de 80kV en utilisant un grossissement de 60 000 X à 120 000 X. Les images sont prises par le système informatique Gatan.

## 2.7 Spectroscopie en dichroïsme circulaire

### *Structure secondaire des protéines*

Selon la protéine, le surnageant ou le culot d'ultracentrifugation fut utilisé. Les protéines sont tout d'abord dialysées dans de l'eau. Le spectre de dichroïsme circulaire des protéines à 20 °C est obtenu avec un spectrophotomètre Olis RSM 1000. L'échantillon est mis dans une cuvette de quartz de 1 mm d'épaisseur et les spectres sont enregistrés en ellipticité entre 190 et 260 nm. Dix spectres par échantillons sont enregistrés et la moyenne de dix est calculée. Les spectres sont ensuite aplanis selon la méthode de Gorry (1990) (Gorry 1990) et les unités sont converties en ellipticité moyenne par résidu « Mean residue weight (MRW) ellipticity » selon la formule :

$$[\theta]_{\text{MRW}} = [\theta] * \text{MRW} / (10 \times c \times l)$$

où  $[\theta]_{\text{MRW}}$  est l'ellipticité molaire moyenne par résidu en  $\text{deg} \times \text{cm}^2 / \text{dmol}$ ,  $[\theta]$  est l'ellipticité en degrés, MRW est la masse moléculaire moyenne des résidus de la protéine (la valeur

utilisée ici est 108 pour toutes les protéines),  $c$  est la concentration de l'échantillon en mg/mL et  $l$  est l'épaisseur de la cuvette en cm (0.1 cm). Les concentrations en protéines sont estimées en prenant une lecture de la  $DO_{280}$  dans un tampon guanidine-HCl 6M. Le coefficient d'extinction molaire des protéines dans ce tampon est tiré d'une analyse sur le site Internet: <http://www.expasy.org/tools/protparam.html>.

## 2.8 EMSA

La capacité des mutants à lier les acides nucléiques est évaluée par des EMSA (Essais de mobilité électrophorétique sur gel de polyacrylamide)

### 2.8.1 Préparation de la sonde marquée au $^{32}P$

Les EMSA sont réalisés en utilisant une sonde d'ARN simple brin. La sonde est composée des 80 premiers nucléotides à l'extrémité 5' dans la région non codante du génome de PapMV. À partir d'un clone contenant les 4500 premiers nucléotides du génome de PapMV dans le vecteur d'expression pNEB 193 (travaux effectués par Dr Nathalie Majeau), un produit PCR contenant le promoteur T7 en amont des 80 premiers nucléotides est généré. Les amorces utilisées pour générer le produit PCR sont présentées au Tableau 6. La carte du vecteur d'expression pNEB 193 a été placée à l'Annexe 4. On effectue ensuite une réaction de transcription *In vitro* en utilisant le kit *T7 RiboMAX™ Large Scale RNA Production System* (Promega, catalogue # P1320) selon les recommandations du fabricant. L'ARN est purifié sur colonne de G-50 de séphadex *Quick Spin Columns for DNA/RNA Purification* (Roche, catalogue # 1 273 965) et vérifié sur gel d'agarose.

#### *Marquage de l'ARN*

L'ARN est ensuite déphosphorylé avec la phosphatase alcaline de crevette (Shrimp alkaline phosphatase, Fermentas, catalogue # EF0511). L'ARN est extrait au Phénol-chloroforme et précipité à l'éthanol avant le marquage au  $\gamma$ - $^{32}P$ . Le marquage de la sonde au  $\gamma$ - $^{32}P$  est fait avec la polynucléotide kinase T4 (New England Biolabs, catalogue # M0201) selon les directions du fabricant. La sonde est ensuite purifiée sur colonne G-50, extraite au phénol-chloroforme et précipitée à l'éthanol. L'ARN est ensuite resuspendu dans de l'eau et est dosé au spectrophotomètre.

Nom	Séquence
PapMV_INI_80F Forward	5'-actggaattctaatacgcactactatagaaaagaacacaaagca-3'
PapMV_INI_80R Reverse	5'-actggaattcgcctgttgattaccaagtcctggccaaa-3'

Tableau 6. Amorces utilisées pour amplifier la séquence d'initiation d'assemblage du PapMV. F : brin sens (forward), R : brin complémentaire (reverse). La séquence correspondant au promoteur T7 est en italique.

### 2.8.2 Réaction EMSA

Les quantités croissantes de protéines sont incubées avec 165 fmol d'ARN marqué (quantité constante) dans un tampon de liaison auquel est ajouté 7,5 U d'inhibiteur de ribonucléase (RNAGuard) (Amersham Biosciences, catalogue # 27-0816-01). Le volume total de la réaction de liaison est de 10 µl. La réaction de liaison se fait à 22 °C pendant 60 minutes. À la fin de la réaction, 2 µl de bleu natif sont ajoutés à la réaction de liaison et le volume total est mis sur un gel de polyacrylamide natif 5 %. La migration est effectuée dans du tampon TBE 0,5X à 10 mAmp par gel pendant 90 minutes. Le gel est ensuite enveloppé dans un papier plastique et incubé 90 minutes en présence d'un écran de phosphore (*Storage Phosphor Screen*) (Amersham Biosciences, catalogue # 63-0034-87). L'écran est ensuite lu à l'aide d'un Typhoon 9200 (Amersham Biosciences, catalogue # 63-0055-74).

## **2.9 Chromatographie par tamisage moléculaire sur FPLC**

Les protéines fraîchement purifiées et dialysées dans du tampon Tris-HCl 10 mM pH 8.0 contenant 150 mM NaCl sont soumises à un tamisage moléculaire afin de déterminer le niveau de multimérisation de chaque protéine. Un appareil ÄKTA explorer™ est requis pour cette opération. Les colonnes Superdex™ 75 26/60 (Amersham Biosciences, catalogue # 17-1070-01) et Superdex™ 200 16/60 (Amersham Biosciences, catalogue # 17-1069-01) ont été employées. Les marqueurs de poids moléculaire utilisés pour la colonne

Superdex™ 75 sont : l'albumine de sérum de bovin (67 kDa), l'ovalbumine (43 kDa), la chymotrypsine-alpha (25 kDa) et la ribonucléase A (13,7 kDa) alors que ceux utilisés pour la colonne Superdex™ 200 sont : la thyroglobuline (669 kDa), la ferritine (440 kDa), la catalase (232 kDa) et l'aldolase (158 kDa). Une régression linéaire simple du volume d'éluion en fonction du poids moléculaire des marqueurs est faite. L'équation dérivée de la régression linéaire simple, de type  $y = mx + b$ , est ensuite utilisée pour déterminer le poids moléculaire des protéines analysées.

## 2.10 Traitement à l'acide acétique

L'obtention de disques à partir des NLPs de CP<sub>6-215</sub>, F13L et F13Y a été réalisée tel que décrit auparavant (Erickson, Bancroft *et al.* 1976). Deux volumes d'acide acétique glacial ont été ajoutés à un volume de NLPs et incubés une heure à 4 °C. L'ARN libéré est enlevé de l'échantillon par une centrifugation de 15 minutes à 10 000 g. Le surnageant fut récupéré et soumis à une ultracentrifugation de 2 heures à 100 000 g en utilisant un rotor Beckman 50.2Ti afin d'éliminer les NLPs non désassemblées. Afin d'enlever toutes traces d'acide acétique, les protéines furent dialysées 48 heures contre 10 mM Tris-HCl pH 8.0

## 2.11 Purification de PapMV à partir de plants infectés

Toutes les étapes de la purification du virus doivent être faites à 4°C ou sur la glace. Les feuilles de papaye infectées sont broyées dans du tampon Tris-HCl 50 mM pH 8,0 contenant 10 mM EDTA 100 ml pour 100 g de feuille. Le broyat est filtré sur plusieurs épaisseurs de coton à fromage. Au volume recueilli, on ajoute du triton X-100 1% goutte-à-goutte sous agitation. On laisse agiter 10 minutes. On ajoute goutte-à-goutte du chloroforme (1/4 de volume) dans la glace. On incube ensuite 25 minutes sous agitation. On centrifuge 20 minutes à 10 000 g. On récupère le surnageant et laisse évaporer le chloroforme sous agitation pendant 12 heures. Ensuite, centrifuger à 10 000 g pendant 20 minutes. On récupère le surnageant et on procède à une ultracentrifugation à 100 000 g pour 2,5 heures avec un rotor Beckman 50.2Ti . On jette le surnageant et on suspend le culot dans 6 à 12 ml de tampon Tris-HCl 50 mM pH 8,0 en utilisant un «putter» de 10 ml. On ultracentrifuge sur un coussin de 30 % de sucrose. On jette le surnageant et on laisse reposer 12 heures dans un

minimum de tampon. On homogénéise et centrifuge 10 minutes à 20 442 x g. Le surnageant est ensuite transféré et on y ajoute du chloroforme (1/10 de volume), on vortexe et laisse décanter. On récupère ensuite le surnageant et ultracentrifuge pendant 3,5 heures à 100 000 g sur un coussin de 30 % de sucrose dans un rotor beckman 70.1Ti. On jette le surnageant et suspend le culot dans 2 ml de tampon Tris-HCl 50 mM pH 8,0. On centrifuge 5 minutes à 20 442 x g dans des microtubes de 1.5 ml (Ultident, catalogue # 24-SCT-150-SS-CS), récupère le surnageant et ultracentrifuge 2,5 heures à 4°C. Le culot est finalement suspendu dans le plus petit volume possible de tampon Tris-HCl 50mM pH 8,0.

## 2.12 Prédictions de structures secondaires

Les prédictions de structure secondaire furent réalisées à l'aide de logiciels informatiques. Trois algorithmes différents furent utilisés pour chaque clone présent dans la section 2.1 : PROFsec (<http://cubic.bioc.columbia.edu/pp>), PSIPREP (<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>) et Sable-2 (<http://sable.cchmc.org/>). Les résultats obtenus pour les trois algorithmes pour chaque protéine sont présentés à l'Annexe 5. Un consensus pour les trois prédictions a été réalisé. Pour les prédictions réalisées avec le logiciel PROFsec : le *L* représente une région non structurée, le *H* représente une région impliquée dans une structure en hélice-alpha et le *E* représente une région impliquée dans une structure en feuillet-bêta. Pour les prédictions réalisées avec les logiciels PSIPREP, Sable-2 et la séquence consensus : le *C* représente une région non structurée, le *H* représente une région impliquée dans une structure en hélice-alpha et le *E* représente une région impliquée dans une structure en feuillet-bêta. Pour les algorithmes PROFsec, PSIPREP et Sable-2, les niveaux de confiance des prédictions sont représentés par des chiffres où 0 correspond au niveau de confiance le plus faible et 9 au plus élevé. Pour la séquence consensus, les lettres majuscules indiquent des prédictions avec des niveaux de confiance élevés et les lettres minuscules, des niveaux de confiance moins élevés.



## 2.13 Échantillons RMN

Cette section des manipulations fut réalisée par Katia Lecours. Le protocole détaillé pour la RMN se trouve dans les sections 2.17 à 2.19 de son mémoire (Lecours 2007). Les informations principales ont été résumées dans la section suivante.

### 2.12.1 Préparation des échantillons pour RMN

Des échantillons ont été préparés pour les différentes protéines monomériques à l'étude. Tous les échantillons furent filtrés dans des microtubes centrifugeables de 0,5 ml, Amicon Ultrafree<sup>®</sup>-MC 0,22 µm (Millipore), en répondant aux recommandations du fabricant. Une sonde cryogénique fut utilisée sur le spectromètre RMN. Des tubes Shigemi (BMS-005TV (Shigemi)) permettant d'obtenir un meilleur rapport signal sur bruit et de meilleurs spectres furent aussi utilisés. Ces tubes nécessitent d'avoir un échantillon de 18 mm de haut afin que l'échantillon soit bien centré au centre de la bobine de l'aimant RMN. Afin de répondre à ce critère, des échantillons de 350 µl ont été préparés. Tous les échantillons ont été composés de 10 % D<sub>2</sub>O. Ceci a pour effet de bloquer le champ magnétique à une seule fréquence ce qui élimine les risques de fluctuation. De plus, 100 µM de DSS, servant de référence de déplacement chimique à 0 ppm, a été ajouté aux les échantillons. Le pH, pour chaque échantillon, fut mesuré à l'aide d'un pH-mètre équipé d'une électrode de 4 mm, en plus d'être évalué grâce au spectres RMN 1D-<sup>1</sup>H par la position du pic d'imidazole à plus haut déplacement chimique. Les conditions d'échantillons sont résumées au Tableau 7.

Protéine	Marquage	Volume (µl)	[protéine] (mM)	Constitution
CP <sub>27-215</sub>	<sup>15</sup> N	350	0,15	H <sub>2</sub> O, 250 mM imidazole, pH 6,18, 10 mM DTT, 1X cocktail d'inhibiteurs de protéases, 0,1 % NaN <sub>3</sub>
CP <sub>14-215</sub>	<sup>15</sup> N	350	0,15	H <sub>2</sub> O, 250 mM imidazole, pH 6,14, 10 mM DTT, 1X cocktail d'inhibiteurs de protéases, 0,1 % NaN <sub>3</sub>

Tableau 7. Constitution des deux échantillons pour RMN de la PapMV CP. Adapté du mémoire de Katia Lecours (Lecours 2007).

### 2.12.2 Enregistrement des spectres RMN

Cette étude est effectuée par RMN en solution. Le spectromètre RMN utilisé est un Varian INOVA 600 MHz, situé au pavillon Marchand de l'Université Laval, à Québec. Ce spectromètre est doté d'un champ à gradient pulsé pour l'axe Z. Également, il était, au tout début du projet, muni d'une sonde à la température de la pièce à triple résonance. Depuis mars 2005, il est équipé d'une sonde cryogénique à triple résonance qui améliore grandement le ratio signal/bruit. Les spectres RMN ont été enregistrés à une température de 283 K. Les paramètres d'acquisition des spectres sont présentés au Tableau 8.

Protéine	Échantillon	Noyau	Nb de points complexes	Largeurs spectrales	Nb d'accumulations de FID
CP <sub>27-215</sub>	<sup>15</sup> N 0,15 mM	<sup>1</sup> H <sup>15</sup> N	1024 128	12000 1680	1 <sup>er</sup> et 5 <sup>e</sup> =176 2 <sup>e</sup> =176 + mode TROSY 3 <sup>e</sup> =48 4 <sup>e</sup> =32
CP <sub>14-215</sub>	<sup>15</sup> N 0.15 mM	<sup>1</sup> H <sup>15</sup> N	1024 64	12000 1680	48

Tableau 8. Paramètres d'acquisition des <sup>15</sup>N-HSQC enregistrés avec les deux protéines monomériques. Le nombre d'accumulations de FID est décrit pour chaque spectre enregistré par échantillon (FID : Free Induction Decay). Adapté du mémoire de Katia Lecours (Lecours 2007).

### 2.12.3 Analyse des spectres RMN

L'analyse des spectres RMN obtenus pour les deux protéines monomériques a été effectuée par Katia Lecours au cours de sa maîtrise. La description détaillée de la façon dont s'est déroulée l'analyse a été présentée à la section 2.19 du mémoire de Katia Lecours (Lecours 2007).

## 3.0 Élaboration d'un protocole standard pour la production des PapMV CP

### 3.1 Introduction

Il a été récemment démontré qu'une délétion des cinq premiers acides aminés de la PapMV CP n'affecte pas sa capacité à former des pseudovirions dans *E.coli* (Tremblay, Majeau *et al.* 2006). Ce système d'expression de protéines utilisant la machinerie cellulaire de la bactérie *E.coli* est très efficace. En effet, pour un litre de culture bactérienne utilisée, de 40 à 50 mg de la protéine CP<sub>6-215</sub> peuvent être purifiées (Tremblay, Majeau *et al.* 2006). La production de protéines en utilisant les bactéries permet aussi l'auto-assemblage des pseudovirions dans le cytoplasme bactérien. Les protéines purifiées ont été caractérisées par des méthodes biochimiques et comparées à la protéine sauvage (Tremblay 2005; Tremblay, Majeau *et al.* 2006) : aucune différence structurale majeure entre les deux CPs n'a pu être détectée. En plus de nous fournir un bon outil pour la production de grandes quantités de protéines, l'expression dans *E.coli* permet une analyse rapide de formes mutantes et ainsi que de leur impact sur l'auto-assemblage dans la bactérie.

Cependant, les conditions optimales de production et de purification de protéines en utilisant la machinerie cellulaire bactérienne de *E.coli* n'ont pas encore été établies et un certain nombre de paramètres peuvent varier d'une production à l'autre, d'une personne à l'autre. Le premier objectif de ce projet était donc d'établir un protocole standard pour la production et la purification des protéines de capsidique qui serait bénéfique pour les protéines et qui serait identique d'une fois à l'autre afin que les résultats soient répétables. En effet, plusieurs facteurs peuvent influencer le rendement en protéines purifiées ainsi que celui en NLPs : la présence de sel, le pH, la température, etc. La prochaine section a été divisée en quatre sous-sections; chaque section correspond à une des quatre grandes parties du protocole, soit la production, l'extraction, la purification et la dialyse des protéines.

## 3.2 Résultats

### 3.2.1 La production de protéines

Les étapes décrites à la section 2.2.1 ainsi que celles précédant l'induction à l'IPTG à la section 2.2.2 ne diffèrent pas de celles décrites plus tôt par Marie-Hélène Tremblay (Tremblay 2005). Dans le but d'optimiser l'expression, un test sur le temps d'induction idéal fut réalisé avec la protéine CP<sub>6-215</sub>. Des temps de 1, 2, 3, 4, 5 et 16 heures ont été testés. Le résultat est présenté à la Figure 24.

La bande correspondante à la protéine CP<sub>6-215</sub>, située entre les marqueurs de poids moléculaires de 25 et 33 kDa, est facilement identifiable sur le gel de la Figure 24. Le puit #1 du gel montre le répertoire protéique de la bactérie avant l'induction de l'expression de la protéine CP<sub>6-215</sub>. Les puits # 2 à 7 témoignent de la quantité accrue de protéines obtenues en fonction du temps. C'est après 16 heures d'induction que la quantité de protéines d'intérêt produites est à son meilleur.

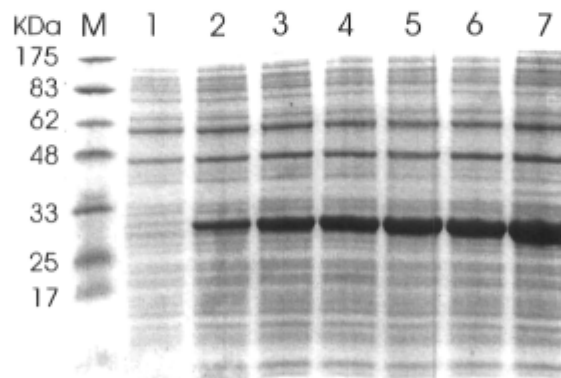


Figure 24. Profil d'expression protéique sur gel dénaturant SDS-PAGE 10 % de la CP<sub>6-215</sub> à différentes durées d'induction. Légende. M : Marqueurs de poids moléculaires; puits 1-7 : Extraits bruts de *E. coli* suivant l'induction de l'expression de la CP<sub>6-215</sub> à 22 °C pour les différentes durées d'induction. 1) 0 heure; 2) 1 heure; 3) 2 heures; 4) 3 heures; 5) 4 heures; 6) 5 heures; 7) 16 heures.

### 3.2.2 L'extraction des protéines (lyse bactérienne)

Cette section concerne le protocole décrit à la section 2.2.3. Afin de réaliser la lyse cellulaire, deux appareils sont mis à notre disposition. La première technique, utilisée par Marie-Hélène Tremblay, est la sonication. Cette technique consiste à faire éclater les

cellules bactériennes par l'utilisation d'ultrasons en faisant de 5 à 10 cycles de 30 secondes d'ultrasons pour 2 minutes de repos sur glace. Les débris cellulaires sont ensuite enlevés par deux centrifugations successives de 20 minutes à 13 000 RPM.

La deuxième technique mise à notre disposition utilise plutôt une Presse de French. Cette fois-ci, la lyse bactérienne est causée par l'application d'une forte pression (750 PSIG) sur les cellules. Les débris cellulaires sont aussi enlevés par deux centrifugations successives de 20 minutes à 13 000 RPM.

La Figure 25 montre une comparaison des deux méthodes pour une même protéine : la protéine CP<sub>14-215</sub>. La CP<sub>14-215</sub> est produite uniquement sous forme monomérique (voir section 5.2.1). De plus, elle est sensible à la dégradation et forme facilement des agrégats non spécifiques lorsque soumise à un stress. La différence de profil obtenu pour les deux méthodes est facilement observable. Lors de l'utilisation de la sonication, plus de 75 % de la protéine se retrouve sous forme d'agrégats non spécifiques (courbe rouge, premier pic à 42.89 ml). Ceci signifie que seulement 25 % de la production totale de protéines pourra être utilisée pour des fins d'analyse. Pour la presse de French, plus de 80 % de la protéine produite se retrouve sous forme monomérique, non agrégée (courbe bleue, pic à 83.34 ml). L'emploi de cette deuxième technique apporte un rendement trois fois supérieur à celui obtenu auparavant.

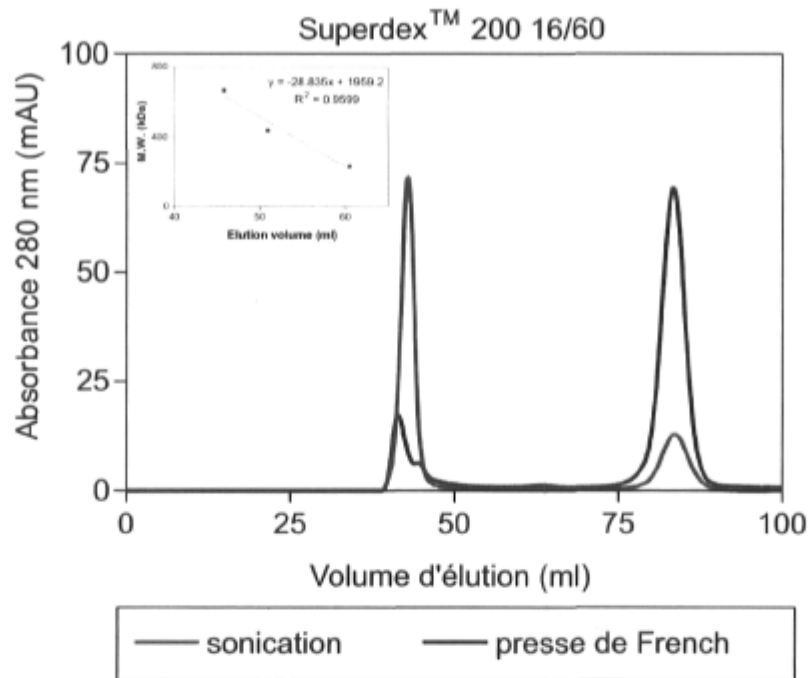


Figure 25. Filtration sur gel de la protéine CP<sub>14-215</sub> sur la colonne Superdex™ 200 16/60. La courbe de couleur rouge représente le profil obtenu pour la protéine purifiée lorsque l'on utilise la sonication pour faire la lyse des cellules bactériennes. La courbe de couleur bleue présente le profil obtenu pour cette même protéine mais lorsque l'on utilise la presse de French pour faire la lyse des cellules bactériennes. Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'élution pour cette colonne a été placé dans le coin supérieur gauche.

L'utilisation de la presse de French semblant augmenter la viscosité de l'échantillon, nous avons pensé faire un traitement DNase afin de détruire l'ADN et ainsi réduire la viscosité. Tel qu'attendu, l'échantillon devient beaucoup plus fluide après un traitement de 15 minutes à 22 °C avec la DNase.

L'impact de l'ajout du traitement DNase fut mesuré par un simple test. La protéine CP<sub>6-215</sub> fut produite en duplicata et purifiée selon le même protocole à l'exception du traitement DNase qui fut exécuté sur seulement un des deux échantillons. Une fois purifiées et dialysées, les protéines furent soumises à une filtration sur gel et comparées. La Figure 26 nous présente les résultats obtenus. Les deux courbes présentées à la Figure 25 ont été réalisées avec 2 mg de protéines chacune et la somme des aires sous la courbe est identique

pour les deux conditions (1087,9 mAU\*ml pour CP<sub>6-215</sub> sans traitement DNase et 1089,3 mAU\*ml pour CP<sub>6-215</sub> avec traitement DNase).

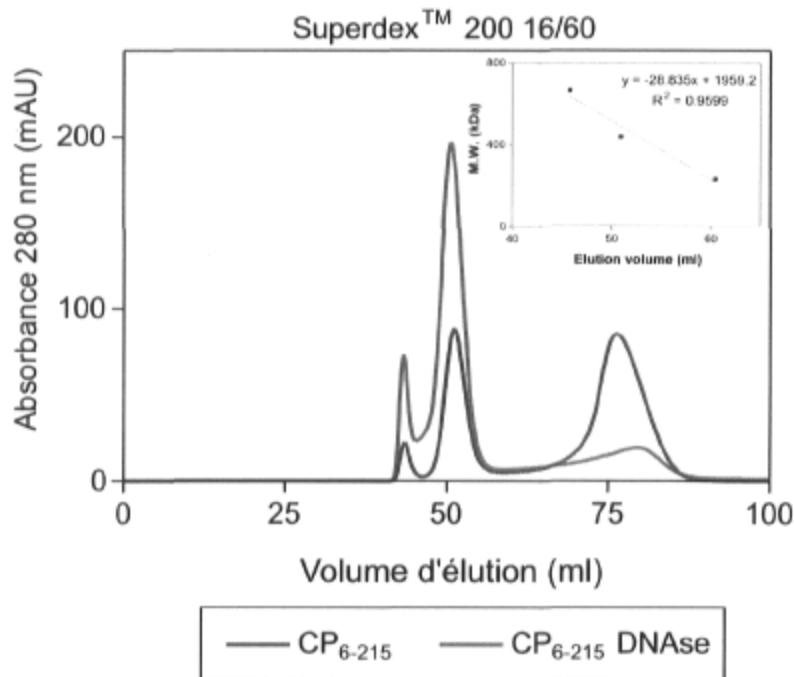


Figure 26. Filtration sur gel de la protéine CP<sub>6-215</sub> sur la colonne Superdex™ 200 16/60. La courbe de couleur mauve représente le profil obtenu pour la protéine sans avoir effectué le traitement DNase (contrôle négatif). La courbe de couleur turquoise présente le profil obtenu pour cette même protéine en ayant subi le traitement DNase tel que décrit à la section 2.2.3. Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'éluion pour cette colonne a été placé dans le coin supérieur droit.

La différence entre les deux profils est facilement observable. Lorsque l'on ne fait pas le traitement DNase, 65 % de la protéine CP<sub>6-215</sub> se retrouvent sous une forme monomérique ou multimérique à faible poids moléculaire (moins de 232 kDa). Le reste de la protéine se trouve sous des formes multimériques à haut poids moléculaires : 30 % correspond à des disques de 450 kDa et 5% correspond aux multimères de plus de 670 kDa dont font partie les pseudovirions. Avec le traitement DNase, le pourcentage de protéines de faible poids moléculaires tombe à 15 %, alors que celui pour les disques augmente à 75 %. La proportion de protéines possédant un poids moléculaire supérieur à 670 kDa double, passant de 5 à 10 % de la production totale.

### 3.2.3 La purification des protéines

Cette section concerne le protocole décrit à la section 2.3. La majorité du protocole demeure intact. Cependant, nous avons voulu vérifier si les conditions de purification que nous utilisions étaient optimales. L'utilisation de haute concentration en imidazole (1M) dans le tampon d'élution est problématique. À plusieurs reprises, les protéines entreposées dans 1M d'imidazole pour plus de 24 heures ont précipité. Les protéines précipitent parfois lors de la dialyse. Nous avons donc voulu connaître la quantité minimale d'imidazole nécessaire pour éluer la PapMV CP.

Un premier test fût réalisé pour la protéine recombinante CP<sub>14-215</sub>. Nous avons diminué la quantité d'imidazole présente dans le tampon d'élution d'un facteur de trois. L'élution a été donc faite avec 300 mM d'imidazole. Après 30 minutes d'incubation avec ce tampon, les billes de nickel étaient à nouveau incubées avec un tampon d'élution, contenant cette fois-ci 1 M imidazole. Cette deuxième étape devait permettre de libérer toutes les protéines encore liées aux billes qui n'ont pas été libérées lors de la première élution. Les deux échantillons furent dosés au spectrophotomètre à une longueur d'onde de 280 nm. La totalité de la protéine avait été éluee avec 300 mM.

Il importait alors de vérifier si les résultats obtenus étaient extrapolables à la CP<sub>6-215</sub>. Nous avons donc procédé de façon identique. L'absorbance à 280 nm a permis de constater que seulement les deux tiers de la protéine étaient éluees avec 300 mM imidazole. Les résultats obtenus pour le monomère CP<sub>14-215</sub> n'étaient donc pas applicables à la protéine CP<sub>6-215</sub>.

Afin de vérifier si les protéines éluees avec une faible quantité d'imidazole possédaient les mêmes caractéristiques que celles éluees avec 1M d'imidazole, un troisième test fut réalisé. Les deux élutions obtenues tel que décrit précédemment ont été soumises à une filtration sur gel sur une colonne Superdex<sup>TM</sup> 200 16/60 et les résultats sont présentés à la Figure 27. La courbe de la CP<sub>6-215</sub> éluee avec 1 M d'imidazole représente le contrôle dans cette expérience. L'élution de la protéine avec 300 mM a donné le profil suivant : 4 % sous forme de multimères de poids moléculaires supérieurs à 670 kDa, 8 % de disques de 450 kDa et 88 % sous forme de multimères de faibles masses moléculaires. L'élution suivante



avec 1 M imidazole a donné un profil quelque peu différent avec des pourcentages de 6 %, 20 % et 74 % respectivement. Cela signifie que plus de 25 % des protéines qui sont restées accrochées à la colonne étaient sous une forme multimérique de haut poids moléculaire. L'éluion avec 300 mM d'imidazole semble donc favoriser le détachement des protéines se trouvant sous un état multimérique de faible poids moléculaires en laissant la plupart des protéines de poids moléculaires supérieurs à 450 kDa accrochées à la colonne.

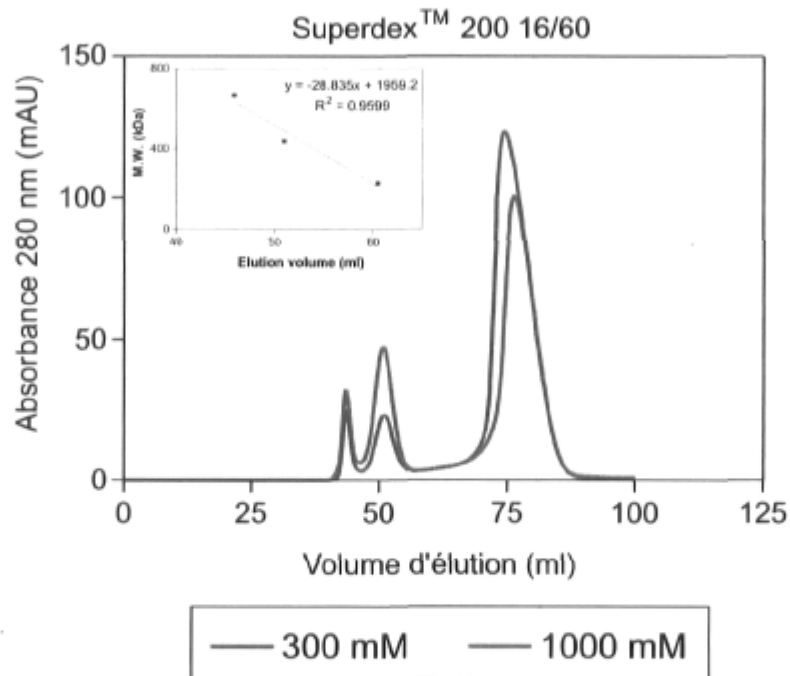


Figure 27. Filtration sur gel de la protéine CP<sub>6-215</sub> sur la colonne Superdex™ 200 16/60. La courbe de couleur orange représente le profil obtenu pour la protéine éluée avec 10 mM tris-HCl pH 8.0 et 300 mM imidazole. La courbe de couleur verte présente le profil obtenu pour cette même protéine éluée une seconde fois avec 10mM tris-HCl pH 8.0 et 1000 mM imidazole. Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'éluion pour cette colonne a été placé dans le coin supérieur gauche.

### 3.2.4 La dialyse des protéines

La précipitation de quelques protéines à plusieurs reprises lors de la dialyse nous a conduits à nous poser des questions sur la pertinence de cette technique. Nous avons soumis la PapMV CP E19P (voire section 4.2.3) à une filtration sur gel sur une colonne Superdex™ 200 16/60 avant et après une dialyse de 18 heures dans du tampon 10 mM Tris-HCl pH 8.0. Les résultats sont présentés à la Figure 28.

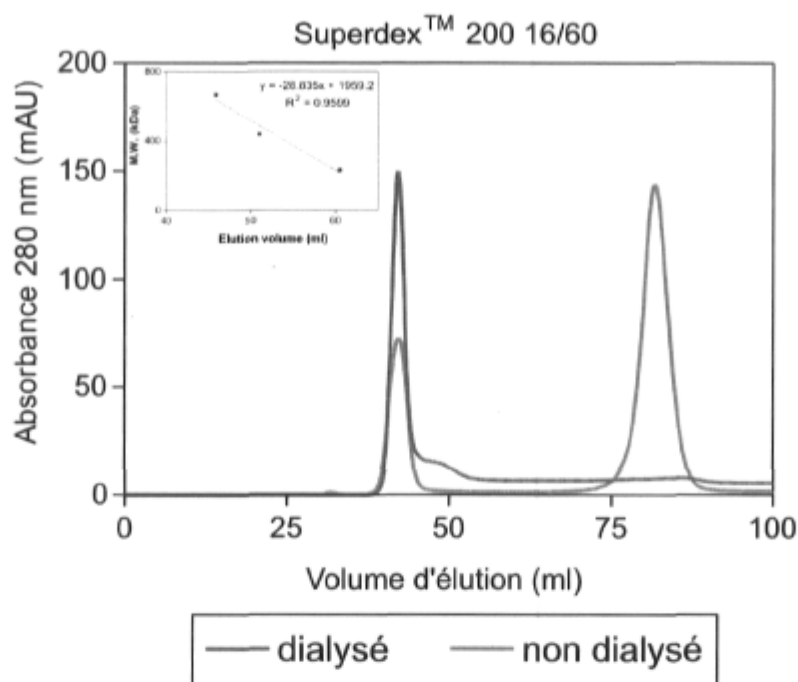


Figure 28. Filtration sur gel de la protéine E19P sur la colonne Superdex™ 200 16/60. La courbe de couleur bleue représente le profil obtenu pour la protéine non dialysée. La courbe de couleur mauve présente le profil obtenu pour cette même protéine mais dialysée. Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'éluion pour cette colonne a été placé dans le coin supérieur gauche.

La différence entre les profils de multimérisation des deux traitements est surprenante. Lorsque la protéine E19P n'est pas dialysée, seulement 25 % de la protéine se trouve sous forme de multimères de plus de 670 kDa, potentiellement des NLPs. Ce pourcentage grimpe à 98 % lorsque une dialyse d'au moins 18 heures dans du tampon 10 mM Tris-HCl pH 8.0 est réalisée. Ce dernier résultat a mis en évidence l'importance de ce traitement dans le processus de purification.

### 3.3 Discussion

#### *La production de protéines*

La réussite de l'expression de la PapMV CP dans *E. coli* étant récente, le temps idéal d'induction n'avait pas encore été vérifié. La Figure 24 présente les résultats d'induction en fonction du temps obtenus pour la CP<sub>6-215</sub>. Des temps de une à cinq heures ont été comparés à celui utilisé précédemment (Tremblay, Majeau *et al.* 2006) de 16 heures. Une

augmentation croissante de CP<sub>6-215</sub> en fonction du temps a pu être remarquée. De plus, le niveau d'expression des autres protéines bactériennes demeure constant tout au long des seize heures d'induction indiquant que l'augmentation de la production de la protéine d'intérêt est réelle et n'est pas due à une présence accrue de bactéries dans l'échantillon.

#### *La lyse bactérienne*

Deux techniques de lyse bactérienne ont été testées et comparées : la sonication utilisant les ultrasons et la presse de French utilisant la pression. Les deux méthodes ont été utilisées en parallèle avec la protéine CP<sub>14-215</sub>. Cette CP monomérique est plus instable et sensible à la dégradation, tout comme la protéine monomérique CP<sub>27-215</sub> étudiée par Katia Lecours (Lecours, Tremblay *et al.* 2006).

L'emploi de la presse de French a permis de recueillir trois fois plus de protéines sous formes monomériques non dégradées et non agrégées que lorsque l'on utilise la sonication. Cette modification au protocole représente une grande économie de temps et d'argent puisque plus de 80 % de la CP<sub>14-215</sub> est désormais utilisable (Figure 25). La sonication produit beaucoup de chaleur qui peut être dommageable pour la protéine d'intérêt. Notre équipe a démontré auparavant (Tremblay, Majeau *et al.* 2006) que le virus de la Mosaïque de la Papaye était particulièrement résistant à la chaleur, pouvant résister à des températures de plus de 80 °C. Les NLPs formés à partir de la CP<sub>6-215</sub> pouvaient quant à eux résister à des températures de 60 °C. Cependant, cette résistance à la chaleur était intimement liée à la forme multimérique sous laquelle la protéine se trouvait. En effet, cette résistance était pratiquement perdue lorsque les disques de 20 sous-unités de la CP étaient testés. On peut donc suggérer qu'elle est quasi inexistante pour les CP sous forme monomérique.

La presse de French, quant à elle, ne produit que très peu de chaleur. Le tube en acier dans lequel les bactéries sont enfermées est conservé à 4 °C et les maintient à cette température tout au long de la procédure. Les protéases sont beaucoup moins actives à cette température qu'à 22 ou 37 °C, diminuant ainsi les risques de dégradation de la protéine.

Une autre modification au protocole a été ajoutée pour cette section. Un traitement DNase de 15 minutes à 22 °C immédiatement après la lyse bactérienne fut ajouté. Cet ajout a un effet bénéfique sur la multimérisation de la CP lors de sa production dans *E. coli*. Plus de 75 % de la CP<sub>6-215</sub> qui se trouvait sous une forme multimérique de faible poids moléculaire a formé des disques ou des multimères de haut poids moléculaire (Figure 26).

Cet impact de la DNase sur le niveau de la multimérisation de la CP peut s'expliquer, en partie, par les propriétés de la CP. Il a été démontré en 1978 que les PapMV CP peuvent s'assembler de façon non spécifique autour d'ADNs hétérologues à un pH de 6.0 (Abouhaidar et Bancroft 1978). Il est donc probable que, lors de la lyse, l'ADN bactérien ainsi libéré se lie de façon non spécifique aux CPs et bloque ainsi l'interaction potentielle entre les CPs et l'ARN. Lorsque l'ADN est dégradé, l'interaction ARN:protéine peut à nouveau ce faire.

#### *La purification des protéines*

La quantité d'imidazole nécessaire à la libération des CPs des billes de nickel fut étudiée pour deux protéines : CP<sub>6-215</sub> et CP<sub>14-215</sub>. Aucune quantité minimale d'imidazole n'a pu être déterminée de façon à convenir à toutes les CPs. Une quantité de 300 mM d'imidazole était suffisante pour éluer la totalité des protéines monomériques CP<sub>14-215</sub> alors que pour CP<sub>6-215</sub>, cette quantité ne suffisait pas. Le niveau de multimérisation des protéines n'est sûrement pas le seul facteur qui influence la libération des protéines, mais son rôle ne peut probablement pas être complètement exclu.

#### *La dialyse des protéines*

Les résultats présentés à la Figure 28 démontrent clairement l'importance de l'étape de la dialyse dans le processus de multimérisation des protéines. Une des raisons qui a été suggérée est que la dialyse permet de se départir des sels et de l'imidazole utilisés lors de la purification (voire section 2.3). Il a été suggéré pour plusieurs potyvirus que des interactions électrostatiques entre des résidus chargés positivement et des résidus chargés négativement hautement conservés entre les virus hélicoïdaux pourraient être impliqués dans le mécanisme d'assemblage des particules virales. Comme des résidus chargés sont

aussi présents chez les potexvirus, ces interactions pourraient aussi être impliquées dans l'assemblage du PapMV. La présence de sels et d'imidazole avec les PapMV CPs purifiées pourrait donc interférer dans les interactions électrostatiques et leur élimination par la dialyse pourrait conduire à un assemblage plus efficace.

## 4.0 Implication de l'hélice Q18-S23 prédite dans l'assemblage du PapMV

### 4.1 Introduction

Notre équipe a récemment démontré que la délétion des cinq premiers acides aminés à l'extrémité N-terminale n'affectait pas la formation de pseudovirions (Tremblay, Majeau *et al.* 2006) alors qu'une délétion des 26 premiers acides aminés abolit complètement la formation de pseudovirions et mène à une forme monomérique de la protéine (Lecours, Tremblay *et al.* 2006). La formation de NLPs est possible grâce à la présence de deux types d'interactions essentielles : 1) les interactions ARN:protéines et 2) les interactions protéines:protéines. Lorsqu'une des deux interactions ne peut pas se faire, l'assemblage n'est pas possible. Dans le cas des virus filamenteux, l'assemblage se déroule en trois étapes (voire section 1.1). La première étape visant la multimérisation de la protéine de coque en dimères, trimères ou disques requière des interactions protéines:protéines. La deuxième étape nécessite quant à elle des interactions ARN:protéines qui permettent le contact entre les multimères formés et les acides nucléiques. La dernière étape, l'élongation, implique les deux types d'interactions à la fois.

L'introduction de mutations à l'intérieur du gène de la protéine de coque peut parfois affecter l'une ou l'autre de ces interactions, comme ce fut le cas pour CP<sub>27-215</sub>. En effet, l'étude de cette protéine selon plusieurs tests de caractérisation a permis de mettre en évidence son incapacité à se multimériser (Lecours, Tremblay *et al.* 2006). Il est important de signaler que cette incapacité à former des multimères n'est probablement pas causée par un mauvais repliement de la protéine, tel que le suggèrent les études en spectroscopie de dichroïsme circulaire (voir Figure 18).

Non seulement cette protéine est produite uniquement sous forme monomérique dans *E. coli*, mais elle échoue aussi au test d'assemblage *in vitro*. Tel que montré à la Figure 16, la protéine CP<sub>27-215</sub> ne possède pas d'affinité pour les 80 premiers acides nucléiques de l'ARN du PapMV. Cependant, il serait faux de conclure que cette délétion affecte la capacité de la protéine à se lier à l'ARN. Selon les trois grandes étapes de l'assemblage chez les virus

hélicoïdaux, qui semblent être valides pour PapMV, la première étape de multimérisation de la protéine de coque en disques est conditionnelle à la deuxième qui permet la liaison des multimères à l'acide nucléique. L'absence d'interaction entre les protéines CP<sub>27-215</sub> et la sonde phosphorylée de la Figure 16 serait donc causée par l'absence de multimérisation.

Ces dernières constatations nous permettent d'affirmer qu'il y a présence d'une région essentielle à la multimérisation comprise entre le sixième et le vingt-sixième acide aminé de la PapMV CP. En premier lieu, nous avons voulu vérifier la présence d'un élément structural reconnu pour être impliqué dans les interactions protéines:protéines. Une prédiction de structure secondaire fut réalisée par Katia et une petite hélice-alpha de six acides aminés fut prédite pour les acides aminés Q18 à S23 (Lecours, Tremblay *et al.* 2006).

Chacun des six acides aminés impliqués dans l'hélice putative fut examiné : cette hélice débute avec la glutamine à la position 18 et possède les acides aminés suivants dans l'ordre : glutamine (Q18), acide glutamique (E19), glutamine (Q20), méthionine (M21), sérine (S22) et sérine (S23). Afin de déterminer lequel ou lesquels de ces acides aminés il serait intéressant de muter, une recherche chez les autres virus végétaux fut effectuée. Tout d'abord, l'alignement des extrémités N-terminale des protéines de capsid de 19 *Potexvirus* fut utilisé pour déterminer le ou les acides aminés les plus importants (Figure 29). L'acide aminé le plus conservé parmi les six impliqués dans l'hélice putative est l'acide glutamique à la position 19 de la PapMV CP. Huit *Potexvirus* sur 19 possèdent cet acide glutamique et trois autres possèdent un acide aspartique à cette position, ce qui implique un changement conservateur. Donc, onze des 19 *Potexvirus* alignés possèdent un acide aminé chargé négativement à cette position.

	6	23
BaMV	GQQAAPQPWE	TKFTK LAA
FoMV	ADVTDATDYK	KPPA TQKA
CymMV	AATYSAADPT	SAPKLA LAA
PAMV	.KQFSASDVR	SSPSLA LDE
NMV	PANADLSDPN	RAPSL LKK
ScaVX	P..ADLSDPT	RAPSLK LQA
PepMV	.APSDFSNPN	TAPSL LKK
WCIMV	~~~HATTTAT	TPPSLT IRA
AltMV	~HSTP...F	PQVTC QMDA
CVX	PRSTPQSGPF	QTLSSQLAA
PIAMV	~~~~~MALN	TAPTA ALAA
TVX	~~~~~MALN	TAPNP ALAA
CsCMV	AATPLSALS	TAPT LSR
CIYMV	~HTDTKKTLF	SAPT QLDT
HVX	PPAAPSPVTF	TAPTC QLTS
LVX	~~~~~MTT	FVPDAKTWAD
SMYEV	NVANQVGDPF	RVLTP LAA
PVX	TPATASGLFT	.IPDG FFST
PapMV	H.STPNIA.F	PAITQ QMSS
Consensus	PA~TP~SDPF	TAPTL LAA

Figure 29. Alignement de l'extrémité N-terminale des CP des Potexvirus. Les CPs de 8 potexvirus sur 19 ont un acide glutamique à la position E19 de la PapMV CP et les CPs de 3 autres *Potexvirus* ont un acide aspartique à la position E19. Légende : acides aminés en rouge : acides aminés chargé négativement; BaMV: *Bamboo mosaic virus*; FoMV: *Foxtail mosaic virus*; CymMV: *Cymbidium mosaic virus*; PAMV : *Potato aucuba mosaic virus*; NMV: *Narcissus mosaic virus*; ScaVX : *Scallion virus X*; PepMV : *Pepino mosaic virus*; WCIMV : *White clover mosaic virus*; AltMV: *Alternanthera mosaic virus*; CVX : *Cactus virus X*; PIAMV : *Plantago asiatica mosaic virus*; TVX : *Tulip virus X*; CsCMV: *Cassava common mosaic virus*; CIYMV: *Clover yellow mosaic virus*; HVX: *Hosta virus X*; LVX: *Lily virus X*; SMYEV : *Strawberry mild yellow edge virus*; PVX : *Potato virus X*; PapMV : *Papaya mosaic virus*. Adapté du mémoire de Marie-Hélène Tremblay (Tremblay 2005).

De plus, 17 des 19 Potexvirus alignés ont un acide aminé chargé négativement situé à proximité de la position 19 de la PapMV CP (Figure 29, acides aminés surlignés en rouge). Étant donné la conservation de la charge négative à cette position, nous avons décidé de vérifier son importance pour l'assemblage. Les acides aminés chargés négativement peuvent parfois interagir avec des acides aminés chargés positivement via des interactions électrostatiques lorsque le contexte est favorable. Quelques études antérieures ont porté sur ces interactions et leur importance pour l'assemblage a été vérifiée chez quelques virus végétaux.

Une étude portant sur l'importance des ponts salins dans l'assemblage du *Physalis mottle tymovirus* (PhMV), un virus végétal icosaédrique de forme T=3, a été récemment réalisée



(Umashankar, Murthy *et al.* 2006). Deux ponts salins ont été découverts et leur rôle a été étudié par mutagenèse. Le premier pont salin étudié se fait via les acides aminés D83 et R159. L'acide aspartique de la position 83 serait exposé à la surface, alors que l'arginine 159 ne le serait que partiellement. Le remplacement simultané de ces deux résidus chargés par des alanines a été démontré pour ne pas affecter la structure des sous-unités ainsi que l'assemblage du PhMV. Les auteurs ont donc conclu que ce pont salin apportait une stabilité additionnelle non essentielle à l'assemblage du virus.

Le second pont salin à l'étude impliquait les acides aminés R68 et D150. Ces résidus se trouvent cachés à l'interface des sous-unités. La mutation de ces acides aminés mène à des protéines partiellement structurées, instables et incapables de s'assembler. Les auteurs ont donc conclu que ce pont salin entre les sous-unités était indispensable à la structure des sous-unités ainsi qu'à l'assemblage (Umashankar, Murthy *et al.* 2006). De plus, ils suggèrent que les ponts ioniques se produisant à la surface sont moins importants pour la stabilité de la particule que ceux impliqués à l'interface des sous-unités.

Une seconde étude fut menée sur le *Potyvirus* Pepper vein banding virus (PVBV) (Anindya et Savithri 2003). Les auteurs ont réalisé plusieurs mutants de délétion aux extrémités N- et C-terminales de la protéine de coque. Les résultats de leurs travaux leur ont permis de proposer un modèle pour l'assemblage du PVBV. Ce modèle est présenté et décrit à la section 1.8. En résumé, ils proposent que l'extrémité N-terminale d'une CP interagirait avec l'extrémité C-terminale d'une deuxième CP et ainsi de suite jusqu'à l'obtention d'un intermédiaire de l'assemblage sous forme de disque, qui pourrait ensuite interagir avec les acides nucléiques pour former la particule virale mature. De plus, ils suggèrent que les interactions entre les extrémités des CPs se feraient au moyen d'interactions électrostatiques. Ce modèle pourrait bien s'appliquer aux autres *Potyvirus* puisque leurs extrémités N- et C-terminale sont exposées à la surface et qu'elles contiennent toutes un bon nombre de résidus chargés et ce, en dépit de leur faible similarité de séquence.

Le deuxième objectif de ce projet était donc de vérifier la présence d'un pont salin chez le PapMV ainsi que de vérifier l'implication et l'importance de l'hélice Q18-S23 prédite à

l'extrémité N-terminale de la PapMV CP. Pour ce faire, des mutants de délétion et des mutants ponctuels furent réalisés. La prochaine section présente les résultats obtenus pour les mutants de délétion CP<sub>15-215</sub> et CP<sub>21-215</sub> suivi de ceux obtenus pour les mutants ponctuels E19K et E19P.

## 4.2 Résultats

### 4.2.1 Prédictions de structures secondaires des mutants de délétion CP<sub>15-215</sub> et CP<sub>21-215</sub>

Afin de vérifier l'importance de l'hélice Q18-S23 prédite par les prédictions de structure secondaire, deux mutants de délétions furent réalisés : CP<sub>15-215</sub> et CP<sub>21-215</sub>. La Figure 30 présente ces prédictions à l'extrémité N-terminale pour les deux mutants de délétion.

La délétion des 14 premiers acides aminés ne semble pas affecter la formation de l'hélice Q18-S23 selon les prédictions de structure secondaire présentées à la Figure 30. Par contre, la délétion des 20 premiers acides aminés à l'extrémité N-terminale semble efficace pour la détruire. La comparaison du niveau de multimérisation de ces deux mutants devrait donc nous permettre de déterminer le rôle de l'hélice Q18-S23 dans la formation de multimères et dans l'assemblage.

AA	10	20	30	40	50	60
Séquence	MASTPNIAFPAIT	<b>QEQMSS</b> IKVDPTSNLLPSQEQLKSVSTLMVAAKVPAAASVTTVALELV				
CP <sub>6-215</sub>	CCCCCCCCCCCC	<b>HHHHHh</b> cCCCCcccCCCHHHHHHHHHHHHHHcCCCChHHHHHHHHHH				
CP <sub>15-215</sub>		CC <b>HHHHHh</b> cCCCCcccCCCHHHHHHHHHHHHHHcCCCChHHHHHHHHHH				
CP <sub>21-215</sub>			CcccCCCCcccCCCHHHHHHHHHHHHHHcCCCChhHHHHHHHHHH			

Figure 30. Prédiction de structure secondaire à l'extrémité N-terminale pour les PapMV CP<sub>6-215</sub>, CP<sub>15-215</sub> et CP<sub>21-215</sub>. Trois algorithmes différents furent utilisés pour chaque clone : PROFsec, PSIPREP et Sable-2 et le consensus de ces trois prédictions est présenté ici. Le C représente une région non structurée et le H représente une région impliquée dans une structure en hélice-alpha. Les lettres majuscules indiquent des prédictions avec des niveaux de confiance élevés et les lettres minuscules, des niveaux de confiance plus faibles. La totalité des prédictions se trouve à l'Annexe 5.

#### 4.2.2 Production et purification des mutants de délétion

Les mutants CP<sub>15-215</sub> et CP<sub>21-215</sub> furent produits dans *E. coli* et purifiés selon le protocole décrit aux sections 2.2 et 2.3. Les protéines induites ainsi que celles éluées ont été vérifiées sur un gel dénaturant SDS-PAGE (Figure 31).

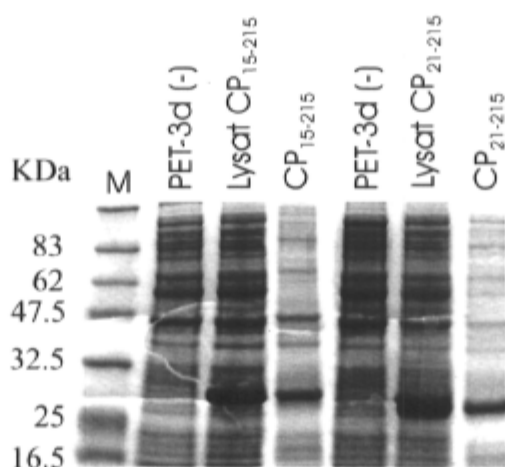


Figure 31. Profil de purification des protéines CP<sub>15-215</sub> et CP<sub>21-215</sub>. Gel dénaturant SDS-PAGE pour les deux protéines. Légende : M : marqueurs de poids moléculaire; pET-3d (-) : lysat bactérien avant induction; Lysat CP<sub>15-215</sub> ou CP<sub>21-215</sub> : lysat bactérien après l'induction à l'IPTG 16h à 22 °C; CP<sub>15-215</sub> ou CP<sub>21-215</sub> : élution des protéines avec 1 M imidazole.

Malgré l'usage du protocole de production et purification amélioré décrit aux sections 2.2 et 2.3, une très faible quantité de protéines CP<sub>15-215</sub> et CP<sub>21-215</sub> fut élue. De plus, le niveau de pureté des éluats était médiocre. Lorsque qu'une dialyse afin d'enlever l'imidazole était effectuée avec les éluats, la majorité des protéines précipitaient et celles non précipitées étaient en partie dégradées (Figure 32). En raison de leur instabilité, les deux mutants de délétion furent mis de côté.

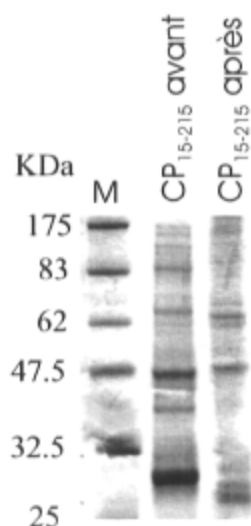


Figure 32. Comparaison de la protéine CP<sub>15-215</sub> avant et après la dialyse sur gel dénaturant SDS-PAGE. Légende : M : marqueurs de poids moléculaire; CP<sub>15-215</sub> avant : CP<sub>15-215</sub> élution avant la dialyse; CP<sub>15-215</sub> après : CP<sub>15-215</sub> élution après la dialyse de 16h dans 10 mM tris-HCl pH 8.0.

#### 4.2.3 Mutants ponctuels

Puisque le rôle de l'hélice Q18-S23 n'a pu être déterminé à l'aide des mutants de délétion, des mutants ponctuels furent réalisés. Afin de déterminer quelles mutations effectuer, les acides aminés impliqués dans cette hélice présumée furent examinés (voire Figure 30). La présence d'un acide aminé chargé négativement, un acide glutamique (E) à la position 19 fut détectée. Deux mutants de cet acide aminé ont été réalisés dans le contexte de la CP<sub>6-215</sub> : E19K et E19P. L'échange d'un acide glutamique pour une lysine (E19K) fut réalisé dans le but de vérifier si la charge négative présente à cet endroit était importante ou non. Le second mutant ponctuel codait pour une proline à la position 19 au lieu d'un acide glutamique (E19P); la proline étant reconnue comme un des acides aminés briseurs d'hélices.

#### 4.2.4 Prédiction de structure secondaire pour E19K et E19P

Afin de vérifier l'impact de ces deux mutations sur la présence de l'hélice putative Q18-S23, des prédictions de structure secondaire furent réalisées pour les deux mutants ponctuels et comparées à celle obtenue pour la CP<sub>6-215</sub>. Les résultats sont présentés à la Figure 33.

AA	10	20	30	40	50	60
Séquence	MASTPNIAFP	AITQEQMSS	IKVDPTS	NLLPSQEQL	KSVSTLMV	AAKVPAA
CP <sub>6-215</sub>	CCCCCCCC	CCCCCH	HHHHh	hCCCCccc	CCCCCH	HHHHHHHHHHHHHHHHh
E19K	CCCCCCCC	CCCCCH	HHHHH	hCCCCccc	CCCCCH	HHHHHHHHHHHHHHh
E19P	CCCCCCCC	CCCCCH	HHHHH	hCCCCccc	CCCCCH	HHHHHHHHHHHHh

Figure 33. Prédiction de structure secondaire à l'extrémité N-terminale pour les PapMV CP<sub>6-215</sub>, E19K et E19P. Trois algorithmes différents furent utilisés pour chaque clone : PROFsec, PSIPREP et Sable-2 et le consensus de ces trois prédictions est présenté ici. Le *C* représente une région non structurée et le *H* représente une région impliquée dans une structure en hélice-alpha. Les lettres majuscules indiquent des prédictions avec des niveaux de confiance élevés et les lettres minuscules, des niveaux de confiance plus faibles. La totalité des prédictions se trouve à l'Annexe 5.

Les résultats obtenus pour les prédictions de structure secondaire furent surprenants. Tout d'abord, le changement de la charge négative à la position 19 pour une charge positive ne semble pas avoir d'impact sur la formation de l'hélice putative. Ensuite, l'emploi de la proline à cette même position ne semble pas non plus avoir d'effet sur cette hélice.

#### 4.2.5 Production et purification des mutants de délétion

Malgré les résultats obtenus avec les prédictions de structure secondaire, les mutants E19K et E19P furent produits dans *E. coli* et purifiés selon le protocole décrit aux sections 2.2 et 2.3. Les protéines induites ainsi que celles éluées ont été vérifiées sur un gel dénaturant SDS-PAGE (Figure 34). De bonnes quantités de protéines E19K et E19P peuvent être produites et purifiées sans problème. De plus, ces protéines peuvent aussi être dialysées 16 heures dans du tampon 10 mM de Tris-HCl pH 8.0 sans causer d'agrégations non spécifiques.

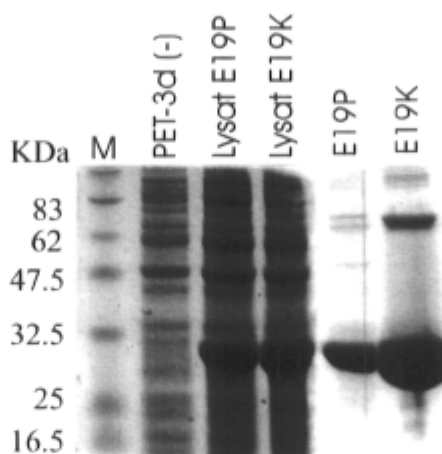


Figure 34. Profil de purification des protéines E19K et E19P. Gel dénaturant SDS-PAGE pour les deux protéines. Légende : M : marqueurs de poids moléculaire; pET-3d (-) : lysat bactérien avant induction; Lysat E19K ou E19P : lysat bactérien après l'induction à l'IPTG 16h à 22 °C; E19K ou E19P : protéines dialysées à la suite d'une élution des protéines avec 1 M imidazole.

#### 4.2.6 Tamisage moléculaire de E19K et E19P

Afin de connaître le profil de multimérisation des protéines purifiées, E19K et E19P furent soumises à du tamisage moléculaire sur une colonne Superdex™ 200 16/60 et comparée avec la protéine CP<sub>6-215</sub>. Le profil de multimérisation obtenu par tamisage moléculaire pour E19K ressemble beaucoup à celui obtenu pour la protéine de comparaison CP<sub>6-215</sub> (Figure 35A). Par contre, la quantité de protéines se trouvant sous une forme multimérique inférieure à 232 kDa est supérieure pour la protéine E19K, diminuant ainsi la quantité de disques et de multimères de plus de 670 kDa formés. Pour ce qui est de E19P, le profil obtenu est complètement différent : la majorité des protéines soumises au tamisage moléculaire se trouvait sous une forme multimérique supérieure à 670 kDa (Figure 35B).

#### 4.2.7 Microscopie électronique de E19K et E19P

Afin de vérifier si les mutations introduites pour l'acide aminé 19 affectaient la capacité des protéines à former des NLPs, des grilles de microscopie furent réalisées tel que décrit à la section 2.6 pour les protéines E19K et E19P. Les résultats sont présentés à la Figure 36. Afin de mieux caractériser les différentes protéines, la longueur moyenne des pseudovirions fut évaluée au microscope électronique en mesurant 200 NLPs pour chaque mutant (Figure 36D).

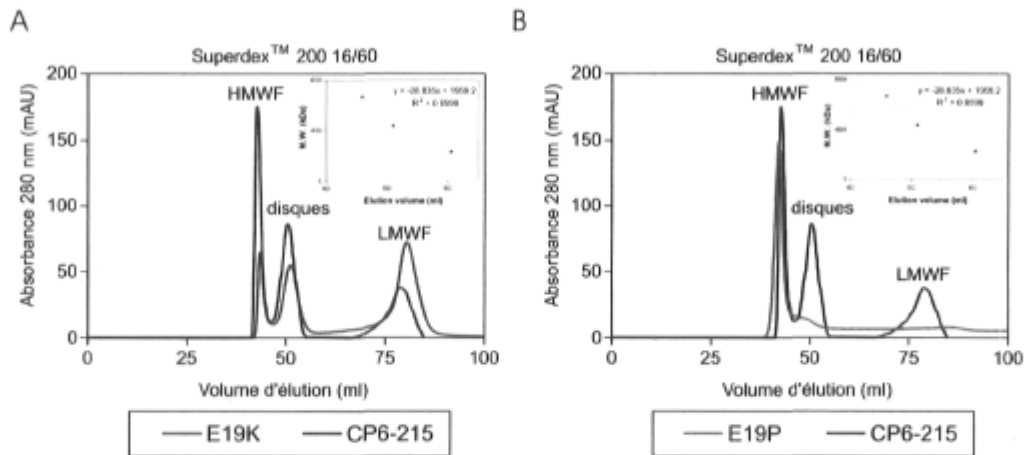


Figure 35. Tamisage moléculaire sur colonne Superdex™ 200 16/60 A) de la protéine E19K (courbe rouge) ou B) de la protéine E19P (courbe verte) en comparaison avec la protéine CP<sub>6-215</sub> (courbe bleue). Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'éluion pour cette colonne a été placé dans le coin supérieur droit. Légende : HMWF : multimères de poids moléculaire supérieur à 670 kDa; disques : disques de 20 sous-unités de la CP de 450 kDa; LMWF : forme monomérique ou multimérique inférieure à 232 kDa.

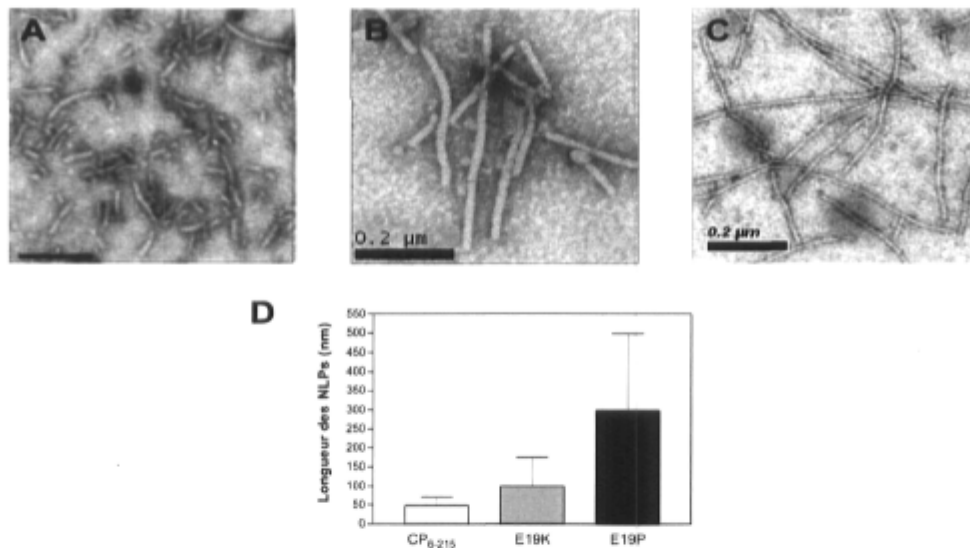


Figure 36. Microscopie électronique des protéines A) CP<sub>6-215</sub>; B) E19K, C) E19P. La barre représente 200 nm. D) Moyenne des longueurs des particules virales de CP<sub>6-215</sub>, de E19K et de E19P. 200 NLPs ont servi à faire les calculs.

Étonnamment, la longueur moyenne des NLPs pour E19K et E19P était deux fois et six fois supérieure à celle obtenue pour la CP<sub>6-215</sub> avec des valeurs de 100 nm, 300 nm et 50 nm respectivement.

### 4.3 Discussion

Quatre mutants furent produits au cours de cette partie du projet de recherche. Tous avaient pour but de déterminer l'importance pour la multimérisation de l'hélice Q18-S23 prédite qui pourrait être impliquée dès les premières étapes de l'assemblage du PapMV. Les deux mutants de délétion produits n'ont permis de produire une protéine stable.

#### *Les mutants ponctuels*

C'est donc dans l'optique de vérifier la présence d'un pont salin chez le PapMV que des mutants ponctuels furent produits. La première mutation qui a été effectuée était E19K où la charge négative apportée par l'acide glutamique était remplacée par une charge positive présente chez la lysine. Il est intéressant de constater que chez deux des 19 *Potexvirus*, c'est une lysine qui est présente à la position 19 (Figure 29, ScaVX et LVX). Ce mutant devait nous permettre de savoir si la charge négative à l'extrémité N-terminale est essentielle à la multimérisation en participant à des interactions électrostatiques. Cependant, les résultats obtenus pour E19K semblent plutôt infirmer l'hypothèse selon laquelle l'acide glutamique à la position 19 de la PapMV CP serait impliqué dans un pont ionique essentiel à la multimérisation. Malgré cette mutation, les protéines de capsid conservent leur capacité à se multimériser en formant des disques et des NLPs. Par contre, tel qu'illustré à la Figure 35A, la mutation semble tout de même légèrement affecter la capacité de la CP à se multimériser puisque qu'une plus faible proportion de CPs se trouvent sous forme multimérique supérieure à 450 kDa en comparaison avec la CP<sub>6-215</sub>. Par contre, les NLPs, une fois formées, étaient en moyenne deux fois plus longues que celles obtenues avec la CP<sub>6-215</sub> (Figure 36D). Il serait donc possible que l'acide aminé E19 soit impliqué dans un pont ionique apportant une stabilité supplémentaire aux multimères mais non essentielle à la formation de particules tout comme c'est le cas pour le pont ionique D83-R159 du PhMV.

Ensuite, nous avons voulu vérifier l'importance de l'hélice Q18-S23 prédite en remplaçant l'acide glutamique de la position 19 par une proline (E19P). Les prolines sont des acides aminés incompatibles avec les structures en hélice-alpha. En effet, elles sont souvent présentes au début ou à la fin de ces hélices. Chez les prolines, tel qu'illustré à la Figure 37,



la chaîne latérale est reliée à la chaîne principale se qui créé ainsi une torsion de la chaîne principale. C'est donc pour cette raison que les prolines sont souvent retrouvées dans les régions non structurées des protéines (Branden et Tooze 1996). En introduisant une proline à l'intérieur de l'hélice putative Q18-S23, nous pensions pouvoir la détruire et ainsi observer son impact.

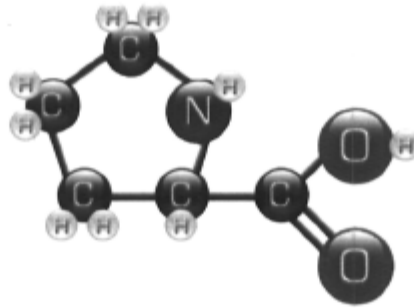


Figure 37. Représentation schématique de la proline. Légende : C : carbone; H : hydrogène; N : azote; O : oxygène. Tiré de <http://upload.wikimedia.org/wikipedia/fr/8/8e/Proline.jpg>.

Contrairement à nos attentes et selon les prédictions de structure secondaire présentées à la Figure 33, l'ajout de la proline à la position 19 ne semble pas affecter la formation de la petite hélice prédite. Nous avons cependant produit quand même ce mutant dans *E.coli* afin d'observer sa capacité à former des pseudovirions. Plus de 98 % de la protéine E19P se trouvaient sous forme de multimères de plus de 670 kDa (Figure 35B). De plus, la longueur moyenne des particules de E19P observées en microscopie électronique était six fois supérieure à celle obtenue pour CP<sub>6-215</sub>, avec des valeurs de 300 et 50 nm respectivement (Figure 36D). Jusqu'à maintenant, nous ne pouvons pas expliquer pourquoi la modification du glutamate 19 en proline apporte un bénéfice pour la formation de particules. Il est possible que cette mutation augmente l'affinité des sous-unités de la CP entre elles se traduisant ainsi par une meilleure capacité à former des NLPs.

En conclusion, les deux mutants ponctuels, tout comme ceux de délétion, n'ont pas permis d'établir l'importance de l'hélice Q18-S23 prédite. Des mutations supplémentaires dans cette région seraient nécessaires pour répondre à cette question. Il serait intéressant de produire des mutants ponctuels ou légèrement délétés qui aboliraient l'hélice prédite.

## 5.0 Étude de l'implication de l'extrémité N-terminale dans la formation de NLPs

### 5.1 Introduction

Récemment, notre équipe a réussi à isoler une forme tronquée de la protéine de coque du PapMV (CP<sub>27-215</sub>) qui se trouve exclusivement sous forme monomérique (Lecours, Tremblay *et al.* 2006). Des études en résonance magnétique nucléaire ont été débutées avec cette protéine. Ces études ont été ralenties à cause de la sensibilité de la protéine à la dégradation (Lecours 2007). Il a été mesuré par des analyses en spectroscopie de masse, que 35 acides aminés situés à l'extrémité C-terminale de la CP<sub>27-215</sub> sont clivés pendant l'acquisition des spectres RMN (Lecours 2007). Le résultat de ce clivage commence à «contaminer» les spectres RMN après seulement deux ou trois jours de collecte de données RMN. Ce problème de stabilité de la protéine monomérique CP<sub>27-215</sub> persiste malgré l'ajout d'inhibiteurs de protéases ou d'azide de sodium à la protéine en solution (Lecours 2007).

Pour pallier à ce problème, il est important d'identifier une autre forme monomérique stable de la protéine de capsid. Les résultats obtenus au cours de la section précédente nous suggèrent que l'hélice-alpha prédite pour les acides aminés 18 à 23 n'est pas impliquée dans la multimérisation à l'extrémité N-terminale. Les protéines tronquées CP<sub>15-215</sub> et CP<sub>21-215</sub> sont sensibles à la dégradation et sont difficiles à étudier. Nous nous intéressons à des formes tronquées plus courtes pour vérifier si nous pouvons augmenter la stabilité des protéines et faciliter entre autre les études en RMN. La prochaine section est divisée en trois sous-sections où chaque sous-section présente un mutant de délétion différent : 1) CP<sub>14-215</sub>; 2) CP<sub>13-215</sub> et 3) CP<sub>14-180</sub>.

## 5.2 Résultats

### 5.2.1 Mutant de délétion CP<sub>14-215</sub>

Le clonage et la production la protéine tronquée CP<sub>14-215</sub> ont été réalisés tel que décrit aux sections 2.1.1 (stratégie # 2) à 2.2. Les protéines induites ont été vérifiées sur gel dénaturant SDS-PAGE (Figure 38).

La purification de la protéine CP<sub>14-215</sub> s'est effectuée selon le protocole décrit à la section 2.3. La protéine purifiée et dialysé a été vérifiée sur gel dénaturant SDS-PAGE (Figure 39). On peut remarquer à la Figure 39 qu'une partie de la protéine purifiée et dialysée a été dégradée (bandes indiquées par une flèche rouge). Une première bande correspondant à une protéine dégradée est apparente à environ 25 kDa et une seconde à moins de 6,5 kDa. Comme la protéine CP<sub>14-215</sub> non dégradée a un poids moléculaire calculé de 22,6 kDa et un poids moléculaire apparent sur gel dénaturant d'environ 30 kDa, il est difficile d'estimer le poids moléculaire réel de la forme dégradée de la CP.

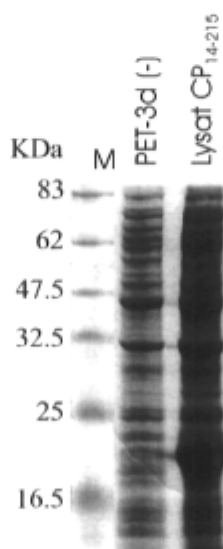


Figure 38. Vérification de l'induction de la CP<sub>14-215</sub> sur gel SDS-PAGE. Légende : M : marqueurs de poids moléculaire; pET-3d (-) : lysat bactérien avant induction; Lysat CP<sub>14-215</sub>: lysat bactérien après l'induction à l'IPTG 16 heures à 22 °C.

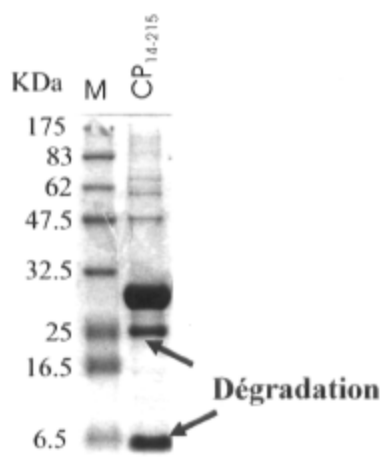


Figure 39. Gel dénaturant SDS-PAGE présentant la protéine CP<sub>14-215</sub> purifiée et dialysée. Légende : M : marqueurs de poids moléculaire; CP<sub>14-215</sub> : CP<sub>14-215</sub> purifiée et dialysée. Les flèches rouges identifient les bandes de dégradation.

#### *Microscopie électronique*

Afin de vérifier la capacité de ce mutant à former des NLPs, des grilles de microscopie électronique furent réalisées avec la protéine CP<sub>14-215</sub> dialysée colorées à l'uranyle acétate 2 %. La présence de NLPs n'a pu être détectée sur aucune des trois grilles préparées (données non présentées ici).

#### *Tamissage moléculaire*

Afin de connaître le profil de multimérisation de la protéine CP<sub>14-215</sub>, un tamissage moléculaire fut utilisé. Puisque que la protéine semblait incapable de former des NLPs, la colonne Superdex<sup>TM</sup> 75 26/60 séparant les protéines ayant des poids moléculaires variant entre 13,7 et 67 kDa fut utilisée (Figure 40).

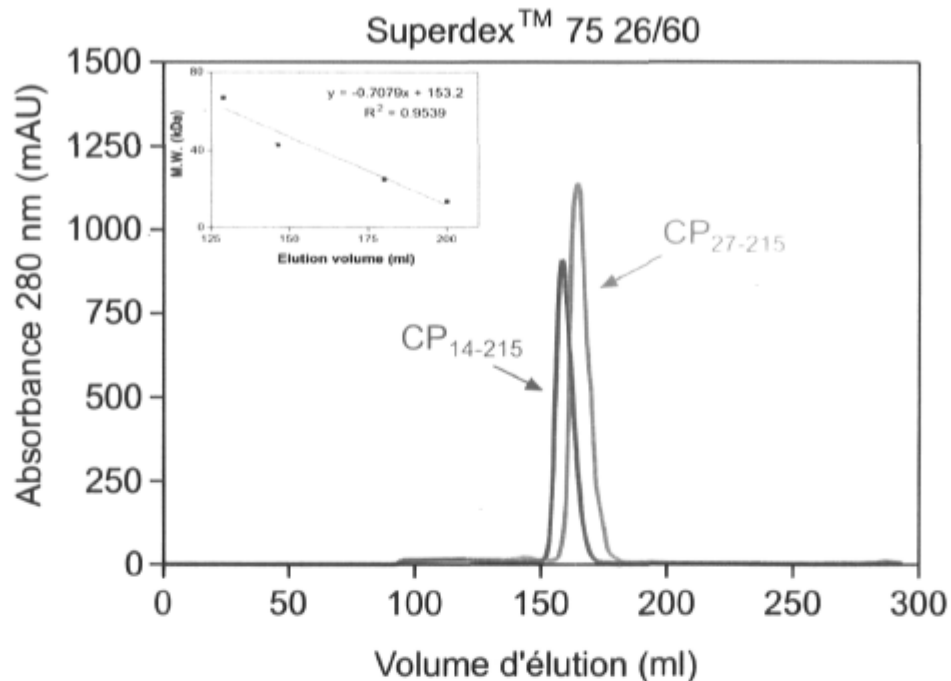


Figure 40. Tamisage moléculaire sur colonne Superdex™ 75 26/60 de la protéine CP<sub>14-215</sub> (courbe turquoise) en comparaison avec le monomère CP<sub>27-215</sub> (courbe orange). Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'élution pour cette colonne a été placé dans le coin supérieur gauche.

La protéine CP<sub>14-215</sub>, tout comme la CP<sub>27-215</sub>, est produite exclusivement sous forme monomérique. Aucun disque de 450 kDa ne fût détecté. En effet, si une portion de la protéine avait été présente sous forme de disques, un pic d'absorbance aurait été visible aux environs de 108 ml puisque ce volume correspond au volume d'exclusion de la colonne. De plus, cette protéine est incapable de former des dimères en solution puisque, sur cette colonne, les formes dimérique et monomérique sont facilement séparables. Il est normal que la protéine CP<sub>14-215</sub> soit éluee avant la CP<sub>27-215</sub> puisqu'elle possède 13 acides aminés de plus, ce qui fait augmenter son poids moléculaire.

#### *Liaison à l'ARN*

La capacité de la protéine CP<sub>14-215</sub> à se lier à l'ARN a été comparée à celle obtenue pour la protéine CP<sub>27-215</sub>. Il a été publié récemment que la protéine monomérique CP<sub>27-215</sub> était incapable d'interagir *in vitro* avec des acides nucléiques (Lecours, Tremblay *et al.* 2006) (voire section 1.7.1). L'utilisation croissante de protéines (passant de 0 ng à 1500 ng) avec

une quantité fixe (165 fmol) de sonde (correspondant aux 80 premiers nucléotides de l'ARN viral du PapMV) marquée au  $\gamma$ - $^{32}\text{P}$  n'a permis d'observer aucune interaction entre l'ARN et la protéine CP<sub>14-215</sub> (Figure 41). Les résultats obtenus en EMSA pour CP<sub>27-215</sub> sont donc applicables à CP<sub>14-215</sub>.

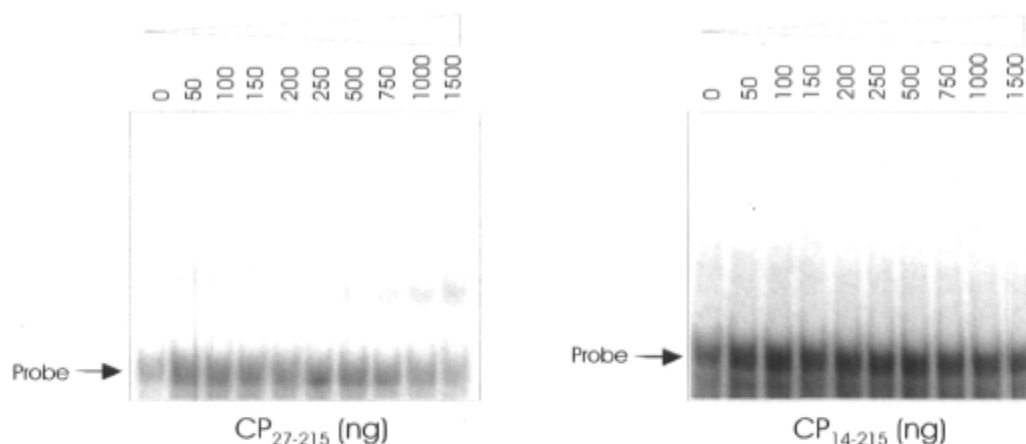


Figure 41. EMSA de la protéine CP<sub>14-215</sub> en comparaison avec celui de la protéine monomérique CP<sub>27-215</sub>. Des quantités croissantes de protéines ont été incubées avec une quantité constante de sonde d'ARN marquée au  $\gamma$ - $^{32}\text{P}$ . Les chiffres dans le haut du gel représentent les quantités croissantes de protéines incubées avec la sonde. Légende : Probe : sonde d'ARN marqué non lié aux protéines.

#### *$^{15}\text{N}$ -HSQC des PapMV CP<sub>14-215</sub> et CP<sub>27-215</sub>*

Afin de comparer les différences de structure entre les deux formes monomérique de la PapMV CP, les deux protéines furent marquées au  $^{15}\text{N}$  par Katia Lecours. Tout d'abord, les conditions optimales pour la RMN furent établies avec la protéine CP<sub>27-215</sub> (Lecours 2007). Ces conditions sont décrites à la section 2.13. Des échantillons RMN possédant les conditions optimisées pour la CP<sub>27-215</sub> ont été préparés. À l'aide de ces échantillons, des spectres 2D  $^{15}\text{N}$ -HSQC TROSY furent enregistrés à la première journée de leur préparation et deux semaines plus tard. Les paramètres d'acquisition ont été décrits préalablement dans la section 2.12. Pour CP<sub>27-215</sub>, le nombre de pics dénombrés dans le HSQC du premier jour correspond au nombre attendu avec quelques pics supplémentaires. Cependant, ce nombre est grandement augmenté après deux semaines. Cette croissance du nombre de pics détectés en fonction du temps est un signe de dégradation de la protéine.

Les résultats obtenus avec la protéine CP<sub>14-215</sub> sont comparables à ceux de CP<sub>27-215</sub>, mais avec un nombre de pics plus élevés qui correspondent aux treize acides aminés supplémentaires. La superposition des spectres enregistrés pour les deux protéines est présentée à la Figure 42. Il importe de mentionner que tous les pics correspondant aux régions structurées pour la CP<sub>27-215</sub> sont aussi présents pour CP<sub>14-215</sub> et qu'ils sont situés aux mêmes déplacements chimiques indiquant que la structure est très semblable pour les deux protéines. Finalement, le spectre obtenu pour CP<sub>14-215</sub> après deux semaines montre des signes de dégradation similaires à CP<sub>27-215</sub>.

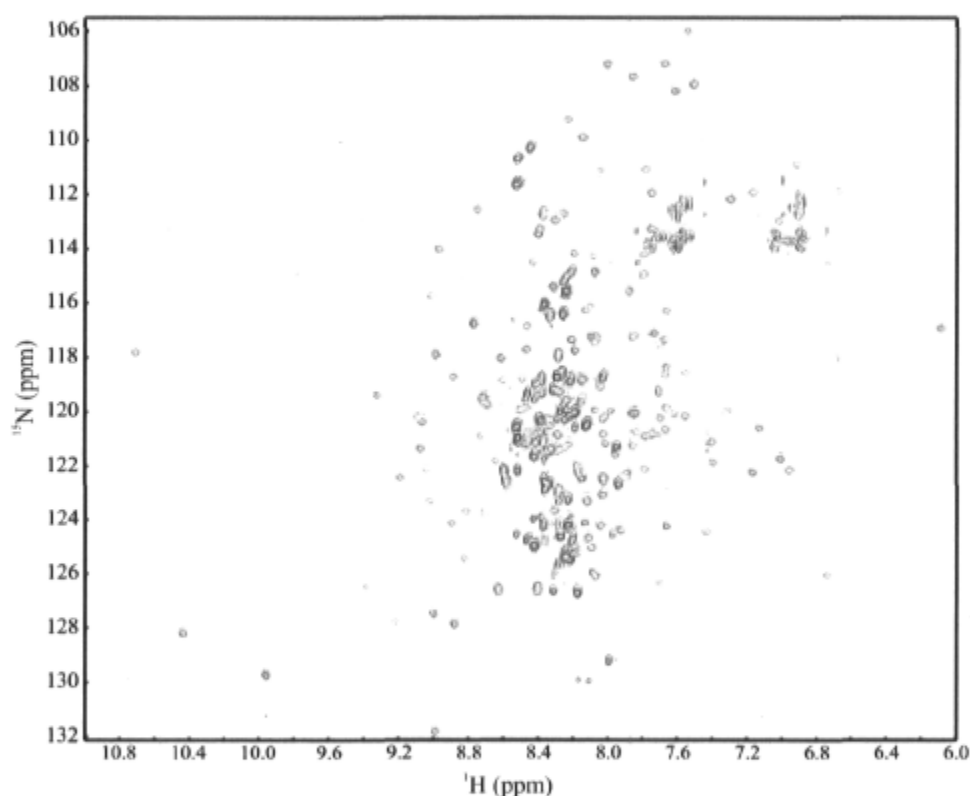


Figure 42. <sup>15</sup>N-HSQC de la PapMV CP<sub>27-215</sub> (en rouge) et CP<sub>14-215</sub> (en bleu) suivant les paramètres détaillés à la section 2.13. Tiré du mémoire de Katia Lecours (Lecours 2007).

### 5.2.2 Mutant de délétion CP<sub>13-215</sub>

Les résultats obtenus à la section précédente indiquent que la suppression des treize premiers acides aminés à l'extrémité N-terminale conduit à une forme monomérique instable. Nous avons donc vérifié les aptitudes de multimérisation d'une forme mutante de la PapMV CP comportant un acide aminé de plus à l'extrémité N-terminale (CP<sub>13-215</sub>). Le

clonage, la production et la purification de ce mutant ont été réalisés tel que décrit aux sections 2.1.1 (stratégie # 2) à 2.3. Les protéines induites et celles éluées ont été vérifiées sur gel dénaturant SDS-PAGE (Figure 43).

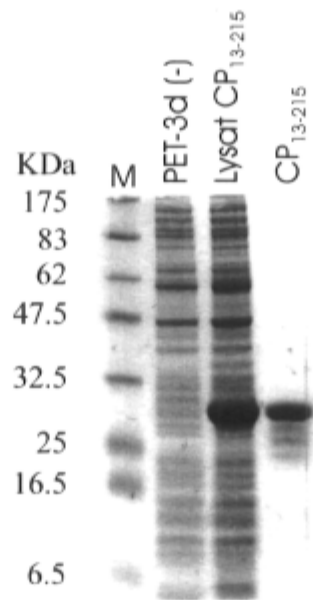


Figure 43. Vérification de l'induction et de l'éluion de la CP<sub>14-215</sub> sur gel SDS-PAGE. Légende : M : marqueurs de poids moléculaire; pET-3d (-) : lysat bactérien avant induction; Lysat CP<sub>14-215</sub>: lysat bactérien après l'induction à l'IPTG 16 heures à 22 °C; CP<sub>13-215</sub> : CP<sub>13-215</sub> purifiée et dialysée.

Immédiatement après l'éluion et la dialyse de la protéine CP<sub>13-215</sub>, on peut remarquer qu'une faible proportion de la protéine se trouve sous des formes dégradées (Figure 44, ligne CP<sub>13-215</sub>). La dégradation observée pour CP<sub>13-215</sub> n'est pas comparable à celle obtenue pour le monomère CP<sub>14-215</sub> (Figure 39). En effet, il semblerait que la dégradation observée pour CP<sub>13-215</sub> ait affectée une plus faible quantité de protéines que pour CP<sub>14-215</sub>. De plus, la dégradation de CP<sub>13-215</sub> semble survenir de façon plus variable, puisqu'au moins trois bandes mineures de dégradation peuvent être détectées entre les marqueurs de poids moléculaires 16,5 et 32,5 kDa alors qu'une seule bande bien définie n'était visible à la Figure 39.



### Microscopie électronique

Afin de vérifier la capacité de ce mutant à former des NLPs, des grilles de microscopie électronique furent réalisées avec la protéine CP<sub>13-215</sub> dialysée et colorées à l'uranyle acétate 2 % (Figure 44).

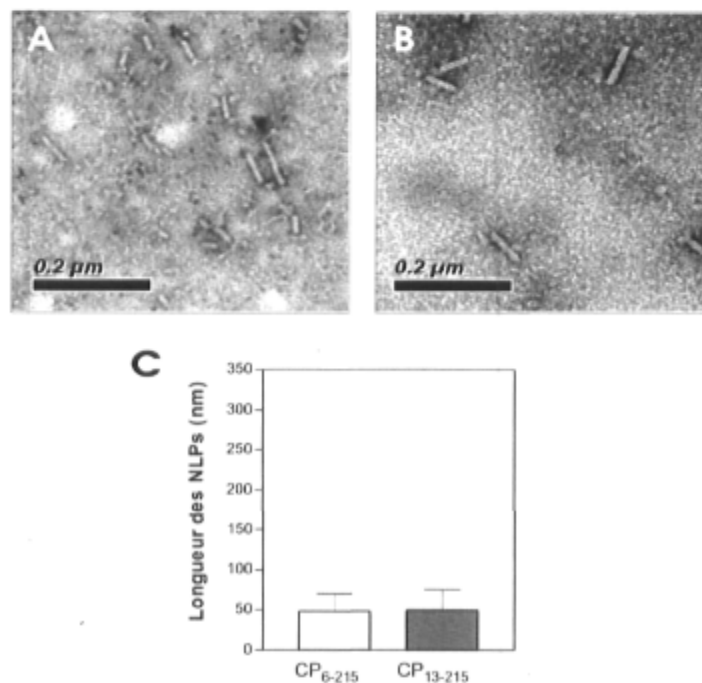


Figure 44. Comparaison de la microscopie électronique des protéines A) CP<sub>6-215</sub> et B) CP<sub>12-215</sub>. La barre représente 200 nm. C) Moyenne des longueurs des particules virales de CP<sub>6-215</sub> et de CP<sub>12-215</sub>. 250 NLPs ont servi à faire les calculs.

Les NLPs observées pour CP<sub>13-215</sub> sont de longueurs similaires à celles obtenues pour CP<sub>6-215</sub>. 250 NLPs par mutant furent mesurées afin d'établir la longueur moyenne des NLPs formées à partir des deux protéines et des valeurs moyennes d'environ 50 nm ont été obtenues pour les deux mutants. Aucune différence particulière entre les NLPs n'a pu être observée en microscopie électronique.

### Tamissage moléculaire

Afin de connaître le profil de multimérisation de la protéine CP<sub>13-215</sub>, un tamisage moléculaire fut utilisé. Puisque que la protéine possède la capacité de former des NLPs, la

colonne Superdex™ 200 16/60 fut utilisée afin de vérifier la présence de disques de 450 kDa et de multimères de poids moléculaires inférieurs à 232 kDa (Figure 45).

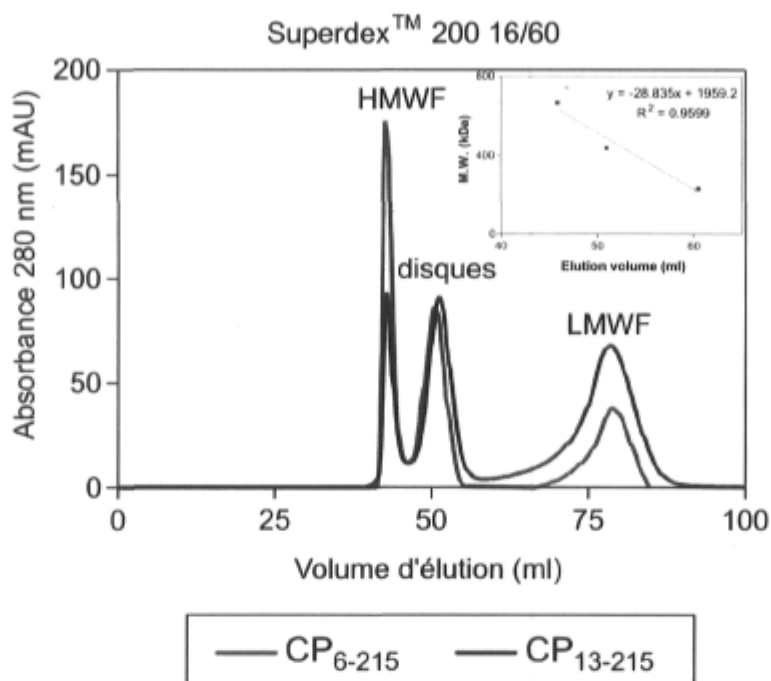


Figure 45. Tamisage moléculaire sur colonne Superdex™ 200 16/60 de la protéine CP<sub>13-215</sub> (courbe bleue) en comparaison avec la CP<sub>6-215</sub> (courbe rouge). Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'éluion pour cette colonne a été placé dans le coin supérieur droit. Légende : HMWF : multimères de poids moléculaire supérieur à 670 kDa; disques : disques de 20 sous-unités de la CP de 450 kDa; LMWF : forme monomérique ou multimérique inférieure à 232 kDa.

La protéine CP<sub>13-215</sub> a été séparée sur la colonne en trois populations distinctes, tout comme la protéine de référence CP<sub>6-215</sub> : multimères de plus de 670 kDa, multimères d'environ 450 kDa et finalement multimères de moins de 232 kDa. Malgré le fait que la délétion des douze premiers acides aminés n'empêche pas l'assemblage de se faire, ce dernier semble toutefois un peu perturbé. En effet, on peut remarquer que la proportion des protéines comprenant les NLPs est plus faible par rapport à la protéine de référence. La proportion des protéines sous forme de disques ne semble toutefois pas affectée. Par contre, la quantité de protéines se trouvant sous une forme multimérique inférieure à 232 kDa est nettement supérieure pour la protéine CP<sub>13-215</sub>.

### Liaison à l'ARN

Afin de vérifier si la proportion réduite en NLPs de CP<sub>13-215</sub> par rapport à CP<sub>5-215</sub> est due à une affinité plus faible pour les acides nucléiques, un EMSA fut réalisé. Nous avons comparé l'affinité des protéines se trouvant dans le surnageant de l'ultracentrifugation, contenant les disques (multimère de 20 sous-unités) et les formes multimériques de poids moléculaires moindre, pour les deux protéines. Des quantités croissantes de CP<sub>13-215</sub> et CP<sub>6-215</sub> (servant de référence) ont été incubées avec 165 fmol d'une sonde d'ARN marqué au  $\gamma$ -<sup>32</sup>P. Les complexes ont été déposés sur un gel de polyacrylamide en condition native. Le signal radioactif a été révélé utilisant un écran au phosphore et un «phosphoimager» (Typhoon). Les résultats obtenus pour les deux protéines sont présentés à la Figure 46. Aucune différence significative n'a pu être décelée entre les profils des deux protéines pour des quantités de protéines supérieures à 50 ng. Ces résultats suggèrent donc que la liaison de la protéine de capsid à l'ARN n'est pas affectée par la délétion des douze premiers acides aminés.

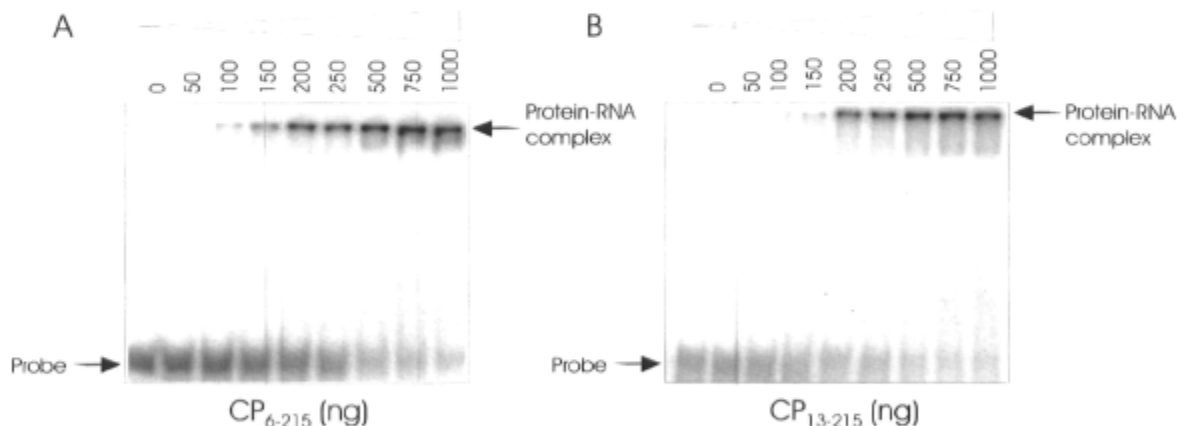


Figure 46. EMSA A) de la protéine CP<sub>13-215</sub> en comparaison avec celui B) de la protéine de référence CP<sub>6-215</sub>. Des quantités croissantes de protéines ont été incubées avec une quantité constante de sonde d'ARN marquée au  $\gamma$ -<sup>32</sup>P. Les chiffres dans le haut du gel représentent les quantités croissantes de protéines incubées avec la sonde. Légende : Probe : sonde d'ARN marqué non lié aux protéines ; Protein:RNA complex : protéines liées aux acides nucléiques marqués.

### Spectroscopie en dichroïsme circulaire

Afin de vérifier la structure secondaire de CP<sub>13-215</sub>, les NLPs récupérées dans le culot de l'ultracentrifugation (voire section 2.5) furent soumises à un test en spectroscopie de dichroïsme circulaire (Figure 47).

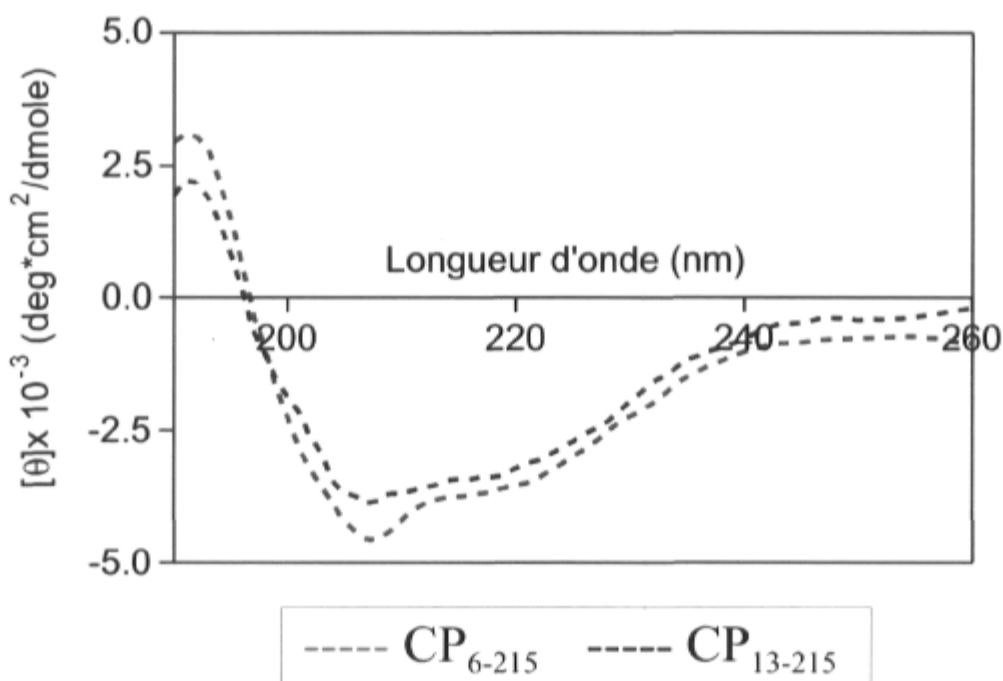


Figure 47. Comparaison du spectre CD des NLPs de la protéine CP<sub>13-215</sub> (courbe bleue) à celui des NLPs de la protéine CP<sub>6-215</sub> (courbe orange).

On peut remarquer que les deux spectres CD sont très similaires. Les deux courbes présentent un minimum à 207 nm et un autre à 222 nm. Cependant, le pic à 207 nm du spectre CD de la protéine CP<sub>13-215</sub> semble moins prononcé que celui obtenu pour la protéine de comparaison. Par contre, les prédictions de la structure secondaire pour les deux protéines ne révèlent aucune différence significative de structure secondaire entre elles (Annexe 5).

#### 5.2.3 Mutant tronqué CP<sub>14-180</sub>

L'utilisation de la protéine CP<sub>14-215</sub> en RMN par Katia Lecours présente aussi des problèmes de dégradation semblable à ceux obtenus pour la protéine CP<sub>27-215</sub>. Une analyse

en spectroscopie de masse effectuée par Katia Lecours sur la forme dégradée de la CP<sub>27-215</sub> a permis d'identifier l'emplacement exact du clivage (Lecours 2007). En effet, le clivage aurait lieu à l'extrémité C-terminale de la protéine et enlèverait les 35 derniers résidus. La forme CP<sub>14-180</sub> a donc été clonée et la protéine a été exprimée. Le clonage de ce mutant est décrit à la section 2.1.1 à la stratégie # 3. Ce clone fut exprimé dans *E. coli* et purifié selon le protocole décrit aux sections 2.2 et 2.3 (Figure 48).

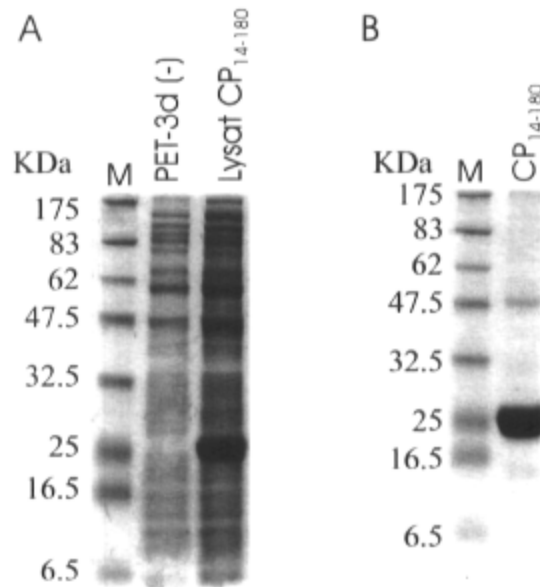


Figure 48. Expression et purification de la protéine CP<sub>14-180</sub> sur gel dénaturant SDS-PAGE. A) Vérification de l'induction de la CP<sub>14-180</sub>. Légende : M : marqueurs de poids moléculaire; pET-3d (-) : lysat bactérien avant induction; Lysat CP<sub>14-180</sub> : lysat bactérien après l'induction à l'IPTG 16 heures à 22 °C. B) Vérification de l'expression de la protéine CP<sub>14-180</sub> purifiée et dialysée. Légende : M : marqueurs de poids moléculaire; CP<sub>14-180</sub> : CP<sub>14-180</sub> purifiée et dialysée.

La protéine CP<sub>14-180</sub> a été purifiée et dialysée sans causer de la dégradation, tel qu'il fut observé pour la protéine CP<sub>14-215</sub> (Figure 39). De plus, le niveau de pureté du double mutant purifié est supérieur à celui observé pour CP<sub>14-215</sub>. Cette protéine se retrouve à une masse moléculaire de 25 kDa lorsque soumise à un gel SDS-PAGE, ce qui correspond à la forme dégradée de CP<sub>14-215</sub> (Figure 39).

### *Microscopie électronique*

Malgré le fait que la protéine CP<sub>14-215</sub> soit incapable de former des NLPs, nous avons voulu vérifier quand même que cette incapacité était aussi présente chez le double mutant CP<sub>14-180</sub>. Des grilles de microscopie électronique furent réalisées avec la protéine CP<sub>14-180</sub> dialysée et furent colorées à l'uranyle acétate 2 %. La présence de NLPs n'a pu être détectée sur aucune des trois grilles préparées (données non montrées ici).

### *Tamissage moléculaire*

Afin de connaître le profil de multimérisation de la protéine CP<sub>14-180</sub>, un tamissage moléculaire fut réalisé. Puisque que la protéine est sûrement produite uniquement sous forme monomérique, la colonne Superdex<sup>TM</sup> 75 26/60 séparant les protéines ayant des poids moléculaires variant entre 13,7 et 67 kDa fut utilisée (Figure 49). Cependant, la protéine fut entreposée à 4 °C pendant 4 semaines avant d'être soumise au tamissage moléculaire. Puisque nous avons enlevé les extrémités qui devaient être clivées lors de l'entreposage, nous avons pensé que la stabilité de cette protéine était une chose acquise. Cependant, les résultats obtenus pour le tamissage moléculaire nous prouvent le contraire. En effet, on peut remarquer à la Figure 49 que le pic d'élution de la colonne est très large, s'étendant sur plus de 30 ml. De plus, en regardant attentivement cette région, on peut remarquer la présence de deux pics se chevauchant dont les sommets seraient situés aux alentours de 176 ml et de 181 ml. Cela suggère donc que la protéine CP<sub>14-180</sub> se trouve sous deux formes différentes de poids moléculaires légèrement différents. Donc, la protéine serait donc encore sensible au clivage lorsque entreposée quelque temps.

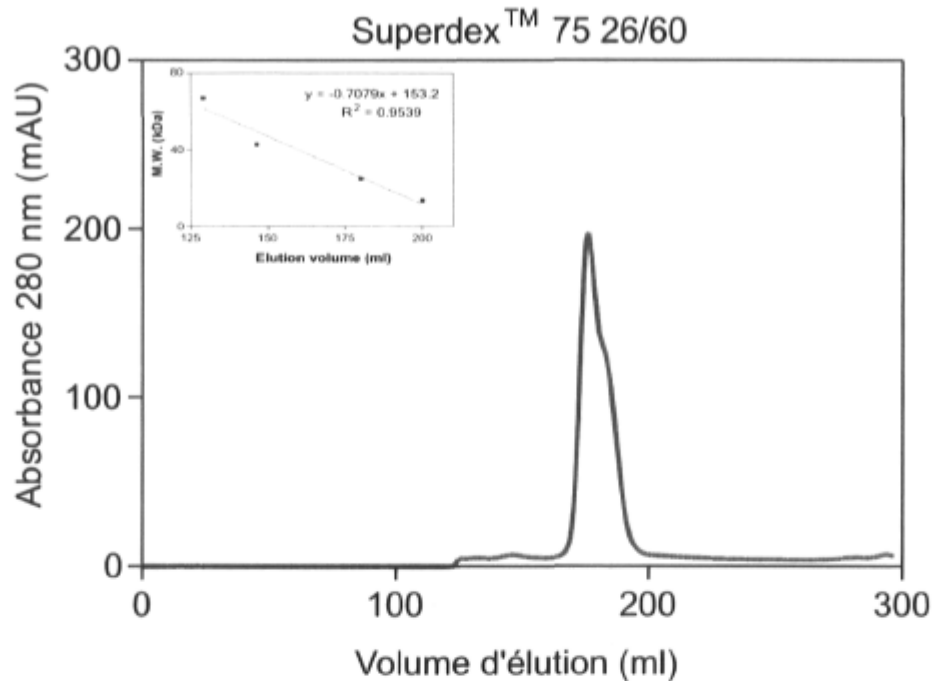


Figure 49. Tamisage moléculaire sur colonne Superdex™ 75 26/60 de la protéine CP<sub>14-180</sub> (courbe mauve). Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'élution pour cette colonne a été placé dans le coin supérieur gauche.

Comme les deux populations protéiques de CP<sub>14-180</sub> sont éluées à des volumes très similaires, il est pratiquement impossible de les séparer en vue d'analyser la protéine clivée par spectroscopie de masse, comme il a été fait avec la CP<sub>27-215</sub> par Katia Lecours (Lecours 2007). Nous avons donc décidé de refaire la même procédure mais avec le double mutant de délétion CP<sub>27-180</sub>, en supposant que la dégradation observée pour CP<sub>14-180</sub> se produit à l'extrémité N-terminale. Les analyses en spectroscopie de masse réalisées avec la CP<sub>27-215</sub> montrent que la dégradation survient à un endroit précis à l'extrémité C-terminale. Nous pouvons donc penser que la dégradation observée pour CP<sub>13-180</sub> est due aux treize acides aminés supplémentaires à l'extrémité N-terminale. Le protocole de clonage de ce nouveau mutant est décrit à la section 2.1.1, stratégie # 3. Cependant, des problèmes de clonage sont survenus à ce moment et le clone sans mutation non désirée n'a pu être obtenu que tout récemment. Le temps nous étant compté, la nouvelle protéine CP<sub>27-180</sub> n'a pas pu être exprimée dans *E. coli*.

### 5.3 Discussion

#### *Les mutants de délétion CP<sub>13-215</sub> et CP<sub>14-215</sub>*

Cette section du projet fut très importante puisqu'elle a permis d'identifier plus précisément la région impliquée dans la multimérisation à l'extrémité N-terminale. Nous avons pu identifier exactement le nombre d'acides aminés essentiels à l'extrémité N-terminale pour permettre la multimérisation. Les douze premiers acides aminés ne sont pas nécessaires à la multimérisation, car ils peuvent être enlevés sans pour autant empêcher la formation de disques et de NLPs. Cependant, il est possible qu'ils jouent un rôle dans la multimérisation et dans l'assemblage puisque la comparaison des profils de multimérisation des protéines CP<sub>6-215</sub> et CP<sub>13-215</sub> relève une différence d'efficacité de multimérisation (Figure 45).

La proportion de multimères de poids moléculaires inférieurs à 232 kDa est nettement plus importante pour CP<sub>13-215</sub>, alors que celle des multimères plus gros que 670 kDa est inférieure. La proportion de disques de 450 kDa est identique pour les deux protéines. La réduction de la quantité de NLPs formés ne peut pas être expliquée par la diminution de l'affinité des disques pour les acides nucléiques comme en témoigne la Figure 46. Cela suggère donc que les douze premiers acides aminés pourraient être impliqués dans les interactions protéines:protéines donc, à la première étape de l'assemblage qui conduit à l'obtention de disques ainsi qu'à la dernière étape d'élongation qui mène à la formation de particules virales matures. Les NLPs formés sont très semblables à celles de la protéine de référence CP<sub>6-215</sub> (Figure 44). La raison de la réduction de l'efficacité de l'assemblage pour la protéine CP<sub>13-215</sub> demeure toujours inconnue.

Lorsque les treize premiers acides aminés à l'extrémité N-terminale sont enlevés, la protéine de capsid perd sa capacité à se multimériser. Le tamisage moléculaire de la CP<sub>14-215</sub> (Figure 40) démontre bien que la totalité de la protéine est produite sous forme monomérique. Tout comme CP<sub>27-215</sub>, la CP<sub>14-215</sub> est incapable de se lier aux acides nucléiques correspondant aux 80 premiers nucléotides de l'ARN du PapMV (Figure 41). Cependant, cette CP est très sensible à la dégradation, car une certaine proportion de la protéine fraîchement purifiée et dialysée se retrouve sous une forme dégradée. Il a été



suggéré que cette dégradation se déroulait de façon similaire à celle observée pour CP<sub>27-215</sub> pendant les tests RMN (Lecours 2007) et qu'elle clivait aussi les 35 derniers acides aminés.

Afin de vérifier l'impact des 13 acides aminés supplémentaires sur la structure de la protéine monomérique, des spectres RMN furent enregistrés pour la CP<sub>14-215</sub>. Selon les prédictions de structure secondaire, une différence devrait être observée entre les spectres RMN des protéines CP<sub>14-215</sub> et CP<sub>27-215</sub>, puisque une hélice-alpha de six acides aminés est prédite pour les résidus 18 à 22 et que ces résidus sont absents pour la CP<sub>27-215</sub> (Annexe 5). La comparaison des spectres RMN des deux protéines semble plutôt infirmer cette hypothèse. On remarque qu'il y a plus de pics présents au centre du spectre, ce qui correspond à une région non structurée (Lecours 2007). Tous les pics correspondant aux régions structurées pour la CP<sub>14-215</sub> sont aussi présents pour la CP<sub>27-215</sub>. De plus, ces pics sont aux mêmes déplacements chimiques ce qui indique que la structure est très semblable pour les deux protéines. Ces résultats nous amènent à supposer que les acides aminés 13 à 26 ne seraient pas structurés contrairement à ce qui est suggéré par les prédictions de structure secondaire pour les résidus Q18 à S23. Évidemment, les résultats préliminaires en RMN sont plus fiables que ceux obtenus pour les prédictions de structure secondaire, qui ne sont fiables qu'à 75 %.

#### *Les doubles mutants de délétion CP<sub>14-180</sub> et CP<sub>27-180</sub>*

Étant donné la sensibilité à la dégradation des protéines monomériques CP<sub>14-215</sub> et CP<sub>27-215</sub>, le double mutant CP<sub>14-180</sub> semblait bien prometteur. Cependant, après un entreposage de la protéine pour un mois à 4 °C, nous avons pu observer à nouveau de la dégradation. Par contre, la portion dégradée doit compter peu d'acides aminés puisqu'il fut impossible de séparer les protéines dégradées de celles intactes par tamisage moléculaire à l'aide d'une colonne Superdex<sup>TM</sup> 75 26/60. En raison de ce résultat, les recherches sur ce mutant furent arrêtées et une méthode alternative fut recherchée.

Puisque les résultats obtenus par Katia Lecours avec la forme dégradée de CP<sub>27-215</sub> semblent indiquer que la dégradation a lieu à l'extrémité C-terminale (Lecours 2007), nous avons donc décidé de produire une protéine tronquée des 26 acides aminés en N-terminal et

de 35 acides aminés en C-terminal. En raison de nombreux problèmes rencontrés au cours du clonage du gène du double mutant, la protéine CP<sub>27-180</sub> n'a pas pu être exprimée dans *E. coli*. Plusieurs résultats obtenus par Katia Lecours (Lecours 2007) nous portent à croire que cette nouvelle protéine sera résistante à la dégradation. Si tel est le cas, cette protéine pourra être utilisée plus facilement que CP<sub>27-215</sub> pour des études en RMN. Selon les prédictions de structure secondaire (Figure 17), les quatre derniers résidus d'une hélice-alpha de 21 acides aminés prédite pour les acides aminés 165 à 185 ainsi qu'une hélice-alpha de 9 acides aminés prédite pour les résidus 192 à 200 sont présents dans les 35 résidus en C-terminal. Une délétion dans cette région risque de diminuer la quantité d'hélices-alpha, si l'on considère que les prédictions de structure secondaire sont justes.

### *Conclusion et perspectives*

À la lumière des résultats obtenus aux sections 5.2.1 et 5.2.2, il apparaît évident que la F13 de la PapMV CP est essentielle à la multimérisation. Nous devons nous demander si c'est la phénylalanine à cette position qui est importante ou si c'est le nombre d'acides aminés à l'extrémité N-terminale qui est important. Il a été suggéré pour le Pepper Vein Banding potyvirus que le processus d'assemblage nécessitait un minimum d'environ 33 résidus à l'extrémité N-terminale exposés à la surface de la protéine (Anindya et Savithri 2003).

Un alignement des extrémités N-terminales de 19 *Potexvirus* fut réalisé et s'est avéré informatif (Figure 50). Les CPs de 6 des 19 *Potexvirus* possèdent une phénylalanine à la position F13 de la PapMV CP et 2 autres CPs possèdent une phénylalanine immédiatement avant ou immédiatement après la F13 de la PapMV CP. De plus, 6 autres *Potexvirus* ont un acide aminé hautement hydrophobe (W, Y, V et L) pour l'acide aminé qui précède la F13 de la PapMV CP. La phénylalanine faisant aussi partie des acides aminés hautement hydrophobes, 14 CPs alignées sur 19 possèdent un acide aminé hydrophobe dans l'environnement immédiat de la F13. L'hydrophobicité semi-conservée à l'extrémité N-terminale des CPs des *Potexvirus* pourrait donc jouer un rôle important dans l'assemblage des NLPs. La section 6.0 présente les résultats obtenus pour des mutations affectant la F13 de la PapMV CP.

	6	23
BaMV	GQQAAPQPWE	TKFTKDDLAA
FoMV	ADVTDATDYK	KPPAETEQKA
CymMV	AATYSAADPT	SAPKLADLAA
PAMV	.KQFSASDVR	SSPSLADLDE
NMV	PANADLSDPN	RAPSLLEDLKK
ScaVX	P..ADLSDPT	RAPSLKELQA
PepMV	.APSDFSNPN	TAPSLSDLKK
WCIMV	~MATTAT	TPPSLTDIRA
AltMV	~MSTP...█	PQVTQEOMDA
CVX	PRSTPQSGP█	QTLSSSQLAA
PIAMV	~~~~~MALN	TAPTADALAA
TVX	~~~~~MALN	TAPNPEALAA
CsCMV	AATTPLSALS	TAPTDEELSR
CIYMV	~MTDTKKTL█	SAPTDEQLDT
HVX	PPAAPSPVT█	TAPTQEQLTS
LVX	~~~~~NTT	█VPAKTWAD
SHYEV	NVANQVGDGP█	RVLTPEELAA
PVX	TPATASGL█T	.IPDGDFST
PapMV	H.STPNIA..█	PAITQEOMSS
Consensus	PA-TP-SDP█	TAPTLEDLAA

Figure 50. Alignement de l'extrémité N-terminale des CPs des *Potexvirus*. Les CPs de 6 potexvirus sur 19 ont une phénylalanine à la position F13 de la PapMV CP. Légende : acides aminés en vert : phénylalanine située autour de la position 13 de la PapMV CP; acides aminés en jaune : acides aminés hautement hydrophobes situés autour de la position 13 de la PapMV CP; BaMV: *Bamboo mosaic virus*; FoMV: *Foxtail mosaic virus*; CymMV: *Cymbidium mosaic virus*; PAMV : *Potato aucuba mosaic virus*; NMV: *Narcissus mosaic virus*; ScaVX : *Scallion virus X*; PepMV : *Pepino mosaic virus*; WCIMV : *White clover mosaic virus*; AltMV: *Alternanthera mosaic virus*; CVX : *Cactus virus X*; PIAMV : *Plantago asiatica mosaic virus*; TVX : *Tulip virus X*; CsCMV: *Cassava common mosaic virus*; CIYMV: *Clover yellow mosaic virus*; HVX: *Hosta virus X*; LVX: *Lily virus X*; SMYEV : *Strawberry mild yellow edge virus*; PVX : *Potato virus X*; PapMV : *Papaya mosaic virus*. Adapté du mémoire de Marie-Hélène Tremblay (Tremblay 2005).

## 6.0 Importance de l'hydrophobicité à l'extrémité N-terminale

### 6.1 Introduction

La protéine de capsid du PapMV, un *Potexvirus*, possède un domaine central similaire aux autres *Potexvirus* (Annexe 1), mais possède aussi des caractéristiques communes avec des *Potyvirus*. Par exemple, il a été démontré par des études biochimiques et immunochimiques que les extrémités N- et C-terminales des CPs des *Potyvirus* et des *Potexvirus* étaient localisées à la surface (Allison, Dougherty *et al.* 1985; Koenig et Torrance 1986; Shukla, Strike *et al.* 1988). Cependant, très peu d'informations sur leur possible rôle sont disponibles. Anindya et Savithri (Anindya et Savithri 2003) ont suggéré que les deux extrémités de la PVBV CP, un *Potyvirus*, étaient impliquées dans une interaction protéines:protéines au tout début du processus d'assemblage. Ils suggèrent aussi que leur modèle d'assemblage pourrait aussi s'appliquer aux autres *Potyvirus* puisque leurs extrémités N- et C-terminales sont aussi exposées à la surface.

Chez les *Potexvirus*, la présence de ces extrémités à la surface du PapMV a été confirmée récemment (Tremblay 2005). Notre équipe a aussi récemment prouvé qu'une partie de l'extrémité N-terminale était essentielle à l'assemblage de la CP en particule virale mature (Lecours, Tremblay *et al.* 2006). Les résultats obtenus au cours de la section précédente de ce travail suggèrent l'importance de l'hydrophobicité de l'acide aminé F13 pour l'assemblage. Sans cet acide aminé, la formation de particules est impossible. L'implication de cet acide aminé dans les toutes premières étapes de l'assemblage semble la meilleure hypothèse, puisque la délétion des treize premiers acides aminés mène à une protéine monomérique incapable d'interagir entre elles.

Le but de cette section est de vérifier l'importance de l'hydrophobicité retrouvée à l'extrémité N-terminale de la PapMV CP ainsi que chez quelques autres *Potexvirus* (Figure 50) dans les processus d'assemblage.

## 6.2 Résultats

Afin de tester l'importance de l'hydrophobicité présente à l'extrémité N-terminale de la PapMV CP, plusieurs mutations ponctuelles furent réalisées. Étant donné que c'est la F13 qui semble importante pour l'assemblage, nous avons changé cet acide aminé par d'autres acides aminés ayant une hydrophobicité équivalente ou moindre. La Figure 51 présente ces mutants ponctuels en ordre décroissant d'hydrophobicité.

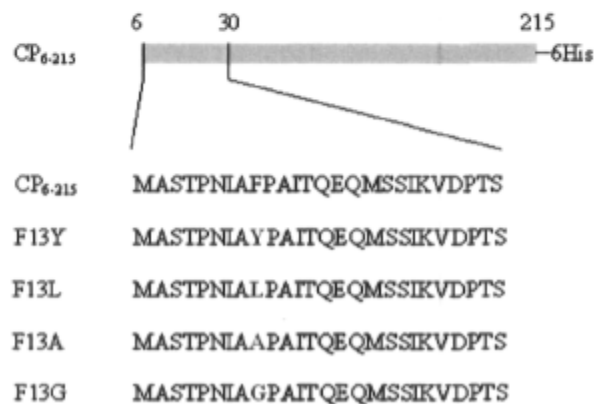


Figure 51. Schématisation des mutants ponctuels de la PapMV CP pour l'acide aminé F13. Les lettres colorées en rouge représentent les acides aminés mutés. Légende : 6His : queue de six histidines.

### 6.2.1 Expression et purification des mutants F13

Les quatre protéines ont été exprimées dans *E. coli* et purifiées selon le protocole décrit aux sections 2.2 à 2.3. La Figure 52 présente la vérification de l'expression des différentes protéines alors que la Figure 53 nous les présente une fois purifiées et dialysées. On peut remarquer que la protéine F13A purifiée et dialysée semble un peu sensible à la dégradation en raison de la légère bande apparaissant à une hauteur de 25 kDa sur le gel. La protéine F13G semble toutefois beaucoup plus sensible à la dégradation en raison de l'intensité de cette bande à 25 kDa. Le niveau de pureté de la F13G semble inférieur à celui des trois autres CPs. Les protéines F13L et F13Y, quant à elles, ne semblent pas être sensibles à la dégradation.

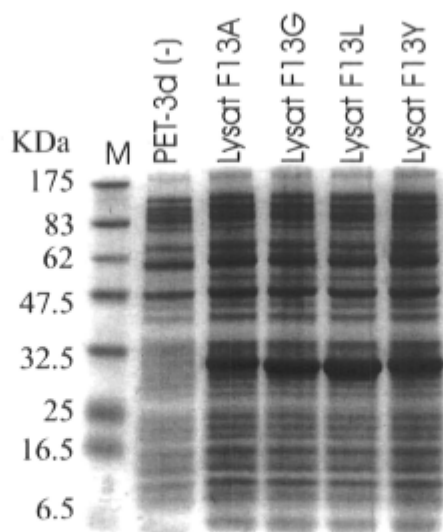


Figure 52. Vérification de l'induction des CPs F13A, F13G, F13L et F13Y sur gel SDS-PAGE. Légende : M : marqueurs de poids moléculaire; pET-3d (-) : lysat bactérien avant induction; Lysat F13: lysat bactérien après l'induction à l'IPTG 16 heures à 22 °C.

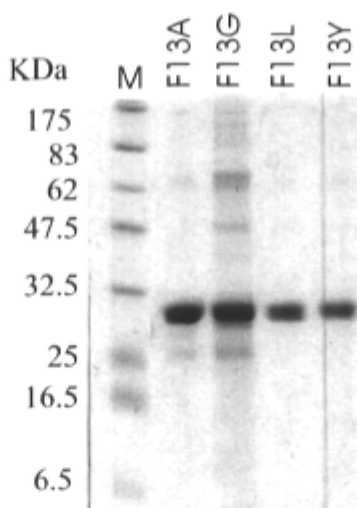


Figure 53. Gel dénaturant SDS-PAGE présentant les protéines F13A, F13G, F13L et F13Y purifiées et dialysées. Légende : M : marqueurs de poids moléculaire; F13 : mutant F13 purifiée et dialysée.

### 6.2.2 Tamisage moléculaire

Afin de connaître le profil de multimérisation des quatre mutants de la F13, un tamisage moléculaire au moyen d'une colonne Supdex<sup>TM</sup> 200 16/60 fut réalisé avec les protéines

purifiées et dialysées. La Figure 54 présente les résultats obtenus en comparaison avec la protéine de référence CP<sub>6-215</sub>.

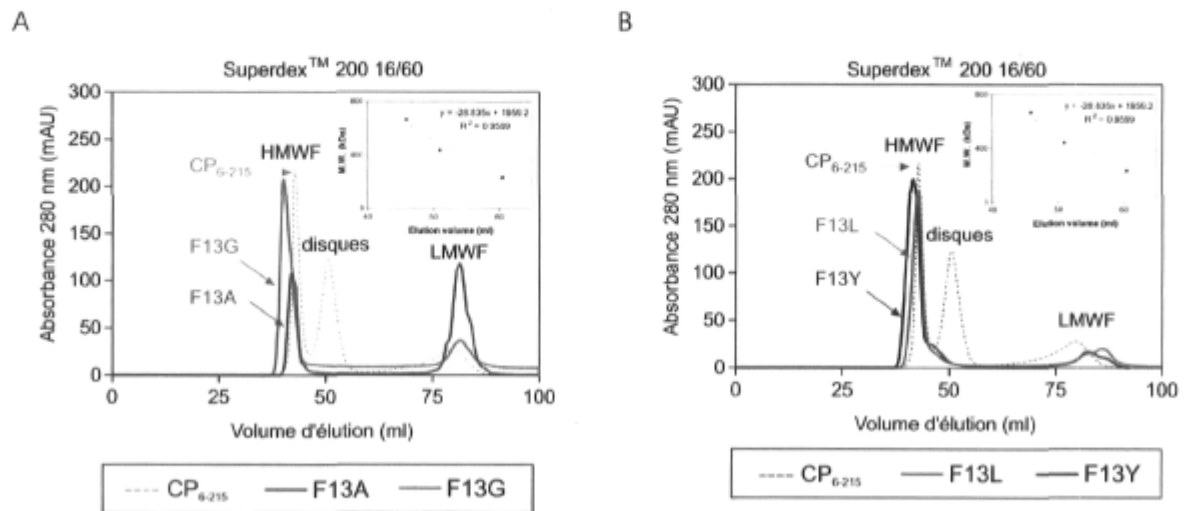


Figure 54. Tamisage moléculaire sur colonne Superdex™ 200 16/60 des mutants F13. A) Comparaison de profils de multimérisation des protéines F13A (courbe mauve), F13G (courbe turquoise) et CP<sub>6-215</sub> (courbe rose). B) Comparaison de profils de multimérisation des protéines F13L (courbe verte), F13Y (courbe bleue) et CP<sub>6-215</sub> (courbe rouge). Le graphique représentant la relation linéaire entre les poids moléculaires et le volume d'élution pour cette colonne a été placé dans le coin supérieur droit. Légende : HMWF : multimères de poids moléculaire supérieur à 670 kDa; disques : disques de 20 sous-unités de la CP de 450 kDa; LMWF : forme monomérique ou multimérique inférieure à 232 kDa.

Des résultats plutôt surprenants ont été obtenus pour les quatre protéines à l'étude. Il est intéressant de noter qu'aucune de ces protéines n'a été produite sous forme de disques de 450 kDa comme c'est le cas pour la protéine de référence CP<sub>6-215</sub>. De plus, les quatre formes mutantes de la CP contiennent un certain pourcentage de protéines se trouvant sous une forme multimérique supérieure à 670 kDa.

Pour F13A, la majorité de la protéine se trouve sous une forme monomérique ou multimérique inférieure à 232 kDa. Le restant de la CP se trouve dans le premier pic d'élution à 42,1 ml et comprend toute les formes de la CP multimérisée ayant un poids moléculaire supérieur à 670 kDa.

Le profil est quelque peu différent dans le cas de la F13G. Une très faible proportion de la CP se trouve à un poids moléculaire inférieur à 232 kDa, alors que la majorité de cette protéine est éluée à 40,3 ml. Il est intéressant de noter ici que ce pic d'éluion est plus large pour cette protéine que pour F13A et CP<sub>6-215</sub>.

Pour ce qui est des protéines F13L et F13Y, leur profil de purification est très semblable. La presque totalité des CPs est éluées sous une forme multimérique de poids moléculaire supérieur à 670 kDa à des volumes de 42,6 ml et 41,7 ml, respectivement. Le reste des protéines se trouve sous une forme multimérique de poids moléculaire inférieur à 232 kDa. On peut remarquer que ces protéines sont éluées quelques millilitres après les protéines correspondantes pour CP<sub>6-215</sub>.

### 6.2.3 Microscopie électronique

Afin de vérifier la présence de NLPs dans les protéines purifiées, le sommet des pics élués aux environs de 41 ml de chacune des protéines fut récolté et déposé sur des grilles de microscopie électronique. Les grilles ont été colorées à l'uranyle acétate 2%. Les photos de microscopie pour les différentes protéines sont présentées à la Figure 55.

On peut remarquer en regardant les photos de microscopie électronique que les NLPs des protéines F13L et F13Y semblent en moyenne plus longues que celles formées par la protéine CP<sub>6-215</sub>. Cette impression est confirmée par la mesure de la longueur de 250 NLPs par protéine (Figure 55F). Des longueurs moyennes de 50 nm, 130 nm et 200 nm furent obtenues pour les protéines CP<sub>6-215</sub>, F13L et F13Y. Il semblerait donc que ces deux mutants soient plus efficaces à former des NLPs.

Pour ce qui est de la protéine F13A, seulement quelques NLPs furent observées pour la totalité des grilles de microscopie. La majorité des protéines observées se trouvaient plutôt sous une forme d'agrégats non spécifiques. Pour la protéine F13G, aucune NLP ne fut observée. La totalité des protéines se trouvait sous une forme d'agrégats non spécifiques. Étant donné que le mutant F13G ne formait ni NLP, ni disque et que la portion de protéines



se trouvant sous une forme multimérique de faible poids moléculaire était hautement sensible à la dégradation, cette protéine fut retirée du reste de l'étude.

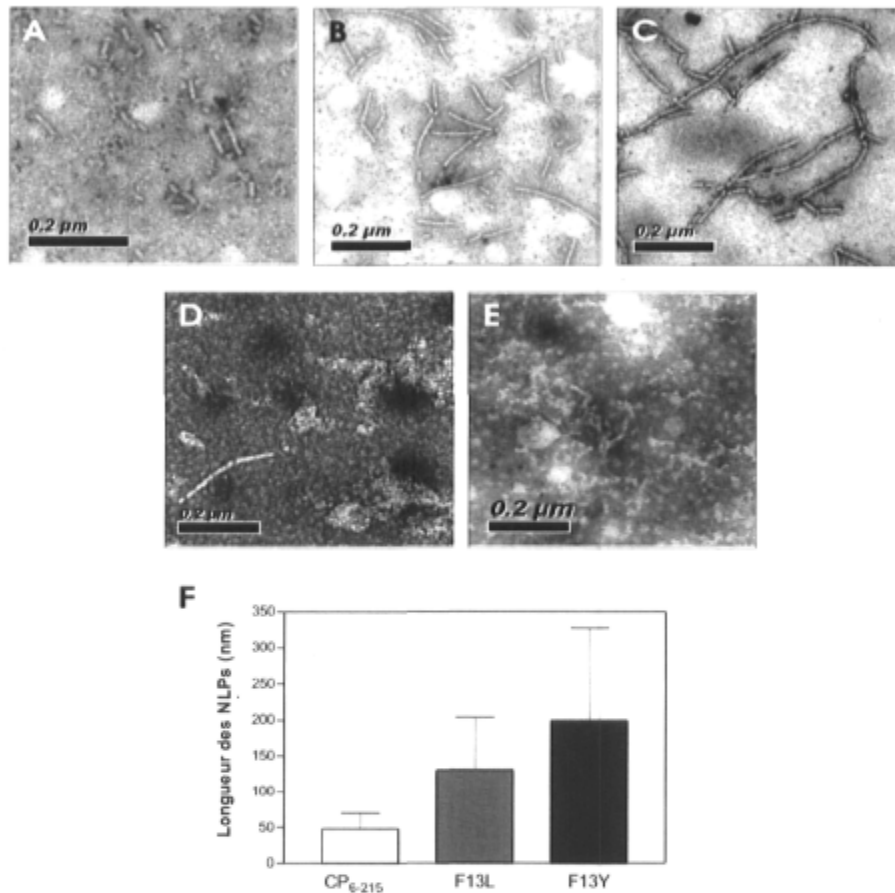


Figure 55. Comparaison de la microscopie électronique des protéines A) CP<sub>6-215</sub>; B) F13L; C) F13Y; D) F13A et E) F13G. La barre représente 200 nm. F) Moyenne des longueurs des particules virales de CP<sub>6-215</sub>, F13L et F13Y. 250 NLPs ont servi à faire les calculs.

#### 6.2.4 Liaison à l'ARN

Afin de confirmer l'incapacité de la protéine F13A à interagir avec les 80 premiers nucléotides de l'ARN du PapMV, un EMSA a été réalisé (Figure 56). Afin d'éviter les interactions non spécifiques dues aux agrégats non structurés de protéines, une ultracentrifugation d'une heure et demi à 35 000 RPM fut réalisée et c'est uniquement le surnageant qui fut incubé avec la sonde marquée. Aucune interaction entre les protéines et la sonde ne fut détectée. Ces résultats suggèrent que la protéine F13A se trouve sous une forme monomérique ou multimérique incapable d'interagir avec les acides nucléiques.

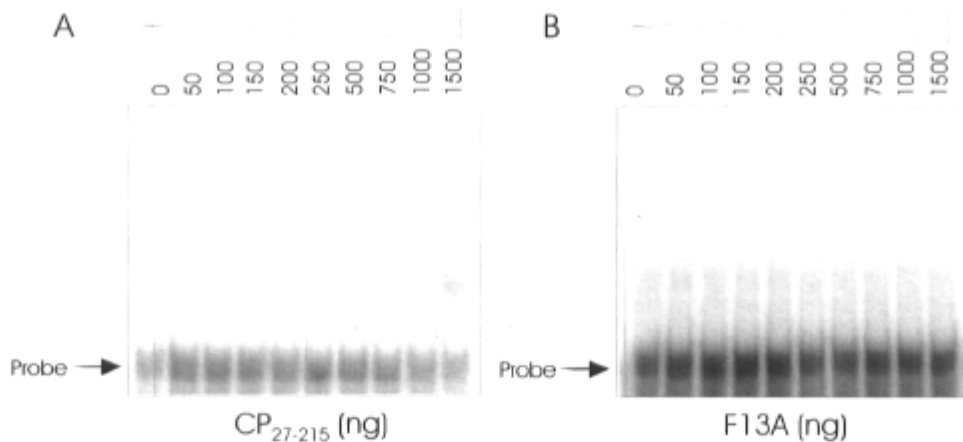


Figure 56. EMSA A) de la protéine monomérique CP<sub>27-215</sub> en comparaison avec celui B) de la protéine F13A. Des quantités croissantes de protéines ont été incubées avec une quantité constante de sonde d'ARN marquée au  $\gamma$ -<sup>32</sup>P. Les chiffres dans le haut du gel représentent les quantités croissantes de protéines incubées avec la sonde. Légende : Probe : sonde d'ARN marqué non lié aux protéines.

Un autre EMSA fut réalisé cette fois-ci pour les protéines F13L et F13Y afin de savoir si leur efficacité accrue à former de longues NLPs étaient causée par une meilleure affinité pour les acides nucléiques, tel que ce fut le cas pour la protéine E128A (section 1.4.2, (Tremblay, Majeau *et al.* 2006)). Nous avons procédé à un traitement à l'acide acétique tel qu'il a été décrit précédemment (section 2.10) (Erickson, Bancroft *et al.* 1976) afin d'obtenir des disques de 450 kDa à partir des NLPs. Le même traitement fut réalisé pour CP<sub>6-215</sub>. Les disques ainsi obtenus pour les trois protéines furent incubés avec la sonde marquée au  $\gamma$ -<sup>32</sup>P et les différentes réactions furent soumises à une migration électrophorétique. Les résultats sont présentés à la Figure 57.

Étonnamment, les deux protéines mutantes pour la F13 possédaient une affinité réduite voire nulle pour les acides nucléiques, alors que la CP<sub>6-215</sub> conservait sa forte affinité pour eux. Deux raisons de cet échec furent suggérées : 1- les protéines F13L et F13Y traitées à l'acide acétique ne forment pas de disques ou bien forment des disques possédant une structure inadéquate et 2- ces protéines possèdent une meilleure affinité entre elles et sont capables de s'assembler sans la présence d'acides nucléiques.

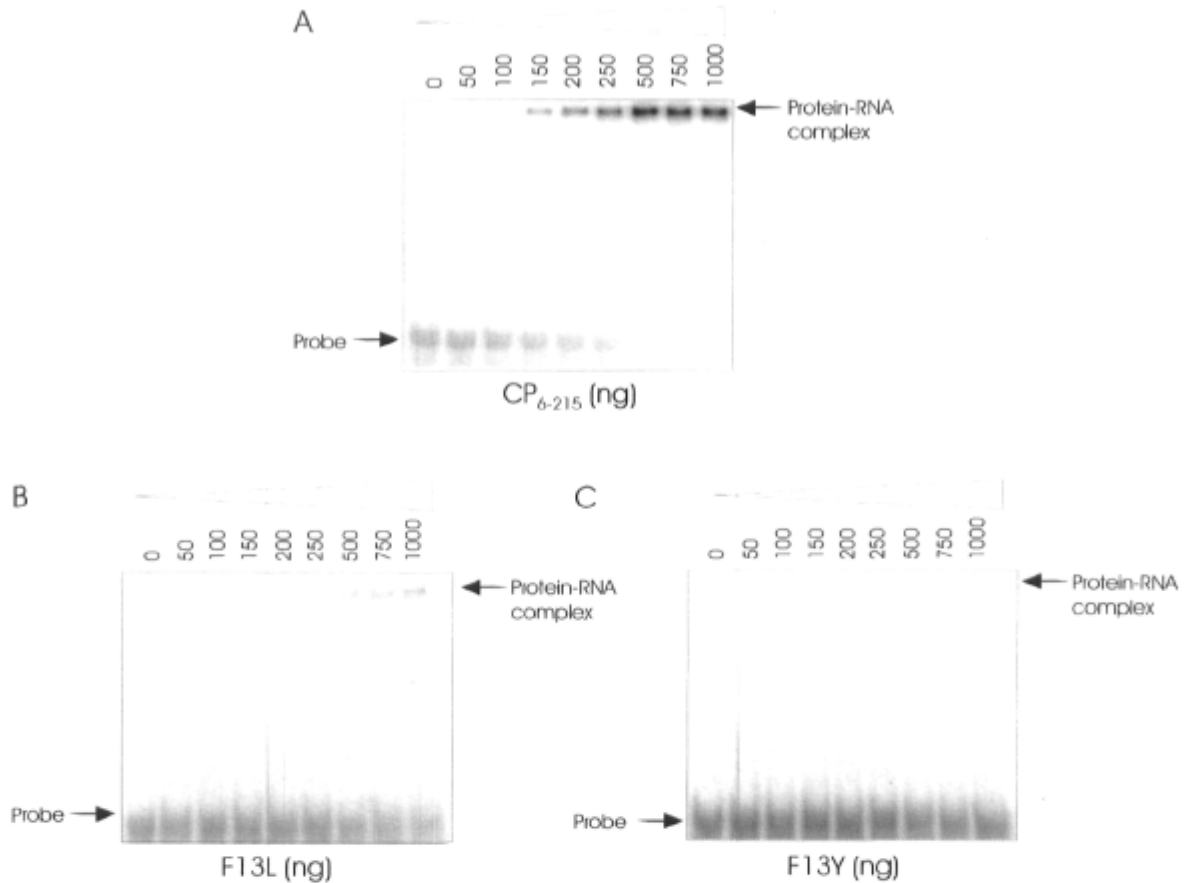


Figure 57. EMSA A) de la protéine CP<sub>6-215</sub> en comparaison avec ceux des protéines B) F13L et C) F13Y. Les disques des différentes protéines furent obtenus par un traitement à l'acide acétique (Erickson, Bancroft *et al.* 1976). Des quantités croissantes de protéines ont été incubées avec une quantité constante de sonde d'ARN marquée au  $\gamma$ - $^{32}$ P. Les chiffres dans le haut du gel représentent les quantités croissantes de protéines incubées avec la sonde. Légende : Probe : sonde d'ARN marqué non lié aux protéines ; Protein:RNA complex : protéines liées aux acides nucléiques marqués.

### 6.2.5 Microscopie électronique

Afin de vérifier laquelle des hypothèses formulées est la bonne, les disques extraits à l'acide acétique des trois protéines furent déposés sur une grille de microscopie électronique, colorés à l'uranyl acétate et observés au microscope électronique (Figure 58). La présence de disques fut décelée pour les trois protéines CP<sub>6-215</sub>, F13L et F13Y.

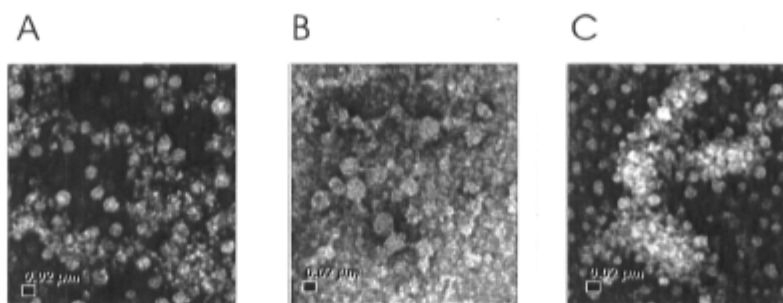


Figure 58. Microscopie électronique des disques extraits à l'acide acétique des protéines A) CP<sub>6-215</sub>, B) F13L et C) F13Y. La barre représente 20 nm.

### 6.2.6 Spectroscopie en dichroïsme circulaire

Afin de vérifier si la structure secondaire des disques de F13L et F13Y traités à l'acide acétique est affectée, un spectre CD fut enregistré pour chacune des deux protéines (Figure 59B et C). Les spectres CD obtenus sont comparables à celui de la protéine native (Figure 59A). En raison de la très faible quantité de disques extraits à l'acide acétique pour les protéines F13L et F13Y, il n'est pas possible de superposer les trois spectres CD. Une nouvelle expérience avec des protéines plus concentrées devrait être réalisée. Il semble donc que la diminution d'affinité pour les acides nucléiques observée à la Figure 57 ne peut pas être expliquée totalement par un problème de structure des disques.

### 6.2.7 Rapport des absorbances à 280 et 260 nm pour les NLPs

Des densités optiques furent prises à 260 nm et 280 nm pour les NLPs de F13L et F13Y afin de vérifier leur contenu en acides nucléiques. Le rapport DO 280/260 fut calculé et comparé à celui des virions de PapMV sauvages ainsi qu'à celui des NLPs de la CP<sub>6-215</sub>. L'expérience fut répétée trois fois à partir de protéines purifiées à différents moments. Le Tableau 9 présente la moyenne des rapports obtenus pour les quatre particules virales. Les valeurs obtenues pour les NLPs de F13L et de F13Y sont comparables à celle obtenue pour le virus sauvage et sont inférieures à celle des NLPs de CP<sub>6-215</sub>. Ce résultat suggère donc que la proportion d'ARN contenue dans les NLPs par rapport au nombre de protéines absorbant à 280 nm pour les mutants F13L et F13Y est plus semblable à celle retrouvée pour le virus sauvage que pour celle des NLPs de CP<sub>6-215</sub> produite dans *E. coli*. Il y a donc présence d'ARN encapsidé dans les NLPs de F13L et de F13Y.

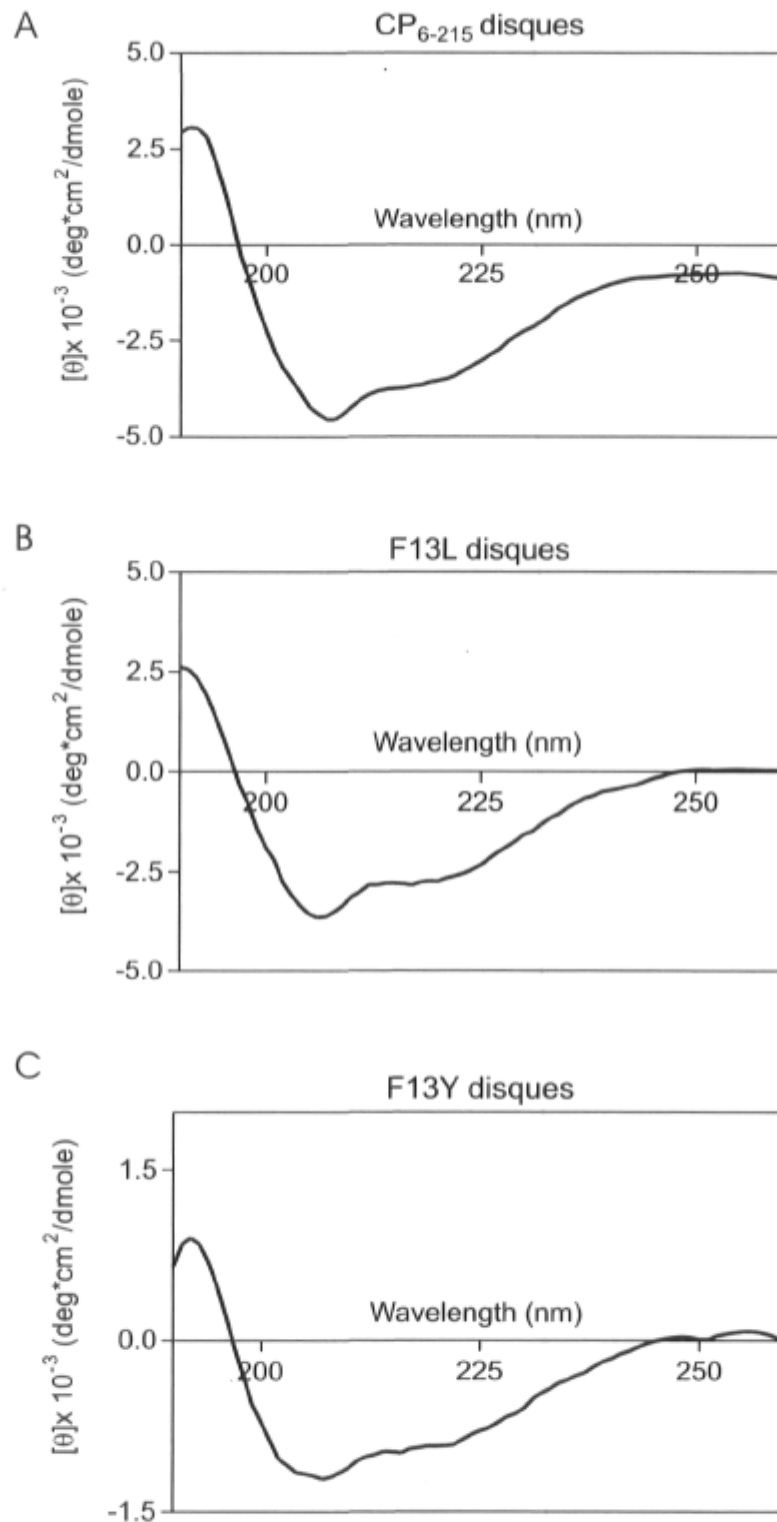


Figure 59. Spectre CD des NLPs traitées à l'acide acétique pour les protéines A) CP6-215; B) F13L et C) F13Y.

	Virus et NLPs			
	PapMV	CP <sub>6-215</sub>	F13L	F13Y
rapport 280/260	0,75	1,1	0,8	0,75

Tableau 9. Comparaison des rapports d'absorbance obtenus en divisant l'absorbance à 280 nm par celle à 260 nm pour les particules virales de PapMV sauvage et les NLPs de CP<sub>6-215</sub>, F13L et F13Y.

#### 6.2.8 Rapport des absorbances à 280 et 260 nm pour les disques

À la lumière des résultats obtenus dans la section précédente, une troisième hypothèse fut formulée : la présence d'ARN contaminant interagissant avec les sites de liaison à l'ARN des disques de F13L et de F13Y extraits à l'acide acétique peut expliquer la diminution d'affinité des protéines F13L et F13Y pour les acides nucléiques. Cette hypothèse fut vérifiée avec les rapports 280/260 obtenus de la même façon que décrit précédemment, mais avec les disques des quatre mêmes protéines provenant du traitement à l'acide acétique. Les rapports obtenus pour les disques de F13L et F13Y sont nettement inférieurs à ceux des deux autres protéines (Tableau 10). Ce résultat signifie donc qu'il y a présence d'acides nucléiques absorbant à 260 nm en plus grande quantité avec les disques des protéines F13L et F13Y qu'avec ceux de la protéine sauvage et de la CP<sub>6-215</sub>.

	Disques extraits			
	PapMV	CP <sub>6-215</sub>	F13L	F13Y
Rapport 280/260	1,5	1,55	0,95	0,9

Tableau 10. Comparaison des rapports d'absorbance obtenus en divisant l'absorbance à 280 nm par celle à 260 nm pour les disques traités à l'acide acétique à partir des virions de PapMV sauvage et des NLPs de CP<sub>6-215</sub>, F13L et F13Y.

### 6.3 Discussion

Au cours de cette partie du projet, l'importance de l'hydrophobicité retrouvée à la position 13 de la PapMV CP fut vérifiée par la production de quatre formes mutantes de la CP : F13A, F13G, F13L et F13Y. Chacune des quatre constructions précédentes comporte un acide aminé d'hydrophobicité différente à la place de la phénylalanine à la position 13. Le choix de ces acides aminés sera expliqué tout au long de cette section.

Tout d'abord, il est important de connaître les caractéristiques de la phénylalanine (abrégée Phe ou F). Cet acide aminé aromatique non polaire se caractérise par la présence d'un noyau benzène dans sa chaîne latérale (Figure 60A), lui conférant ainsi une chaîne latérale hydrophobe. La phénylalanine est souvent retrouvée dans le cœur hydrophobe des protéines. Elle est parfois impliquée dans les interactions protéines:protéines en interagissant avec d'autres acides aminés aromatiques par l'empilement de leur noyau benzène (Branden et Tooze 1996).

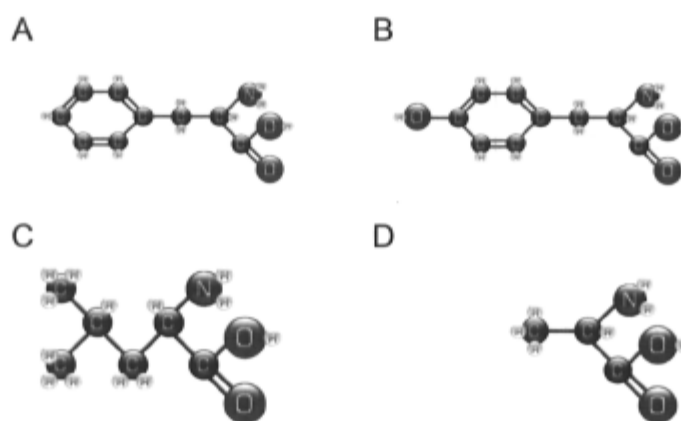


Figure 60. Représentation schématique de A) la Phénylalanine; B) la Tyrosine; C) la Leucine et D) l'Alanine. Légende : C : carbone; H : hydrogène; N : azote; O : oxygène. Tiré de <http://fr.wikipedia.org/wiki>.

L'acide aminé le plus près chimiquement de la phénylalanine est la tyrosine (abrégée Tyr ou Y). La chaîne latérale de la tyrosine est identique à celle de la phénylalanine à l'exception de son groupement OH présent sur l'un des carbones participant à l'anneau benzène (Figure 60B). Ce groupement OH lui confère la possibilité d'acquies une charge

au bout de la chaîne latérale faisant de la tyrosine un acide aminé aromatique polaire. Elle peut aussi participer à des interactions protéines:protéines par empilement (Branden et Tooze 1996). Un des variants de la PapMV CP réalisés au cours de cette partie du projet possède la F13 mutée pour une Y. Cette mutation avait pour but de vérifier l'importance des interactions protéiques par empilement et de vérifier la conséquence de l'introduction potentielle d'une charge à l'extrémité N-terminale tout en conservant le noyau hydrophobe de la F13.

Les résultats obtenus pour F13Y furent inattendus. La purification de cette protéine produite dans *E. coli* presque exclusivement sous forme de NLPs fut surprenante car la protéine de référence CP<sub>6-215</sub> en est incapable. De plus, les NLPs de la F13Y sont en moyenne quatre fois plus longues que celle de la CP<sub>6-215</sub> en demeurant toutefois trois fois plus courte que le virus sauvage. Jusqu'à aujourd'hui, nous ne pouvons expliquer entièrement ce résultat. Les séquences en acides aminés des protéines F13Y et CP<sub>6-215</sub> étant identique à l'exception du treizième acide aminé, les différences de profils observées entre les deux CPs ne peuvent s'expliquer que par la mutation effectuée. Il est possible que l'ajout du groupement OH au noyau benzène de la chaîne latérale de la F13 soit responsable de l'efficacité accrue de l'assemblage de la protéine. Il serait intéressant de faire un nouveau mutant ponctuel pour la F13 en introduisant cette fois-ci un acide aminé chargé négativement (un acide aspartique ou un acide glutamique) à la place de la phénylalanine.

Un des vingt acides aminés reconnus comme étant des plus hydrophobes est la leucine (abrégée Leu ou L). Cet acide aminé non polaire possède une longue chaîne latérale constituée uniquement de CH<sub>2</sub> et de CH<sub>3</sub> (Figure 60C). La leucine est connue pour son hydrophobicité qui joue souvent un rôle dans les interactions protéines:protéines en formant un «leucine zipper» (Figure 61). Ce motif structural est retrouvé dans les interactions entre deux hélices-alpha parallèles («coiled-coil»). La caractéristique majeure de cette structure est la prédominance de la leucine à la position *d* de l'hélice. C'est-à-dire que la leucine est souvent présente à tous les 7 acides aminés participants à l'hélice (Branden et Tooze 1996).



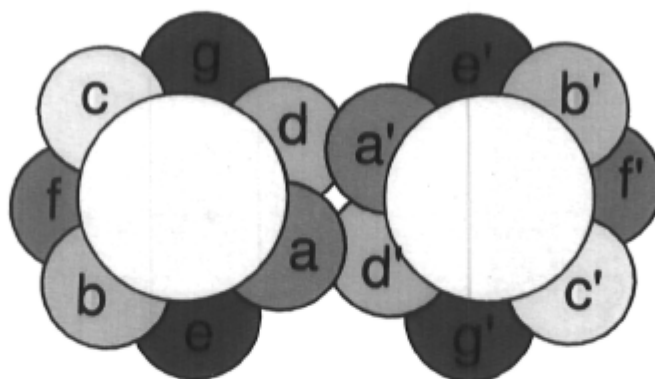


Figure 61. Représentation schématique du motif structural appelé «Leucine zipper». Tiré de [http://en.wikipedia.org/wiki/Leucine\\_zipper](http://en.wikipedia.org/wiki/Leucine_zipper).

En raison de l'hydrophobicité de sa chaîne latérale, une leucine fut introduite dans la séquence de la PapMV CP à la place de la F13. Les résultats obtenus pour ce mutant sont tout aussi surprenants que ceux obtenus pour F13Y. La CP F13L produite et purifiée selon le protocole décrit aux sections 2.2 et 2.3 est aussi retrouvée presque exclusivement sous forme de NLPs. Ces pseudovirions produits sont aussi plus longs (environ 2,6 fois) que ceux obtenus pour la protéine de référence CP<sub>6-215</sub> mais en étant plus courts que ceux observés pour F13Y. Comme pour F13Y, nous ne pouvons expliquer la raison de cette efficacité d'assemblage accrue.

La troisième mutation effectuée fut de changer la F13 pour une alanine. L'alanine (abrégée Ala ou A) est un acide aminé hydrophobe retrouvé un peu partout dans les protéines. Sa courte chaîne latérale est constituée uniquement d'un groupement méthyle (CH<sub>3</sub>) (Figure 60D). Le remplacement de la F13 dans la PapMV CP par une alanine s'est caractérisé par une diminution de l'efficacité de la multimérisation. La plupart de la protéine F13A se trouve sous une forme multimérique de faible poids moléculaire et très peu de NLPs furent observées parmi les multimères de poids moléculaires supérieurs à 670 kDa. Cependant, la proportion de NLPs par rapport à celle des agrégats non spécifiques n'a pas pu être déterminée. Ces résultats viennent appuyer l'hypothèse selon laquelle l'hydrophobicité de la F13 serait importante pour la multimérisation. De plus, en raison de la grande quantité de monomères ou multimères de faible poids moléculaires, il est probable que

l'hydrophobicité de l'extrémité N-terminale joue un rôle seulement dans les premières étapes de l'assemblage. La présence de NLPs dans l'échantillon appuie la supposition qu'une fois que les CPs sont correctement multimérisées, l'assemblage se fait de façon efficace.

La dernière substitution d'acides aminés réalisée au cours de cette section fut de remplacer la F13 par une glycine (abrégée Gly ou G). La glycine est l'acide aminé le plus simple des 20 acides aminés. Elle possède uniquement un hydrogène en guise de chaîne latérale. Le but de la mutation F13G était de confirmer les résultats obtenus avec la F13A avec un acide aminé encore moins hydrophobe que l'alanine. La forte tendance de la protéine à s'agréger de façon non spécifique ainsi que l'instabilité de la faible proportion de la CP non agrégée confirme l'hypothèse de l'importance de l'hydrophobicité pour la multimérisation.

À la lumière des résultats obtenus au cours de cette partie du projet, il ne fait aucun doute que l'hydrophobicité présente à la position 13 de la PapMV CP est importante pour la multimérisation de la CP et l'assemblage en particules matures.

## 7.0 Conclusion

Tout au long de ce projet de recherche, des informations pertinentes concernant l'implication de l'extrémité N-terminale de la PapMV CP dans la multimérisation et l'assemblage ont été mises à jour. En premier lieu, l'élaboration d'un protocole standard de production et de purification des CPs exprimées dans *E. coli* a permis de faire avancer plus rapidement les recherches en standardisant la méthode, mais surtout en permettant des niveaux d'expression de protéines plus élevés.

Les travaux avec les mutants de délétions ont permis de mieux cibler la région essentielle à la multimérisation à l'extrémité N-terminale. Auparavant, nous savions qu'il y avait quelque chose d'important dans les 26 premiers acides aminés mais nous ignorions quoi et comment cela se déroulait. Les travaux en RMN de Katia Lecours en collaboration avec nos résultats ont permis d'écarter l'hypothèse de l'implication essentielle de la petite hélice Q18-S22 dans la multimérisation. Son existence est aussi remise en doute à la lumière des résultats obtenus. Les mutants ponctuels E19P et E19K ont permis de démontrer l'absence d'un pont ionique essentiel à la multimérisation à l'extrémité N-terminale. Cependant, il est possible que ce pont existe vraiment, mais qu'il serve plutôt à stabiliser la structure. L'existence d'un pont ionique essentiel à l'assemblage impliquant deux acides aminés chargés situés ailleurs dans la CP n'est pas exclue : l'hypothèse du pont ionique impliquant une arginine et un acide aspartique hautement conservé chez les virus filamenteux flexibles postulée par Dolja *et al.* en 1991 demeure toujours valide (Dolja, Boyko *et al.* 1991).

Des travaux réalisés sur les formes tronquées CP<sub>13-215</sub> et CP<sub>13-180</sub> ont permis de créer un double mutant de délétion qui, selon les résultats préliminaires, devrait être plus stable et résistant à la dégradation. L'obtention d'un tel mutant permet d'avoir de meilleurs résultats en RMN, ce qui peut mener à la découverte de la structure tridimensionnelle de la première CP de virus filamenteux flexibles.

Finalement, l'étude entourant l'acide aminé F13 de la PapMV CP a permis de mettre à jour l'importance de l'hydrophobicité à cette position. Une diminution du caractère hydrophobe

de l'acide aminé à cette position résulte en une difficulté, voire en une incapacité à s'assembler de façon efficace.

Il est maintenant évident que l'extrémité N-terminale joue un rôle dans les toutes premières étapes de l'assemblage du PapMV. Lorsque que l'hydrophobicité autour de la position 13 est grandement réduite ou absente, aucune multimérisation efficace ne peut être effectuée. L'extrémité N-terminale est donc essentielle à la dimérisation de la protéine de coque. Le même constat fut fait pour un autre virus filamenteux, le PVBV. Ce potyvirus est incapable d'initier son assemblage lorsque les extrémités N- ou C-terminales sont absentes. Un modèle d'assemblage pour ce virus fut alors proposé (Figure 62) (Anindya et Savithri 2003). Ce modèle soutient que l'extrémité N-terminale d'une CP interagirait avec l'extrémité C-terminale d'une autre CP via des forces électrostatiques. Ce processus se poursuivrait jusqu'à l'obtention d'un intermédiaire de l'assemblage de coefficient de sédimentation de 16S. Ce «disque» serait indispensable à l'obtention de particules virales complètes. Une fois le disque formé, les extrémités N- et C-terminales peuvent être enlevées sans interférer dans le processus d'assemblage.

Il est possible que ce modèle d'assemblage soit applicable, du moins en partie, à ce qui se produit pour le PapMV. Dans le cas du PapMV, nous proposons que les interactions des extrémités de la CP interagissent entre elles via des interactions hydrophobes. Selon des résultats préliminaires obtenus par Marie-Hélène Tremblay, les 35 derniers acides aminés à l'extrémité C-terminale ne sont pas nécessaires à la formation d'intermédiaires d'assemblage et de disques. Cependant, ces acides aminés sont essentiels à la formation de NLPs (Tremblay 2005). Un travail similaire à celui qui a été fait pour l'extrémité N-terminale, impliquant la création de plusieurs mutants de délétions et ponctuels, serait nécessaire pour valider l'hypothèse de l'interaction entre les deux extrémités de la protéine de coque. De plus, cette étude pourrait permettre d'identifier le ou les acides aminés cruciaux pour l'assemblage présent à l'extrémité C-terminale.

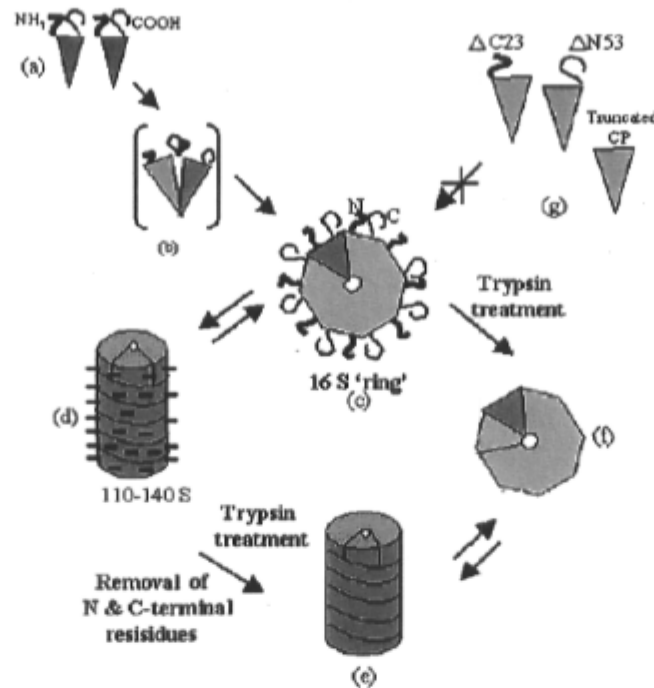


Figure 62. Modèle d'assemblage du Pepper Vein Banding potyvirus. (a) les extrémités N- et C-terminales interagissent ensemble (b) pour former un intermédiaire 16S en forme de disque (c) avec les extrémités exposées à la surface disponibles pour une digestion par la trypsine (f). Ces intermédiaires forment des NLPs à bas pH et à faible concentration ionique (d) et peuvent être reconvertis en intermédiaire 16S. Les extrémités N- et C-terminales exposées à la surface peuvent aussi être digérées par la trypsine à partir des NLPs (e). Ces particules trypsinisées peuvent également être dissociées en intermédiaire 16S (f). Les mutants CPΔC23, CPΔN53 (g) ne peuvent pas s'assembler en intermédiaire 16S. Tiré de (Anindya et Savithri 2003).

La longueur variable des NLPs qui sont produites dans *E. coli* est problématique. Cette variabilité serait probablement due à la différence de longueur des ARNs bactériens encapsidés. La nature de ces ARNs bactériens demeure toujours inconnue jusqu'à ce jour. La coexpression de la PapMV CP et d'un ARN comportant le signal d'encapsidation de l'ARN viral de PapMV, devrait permettre l'obtention de NLPs formées de l'ARN viral et de la PapMV CP. En effet, il est possible de supposer que l'ARN viral sera encapsidé préférentiellement à l'ARN bactérien. De plus, la coexpression de la CP et de l'ARN viral devrait permettre de produire des NLPs de longueur similaire, puisque tous les ARNs encapsidés auront la même longueur.

Finalement, il reste encore beaucoup de travail à faire avant de pouvoir déterminer comment se déroule l'assemblage du PapMV, quels domaines de sa CP sont impliqués et à quel moment. Une meilleure compréhension des processus régissant l'assemblage du PapMV serait très utile à notre équipe pour l'élaboration de nouveau vaccin utilisant comme plateforme vaccinale le PapMV. En effet, la capacité de contrôler la proportion de NLPs produites par rapport aux autres formes permettrait de diminuer les coûts de production tout en augmentant l'efficacité des recherches. Les mutants F13L et F13Y, produits à presque 100 % sous forme de NLPs, sont présentement à l'étude afin de vérifier s'ils peuvent être utilisés.

## Bibliographie

- Abouhaidar, M. and J. B. Bancroft (1978). "The initiation of papaya mosaic virus assembly." *Virology* **90**(1): 54-9.
- Abouhaidar, M. G. and J. B. Bancroft (1980). "The Polarity of Assembly of Papaya Mosaic-Virus and Tobacco Mosaic-Virus RNAs with Pmv-Protein under Conditions of Nonspecificity." *Virology* **107**(1): 202-207.
- Abouhaidar, M. G. and R. Lai (1989). "Nucleotide sequence of the 3'-terminal region of clover yellow mosaic virus RNA." *J Gen Virol* **70** ( Pt 7): 1871-5.
- Adams, M. J., J. F. Antoniw, *et al.* (2004). "The new plant virus family Flexiviridae and assessment of molecular criteria for species demarcation." *Arch Virol* **149**(5): 1045-60.
- Allison, R. F., W. Dougherty, *et al.* (1985). "Biochemical analysis of the capsid protein gene and capsid protein of tobacco etch virus: N-terminal amino acids are located on the virion's surface." *Virology* **147**(2): 309-16.
- Anindya, R. and H. S. Savithri (2003). "Surface-exposed amino- and carboxy-terminal residues are crucial for the initiation of assembly in Pepper vein banding virus: a flexuous rod-shaped virus." *Virology* **316**(2): 325-36.
- Baratova, L. A., A. V. Efimov, *et al.* (2001). "In situ spatial organization of Potato virus X coat protein subunits as assessed by tritium bombardment." *J Virol* **75**(20): 9696-702.
- Baratova, L. A., N. I. Grebenshchikov, *et al.* (1992). "The topography of the surface of potato virus X: tritium planigraphy and immunological analysis." *J Gen Virol* **73** ( Pt 2): 229-35.
- Bawden, F. C., Pirie, N.W., Bernal, J. D. and Fankuchen, I (1936). "Liquid crystalline substances from virus-infected plants." *Nature* **138**: 1051-1055.
- Bernal, J. D. a. I. F. (1941). "X-ray and crystallographic studies of plant virus preparations." *J Gen Physiol* **25**: 111-165.
- Bhyravbhatla, B., S. J. Watowich, *et al.* (1998). "Refined atomic model of the four-layer aggregate of the tobacco mosaic virus coat protein at 2.4-Å resolution." *Biophys J* **74**(1): 604-15.
- Bink, H. H. and C. W. Pleij (2002). "RNA-protein interactions in spherical viruses." *Arch Virol* **147**(12): 2261-79.
- Blanch, E. W., D. J. Robinson, *et al.* (2002). "Solution structures of potato virus X and narcissus mosaic virus from Raman optical activity." *J Gen Virol* **83**(Pt 1): 241-6.
- Branden, C. and J. Tooze (1996). *Introduction à la structure des protéines.*
- Campos-Olivas, R., J. L. Newman, *et al.* (2000). "Solution structure and dynamics of the Rous sarcoma virus capsid protein and comparison with capsid proteins of other retroviruses." *J Mol Biol* **296**(2): 633-49.
- Canizares, M. C., L. Nicholson, *et al.* (2005). "Use of viral vectors for vaccine production in plants." *Immunol Cell Biol* **83**(3): 263-70.
- Chapman, S., G. Hills, *et al.* (1992). "Mutational analysis of the coat protein gene of potato virus X: effects on virion morphology and viral pathogenicity." *Virology* **191**(1): 223-30.
- Culver, J. N. (2002). "Tobacco mosaic virus assembly and disassembly: determinants in pathogenicity and resistance." *Annu Rev Phytopathol* **40**: 287-308.

- Denis, J., N. Majeau, *et al.* (2007). "Immunogenicity of papaya mosaic virus-like particles fused to a hepatitis C virus epitope: Evidence for the critical function of multimerization." *Virology*.
- Dolja, V. V., V. P. Boyko, *et al.* (1991). "Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: two families with distinct patterns of sequence and probably structure conservation." *Virology* **184**(1): 79-86.
- Dolja, V. V., R. Haldeman, *et al.* (1994). "Distinct functions of capsid protein in assembly and movement of tobacco etch potyvirus in plants." *Embo J* **13**(6): 1482-91.
- Dolja, V. V., R. Haldeman-Cahill, *et al.* (1995). "Capsid protein determinants involved in cell-to-cell and long distance movement of tobacco etch potyvirus." *Virology* **206**(2): 1007-16.
- Dolja, V. V. and E. V. Koonin (1991). "Phylogeny of capsid proteins of small icosahedral RNA plant viruses." *J Gen Virol* **72**(Pt 7): 1481-6.
- Erickson, J. W. and J. B. Bancroft (1978). "The self-assembly of papaya mosaic virus." *Virology* **90**(1): 36-46.
- Erickson, J. W., J. B. Bancroft, *et al.* (1976). "The assembly of papaya mosaic virus protein." *Virology* **72**(2): 514-7.
- Gorry, P. A. (1990). "General Least-Squares Smoothing and Differentiation by the Convolution (Savitzky-Golay) Method." *Analytical Chemistry* **62**(6): 570-573.
- Gregory, J. a. K. C. H. (1965). "Methods of preparing oriented tobacco mosaic virus sols for x-ray diffraction." *J Mol Biol* **13**: 769-801.
- Huisman, M. J., H. J. Linthorst, *et al.* (1988). "The complete nucleotide sequence of potato virus X and its homologies at the amino acid level with various plus-stranded RNA viruses." *J Gen Virol* **69** ( Pt 8): 1789-98.
- Kaftanova, A. S., N. A. Kiselev, *et al.* (1975). "Structure of products of protein reassembly and reconstruction of potato virus X." *Virology* **67**(1): 283-7.
- Khorasanizadeh, S., R. Campos-Olivas, *et al.* (1999). "Solution structure of the capsid protein from the human T-cell leukemia virus type-I." *J Mol Biol* **291**(2): 491-505.
- Koenig, R. and L. Torrance (1986). "Antigenic Analysis of Potato Virus X by Means of Monoclonal Antibodies." *J Gen Virol* **67**: 2145-51.
- Koonin, E. V. and V. V. Dolja (1993). "Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences." *Crit Rev Biochem Mol Biol* **28**(5): 375-430.
- Kozlovsky, S. V., O. V. Karpova, *et al.* (2003). "Effect of the N-terminal domain of the coat protein of potato virus X on the structure of viral particles." *Dokl Biochem Biophys* **391**: 189-91.
- Lecours, K. (2007). "Caractériser la structure et la dynamique par RMN en solution pour la protéine de capsid du Virus de la Mosaïque de la Papaye." *Mémoire à l'Université Laval*.
- Lecours, K., M. H. Tremblay, *et al.* (2006). "Purification and biochemical characterization of a monomeric form of papaya mosaic potexvirus coat protein." *Protein Expr Purif* **47**(1): 273-80.
- Melcher, U. (2000). "The '30K' superfamily of viral movement proteins." *J Gen Virol* **81**(Pt 1): 257-66.
- Namba, K., R. Pattanayek, *et al.* (1989). "Visualization of protein-nucleic acid interactions in a virus. Refined structure of intact tobacco mosaic virus at 2.9 Å resolution by X-ray fiber diffraction." *J Mol Biol* **208**(2): 307-25.



- Parker, L., A. Kendall, *et al.* (2002). "Surface features of potato virus X from fiber diffraction." *Virology* **300**(2): 291-5.
- Porta, C. and G. P. Lomonosoff (1998). "Scope for using plant viruses to present epitopes from animal pathogens." *Rev Med Virol* **8**(1): 25-41.
- Purcifull, D. E. and E. Heibert (1971). "Papaya Mosaic Virus." *CMI/AAB no.56*.
- Shukla, D. D., P. M. Strike, *et al.* (1988). "The N-Termini and C-Termini of the Coat Proteins of Potyviruses Are Surface-Located and the N-Terminus Contains the Major Virus-Specific Epitopes." *Journal of General Virology* **69**: 1497-1508.
- Shukla, D. D. and C. W. Ward (1989). "Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group." *Adv Virus Res* **36**: 273-314.
- Sit, T. L., M. G. Abouhaidar, *et al.* (1989). "Nucleotide sequence of papaya mosaic virus RNA." *J Gen Virol* **70**(Pt 9): 2325-31.
- Skryabin, K. G., A. S. Kraev, *et al.* (1988). "The nucleotide sequence of potato virus X RNA." *Nucleic Acids Res* **16**(22): 10929-30.
- Stubbs, G. (1999). "Tobacco mosaic virus particle structure and the initiation of disassembly." *Philos Trans R Soc Lond B Biol Sci* **354**(1383): 551-7.
- Tollin, P., J. B. Bancroft, *et al.* (1979). "Diffraction Studies of Papaya Mosaic-Virus." *Virology* **98**(1): 108-115.
- Tremblay, M.-H. (2005). "Cartographie des domaines de la protéine de coque du virus de la mosaïque de la papaye impliqués dans l'assemblage viral." *Mémoire à l'Université Laval*.
- Tremblay, M. H., N. Majeau, *et al.* (2006). "Effect of mutations K97A and E128A on RNA binding and self assembly of papaya mosaic potyvirus coat protein." *Febs J* **273**(1): 14-25.
- Umashankar, M., M. R. Murthy, *et al.* (2006). "The role of inter-subunit ionic interactions in the assembly of Physalis mottle tymovirus." *Arch Virol* **151**(10): 1917-31.
- Urcuqui-Inchima, S., A. L. Haenni, *et al.* (2001). "Potyvirus proteins: a wealth of functions." *Virus Res* **74**(1-2): 157-75.
- Verde, C., A. Malorni, *et al.* (1989). "The primary structure of papaya mosaic virus coat protein: a revision." *J Protein Chem* **8**(6): 795-805.
- Zhang, H., E. Todderud, *et al.* (1993). "Crystallization and preliminary X-ray analysis of papaya mosaic virus coat protein." *J Mol Biol* **234**(3): 885-7.

## Annexe 1. Alignement des CPs des *Potexvirus*

Name: BAMBOO\_MOSAIC\_VIRUS\_CP Len: 276 Check: 2224 Weight: 1.00  
 Name: FOXTAIL\_MOSAIC\_VIRUS Len: 276 Check: 1995 Weight: 1.00  
 Name: CYMBIDIUM\_MOSAIC\_VIRUS Len: 276 Check: 2111 Weight: 1.00  
 Name: POTATO\_AUCUBA\_MOSAIC\_VIRUS Len: 276 Check: 8389 Weight: 1.00  
 Name: NARCISSUS\_MOSAIC\_VIRUS\_CP Len: 276 Check: 9662 Weight: 1.00  
 Name: SCALLION\_VIRUS\_X Len: 276 Check: 7794 Weight: 1.00  
 Name: PEPINO\_MOSAIC\_VIRUS\_CP Len: 276 Check: 7733 Weight: 1.00  
 Name: WHITE\_CLOVER\_MOSAIC\_VIRUS Len: 276 Check: 8132 Weight: 1.00  
 Name: Altranathera\_Potex Len: 276 Check: 800 Weight: 1.00  
 Name: Cactus\_virus\_X\_CP Len: 276 Check: 9357 Weight: 1.00  
 Name: PLANTAGO\_ASIATICA\_MOSAIC\_POTEXVIRUS Len: 276 Check: 2033 Weight: 1.00  
 Name: TULIP\_VIRUS\_X Len: 276 Check: 1230 Weight: 1.00  
 Name: CASSAVA\_COMMON\_MOSAIC\_VIRUS Len: 276 Check: 4587 Weight: 1.00  
 Name: CLOVER\_YELLOW\_MOSAIC\_VIRUS Len: 276 Check: 6966 Weight: 1.00  
 Name: HOSTA\_VIRUS\_X\_CP Len: 276 Check: 2065 Weight: 1.00  
 Name: LILY\_VIRUS\_X\_CP Len: 276 Check: 3398 Weight: 1.00  
 Name: STRAWBERRY\_MILD\_YELLOW\_EDGE-ASSOCIATED Len: 276 Check: 7751 Weight: 1.00  
 Name: POTATO\_VIRUS\_X\_CP Len: 276 Check: 127 Weight: 1.00  
 Name: PapMV\_CLONE-Pet\_MH Len: 276 Check: 9963 Weight: 1.00

//

	1				50
BAMBOO_MOSAI	~~~~~MS	GTGTGTGRGT	GTGVGGTGGT	GGTGGGGTGR	GQQAAPQPWE
FOXTAIL_MOSA	~~~~~	~~~~~	~~~~~	~~~~~MATQN	ADVTDATDYK
CYMBIDIUM_MO	~~~~~	~~~~~	~~~~~	~~MGEPTPTP	AATYSAADPT
POTATO_AUCUB	MVDSKKTETP	QVVDASKKAE	NSKTSQAGRI	QFLSAP....	.KQFSASDVR
NARCISSUS_MO	~~~~~	~~~~~	~~~~~MA	TPSTQTDDPK	PANADLSDPN
SCALLION_VIR	~~~~~	~~~~~	~~~MDKLDAG	QPQRKQAEPV	P..ADLSDPT
PEPINO_MOSAI	~~~~~	~~~~~MPD	TTPVAATSSA	PPTAKDAGAK	.APSDFSNPN
WHITE_CLOVER	~~~~~	~~~~~	~~~~~	~~~~~	~~~MATTTAT
Altranathera	~~~~~	~~~~~	~~~~~	~~~~~	~MSTP....F
Cactus_virus	~~~~~	~~~~~	~~~~~MST	TGVQSSQSSG	PRSTPQSGPF
PLANTAGO_ASI	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~MALN
TULIP_VIRUS	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~MALN
CASSAVA_COMM	~~~~~	~~~~~	~~~~~MATPTS	TTPTTATITQ	AATTPLSALS
CLOVER_YELLOW	~~~~~	~~~~~	~~~~~	~~~~~	~MTDTKKTLF
HOSTA_VIRUS	~~~~~	~~~~~	~~~~~	~~~MASDAPT	PPAAPSPVTF
LILY_VIRUS_X	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~MTT
STRAWBERRY_M	~~~~~MGDQPR	PPVPPAPGSN	PLPMGSTPPV	LPGRTPNPNA	NVANQVGDPPF
POTATO_VIRUS	~~~~~MSAP	ASTTQATGST	TSTTTKTAGA	TPATASGLFT	
PapMV_CLONE-	~~~~~	~~~~~	~~~~~	~~~~~	MASTPNIA.F
Consensus	-----	--V-----A-	--P-GATGST	TPTTTTPT-P-	PA-TP-SDPF

	51				100
BAMBOO_MOSAI	TKFTKDDLAA	IEPKPASANV	PNTKQWISIQ	AGLIKA.GAT	DANFMKVLG
FOXTAIL_MOSA	KPPAETEQA	LTIQPRSNKA	PSDEELVRII	NAAQKR.GLT	PAAFVQAAIV
CYMBIDIUM_MO	SAPKADLAA	IKYSPVTSSI	ATPEEIKAIT	QLWVNNLGLP	ADTVGTAAI.
POTATO_AUCUB	SSPSLADLDE	IAYEVRTTSI	ASPAEIEAVC	QLWIRNTEIP	ADKVALIAI.
NARCISSUS_MO	RAPSLLEDLKK	IKYESTTTAV	ATPAEIQLLG	DLF.KKLGLD	ANSVAP.AMW
SCALLION_VIR	RAPSLKELQA	VKYVSTTTSV	ATPDEIKQLG	ELF.QKLGVD	GSSIGP.AMW
PEPINO_MOSAI	TAPSLSDLKK	VKYVSTVTSV	ATPAEIEALG	KIF.TAMGLA	ANETGP.AMW
WHITE_CLOVER	TPPSLTDIRA	LKYTSSTVSV	ASPAEIEAIT	KTW.AETFKI	PNDVLPLACW
Altranathera	PQVTQEOMDA	FTPHTTNNLL	PSPEQLTTIA	SLL.VAAKVP	AASTTTIAL.
Cactus_virus	QTLSSSQLAA	LSLGVTSLL	PSPAELVSIS	QAL.TTLGAS	ATNLTPLSL.
PLANTAGO_ASI	TAPTADALAA	MAFPVSSPSV	PTAQELDTIT	SGL.TTLGVP	TDSLSSHAL.
TULIP_VIRUS	TAPNPEALAA	MTLEVSSPAV	PTPAELDTIA	AGL.TTLGVP	ADSLISHAL.
CASSAVA_COMM	TAPTDEELSR	LDLKPASNLV	ASADALSAIA	ADW.ASLKVP	TAQLMRHAL.

CLOVER_YELLO	SAPTDEQLDT	LTLTIESNLV	PSISELEAIA	KDW.KTLGLQ	EADFTANAI.
HOSTA_VIRUS_	TAPTQEQLTS	LALPIISTR	PSPDVLNQIS	VKW.QELGVP	TASISSTAI.
LILY_VIRUS_X	FVPDAKTWAD	TAYTAQSES	ATAEELQSIA	TLW.EGIGIP	AANFFDVAF.
STRAWBERRY_M	RVLTPPEELAA	.PISAASNKV	ATREQILGIV	AD.LNALGFV	GDP..ALGLF
POTATO_VIRUS	.IPDGDFFST	ARAVVASDAV	ATNEDLSEIE	AVW.KDMKVP	TDTMAQAAW.
PapMV_CLONE-	PAITQEQMSS	IKVDPTSPLL	PSQEQLKSVS	TLM.VAAKVP	AASVTTVAL.
Consensus	TAPTTLEDLAA	LKY--TSNSV	A-PEELEAIA	ALWIKALGVP	AASVTP-ALW

101

BAMBOO_MOSAI	.LSLEAFDRG	SSEATTWDG.	..ITEGVEHR	AAANAIKEAN	CPIHKVTTYLL
FOXTAIL_MOSA	.FTM...DKG	ATDSTIFTG.	..KYNTFPMK	SLALRCKDAG	VPVHKLCYFY
CYMBIDIUM_MO	DLARAYADVG	ASKSATLLGF	CPTKPDVRR	ALAGRSLWPT	SPPASFCAYY
POTATO_AUCUB	DMARAYADVG	ASRKAVLLDA	PTLAPTVAR	RLAQLKAGAG	ISPRQFCSYY
NARCISSUS_MO	DLARAYADVQ	ASRSAVLSGT	TPSNPAITRQ	ALARQFYVIN	ITPRQFCMYF
SCALLION_VIR	DLARAYADVQ	SSRSAMLGT	TPSNPAITRQ	ALARQFYIVN	ITPRQFCMYF
PEPINO_MOSAI	DLARAYADVQ	SSKSAQLIGA	TPSNPALSRR	ALAAQFDRIN	ITPRQFCMYF
WHITE_CLOVER	DLARAFADVG	ASSKSELTDG	SAALAGVSRK	QLA.QAIKIH	CTIRQFCMYF
Altranathera	ELVNFCYDNG	SSAYTVVVG	SS.LAEVSL	QVANIVKASG	TSLRKFCRFF
Cactus_virus	EIVNYCFDNG	SSPETVFKGD	ST.V..LQMP	SPKSPCHHPI	TTLRQFCRYF
PLANTAGO_ASI	ALVNACFDAG	SSSFVTLSPG	SP.TPTISLA	QIAGVVKVT.	TTLRQFCRYF
TULIP_VIRUS	ALVNACFDAG	SSQFTTISGP	SP.TPTITLA	QLAGVVKVS.	TTLRQFCRYF
CASSAVA_COMM	DLVNFCFDSG	SSKYTTVEGS	SP.TPTIPRA	ALAGAVRKH.	TTLRQFCRYF
CLOVER_YELLO	KIAWFCYHSG	SSSVQVQGN	ST.SDKIPLY	QLAGVVRQH.	STLRRFCRYF
HOSTA_VIRUS_	ALCMACYHSG	SSGSTLIPGL	AP.GTTVNYT	SLAAAVK.SL	ATLREFARYF
LILY_VIRUS_X	QLAMRCSGDH	ASSLTVLSGN	CTVAPTPTLK	AAAGLVKA.V	LPLRQFCRYF
STRAWBERRY_M	DLAFHCYDIG	SSPSAQVGP	SPF..GCSRM	QVAAVVR.NH	CTLRQFCMYF
POTATO_VIRUS	DLVRHCADVG	SSAQTEMIDT	GPYSNGISRA	RLAAAIKE.V	CTLRQFCMKY
PapMV_CLONE-	ELVNFCYDNG	SSAYTTVTGP	SS.IPEISLA	QLASIVKASG	TSLRKFCRYF
Consensus	DLARACADVG	SS-STTSGP	SPSTPTISRA	-LAG-VK-S-	TTLRQFCRY-

150

151

BAMBOO_MOSAI	AKPTFAIRQS	KNLPPANFAK	KNVPSQYKWC	AFDAFDGLYD	PTCL.ASELP
FOXTAIL_MOSA	TKPAYANRRV	ANQPPARWTN	ENVPKANKWA	AFDTFDALLD	PYVV.PSSVP
CYMBIDIUM_MO	AKVWNLMMLA	TNDPPANWAK	AGFQEDTRFA	AFDFFDAVDS	TAALEPAE.W
POTATO_AUCUB	AKIVWNMLH	KNEPPANWAK	IGFKEDYKFA	AFDFFDAVDS	PAALEPSQ.W
NARCISSUS_MO	AKVVWNLLLD	SNVPPAGWAK	QGLPDDCKFA	GDFDFEGVLS	PAALDPADGL
SCALLION_VIR	AKVVWNMMID	SNVPPAGWVK	HGLPEDCKFA	GFVDFEGVLS	PSSLDPADGL
PEPINO_MOSAI	AKVVWNILLD	SNIPPANWAK	LGYQEDTKFA	AFDFFDGVTN	PASLQPADGL
WHITE_CLOVER	ANVVWNIMLD	TKTPPASWSK	LGYKEESKFA	GDFDFDGVNH	PAALMPADGL
Altranathera	APIIWNLRTD	K.TPPANWEA	NGFKPTEKFA	AFDFFDGVEN	PAAMQPPGGL
Cactus_virus	AKIWNRYRS	KNLPPAAWEA	WAYKPEQKFA	AFDFFDGVLN	EAALNPIDGL
PLANTAGO_ASI	AKIWNARLA	RNLPPAGFAR	ANIKFEHRWA	GDFDFDGLLN	PAALEPPGGL
TULIP_VIRUS_	AKLIWNARLS	RNRPPAGFAR	AYVKTGQKWA	GDFDFDGLLN	PAALEPLGGL
CASSAVA_COMM	AKIWNARVK	ANIPPAGYAN	AHIKPEQAF	GDFDFDGMN	VAALEPSGGL
CLOVER_YELLO	AKVIWNYALR	KNQPPANWAS	QNYKEADRFA	AFDFFEVSS	SAALSPPGGL
HOSTA_VIRUS_	APIIWNYAIE	HKIPPANWAA	MGYKENTKYA	AFDTFDSILN	PAALQPTGGL
LILY_VIRUS_X	AKFVWNWRLS	HDLPPANWAD	SQFPAEARFA	AFDFFDGVTN	SAAPQPPDGL
STRAWBERRY_M	APSVWNKAVR	DNRPPGNWSN	LQFTPETKFA	AFDFFDGVLN	PASQEV..L
POTATO_VIRUS	APVVWNWMLT	NNSPPANWQA	QGFKPEHKFA	AFDFFDGVTN	PAAIMPKEGL
PapMV_CLONE-	APIIWNLRTD	K.MAPANWEA	SGYKPSAKFA	AFDFFDGVEN	PAAMQPPSGL
Consensus	AK-VWNLRLD	KNLPPANWAK	AG-KEETKFA	AFDFFDGVLN	PAALEPP-GL

200

201

BAMBOO_MOSAI	YDAPSEIDRM	ASATFKTIQI	KIANDQKGFN	L.NYNPNVTQ	ARLPNA....
FOXTAIL_MOSA	YDEPTPEDRQ	VNEIFKKDNL	SQAASRNQL.	L.GTQASITR	GRINGA....
CYMBIDIUM_MO	QRRPTDRERA	AHSIGKYGAL	ARQRIQ.NGG	LITNIAEVNQ	GPSWSTNTLN
POTATO_AUCUB	VRHPTDKERA	AHGVVKWASL	SRERLQ.EGT	SITTVaelNK	GHLGGYNNLP

250

```

NARCISSUS_MO IRPPSQREIQ AHSTAKYGAL ARQRYRMETS FPPWLKSLT. ....VGSVA
SCALLION_VIR IRHPSQREIQ AHSTAKYGAL ARQRIQ.NGN FVSNLAEVTH GRAGGVNSMY
PEPINO_MOSAI IRQPNEKELA AHSVAKYGAL ARQKI.STGN YITTLGEVTR GHMGGANTMY
WHITE_CLOVER IRGPSEAELL AHQTAKQVAL HRDAKRRGTN VVNSV.EITN GRSDPIGPLI
Altranathera VRAPSQAERI ANATNKQVNL FQAAAQDN.N FASNSAFITK GQL..SSNSP
Cactus_virus VRRPNEAERL ANQTNRNVHL FESNAQKN.R ALTTALVTK GLQ..GSESP
PLANTAGO_ASI SRTPTPDEV T ANETARSLNL FEARASYS.N LASTSTQFTR GQL..SNTAP
TULIP_VIRUS_ TREPTPDEIT ANETARSLGL FESRANSN.N LATTSTQFTR GQL..SNTSP
CASSAVA_COMM VREPTPQEII AAETARSLNL FEAQSKGN.N LATNATQVTR GRL..SSSEP
CLOVER_YELLOW IREPSPNERM ANETNKNVHL YQTASRGS.N LATTSTVATK GAY..STNAS
HOSTA_VIRUS_ IRQPTEEELL AHQANSALHI FDSLRL..N.D FASTDGRVTR GHI..TSNVN
LILY_VIRUS_X IRPPTELELS AAQTAKFAAL ..ARVRGS.G FVTTAAEITH GRAEVSRT..
STRAWBERRY_M WRQPTPQEII ASATHKDVAT YRAASKAHDR .ISNSTLLTK G..ASRSTPP
POTATO_VIRUS IRPPSEAEMN AAQTAAFVKI TKARAQSN.D FASLDAVTR GRITGTTTAE
PapMV_CLONE- TRSPTQEERI ANATNKQVHL FQAAAQDN.N FASNSAFITK GQI..SGSTP
Consensus IREPTAER- A--TAKYVAL FRARAQ-NGN -ATTSAEVTR GRLGGSNTAP

```

```

                251                                276
BAMBOO_MOSAI PLPALPEPTS D~~~~~
FOXTAIL_MOSA ..PALPNNQ YFIEAPQ~~~
CYMBIDIUM_MO ALPAP~~~~~
POTATO_AUCUB ALMAPP~~~~~
NARCISSUS_MO STPCTPLKHL QNCNRNTSKL KLVCGL
SCALLION_VIR AIEAPPEL~~~
PEPINO_MOSAI AIDAPPEL~~~
WHITE_CLOVER TYPQ~~~~~
Altranathera TIQYLPPPE~
Cactus_virus RIQFLPGPE~
PLANTAGO_ASI QVQFLPAPSD
TULIP_VIRUS_ TVQFLPSPED
CASSAVA_COMM QVQFLTGVDE
CLOVER_YELLOW NAGFPYHRPE
HOSTA_VIRUS_ SLNYLPAPEG SS~~~~~
LILY_VIRUS_X ..MLLSPP~~~
STRAWBERRY_M ALLPGP~~~~~
POTATO_VIRUS AVVTLPPP~~~
PapMV_CLONE- TIQFLPPPEH HHHHH~~~~~
Consensus A-Q-LPPE- -----

```

## Annexe 2. Milieux et tampons utilisés

### *Clonage :*

- Gel d'agarose 0,8 % (1 L) :
  - 8g d'agarose
  - 40µl de bromure d'éthidium
  
- Bleu à ADN 3X :
  - 40% glycérol
  - 0,25% bleu de bromophénol
  
- Milieu LB (1 L) :
  - 10g de peptone
  - 5g d'extrait de levure
  - 10g de NaCl
  - 900 ml d'eau
  - Le pH 7,0

### *Production de protéines :*

- Milieu 2XYT (1 L) :
  - 16g de peptone
  - 10g d'extrait de levure
  - 5g de NaCl
  - 800 ml d'eau distillée
  - Le pH 7,0
  
- Milieu agar 2XYT avec ampicilline :
  - Milieu 2XYT liquide
  - 20g d'agar
  - ampicilline 50 µg/ml \*

\* l'ampicilline est ajoutée après l'autoclavage, lorsque le milieu est tiède

### *Purification de protéines :*

- Tampon de lyse (Lysis Buffer) :
  - 50 mM NaH<sub>2</sub>PO<sub>4</sub>
  - 300 mM NaCl
  - 10 mM imidazole
  - pH 8,0

- Tampon de lavage 1 : 50 mM NaH<sub>2</sub>PO<sub>4</sub>  
300 mM NaCl  
20 mM imidazole  
pH 8,0
- Tampon de lavage 2 : 50 mM NaH<sub>2</sub>PO<sub>4</sub>  
300 mM NaCl  
50 mM imidazole  
pH 8,0
- Tampon de lavage 3 : 10 mM Tris-HCl  
50 mM imidazole  
pH 8,0
- Tampon d'élution : 10 mM Tris-HCl  
1 M imidazole  
pH 8,0
- Tampon d'élution : 10 mM Tris-HCl  
300 mM imidazole  
pH 8,0

*Gels Shifts (EMSA) :*

- Tampon de liaison : 10 mM Tris-HCl  
4 % glycérol  
1 mM MgCl<sub>2</sub>  
0,5 mM EDTA  
0,5 mM DTT  
20 mM NaCl  
pH 8,0
- Tampon TBE 5 X (pour 1 L) : 54 g base Tris  
27,5 g d'acide borique  
4,65 g EDTA  
pH 8,3
- Tampon bleu natif 3X : 50 % glycérol  
0,3 % bleu de bromophénol  
TBE 0,5X
- Gel de polyacrylamide natif 5% (pour 10 ml) :  
  
1 ml de tampon TBE 5X  
1,25 ml de mélange d'acrylamide pour gel natif\*

7,75 ml d'eau  
 50 µl de persulfate d'ammonium 10 %  
 10 µl de TEMED

\*Solution de mélange d'acrylamide pour gel natif : 38,7 % d'acrylamide  
 1,3 % de bisacrylamide

*Préparation de gels de polyacrylamide tris-tricine (2 gels):*

- Gel d'empilement :
  - 0,5 ml de mélange d'acrylamide\*\*
  - 1,55 ml de tampon de gel\*\*\*
  - 4,2 ml d'eau
  - 125 µl de persulfate d'ammonium 10 %
  - 12 µl de TEMED
  
- Gel de séparation 10% :
  - 3,05 ml de mélange d'acrylamide\*\*
  - 5 ml de tampon de gel\*\*\*
  - 1,6 ml de glycérol
  - 5,35 ml d'eau
  - 125 µl de persulfate d'ammonium 10 %
  - 12 µl de TEMED

\*\* Solution de mélange d'acrylamide : 49,5 % d'acrylamide  
 3 % de bisacrylamide

\*\*\* Solution de tampon de gel : 3 M Tris-HCl pH 8,45  
 0,3 % SDS

- Bleu dénaturant à protéine 3X :
  - 6% SDS
  - 50% glycérol
  - 0,2M β-mercaptoéthanol
  - 0,3% bleu de bromophénol
  
- Solution de décoloration pour gel de protéine :
  - 360 ml 99% éthanol
  - 360 ml d'eau
  - 80 ml d'acide acétique glacial

## Annexe 3. Séquences des clones

### PapMV CP<sub>(6-215)</sub>

#### *Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCTTCCCCGCCATCACCCAGGAACAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTCCGGGACTAACCAGGTCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCCGAACACCATCACCATCACCATTAG

#### *Séquence en acides aminés*

MASTPNIAFP AITQEQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVKASGTSR  
 KFCRYFAPII WNLRTDKMAP ANWEASGYKP SAKFAAFDFD DGVENPAAMQ  
 PPSGLTRSPT QEERIANATN KQVHLFQAAA QDNNFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH



**PapMV CP<sub>(13-215)</sub>***Séquence en acides nucléiques*

ATGTTCCCCGCCATCACCCAGGAACAAATGAGCTCGATTAAGGTCGATCCAACGTCCAATC  
 TTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCTCATGGTAGCTGCTAAGGTTCC  
 AGCAGCCAGTGTTACAACGTGGCATTGGAGTTGGTTAACTTCTGCTATGACAATGGGTCC  
 AGCGGTACACCACAGTGACTGGCCCATCATCAATACCGGAGATATCACTGGCACAATTGG  
 CCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATTCTGCCGGTACTTCGCGCCAAT  
 AATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAATTGGGAGGCCTCAGGATACAAG  
 CCAAGCGCAAATTTGCCGCGTTCGACTTCTTCGACGGGGTGGAGAATCCGGCGGCCATGC  
 AACCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGAGCGGATTGCCAATGCCACCAA  
 CAAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAACAACCTTGCCAGCAACTCCGCC  
 TTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCATCCAATTCCTTCCACCCCCG  
 AACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MFPAITQEQM SSIKVDPTSN LLPSQEQLKS VSTLMVAAKV PAASVTTVAL  
 ELVNFCYDNG SSAYTTVTGP SSIPEISLAQ LASIVKASGT SLRKFCRYFA  
 PIIWNLRTDK MAPANWEASG YKPSAKFAAF DFFDGVENPA AMQPPSGLTR  
 SPTQEERIAN ATNKQVHLEQ AAAQDNNFAS NSAFITKGQI SGSTPTIQFL  
 PPPEHHHHHH

**PapMV CP<sub>(14-215)</sub>***Séquence en acides nucléiques*

ATGCCCCGCATCACCCAGGAACAAATGAGCTCGATTAAGGTCGATCCAACGTCCAATCTTC  
 TGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCTCATGGTAGCTGCTAAGGTTCCAGC  
 AGCCAGTGTTACAACGTGGCATTGGAGTTGGTTAACTTCTGCTATGACAATGGGTCCAGC  
 GCGTACACCACAGTGA CTGGCCCATCATCAATACCGGAGATATCACTGGCACAATTGGCCA  
 GCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATTCTGCCGGTACTTCGCGCCAATAAT  
 CTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAATTGGGAGGCCTCAGGATAACAAGCCA  
 AGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGGTGGAGAATCCGGCGGCCATGCAAC  
 CCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGAGCGGATTGCCAATGCCACCAACAA  
 ACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAACA ACTTTGCCAGCAACTCCGCCTTC  
 ATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCATCCAATTCCTTCCACCCCCGAAC  
 ACCATCACCATCACCATTAG

*Séquence en acides aminés*

MPAITQEOMS SIKVDPTSNL LPSQEQLKSV STLMVAAKVP AASVTTVALE  
 LVNFCYDNGS SAYTTVTGPS SIPEISLAQL ASIVKASGTS LRKFCRYFAP  
 IIWNLRTDKM APANWEASGY KPSAKFAAFD FFDGVENPAA MQPPSGLTRS  
 PTQEERIANA TNKQVHLFQA AAQDNNFASN SAFITKGQIS GSTPTIQFLP  
 PPEHHHHHH

**PapMV CP<sub>(15-215)</sub>***Séquence en acides nucléiques*

ATGGCCATCACCCAGGAACAAATGAGCTCGATTAAGGTCGATCCAACGTCCAATCTTCTGC  
 CCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCTCATGGTAGCTGCTAAGGTTCCAGCAGC  
 CAGTGTTACAACCTGTGGCATTGGAGTTGGTTAACTTCTGCTATGACAATGGGTCCAGCGCG  
 TACACCACAGTGACTGGCCCATCATCAATACCGGAGATATCACTGGCACAATTGGCCAGCA  
 TTGTCAAAGCTTCGGGCACTTCCCTTAGGAAATTCTGCCGGTACTTCGCGCCAATAATCTG  
 GAATCTGAGGACGGACAAAATGGCTCCTGCCAATTGGGAGGCCTCAGGATACAAGCCAAGC  
 GCCAAATTTGCCGCGTTCGACTTCTTCGACGGGGTGGAGAATCCGGCGGCCATGCAACCCC  
 CTTCCGGGACTAACCAGGTCGCCGACCCAGGAAGAGCGGATTGCCAATGCCACCAACAAACA  
 GGTGCATCTCTTCCAAGCCGCGGCACAGGACAACAACCTTTGCCAGCAACTCCGCCTTCATC  
 ACCAAAGGCCAAATTTCTGGGTCAACCCCAACCATCCAATTCCTTCCACCCCCGAACACC  
 ATCACCATCACCATTAG

*Séquence en acides aminés*

MAITQEQMSS IKVDPTSNLL PSQEQLKSVS TLMVAAKVPA ASVTTVALEL  
 VNFCYDNGSS AYTTVTGPSS IPEISLAQLA SIVKASGTSL RKFCRYFAPI  
 IWNLRDCKMA PANWEASGYK PSAKFAAFDF FDGVENPAAM QPPSGLTRSP  
 TQEERIANAT NKQVHLFQAA AQDNNFASNS AFITKGQISG STPTIQFLPP  
 PEHHHHHH

**PapMV CP<sub>(21-215)</sub>***Séquence en acides nucléiques*

ATGAGCTCGATTAAGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGT  
 CAGTGTCCACCCTCATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGCATT  
 GGAGTTGGTTAACTTCTGCTATGACAATGGGTCCAGCGCGTACACCACGGTGACTGGCCCA  
 TCATCAATACCGGAGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGGCACTT  
 CCCTTAGGAAATTCTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAAT  
 GGCTCCTGCCAATTGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGAC  
 TTCTTCGACGGGGTGGAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTGCG  
 CGACCCAGGAAGAGCGGATTGCCAATGCCACCAACAACAGGTGCATCTCTTCCAAGCCGC  
 GGCACAGGACAACAACCTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGG  
 TCAACCCCAACCATCCAATTCCTTCCACCCCCGAACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MSSIKVDPTS NLLPSQEQLK SVSTLMVAAK VPAASVTTVA LELVNFCYDN  
 GSSAYTTVTG PSSIPEISLA QLASIVKASG TSLRKFCRYF APIIWNLRD  
 KMAPANWEAS GYKPSAKFAA FDFFDGVENP AAMQPPSGLT RSPTQEERIA  
 NATNKQVHLF QAAAQDNNFA SNSAFITKGQ ISGSTPTIQF LPPPEHHHHH  
 H

**PapMV CP<sub>(27-215)</sub>***Séquence en acides nucléiques*

ATGGCCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCC  
TCATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACCTGTGGCATTGGAGTTGGTTAA  
CTTCTGCTATGACAATGGGTCCAGCGGTACACCACGGTGACTGGCCCATCATCAATACCG  
GAGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAAT  
TCTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAA  
TTGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGG  
GTGGAGAATCCGGCGGCCATGCAACCCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAG  
AGCGGATTGCCAATGCCACCAACAAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAA  
CAACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACC  
ATCCAATTCCTTCCACCCCCGAACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MADPTSNLLP SQEQLKSVST LMVAAKVPAA SVTTVALELV NFCYDNGSSA  
YTTVTGPSSI PEISLAQLAS IVKASGTSR KFCRYFAPII WNLRTDKMAP  
ANWEASGYKP SAKFAAFDFD DGVENPAAMQ PPSGLTRSPT QEERIANATN  
KQVHLFQAAA QDNNFASNSA FITKGQISGS TPTIQFLPPP EHHHHHH

**PapMV CP<sub>(14-180)</sub>***Séquence en acides nucléiques*

ATGCCCCGCCATCACCCAGGAACAAATGAGCTCGATTAAGGTCGATCCAACGTCCAATCTTC  
TGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCTCATGGTAGCTGCTAAGGTTCCAGC  
AGCCAGTGTTACAACCTGTGGCATTGGAGTTGGTTAACTTCTGCTATGACAATGGGTCCAGC  
GCGTACACCACAGTGACTGGCCCATCATCAATACCGGAGATATCACTGGCACAATTGGCCA  
GCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATTCTGCCGGTACTTCGCGCCAATAAT  
CTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAATTGGGAGGCCTCAGGATACAAGCCA  
AGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGGTGGAGAATCCGGCGGCCATGCAAC  
CCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGAGCGGATTGCCAATGCCACCAACAA  
ACAGGTGCATCTCTTCCACCATCACCATCACCATTAG

*Séquence en acides aminés*

MPAITQEQMS SIKVDPTSNL LPSQEQKSV STLMVAAKVP AASVTTVALE  
LVNFCYDNGS SAYTTVTGPS SIPEISLAQL ASIVKASGTS LRKFCRYFAP  
IIWNLRTDKM APANWEASGY KPSAKFAAFD FFDGVENPAA MQPPSGLTRS  
PTQEERIANA TNKQVHLFHH HHHH

**PapMV CP<sub>27-180</sub>***Séquence en nucléotides*

ATGGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCTCA  
TGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACCTGTGGCATTGGAGTTGGTTAACTT  
CTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTGGCCCATCATCAATACCGGAG  
ATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCGGGCACTTCCCTTAGGAAATTCT  
GCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAATTG  
GGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGGTG  
GAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGAGC  
GGATTGCCAATGCCACCAACAAACAGGTGCATCTCTTCCACCATCACCATCACCATTAG

*Séquence en acides aminés*

MDPTSNLLPSQEQQLKSVSTLMVAAKVPAASVTTVALELVNFCYDNGSSAYTTVTGPSSIPE  
ISLAQLASIVKASGTSLRKFCRYFAPIIWNLRDVKMAPANWEASGYKPSAKFAAFDFFDGV  
ENPAAMQPPSGLTRSPTQEERIANATNKQVHLFHHHHH

**PapMV F13A***Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCGCGCCCGCCATCACCCAGGAACAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGAAGTGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTGCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCGAACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MASTPNIAAP AITQEQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVKASGTSR  
 KFCRYFAPII WNLRTDKMAP ANWEASGYKP SAKFAAFDFD DGVENPAAMQ  
 PPSGLTRSPT QEERIANATN KQVHLEQAAA QDNNFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH



**PapMV F13G***Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCGGTCCCGCCATCACCCAGGAACAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTIONGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAAACAGGTGCATCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCGAACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MASTPNIAGP AITQEQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVKASGTSLR  
 KFCRYFAPII WNLRTDKMAP ANWEASGYKP SAKFAAFDFE DGVENPAAMQ  
 PPSGLTRSPT QEERIANATN KQVHLFQAAA QDNNFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH

**PapMV F13L***Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCCTGCCCGCCATCACCCAGGAACAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCGAACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MASTPNIALP AITQEQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVKASGTSR  
 KFCRYFAPII WNLRTDKMAP ANWEASGYKP SAKFAAFDFE DGVENPAAMQ  
 PPSGLTRSPT QEERIANATN KQVHLFQAAA QDNNFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH

**PapMV F13Y***Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCTATCCCGCCATCACCCAGGAACAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCGAACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MASTPNIAYP AITQEQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVKASGTSR  
 KFCRYFAPII WNLRTDKMAP ANWEASGYKP SAKFAAFDEF DGVENPAAMQ  
 PPSGLTRSPT QEERIANATN KQVHLFQAAA QDNNEFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH

**PapMV E19K***Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCTTCCCCGCCATCACCCAGAAACAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCCGAACCCATCACCATCACCATTAG

*Séquence en acides aminés*

MASTPNIAFP AITQKQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVKASGTSLR  
 KFCRYFAPII WNLRTDKMAP ANWEASGYKP SAKFAAFDFE DGVENPAAMQ  
 PPSGLTRSPT QEERIANATN KQVHLFQAAA QDNNFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH

**PapMV E19P***Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCTTCCCCGCCATCACCCAGCCGCAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCCGAACCCATCACCATCACCATTAG

*Séquence en acides aminés*

MASTPNIAFP AITQPQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVKASGTSLR  
 KFCRYFAPII WNLRTDKMAP ANWEASGYKP SAKFAAFDFF DGVENPAAMQ  
 PPSGLTRSPT QEERIANATN KQVHLFQAAA QDNNFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH

**PapMV K97A***Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCTTCCCCGCCATCACCCAGGAACAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTGTTCGCGGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGAGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTTCGGGACTAACCAGGTCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCCGAACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MASTPNIAFP AITQEQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVAASGTSLR  
 KFCRYFAPII WNLRTDKMAP ANWEASGYKP SAKFAAFDFE DGVENPAMQ  
 PPSGLTRSPT QEERIANATN KQVHLFQAAA QDNNEFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH

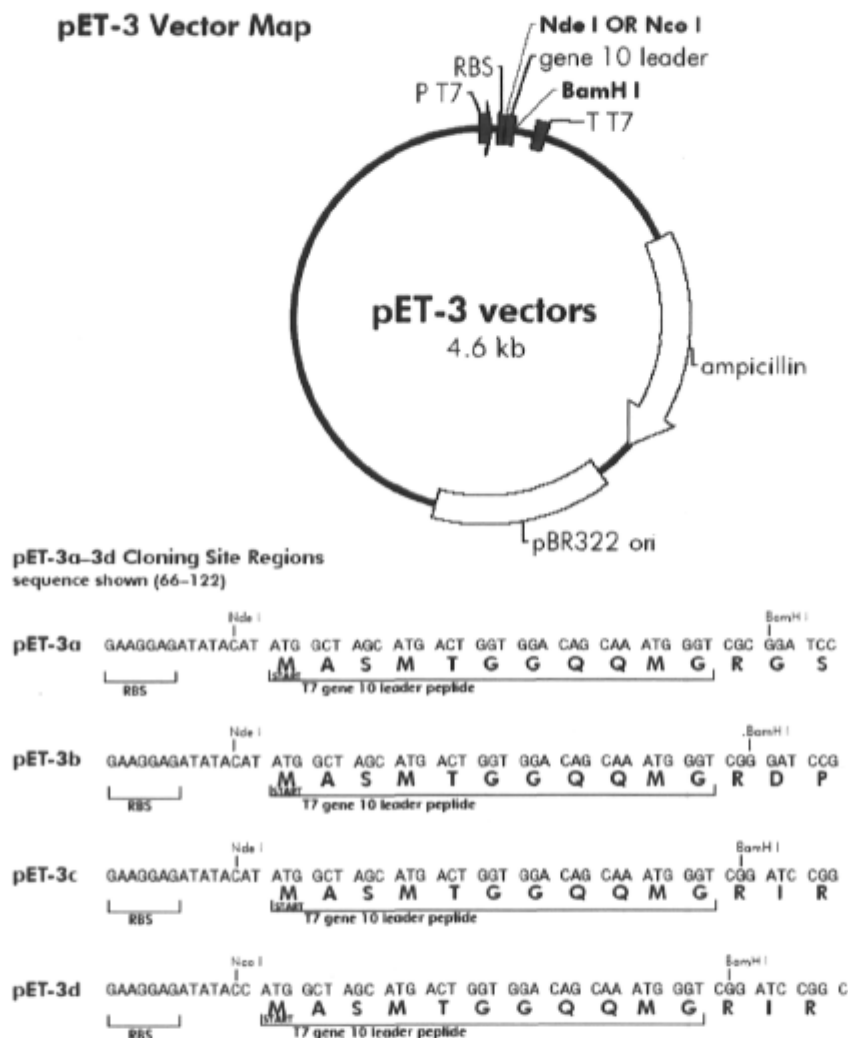
**PapMV E128A***Séquence en acides nucléiques*

ATGGCATCCACACCCAACATAGCCTTCCCCGCCATCACCCAGGAACAAATGAGCTCGATTA  
 AGGTCGATCCAACGTCCAATCTTCTGCCCTCCCAAGAGCAGTTAAAGTCAGTGTCCACCCT  
 CATGGTAGCTGCTAAGGTTCCAGCAGCCAGTGTTACAACGTGGGCATTGGAGTTGGTTAAC  
 TTCTGCTATGACAATGGGTCCAGCGCGTACACCACAGTGACTGGCCCATCATCAATACCGG  
 AGATATCACTGGCACAATTGGCCAGCATTGTCAAAGCTTCCGGCACTTCCCTTAGGAAATT  
 CTGCCGGTACTTCGCGCCAATAATCTGGAATCTGAGGACGGACAAAATGGCTCCTGCCAAT  
 TGGGCGGCCTCAGGATACAAGCCAAGCGCCAAATTTGCCGCGTTCGACTTCTTCGACGGGG  
 TGGAGAATCCGGCGGCCATGCAACCCCTTCGGGACTAACCAGGTCGCCGACCCAGGAAGA  
 GCGGATTGCCAATGCCACCAACAAACAGGTGCATCTCTTCCAAGCCGCGGCACAGGACAAC  
 AACTTTGCCAGCAACTCCGCCTTCATCACCAAAGGCCAAATTTCTGGGTCAACCCCAACCA  
 TCCAATTCCTTCCACCCCCCGAACACCATCACCATCACCATTAG

*Séquence en acides aminés*

MASTPNIAFP AITQEQMSSI KVDPTSNLLP SQEQLKSVST LMVAAKVPAA  
 SVTTVALELV NFCYDNGSSA YTTVTGPSSI PEISLAQLAS IVKASGTSR  
 KFCRYFAPII WNLRTDKMAP ANWAASGYKP SAKFAAFDFE DGVENPAAMQ  
 PPSGLTRSPT QEERIANATN KQVHLFQAAA QDNNFASNSA FITKGQISGS  
 TPTIQFLPPP EHHHHHH

## Annexe 4. Carte Vecteur d'expression pET-3d



Feature	Nucleotide Position			
	pET-3a	pET-3b	pET-3c	pET-3d
T7 promoter	1-19	1-19	1-19	1-19
ribosome binding site (RBS)	66-72	66-72	66-72	66-72
Nde I (pET-3a-c) or Nco I (pET-3d) cloning site	78-83	78-83	78-83	78-83
T7 gene 10 translated leader	81-113	81-113	81-113	80-112
BamH I cloning site	117-122	116-121	115-120	114-119
T7 terminator	191-237	190-236	189-235	188-234
ampicillin resistance (bla) ORF	840-1697	839-1696	838-1695	837-1694
pBR322 origin of replication	1848-2515	1847-2514	1846-2513	1845-2512

FIGURE 1 The pET-3 vectors



## Annexe 5. Prédications de structure secondaire

### Prédiction CP<sub>6-215</sub>

AA	10	20	30	40	50	60
Séquence	MASTPNIAFPAITQEQMSSIKVDPTSNNLLPSQEQLKSVSTLMVAAKVPAAASVTTVALELV					
PROF_sec	LLLLLLLLLLLLLHHH...LLL.LLLLHHHHHHHHHHH.LLLL.HHHHHHHHHH					
	987776766788857774224675545568888999899898735888145678899998					
PSIPREP	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	998867767777889997348887748899999999999999984988568999999999					
Sable-2	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	788987888887889999769987777789989999999999977999567999999999					
Consensus	CCCCCCCCCCCCCHHHHhcCCCcccCCCCCHHHHHHHHHHHHhCCCchHHHHHHHHHH					
AA	70	80	90	100	110	120
Séquence	NFCYDNGSSAYTTVTGPSSIPEISLAQLASIVKASGTSLRKFCRYFAPIIWNLRDVKMAP					
PROF_sec	HHH.LLL.....LLLLL.LLHHHHHHHHH...L.HHHHHHHHHHH.....LLL.					
	875236665221223566666456788988886204536888898878665444205674					
PSIPREP	HHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHH					
	998278885444223224888841999999987628889998898889986698636898					
Sable-2	HHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHH					
	987688887767788898888777999999998768988999999889999987678995					
Consensus	HHHccCCcccccccccccCCCccHHHHHHHHHHhCCCHHHHHHHHHHHhhHHhccCCc					
AA	130	140	150	160	170	180
Séquence	ANWEASGYKPSAKFAAFDFDGVENPAAMQPPSGLTRSPTQEERIANATNKQVHLFQAAA					
PROF_sec	.....LLL.....LL...LLLLL...LLHHHHHHH.....HHH...					
	012202788542021001101246742255655632478868887654333334555431					
PSIPREP	HHHHHCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHH					
	56886598710201333202676687524887455569798999999997893999975					
Sable-2	HHHHHCCCCCHHEHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH					
	889985899655434444435567877778888777899889999999999885788874					
Consensus	hHHHHcCCcchhhccccCCCcccCCccccCCCHHHHHHHHHHHhhHHHHHH					
AA	190	200	210	220		
Séquence	QDNNFASNSAFITKGQISGSTPTIQFLPPPEHHHHHH					
PROF_sec	..LL.....L..LLL.....LLL...LLL					
	0465333320220451267554000037863234578					
PSIPREP	CCCCCHHHHHHCHHCCCCCCCCCCCCCCCCCCCC					
	01885122887510010789876658889832245689					
Sable-2	CCCCCCCCCHHCCCCCCCCCCCCCCCCCCCC					
	678866764444357778998864457898877777					
Consensus	ccCCcccchhhccccCCCcccCCccccCC					

Prédiction CP<sub>13-215</sub>

AA	20	30	40	50	60	70
Séquence	MFPAITQEQMSSIKVDPTS	NLLPSQEQ	QLKSVSTLMVAAKVPAASVTTVALELVNFCYDNG			
PROF_sec	LLLLL.HHH...LLL...LLL	HHHHHHHHHHHHHHHHHH	LLL...HHHHHHHHHHHHHHHH	LL		
PSIPREP	CCCCCHHHHHHCCCCCCCC	CCCCCHHHHHHHHHHHHHHHHH	CCCCCHHHHHHHHHHHHHHHHH	CCCC		
Sable-2	CCCCCHHHHHHCCCCCCCC	CCCCCHHHHHHHHHHHHHHHHH	CCCCCHHHHHHHHHHHHHHHHH	CCCC		
Consensus	CCCCCHHHHHhC	CCcCCCC	CCCHHHHHHHHHHHHHHHhC	CCCh	HHHHHHHHHHHHHHHH	CC

AA	80	90	100	110	120	130
Séquence	SSAYTTVTGPSSIPEISLAQLASIVKASGTS	SLRKFCRYFAP	IWNLR	TDKMAPANWEASG		
PROF_sec	LL.....LLLLL..L	HHHHHHHHHH...L	HHHHHHHHHHHHHHHHHH	LL	.....L	
PSIPREP	CCCCCCCCCCCCCCCC	CHHHHHHHHHHH	CCCCCHHHHHHHHHHHHHHHHH	CCCC	CHHHHH	HCC
Sable-2	CCCCCCCCCCCCCCCC	CHHHHHHHHHHH	CCCCCHHHHHHHHHHHHHHHHH	CCCC	CHHHHH	HCC
Consensus	CccccccccC	CCcChHHHHHHHHhC	CCCHHHHHHHHHHHHHHHHHhC	CCCh	HHHHH	hCC

AA	140	150	160	170	180	190
Séquence	YKPSAKFAAFDFFDGVENPAAMQPPSGL	TRSPTQEERIANATNKQVHLFQAAAQDNNFAS				
PROF_sec	LLL.....LL...LLLLL..LLL	HHHHHHHH...HHH...LL...				
PSIPREP	CCHHHHHHHHHH	CCCCCCCCCCCCCCCC	CCCCCHHHHHHHHHHHHHHHHH	CCCC		
Sable-2	CCCCCHHEEEEE	CCCCCCCCCCCCCCCC	CCCCCHHHHHHHHHHHHHHHHH	CCCC		
Consensus	CCcchheeeee	CCCCC	CCCCCC	CCCHHHHHHHHHHHHHHHHHh	HHHHHh	CCcC

AA	200	210	220
Séquence	NSAFITKGQISGSTPTIQFLPPPEHHHHHH		
PROF_sec	.....L..LL.....LL...LL		
PSIPREP	HHHHHHHHHH	CCCCCCCCCCCCCCCC	
Sable-2	CCHHHCCCCCCCC	CCCCCCCCCCCCCCCC	
Consensus	cchhhcccc	CCCCcccc	CCccccCC

**Prédiction CP<sub>14-215</sub>**

AA	20	30	40	50	60	70
Séquence	MPAITQEQMSSIKVDPTSNNLPSQEQLKSVSTIMVAAKVPAASVTTVALELVNFCYDNGS					
PROF_sec	LLLLLHHHH...LLL...LLLHHHHHHHHHHHH.LLLL.HHHHHHHHHHHHH.LLL					
	988885677432567543455888899989989872588813567889999887523566					
PSIPREP	CCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHCCCC					
	97889899973588877688999999999999999849885689999999999827888					
Sable-2	CCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHCCCC					
	778668899976998777878899999999999997799956799999999998758888					
Consensus	CCCCCHHHHHhCccCccccCCHHHHHHHHHHHhCChhHHHHHHHHHHhccCCC					

AA	80	90	100	110	120	130
Séquence	SAYTTVTGPPSSIPEISLAQLASIVKASGTSLRKFCRYFAPIIWNLR'VDKMAPANWEASGY					
PROF_sec	L.....LLLLL..LHHHHHHHH...L.HHHHHHHHHHHHH...LLL.....LL					
	522122366666544578898898630453688889887766555420667401220378					
PSIPREP	CCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCCCHHHHHCCC					
	444322322478884199999987652888999889988998669874679845887598					
Sable-2	CCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCCCHHHHHCCC					
	776778888888887799999999876998899999988999998877899588998589					
Consensus	ccccccccCCCCcCCHHHHHHHHHhCChHHHHHHHHHHHHHHHHhccCChhHHhCCC					

AA	140	150	160	170	180	190
Séquence	KPSAKFAAFDFFDGVENPAAMQPPSGLTRSP'QEERIANATNKQVHLFQAAAQDNNFASN					
PROF_sec	LL.....LL...LLLLL...LLHHHHHHH.....HHHHH...LL....					
	854112100110124674225565563247886888765433444556653104653334					
PSIPREP	CHHHHHHHHHHCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHCCCCCCH					
	71020133320156568764689701356989899999999789499997501885123					
Sable-2	CCCCEEEEEECCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHCCCCCCCC					
	96544344443455678777788888778998899999999998788887567887676					
Consensus	CcccccccccCCCCcCCCCCCCCCCHHHHHHHHHHHHHhHHHHhccCCCC					

AA	200	210	220
Séquence	SAFITKGQISGSTPTIQFLPPPEHHHHHH		
PROF_sec	.....L..LL.....LL....LL		
	30220451166544000037864234678		
PSIPREP	HHHHCHHCCCCCCCCCCCCCCCCCCCC		
	88851001078987665789832245689		
Sable-2	CHHECCCCCCCCCECCCCCCCCCCCC		
	4433457678998754468898777777		
Consensus	chhccccCCCCccecCCCCccccCC		

**Prédiction CP<sub>15-215</sub>**

AA	20	30	40	50	60	70
Séquence	MAITQEQMSSIKVDPTSNNLLPSQEQLKSVSTLMVAAKVPAAASVTTVALELVNFCYDNGSS					
PROF_sec	LLLLHHHHHHLL					
PSIPREP	988857775315675434558888999899898725888145678899998875235665					
Sable-2	CCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHCCCCC					
Consensus	98787899726888776889999999999999998498856899999999998278885					
	77648889876999877778899999999999987999477999999999987588887					
	CCCCHHHHHhCccCccccCCCHHHHHHHHHHHHHhCChHHHHHHHHHHHHhccCCCC					

AA	80	90	100	110	120	130
Séquence	AYTTVTGPPSSIPEISLAQLASIVKASGTSLRKFCRYFAPIIWNLRDTPKMAPANWEASGYK					
PROF_sec	LLLLLLLLLLLLLLLLHHHHHHHHHHHHLLLLHHHHHHHHHHHHHHHHHHHHLLLLLHHHHLLLL					
PSIPREP	221223666665445788988886304526888898877665454205674012203788					
Sable-2	CCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHCCCC					
Consensus	434323224778821999999876628889998898889987798636898568965987					
	76778888888877799999998768988999998889999987678994889985899					
	ccccccccCCCCcchHHHHHHHHHHhCChHHHHHHHHHHHHHHHHHHhccCCChHHHhCCC					

AA	140	150	160	170	180	190
Séquence	PSAKFAAFDFFDGVENPAA MQPPSGLTRSPTQEERIANATNKQVHLFQAAAQDNNFASNS					
PROF_sec	LLLLEEHHHHLL					
PSIPREP	541111001100146743355655633578868887654344445566531046533343					
Sable-2	HHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHCCCCCHH					
Consensus	102113442104656876469973466797989999999998895999975018851128					
	CCCHEEEEECCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCC					
	655434444345667877778888877899889999999999986888876579876764					
	ccccheeeeccccCccccCCCCccccCChHHHHHHHHHHHHhHHHHHhccCCcchh					

AA	200	210	220
Séquence	AFITKGQISGSTPTIQFLPPPEHHHHHH		
PROF_sec	LELLLLLLLLLLLLLEEEELLLLLLLLLL		
PSIPREP	0220451167544010037863234578		
Sable-2	HHHHHHHHCCCCCCCCCCCCCCCCCCCC		
Consensus	8751012078987644789843245689		
	HHCCCCCCCCCCCCCCCCCCCCCCCC		
	4444577789987644678987777777		
	hhhccccCCCCccecCccCccccCC		

Prédiction CP<sub>21-215</sub>

AA	30	40	50	60	70	80	
Séquence	MSSIKVDPTS	NLLPSQEQ	LKSVSTLM	VAAKVPAAS	VTTVALEL	VNFCYDNG	SSAYTTV
PROF_sec	LLLLLLLLL	LLLLLLLLL	LLLLLLLLL	LLLLLLLLL	LLLLLLLLL	LLLLLLLLL	LLLLLLLLL
	96211143	43455888	99998998	98725888	1246688	999988	75225665
PSIPREP	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC
	9642344	5658899	9999999	9884988	6689999	9999999	82788854
Sable-2	CCECCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC
	5445767	777888	9999999	9779995	7799999	9999998	7588887
Consensus	CcccCC	ccccCC	CCCCCCCC	CCCCCc	hCCCCCCCC	CCCCCCCC	CcCCCCCCCC

AA	90	100	110	120	130	140
Séquence	PSSIPEISL	AQLASIV	KASGTS	LRKFCRY	FAPIIWN	LRTDKMA
PROF_sec	LLLLLLLL	LLLLLLLL	LLLLLLLL	LLLLLLLL	LLLLLLLL	LLLLLLLL
	6666644	6788988	8863035	368888	988887	6545420
PSIPREP	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC
	2588884	1999998	7553888	99988	9888998	7698636
Sable-2	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC
	9898877	7999999	9876998	89999	9887788	9588998
Consensus	ccCCCC	cC	CCCCCCCC	CCCCCc	hCCCCCCCC	CcCCCCCCCC

AA	150	160	170	180	190	200
Séquence	FDFFDGVEN	PAAMQP	PSGLTRS	PTQEERIA	NATNKQV	HLFQAAAQ
PROF_sec	HHHHLL	LLLLLLLL	LLLLLLLL	LLLLLLLL	LLLLLLLL	EEEEEEEE
	0010124	674225	565632	478868	887654	334445
PSIPREP	HHCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC
	320156	66876	458971	11345	98989	999998
Sable-2	EECCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC	CCCCCCCC
	44445	56787	777888	87789	99899	99998
Consensus	eecccc	Ccccc	CCCCCCCC	CCCCCc	hCCCCCCCC	CcCCCCCCCC

AA	210	220
Séquence	ISGSTPTIQ	FLPPE
PROF_sec	LLLLLLEE	LLLLLLLL
	1665430	100378
PSIPREP	HCCCCC	CCCCCCCC
	078986	65478
Sable-2	CCCCCCE	CCCCCCCC
	789988	644578
Consensus	cCCCCc	ceecCC

Prédiction F13A

AA	10	20	30	40	50	60
Séquence	MASTPNIAAPAITQEQMSSSIKVDPTSNNLLPSQEQLKSVSTLMVAAKVPAASVTTVALELV					
PROF_sec	LLLLLLLLLLLLLHHHH...LLLLL.LLLLLHHHHHHHHHHHHHHHH,LLL...HHHHHHHHHH					
	987777777788857874225776545568888999899898735888135678899998					
PSIPREP	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	988867777788898999635887563889999999999999999983988568999999999					
Sable-2	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	6788878889888999998769997777888999999999999999879995678999999999					
Consensus	CCCCCCCCCCCCCHHHHhCCcCCCCCCHHHHHHHHHHHHHhCCcCchHHHHHHHHHH					

AA	70	80	90	100	110	120
Séquence	NFCYDNGSSAYTTVTGPPSSIPEISLAQLASIVKASGTSLRKFCRYFAPIIWNLRDTPKMAP					
PROF_sec	HHH...LLL...LL...LHHHHHHHH...L.HHHHHHHHHHH.H...LLL.					
	875235665211223666666446788988886204536888898877665454206674					
PSIPREP	HHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCC					
	998278885443112124888832999999887628889998898889986698637898					
Sable-2	HHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCC					
	987588887767788898888767999999998769988999998889999987578995					
Consensus	HHhCCcCCCCccccCCCCcCHHHHHHHHhCCCHHHHHHHHHHHHHHHHHhCCcCC					

AA	130	140	150	160	170	180
Séquence	ANWEASGYKPSAKFAAFDFFDGVENPAAMQPPSGLTRSPQTQEERIANATNKQVHFLFOAAA					
PROF_sec	.....LLL.....LL...LLLLL...LLLHHHHHHH.....HHHH..					
	012202788541121001101246742255655632478868887654334444566531					
PSIPREP	HHHHHCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHH					
	45887598710201333201566687646897346669898999999999999996999975					
Sable-2	HHHHHCCCCCCEEEEECCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH					
	8899858996553344443355678777888887789988999999999999987888875					
Consensus	hhHHhCCcCCCCeEEEEccccccccccccCCCCccccCCHHHHHHHHHHHHHhHHHHHH					

AA	190	200	210	220
Séquence	QDNFASNSAFITKGQISGSTPTIQFLPPPEHHHHHH			
PROF_sec	..LL.....L...LLL.....LL...LLL			
	0465333330220451267554000037863234578			
PSIPREP	CCCCCCHHHHHHCHHHCCCCCCCCCCCCCCCCCCCC			
	0188512287750101078987433789855355689			
Sable-2	CCCCCCHHHHCCCCCCCCCCECCCCCCCCCCCC			
	678876764544457678998754457888777777			
Consensus	ccCCcccchhhccccCCCCceecCCcCCCCcCC			

Prédiction F13G

AA	10	20	30	40	50	60
Séquence	MASTPNIAGPAITQEQMSSIKVDPTSNNLLPSQEQLKSVSTLMVAAKVPAA SVTTVALELV					
PROF_sec	LLLLLLLLLLLLLHHHH...LLLLL.LLLLLHHHHHHHHHHHHHH.LLLL..HHHHHHHHHH					
	98777777788898578742256755455588888999899898735888145678899998					
PSIPREP	CCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	998867889887989997358876658889999999999999999849888457999999999					
Sable-2	CCCCCCCCCCCCCHHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	7888878898888899987699876777889999999999999987999577999999999					
Consensus	CCCCCCCCCCCCCHHHHhCCcCCcCCCHHHHHHHHHHHHHhCCChhHHHHHHHHHH					

AA	70	80	90	100	110	120
Séquence	NFCYDNGSSAYTTVTGPPSSIPEISLAQLASIVKASGTS LRKFCRYFAPIIWNLR LTKMAP					
PROF_sec	HHH..LLL.....LLLLL..LHHHHHHHHH...L.HHHHHHHHHHHH.H...LLL.					
	875225675221223566666446788988886204536888898877665454206674					
PSIPREP	HHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCC					
	998278885444223235888841999999877618889998898889986688626898					
Sable-2	HHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCC					
	998688887767778898988877999999987899889999888999987578995					
Consensus	HHHCCcCCcCCCCCCCCCCCCcCHHHHHHHHHhCCCHHHHHHHHHHHHHHHhCCcCC					

AA	130	140	150	160	170	180
Séquence	ANWEASGYKPSAKFAAFDFFDGVENPAAMQPPSGLTRSPTQEERIANATNKQVHLFQAAA					
PROF_sec	.....LLL.....LL...LLLLL...LLHHHHHHH.....HHH...					
	012203788542022000101246742255655632468868877654334434555431					
PSIPREP	HHHHHCCCCCHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH					
	568966987102013332014656876468973465697989999999998895999975					
Sable-2	HHHHHCCCCCCEEEEECCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH					
	899985899655434444345567877788888778998899999999999875788875					
Consensus	hHHHHcCCcCCcEEEEeCCCCcCCcCCcCCcCCcCCCHHHHHHHHHHHHHhHHHHHh					

AA	190	200	210	220
Séquence	QDNNFASNSAFITKGQISGSTPTIQFLPPPEHHHHHHH			
PROF_sec	.LL.....L..LLL.....LL...LLL			
	0465333320220451166554011037864234578			
PSIPREP	HCCCCCHHHHHHHHHCCCCCCCCCCCCCCCCCCCC			
	0088512288751001078987555788865355689			
Sable-2	CCCCCCCCCHHHCCCCCCCCCCCCCECCCCCCCCCCCC			
	6788767644444576789987544578987777777			
Consensus	ccCCcccchhhcCCCCcCCcCCcCCCCcCC			

Prédiction F13L

AA	10	20	30	40	50	60
Séquence	MASTPNIALPAITQEQMSSIKVDPTSNNLLPSQEQLKSVSTLMVAAKVPAAASVTTVALELV					
PROF_sec	LLLLLLLLLLLLLHHHH...LLLLL.LLLLLHHHHHHHHHHHH.LLLL..HHHHHHHHHH					
	988776766888857774225676545568888999899898735888145678899998					
PSIPREP	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	988834556778989997438887748899999999999999984988458999999999					
Sable-2	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	678886788887899999769998777789989999999999987999577999999999					
Consensus	CCCCCCCCCCCCCHHHHhCccCccCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					
AA	70	80	90	100	110	120
Séquence	NFCYDNGSSAYTTVTGPSSIPEISLAQLASIVKASGTSLRKFCRYFAPIIWNLRDVKMAP					
PROF_sec	HHH...LLL.....LLLLL..LHHHHHHHH...L.HHHHHHHHHHH.H...LLL.					
	875236675221223566666446788988886204536888898888665454106674					
PSIPREP	HHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCC					
	998177885443222124888841999999887628889998898889987698636898					
Sable-2	HHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCCC					
	998588887767788898888766999999987689889999988999987578995					
Consensus	HHHccCCccccccccCCCCcCHHHHHHHHHhCCHHHHHHHHHHHHHHHhccCCc					
AA	130	140	150	160	170	180
Séquence	ANWEASGYKPSAKFAAFDFDFDGVENPAAMQPPSGLTRSPTQEERIANATNKQVHLFQAAA					
PROF_sec	.....LLL.....LL...LLLLL...LLHHHHHHH.....HHH...					
	013202788541121001100145732255655632478868887654333334555431					
PSIPREP	HHHHHCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHH					
	55886598710201333201566687646897346669898999999999995999975					
Sable-2	HHHHHCCCCCHHEEEHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHH					
	88998589965543444333556787778888877899889999999999987888875					
Consensus	hHHHhCccCchheeehccccCCcccCCccccCCHHHHHHHHHHHHHhHHHHHh					
AA	190	200	210	220		
Séquence	QDNNFASNSAFITKGQISGSTPTIQFLPPPEHHHHHH					
PROF_sec	..LL.....L..LLL.....LL...LLL					
	0465333430220451267554000037863234578					
PSIPREP	HCCCCCHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC					
	0188512177751101078987665888855355689					
Sable-2	CCCCCCCCCHHHCCCCCCCCCCCCCECCCCCCCCCCC					
	6788767644434576789987544578987777777					
Consensus	ccCCcccchhhccccCCCCceecCCccccccC					



Prédiction F13Y

AA	10	20	30	40	50	60
Séquence	MASTPNIAYPAITQEOMSSIKVDPTS NLLPSQEQLKSVSTLMVAAKVPAA SVTTVALELV					
PROF_sec	LLLLLLLLLLLLLHHHH...LLL...LLLLLHHHHHHHHHHHH.LLLL...HHHHHHHHHH					
	97777677788985776422567544558888999899898735888135678899998					
PSIPREP	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	998878888878989997438877767899999999999999984988568999999999					
Sable-2	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHH					
	78898788888678999876999877778998999999999977999677999999999					
Consensus	CCCCCCCCCCCCCHHHHHhCccCcccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					

AA	70	80	90	100	110	120
Séquence	NFCYDNGSSAYTTVTGPPSIP EISLAQLASIVKASGTS LRKFCRYFAPI IWNLR TDKMAP					
PROF_sec	HHH...LLL.....LLLLL...LHHHHHHHH...L.HHHHHHHHHHH.H...LLL.					
	875225675221223566665446788988886204536888898877665454205674					
PSIPREP	HHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCC					
	998378885434223235888841999999887618889998898889986698626898					
Sable-2	HHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHCCCC					
	98758888776777889898876799999999876898899999888999998778995					
Consensus	HHHhCccCccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhccCCc					

AA	130	140	150	160	170	180
Séquence	ANWEASGYKPSAKFAAFDFFDGVENPAA MQPPSGLTRSPTQEERIANATNKQVHLFQAAA					
PROF_sec	.....LLL.....LL...LLLLL...LLLHHHHHHH.....HHH...					
	01330278854202100110124674225565632478868887654334444566431					
PSIPREP	HHHHHCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHH					
	56886598710201333201576687534897445469798999999997894999975					
Sable-2	HHHHHCCCCCHHEEEEECCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHH					
	7899858996554444443355678777888887789988999999999987888875					
Consensus	hHHHhCccCccchheeeeccccCCcCccCccCccccCCHHHHHHHHHHHHHhHHHHHh					

AA	190	200	210	220
Séquence	QDNNFASNSAFITKGQISGSTPTIQFLPPPEHHHHHH			
PROF_sec	..LL.....L..LLL.....LLL...LLL			
	0465333320220451267554000037863234578			
PSIPREP	CCCCCCHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC			
	0188501287751100078987554789844355689			
Sable-2	CCCCCCHHHCCCCCCCCCCCCCECCCCCCCCCCCC			
	678876764443467778998754457898777777			
Consensus	ccCCcccchhhccccCCCCceecCCcCccccCC			

Prédiction E19K

AA	10	20	30	40	50	60
Séquence	MASTPNIAFFPAITQKQMSSIKVDPTSNNLLPSQEQLKSVSTLMVAAKVPAASVTTVALELV					
PROF_sec	LLLLLLLLLLLLLLLLHHHHHHHHLLLLLLLLLLLLLLLLHHHHHHHHHHHHHHHHHHLLLLLLLLHHHHHHHHHHHH					
	987776766888857774225675545568888999899898735888135678899998					
PSIPREP	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHH					
	9988236567678899973388877588999999999999999849885689999999999					
Sable-2	CCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHH					
	6888857888878999987699987777899899999999998799956789999999999					
Consensus	CCCCCCCCCCCCCHHHHHhCCCCcccCCCHHHHHHHHHHHHHHHhhCCCCchHHHHHHHHHHHHHHHH					

AA	70	80	90	100	110	120
Séquence	NFCYDNGSSAYTTVTGPPSSIPEISLAQLASIVKASGTS LRKFCRYFAPIIWNLRTDKMAP					
PROF_sec	HHHHLLLLLLLLLLLLLLLLLLLLLLLLLLLLHHHHHHHHHHHHHHHHHHLLLLLHHHHHHHHHHHHHHHHHHHHHHHHHH					
	875226665221223566665446788988886204536888898878665454206674					
PSIPREP	HHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHH					
	998278885433223235888852999999877628889998898889986698637898					
Sable-2	HHHCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHH					
	998588887667788898888665999999998768988999999889999988778994					
Consensus	HHHhccccccccccccCCCCcchHHHHHHHHHHHHcCCCHHHHHHHHHHHHHHHHHHHcCCCCc					

AA	130	140	150	160	170	180
Séquence	ANWEASGYKPSAKFAAFDFFDGVENPAAMQPPSGLTRSPTQEERIANATNKQVHLFQAAA					
PROF_sec	HHHHHLLLLLLLLLEEEHHHHHHLLLLLLLLLLLLLLLLLLLLLLLLLLLLHHHHHHHHHHHHHHHHHHHHHHHHHH					
	013302788542022000100146732255655632478868887654333334555431					
PSIPREP	HHHHHCCCCCHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHH					
	45897598710201333211465687646897346669798999999998895999975					
Sable-2	HHHHHCCCCCCCCEEEEEECCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHHHHHHHHHHHHHHHHHHH					
	8999858996553344443355678777888887789988999999999886788875					
Consensus	hhHHhCCCCccccccccccecccCCcCCcCCcCCCCCCHHHHHHHHHHHHHHHhhHHHHHH					

AA	190	200	210	220
Séquence	QDNMFASNSAFITKGQISGSTPTIQFLPPEHHHHHH			
PROF_sec	LLLLLLLLLLLLLEELLLLLLLLLLLLLLEELLLLLLLLLLLLL			
	046533330220451267554000037864234678			
PSIPREP	HCCCCCCHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCC			
	0188512288751002078987433789854355689			
Sable-2	CCCCCCCCHHHHCCCCCCCCCCCCCCCCCCCCCCCC			
	6788767645444676789988644688987777777			
Consensus	cCCccccchhhccccCCCCccccCCCcccccc			



Comparaison prédictions structures secondaires des différents mutants

AA	10	20	30	40	50	60
Séquence	MASTPNIAFPAITQEQMSSIKVDPTSNLLPSQEQLKSVSTLMVAAKVPAASVTTVALELV					
CP <sub>6-215</sub>	CCCCCCCCCCCCCHHHHHhCccCccccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					
CP <sub>13-215</sub>	CCCCCHHHHHhCccCccccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHHHH					
CP <sub>14-215</sub>	CCCCCHHHHHhCccCccccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHHHH					
F13A	CCCCCCCCCCCCCHHHHHhCccCccccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					
F13G	CCCCCCCCCCCCCHHHHHhCccCccccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					
F13L	CCCCCCCCCCCCCHHHHHhCccCcccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					
F13Y	CCCCCCCCCCCCCHHHHHhCccCccccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					
E19K	CCCCCCCCCCCCCHHHHHhCccCccccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					
E19P	CCCCCCCCCCCCCHHHHHhCccCccccCCCCCHHHHHHHHHHHHHhCccChHHHHHHHHHH					

AA	70	80	90	100	110	120
Séquence	NFCYDNGSSAYTTVTGPPSSIP EISLAQLASIVKASGTSLRKFCRYFAPIIWNLRD KMAP					
CP <sub>6-215</sub>	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHhHHHccCCc					
CP <sub>13-215</sub>	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhCCCc					
CP <sub>14-215</sub>	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhccCCc					
F13A	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhccCCc					
F13G	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhccCCc					
F13L	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhccCCc					
F13Y	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhccCCc					
E19K	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhCCCc					
E19P	HHHccCCCccccccccccccCCCCcCHHHHHHHHHhCccCHHHHHHHHHHHHHHHhCCCc					

AA	130	140	150	160	170	180
Séquence	ANWEASGYKPSAKFAAFDFFDGVENPAAMQPPSGLTRSPTQEERIANATNKQVHLFQAAA					
CP <sub>6-215</sub>	hHHHHcCCCcccheeehhccccCCCCcCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHH					
CP <sub>13-215</sub>	hhHHHcCCCccchheeeecccccCccccCCCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHh					
CP <sub>14-215</sub>	hhHHHcCCCccceeeeeecccccCccccCCCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHh					
F13A	hhHHHcCCCccceeeeeecccccCCCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHH					
F13G	hHHHHcCCCccceeeeeeccccCCCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHH					
F13L	hHHHHcCCCccchheeehccccCCCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHH					
F13Y	hHHHHcCCCccchheeeecccccCCCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHH					
E19K	hhHHHcCCCccceeeeeeccccCCCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHH					
E19P	hhHHHcCCCccceeeeeeccccCCCCcCCCCcCCCHHHHHHHHHHHHHhHHHHHH					

AA	190	200	210	220
Séquence	QDNNFASNSAFITKGQISGSTPTIQFLPPPEHHHHHH			
<b>CP<sub>6-215</sub></b>	ccCCCCcchhhhccccCCCCCCCCCCCCCCCCcC			
<b>CP<sub>13-215</sub></b>	ccCCCCcchhhhccccCCCCCCCCCCCCCCCCcC			
<b>CP<sub>14-215</sub></b>	ccCCCCcchhhhccccCCCCccecCCCCCCCCcC			
<b>F13A</b>	ccCCCCcchhhhccccCCCCccecCCCCCCCCcC			
<b>F13G</b>	ccCCCCcchhhhccccCCCCccecCCCCCCCCcC			
<b>F13L</b>	ccCCCCcchhhhccccCCCCccecCCCCCCCCcC			
<b>F13Y</b>	ccCCCCcchhhhccccCCCCccecCCCCCCCCcC			
<b>E19K</b>	ccCCCCcchhhhccccCCCCCCCCCCCCCCCCcC			
<b>E19P</b>	ccCCCCcchhhhccccCCCCCCCCCCCCCCCCcC			